


## RESEARCH ARTICLE

## Open Access



# Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data

Elisa Peripolli<sup>1</sup>, Christian Reimer<sup>2,3</sup>, Ngoc-Thuy Ha<sup>2,3</sup>, Johannes Geibel<sup>2,3</sup>, Marco Antonio Machado<sup>4,5</sup>, João Cláudio do Carmo Panetto<sup>5</sup>, Andréa Alves do Egito<sup>6</sup>, Fernando Baldi<sup>1</sup>, Henner Simianer<sup>2,3</sup> and Marcos Vinícius Gualberto Barbosa da Silva<sup>4,5\*</sup> 

## Abstract

**Background:** The cattle introduced by European conquerors during the Brazilian colonization period were exposed to a process of natural selection in different types of biomes throughout the country, leading to the development of locally adapted cattle breeds. In this study, whole-genome re-sequencing data from indicine and Brazilian locally adapted taurine cattle breeds were used to detect genomic regions under selective pressure. Within-population and cross-population statistics were combined separately in a single score using the de-correlated composite of multiple signals (DCMS) method. Putative sweep regions were revealed by assessing the top 1% of the empirical distribution generated by the DCMS statistics.

**Results:** A total of 33,328,447 biallelic SNPs with an average read depth of 12.4X passed the hard filtering process and were used to access putative sweep regions. Admixture has occurred in some locally adapted taurine populations due to the introgression of exotic breeds. The genomic inbreeding coefficient based on runs of homozygosity (ROH) concurred with the populations' historical background. Signatures of selection retrieved from the DCMS statistics provided a comprehensive set of putative candidate genes and revealed QTLs disclosing cattle production traits and adaptation to the challenging environments. Additionally, several candidate regions overlapped with previous regions under selection described in the literature for other cattle breeds.

**Conclusion:** The current study reported putative sweep regions that can provide important insights to better understand the selective forces shaping the genome of the indicine and Brazilian locally adapted taurine cattle breeds. Such regions likely harbor traces of natural selection pressures by which these populations have been exposed and may elucidate footprints for adaptation to the challenging climatic conditions.

**Keywords:** *Bos taurus indicus*, *Bos taurus taurus*, Signatures of selection, Local adaptation, Next-generation sequencing

\* Correspondence: [marcos.vb.silva@embrapa.br](mailto:marcos.vb.silva@embrapa.br)

<sup>4</sup>National Council for Scientific and Technological Development (CNPq), Lago Sul 71605-001, Brazil

<sup>5</sup>Embrapa Dairy Cattle, Juiz de Fora 36038-330, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The first cattle herds were brought to Brazil by Portuguese conquerors in 1534 during the Brazilian colonization period [1]. These cattle have undergone to a process of natural selection for more than 450 years in a wide range of ecosystems throughout the country [2]. Natural selection in a remarkably diverse set of environments together with recurring events of breed admixture led to the development of locally adapted cattle breeds, i.e. Curraleiro Pé-Duro, Pantaneiro, Crioulo Lageano, Caracu, and Mocho Nacional [3]. By the end of the nineteenth century, the increasing demand for food supply triggered the imports of exotic and more productive breeds of indicine origin [3, 4]. As a consequence, a reduction in locally adapted cattle breed populations has occurred to such an extent that nowadays, most of them are threatened with extinction [3, 5].

Brazilian locally adapted cattle breeds have been subjected to strong environmental pressures and faced several difficulties including hot, dry or humid tropical climate conditions, scarce food availability, diseases, and parasite infestations without any significant selective pressure imposed by man [2]. Influenced by the environment and shaped by natural selection, these animals acquired very particular traits to thrive in distinct ecosystems, which has presumably left detectable signatures of selection within their genomes. In this regard, Brazilian locally adapted cattle breeds represent an important genetic resource for the understanding of the role of natural selection in diverse environments, providing new insights into the genetic mechanisms inherent to adaptation and survivorship [6]. Although their productivity is much lower compared to highly-specialized breeds under intensive production systems [7, 8], great efforts have been made to improve our knowledge of locally adapted breeds [5, 9, 10] and their use in crossbred schemes.

According to Utsunomiya et al. [11], signatures of selection studies should strongly focus on small local breeds given their endangered status and the putative importance of their genomes in unraveling footprints of selection by elucidating genes and structural variants underlying phenotypic variation. Advances in molecular genetics and statistical methodologies together with the availability of whole-genome re-sequencing has notably improved the accuracy to disentangle the effects of natural and artificial selection in the genome of livestock [12–14]. However, despite the recent achievements in high-throughput sequencing, studies to detect positive selection in endangered Brazilian locally adapted cattle breeds are incipient. Previous studies on such breeds have mainly focused on population structure and genetic diversity using Random Amplified Polymorphic DNA (RAPD), pedigree data, microsatellite, and Single-Nucleotide Polymorphism (SNP) arrays [15–19].

In this study, we report for the first-time signatures of selection derived from whole-genome re-sequencing data in three Brazilian locally adapted taurine cattle breeds as well as in one indicine breed. Potential biological functions of the genes screened within the putative candidate regions were also examined to better elucidate the phenotypic variation related to adaptation shaped by natural selection.

## Results

### Data

DNA samples from 13 Gir (GIR), 12 Caracu Caldeano (CAR), 12 Crioulo Lageano (CRL), and 12 Pantaneiro (PAN) re-sequenced to 15X genome coverage were used. An average alignment rate of 99.59% was obtained. After SNP calling and filtering, a total of 33,328,447 SNPs distributed across all 29 autosomes were retained for subsequent analyses with an average read depth of 12.37X (9.57 ~ 17.52X).

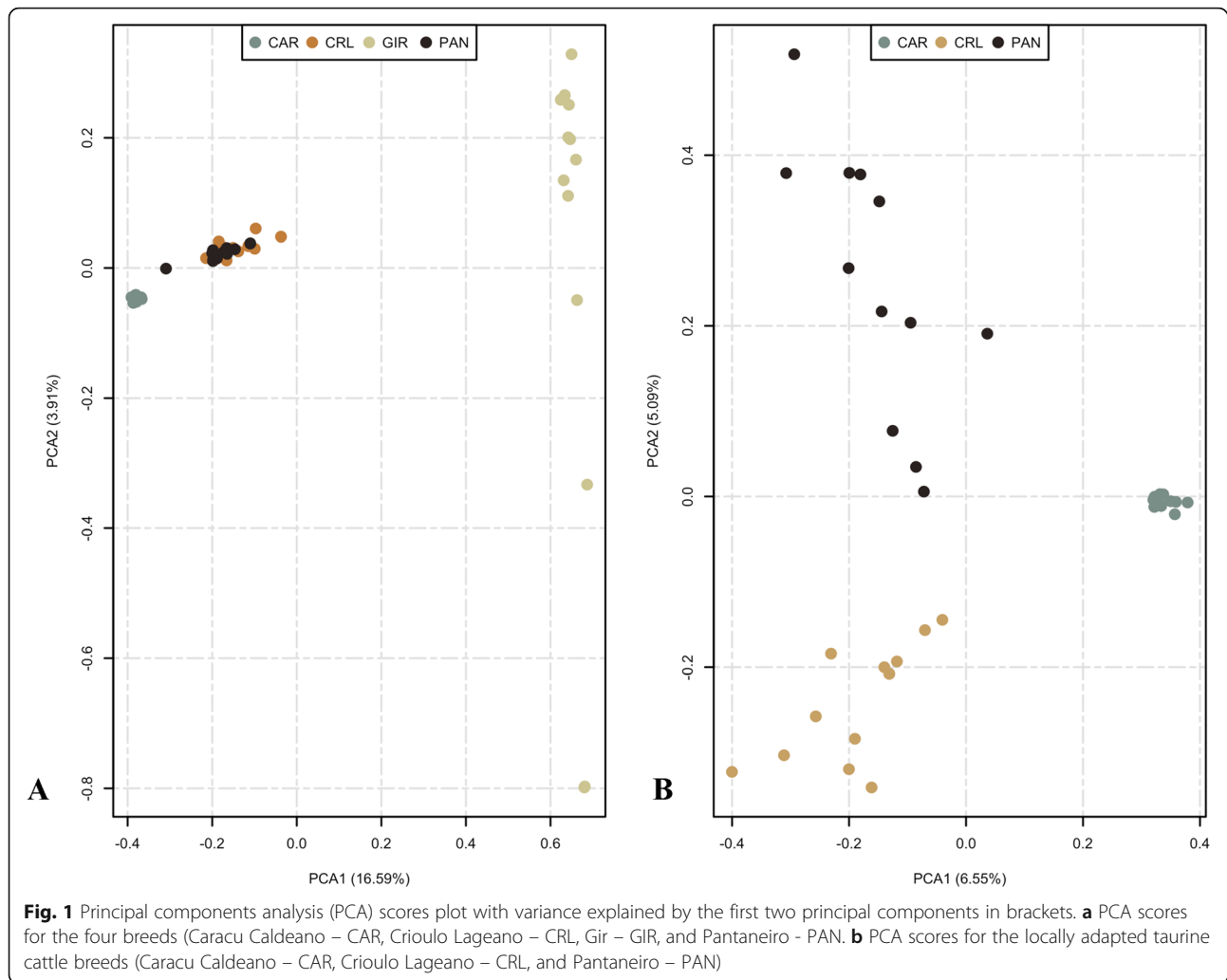
### Variant annotation and enrichment

Of the total SNPs identified ( $n = 33,328,447$  SNPs), most of them were located in intergenic (67.17%) and intronic (25.85%) regions (Additional file 1). A total of 1,065,515 (3.19%) variants were located in the 5-kb regions upstream from genes, and 928,061 (2.78%) in the 5-kb regions downstream from genes. Several variants with high consequence on protein sequence were identified, including splice acceptor variant ( $n = 471$ ), splice donor variant ( $n = 481$ ), stop gained ( $n = 1111$ ), stop lost ( $n = 58$ ), and start lost ( $n = 208$ ). According to SIFT scores, 24,159 variants (23,428 missense, 578 splice region, and 143 start lost) were classified as deleterious.

Following variant annotation, we further investigated the gene content within the predicted variants to cause relevant biological functions. A total of 1189 genes were described within variants with high consequence on protein sequence and 7373 genes within those causing a deleterious mutation based on the SIFT score. Functional enrichment analysis revealed several gene ontology (GO) terms and one Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway overrepresented ( $p < 0.01$ ) for the set of genes previously described (Additional files 2 and 3), however, none of them have been associated with the traits/phenotypes that could be affected by the natural selection which those breeds have been subjected to.

### Population structure

The population structure among breeds was dissected by analyzing the first two principal components, which accounted for roughly 20% of the genetic variability and divided the populations into three clusters (Fig. 1a). A clear separation could be observed between indicine



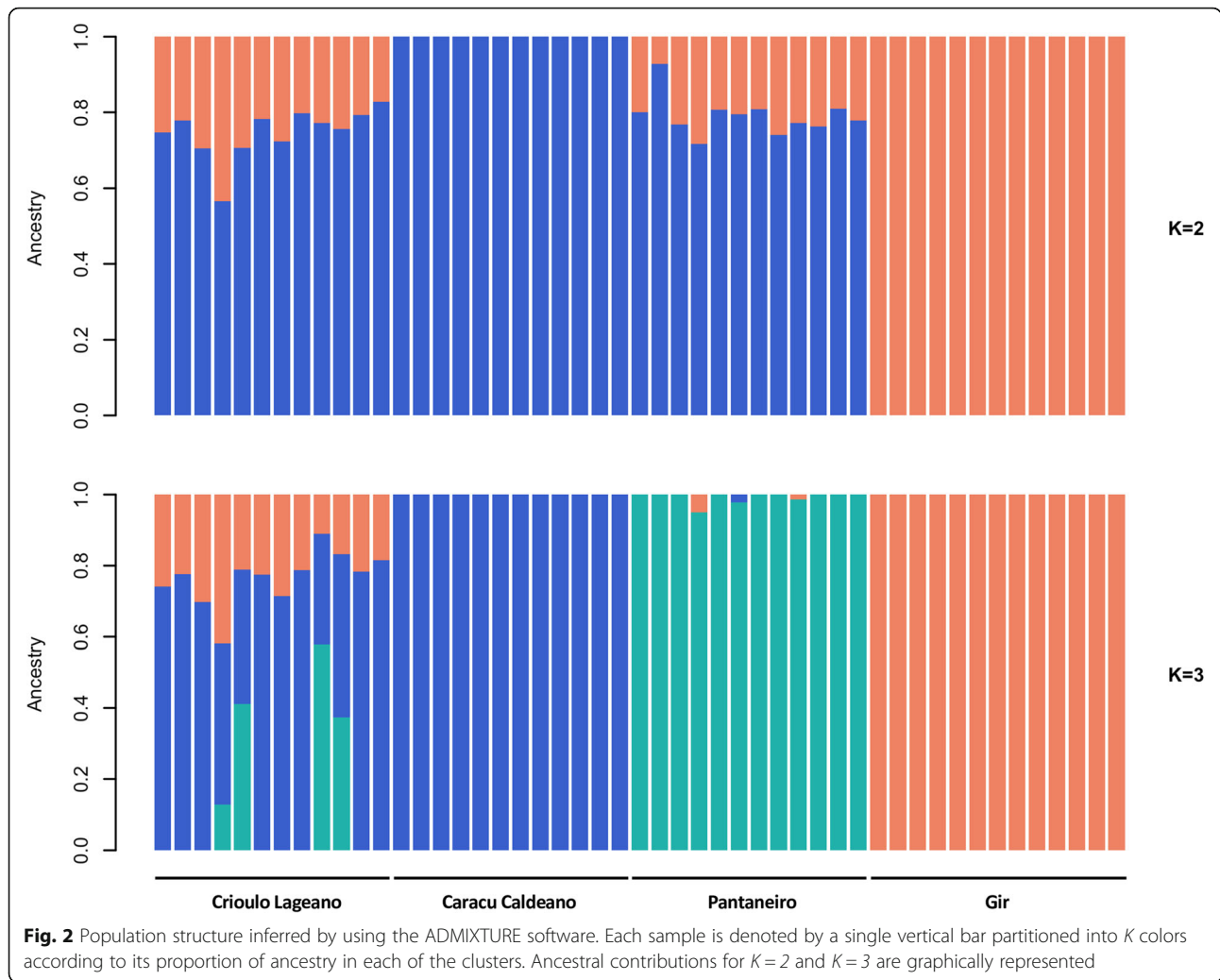
(*Bos taurus indicus*) and locally adapted taurine (*Bos taurus taurus*) populations. Within the taurine populations, the greatest overlap of genetic variation was observed between CRL and PAN breeds. Despite clustering together, the analysis of molecular variance (AMOVA) revealed genetic differentiation between those two breeds ( $p < 0.001$ , Additional file 4), indicating that all four breeds could be considered as genetically independent entities. Further, when analyzing the first two principal components encompassing the locally adapted taurine cattle breeds (Fig. 1b), an evident separation could be observed between CAR and the remaining two populations. The analysis also distinguished CRL from PAN, agreeing with the AMOVA results.

Admixture analysis was performed to further estimate the proportions of ancestry ( $K$ ) in each population (Fig. 2). The lowest cross-validation error (0.387) was observed for  $K = 2$ , revealing the presence of two main clusters differentiating the locally adapted taurine populations from the indicine population. Within the taurine

populations, the CAR breed did not show admixed ancestry while CRL and PAN breeds showed 77% of taurine and 23% of indicine ancestry on average. When  $K = 3$  was assumed, CRL samples revealed evidence of admixed ancestry from other breeds, whereas PAN samples were quite homogeneous, with little indication of introgression from other breeds. CAR and GIR breeds displayed a greater uniformity and did not reveal major signs of admixture of other breeds, being consistent with  $K = 2$ .

#### Genomic inbreeding

Descriptive statistics for runs of homozygosity-based inbreeding coefficients ( $F_{ROH}$ ) are shown in Table 1. The average inbreeding coefficients did not differ significantly ( $p < 0.05$ ) among breeds, with the exception of CAR animals. It is worth to highlight that these animals also displayed the smallest inbreeding variability among all breeds, supported by the lowest coefficient of variation.



**Selective sweeps**

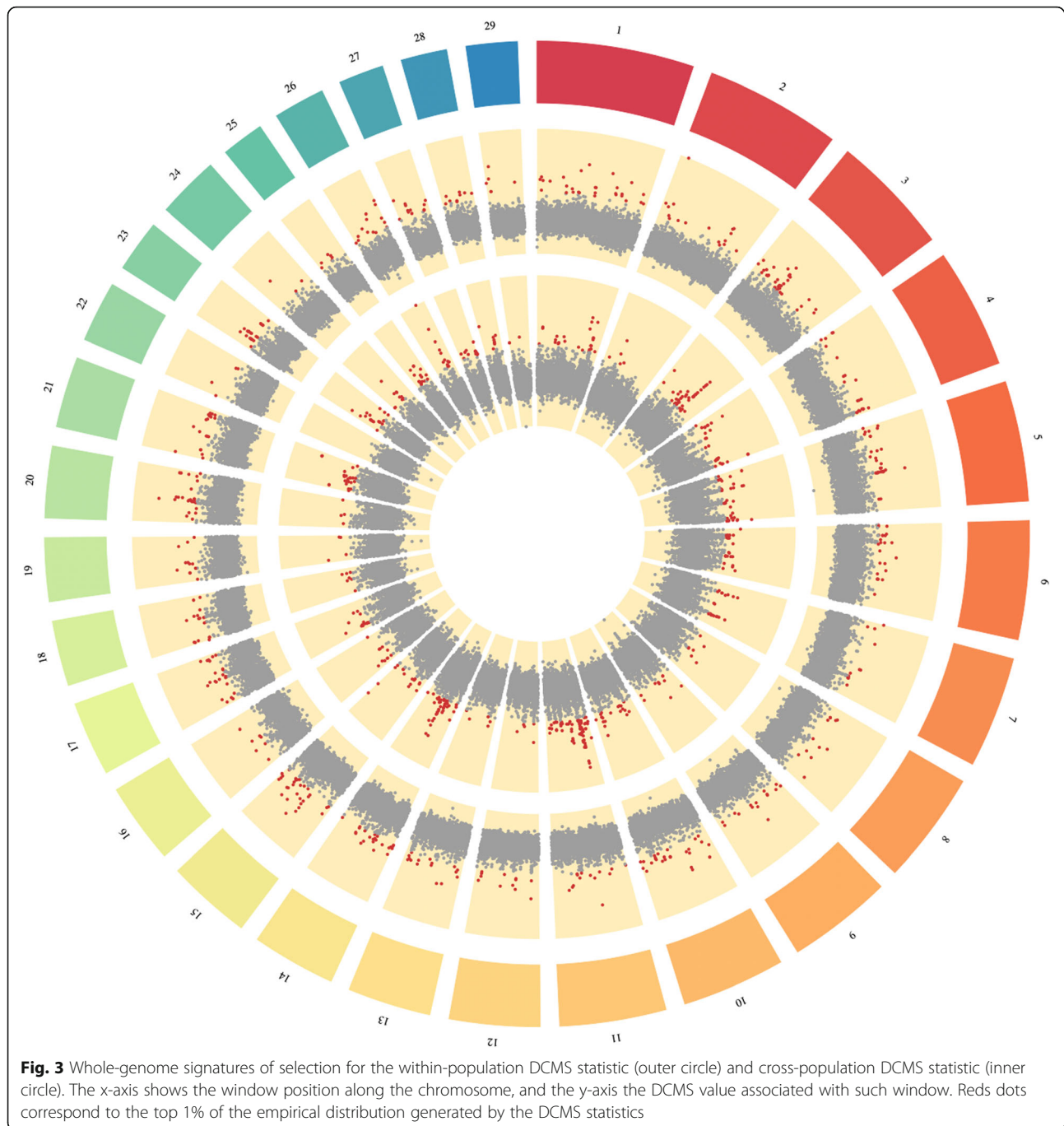
A total of 499 putative sweep regions encompassing 221 genes were identified from the top 1% of the empirical distribution generated by the within-population decorrelated composite of multiple signals (DCMS) statistic [20] (Fig. 3, Additional file 5). For the cross-population DCMS statistic, the top 1% of the empirical distribution revealed 503 putative sweep regions comprehending 242 genes (Additional file 6). The *Bos taurus* autosome (BTA)

3 displayed the highest number of putative sweep regions for the within-population DCMS statistic ( $n = 33$ ), while BTA11 did for the cross-population DCMS statistic ( $n = 67$ ). The functional importance of the annotated genes was assessed by performing GO and KEGG pathway enrichment analysis separately for each DCMS statistic and its respective retrieved gene list. No overall significant enrichment of any particular GO nor KEGG was found after adjusting the  $p$ -values for False Discovery Rate [21].

**Table 1** Descriptive statistics of runs of homozygosity-based inbreeding coefficient ( $F_{ROH}$ ) for Gir (GIR), Crioulo Lageano (CRL), Caracu Caldeano (CAR), and Pantaneiro (PAN) cattle breeds

Breed	Mean	Median	Minimum	Maximum	Coefficient of variation (%)
Gir	0.040 <sup>b</sup>	0.038	0.020	0.060	29.37
Crioulo Lageano	0.036 <sup>b</sup>	0.028	0.017	0.082	53.69
Caracu Caldeano	0.138 <sup>a</sup>	0.140	0.121	0.153	8.63
Pantaneiro	0.045 <sup>b</sup>	0.042	0.022	0.096	43.56

Means sharing a common letter within a column were not significantly different ( $p < 0.05$ ) from one another



Five genomic regions overlapped between the candidate sweep regions of the within-population and cross-population DCMS statistics (BTA4:101600000–101,650,000, BTA5:3700000–3,750,000, BTA9:98650000–98,700,000, BTA11:22300000–22,350,000, and BTA11:53900000–53,950,000). When inspecting in detail, the region on BTA4:101600000–101,650,000 harbored two quantitative trait locus (QTL) with functions related to the bovine respiratory disease [22] and body condition

score [23]. The remaining four regions have not been associated with any QTL in cattle so far, however, they were found to be in close vicinity (~ 15 to 237 kb) with specific QTLs for beef cattle production traits. Such QTLs included body weight at yearling, calving ease, body weight gain, and marbling score [24–26]. Further, among the five overlapping candidates sweep regions, only the one on BTA9 was found to harbor a gene, the *PRKN*.

### Selective sweeps and runs of homozygosity

Shared genomic regions harboring several protein-coding genes were identified between runs of homozygosity (ROH) hotspots and the putative sweep regions retrieved from the DCMS statistics (Table 2). ROH hotspots for each breed are described in Additional file 7. For the shared regions disclosed when considering the within-population DCMS statistic, the ones located on BTA1:8300000–8,350,000 and BTA1:41600000–41,650,000 coincided with a QTL for somatic cell score [27] and maturity rate [28], respectively. It is noteworthy to underscore that despite not displaying any overlapping QTL, the region on BTA8:15700224–15,700,228 was described nearby (~ 99 kb) a QTL for tick resistance [29], and those on BTA21:6550000–6,600,000 and BTA21:

63250000–63,300,000 were very close (< 14 kb) to QTLs for reproductive-related traits [30, 31]. When considering the cross-population DCMS statistic, the candidate regions overlapped previously identified QTLs formerly implicated in dairy-related [35–37, 39] and body-related (weight [24], energy content [34], and conformation [35]) traits. Further, several QTLs associated with body conformation and growth [23, 24, 33], reproductive-related traits [28, 32], and coat texture [38] were described to be in very close proximity (~ 18.98 to 88.38 kb).

### Overlap with candidate regions under positive selection in other cattle populations

Several putative sweep regions identified from the top 1% of the empirical distribution generated by the within-

**Table 2** Gene annotation and reported QTLs for the shared genomic regions between runs of homozygosity (ROH) hotspots and the putative sweep regions retrieved from the within-population and cross-populations DCMS statistics

BTA <sup>1</sup>	Start	End	Genes	QTL <sup>2</sup>
<i>Within-population DCMS statistic x ROH</i>				
1	8,300,000	8,350,000	–	<b>Somatic cell score [27]</b>
1	41,600,000	41,650,000	<i>EPHA6, ARL6</i>	<b>Maturity rate [28]</b>
1	112,250,000	112,300,000	<i>KCNAB1</i>	–
8	15,800,000	15,850,000	–	Tick resistance [29]
15	35,365,655	35,399,999	<i>OTOG</i>	–
15	35,400,001	35,450,000	–	–
18	34,718,675	34,750,000	<i>CDH16, RRAD</i>	–
21	6,550,000	6,600,000	<i>ADAMTS17</i>	Calving ease [30]
21	63,250,000	63,300,000	<i>VRK1</i>	Interval to first estrus after calving [31]
<i>Cross-population DCMS statistic x ROH</i>				
3	77,250,000	77,300,000	–	Body condition score [23]
5	31,800,000	31,850,000	–	Body weight (yearling) [24], Conception rate [32]
5	38,761,637	38,761,745	<i>YAF2</i>	–
7	57,050,000	57,100,000	–	Rump angle [33]
11	67,450,000	67,500,000	<i>ANTXR1, GFPT1</i>	<b>Body weight (yearling) [24], Body energy content [34]</b>
11	67,700,000	67,749,999	–	–
11	67,750,001	67,800,000	<i>NFU1</i>	–
11	68,550,000	68,600,000	<i>PCYOX1</i>	–
14	52,900,000	52,914,848	–	Maturity rate [28]
15	10,150,000	10,200,000	–	–
15	10,900,000	10,950,000	–	<b>Calving ease (maternal) [35], Daughter pregnancy rate [35], Foot angle [35], Milk fat percentage [35], Milk fat yield [35], Net merit [35], Length of productive life [35], Milk protein percentage [35], Milk protein yield [35], Calving ease [35], Somatic cell score [35]</b>
20	38,000,000	38,050,000	<i>RANBP3L, NADK2</i>	<b>Milk protein percentage [36], Milk protein yield [37], Milk yield [37], Coat texture [38]</b>
21	200,000	250,000	–	–
25	1,345,564	1,350,000	<i>NME3, MRPS34</i>	<b>Milk fat yield [39]</b>

<sup>1</sup> BTA: *Bos taurus* autosome; <sup>2</sup> QTLs within the candidate genomic regions are highlighted in bold. Non-bold QTLs were the closest and most suitable candidate QTL for the given candidate region



population and cross-population DCMS statistics were in agreement with previous research on signatures of selection in cattle (Additional files 8 and 9, respectively). Such studies included indigenous African and Spanish [6, 40–43], native [44–46], tropical-adapted [6, 47–49], Chinese [49, 50], and commercial beef and dairy [13, 41, 49, 51–54] cattle breeds. For the five genomic regions identified overlapping in between the DCMS statistics, the one on BTA9:98650000–98,700,000 matched with a previous study on cattle breeds selected for dairy production [54]. Besides, common signals found between ROH hotspots and the within-population and cross-population DCMS statistics were also supported by previously published data on signatures of selection [6, 41, 43, 44, 46, 50, 53] (Additional files 10 and 11, respectively).

## Discussion

### Population structure

The segregation between indicine and taurine cattle populations described in both principal component and admixture analysis ( $K=2$ ) reflects the divergence and evolutionary process started roughly two million years ago [55, 56]. As a result of the domestication process and selective breeding over time, the cattle can be classified into temperate (*Bos taurus taurus* or taurine) and tropical (*Bos taurus indicus* or indicine) based on the common adaptive and evolutionary traits they have acquired [57]. Within the Brazilian locally adapted taurine breeds, the principal component analysis (PCA) indicates the highest relatedness between CRL and PAN breeds and their divergence from the CAR breed may be explained by the European cattle type introduced in Brazil during the colonization period [58]. These results were similar to those obtained using RAPD [17] and microsatellites [19]. Portuguese purebred cattle brought to Brazil belonged to three different bloodlines: *Bos taurus aquitanicus*, *Bos taurus batavicus*, and *Bos taurus ibericus*. In this regard, CRL and PAN breeds descended from a common ancestral pool and have their origin in breeds from *Bos taurus ibericus* cattle, while the CAR cattle is derived from the *Bos taurus aquitanicus* cattle [17]. Further, the divergence within the locally adapted cattle breeds may be a result of artificial selection events over time since the CAR cattle have been selected for milk production for the past 100 years, while CRL and PAN started recently to be artificially selected.

Levels of introgression of indicine genes in taurine breeds described herein are consistent with previous studies on Brazilian locally adapted taurine breeds [16, 17, 19]. This gene flow reinforces the concept that the import of exotic breeds at the beginning of the twentieth century [3] led to the miscegenation of the locally adapted breeds due to crossbreeding practices, resulting nearly in their extinction [4]. In this

regard, the CRL breed experienced some introduction of Nellore (*Bos taurus indicus*) genes for a short period in the eighties [17], which can be visualized when assuming  $K=2$  and  $K=3$ . Concurring with our findings, Egito et al. [19] also revealed that CRL and PAN animals were the closest to the indicine cattle among four Brazilian locally adapted cattle breeds, displaying the highest frequency of indicine gene introgression. A cytogenetic analysis study on the PAN cattle also revealed absorbing crosses with the indicine cattle [59]. In addition, the absence of admixture patterns in CAR individuals has been previously described by Campos et al. [16] and Egito et al. [21]. The homogeneity of such population most likely reflects its formation process and the objective of selection for dairy traits since 1893 [60], which may have distinguished them from other locally adapted taurine breeds when taking into consideration the genetic structure integrity.

### Genomic inbreeding

As already stated, the Brazilian locally adapted cattle breeds nearly disappeared between the late 19th and beginning of the twentieth century, and most of them are nowadays threatened with extinction [3, 5]. It is worth to stress out that the CAR cattle are an exception, and they can be considered as an established breed [5, 61]. In this regard, animals comprising our dual purpose cattle populations, which were exploited for meat production in former times [62], are nowadays mainly used in animal genetic resources conservation programs (in situ and ex situ) and as a germplasm reservoir to preserve the genetic variability [4, 63]. Different from the dual-purpose cattle populations, the dairy populations are no longer considered endangered, and such animals have been selected for milk production traits in the southeastern region of Brazil since 1893 (CAR, [60]) and the early nineties (GIR, [64]).

Most of the locally adapted cattle breeds in Brazil developed from a narrow genetic base, and in such cases, inbreeding can increase over generations and reduce genetic variability [65]. Despite their population background, CRL and PAN animals displayed low  $F_{ROH}$  estimates, concurring with heterozygosity estimates (Results not shown). Decreased levels of inbreeding and high genetic variability have been previously described for both breeds, probably resulting from a slight selection pressure and herd management focused on maintaining genetic diversity by using a male:female relationship larger than usual [19]. Egito et al. [15] attributed such results to the formation of new PAN herds from 2009 onwards while Pezzini et al. [18] associated it with the diversification in the use of CRL sires. Further, Egito et al. [19] stated that CRL and PAN cattle were the most

diverse population with the highest mean allelic richness among four locally adapted cattle breeds investigated. Such results are consistent with  $F_{ROH}$  estimates found in this current work, reflecting mild selection pressure in our dual-purpose cattle populations together with rationale mating decisions and herd management taken by the breeders and associations.

The highest  $F_{ROH}$  found for the CAR population most likely reflects its history of selective breeding for milk-related traits from a limited genetic base and the occurrence of a population decrease in the sixties, as discussed by Egito et al. [19]. According to Marras et al. [66], it is not unusual to disclose a higher sum of ROH in dairy than in beef populations. In this regard, the reduction of genetic variability through the increase of autozygosity in dairy breeds can be explained by the intense artificial selection with the use of a relatively small number of proven sires [67]. Despite being also specialized for milk-related traits, it is not surprising that the GIR population did not show as high  $F_{ROH}$  levels as did CAR. Previous studies have also shown low inbreeding rates for the GIR cattle considering pedigree-based inbreeding coefficient [68, 69] and  $F_{ROH}$  [70, 71]. A trend in the decrease of inbreeding has been previously described [68, 70], and it happens along with the establishment of the Brazilian Dairy Gir Breeding Program (PNMGL) and the Gir progeny testing. Presumptively, these two concomitant events led to the dissemination of the breed, allowing formerly closed herds to start using semen of proven sires, increasing the overall genetic exchange and reducing the average inbreeding over time.

#### Candidate regions under positive selection

After combining the top 1% putative sweep regions retrieved from the within-population and cross-population DCMS statistics, five candidate regions harboring two QTLs and only one protein-coding gene were identified. Such results allowed us to highlight the body condition score QTL [23] on BTA4:101600000–101,650,000, which can be defined as the amount of metabolized energy stored in fat and muscle of a live animal [72]. During periods of energy shortage, key hormones expression and tissue responsiveness adjust to increase lipolysis to meet energy requirements and maintain physiological equilibrium [73, 74]. Regulation and coordination of energy partitioning and homeostasis is a challenge to sustainable intensification of cattle productivity in the tropics. The variation in the animal's nutritional and energetic balance may explain the observed variability in performance between animals in different environments [75]. Negative energy balance most likely reduce energy expenditure, impairing reproductive performance [76], and increasing the susceptibility to infections [77]. As formerly described, the Brazilian locally adapted cattle

breeds faced several environmental pressures to thrive in the tropics under harsh environmental conditions, suggesting that animals that were able to minimize the mobilization of adipose tissue reserves in response to the energy deficit might have conferred fitness advantage than the average individual in the given population.

The *PRKN* (also known as *PARK2*) was the only annotated gene identified in between the DCMS statistics, and its functions have been associated with adipose metabolism and adipogenesis [78]. Remarkably, it is considered a strong positional candidate for adiposity regulation in chicken [79].

We also explored common signals between ROH hotspots and the top 1% putative sweep regions retrieved from both DCMS statistics to increase the power of signals. Among the genes identified when considering the within-population DCMS statistic, we revealed the presence of two interesting genes that have been described to have effects on temperament (*EPHA6*) [80] and body size (*ADAMTS17*) [81] in cattle. Further, one gene associated with temperament (*ANTXR1*) [82] was also highlighted when considering the cross-population DCMS statistic.

In tropical and subtropical regions, cattle productivity depends not only on the inherent ability of animals to grow and reproduce but also on their ability to overcome environmental stressors that impact several aspects of cattle production [83]. In cattle, stress responsiveness has been associated with cattle behavior, more specifically, temperament. Temperament can adversely affect key physiological processes involved in cattle growth, reproduction, and immune functions [84]. Studies have shown that non-temperamental cattle tend to gain weight faster [85–87], spend more time eating [87], and have a higher dry matter intake and average daily gain [85, 88] than temperamental cattle. Further, studies have discussed the negative impacts of temperamental animals on immune-related functions (reviewed by [84]). Two reasons might explain those genes associated with temperament located on ROH hotspots overlapping regions on BTA1:41600000–41,650,000 and BTA11:67450000–67,500,000. The first reason is that such genes likely reflect levels of introgression of indicine genes in locally adapted taurine cattle breeds, as confirmed by admixture analysis. *Bos taurus indicus* and their crosses have been reported to be more temperamental than *Bos taurus taurus* cattle when reared under similar conditions [89]. The second reason is that the locally adapted taurine cattle breeds were able to overcome environmental stressors through natural selection over time and could prosper in such harsh tropical environment.

The *ADAMTS17* gene, described enclosing a ROH hotspot overlapping region on BTA21:6550000–6,600,000, is a well-known candidate gene with a major impact on body size [81, 90, 91]. Much has been discussed about the relationship between body size and environmental



adaptation. Variations in body size may be explained as an adaptive response to climate and/or can be driven by changes in feed resources and seasonal influences [92, 93]. In this regard, large body size animals can better tolerate austere conditions, having advantages under cold stress as well as in the use of abundant forage resources [94]. On the other hand, smaller animals exhibit better adaptation to warmer and dry climates [95–97] and are more efficient for grazing under seasonal and scarce forage resources [98]. Based on morphological measurements, it should be noted that the indicine and Brazilian locally adapted taurine cattle breeds are small to medium-sized breeds. Both GIR, CRL, and PAN have reduced body size and light-weight, in which females exhibit an average adult live weight of 418 kg [99], 430 kg [100], and 298 kg [101], respectively. CAR animals have a greater body size among the locally adapted cattle breeds, with females displaying an average live weight of 650 kg [102].

Two intersecting QTLs associated with productivity traits usually favored in commercial breeds (somatic cell score and maturity rate QTLs) were found in ROH hotspots overlapping regions when considering the within-population DCMS statistic. Among the QTLs identified when considering the cross-population DCMS statistic, the one associated with body energy content [34] must be highlighted given its importance in energy partitioning and homeostasis, as previously discussed. Additionally, several remarkably QTLs neighboring the candidate regions intervals were identified. These QTLs have been associated with different biological functions linked to local environment adaptation, such as parasite vector resistance (tick resistance QTL), reproductive-related traits (calving ease, interval to first estrus after calving, conception and maturity rate QTLs), body conformation and morphology traits (body condition score, body weight at yearling, rump angle QTLs), and coat color (coat texture QTL).

The genes and QTLs identified within the candidate regions provide a hint about the selective forces shaping the genome of the indicine and Brazilian locally adapted taurine cattle breeds. Such selective forces were described to be likely associated with adaptation to a challenging environment and environmental stressors. Further, several QTLs identified nearby the candidate regions intervals were also associated to a lesser extent with beef cattle production traits, while others with various biological functions presumably linked to selection to environmental resilience as well.

#### **Overlap with candidate regions under positive selection in other cattle populations**

The greatest number of the putative sweep regions identified from the top 1% of the within-population DCMS statistic overlapped with candidate regions under positive selection previously reported in five cattle breeds

selected for dairy production [54], comprehending roughly 22% ( $n = 52$ ) of the overlapping regions. For the top 1% of the cross-population DCMS statistic, the greatest number was described for native cattle breeds from Siberia, eastern and northern Europe [46], totaling nearly 17% ( $n = 50$ ) of the overlapping regions. Remarkably, in both statistics, the majority of the shared signals within those reported in the literature was found associated with specialized cattle breeds (i.e. dairy and beef). We also identified signatures of selection within those reported in the literature shared by breeds showing different production selection within the same candidate region. According to Gutiérrez-Gil et al. [103], such genomic regions may reflect selection for general traits such as metabolic homeostasis, or they might disclose the pleiotropic effects of genes on relevant traits underlying specialized cattle breeds.

The greater number (seven out of 11) of the putative sweep regions shared between ROH hotspots and the top 1% putative sweep regions retrieved from both DCMS statistics overlapped with regions previously described on local and native cattle breeds [41, 43, 44, 46]. Such results allow us to assume that the same selective forces are most likely acting across these populations, and such regions might have been shaped by selection events rather than genetic drift or admixture events.

It is noteworthy to underscore that the regions under positive selection for other cattle populations reported herein were mainly obtained through medium and high-density SNP arrays. SNP genotyping arrays suffer from SNP ascertainment bias, and it strongly influences population genetic inferences (reviewed by Lachance and Tishkoff [104]). Besides, some scan methodologies based on site frequency spectrum and population differentiation may be more likely to ascertainment bias than others [105, 106], compromising the power of the tests and may yielding to flawed results [107] when compared to those obtained from whole-genome re-sequencing data.

#### **Conclusions**

By using whole-genome re-sequencing data, we identified candidate sweep regions in indicine and Brazilian locally adapted taurine cattle breeds, of which the latter have been exposed to a process of natural selection for several generations in extremely variable environments. The signatures of selection across the genome could provide important insights for the understanding of the adaptive process and the differences in the breeding history underlying such breeds. Our findings suggest that admixture has occurred in some locally adapted taurine populations due to the introgression of exotic breeds, and the stratification results revealed the genetic structure integrity of the dairy populations sampled in this study. Candidate sweep regions, most of which overlapped with or were

nearby reported QTLs and candidate genes closely linked to cattle production traits and environmental adaptation. Putative sweep regions together with ROH hotspots also provided valuable shreds of evidence of footprints for adaptation to the challenging climatic conditions faced by the breeds. The candidate sweeps regions and the gene list retrieved from them can improve our understanding of the biological mechanisms underlying important phenotypic variation related to adaptation to hostile environments and selective pressures events to which these breeds have undergone. Furthermore, the study provides complementary information which could be used in the implementation of breeding programs for the conservation of such breeds.

## Methods

### Samples, sequencing, and raw data preparation

Sequencing analysis was based on data from 13 Gir (*Bos taurus indicus*, dairy production use), 12 Caracu Caldeano (*Bos taurus taurus*, dairy production use), 12 Crioulo Lageano (*Bos taurus taurus*, dual purpose use), and 12 Pantaneiro (*Bos taurus taurus*, dual purpose use) animals. The studied breeds can be classified into two groups: (i) indicine breeds represented by the Gir (GIR) cattle; and (ii) locally adapted taurine cattle breeds encompassing Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle. Animals were sampled from three Brazilian geographical regions, including the south (CRL), southeast (GIR and CAR), and mid-west (PAN) (Additional file 12).

DNA was extracted from semen samples that were collected from GIR bulls and blood samples from the remaining breeds. The semen straws were acquired from three commercial artificial insemination centers (American Breeders Service (ABS), Cooperatie Rundvee Verbetting (CRV), and Alta Genetics) and the DNA samples from the Animal Genetics Laboratory (AGL) at EMBRAPA Genetic Resources and Biotechnology (Cenargen, Brasília-DF, Brazil). Paired-end whole-genome resequencing with  $2 \times 100$  bp reads (CRL) and  $2 \times 125$  bp reads (GIR, CAR, and PAN) was performed on the Illumina HiSeq2500 platform with an aimed average sequencing depth of 15X.

Pair-end reads were aligned to the *Bos taurus taurus* genome assembly UMD 3.1 using Burrows-Wheeler Alignment MEM (BWA-MEM) tool v.0.7.17 [108] and converted into a binary format using SAMtools v.1.8 [109]. Polymerase chain reaction (PCR) duplicates were marked using Picard tools (<http://picard.sourceforge.net>, v.2.18.2). For downstream processing, GATK v.4.0.10.1 [110–112] software was used. Base quality score recalibration was performed using a SNP database (dbSNP Build 150) retrieved from the NCBI [113] followed by SNP calling using the HaplotypeCaller algorithm. To

remove unreliable SNP calls and reduce the false discovery rate, hard filtering steps were applied on the variant call. Insertions and deletions polymorphism (Indels) and multi-allelic SNPs were filtered out, and then hard filtering was applied for clustered SNPs ( $> 5$  SNPs) in a window size of 20 bp. An outlier approach was used and values above 14.44 (highest 5%) for Fisher strand test were removed. The same was applied for the highest and lowest 2.5% values for base quality rank sum test ( $-2.26$  and  $3.04$ ), mapping quality rank sum test ( $-2.46$  and  $1.58$ ), read position rank sum test ( $-1.64$  and  $2.18$ ), and read depth (267 and 883). Variants with a mapping quality value lower than 30 (0.1% error probability) were also removed from the call set. SNPs that passed the filtering process and located on autosomal chromosomes were retained for subsequent analysis.

### Variant annotation and predicted functional impacts

A functional annotation analysis of the called variants was performed to assess their possible biological impact using the Variant Effect Predictor (VEP, [114]) together with the Ensembl cow gene set 94 release. Variants are categorized according to their consequence impact on protein sequence as high, moderate, low, or modifier (more severe to less severe). Variants with 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. The impact of amino acid substitutions on protein function were predicted using the sorting intolerant from tolerant (SIFT) scores implemented on VEP tool, and variants with SIFT scores lower than 0.05 were considered as deleterious to protein function.

Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool [115, 116] was used to identify overrepresented GO terms and KEGG pathways using the list of genes retrieved from the variants classified with high consequence on protein sequence and as deleterious, and the *Bos taurus taurus* annotation file as a background. The  $p$ -values were adjusted by False Discovery Rate [21], and significant terms and pathways were considered when  $p < 0.01$ .

### Population differentiation analysis

A PCA implemented with a custom R script was used to examine the genetic structure of the four breeds. AMOVA [117] was also implemented to test for genetic differentiation among breeds. Such method consists in assessing population differentiation using molecular markers together with a pairwise distance matrix, and it can easily incorporate additional hierarchical levels of population structure. AMOVA computations were conducted using the 'amova' function in R package pegas

[118]. The analyses were based on pairwise squared Euclidean distances using the ‘dist’ function implemented in R [119] and the statistical significances were tested by permutations ( $n = 1000$ ). Additionally, the software ADMIXTURE v1.3 [120] was used to reveal admixture patterns among breeds by measuring the proportion of individual ancestry from different numbers of hypothetical ancestral populations ( $K$ ). Linkage disequilibrium (LD) pruning for admixture analysis was performed on PLINK v1.90 software [121] to remove SNP with a  $R^2$  value greater than 0.1 with any other SNP within a 50-SNP sliding window. The optimal number of  $K$  was defined based on the cross-validation error value ( $K = 1$  to 5) implemented in ADMIXTURE.

### Genomic inbreeding coefficient estimation

Genomic inbreeding coefficients based on runs of homozygosity ( $F_{ROH}$ ) were estimated for every animal according to the genome autozygotic proportion described by McQuillan et al. [122]:

$$F_{ROH}^i = \frac{S_{ROH}^i}{L_{GEN}}$$

where  $S_{ROH}^i$  is the sum of ROH across the genome for the  $i^{th}$  animals and  $L_{GEN}$  is the total length of the autosomes covered by SNPs.  $L_{GEN}$  was taken to be 2511.4 Mb based on the *Bos taurus taurus* genome assembly UMD 3.1. ROH were identified in every individual using PLINK v1.90 [121] software in non-overlapping sliding windows of 50 SNPs. The minimum length of a ROH was set to 500 kb. A maximum of three SNPs with missing genotypes and three heterozygous SNPs were admitted in each window, as discussed by Ceballos et al. [123]. Tukey’s post-hoc test [124] was used to identify significant pairwise comparisons ( $p < 0.05$ ).

### Selective sweeps detection

Four statistical methods were implemented to detect genomic regions under selective pressure. Cross-population methods encompassed the Wright’s fixation index ( $F_{ST}$ ) and the Cross-Population Extended Haplotype Homozygosity (XP-EHH). Within-population methods included the Composite Likelihood Ratio (CLR) statistic and the integrated Haplotype Score (iHS).

$F_{ST}$  [125] was calculated between all six pairwise combinations of the four breeds with custom R scripts as follows:

$$F_{ST} = \frac{\bar{p}(1 - \bar{p}) - \sum c_i p_i (1 - p_i)}{\bar{p}(1 - \bar{p})}$$

where  $\bar{p}$  is the average frequency of an allele in the

total population,  $p_i$  is the allele frequency in the  $i^{th}$  population, and  $c_i$  is the relative number of SNPs in the  $i^{th}$  population.  $F_{ST}$  scores were then averaged in non-overlapping sliding windows of 50 kb. SweepFinder2 software [126] was used to calculate the CLR statistic [127] within each breed in non-overlapping sliding windows of 50 kb across the genome. The ancestral allele information was assessed from a cattle reference allele list retrieved from Rocha et al. [128]. The CLR analysis was performed considering only SNPs containing the ancestral allele information ( $n = 11,260,629$  SNPs). The iHS [129] and XP-EHH [130] statistics were calculated using the program selscan v1.2.0a [131] with default parameters. Within each population, haplotype phasing was performed using Beagle 5.0 [132] and the genetic distances were determined by assuming that 1 Mb  $\approx$  1 centiMorgan (cM). The iHS scores were calculated within each breed and XP-EHH between all six pairwise combinations of the four breeds. The unstandardized iHS and XP-EHH scores were standard normalized using the script norm with default parameters, as provided by selscan. Absolute iHS and XP-EHH values were averaged in non-overlapping sliding windows of 50 kb. To compute the iHS statistic, the same subset of SNPs ( $n = 11,260,629$  SNPs) applied in the CLR statistic was used, however, without considering any ancestral allele information. Independent results for each statistical method and population implemented herein are presented in Additional file 13.

Selective sweeps detection can be enhanced by combining multiple genome-wide scan methodologies, benefiting from advantageous complementarities among them together with the increase in the statistical power [20, 133–136]. Further, combining within-population statistics from multiple breeds may decrease false-positive signals that arise due to population stratification (reviewed by Hellwege et al. [137]). Accordingly, within-population and cross-population statistics were combined separately in a single score using the DCMS statistic [20]. The DCMS statistic was calculated for each 50 kb window using the MINOTAUR package [138] and the empirical  $p$ -values of each statistic were derived from a skewness normal distribution with an appropriate one-tailed test (Additional file 14). Candidate sweep regions under selection were revealed by assessing the top 1% of the empirical distribution generated by the DCMS statistics.

Candidate regions identified herein were compared with previous regions under selection described in the literature for other cattle breeds. Overlap analysis was carried out using the Bioconductor package *GenomicRanges* [139].

### Selective sweeps and runs of homozygosity

Candidate sweep regions revealed from the top 1% of the empirical distribution generated by the DCMS

statistics were intersected with ROH hotspots to identify common signals between both methodologies. ROH formerly identified to estimate  $F_{ROH}$  were applied, and ROH hotspots were determined by selecting segments shared by more than 50% of the samples within each breed.

Overlap analysis was performed separately for each DCMS statistic using the Bioconductor package *GenomicRanges* [139].

### Functional annotation of the candidate regions

Genes were annotated within the candidate sweep regions using the cow gene set Ensembl release 94 fetched from the Biomart tool [140]. BEDTools [141] was used to identify overlaps between the retrieved gene set list and the putative sweep regions. DAVID v6.8 tool [115, 116] was used to identify overrepresented GO terms and KEGG pathways using the list of genes from the putative sweep regions and the *Bos taurus taurus* annotation file as a background. The  $p$ -values were adjusted by False Discovery Rate [21], and significant terms and pathways were considered when  $p < 0.01$ . QTLs retrieved from the CattleQTL database [142] were overlapped with the candidate sweep regions using BEDtools [141].

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-020-07035-6>.

**Additional file 1.** Distribution of the functional consequences of the called variants ( $n = 33,328,447$  SNPs) using the Variant Effect Predictor (VEP) tool.

**Additional file 2.** Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis enriched ( $p < 0.01$ ) based on variants with high consequence on protein sequence set of genes

**Additional file 3.** Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis enriched ( $p < 0.01$ ) based on deleterious variants (SIFT score  $< 0.05$ ) set of genes

**Additional file 4.** Analysis of Molecular Variance results.

**Additional file 5.** Annotated candidate sweep regions for the within-population statistic retrieved from the top 1% of the empirical distribution generated by the DCMS statistic.

**Additional file 6.** Annotated candidate sweep regions for the cross-population statistic retrieved from the top 1% of the empirical distribution generated by the DCMS statistic.

**Additional file 7.** Runs of homozygosity (ROH) hotspots for Gir (GIR), Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds.

**Additional file 8.** Overlapping of the putative sweep regions identified from the top 1% of the within-population DCMS statistic with candidate regions under positive selection previously reported in other cattle populations.

**Additional file 9.** Overlapping of the putative sweep regions identified from the top 1% of the cross-population DCMS statistic with candidate regions under positive selection previously reported in other cattle populations.

**Additional file 10.** Overlapping between ROH hotspots and the top 1% of the within-population DCMS statistic with the candidate regions under positive selection previously reported in other cattle populations.

**Additional file 11.** Overlapping between ROH hotspots and the top 1% of the cross-population DCMS statistic with the candidate regions under positive selection previously reported in other cattle populations.

**Additional file 12.** Brazilian geographical regions of the four cattle breeds sampled in the study (Adapted from [https://pt.wikipedia.org/wiki/Ficheiro:Brazil\\_Labelled\\_Map.svg](https://pt.wikipedia.org/wiki/Ficheiro:Brazil_Labelled_Map.svg)).

**Additional file 13.** Manhattan plot of the independent results for each selective sweep statistical method and population.

**Additional file 14.** Histogram and quantile-quantile (Q-Q) plots of statistical scores calculated for all four methods derived from a skewness normal distribution.

### Abbreviations

DCMS: De-correlated composite of multiple signals; RAPD: Random amplified polymorphic DNA; SNP: Single-nucleotide polymorphism; GIR: Gir; CAR: Caracu Caldeano; CRL: Crioulo Lageano; PAN: Pantaneiro; AMOVA: Analysis of molecular variance; K: Proportions of ancestry;  $F_{ROH}$ : Runs of homozygosity-based inbreeding coefficients; BTA: *Bos taurus* autosome; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; QTL: Quantitative trait locus; ROH: Runs of homozygosity; PCA: Principal component analysis; PNMGL: Brazilian Dairy Gir Breeding Program; ABS: American Breeders Service; CRV: Cooperatie Rundvee Verbetering; BWA-MEM: Burrows-Wheeler Alignment MEM; PCR: Polymerase chain reaction; VEP: Variant Effect Predictor; SIFT: Sorting intolerant from tolerant; DAVI D: Database for Annotation, Visualization, and Integrated Discovery; LD: Linkage disequilibrium;  $F_{ST}$ : Wright's fixation index; XPEHH: Cross-population extended haplotype homozygosity; CLR: Composite likelihood ratio; iHS: integrated haplotype score; cM: centimorgan

### Acknowledgements

The authors would like to thank the Brazilian Association of Dairy Gir Breeders, the Brazilian National Council for Scientific and Technological Development (CNPq), and the National Institute of Science and Technology - Animal Science (INCTCA) for all support as well as EMBRAPA Dairy Cattle Research Center for providing the data.

### Authors' contributions

EP, CR, HS, FB, and MVGBS contributed to the conceptualization of the manuscript. EP, CR, and HS designed the experiment. EP and CR carried out the data analysis and EP mainly wrote the manuscript. MVGBS, JCCP, and MAM provided the samples and DNA extraction. MVGBS financed the DNA preparation and re-sequencing analysis. N-TH helped to assess the best distribution to generate the empirical  $p$ -values. JG contributed with R and bash scripts to analyze the data. AAE contributed to the discussion section regarding the population structure and genomic inbreeding comprehending the locally adapted cattle breeds. All authors revised the manuscript. All authors read and approved the final manuscript.

### Funding

E.P. was supported by São Paulo Research Foundation (FAPESP) grant #2017/27148-9. 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.

### Availability of data and materials

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 corresponding author Marcos Vinícius Gualberto Barbosa da Silva, e-mail: [marcos.vb.silva@embrapa.br](mailto:marcos.vb.silva@embrapa.br)

The *Bos taurus taurus* genome assembly (UMD 3.1) used in this study can be found in [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000003055.4/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000003055.4/) (RefSeq assembly accession: GCF\_000003055.4). The SNP database (dbSNP Build 150) used in this study can be found in [https://ftp.ncbi.nih.gov/snp/organisms/archive/cow\\_9913/VCF/](https://ftp.ncbi.nih.gov/snp/organisms/archive/cow_9913/VCF/). Overrepresented GO terms and KEGG pathways



described in this study were retrieved from DAVID (Database for Annotation, Visualization, and Integrated Discovery) v6.8 tool [115, 116].

#### Ethics approval and consent to participate

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.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>São Paulo State University (Unesp), School of Agricultural and Veterinarian Sciences, Jaboticabal 14884-900, Brazil. <sup>2</sup>Animal Breeding and Genetics Group, Department of Animal Sciences, University of Goettingen, Albrecht-Thaer-Weg 3, 37075 Goettingen, Germany. <sup>3</sup>Center for Integrated Breeding Research, University of Goettingen, Albrecht-Thaer-Weg 3, 37075 Goettingen, Germany. <sup>4</sup>National Council for Scientific and Technological Development (CNPq), Lago Sul 71605-001, Brazil. <sup>5</sup>Embrapa Dairy Cattle, Juiz de Fora 36038-330, Brazil. <sup>6</sup>Embrapa Beef Cattle, Campo Grande 79106-550, Brazil.

Received: 30 March 2020 Accepted: 27 August 2020

Published online: 11 September 2020

#### References

- Primo A. El ganado bovino ibérico en las Américas: 500 años después. *Arch Zootec.* 1992;41:421–32.
- Mariante A, Cavalcante N. Animais do descobrimento: raças domésticas da história do Brasil. Centro de Pesquisa Agropecuária do Pantanal: Empresa Brasileira de Pesquisa Agropecuária; 2000.
- Egito AA, Mariante AS, Albuquerque MSM. Programa brasileiro de conservação de recursos genéticos animais. *Arch Zootec.* 2002;51:7.
- da Mariante A. S, Albuquerque M do SM, do Egito AA, McManus C. Advances in the Brazilian animal genetic resources conservation programme. *Anim Genet Resour Inf.* 1999;25:107–21.
- Felix G, Piovezan U, Juliano R, Silva M, Fioravanti M. Potencial de uso de raças bovinas locais brasileiras: Curraleiro Pé-duro e Pantaneiro. *Enciclopédia Biosf.* 2013;9:1715–41.
- Kim J, Hanotte O, Mwai OA, Dessie T, Salim B, Diallo B, et al. The genome landscape of indigenous African cattle. *Genome Biol.* 2017;18:34.
- Zander KK, Signorello G, De Salvo M, Gandini G, Drucker AG. Assessing the total economic value of threatened livestock breeds in Italy: Implications for conservation policy. *Ecol Econ.* 2013;93:219–29.
- Ugarte E, Ruiz R, Gabia D, Beltrán de Heredia I. Impact of high-yielding foreign breeds on the Spanish dairy sheep industry. *Livest Prod Sci.* 2001;71:3–10.
- Carvalho GMC, Fé Da Silva LR, Almeida MJO, Lima Neto AF, Beffa LM. Phenotypic evaluation of Curraleiro Pé-duro breed of cattle from semi-arid areas of Brazil. *Arch Zootec.* 2013;62:23–25.
- Cardoso CC, Lima FG, Fioravanti MCS, Egito AA, Paula e Silva FC, Tanure CB, et al. Heat tolerance in curraleiro pe-duro, pantaneiro and nelore cattle using thermographic images. *Animals.* 2016;6.
- Utsunomiya YT, Pérez O'Brien AMP, Sonstegard TS, Sölkner J, Garcia JF. Genomic data as the "hitchhiker's guide" to cattle adaptation: Tracking the milestones of past selection in the bovine genome. *Front Genet.* 2015;6.
- Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet.* 2014;46.
- Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.* 2014;10:e100414.
- Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci Rep.* 2016;6:38932.
- Egito AA, Martinez AM, Juliano RS, Landi V, Moura MI, Silva MC, et al. Population study of Pantaneiro cattle herds aiming the management and genetic handling of the breed. *Actas Iberoam en Conserv Anim.* 2016;7:59–63.
- Campos BM, Carmo AS, Egito AA, Mariante AS, Albuquerque MSM, Gouveia JJS, et al. Genetic diversity, population structure, and correlations between locally adapted zebu and taurine breeds in Brazil using SNP markers. *Trop Anim Health Prod.* 2017;49:1677–84.
- Serrano G, Egito A, McManus C, Mariante A. Genetic diversity and population structure of Brazilian native bovine breeds. *Pesqui Agropecu Bras.* 2004;39:543–9.
- Pezzini T, Mariante AS, Martins E, Paiva S, Seixas L, Costa JBG, et al. Population structure of Brazilian Crioula lageana cattle (*Bos taurus*) breed. *Rev Colomb Ciencias Pecu.* 2018;31:93–102.
- Egito A, Paiva S, Albuquerque M do S, Mariante A, Almeida L, Castro S, et al. Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. *BMC Genet.* 2007;8:83.
- Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. *Heredity (Edinb).* 2015;115:426–36.
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc.* 1995;57:289–300.
- Kiser JN, Lawrence TE, Neupane M, Seabury CM, Taylor JF, Womack JE, et al. Rapid communication: Subclinical bovine respiratory disease - loci and pathogens associated with lung lesions in feedlot cattle. *J Anim Sci.* 2017; 95:2726–31.
- Veerkamp RF, Coffey MP, Berry DP, De Haas Y, Strandberg E, Bovenhuis H, et al. Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. *Animal.* 2012;6:1738–49.
- Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, et al. Genome-wide association study of growth in crossbred beef cattle. *J Anim Sci.* 2010;88:837–48.
- Purfield DC, Bradley DG, Evans RD, Kearney FJ, Berry DP. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. *Genet Sel Evol.* 2015;47:47.
- Mateescu RG, Garrick DJ, Reecy JM. Network analysis reveals putative genes affecting meat quality in Angus cattle. *Front Genet.* 2017;8.
- Strillacci MG, Frigo E, Schiavini F, Samoré AB, Canavesi F, Vevey M, et al. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. *BMC Genet.* 2014;15:106.
- Crispim AC, Kelly MJ, Guimarães SEF, E Silva FF, Fortes MRS, Wenceslau RR, et al. Multi-trait GWAS and new candidate genes annotation for growth curve parameters in brahman cattle. *PLoS One.* 2015;10:e0139906.
- Mapholi NO, Maiwashe A, Matika O, Riggio V, Bishop SC, MacNeil MD, et al. Genome-wide association study of tick resistance in South African Nguni cattle. *Ticks Tick Borne Dis.* 2016;7:487–97.
- Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics.* 2017;18.
- Hawken RJ, Zhang YD, Fortes MRS, Collis E, Barris WC, Corbet NJ, et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. *J Anim Sci.* 2012;90:1398–410.
- Parker Gaddis KL, Null DJ, Cole JB. Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *J Dairy Sci.* 2016;99:6420–35.
- Wu X, Fang M, Liu L, Wang S, Liu J, Ding X, et al. Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. *BMC Genomics.* :897.
- Tetens J, Seidenspinner T, Buttcherer N, Thaller G. Whole-genome association study for energy balance and fat/protein ratio in German Holstein bull dams. *Anim Genet.* 2013;44:1–8.
- Cole JB, Wiggins GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics.* 2011;12:408.
- Nayeri S, Sargolzaei M, Abo-Ismael MK, May N, Miller SP, Schenkel F, et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* 2016;17:75.
- Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP, et al. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genet.* 2012;13:21.



38. Huson HJ, Kim E-S, Godfrey RW, Olson TA, McClure MC, Chase CC, et al. Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Front Genet.* 2014;5.
39. Iso-Touru T, Sahana G, Guldbandsen B, Lund MS, Vilkki J. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genet.* 2016;17:55.
40. Bahbahani H, Clifford H, Wragg D, Mbole-Kariuki MN, Van Tassel C, Sonstegard T, et al. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Sci Rep.* 2015;5:11729.
41. Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Van Tassel CP, et al. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol.* 2015;32:711–25.
42. Makina SO, Muchadeyi FC, Van Marle-Köster E, Taylor JF, Makgahlela ML, Maiwashe A. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet Sel Evol.* 2015;47:92.
43. González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Díaz C, Piedrafita J, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. *Genet Sel Evol.* 2016;48:81.
44. Rothhammer S, Seichter D, Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics.* 2013;14:908.
45. Pitt D, Bruford MW, Barbato M, Orozco-terWengel P, Martínez R, Sevane N. Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evol Appl.* 2019;12:105–22.
46. Iso-Touru T, Tapio M, Vilkki J, Kiseleva T, Ammosov I, Ivanova Z, et al. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim Genet.* 2016;47:647–57.
47. Somavilla AL, Sonstegard TS, Higa RH, Rosa AN, Siqueira F, Silva LOC, et al. A genome-wide scan for selection signatures in Nelore cattle. *Anim Genet.* 2014;45:771–81.
48. Liao X, Peng F, Forni S, McLaren D, Plastow G, Stothard P. Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. *Genome.* 2013;56:592–8.
49. Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, et al. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol Biol Evol.* 2018;35:688–99.
50. Wang Z, Ma H, Xu L, Zhu B, Liu Y, Bordbar F, et al. Genome-Wide Scan Identifies Selection Signatures in Chinese Wagyu Cattle Using a High-Density SNP Array. *Animals.* 2019;9.
51. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol.* 2015;47:49.
52. Pérez O'Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassel CP, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol.* 2014;46:19.
53. Boitard S, Boussaha M, Capitan A, Rocha D, Servin B. Uncovering adaptation from sequence data: Lessons from genome resequencing of four cattle breeds. *Genetics.* 2016;203:433–50.
54. Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics.* 2010;185:1451–61.
55. Machugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (*Bos Taurus* and *Bos Indicus*). *Genetics.* 1997;146:1071–86.
56. Hiendleder S, Lewalski H, Janke A. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet Genome Res.* 2008;120:150–6.
57. Chan EKF, Nagaraj SH, Reverter A. The evolution of tropical adaptation: Comparing taurine and zebu cattle. *Anim Genet.* 2010;41:467–77.
58. Mazza M, Mazza C, Sereno J, Santos S, Pellegrin A. Etnobiologia e conservação do bovino Pantaneiro. Centro de Pesquisa Agropecuária do Pantanal: Empresa Brasileira de Pesquisa Agropecuária; 1994.
59. Issa EC, Jorge W, Sereno JRB. Cytogenetic and molecular analysis of the Pantaneiro cattle breed. *Pesqui Agropecu Bras.* 2006;41:1609–15.
60. Queiroz SA, Pelicioni LC, Silva BF, Sesana JC, Martins MIEG, Sanches A. Selection indices for a dual purpose breed Caracu. *Rev Bras Zootec.* 2005;34:827–37.
61. Mariante AS, Egito AA, Albuquerque M do SM, Paiva SR, Ramos AF. Managing genetic diversity and society needs. *Rev Bras Zootec.* 2008;37:127–36.
62. Mazza MCM, Mazza CA, Sereno JRB, Santos SAL, Mariante AS. Conservation of Pantaneiro cattle in Brazil: Historical origin. *Arch Zootec.* 1992;41:443–53.
63. Mariante AS, Albuquerque M do SM, Egito AA, McManus C, Lopes MA, Paiva SR. Present status of the conservation of livestock genetic resources in Brazil. *Livest Sci.* 2009;120:204–12.
64. Queiroz SA, Lôbo RB. Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil. *J Anim Breed Genet.* 1993;110:228–33.
65. Wright S. Coefficients of Inbreeding and Relationship. *Am Nat.* 1922;56:330–8.
66. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet.* 2014;46:110–21.
67. Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassel CP, Crooker BA, et al. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One.* 2013;8:e80813.
68. Reis Filho JC, Lopes PS, Verneque R da S, Torres R de A, Teodoro RL, Carneiro PLS. Population structure of Brazilian Gyr dairy cattle. *Rev Bras Zootec.* 2010;39:2640–5.
69. Santana Junior ML, Pereira RJ, Bignardi AB, El Faro L, Tonhati H, Albuquerque LG. History, structure, and genetic diversity of Brazilian Gir cattle. *Livest Sci.* 2014;163:26–33.
70. Peripolli E, Baldi F, da Silva MVGB, Irgang R, Lima ALF, R. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics.* 2018;19:34.
71. Neves HHR, Scaletz DCB, Queiroz SA, Desidério JA, Pimentel ECG. Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle. *Arch Zootec.* 2015;64:99–108.
72. Ferguson JD, Galligan DT, Thomsen N. Principal Descriptors of Body Condition Score in Holstein Cows. *J Dairy Sci.* 1994;77:2695–703.
73. Bauman DE, Bruce CW. Partitioning of Nutrients During Pregnancy and Lactation: A Review of Mechanisms Involving Homeostasis and Homeorhesis. *J Dairy Sci.* 1980;63:1514–29.
74. Bell AW. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. *J Anim Sci.* 1995;73:2804–19.
75. Whitaker DA, Goodger WJ, Garcia M, Perera BMAO, Wittwer F. Use of metabolic profiles in dairy cattle in tropical and subtropical countries on smallholder dairy farms. *Prev Vet Med.* 1999;38:119–31.
76. Stockdale CR. Body condition at calving and the performance of dairy cows in early lactation under Australian conditions: A review. *Aust J Exp Agric.* 2001;41:823–39.
77. Collard BL, Boettcher PJ, Dekkers JCM, Petitclerc D, Schaeffer LR. Relationships between energy balance and health traits of dairy cattle in early lactation. *J Dairy Sci.* 2000;83:2683–90.
78. Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, et al. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genet.* 2017;18:11.
79. Roux PF, Boitard S, Blum Y, Parks B, Montagner A, Mouisel E, et al. Combined QTL and selective sweep mappings with coding SNP annotation and cis-eQTL analysis revealed PARK2 and JAG2 as new candidate genes for adiposity regulation. *G3 Genes, Genomes, Genet.* 2015;5:517–29.
80. dos Santos FC, Peixoto MGCD, Fonseca PA de S, Pires M de FÁ, Ventura RV, Rosse I da C, et al. Identification of Candidate Genes for Reactivity in Guzerat (*Bos indicus*) Cattle: A Genome-Wide Association Study. *PLoS One* 2017;12:e0169163.
81. Lee YL, Bosse M, Mullaart E, Groenen MAM, Veerkamp RF, Bouwman AC. Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC Genomics.* 2020;21:89.
82. Valente TS, Baldi F, Sant'Anna AC, Albuquerque LG, Costa MJRP Da. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nelore cattle *PLoS One.* 2016;11:e0156956.
83. Burrow HM, Prayaga KC. Correlated responses in productive and adaptive traits and temperament following selection for growth and heat resistance in tropical beef cattle. *Livest Prod Sci.* 2004;86:143–61.
84. Burdick NC, Randel RD, Carroll JA, Welsh TH. Interactions between temperament, stress, and immune function in cattle. *Int J Zool.* 2011;2011.
85. Voisin BD, Grandin T, Tatum JD, O'Connor SF, Struthers JJ. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. *J Anim Sci.* 1997;75:892–6.
86. Silveira IDB, Fischer V, Farinatti LHE, Restle J, Filho DCA, de Menezes LFG. Relationship between temperament with performance and meat quality of

- feedlot steers with predominantly Charolais or Nellore breed. *Rev Bras Zootec.* 2012;41:1468–76.
87. Cafe LM, Robinson DL, Ferguson DM, McIntyre BL, Geesink GH, Greenwood PL. Cattle temperament: Persistence of assessments and associations with productivity, efficiency, carcass and meat quality traits. *J Anim Sci.* 2011;89:1452–65.
  88. Petherick JC, Holroyd RG, Swain AJ. Performance of lot-fed Bos indicus steers exposed to aspects of a feedlot environment before lot-feeding. *Aust J Exp Agric.* 2003;43:1181–91. <https://doi.org/10.1071/EA02118>.
  89. Burrow HM. Measurement of temperament and their relationship with performance traits of beef cattle. *Anim Breed Abstr.* 1997;65:478–95.
  90. Frischknecht M, Flury C, Leeb T, Rieder S, Neuditschko M. Selection signatures in Shetland ponies. *Anim Genet.* 2016;47:370–2.
  91. Avila F, Mickelson JR, Schaefer RJ, McCue ME. Genome-wide signatures of selection reveal genes associated with performance in American Quarter Horse subpopulations. *Front Genet.* 2018;9.
  92. Gardner JL, Peters A, Kearney MR, Joseph L, Heinsohn R. Declining body size: A third universal response to warming? *Trends Ecol Evol.* 2011;26:285–91.
  93. Martin JM, Mead JL, Barboza PS. Bison body size and climate change. *Ecol Evol.* 2018;8:4564–74.
  94. Dickerson GE. Animal size and efficiency: Basic concepts. *Anim Prod.* 1978;27:367–79.
  95. McCain CM, King SRB. Body size and activity times mediate mammalian responses to climate change. *Glob Chang Biol.* 2014;20:1760–9.
  96. Pacifici M, Visconti P, Butchart SHM, Watson JEM, Cassola FM, Rondinini C. Species' traits influenced their response to recent climate change. *Nat Clim Chang.* 2017;7:205–8.
  97. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nat Rev Genet.* 2013;14:807–20.
  98. Taylor CR, Caldwell SL, Rowntree VJ. Running up and down hills: Some consequences of size. *Science (80- ).* 1972;178:1096–7.
  99. Araújo Teixeira RM, Lana R de P, Fernandes L de O, de Oliveira AS, de Queiroz AC, de Oliveira Pimentel JJ. Desempenho produtivo de vacas da raça Gir leiteira em confinamento alimentadas com níveis de concentrado e proteína bruta nas dietas. *Rev Bras Zootec.* 2010;39:2527–34.
  100. Mcmanus C, Seixas L. A Raça Crioula Lageana. 2010. [www.animal.unb.br](http://www.animal.unb.br).
  101. Issa EC, Jorge W, Egito AA, Sereno JRB. Cytogenetic analysis of the Y chromosome of native Brazilian bovine breeds: preliminary data. *Arch Zootec.* 2009;58:93–101.
  102. Araújo AM de, Ramos AF, Egito AA do, Mariante A da S, Varela ES, Figueiredo EAP de, et al. Núcleos de conservação de Bovinos. In: Albuquerque M do SM, Ianella P, editors. *Inventário de Recursos Genéticos Animais da Embrapa*. Brasília: Empresa Brasileira de Pesquisa Agropecuária; 2016. p. 17–23.
  103. Gutiérrez-Gil B, Arranz JJ, Wiener P. An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. *Front Genet.* 2015;6:167.
  104. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays.* 2013;35:780–6.
  105. Jakobsson M, Edge MD, Rosenberg NA. The relationship between FST and the frequency of the most frequent allele. *Genetics.* 2013;193:515–28.
  106. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 2005;15:1496–502.
  107. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. *Livest Sci.* 2014;166:133–43.
  108. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv.* 2013;1303.
  109. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
  110. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
  111. DePristo MA, Rivas MA, McKenna A, Hartl C, del Angel G, Sivachenko AY, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
  112. Garimella KV, Levy-Moonshine A, Jordan T, Van der Auwera GA, Hartl C, del Angel G, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinforma.* 2013;11:11.10.1–11.10.33.
  113. Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308–11.
  114. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:122.
  115. Huang DW, Sherman BT, Lempicki R a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44–57.
  116. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1–13.
  117. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics.* 1992;131:479–91.
  118. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics.* 2010;26:419–20.
  119. R Core Team R. R: A Language and Environment for Statistical Computing. Available online at <https://www.R-project.org/>; 2015.
  120. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
  121. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
  122. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. *Am J Hum Genet.* 2008;83:359–72.
  123. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics.* 2018;19:106.
  124. Tukey JW. Comparing Individual Means in the Analysis of Variance. *Biometrics.* 1949;5:99–114.
  125. Wright S. The Genetical Structure of populations. *Nature.* 1950;166:247–9.
  126. DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics.* 2016;32:1895–7.
  127. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome Res.* 2005;15:1566–75.
  128. Rocha D, Billerey C, Samson F, Boichard D, Boussaha M. Identification of the putative ancestral allele of bovine single-nucleotide polymorphisms. *J Anim Breed Genet.* 2014;131:483–6.
  129. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4:e72.
  130. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature.* 2007;449:913–8.
  131. Szpiech ZA, Hernandez RD. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol.* 2014;31:2824–7.
  132. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet.* 2018;103:338–48.
  133. Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Van Tassel CP, do Carmo AS, Mészáros G, et al. Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods *PLoS One.* 2013;8:e64280.
  134. Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW. Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genet.* 2014;15:34. <https://doi.org/10.1186/1471-2156-15-34>.
  135. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, et al. A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science (80- ).* 2010;327:883–6.
  136. Lin K, Li H, Schlötterer C, Futschik A. Distinguishing positive selection from neutral evolution: Boosting the performance of summary statistics. *Genetics.* 2011;187:229–44.
  137. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet.* 2017;95:1.22.1–1.22.23.
  138. Verity R, Collins C, Card DC, Schaal SM, Wang L, Lotterhos KE. MINOTAUR: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. *Mol Ecol Resour.* 2017;17:33–43.
  139. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol.* 2013;9:e1003118.

140. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.* 2009;37:W23–7.
141. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
142. Hu ZL, Park CA, Reecy JM. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* 2016;44:D827–33.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

