



# Pedigree reconstruction and population structure using SNP markers in Gir cattle

Arielly Oliveira Garcia<sup>1</sup> · Pamela Itajara Otto<sup>2</sup> · Luiz Afonso Glatzl Junior<sup>3</sup> · Renata de Fátima Bretanha Rocha<sup>1</sup> · Mateus Guimarães dos Santos<sup>1</sup> · Daniele Alves de Oliveira<sup>2</sup> · Marcos Vinicius Gualberto Barbosa da Silva<sup>4</sup> · João Cláudio do Carmo Panetto<sup>4</sup> · Marco Antônio Machado<sup>4</sup> · Rui da Silva Verneque<sup>4</sup> · Simone Eliza Facioni Guimarães<sup>1</sup>

Received: 20 July 2022 / Revised: 24 November 2022 / Accepted: 3 January 2023 / Published online: 16 January 2023  
© The Author(s), under exclusive licence to Institute of Plant Genetics Polish Academy of Sciences 2023

## Abstract

Our objective was to establish a SNPs panel for pedigree reconstruction using microarrays of different densities and evaluate the genomic relationship coefficient of the inferred pedigree, in addition to analyzing the population structure based on genomic analyses in Gir cattle. For parentage analysis and genomic relationship, 16,205 genotyped Gir animals (14,458 females and 1747 males) and 1810 common markers to the four SNP microarrays were used. For population structure analyses, including linkage disequilibrium, effective population size, and runs of homozygosity (ROH), genotypes from 21,656 animals were imputed. Likelihood ratio (LR) approach was used to reconstruct the pedigree, deepening the pedigree and showing it is well established in terms of recent information. Coefficients for each relationship category of the inferred pedigree were adequate. Linkage disequilibrium showed rapid decay. We detected a decrease in the effective population size over the last 50 generations, with the average generation interval around 9.08 years. Higher ROH-based inbreeding coefficient in a class of short ROH segments, with moderate to high values, was also detected, suggesting bottlenecks in the Gir genome. Breeding strategies to minimize inbreeding and avoid massive use of few proven sires with high genetic value are suggested to maintain genetic variability in future generations. In addition, we recommend reducing the generation interval to maximize genetic progress and increase effective population size.

**Keywords** Inbreeding · Linkage disequilibrium · Parentage · Relationship · Runs of homozygosity

## Introduction

Taurine breeds are adapted in temperate or subtropical climates, but do not maintain performances in tropical regions, especially due to heat stress and parasite

infestations. Thus, better adapted cattle genetics is needed in tropical herds, such as indicine breeds. Gir breed is a major choice in tropical milk production, either in cross-breeding or as pure animals.

Considering genomic evaluation in commercial breeds, there are several available genotyping panels. Moreover, genotypes obtained at different densities microarrays are used to estimate relationship coefficients with great accuracy in an unobserved pedigree or to identify genomic regions that are not affected by recombination (Speed and Balding 2014).

The success of breeding programs depends on the knowledge of several factors that interfere with selection, such as effective population size, generation interval, and genetic variability (Malhado et al. 2010). Population structure combined with information about genetic changes can guide management decisions, allowing development of strategies to promote genetic gain.

---

Communicated by: Maciej Szydlowski

✉ Simone Eliza Facioni Guimarães  
sfacioni@ufv.br

<sup>1</sup> Department of Animal Science, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

<sup>2</sup> Department of Animal Science, Universidade Federal de Santa Maria, Santa Maria, RS 97105-900, Brazil

<sup>3</sup> Department of Computer Science, Universidade Federal de Juiz de Fora, Juiz de Fora, MG 36036-900, Brazil

<sup>4</sup> EMBRAPA - Dairy Cattle Research Center, Juiz de Fora, MG 36038-330, Brazil

The main studies evaluating population structure are based on pedigree records (Leroy et al. 2013; Santana Jr et al. 2014). Recent genomic approaches provide information on genetic diversity, population history, and inbreeding coefficient (Peripolli et al. 2018; Ospina et al. 2019). In the current research, our objectives are as follows: (1) to establish a SNP panel for parentage testing and pedigree reconstruction using different densities microarrays and (2) to evaluate genomic relationship coefficient of inferred pedigrees, in addition to disentangle population structure in Gir cattle.

## Materials and methods

### Data

Our data includes files from the Brazilian Dairy Gir Breeding Program (PNMGL), coordinated by Brazilian Agricultural Research Corporation (Embrapa) Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil) in cooperation with the Brazilian Association of Zebu Breeders (ABCZ) and the Brazilian Association of Dairy Gir Breeders (ABCGIL).

For parentage and genomic relationship analyses, genotypes and genealogical records of 16,983 animals born between 1960 and 2020 were considered. DNA samples were analyzed using four different densities of SNP microarrays: 597 animals genotyped using Illumina Bovine HD BeadChip with 777 K (Illumina, San Diego, CA, USA), 11,207 animals using GeneSeek Genomic Profiler Indicus (GGPi) BeadChip, with 35 K; 3,653 animals with GGPI 50 K, and 1526 animals using GGPI-LD BeadChip with 27 K (GeneSeek, Inc., Lincoln, NE, USA). Map positions were based on ARS-UCD1.2 bovine genome assembly (GenBank project accession: NKLS00000000.2), being initially 7372 SNPs common to these four microarrays.

Quality control for genotypes was implemented using snpStats package (Clayton 2019) in the R software (R version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria). Samples with call rate < 0.90 and average GenCall score (GC) < 0.70 were removed. SNPs at the same position or not assigned to any chromosome were removed from the dataset. Only autosomal SNPs with Call Rate > 0.90, at Hardy–Weinberg equilibrium ( $p$  value >  $10^{-6}$ ), minor allele frequency (MAF) > 0.05, and mean GC score > 0.70 present in all four different SNP microarrays were maintained for further analysis. SNPs in linkage disequilibrium, with  $r^2 > 0.20$ , were removed. After quality control, 16,205 Gir animals, 14,458 females and 1747 males, remained for parentage and genomic relationship analyses and 1810 SNPs common to the four SNP microarrays.

For population structure analyses, which includes linkage disequilibrium, effective population size and runs of

homozygosity, the genotypes of 21,656 animals genotyped with five microarrays of different densities were imputed. A detailed description of this analysis is done in the next section.

### Imputation

For population structure analyses, genotypes and genealogical records from animals born between 1960 and 2020 were considered. Imputation was implemented using FIMPUTE 2.2 software (Sargolzaei et al. 2014), in which the lower density panels were imputed to the high-density level (HD). Animals genotyped with the Bovine HD BeadChip (Illumina Inc., San Diego, CA, USA) were used as reference population for imputation. After imputation, the 21,656 animals had information of 416,155 SNPs. Imputation accuracy was 0.9669.

### Parentage analysis and pedigree reconstruction

Pedigree reconstruction was performed by likelihood method, using sequoia package (Huisman 2017) in the R software (R version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria), which calculates the likelihood ratio (LR) of the offspring genotype. LR gives more weight to rare alleles, besides considering that all loci are independent; so the total likelihood ratio is multiplied by all loci (Grashei et al. 2018). We used the logarithm (base 10) of likelihood ratios (LLR), which indicates that a given animal is the parent versus the next most likely relationship between the focal individual and this animal (Marshall et al. 1998), with parentage attributed to the most likely parent. LLR values, while useful in obtaining a solution of maximum likelihood, cannot be interpreted statistically or biologically.

The quantity that was maximized was the total likelihood of the pedigree configuration over all genotyped individuals (N) (Huisman 2017):

$$\mathcal{L}(P) = \prod_{A=1}^N \mathcal{L}(A, D_A, S_A) \approx \prod_{A=1}^N \prod_l P(A_l = X | D_A, S_A)$$

where  $P(A_l = X | D_A, S_A)$  is the probability of observing genotype  $X$  at locus  $l$  in individual  $A$ , conditional only on its parents  $D_A$  and  $S_A$  in pedigree  $P$ . Which were then multiplied over all individuals and multiplied over all loci.

Although pedigree files were available initially, to reconstruct the pedigree, we considered the worst-case scenario, assuming no available pedigree information. This implied that all peer relationships would be tested. The following relationships were attributed: parent-offspring (PO); full siblings (FS); half siblings (HS); grand-offspring (GO); full avuncular (FA); half avuncular (HA); full nephew/niece (FN); and half nephew/niece (HN).

In well-established pedigrees, the sequoia package (Huisman 2017) estimates the probability of confidence through reference pedigrees. For this analysis, the reference pedigree provided by the Brazilian Association of Zebu Breeders (ABCZ) was used, with prior verification of incompatibilities between parents and offspring based on the count of Mendelian conflicts, using SeekParentF90 software (Aguilar et al. 2014). Conflict threshold to exclude any parent was set to one percent (1.0%) of SNP markers. Confidence probability, obtained through ten simulations, was considered the number of correct assignments, which coincided with the reference pedigree, divided by all assignments made in simulated pedigrees.

### Genomic relationship coefficient

Genomic relationship matrix was calculated according to VanRaden (2008), using preGSf90 (Aguilar et al. 2010):

$$G = \frac{ZZ'}{2 \sum p_j(1 - p_j)}$$

where  $Z$  is an incidence matrix for SNP effects with elements for animal  $i$  and SNP  $j$  with allele frequency  $p_j$  (VanRaden 2008).

### Linkage disequilibrium (LD) and LD decay

Linkage disequilibrium between SNPs is computed as correlation of gene frequencies ( $r_{ij}^2$ ) (Hill and Robertson 1968) using PLINK 1.9 software (Purcell et al. 2007):

$$r_{ij}^2 = \frac{(p_{ij} - p_i q_j)^2}{p_1 p_2 q_1 q_2}$$

where  $p_{ij}$  is the probability of the marker allele pair  $i$  and  $j$ ;  $p_i$  and  $q_j$  are marginal allelic frequencies at  $i$  and  $j$ ; and  $p_1 p_2 q_1 q_2$  is the product of four allele frequencies at both loci.

To assess LD decay,  $r_{ij}^2$  will be regressed in the distance between pairs of markers based on the nonlinear parametric model described by Sved (1971):

$$LD_{ij} = ([1 + 4\beta d_{ij}] + e_{ij})^{-1}$$

where  $LD_{ij}$  is the  $r_{ij}^2$  observed between SNPs  $i$  and  $j$ ;  $d_{ij}$  is the distance in Kb between SNPs  $i$  and  $j$ ;  $\beta$  is the coefficient that describing LD decay with distance; and  $e_{ij}$  is a random residue defined as  $e_{ij} \sim N(0, \sigma^2)$ .

### Effective population size

Effective population size ( $N_e$ ) was estimated based on relationship between  $N_e$  and LD without mutation and

recombination rate ( $c$ ) proposed by Sved (1971), using the equation (Hayes et al. 2003; Ospina et al. 2019):

$$N_T = (1/4c) \times ((1/r^2) - 1)$$

where  $c$  represents the distance from map in Morgans, assuming 1 cM is equal to 1 Mb, and  $N_T$  is effective population size in the  $T^{th}$  generation, and  $T$  is  $1/2c$ .

### Runs of homozygosity and inbreeding coefficient

Runs of homozygosity (ROH) were identified in 21,656 individuals, using the genotypes imputed, by sliding window methodology of specified length, or a number of homozygous SNPs. For this, we used PLINK 1.9 software (Purcell et al. 2007). The parameters applied to define ROHs were as follows: (i) a sliding window of 50 SNPs throughout the genome; (ii) proportion of homozygous overlapping windows was 0.05; (iii) minimum number of consecutive SNPs included in an ROH was 120; (iv) minimum ROH length was 1 Mb; (v) maximum interval between consecutive homozygous SNPs was 1000 Kb; and (vi) density of one SNP per 50 Kb; vii. maximum of 1 SNP with missing genotypes and even one heterozygous genotype was allowed in every ROH. ROHs were defined by a minimum of 1 Mb in length to avoid short and common ROHs, which occur throughout genomes due to LD. ROHs were classified into five length classes: 1–2, 2–4, 4–8, 8–16, and > 16 Mb, identified as ROH<sub>1–2 Mb</sub>, ROH<sub>2–4 Mb</sub>, ROH<sub>4–8 Mb</sub>, ROH<sub>8–16 Mb</sub>, and ROH<sub>>16 Mb</sub>, respectively.

Genomic inbreeding coefficients based on ROH ( $F_{ROH}$ ) were estimated for each animal, using detectRUNS package (Biscarini et al. 2019) in R software (R version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria), according to McQuillan et al. (2008):

$$F_{ROH} = \frac{\sum_{j=1}^n L_{ROHj}}{L_{total}}$$

where  $L_{ROHj}$  is the length of ROH<sub>j</sub> and  $L_{total}$  is total size of each class, chromosome, or wide genome, covered by markers.  $L_{total}$ , for wide genome, was taken as 2,487,068,108 bp, based on consensus map ARS-UCD1.2 (GenBank project accession: NKLS00000000.2).

For each animal,  $F_{ROH}$  ( $F_{ROH1-2 Mb}$ ,  $F_{ROH2-4 Mb}$ ,  $F_{ROH4-8 Mb}$ ,  $F_{ROH8-16 Mb}$ , and  $F_{ROH > 16 Mb}$ ) was calculated based on the distribution of ROH of the five different lengths (ROH<sub>j</sub>).  $F_{ROH}$  for each generation was calculated using the expected length of autozygous segments in distributions with average equal to  $1/2g$  Morgans, where  $g$  is the number of generations since common ancestors (Howrigan et al. 2011). Generation intervals were calculated using optiSel package (Wellmann

2021) in R software (R version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria).

## Results

In this work, for parentage and genomic relationship analyses, SNPs in common from four different genotyping microarrays were used. Markers present at all microarrays initially resulted in a set of 7372 common SNPs (Fig. 1A). After quality control, a panel with 1810 common SNPs was considered (Fig. 1B).

MAF values of the 1810 SNPs ranged from 0.06 to 0.50, with mean MAF of  $0.35 \pm 0.115$ . MAF values for each marker, as well as name, chromosome, and position, according to the ARS-UCD1.2 bovine genome assembly, are presented in Online Resource 1. The distribution of SNPs number along the 29 autosome chromosomes can be seen in Online Resource 2. From ISAG SNP panel (ISAG, 2012), after quality control 46 markers (23%), were polymorphic. MAF values of these SNPs ranged from 0.07 to 0.50, with mean and standard deviation of  $0.24 \pm 0.125$ , respectively.

### Parentage analysis and pedigree reconstruction

LLR values ranged from 3.84 to 84.56 for dams and 1.92 to 128.72 for sires, being mean and standard deviation for dams and sires  $42.44 \pm 6.08$  and  $33.77 \pm 12.69$ , respectively. For pairs of parents, the LLR values ranged from 36.38 to 89.50, with mean and standard deviation of  $63.52 \pm 10.77$ . The distribution of LLR frequencies for pairs, dams, and sires can be seen in Fig. 2.

The opposite homozygous loci (OH) were checked between a given individual and each of the assigned parents (Fig. 3), being

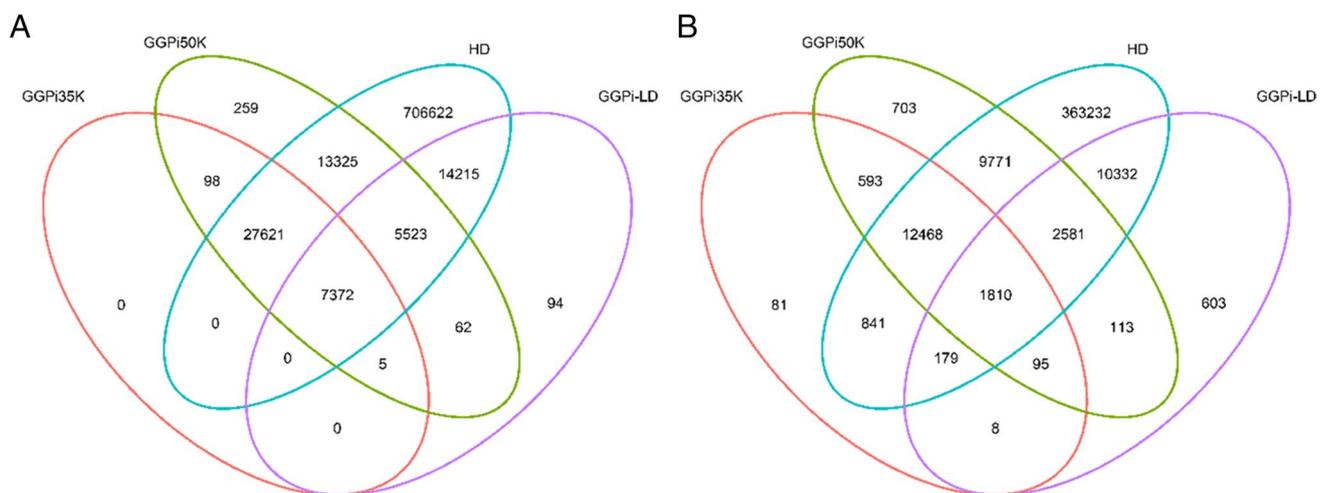
OH calculated for the parent–offspring relationship and the Mendelian Error (ME) for the trio information (sire–offspring–dam).

There should be no Mendelian inconsistencies in true parent–offspring relationships, and no locus with OH should occur between both evaluated individuals, except for mutation events and genotyping errors. Regarding 9238 dam assignments, 85.90% ( $n = 7935$ ) did not present opposite homozygosity, 13.21% ( $n = 1220$ ) presented OH equal to one, and for loci in OH equal to two and three, the percentage was 0.89% ( $n = 80$  and  $n = 2$ , respectively). For sires, out of the 15,066 parentage assignments, 90.21% ( $n = 13,591$ ) were detected OH equal to zero, 9.49% for OH equal to one ( $n = 1430$ ), and for loci in OH equal to two and three, the percentage was 0.30% ( $n = 43$  and  $n = 2$ , respectively).

Considering ME count, only assignments for parent pairs were used. For pairs of parents, 8799 assignments were made, in which 75.33% ( $n = 6628$ ) did not present ME, 21.98% ( $n = 1934$ ) presented one ME, 2.48% presented ME equal to two ( $n = 218$ ), and the sum of ME equal to three and four represented 0.22% ( $n = 17$  and  $n = 2$ , respectively).

Estimation of parentage assignments confidence to the most likely parent was achieved by comparison with the reference pedigree (Huisman 2017). Confidence probability values in the current study equals to 99.9970% for pairs of parents, 99.9980% for sires, and 99.9956% for dams.

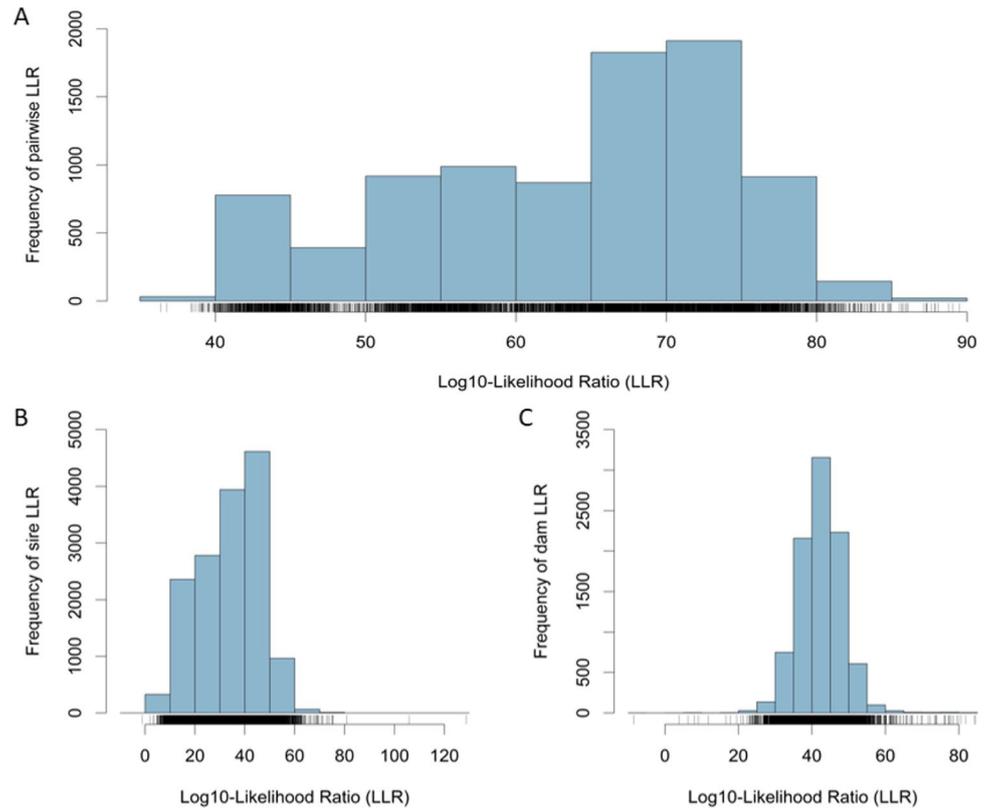
In the reference pedigree, there was 13,460 information about sires ( $\text{Info}_{\text{Ref}}$ ) (Table 1), while in the inferred pedigree, we assigned 15,066 ( $\text{Info}_{\text{Inf}}$ ), with 1612 information occurring only in the inferred pedigree ( $\text{Info}_{\text{only}}$ ). In the dam–offspring category, the inferred pedigree assigned 9238 dams, while our reference pedigree contained 8051 assigned dams, resulting in 7.33% more information in the inferred pedigree. For trio data,



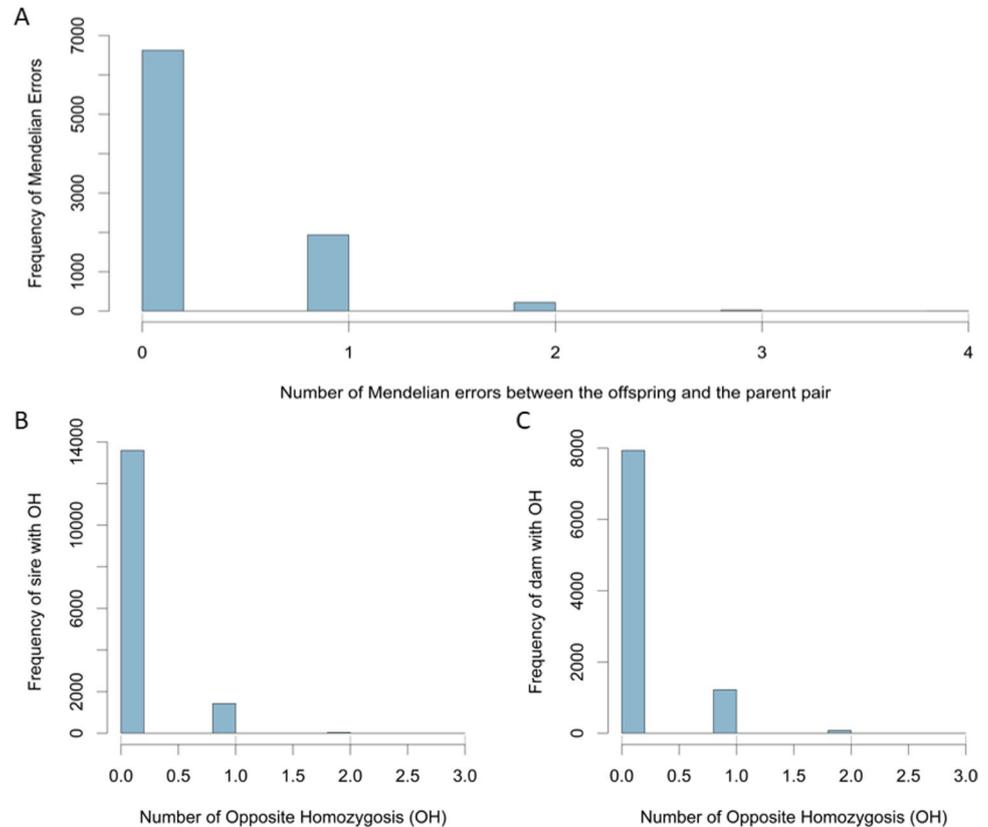
**Fig. 1** Venn diagram with common SNPs to the four microarrays, **A** before quality control and **B** after quality control. Bovine genome version ARS-UCD1.2 was used. GGPi 35 K, GeneSeek Genomic Profiler Indicus (GGPi) with 35 K (Geneseek, Lincoln, NE); GGPI 50 K,

GGPi BeadChip with 50 K; GGPI-LD, GGPI-LD BeadChip with 27 K; and HD, Illumina Bovine HD with 777 K (Illumina, San Diego, CA)

**Fig. 2** **A** Log10-Likelihood Ratio (LLR) for parental pair versus the next most likely configuration between the three individuals (with one or neither parent assigned). **B** LLR of the male being sire versus the next most likely relationship between the focal individual and this male. **C** Same as previous, but for female parents



**Fig. 3** **A** number of Mendelian errors between offspring and pair of parents and **B, C** number of loci in which the offspring presented opposite homozygous loci with the sire and dam, respectively



the reference pedigree contained 7045 information, while we assigned 8799 sire-offspring-dam in the inferred pedigree.

Information, or lack of information, from one or both parents, was not identical in only 16.02% of the 16,205 animals. Comparing information from reference and inferred pedigrees, 90.01%, 92.65%, and 83.98% are respectively coincident for sires, dams, and trios (sire-offspring-dam).

## Genomic relationship coefficient

We also constructed a categorical relationship matrix, containing up to two generations, with dimension equal to  $N \times N$  (where  $N$  = number of animals) from the inferred pedigree file (Fig. 4). This analysis was performed with the 1810 SNPs common to the four microarrays.

Descriptive statistics of the number of assignments made in our inferred pedigree, genomic relationship values, and

pedigree-based expected relationship coefficient values for each category are described in Table 2.

## Linkage disequilibrium and effective population size

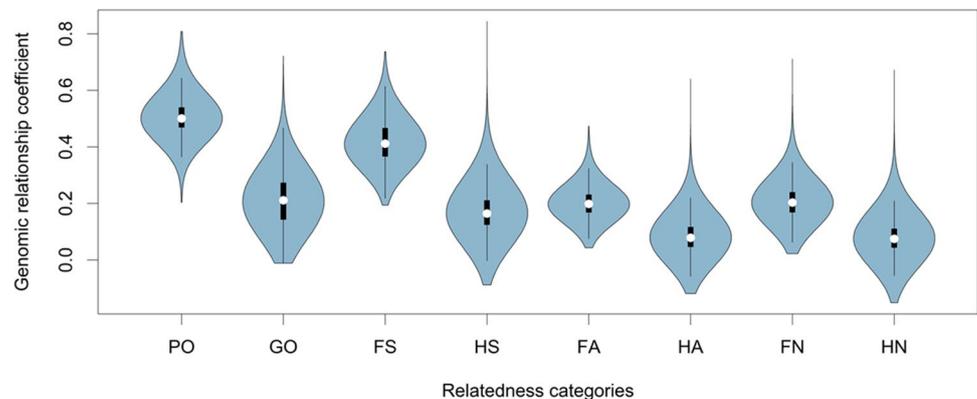
Imputed genotype data for 21,656 animals was used for population structure analyses. LD decay pattern in Gir breed was created based on estimated  $r^2$  measurements, using the equation described by Sved (1971), and can be seen at Fig. 5.

Estimated  $r^2$  between pairs of markers equals to 0.208 for 50 Kb distance. Average  $r^2$  value, observed at distances less than 10 Kb, equals to  $0.37 \pm 0.35$ . Average  $r^2$  values decreased with increasing distance, as well as the percentage of  $r^2$  greater than 0.3 (Online Resource 3). The highest average  $r^2$  value was equal to  $0.40 \pm 0.36$ , and the lowest value was equal to  $0.06 \pm 0.09$ , at distances of 0–5 Kb and 900–1000 Kb, respectively. The  $r^2$  values differed among the autosomes (Online Resource 4).

**Table 1** Comparison between reference (Ref) and inferred (Inf) pedigree and total information contained in reference and inferred pedigrees for the 16,205 Gir animals

	Pedigree compare				Total information			
	Ref <sub>only</sub>	Inf <sub>only</sub>	Match	Mismatch	Info <sub>Ref</sub>	Info <sub>Inf</sub>	% Info <sub>Ref</sub>	%Info <sub>Inf</sub>
Sire-offspring	6	1.612	13.453	1	13.460	15.066	83.06	92.97
Dam-offspring	2	1.189	8.049	0	8.051	9.238	49.68	57.01
Sire-offspring-dam	3	1.757	7.042	0	7.045	8.799	43.47	54.30

**Fig. 4** Coefficient of genomic relationship values for each one of the eight analyzed categories: PO = parent-offspring; GO = grand-offspring; FS = full sibling; HS = half sibling; FA = full avuncular; HA = half avuncular; FN = full nephew/niece; and HN = half nephew/niece

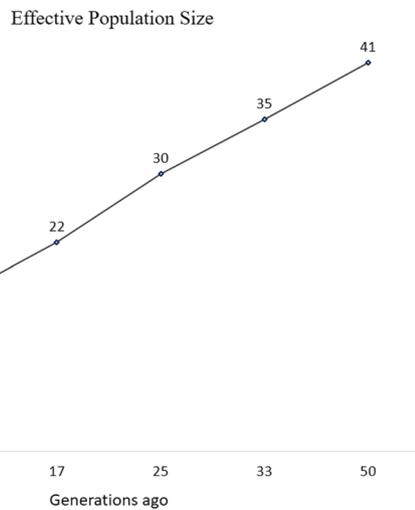
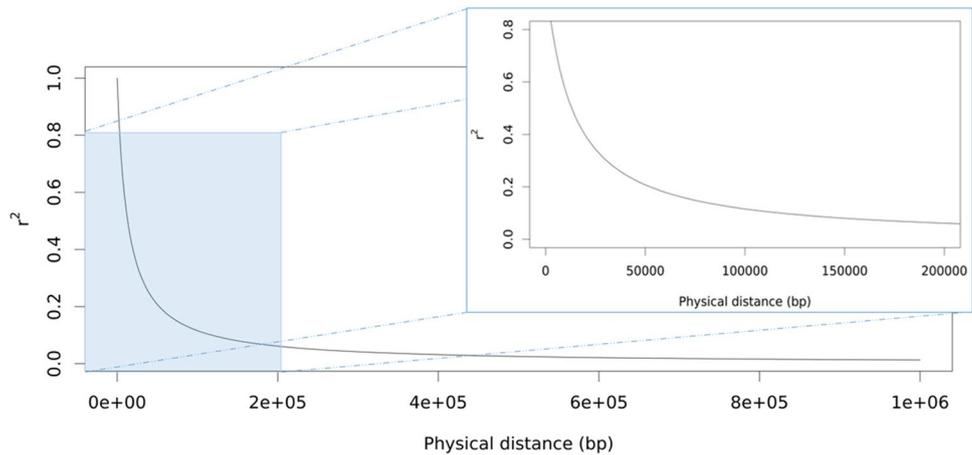


**Table 2** Descriptive statistics, regarding number of assignments ( $N_{\text{assig}}$ ), genomic relationship coefficients ( $r_{\text{GEN}}$ ), and pedigree-based relationship coefficients ( $r_{\text{PED}}$ ) in Gir cattle according to each relationship category

Relationship categories	Acronym	$N_{\text{assig}}$	$r_{\text{GEN}} \text{ min}$	$r_{\text{GEN}} \pm \text{sd}^a$	$r_{\text{GEN}} \text{ max}$	$r_{\text{PED}}$
Parent-offspring	PO	24,304	0.203	$0.508 \pm 0.0589$	0.808	0.500
Grand-offspring	GO	27,802	-0.011	$0.211 \pm 0.088$	0.721	0.250
Full sibling	FS	27,910	0.194	$0.419 \pm 0.0738$	0.737	0.500
Half sibling	HS	16,879,484	-0.088	$0.171 \pm 0.0663$	0.844	0.250
Full avuncular	FA	21,933	0.043	$0.202 \pm 0.0506$	0.474	0.250
Half avuncular	HA	16,924,939	-0.119	$0.084 \pm 0.0525$	0.640	0.125
Full nephew/niece	FN	21,933	0.022	$0.207 \pm 0.0541$	0.711	0.250
Half nephew/niece	HN	16,197,925	-0.151	$0.080 \pm 0.0509$	0.673	0.125

<sup>a</sup> $r_{\text{GEN}} \pm \text{sd}$  = average  $\pm$  standard deviation.

**Fig. 5** Linkage disequilibrium decay in Gir cattle



**Fig. 6** Estimated effective population size, based on LD, in Gir cattle in the last fifty generations

The estimated Gir cattle population  $N_e$  for the previous 50 generations is illustrated in Fig. 6.  $N_e$  estimated was higher in generation 50, with 41 animals, and has declined dramatically over generations, in which the last generation (5) presented 11 animals. Analyzing from the older to the most recent generation,  $N_e$  estimates for Gir cattle show a decreasing trend. Mean generation interval was calculated in 9.08 years.

**Runs of homozygosity and inbreeding coefficient**

A total of 1,156,117 ROHs were found among the 21,656 Gir animals using 416,155 markers. Per animal  $53 \pm 9.0$  ROHs were detected, with a maximum number of 284 and

**Table 3** Runs of homozygosity (ROH) statistics, number and length (in Mb) by length class (ROH<sub>1–2 Mb</sub>, ROH<sub>2–4 Mb</sub>, ROH<sub>4–8 Mb</sub>, ROH<sub>8–16 Mb</sub>, and ROH<sub>>16 Mb</sub>)

Class	$N_{ROH}^a$	Percentage (%)	$L_{ROH}$ (Mb) <sup>b</sup>	$S_{ROH}^c$	$S_{GCOverage}^d$ (%)
ROH <sub>1–2 Mb</sub>	709,782	61.39	1.18	32.78	1.56%
ROH <sub>2–4 Mb</sub>	246,853	21.35	2.77	11.40	1.27%
ROH <sub>4–8 Mb</sub>	116,471	10.08	5.53	5.38	1.20%
ROH <sub>8–16 Mb</sub>	56,489	4.89	11.14	2.82	1.26%
ROH <sub>&gt;16 Mb</sub>	26,522	2.29	25.68	2.16	2.16%

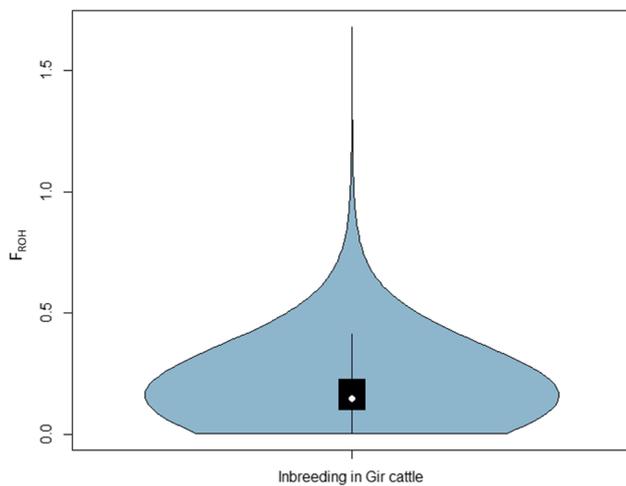
<sup>a</sup> $N_{ROH}$ =runs of homozygosity number, from 21,656 Gir animals. <sup>b</sup> $L_{ROH}$ =average length of ROHs. <sup>c</sup> $S_{ROH}$ =average ROH number per animal. <sup>d</sup> $S_{GCOverage}$ =genome ROH coverage per animal.

a minimum of 6 ROHs. ROH length per individual was  $3.24 \text{ Mb} \pm 1.06 \text{ Mb}$ . A total of 12,272 animals presented ROHs in the class of > 16 Mb, with an average of 2 ROHs per animal in this class, in which minimum and maximum equal to 1 and 19 ROHs, respectively (Table 3).

Distribution and mean length, as well as ROH coverage, by chromosome is shown in Online Resource 5 and Online Resource 6, respectively.

The extent and frequency of ROH are used to infer ancestry at individual and breed levels. Inbreeding, measured as the proportion of ROH-covered genome, results in an increase of homozygosity. We found, in our work, a mean  $F_{ROH}$  of  $0.19 \pm 0.13$  for Gir cattle population, the distribution can be seen in Fig. 7.

To differentiate old and recent inbreeding, we calculate  $F_{ROH}$  based on five different ROH length classes as shown in Table 4. We consider  $F_{ROH}$  of 1–2 Mb in length equal to  $\approx 50$  generations, 2–4 Mb  $\approx 25$  generations, 4–8 Mb  $\approx 12$  generations, 8–16 Mb  $\approx 6$  generations, and 16 Mb  $\approx 3$  generations (Peripolli et al. 2018). Inbreeding showed large variation among chromosomes, both in relation to mean values and level of inbreeding per individual (Fig. 8).



**Fig. 7** Inbreeding coefficient based on runs of homozygosity ( $F_{ROH}$ ) in Gir cattle

## Discussion

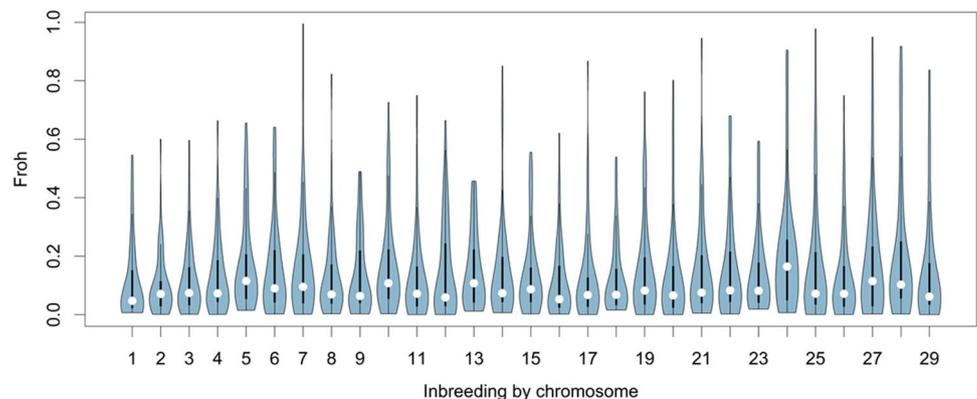
Commercial microarrays, from different platforms, are currently used in livestock genotyping. The most popular microarray in cattle were created initially on *Bos taurus* European genome information, which can result in bias when microarrays are used in distant breeds (Lachance and Tishkoff 2013) or in *Bos indicus* animals. Specific microarrays are developed for *Bos indicus* cattle, such as GeneSeek genomic Profiler Indicus (GGPi) (GeneSeek Inc., Lincoln, NE, USA).

**Table 4** Descriptive statistics of the genomic inbreeding coefficients based on runs of homozygosity ( $F_{ROH}$ ) in different lengths ( $F_{ROH1-2 Mb}$ ,  $F_{ROH2-4 Mb}$ ,  $F_{ROH4-8 Mb}$ ,  $F_{ROH8-16 Mb}$ , and  $F_{ROH > 16 Mb}$ ) for genotyped Gir animals (N)

Class	N	Mean $\pm$ sd <sup>a</sup>	Median	Minimum	Maximum	CV <sup>b</sup>
$F_{ROH1-2 Mb}$	21,656	0.065 $\pm$ 0.029	0.057	0.003	0.372	44.43%
$F_{ROH2-4 Mb}$	21,655	0.049 $\pm$ 0.028	0.041	0.001	0.361	58.23%
$F_{ROH4-8 Mb}$	21,630	0.036 $\pm$ 0.028	0.028	0.002	0.348	76.15%
$F_{ROH8-16 Mb}$	20,014	0.026 $\pm$ 0.025	0.018	0.003	0.334	94.91%
$F_{ROH > 16 Mb}$	12,272	0.022 $\pm$ 0.021	0.015	0.006	0.281	94.56%

<sup>a</sup>sd = standard deviation. <sup>b</sup>CV = coefficient of variation.

**Fig. 8** Distribution of inbreeding coefficient based on runs of homozygosity ( $F_{roh}$ ) by chromosome in Gir cattle



The panel proposed by ISAG for parentage testing consists of 100 markers derived from European taurine breeds, and additional 100 markers from indicine and synthetic breeds (ISAG 2012). In our study, only 46 SNPs were informative, indicating that ISAG panel is not fully representative in all breeds and populations. Strucken et al. (2014) in a study with Hanwoo cattle, which, despite being a taurine breed, differs significantly from European taurines, showed that breed-specific marker panels perform better for parentage testing than the panel established by ISAG.

Kopps et al. (2015) conducted a study testing type and quantity of markers required for high accuracy in parentage tests, concluding this vary according to relationship categories, mating system, and number of overlapping generations. There are few studies reporting parentage testing in cattle, and especially for indicine breeds, there is a lack of studies regarding pedigree analysis tackling genomic relationship. This point must be overcome, since indicine breeds are essential resources in tropical countries.

## Parentage analysis and pedigree reconstruction

Opposing homozygotes occur if one parent is homozygous for one allele, the other parent is homozygous for the other allele, and the offspring is not heterozygous. In true parent–offspring relationships, there should be no Mendelian inconsistencies and no locus with OH should occur between both parents, except for mutation events or genotyping errors.

Even if two individuals are considered unrelated, according to the pedigree, it is possible to obtain genetic information supporting the hypothesis that they share identical alleles; markers might detect incorrect relationships in a pedigree, when these are monomorphic. The uncertainty is due to the random nature of allelic inheritance from parents to offspring during meiosis, and whether alleles are identical by descent or state, but the use of likelihood ratio allows increased information about relationship between individuals to be extracted from observed genotypes.

### Genomic relationship coefficient

Unlike the relationship matrix calculated using pedigree information, which contains only positive values, genomic relationship matrix can also contain negative values, since these values are standardized correlations (VanRaden 2008). The explanation is that there is an average relationship inside each population and there are animals presenting values above or below the average. Animals with divergent genotypes will show negative values (Legarra et al. 2018).

When we compared in our study, the genomic relationship coefficients, and the average relationship coefficient values expected for each class, the calculated values are lower in some classes. This fact can be explained by the inbreeding control that Gir breed has had over the last few years, which is supported by the inbreeding coefficient calculated for the  $ROH_{>16\text{ Mb}}$  class, which represents the most recent inbreeding. Therefore, when compared to the population mean, we found a lower than expected degree of homozygosity.

### Linkage disequilibrium and effective population size

O'BRIEN et al. (2014) evaluated taurine and indicine breeds, including Gir cattle, and observed that lower levels of  $r^2$  were found for indicine breeds. Within indicine breeds, the levels of  $r^2$  at a distance of 0–1 Kb is approximately 0.55, in which Gir breed presented a  $r^2$  average equal to 0.3 with only 16.5 Kb, which indicates a rapid decay of the LD curve (O'Brien et al. 2014). In our study, the average value of  $r^2$  observed at 0–1 Kb was  $0.387 \pm 0.3566$ .

The LD decay graph showed a  $r^2 = 0.3$  at a distance of 29 Kb, similar to taurine breeds at 33 Kb (O'Brien et al. 2014). Gir breed presented an estimated  $r^2$  between pairs of markers equal to 0.208 at 50 Kb, showing a fast decay. Porto-Neto et al. (2014) evaluated LD decay for taurine, indicine and crossbreeds animals. They noted that decay was rapid, with  $r^2$  below 0.2 for pairs of markers separated by 50 Kb for all breeds, but at distances less than 10 Kb, taurine breeds showed higher LD ( $r^2 = 0.45$ ) than indicine breeds ( $r^2 = 0.25$ ) and crossbreeds ( $r^2 = 0.32$ ).

Our average value of  $r^2 = 0.376 \pm 0.3538$ , observed at distances of less than 10 Kb, is higher than found by

Porto-Neto et al. (2014) for indicine breeds. This higher LD value can be explained by the smaller effective population size in our population, which possibly reflects a stronger bottleneck during breed formation. Considering greater distances (100–1000 Kb), the values found by O'Brien et al. (2014), for indicines in a high-density dataset, were similar to ours, with an observed  $r^2$  mean approximately from 0.15 to 100 Kb, and less than 0.1 to 1 Mb, showing smooth decay at greater distances.

Differences in LD levels of indicine breeds can be attributed to historical population events, including mixing with wild-type ancestors (Murray et al. 2010), lower selection pressure (Thévenon et al. 2007) and larger effective population sizes (Ospina et al. 2019).  $N_e$ , according genomic data, can be estimated based on LD standards (Barbato et al. 2015; Jasielczuk et al. 2020). The strength of LD among loci at different genetic distances allowed us to infer about the effective size of the ancestral population;  $N_e$  estimated in Gir cattle population for the previous 50 generations is illustrated in Fig. 6.

$N_e$  estimates show a decreasing trend over generations, analyzing from older to the most recent generation. Malhado et al. (2010) reported  $N_e$ , based on the sex ratio, for Gir raised in Northeastern Brazil estimated at 100 to 150, in the period from 2002 to 2006, corroborating with Reis-Filho et al. (2010) and Santana Jr et al. (2014), who described similar values of  $N_e$ , based on inbreeding coefficients. The current Gir population presented a much smaller effective size than that found by these authors. The difference was possibly due to the methodology, since unlike the others, our information come from molecular markers, resulting in more accurate values.

The mean generation interval (9.08 years) was higher than described by Malhado et al. (2010), 7.9 years, and Reis-Filho et al. (2010), 8.41 years. Longer generation intervals result in reduced annual genetic gain and consequently lower economic returns. Generally, long generation intervals are due to the longevity of the Gir breed, in which animals are usually kept for reproduction until older ages, in addition to the continuous use of specific bulls, without replacements. Nowadays, considering the genomic approach in breeding programs, younger bulls are being included in progeny tests, fastening genetic gain, and replacement of the proven older sires, decreasing generation interval.

### Runs of homozygosity and inbreeding coefficient

Santana Jr. et al. (2014) and Reis-Filho et al. (2010), analyzing the pedigree based inbreeding coefficient ( $F_{PED}$ ) in Gir cattle, found values of 0.0192 and 0.0280, respectively. These studies described inbreeding values close to those found by us at distances above 8 Mb ( $F_{ROH8-16\text{ Mb}} = 0,026 \pm 0,025$ , e  $F_{ROH > 16\text{ Mb}} = 0,022 \pm 0,021$ ). This was explained by Zavarez

et al. (2015), who observed that incomplete pedigree does not capture remote inbreeding and estimates based on pedigree information are comparable only with the  $F_{ROH}$  calculated based on long ROH lengths. The variation between these two estimates occurs since  $F_{PED}$  does not assume selection and genetic recombination events; therefore, it does not consider the potential bias of these events (Peripolli et al. 2018). We emphasize that pedigree relationship is obtained from statistical expectations of the probable genomic proportion of identical by descent alleles, while estimates based on genotypes use the information of identical by state alleles, resulting in greater accuracy (Visscher et al. 2006).

Long ROHs are indicative of recent inbreeding, while short ROHs capture ancient inbreeding derived from older ancestors and detect population bottlenecks. Under the assumption that 1 cM is equivalent to approximately 1 Mb, ROH can be separated into length classes to express different points in the time at each inbreeding occurred (Curik et al. 2014). To differentiate old and recent inbreeding, we calculate  $F_{ROH}$  based on five different ROH length classes (Table 4). Inbreeding events that took place  $\approx 25$  generations ago were presented by the  $F_{ROH}$  calculated by length below or equal to 2 Mb. The degree of inbreeding based on the longest ROH class presented mean inbreeding equal to  $0.022 \pm 0.021$ , reflecting inbreeding events occurring  $< 3$  generations ago.

The Gir population used in this study showed a reduction in the  $N_e$  based on LD, when calculated over the last 50 generations, suggesting a recent inbreeding, which is supported by the occurrence of long ROH segments. Inbreeding levels, especially those based on short ROH segments, were moderate to high, suggesting not only a small  $N_e$  but also existence of bottlenecks in the genome of Gir cattle.

## Conclusions

The panel with 1810 SNPs was satisfactory for parentage attribution, in which the number of attributions was higher than the number of information already contained in the reference pedigree, with a probability of confidence greater than 99,995%. The methodology added information that was not in the pedigree of the breeders' association, suggesting that when there is a large number of genotyped individuals, the approach is advantageous. For Gir breed, due to the ease and speed of execution and interpretation, confirmation of parentage by the exclusion method would still be the most suitable method for routine analysis. The relationship coefficients allowed us to evaluate the distribution of values for each relationship category of the inferred pedigree, showing appropriate values for each category.

Regarding population structure, which we used the imputed genotypes, considering the  $N_e$  decrease based on

LD over the past 50 generations, we recommend monitoring and conservation measures to minimize loss of genetic diversity in the breed. We suggest the application of breeding strategies to minimize inbreeding and avoid the use of the most used bulls with high genetic value to maintain genetic variability in future generations. We also recommend increasing the effective population size and reducing generation interval to maximize genetic progress.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13353-023-00747-x>.

**Acknowledgements** We thank Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and the Ministério da Ciência Tecnologia e Inovação (MCTI)/Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)/Instituto Nacional de Ciência e Tecnologia—Ciência Animal (INCTCA) for their financial support and Embrapa Dairy Cattle Research Center and Embrapa Digital Agriculture (CNPTIA) for providing the data and digital means necessary to carry out this study.

**Author contribution** All authors contributed to the study conception. Material preparation and data collection were performed by João Cláudio do Carmo Panetto, Marcos Vinicius Gualberto Barbosa da Silva, Marco Antônio Machado, and Rui da Silva Verneque. Material preparation and data analysis were performed by Arielly Oliveira Garcia and Pamela Itajara Otto. Resources, conceptualization, and supervision were performed by Simone Eliza Facioni Guimarães. Editing and data analysis were performed by Luiz Afonso Glatzi Junior, Renata de Fátima Bretanha Rocha, Mateus Guimarães dos Santos, and Daniele Alves de Oliveira. The first draft of the manuscript was written by Arielly Oliveira Garcia, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Proex 1246/2020). Author AOG has received research support from Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). Author MGS has received research support from CAPES. Author RFBR has received research support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

**Data Availability** The data sets analysed in the current study might be available under reasonable request.

## Declarations

**Ethics approval** This study did not make direct use of animals; a database from the breeders' association was used. The Universidade Federal de Vicosa Research Ethics Committee has confirmed that no ethical approval is required.

**Competing interests** The authors declare no competing interests.

## References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full

- pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 93(2):743–752. <https://doi.org/10.3168/jds.2009-2730>
- Aguilar I, Misztal I, Tsuruta S, Legarra A, Wang H (2014) PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Pages 1–4 in World Congress on Genetics Applied to Livestock Production (WCGALP). American Society of Animal Science, Vancouver, Canada.
- Barbato M, Orozco-terWengel P, Tapio M, Bruford MW (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet* 6:109. <https://doi.org/10.3389/fgene.2015.00109>
- Biscarini F, Cozzi P, Gaspa G, Marras G (2019) *detectRUNS*: detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.6. <https://CRAN.R-project.org/package=detectRUNS>.
- Calus MP, Mulder HA, Bastiaansen JW (2011) Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. *Genet Sel Evol* 43:1–12. <https://doi.org/10.1186/1297-9686-43-34>
- Clayton D (2019) *snpStats*: SnpMatrix and XSNpMatrix classes and methods. R package version 1.36.0. <https://doi.org/10.18129/B9.bioc.snpStats>.
- Curik I, Ferenčaković M, Sölkner J (2014) Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest Sci* 166:26–34. <https://doi.org/10.1016/j.livsci.2014.05.034>
- FAO (1992) Monitoring animal genetic resources and criteria for prioritization of breeds. K. Majjala. In: Hodges, J. The management of global animal genetic resources, Proceedings of an FAO Expert Consultation, Rome, N° 104.
- FAO. World Food and Agriculture - Statistical (2021) Available in <<http://www.fao.org/faostat/en/#data/QL>>. Accessed May 7, 2021.
- Flanagan SP, Jones AG (2019) The future of parentage analysis: from microsatellites to SNPs and beyond. *Mol Ecol* 28:544–567. <https://doi.org/10.1111/mec.14988>
- GeneBank. Assembly ARS-UCD1.2 (2018) Available in <[https://www.ncbi.nlm.nih.gov/assembly/GCA\\_002263795.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2)>. Accessed in May 8, 2021.
- Grashei KE, Ødegård J, Meuwissen TH (2018) Using genomic relationship likelihood for parentage assignment. *Genet Sel Evol* 50:1–11. <https://doi.org/10.1186/s12711-018-0397-7>
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13:635–643. <https://doi.org/10.1101/gr.387103>
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231. <https://doi.org/10.1007/BF01245622>
- Howrigan DP, Simonson MA, Keller MC (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:1–15. <https://doi.org/10.1186/1471-2164-12-460>
- Huisman J (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Mol Ecol Resour.* 17:1009–1024. R package version 207. <https://doi.org/10.1111/1755-0998.12665>.
- International Society for Animal Genetics (ISAG) (2012) Cattle molecular markers and parentage testing committee. <http://www.isag.us/committees.asp?autotry=true&ULnotkn=true>. Accessed 10 Feb 2021.
- Jasielczuk I, Gurgul A, Szmatoła T, Semik-Gurgul E, Pawlina-Tyszko K, Szyndler-Nędzka M, Blicharski T, Szulc K, Skrzypczak E, Bugno-Poniewierska M (2020) Comparison of linkage disequilibrium, effective population size and haplotype blocks in Polish Landrace and Polish native pig populations. *Livest Sci* 231:103887. <https://doi.org/10.1016/j.livsci.2019.103887>
- Kopps AM, Kang J, Sherwin WB, Palsboll PJ (2015) How well do molecular and pedigree relatedness correspond, in populations with diverse mating systems, and various types and quantities of molecular and demographic data? *G3 Genes Genomes Genetics* 5:1815–1826. <https://doi.org/10.1534/g3.115.019323>
- Lachance J, Tishkoff SA (2013) SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *BioEssays* 35:780–786. <https://doi.org/10.1002/bies.201300014>
- Legarra A, Lourenco DA, Vitezica Z (2018) Bases for genomic prediction. Paris, França: INRA, Animal Genetics Department.
- Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E, Danchin-Burge C (2013) Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genet Sel Evol* 45:1–10. <https://doi.org/10.1186/1297-9686-45-1>
- Lu D, Sargolzaei M, Kelly M, Li C, Vander Voort G, Wang Z, Plastow G, Moore S, Miller S (2012) Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Front Genet* 3:152. <https://doi.org/10.3389/fgene.2012.00152>
- Malhado CHM, Carneiro PLS, Malhado ACM, Martins JAM, Martins Filho R, Bozzi R (2010) History of registered Gir breed in Brazilian Northeast: population structure and genetic improvement of growth traits. *Cienc Rural* 40:1385–1391. <https://doi.org/10.1590/S0103-84782010000600023>
- Marshall TC, Slate JBKE, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639–655. <https://doi.org/10.1046/j.1365-294x.1998.00374.x>
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppeters W, Crews D, Dias-Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassel CP, Williams JL, Taylor JF, Moore SS (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genet* 8:74. <https://doi.org/10.1186/1471-2156-8-74>
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, MacLeod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF (2008) Runs of homozygosity in European populations. *Am J Hum Genet* 83:359–372. <https://doi.org/10.1016/j.ajhg.2008.08.007>
- Montgomery GW, Campbell MJ, Dickson P, Herbert S, Siemerling K, Ewen-White KR, Visscher PM, Martin NG (2005) Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. *Twin Res Hum Genet* 8:346–352. <https://doi.org/10.1375/twin.8.4.346>
- Murray C, Huerta-Sanchez E, Casey F, Bradley DG (2010) Cattle demographic history modelled from autosomal sequence variation. *Philosophical Transactions of the Royal Society b: Biological Sciences* 365:2531–2539. <https://doi.org/10.1098/rstb.2010.0103>
- O'Brien AMP, Mészáros G, Utsunomiya YT, Sonstegard TS, Garcia JF, Van Tassel CP, Carvalheiro R, da Silva MVB, Sölkner J (2014) Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest Sci* 166:121–132. <https://doi.org/10.1016/j.livsci.2014.05.007>
- Ospina AMT, Maiorano AM, Curi RA, Pereira GL, Zerlotti-Mercadante ME, Gonçalves Cyrillo JN, dos S, Aspilcueta-Borquis RR, Silva JAI de V, (2019) Linkage disequilibrium and effective population size in Gir cattle selected for yearling weight. *Reprod in Domestic Anim* 54:1524–1531. <https://doi.org/10.1111/rda.13559>
- Panetto JC, do C, Machado MA, da Silva MVG, Barbosa RS, dos Santos GG, Leite R de MH, Peixoto MGC, (2017) Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. *Livest Sci* 204:33–38. <https://doi.org/10.1016/j.livsci.2017.08.008>

- Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, Panetodo C, Ventura RV, Baldi F, da Silva MVGB, JC (2018) Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gir (*Bos indicus*) dairy cattle. *BMC Genomics* 19:1–13. <https://doi.org/10.1186/s12864-017-4365-3>
- Porto-Neto LR, Kijas JW, Reverter A (2014) The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet Sel Evol* 46:1–5. <https://doi.org/10.1186/1297-9686-46-22>
- Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet* 11:800–805. <https://doi.org/10.1038/nrg2865>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Reis-Filho JC, Lopes PS, Verneque RDS, Torres RDA, Teodoro RL, Carneiro PLS (2010) Population structure of Brazilian Gir dairy cattle. *Rev Bras Zootec* 39:2640–2645. <https://doi.org/10.1590/S1516-35982010001200012>
- Santana ML Jr, Pereira RJ, Bignardi AB, El Faro L, Tonhati H, Albuquerque LG (2014) History, structure, and genetic diversity of Brazilian Gir cattle. *Livest Sci* 163:26–33. <https://doi.org/10.1016/j.livsci.2014.02.007>
- Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. <https://doi.org/10.1186/1471-2164-15-478>
- Saunders IW, Brohede J, Hannan GN (2007) Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. *Genomics* 90:291–296. <https://doi.org/10.1016/j.ygeno.2007.05.011>
- Strucken EM, Gudex B, Ferdosi MH, Lee HK, Song KD, Gibson JP, Kelly M, Piper EK, Porto-Neto LR, Lee SH, Gondro C (2014) Performance of different SNP panels for parentage testing in two East Asian cattle breeds. *Anim Genet* 45(4):572–575. <https://doi.org/10.1111/age.12154>
- Strucken EM, Lee SH, Lee HK, Song KD, Gibson JP, Gondro C (2016) How many markers are enough? Factors influencing parentage testing in different livestock populations. *J of Ani Breed Genet* 133:13–23. <https://doi.org/10.1111/jbg.12179>
- Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol* 2:125–141. [https://doi.org/10.1016/0040-5809\(71\)90011-6](https://doi.org/10.1016/0040-5809(71)90011-6)
- Thévenon S, Dayo GK, Sylla S, Sidibé I, Berthier D, Legros H, Boichard D, Eggen A, Gautier M (2007) The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. *Anim Genet* 38:277–286. <https://doi.org/10.1111/j.1365-2052.2007.01601.x>
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2:0316–0325. <https://doi.org/10.1371/journal.pgen.0020041>
- Wellmann R (2021) optiSel: optimum contribution selection and population genetics. R package version 2.0.5. <https://CRAN.R-project.org/package=optiSel>
- Wiggans GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF, Schenkel FS, Van Tassell CP (2009) Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J Dairy Sci* 92:3431–3436. <https://doi.org/10.3168/jds.2008-1758>
- Wiggans GR, VanRaden PM, Bacheller LR, Tooker ME, Hutchison JL, Cooper T, Sonstegard TS (2010) Selection and management of DNA markers for use in genomic evaluation. *J Dairy Sci* 93:2287–2292. <https://doi.org/10.3168/jds.2009-2773>
- Zavarez LB, Utsunomiya YT, Carmo AS, Neves HH, Carneiro R, Ferenčaković M, O'Brien AMP, Curik I, Cole JB, Van Tassell CP, da Silva MVGB, Sonstegard TS, Sölkner J, Garcia JF (2015) Assessment of autozygosity in Nelore cows (*Bos indicus*) through high-density SNP genotypes. *Front Genet* 6:5. <https://doi.org/10.3389/fgene.2015.00005>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.