Livestock Science xxx (2016) xxx-xxx

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com

Retelling the recent evolution of genetic diversity for Guzerá: Inferences from LD decay, runs of homozygosity and Ne over the generations

Pablo Augusto de Souza Fonseca [a], Fernanda Caroline dos Santos [a], Izinara Cruz Rosse [a], Ricardo Vieira Ventura [b, c], Frank Ângelo Tomita Brunelli [d], Vânia Maldini Penna [e], Rui da Silva Verneque [d], Marco Antônio Machado [d], Marcos Vinícius Gualberto Barbosa da Silva [d], Maria Raquel Santos Carvalho [a, *], Maria Gabriela Campolina Diniz Peixoto [d]

[a] *Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil*
[b] *Center for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada N1G 2W1*
[c] *Beef Improvement Opportunities, Guelph, ON, Canada N1K 1E5*
[d] *Embrapa Gado de Leite, Juiz de Fora 36038-330, Brazil*
[e] *Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil*

ARTICLE INFO

Article history:
Received 6 April 2016
Received in revised form 28 June 2016
Accepted 12 October 2016
Available online xxx

Keywords:
Genetic diversity
Genome-wide genotyping
Genetic structure
Bovine
Guzerá

ABSTRACT

Genetic diversity is the one of the most important issues in conservation studies of livestock breeds or endangered species. In the present study, we tested the feasibility of describing the recent evolution in genetic diversity through genome-wide SNP genotyping and estimates of linkage disequilibrium decay patterns, effective population size, inbreeding coefficient based on runs of homozygosity and population structure. We choose the bovine indicine breed Guzerá because it has suffered recent bottlenecks which have been registered historically. A sample of 1036 females was genotyped using Illumina BovineSNP50. A resampling strategy was applied to correct for sampling biases caused by the population structure in herds, and by the extensive use of some sires for artificial reproduction. A subsample of 210 animals and 32,806 markers with MAF>0.01 was used. Very low linkage disequilibrium was detected for distances greater than 120 Kb between two markers. Furthermore, three points of decrease in effective population size between generations were detected, which coincide with the historically registered bottlenecks. The inbreeding coefficient, based on runs of homozygosity, confirmed a strong contribution of the last 20–30 generations to current inbreeding. In the population structure analysis, the most probable number of sub-populations is 2, reflecting selection purpose (beef or dual-purpose). Taken together, these results allow a retelling of the recent evolution of this breed. The strategy described here will be useful for other breeds or even species for which a careful historical registry is not available for conservation proposals.

1. Introduction

Studies aiming at the conservation of genetic diversity provide the best approaches to selection strategies for the long time survival of a breed or population under artificial selection. Monitoring of genetic diversity is fundamental in any artificial selection process for any population. Over time, some breeds have been moved to different parts of the world, some have disappeared, and some new breeds have been established.

Many studies have established tools for analyzing the evolution of genetic diversity over long periods of time. However, the efficiency of such tools for describing the recent evolution of genetic diversity is a topic that appears less frequently in literature. In Brazil, the establishment of new breeds derived from the Zebu took place mostly since 1860 and these newly established breeds experienced many fluctuations in population size. These fluctuations offered interesting models for evaluating the efficiency of these tools to describe the recent evolution in genetic diversity.

In studies based on microsatellites, a genomic scale estimate of genetic diversity is not possible due to the low number of individuals and loci evaluated. Studies based on pedigree data generally are based on the coefficient developed by Wright (1922). Studies of inbreeding based on pedigree data (F_{PED}) have failed to detect the influence of the founder population on a current population due to the assumption of a relationship between them. Moreover, these studies do not take into account the stochastic nature of recombination, sometimes resulting in research results that are difficult to interpret. When evaluating the genetic diversity of animals, selected based on pedigree data, there are two aspects that should be pointed out: (1) pedigree errors due to misidentification, misinterpretation and wrong registry are common (Ron et al., 1996; Carneiro et al., 1999), thereby reducing selection speed; (2) F_{PED} based studies assume that the entire

* Corresponding author.
Email addresses: pasf2009@ufmg.br (P.A. de Souza Fonseca); fernandacds1990@ufmg.br (F.C. dos Santos); izinara.rosse@gmail.com (I.C. Rosse); rventura@uoguelph.ca (R.V. Ventura); frank.bruneli@embrapa.br (F.Â.T. Brunelli); cbmg@cbmgguzera.com.br (V.M. Penna); rui.verneque@embrapa.br (R. da Silva Verneque); marco.machado@embrapa.br (M.A. Machado); marcos.vb.silva@embrapa.br (M.V.G.B. da Silva); mraquel@icb.ufmg.br (M.R.S. Carvalho); gabriela.peixoto@embrapa.br (M.G.C.D. Peixoto)

http://dx.doi.org/10.1016/j.livsci.2016.10.006
1871-1413/

genome is under neutral selection and do not take into account the effects of artificial selection which may cause some biases of evaluation.

Several specific tools for assessing the genetic diversity have been developed in recent years. Due to the development of Next-Generation Sequencing techniques, a large number of SNP markers has been obtained and platforms for genotyping thousands of markers distributed over all chromosomes have been developed (Matukumalli et al., 2009). The introduction of such platforms has allowed genome-wide estimates of genetic diversity for the first time.

For example, inbreeding coefficients derived from runs of homozygosity (ROH) are good measures to determine the stretches of autozygosity over the genome, as well as the effects of inbreeding (Keller et al., 2011). This approach can be very useful to estimate the inbreeding coefficient, even in very distant generations. This estimate could be very inaccurate using approaches based on pedigree and microsatellites.

In addition, linkage disequilibrium (LD) is a very important tool in association studies as well as studies aiming to evaluate genetic diversity. Thus, LD has been used in several studies to determine the diversity and history, signatures of selection, recombination rates, effective population size and other population events in cattle or other species (Waples and Do, 2010). Currently, the majority of the markers used for genomic selection are derived from taurine sequences. However, the pattern of LD is different between taurine and indicine breeds (Villa-Angulo et al., 2009; O'Brien et al., 2014) and, therefore, results obtained in taurine breeds cannot be applied directly to indicine breeds. Therefore, the measures of whole-genome LD in indicine breeds are important to ensure the effectiveness of breeding programs that use or aim to use genomic selection.

The relationship between the inter-marker distance in Morgans, c, the r^2 and the effective population size (Ne), in the absence of mutation, allows the determination of Ne through LD measures. Along with Ne, the estimates of population genetic structure assist in the analyses of genetic variability observed in a population over a retrospective viewpoint, enabling the prediction of genetic diversity loss and the survival of small populations (Wang, 2005).

Despite the advantages of the genome-wide approaches to account for the genetic variability of a specific population, there are few delineated strategies for performing whole genome genetic diversity estimates. In the present study, a flowchart of analysis is presented, in order to estimate genetic diversity across recent generations using a genomic approach. The Guzerá breed was used as a model due to the good quality of the historical registry that has been kept of the population. However, this flowchart can be applied to any breed or specie in order to estimate the genomic genetic diversity and the contribution of past generations to current inbreeding levels.

2. Materials and methods

2.1. Sampling, genotyping and quality filtering

Initially, 1036 females from six herds in the two Brazilian cattle breeding programs were genotyped using the Illumina Bovine SNP50 v2 BeadChip (Matukumalli et al., 2009). The Brazilian Guzerá population was formed from a small number of animals imported from India at the end of the 1800s. Currently, in the period from 1939 to 2015 over 400,000 Guzerá births were registered officially; of these, 8115 were in 2015. It is estimated that the registered population is currently around 60,000 head. Today there are about 47 herds based on dual purpose breeding, selected at different levels for both milk production and meat. The criteria for the sample selection were based on: (1) a known pedigree record; (2) the females must be dams from bulls which belong to the National Progeny test or Multiple ovulation embryo transfer (MOET) nucleus of the breed; (3) the females must represent different lineages of the breed; and, (4) the females must have production records from the first calving. A strategy, to correct for sampling biases caused by the population structure in herds and due to the extensive use of some sires for artificial reproduction, was adopted prior to performing the genetic estimates. The initial sample was analyzed using a node selection algorithm based on a network's degree of centrality to resample individuals through a Kinship coefficient obtained by the genomic information (Kehdy et al., 2015). After this analysis, only the animals that presented a Kinship coefficient $\Phi ij \geq 0.1$ to any other individual remained in the final sample. Thus, a sample containing 210 individuals was obtained and the present study was performed using this sample. Only markers that had a known map position, minor allelic frequency (MAF)≥0.01, samples with call rate >0.95 were used for further analysis, amounting to 32,806 markers.

2.1.1. Linkage disequilibrium

The SNP and Variation Suite 7 (SVS7) package (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com) was used to estimate the LD throughout the genome. The haplotype frequencies were assessed using the expectation-maximization (EM) algorithm, formalized by Dempster et al. (1977). The LD pattern was estimated using the r^2 statistic (Formula 1 in Supplementary Box 1) (Hill and Robertson, 1968) as well as the absolute value of D' (Formula 2 in Supplementary Box 1) (Lewontin, 1964). This was done using the LD pairwise analysis algorithm implemented in SVS7, where *freq. AB* and *freq. ab* are the haplotype frequencies of coupling haplotypes and *freq. Ab* and *freq. aB* are the haplotype frequencies of repulsion haplotypes. The allelic frequency of A, a, B and b alleles are represented by *freq. A*, *freq. a*, *freq. B* and *freq. b*, respectively. Where, A, a, B, and b are the alleles of two loci. The mean values of r^2 and |D'| and the proportion of markers with an r^2 and |D'| higher than 0.3 and 0.8 (indicators of high LD), respectively, were calculated for each chromosome. To observe the pattern of LD decay, the means of r^2 and |D'| were also calculated for the following distances between markers (in Kb): 0–1, 1–2, 2–3, 3–4, 4–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, 90–100, 100–120, 120–140, 140–160, 160–180, 180–200, 200–220, 220–250, 250–275, 275–300, 300–350, 350–400, 400–500, 500–600, 600–700, 700–800, 800–900, 900–1000, >1000.

2.2. Effective population size (Ne)

The software PLINK v1.07 (Purcell et al., 2007) was used to calculate the synthetic Pairwise LD across the chromosomes, based on the correlation coefficient (r^2). For all pairs of autosomal SNPs, r^2 measures were calculated using the --r2-ld-window 99999 --ld-window-r2 0 --ld-window-kb 100000 command line. Ne was calculated using the relationship between the distance c, r^2 and Ne, assuming the absence of mutation (Sved, 1971) (Formula 3 in Supplementary Box 1). In order to avoid sampling biases, leading to artifact induced LD, an adjustment of expected r^2 value between to loci in absence of mutation, $E(r^2)$, to account for restricted sample size was used (Formula 4 in Supplementary Box 1) (Weir and Hill, 1980). Ne was estimated $(2c)^{-1}$generations ago, where c is the distance in Morgans between two markers for which LD was estimated.

2.3. Inbreeding coefficient based on runs of homozygosity (F_{ROH})

The SNP and Variation Suite 7 (SVS7) package was used to estimate the ROH across the genome of each individual in the sample, using the 32,806 markers obtained after quality control filtering. The following criteria were used to determine a homozygotic segment as a ROH: minimum number of SNPs equal to 15; number of missing calls allowed equal to 5; number of heterozygous calls allowed equal to 0; maximum gap between consecutive homozygous SNPs equal to 1 Mb; and at least 20 animals sharing the same segment as a ROH.

F_{ROH} was defined as shown below (Formula 5 in Supplementary material 1) (Leutenegger et al., 2003), where the F_{ROH} for the *j* individual is calculated as the sum of the length of k ROH present across the genome divided by the overall length of the genome covered by SNPs (L). The overall length was obtained from the reference UMD 3.1 (http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml) from the taurine genome, excluding the sexual and mitochondrial chromosomes, totaling 2,512,082,506 bp. For each individual, the genomic inbreeding coefficients (F_{ROH}>1 Mb, F_{ROH}>2 Mb, F_{ROH}>8 Mb and F_{ROH}>16 Mb) derived from the ROH with different lengths (>1, >2, >8 and >16 Mb) were calculated, as described by Ferenčaković et al. (2013a). Moreover, to investigate the inbreeding evolution during the recent history of the Guzerá, F_{ROH} were calculated for every even generation, in the interval starting 30 generations ago to 2 generations ago, totaling 15 F_{ROH} calculations. The F_{ROH} for each generation was calculated using the expected length of an autozygous segment in a distribution with mean equal to 1/2 g Morgans, where g is the number of generations since the common ancestor (Howrigan et al., 2011). The genomic inbreeding coefficient (F_{HOM}) was calculated with the SVS7 package, using the difference between observed and expected numbers of homozygous genotypes. The F_{PED} coefficient was calculated using the software Endog 4.8 (Gutiérrez and Goyache, 2005). Moreover, the genomic inbreeding coefficient proposed by VanRaden et al. (2008) was calculated, here called F_G. The cor.test function in R (R Core Team, 2015) was used to estimate the Pearson correlation coefficient between the F_{ROH} X F_{HOM} F_{ROH} X F_{PED}, F_{ROH} X F_G, F_{HOM} X F_{PED}, F_{HOM} X F_G and F_G X F_{PED}.

2.4. Population structure analyses

ADMIXTURE 1.23 (Alexander et al., 2009) was used to evaluate ancestry proportions for K ancestral populations. A 5-fold cross-validation analysis was conducted for the K values from 1 through 10, to examine patterns of ancestry and admixture in our data. The graphic of the proportion of each K for each individual was generated using a script in R (R Core Team, 2015), developed in-house.

3. Results

3.1. Patterns of linkage disequilibrium across the genome

Means, medians and proportions of markers in strong LD (r^2>0.3 and |D'|>0.8) are shown in Table 1. Chromosomes 6 and 7 showed the highest mean r^2 values (0.18), while chromosome 14 showed the highest mean |D'| value (Table 2). However, when the proportion of markers in strong LD was evaluated for each chromosome, chromosome 1 followed by chromosome 6 had the higher proportions of r^2>0.3 and chromosomes 1, 2 and 6 the highest proportions of

Table 1
Linkage disequilibrium through different inter-marker distances for the Guzerá obtained using the SVS7 package.

Distance (Kb)	Number of pairs of markers	r^2 mean	r^2 median	|D'| mean	|D'| median	% r^2>0.3	% |D'|>0.8
0–5	258	0.20(±0.3364)	0.03	0.64(±0.3688)	0.77	20.15	75.58
5–10	131	0.15(±0.1578)	0.01	0.61(±0.4090)	0.40	15.27	46.56
10–15	179	0.16(±0.2913)	0.04	0.72(±0.2871)	0.97	18.44	59.22
15–20	338	0.19(±0.2424)	0.06	0.75(±0.3693)	0.96	22.49	63.91
20–25	3427	0.21(±0.2943)	0.09	0.78(±0.3555)	0.90	24.16	65.28
25–30	3208	0.19(±0.1785)	0.06	0.75(±0.3133)	0.79	21.23	61.22
30–40	5110	0.18(±0.2946)	0.02	0.74(±0.3745)	0.97	19.92	59.86
40–50	3491	0.16(±0.2946)	0.07	0.71(±0.2441)	1.00	17.99	55.17
50–60	2674	0.15(±0.1863)	0.01	0.69(±0.3401)	0.91	16.72	52.69
60–70	2259	0.13(±0.1917)	0.04	0.67(±0.3352)	0.72	14.70	50.24
70–80	1914	0.12(±0.2805)	0.02	0.67(±0.3541)	0.77	12.90	49.48
80–90	1435	0.12(±0.1904)	0.01	0.66(±0.3557)	0.71	13.24	48.71
90–100	1210	0.13(±0.3237)	0.03	0.66(±0.3540)	0.75	14.63	47.02
100–120	1877	0.12(±0.0886)	0.05	0.64(±0.3450)	0.87	12.79	46.24
120–140	1304	0.11(±0.2016)	0.02	0.63(±0.3620)	0.64	10.89	44.79
140–160	943	0.10(±0.3039)	0.03	0.61(±0.3765)	0.51	9.44	42.31
160–180	711	0.09(±0.1623)	0.03	0.61(±0.3199)	0.91	8.16	42.62
180–200	495	0.09(±0.2053)	0.04	0.61(±0.3137)	0.89	9.29	41.82
200–220	407	0.07(±0.0955)	0.02	0.59(±0.3481)	0.63	6.63	41.52
220–250	442	0.08(±0.2299)	0.02	0.58(±0.3524)	0.57	5.66	36.20
250–275	247	0.07(±0.1214)	0.01	0.56(±0.3811)	0.51	6.07	34.01
275–300	156	0.09(±0.1787)	0.02	0.56(±0.3436)	0.54	10.26	38.46
300–350	253	0.08(±0.0390)	0.01	0.61(±0.3789)	0.85	7.11	41.90
350–400	142	0.09(±0.2742)	0.01	0.61(±0.3685)	0.74	9.86	40.85
400–500	125	0.06(±0.0598)	0.01	0.55(±0.3823)	0.59	4.80	29.60
500–600	61	0.07(±0.2866)	0.01	0.53(±0.3759)	0.70	6.56	39.34
600–700	24	0.12(±0.2527)	0.03	0.61(±0.3608)	0.61	8.33	41.67
700–800	14	0.08(±0.1004)	0.04	0.45(±0.3467)	0.46	7.14	21.43
800–900	13	0.12(±0.2569)	0.01	0.66(±0.3898)	0.88	15.38	53.85
900–1000	10	0.07(±0.0836)	0.02	0.74(±0.2540)	0.81	0.00	50.00
>1000	21	0.06(±0.1368)	0.01	0.55(±0.3799)	0.52	9.52	38.09

Table 2
Linkage disequilibrium across chromosomes for the Guzerá obtained using the SVS7 package.

Chr	Pairs of markers (N)	r^2 mean	r^2 median	|D'| mean	|D'| median	% r^2>0.3	% |D'|>0.8
1	2118	0.15(±0.27)	0.03	0.71(±0.36)	0.90	16.24	56.14
2	1789	0.17 (±0.23)	0.07	0.70(±0.33)	0.75	14.83	46.08
3	1574	0.15(±0.21)	0.03	0.71(±0.36)	0.48	12.18	41.60
4	1539	0.14(±0.20)	0.02	0.68(±0.32)	0.86	11.43	38.24
5	1190	0.15(±0.20)	0.04	0.70(±0.35)	0.58	8.69	31.26
6	1750	0.18(±0.30)	0.05	0.71(±0.32)	0.71	16.24	45.80
7	1488	0.18(±0.27)	0.05	0.71(±0.35)	0.99	13.64	39.19
8	1507	0.16(±0.26)	0.07	0.70(±0.32)	0.66	14.02	38.53
9	1322	0.15(±0.33)	0.04	0.69(±0.36)	0.99	10.86	33.10
10	1391	0.16(±0.26)	0.04	0.71(±0.28)	0.80	11.38	37.44
11	1404	0.15(±0.15)	0.01	0.69(±0.27)	0.55	10.43	34.23
12	982	0.15(±0.15)	0.05	0.70(±0.30)	0.92	7.46	25.26
13	1040	0.14(±0.34)	0.03	0.69(±0.37)	0.99	7.65	25.64
14	1202	0.17(±0.24)	0.07	0.72(±0.27)	0.77	10.58	31.96
15	1113	0.13(±0.26)	0.03	0.67(±0.38)	0.72	7.32	25.68
16	1145	0.15(±0.16)	0.03	0.68(±0.29)	0.87	8.88	28.47
17	987	0.14(±0.22)	0.03	0.69(±0.32)	0.83	7.41	24.03
18	868	0.15(±0.28)	0.05	0.70(±0.35)	0.59	7.37	21.58
19	820	0.13(±0.19)	0.05	0.68(±0.28)	0.97	5.57	19.83
20	961	0.15(±0.19)	0.02	0.68(±0.37)	0.61	7.37	23.51
21	915	0.13(±0.25)	0.10	0.69(±0.32)	0.91	6.33	23.09
22	872	0.15(±0.21)	0.01	0.69(±0.37)	0.81	6.75	21.34
23	710	0.12(±0.34)	0.03	0.64(±0.40)	0.77	4.25	15.44
24	799	0.15(±0.39)	0.15	0.69(±0.37)	0.96	6.42	19.31
25	633	0.14(±0.35)	0.09	0.69(±0.34)	0.99	4.30	15.96
26	750	0.14(±0.07)	0.01	0.67(±0.40)	0.48	5.62	17.75
27	641	0.13(±0.33)	0.09	0.68(±0.33)	0.87	4.20	16.19
28	593	0.14(±0.28)	0.08	0.64(±0.37)	0.93	4.06	13.36
29	674	0.16(±0.16)	0.03	0.70(±0.33)	0.64	5.62	17.00
Total	32,777	0.14(±0.24)	0.03	0.68(±0.34)	0.85	16.61	53.43

|D'|>0.8. The overall means among the autosomes were r^2=0.14 and |D'|=0.68 (Table 2).

LD decay analysis suggested that for both r^2 and |D'| measures, in distances greater than 120 Kb, the LD reaches very low values (r^2<0.1 and |D'|<0.6) (Fig. 1). This corroborates the results obtained in the analysis of LD means over distances between markers.

The distance intervals of 800–900 Kb and 900–1000 Kb showed discrepant LD decay patterns, mainly for the |D'| measures. In these distance intervals, a low number of pairs of markers were observed (13 and 10, respectively), which could explain these discrepant patterns.

3.2. Effective population size (Ne) through the generations

Table 3 shows the Ne for each generation (in two-generation interval) for the corrected formula of the sample size and the non-corrected estimate of Ne. A clear decrease in Ne value is observed from 30 to 2 generations ago (Fig. 3a), reaching values less than 100 individuals for 2 generations ago in the non-corrected estimate. However, the most interesting result is observed when the difference between the Ne for two consecutive generations was analyzed (Fig. 3b). The highest differences between generations were observed between 26–24, 22–20, 14–12 and from 6 to 2 generations ago, indicating that, at this moment, the highest reductions in Ne for the Guzerá were observed.

3.3. F_{ROH} through the generations

Purfield et al. (2012) have shown that the Illumina Bovine SNP50 v2 BeadChip fails to detect ROH<1 Mb. In the present work, only ROH>1 Mb were evaluated. The observed mean number of ROH per individual was 389.81, the total length of the genome in ROH>1 Mb was 771.87 Mb and the median length of ROH was equal to 1980.07 Kb (Table 4). The F_{ROH} originating from ~50, ~25, ~6 and ~3 generations ago (F_{ROH}>1 Mb, F_{ROH}>2 Mb, F_{ROH}>8 Mb and F_{ROH}>16 Mb, respectively), calculated for the Guzerá, were compared to the F_{ROH} estimates obtained for the Fleckvieh, Brown Swiss, Norwegian Red and Tyrol Grey (Ferenčaković et al., 2013a). The F_{ROH}>1 Mb estimated for Guzerá was higher than the F_{ROH}>1 Mb estimated for the four other breeds (Supplementary Table 1). However, similar values were obtained for F_{ROH}>2 Mb for Guzerá and Brown Swiss (0.148 and 0.129); and, for F_{ROH}>8 Mb and F_{ROH}>16 Mb among Guzerá, Norwegian Red and Tyrol Grey (Supplementary Table 1). It is important to highlight that each population have their own characteristics and the results of these comparison must the interpreted carefully.

The analyses of F_{ROH} obtained for 30 generations until 2 generations ago indicated that there is a major contribution of the 30–20 past generations to current inbreeding (Fig. 2). Furthermore, Fig. 2 also shows that there is a decrease in F_{ROH} between the generations, until 12 generations ago. However, for the other inbreeding estimates the mean values did not change across significantly across the subsamples of each generation (Supplementary Table 2). Moreover,

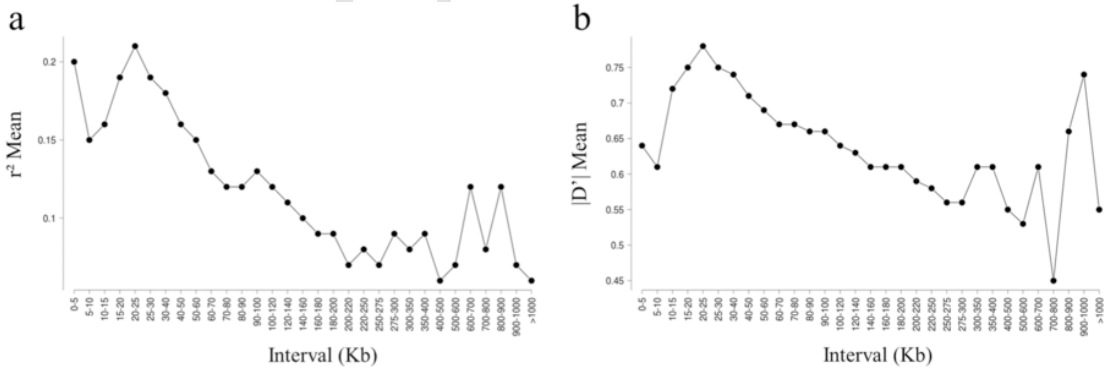

Fig. 1. LD decay in function of the distance between pairs of markers. (a) LD decay for the mean of r^2 values for each distance between pairs of markers; (b) LD decay for the mean of |D'| values for each distance between pairs of markers. The r^2 and |D'| for each pair of markers were obtained using the SVS7 package.

Table 3
Effective population size across the generations obtained for the Guzerá.

Generations	Interval (Morgans)	E (r^2)	Pairs of markers (n)	Ne_corrected[a]	Ne_non-corrected[b]
30	0.016–0.017	0.035(±0.077)	70,234	429.26	399.13
28	0.017–0.019	0.034(±0.076)	54,830	407.14	378.12
26	0.019–0.02	0.033(±0.074)	70,234	386.57	358.47
24	0.02–0.022	0.033(±0.073)	76,498	360.03	333.54
22	0.022–0.024	0.032(±0.072)	84,220	340.51	314.77
20	0.025–0.027	0.031(±0.070)	110,733	316.61	292.01
18	0.027–0.031	0.030(±0.068)	131,643	296.49	272.51
16	0.031–0.035	0.028(±0.064)	178,527	276.55	252.89
14	0.035–0.041	0.027(±0.061)	178,899	258.23	235.13
12	0.041–0.049	0.025(±0.057)	341,787	229.02	207.11
10	0.05–0.062	0.023(±0.052)	338,526	213.33	191.18
8	0.062–0.083	0.021(±0.064)	466,832	196.60	174.12
6	0.083–0.125	0.018(±0.042)	667,568	179.47	156.38
4	0.125–0.25	0.015(±0.033)	1,552,367	152.74	128.67
2	0.25–0.5	0.010(±0.020)	4,094,490	122.25	94.29

[a] Results obtained using Formula 3

[b] Results obtained using Formula 4

Table 4
Descriptive statistic of ROH (>1 Mb) for the Guzerá sample.

Parameter	ROH length (Kb)	Markers in a ROH	Number of ROH	Total length of the genome in ROH>1 Mb
Mean±Standard deviation	1980.07±2036.99	25.58±26.85	389.81±18.74	771.87±67.85
Median	1578.40	21	391	765.19
Minimum	1000.00	14	302	538.13
Maximum	85986.68	1298	432	1062.51

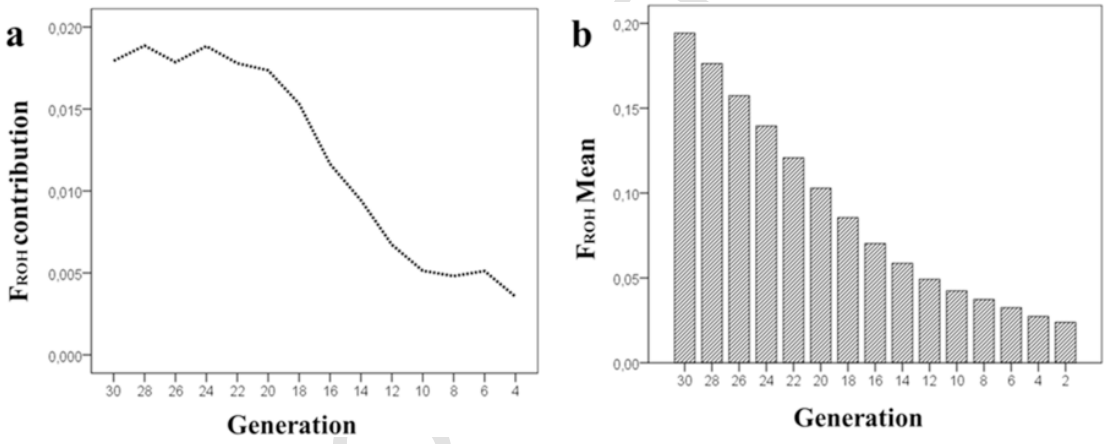

Fig. 2. F_{ROH} across the generations for the Guzerá. (a) Contribution of each generation (F_{ROH} of the previous generation – F_{ROH} of the current generation) to F_{ROH}; (b) The mean of F_{ROH} across the generations.

highly significant Pearson correlation coefficient was observed between F_{ROH} and the F_{HOM} in all generations, except between the $F_{ROH[10\ generations]}$ and $F_{HOM[10\ generations]}$ (Supplementary Table 3). The Pearson correlation for the other genomic inbreeding estimates were smaller than observed between F_{ROH} and F_{PED}. These results reflect the history of the introduction of the Guzerá in Brazil. They also corroborate the pattern of length and number of ROH observed for each individual, which reflects the high levels of inbreeding through which the Guzerá have passed during its recent history.

3.4. Population genetic structure analyses

Cross-validation error analysis suggested K=2 as the most probable value (Supplementary Fig. 2). This K-value probably reflects the selection purposes to which the breed has been subjected in Brazil. In the sample used in the present study, there are two kinds of herds: (1) dual-purpose (milk and beef) herds (Fig. 4, first population); (2) herds used for beef production (Fig. 4, second population). In the first cluster, the dual-purpose lineages are grouped. It is noteworthy that, in the herds where the dual-purpose selection was recently adopted (derivate from beef herds), there is a high proportion of the second component (red in Fig. 4).

4. Discussion

Evaluate the recent evolution of genetic diversity is a challenge, but it fundamental for the understanding of the processes that an endangered breed or species is being going through. Here we test a

strategy based on genome-wide SNP genotyping and analysis of LD decay, F_{ROH}, and evolution of effective population size through the generations, to recover recent evolution of genetic diversity The Zebu breed Guzerá was used as a model, because its history in Brazil is well known.

Until the present study, all other studies aiming to assess the genetic diversity of the Guzerá were based on pedigree or on microsatellite data. Pedigree is frequently limited to only a few generations back; microsatellites are limited due to their low coverage of the genome. The Guzerá passed through some bottlenecks during evolution in Brazil. It was introduced in Brazil at the end of the 19th century, when a small contingent of animals was imported from India. In Brazil, the Guzerá was prevalent until 1939. After this period, a substantial reduction was observed in the purebred population size to uncomfortable levels, due to the widespread use of this breed to produce crossbred animals. After1950, the population size stabilized and began to increase.

In the last decade, a Multiple Ovulation and Embryo Transfer (MOET) scheme was implemented for the Guzerá. This scheme allows the formation of large families of full- and half-sisters, thus permitting the evaluation of the productive potential of a sire by the yield of his sisters (Teodoro et al., 2003). However, despite the adoption of mating planning, the MOET nucleus demands constant evaluation of genetic diversity to avoid the use of a small number of animals selected by family index and subjected to intensive breeding processes. This way, the success of the genetic improvement can be hampered by the inbreeding, caused by the small effective number of animals used in the breeding program. The deleterious effects of endogamic depression have been detected in both taurine and Zebu cattle, affecting characteristics such as embryonic development in the first two post-fertilization weeks (Lazzari et al., 2011), daily milk yield, age at first calving and calving interval (Panetto et al., 2010). Therefore, the evolution of genetic diversity must be monitored constantly in order to ensure the efficiency of the genetic improvement programs and the conservation of the breeds under selection.

LD is an important tool in association studies as well as in studies aiming to evaluate genetic diversity. Thus, LD has been used in several studies to determine the diversity and history, signatures of selection, recombination rates, effective population size and other population events in cattle (Waples and Do, 2010; Bolormaa et al., 2013; Loh et al., 2013; Ramey et al., 2013; O'Brien et al., 2014; Schiavo et al., 2015). Moreover, for high-resolution association mapping, it is also necessary to identify haplotype-block structures and a minimal set of polymorphisms, for example, haplotype tag-SNPs. The results obtained for taurine breeds cannot be applied directly to indicine breeds because the LD pattern is different between these two breeds (Villa-Angulo et al., 2009; O'Brien et al., 2014). However, most of the markers used for genomic selection are derived from taurine sequences.

This is the first study to evaluate the LD pattern in the genome of the Guzerá. The results obtained here are very similar to those described for O'Brien et al. (2014) in taurine and indicine breeds. After quality control filtering, the average gap between markers was 76 Kb, an inconsistent pattern of LD decay was observed in the bins of markers up to 20 Kb apart. Moreover, for Nelore, Gir and Brahman, r^2 reached values below 0.1 for distances larger than 200 Kb, similar to those observed in the present study (120–140 Kb) (O'Brien et al., 2014).

Genome-wide, ROH based analyses of genetic diversity has enabled the identification of genetic relationships that would not be detected otherwise. F_{ROH} has been used in several species to assess inbreeding levels, population evolutionary dynamics and to map diseases. An interesting result emerged when F_{ROH}>1 Mb, F_{ROH}>2 Mb, F_{ROH}>8 Mb and F_{ROH}>16 Mb, obtained for the Guzerá, were compared with these same measures in other breeds.

F_{ROH}>2 Mb is very similar between Guzerá and Brown Swiss breeds (0.148 and 0.129). The Brown Swiss population used by Ferenčaković et al. (2013a) is derived from the US Brown Swiss population, which is genetically small, reflecting the importation of a small number of individuals and subsequent interbreeding. Similarly, the Brazilian Guzerá population originated from a small number of individuals imported from India at the end of the 19th century. Moreover, F_{ROH}>2 Mb reflects the inbreeding originating from approximately 25 generations; which, using a 6-year generation interval (Peixoto, MGCD personal communication), corresponds to the end of the 19th century (c. 1870). These findings reinforce that Guzerá genetic diversity still reflects the importation bottleneck. It is very important to highlight that, as informed by Ferenčaković et al. (2013b), the results obtained through the evaluation of ROH with lengths <4 Mb must be interpreted carefully. This is necessary due to the overestimation of these ROHs, caused by the small number of closely positioned markers, when Illumina Bovine SNP50 v2 BeadChip data are analyzed. The same precaution is necessary when interpreting LD estimates in distances where inconsistent LD decay patterns are observed (Fig. 1).

Similarity, in the F_{ROH}>8 (~6 generations ago) values was observed between the Guzerá and Tyrol Grey breeds. The Tyrol Grey population used in the studies by Ferenčaković et al. (2013a) has a small population size (<5000) of registered cows, which have been the target of a breeding program that involves a bull testing scheme using both artificial insemination and natural mating. Moreover, this population had high inbreeding levels (Sölkner et al., 1998). F_{ROH}>8 reflects the inbreeding originating ~6 generations ago, which corresponds to approximately 40 years ago. At the end of the 70s, the herdbooks of the Guzerá were closed in Brazil. For this reason, several animals no longer met the criteria to be recognized as belonging to the breed, thus reducing the purebred population.

F_{ROH}>16 Mb, which corresponds to the inbreeding originating ~3 generations back (~18 years), obtained for the Guzerá was very similar to that obtained for the Norwegian Red breed by Ferenčaković et al. (2013a). The Norwegian Red population used in this study is highly heterogeneous as a result of the historic admixture and, therefore, has an elevated Ne estimate (Sodeland et al., 2011). This elevated Ne was maintained through the active control of inbreeding and gene flow by importing sires from other Nordic countries. In the last decade, the Guzerá was submitted to an intensive breeding program, which aimed at reducing inbreeding levels and stimulating the increase of genetic diversity. This may explain the similarity of the F_{ROH}>16 Mb observed between the two breeds.

Ne and F_{ROH} across the generations match what is known about the genetic variability and the evolution of the Guzerá during the last century in Brazil. In the present study, it was observed that most of the inbreeding in the current Guzerá population originated from 30 to 20 generations ago. Moreover, analysis of the inbreeding contribution of each generation (F_{ROH} of the previous generation–F_{ROH} of the present generation) shown in Fig. 2a suggests that, up to 20 generations ago, the inbreeding levels did not change drastically. However, a drastic reduction in the Ne at 26 and 22 generations ago were observed (Fig. 3b). These findings reflect the formation of the Brazilian Guzerá herds from the importation of a few animals at the end of the 19th century (20 generations ago). Two other drastic reductions on

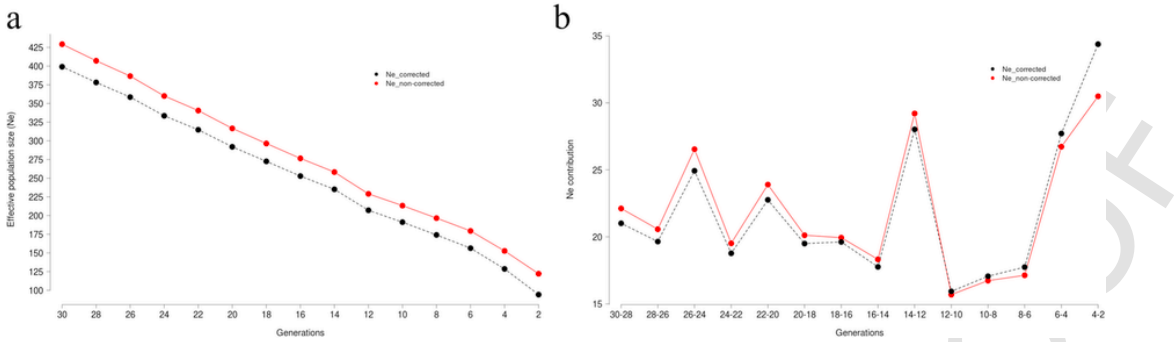

Fig. 3. Effective population size (Ne) across the generations. (a) Ne calculated for each two generation interval, from 30 to 2 generations ago. (b) Ne contribution for the observed Ne decay ($Ne_{[past\ generation]} - Ne_{[current\ generation]}$) for each generation from 30 to 2 generations ago. For both analysis, 6-years was considered as the inter-generation interval.

Ne were observed at 14 and from 6 until 2 generations ago (approximately the end of the 1930s and 1970s, respectively). These were two important moments in the history of Brazilian Guzerá herds: (1) from 1940 until 1950, the Guzerá passed through a drastic size reduction due to its intensive use in the formation of crossbreds; (2) at the end of the 1970s, the closing of the "herdbooks" for the Guzerá resulted in more stringent rules for classifying an animal as a Guzerá and a consequent reduction on the purebred herds. Peixoto et al. (2010) obtained values of inbreeding coefficients assessed by pedigree data, up to 6 generations ago, (from 0.0162 to 0.0283) very similar to the F_{ROH} results for more recent generations (from 10 to 2 generations ago) obtained in the present study. Even though the two approaches are different, high correlation was observed between the inbreeding coefficients assessed by F_{ROH} and F_{PED} (Ferenčaković et al., 2013a; Curik et al., 2014). However, in the present study, a small correlation was observed between F_{ROH} and F_{PED}, as were also observed for the other genomic inbreeding estimates, F_{HOM} and F_G (Supplementary Table 3). It is important to highlight that F_G showed a small correlation between F_{ROH} and F_{HOM}. This can be explained by the fact that the inbreeding coefficient obtained using the method described by VanRaden et al. (2008) is weighted by the allelic frequency of the markers. Therefore, individuals sharing rare alleles will have inbreeding coefficients higher than those of individuals sharing common alleles. The F_{ROH} only takes into account the sum of the ROH across the genome, without any weighting by the ROH frequency. Furthermore, this can be the reason for the low correlation observed between F_{ROH} and F_G. The same reasoning is applicable to the low correlation observed between F_{HOM} and F_G. However, it is important to highlight that the smallest correlation was observed between F_G and F_{PED}, once again reinforcing the differences between pedigree based and genomic based estimates.

Genetic structure analyses showed clear separation between the two lineages composing the sample. The dual-purpose lineages were grouped in the first cluster (Fig. 4). These lineages are early derivatives from beef lineages (second cluster). It is possible to note that, in herds where dual-purpose selection was recently adopted (last bars in first cluster), there is a high proportion of the second component (red in Fig. 4). These results suggest that the Guzerá population comprises a small number of specialized lineages (for example, beef) and, mainly, lineages which are in the process of specialization. This provides an important tool for selecting individuals to be included in breeding programs. Moreover, these results suggest that there are lineages, which can be kept as genetic diversity reservoirs, to prevent endogamic depression, particularly those lineages composed equally of beef and dual-purpose components. Furthermore, considering that,

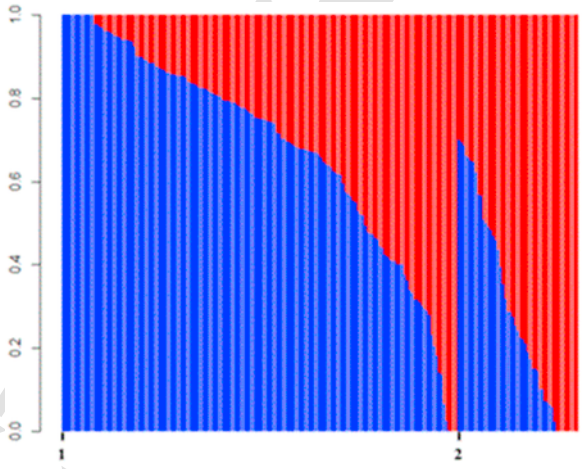

Fig. 4. Probability of each animal belonging to each subpopulation estimated with ADMIXTURE 1.23. Under the hypothesis of K=2, blue represents the probability of the animal belonging to the first population and red, to the second population. Animals were grouped according to the selection purpose: (1) dual-purpose and (2) beef production. (For interpretation of the references to color in this figure legend,the reader is referred to the web version of this article)

in the present study, only minimally related animals were used, important individuals could be identified for use in increasing the genetic diversity levels in specialized Guzerá populations.

In recent years, a growing number of studies evaluating the higher reliability and precision of genetic studies based on genomic, rather than pedigree, data have been published (Kardos et al., 2015; Gorjanc et al., 2015). However, most studies are limited to characterizing contemporary genetic diversity or estimating the evolution of genetic diversity over long periods of time. To our knowledge, this is the first paper to show that it is possible to recover recent evolutionary events through genome-wide data. Precise estimates of effective population size and current inbreeding levels are important aspects for developing strategies which aim avoid the deleterious effects of endogamic depression. The results obtained in the present study corroborate the Guzerá historical record, and reinforce the impact of past generations and recent bottlenecks on current inbreeding levels. Moreover, it was possible to observe a constant decrease in Ne across the generations, especially from 6 to 2 generations ago, which may reflect the intensive selection processes to which the breed has been subjected. Additionally, the Ne values drop below 100 animals in the most recent generations. This estimate indicates a worrisome scenario in which the breed can suffer the deleterious effect of endogamic depression. These results highlight the necessity of developing new strategies for

the genetic diversity management for the breed focusing on the influence of past generations and recent fluctuations in effective population size.

5. Conclusions

The strategy applied here, based on genome-wide genotyping followed by LD decay, Ne and F_{ROH} across the generations, enabled the study of the recent evolutionary history of the Guzerá. The results obtained in the present study demonstrated an intensive contribution of recent evolutionary events to the formation of the current inbreeding and genetic diversity levels in the Guzerá population. Although Guzerá had been used as a model, the findings reported here may be considered for any livestock specie. The intensive selection processes or recent bottlenecks can contribute considerably to the formation of current inbreeding in any population. Therefore, the application of methodologies which are useful for estimating the impact of recent evolutionary events on current genetic diversity is a crucial step in the development of management strategies. Consequently, this can avoid the effects of endogamic depression. Furthermore, the methodology and the results described here can be useful to help the selection of conservation strategies for endangered species, for which historical records are not available.

Conflict of Interest

The authors declare no conflict of interest.

Uncited references

Espigolan et al. (2013), Gasparin et al. (2007), Gonda et al. (2004), Lu et al. (2012), Marques et al. (2011) McKay et al. (2007), Weikard et al. (2012).

Acknowledgement

We thank to the farmers, who allowed the development of this project in their farms. We thank to Mr. Peter Laspina for reviewing language review and valuable comments. This study was supported by Fundação de Amparo a Pesquisa de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Marcos V. B. Silva was supported by the Embrapa – SEG 02.09.07.008.00.00 "Genomic Selection in Dairy Cattle in Brazil", CNPq PVE 407246/2013-4 "Genomic Selection in Dairy Gyr and Girolando Breeds", and FAPEMIG CVZ PPM 00395/14 "Genomic Selection in Brazilian Dairy Breeds" appropriated projects. MRSC has a fellowship from the CNPq – 307975/2010-0 and was supported by CNPq – 505338/2008-A and 481018/2008-5 projects. MGCDP, RVV, MAM have fellowships from FAPEMIG. PASF, FCS and ICR have CAPES fellowships.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.livsci.2016.10.006.

References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Bolormaa, S., Pryce, J.E., Kemper, K.E., Hayes, B.J., Zhang, Y., Tier, B., Barendse, W., Reverter, A., Goddard, M.E., 2013. Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. Genet. Sel. Evol.: GSE 45, 43.

Carneiro, P.L.S., Euclydes, R.F., Silva, Md.A., Lopes, P.S., Torres, Rd.A., Carneiro, A.P.S., Torres Filho, Rd.A., 1999. Effect of pedigree errors on the selection. Rev. Bras. Zootec. 28, 269–274.

Curik, I., Ferenčaković, M., Sölkner, J., 2014. Inbreeding and runs of homozygosity: a possible solution to an old problem. Livest. Sci. 166, 26–34.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B 1–38.

.

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet. Sel. Evol. 45, 1.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I., Sölkner, J., 2013. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J. Anim. Breed. Genet. 130, 286–293.

.

.

Gorjanc, G., Bijma, P., Hickey, J.M., 2015. Reliability of pedigree-based and genomic evaluations in selected populations. Genet. Sel. Evol. 47 (1), 1–15.

Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. J. Anim. Breed. Genet. 122 (3), 172–176.

Hill, W., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231.

Howrigan, D.P., Simonson, M.A., Keller, M.C., 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genom. 12, 1.

Kardos, M., Luikart, G., Allendorf, F.W., 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity

Kehdy, F.S., Gouveia, M.H., Machado, M., Magalhães, W.C., Horimoto, A.R., Horta, B.L., Moreira, R.G., Leal, T.P., Scliar, M.O., Soares-Souza, G.B., 2015. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. Proc. Natl. Acad. Sci. 112, 8696–8701.

Keller, M.C., Visscher, P.M., Goddard, M.E., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189, 237–249.

Lazzari, G., Colleoni, S., Duchi, R., Galli, A., Houghton, F.D., Galli, C., 2011. Embryonic genotype and inbreeding affect preimplantation development in cattle. Reproduction 141, 625–632.

Leutenegger, A.-L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., Thompson, E.A., 2003. Estimation of the inbreeding coefficient through use of genomic data. Am. J. Hum. Genet. 73, 516–523.

Lewontin, R., 1964. The interaction of selection and linkage 1 general considerations; heterotic models. Genetics 49, 49–67.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., Berger, B., 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193, 1233–1254.

.

.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS One 4, e5350.

.

O'Brien, A.M.P., Mészáros, G., Utsunomiya, Y.T., Sonstegard, T.S., Garcia, J.F., Van Tassell, C.P., Carvalheiro, R., da Silva, M.V., 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.

Panetto, J., Gutiérrez, J., Ferraz, J., Cunha, D., Golden, B., 2010. Assessment of inbreeding depression in a Guzerat dairy herd: effects of individual increase in inbreeding coefficients on production and reproduction. J. Dairy Sci. 93, 4902–4912.

Peixoto, M., Poggian, C., Verneque, R., Egito, A., Carvalho, M., Penna, V., Bergmann, J., Viccini, L., Machado, M., 2010. Genetic basis and inbreeding in the Brazilian Guzerat (Bos indicus) subpopulation selected for milk production. Livest. Sci. 131, 168–174.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., 2007. PLINK: a tool set for

whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.
Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70.
R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (URL) ⟨https://www.R-project.org/⟩.
Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D., Taylor, J.F., 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genom. 14, 1.
Ron, M., Blanc, Y., Band, M., Ezra, E., Weller, J., 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci. 79, 676–681.
Schiavo, G., Galimberti, G., Calò, D., Samorè, A., Bertolini, F., Russo, V., Gallo, M., Buttazzoni, L., Fontanesi, L., 2015. Twenty years of artificial directional selection have shaped the genome of the Italian Large White pig breed. Anim. Genet.
Sodeland, M., Kent, M., Hayes, B.J., Grove, H., Lien, S., 2011. Recent and historical recombination in the admixed Norwegian Red cattle breed. BMC Genom. 12, 33.
Sölkner, J., Filipcic, L., Hampshire, N., 1998. Genetic variability of populations and similarity of subpopulations in Austrian cattle breeds determined by analysis of pedigrees. Anim. Sci. 67, 249–256.
Sved, J., 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2, 125–141.
Teodoro, R.L., da Silva Verneque, R., Martinez, M.L., da Silva, M.V.G.B., Penna, V.M., Peixoto, M.G.C.D., 2003. Programa nacional de melhoramento do guzerá para leite: resultados do teste de progênie, do arquivo zootécnico nacional e do núcleo moet. Embrapa Gado de. Leite
VanRaden, P., 2008. Efficient methods to compute genomic predictions. Journal of dairy science 91, 4414-4423.Villa-Angulo, R., Matukumalli, L.K., Gill, C.A., Choi, J., Van Tassell, C.P., Grefenstette, J.J., 2009. High-resolution haplotype block structure in the cattle genome. BMC Genet. 10, 1.
Wang, J., 2005. Estimation of effective population sizes from data on genetic markers. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 360, 1395–1409.
Waples, R.S., Do, C., 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolut. Appl. 3, 244–262.
.
Weir, B., Hill, W.G., 1980. Effect of mating structure on variation in linkage disequilibrium. Genetics 95, 477–488.
Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56, 330–338.