Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle

Neves, H.H.R.[1]; Desidério, J.A.[2]; Pimentel, E.C.G.[3]; Scalez, D.C.B.[1@] and S.A. Queiroz[1]

[1]Departamento de Zootecnia. Universidade Estadual Paulista. Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal. SP. Brazil.
[2]Departamento de Biologia Aplicada à Agropecuária. Universidade Estadual Paulista. Faculdade de Ciências Agrárias e Veterinárias. Jaboticabal. SP. Brazil.
[3]Bayerische Landesanstalt für Landwirtschalft. Institut für Tierzucht. Poing-Grub. Bayern. Germany.

SUMMARY

Genotypes of 25 artificial insemination sires were used to study the extent of linkage disequilibrium (LD) and the correspondence between pedigree and SNP-based estimators of inbreeding in Brazilian Gyr dairy cattle. Overall, 24,020 SNPs had minor allele frequencies (MAF) greater than 5% and were used to calculate two measures of LD (r^2 and D') for all pairs of markers in each autosome. LD was also used to estimate the effective population size (Ne) at different prior generations. Individual inbreeding coefficients (F) were estimated using either pedigree information (Fped, pedigree traced back up to 9 generations) or marker information. Marker-based estimates of F were derived based on the excess homozygosity (Fhet) in SNP markers and the estimated proportion of the genome located in runs of homozygosity (Froh). The mean LD between adjacent markers averaged across all autosomes was approximately 0.20 and 0.75, measured using r^2 and D'. Useful LD was identified between markers separated by up to 100 kb when screening this sample of Gyr dairy cattle. The effective population size showed a consistent trend of decay along time, falling below 56 in the last three generations. Weaker correspondence between individual inbreeding estimates based on runs of homozygosity and pedigree was verified in the present study (estimated correlations between Fped and Froh varied from 0.32 to 0.42). It appears to be feasible to apply genomic selection to Gyr cattle in Brazil, but further studies on the extent of linkage disequilibrium using a larger sample of this population are needed.

Estudo preliminar sobre a extensão do desequilíbrio de ligação e estimativas de autozigose em bovinos leiteiros da raça Gir

RESUMO

Genótipos de 25 touros usados em inseminação artificial foram utilizados para estudar a extensão do desequilíbrio de ligação (LD) e a correspondência entre estimadores de endogamia baseados na informação de pedigree e de SNP da raça Gir no Brasil. No total, 24.020 SNPs tiveram frequências do alelo menor (MAF) maiores que 5% e foram usados para calcular duas medidas de LD (r^2 and D') para todos os pares de marcadores em cada autossomo. LD também foi utilizado para estimar o tamanho efetivo populacional (Ne) da população ancestral em diferentes gerações passadas. Coeficientes de endogamia individuais (F) foram estimados usando informação genealógica (Fped, pedigree de 9 gerações) ou informação de marcadores. As estimativas de F baseadas em marcadores foram obtidas com base no excesso de homozigose (Fhet) em marcadores SNP e a proporção estimada do genoma localizada em *trilhas de homozigose* (*runs of homozigosity*, Froh). O LD médio entre marcadores adjacentes de todos os autossomos foi de aproximadamente 0,20 e 0,75, estimado de acordo com as estatísticas r^2 e D'. Nesta amostra analisada da raça Gir, verificou-se a existência de LD potencialmente útil para seleção genômica (r^2>0,30) no caso de pares de marcadores separados por até 100kb. O tamanho efetivo populacional apresentou uma tendência consistente de queda ao longo do tempo, chegando abaixo de 56 nas últimas três gerações. Uma fraca correspondência entre as estimativas de endogamia individuais baseadas nas trilhas de homozigose e pedigree foi verificada neste estudo (correlações estimadas entre Fped e Froh variaram de 0,32 a 0,42). Os presentes resultados sugerem a possibilidade de aplicação da seleção genômica na raça Gir no Brasil, mas são necessários mais estudos sobre a extensão do desequilíbrio de ligação usando uma maior amostra desta população.

INTRODUCTION

The Gyr dairy cattle breed has been widely used for dairying in tropical regions of Brazil, primarily by crossbreeding with breeds specialized for milk production (predominantly Holstein). These *Bos taurus x B. indicus* cows have shown to excel at profitability in the low to medium-input systems that are predomi-

nant in this country (Madalena *et al.*, 1990; Guimarães *et al.*, 2006), which can be attributed to heterotic and breed complementarity gains in traits such as milk yield, reproductive efficiency, productive longevity and survivability.

The recent sequencing of genomes has led to the discovery of many bi-allelic markers known as single nucleotide polymorphisms (SNPs), which are the most abundant form of sequence variation in DNA. With the availability of very high-throughput genotyping technologies, SNPs have become the genetic markers of choice for high-resolution studies and genome-wide association studies (de Koning *et al.*, 2007). According to Hayes *et al.* (2009), the discovery of thousands of SNPs in the bovine genome was accompanied by a dramatic reduction in the cost of genotyping, which is decisive for the cost-effectiveness of this technology.

Foreseeing this scenario, Meuwissen *et al.* (2001) demonstrated that it would be possible to make very accurate selection decisions if the breeding values were predicted by using only the information from genome-wide dense markers, a method known as genomic selection (GS). According to Jannink *et al.* (2010), genomic selection has shifted the paradigm that marker-assisted selection would be generally ineffective for complex traits and is thus revolutionizing animal and plant breeding.

In dairy cattle breeding, genomic selection is expected to noticeably increase the rates of genetic gain, mostly due to the possibility of reducing the generation interval while maintaining a high accuracy of selection. Hayes *et al.* (2009) argued that this strategy should double the rate of genetic gain in the dairy industry and that Gyr dairy production could benefit from this technology. In addition, Schaeffer (2006) calculated that breeding companies could save up to 92 % of their costs by considering the hypothesis that GS could allow the elimination of traditional progeny testing, especially in breeds that are less numerous than Holstein.

Conversely, the effectiveness of GS depends on the strength of linkage disequilibrium (LD) and how it declines with distance between markers and QTLs in a population (Zhao *et al.*, 2007; Sargolzaei *et al.*, 2008). Markers must be in enough LD with a QTL to predict their effects across the population and across generations, in such a way that the extent of the within-population LD determines the marker density required for association studies and the subsequent implementation of GS.

The amount of LD is equally important as a source of information about historical events of recombination, allowing inferences of genetic diversity, geographic subdivision and genomic regions that have undergone selection (McKay *et al.*, 2007; Slatkin, 2008).

According to Sargolzaei *et al.* (2008), the measures that are most commonly employed to quantify the LD are the multiallelic D' (Lewontin, 1964), r^2 (Hill and Robertson, 1968) and standardized $\chi 2$ (Yamazaki, 1977) (in the case of bi-allelic markers, the last two are equivalent). Though D' has been commonly used in studies of LD in cattle, this measure was found to be biased

in the cases of small sample sizes and markers with low allelic frequencies (Du *et al.*, 2007; Sargolzaei *et al.*, 2008). In addition, simulations performed by Zhao *et al.* (2007) revealed that D' overestimated the amount of LD for bi-allelic markers and that r^2 would be more suitable to estimate usable LD in this case.

An important concern of animal breeding programs is the maintenance of genetic diversity, as future genetic gains are dependent on the existence of enough genetic variability to allow breeding programs to cope with changes in breeding goals, market preferences and environmental conditions (Melka and Schenkel, 2010). One of the key factors associated with the within-breed loss of genetic diversity is inbreeding, and inbred animals tend to have poorer performance in terms of reproduction, survivability and disease resistance, a phenomenon known as *inbreeding depression* (Keller *et al.*, 2011). Though animal breeders aim to improve the performance of the herds undergoing selection, the large influence of a few animals and/or families in such herds increases the probability of obtaining inbred animals, which reinforces the need for developing strategies to monitor and control inbreeding.

Recent studies have suggested that marker-based estimates of individual inbreeding coefficients could outperform the traditional pedigree-based estimators of inbreeding. One of the major advantages of using marker information for this task is the possibility of identifying autozygous segments that are due to common ancestors at a much greater number of generations in the past than is possible to identify by tracing pedigree records (McQuillan *et al.*, 2008).

At the moment, there is little knowledge of the extent of LD in Brazilian Gyr dairy cattle, and this information is necessary to determine the feasibility of genomic selection to improve this breed. Thus, this study was carried out to provide a preliminary assessment of the extent of linkage disequilibrium in Brazilian Gyr dairy cattle, using genotypes for dense SNP markers of progeny-tested sires used in artificial insemination in Brazil, and to investigate the correspondence between different estimators of inbreeding, using either the marker information or pedigree records available for this study.

MATERIAL AND METHODS

The data consisted of genotypes from 25 progeny tested used for AI. These males are representative sires of the Brazilian Gyr dairy cattle breed whose semen is currently marketed in Brazil. Genotypes were obtained using Illumina's BovineSNP50 beadchip (Illumina, San Diego, CA, USA), which includes 54,001 SNP markers. Of the total SNPs genotyped, 24020 markers had minor allele frequencies (MAF) greater than 5 % and were considered in subsequent computations of linkage disequilibrium measures.

To impute missing genotypes, SNPs were ordered by chromosome position and then submitted to fastPHASE (Scheet and Stephens, 2006), chromosome by chromosome. Based on the phased genotypes, two measures of linkage disequilibrium (r^2 and D') were

calculated for all possible pairs of markers in each chromosome, using the GOLD software (Abecasis and Cookson, 2000).

For a brief description of these measures, we denote $p(A_1)$ and $p(A_2)$ as the frequencies of alleles A_1 and A_2 at a given locus A, respectively. In the same manner, $p(B_1)$ and $p(B_2)$ are the allelic frequencies at a given locus B, and thus the frequencies of the possible haplotypes are represented by $p(A_1B_1)$, $p(A_2B_1)$, $p(A_1B_2)$ and $p(A_2B_2)$.

Following this notation, a quantity D was defined as

$$D= p(A_1B_1)* p(A_2B_2) - p(A_1B_2)* p(A_2B_1)$$

Hence, D′ was calculated as

$$D'=|D|/D_{max'}$$

where D_{max} was calculated as

$$D_{max}= min [p(A_1)*p(B_1), p(A_2)*p(B_2)], \text{ if } D<0, \text{ and}$$

$$D_{max}= min [p(A_1)*p(B_2), p(A_2)*p(B_1),] \text{ otherwise.}$$

The measure r² was calculated as

$$r^2= (D)^2/[p(A_1) p(B_1) p(A_2) p(B_2)].$$

Within and inter-chromosomal heterogeneity in LD were investigated by fitting a general linear model to analyze r² data for all syntenic pairs of markers. The following model was fitted using the SAS GLM procedure (SAS Institute Inc 2002):

$$r_{ij}^2= \mu + C_i + \beta_1 (ld_j) + \beta_2 (ld_j)^2 + e_{ij},$$

where r_{ij}^2 is the measure of r² for the jth pair of markers in the ith chromosome, μ is the overall mean for r² among syntenic pairs of markers, C_i is the effect of ith chromosome, β_1 and β_2 are the linear and quadratic coefficients of the regression of r² on the log transformed physical distance (ld_j) and e_{ij} is a residual effect. Because of the non-linear relationship between LD and the physical distance, the log transformed physical distance was considered to improve the fit of the model.

The effective population size (Ne) in the Gyr dairy cattle breed was also estimated at different generations by assuming that LD in any range c (in Morgans) is expected to reflect effective population size at approximately (1/2c) generations ago (Hayes *et al.*, 2003). Hence, at different genetic distances, effective population size was calculated as

$$Ne= (1/r^{2*})-1)/(4c),$$

where *c* is the mean recombination distance and r^{2*} is r² averaged across all SNP pairs in a given *c* range. It was assumed that 1 centimorgan (cM) equals 1 Mb. In this way, the average r² was calculated for all SNP pairs in each of 24 distance classes (ranging from 0.1 to 34 cM), to estimate Ne from 2 to 500 generations back.

The individual inbreeding coefficients were obtained using either pedigree or marker information. The pedigree-based coefficients of inbreeding (Fped) were derived using path coefficient methodology (Wright, 1923). Two types of marker-based individual inbree-

ding coefficients were derived following Keller *et al.* (2011):

1.- Fhet: estimate of genomic autozygosis based on the difference between the observed and expected numbers of homozygous genotypes and,

2.- Froh: estimated proportion of the genome located in regions known as runs of homozygosity (ROH).

When using SNP markers, ROH are defined as continuous segments of homozygous markers, which are highly likely to be autozygous (Ferencakovic *et al.*, 2012), i.e., the presence of a ROH in a given animal is very likely to have occurred because both parents inherited the same haplotype from a common ancestor. In this way, the size of a ROH segment is related to the number of generations (g) until the common ancestor, following an exponential distribution with mean equal to 1/2 g Morgans (Keller *et al.*, 2011). Because recombination events break long chromosome segments over time, long ROH segments are expected to be autozygous segments originating from recent common ancestors, whereas shorter ROH segments are expected to be due to more remote ancestors, though they can also include some non-IBD segments (Ferencakovic *et al.*, 2012).

In the present study, seven thresholds for the minimum length to define a ROH were adopted: 2 Mb, 4 Mb, 6 Mb, 8 Mb, 10 Mb, 12 Mb and 15 Mb, resulting in seven different estimators of inbreeding (Froh2, Froh4, Froh6, Froh8, Froh10, Froh12 and Froh15, respectively). Such estimators are expected to track autozygous segments due to common ancestors at different generations, from more remote (Froh2) to more recent (Froh15).

Both Fhet and Froh were calculated using the PLINK software (Purcell *et al.*, 2007). The correspondence between Fped and marker-based estimators of individual inbreeding coefficients was evaluated using the Pearson's correlation coefficient.

RESULTS

Descriptive statistics regarding the number of SNPs, inter-marker distances and linkage disequilibrium measures are presented in **table I**. Because the markers in the panel are approximately evenly distributed, the autosomes differed in the number of markers that they contained, such that BTA25 had the smallest number (394 markers) and BTA1 had the largest number (1483 markers). The SNP loci density varied among chromosomes, ranging from 10.67 SNPs/Mb (BTA6) to 5.81 SNPs/Mb (BTA13). To give an idea of the dimensionality issue regarding pairwise combinations of syntenic markers, at the current density, LD measures were calculated for 10.299.034 pairs of markers.

Even after filtering the SNP data, a considerable proportion of SNPs had an MAF below 20 % **(figure 1)**. The overall mean of r² for all syntenic pairs was 0.027. In contrast, the extent of LD between adjacent markers was considerable. The overall estimate of LD among adjacent pairs of markers was approximately 0.198 and

Table I. Summary of the analyzed SNP markers for each autosome (BTA) in a sample of Brazilian Gyr dairy cattle (Resumo dos marcadores SNP analisados para cada autossomo (BTA) em uma amostra da raça Gir).

BTA	Length (Mb)	SNP (n)	Longest gap (Mb)[1]	Mean distance ±SD[1]	Mean r² ±SD[1]	Mean D'±SD[1]
1	161.02	1483	1.2044	0.109±0.117	0.193±0.263	0.755±0.330
2	140.55	1192	1.7384	0.118±0.140	0.213±0.279	0.771±0.323
3	127.91	1139	0.9635	0.112±0.127	0.196±0.259	0.755±0.331
4	124.04	1110	1.0998	0.112±0.122	0.197±0.260	0.764±0.327
5	125.80	898	1.5914	0.140±0.177	0.215±0.281	0.750±0.334
6	122.47	1307	0.8262	0.094±0.097	0.222±0.275	0.761±0.323
7	112.06	1006	2.8671	0.112±0.146	0.210±0.281	0.749±0.331
8	116.91	1078	1.5130	0.109±0.129	0.207±0.275	0.764±0.321
9	107.62	975	0.8887	0.110±0.116	0.205±0.269	0.753±0.319
10	106.20	969	2.0815	0.110±0.134	0.186±0.251	0.749±0.331
11	110.17	998	0.8853	0.110±0.121	0.191±0.256	0.746±0.329
12	85.09	688	1.9673	0.124±0.159	0.183±0.261	0.773±0.325
13	127.94	743	1.0540	0.114±0.119	0.221±0.288	0.763±0.332
14	81.32	788	0.7978	0.103±0.119	0.201±0.268	0.718±0.334
15	84.54	786	0.8492	0.108±0.111	0.197±0.256	0.743±0.324
16	77.74	760	1.3020	0.102±0.114	0.196±0.268	0.730±0.341
17	76.32	730	1.5503	0.105±0.113	0.175±0.242	0.728±0.338
18	65.96	595	1.0907	0.111±0.117	0.188±0.252	0.712±0.336
19	65.18	544	1.6434	0.120±0.155	0.190±0.250	0.773±0.319
20	75.56	728	1.5007	0.104±0.117	0.206±0.268	0.747±0.320
21	69.17	620	0.8494	0.112±0.116	0.181±0.246	0.736±0.340
22	61.83	601	1.5310	0.103±0.120	0.209±0.261	0.740±0.333
23	53.27	510	0.8726	0.105±0.113	0.177±0.248	0.714±0.346
24	64.92	567	1.0297	0.114±0.116	0.189±0.265	0.731±0.342
25	43.13	394	0.6059	0.110±0.109	0.168±0.236	0.698±0.351
26	51.67	486	1.0018	0.107±0.115	0.201±0.279	0.754±0.333
27	48.56	462	1.6778	0.105±0.128	0.167±0.230	0.714±0.325
28	46.00	445	1.0455	0.104±0.124	0.164±0.234	0.703±0.345
29	51.58	461	2.0018	0.112±0.145	0.219±0.282	0.746±0.330
Overall*	2584.55	23063	2.8671	0.110±0.127	0.198±0.264	0.747±0.330

[1]Between adjacent SNP. *Overall mean. SNP= single nucleotide polymorphism. ([1]Entre SNPs adjacentes. *Média geral. SNP= polimorfismo de nucleotídeo único).
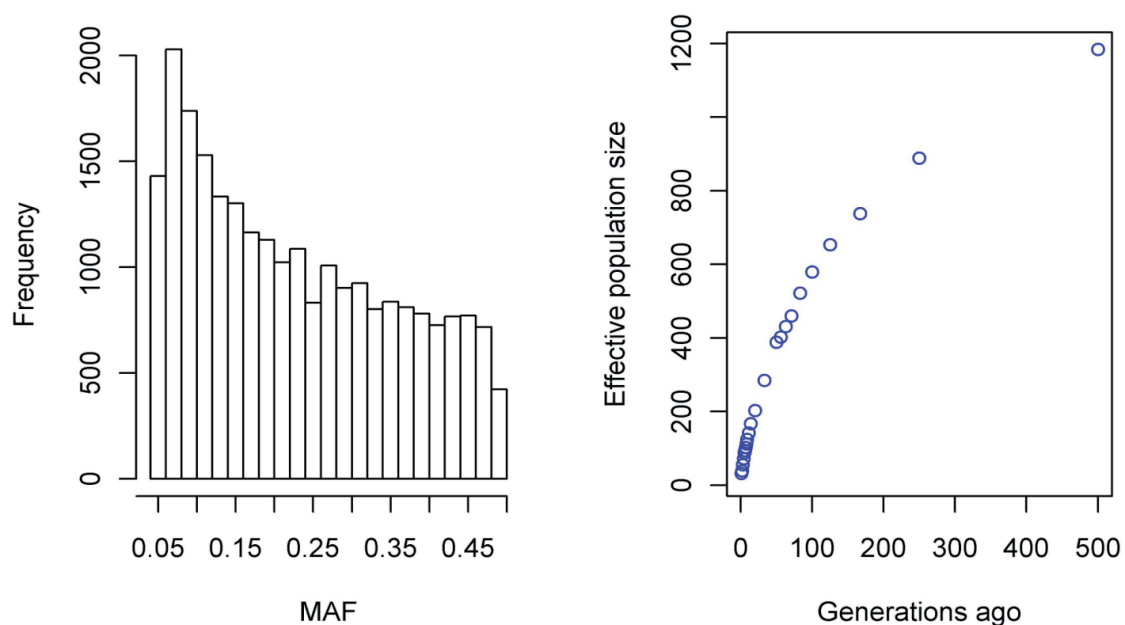

Figure 1. Left.- Histogram of minor allele frequency (MAF) of SNPs. Right.- Estimated effective population size (Ne) over time in Brazilian Gyr dairy cattle (Esquerda.- Histograma do alelo de frequência menor (MAF) de SNPs. Direita.- Tamanho efetivo populacional estimado (Ne) ao longo do tempo em bovinos leiteiros da raça Gir).

0.747, according to the statistics r^2 and D', respectively **(table I)**.

In **figure 2**, estimates of LD are plotted against physical distance between syntenic pairs. For better visualization, the values of LD were averaged in intervals of 0.5 Mb (top) and 10 kb (bottom) for both r^2 and D'. A clear exponential trend of decay with physical distance is observed for LD (Figure 2, top). In addition, by comparing the decay of LD with distance, it is apparent that, on average, the extent of LD estimated according to D' was consistently higher than the values estimated by r^2. Analyzing the averages of LD within a smaller range (**figure 2**, bottom), it can be observed that the useful values of LD (above 0.20, according r^2) did not extend to more than 100 kb. A higher dispersion of LD (especially in the case of D') is observed above 150 Mb, which could be related to the small number of pairs considered in this interval compared with the number of pairs separated by shorter distances.

Table II presents the mean LD measured according to r^2 and D' between markers that were separated by less than 1 Mb. In this range, the markers were approximately evenly spaced within each interval of 0.1 Mb **(table II)**, and higher levels of LD were observed for SNPs in close proximity. For instance, at the ranges 0 to 0.1 Mb, 0.1 to 0.2 Mb, 0.4 to 0.5 Mb and 0.9 to 1 Mb, the average r^2 was 0.215, 0.143, 0.083 and 0.063, respectively. In these same ranges, 24.43 %, 14.61 %, 5.91 % and 3.36 % of the SNP pairs exhibited an r^2 larger than 0.3 **(table II)**.

The current estimate of effective population size obtained using information about the LD falls below 56 for the last three generations. A pattern of reduction in Ne along time can be observed, and the estimates of Ne in ancient generations are expected to be above 500 individuals (**figure 1**, right).

The results regarding intra and inter-chromosomal heterogeneity showed that significant effects of chromosome (p<0.0001) and log transformed physical distance influenced r^2 values (p<0.0001 for both linear and quadratic coefficients) (R^2= 8.996 %). In contrast, most of the observed variation in LD (above 90 % of the variance explained by the model) was due to the effect of log distance.

The level of inbreeding, according to marker and pedigree-based inbreeding coefficients, ranged from 0.008 (Fped) to 0.031 (Froh2). The observed levels of Froh2 and Froh4 (0.030) were much higher when compared with that of Fped.

Pearson's correlations between Fped and marker-based estimators of inbreeding varied from 0.32 to 0.42 **(figure 3)**. Inbreeding coefficients estimated by Froh were highly correlated, regardless of the minimum threshold considered to define a ROH, whereas smaller correlations were verified between Froh and Fhet, suggesting larger differences between both types of estimators.

The threshold applied to define a ROH had small influence on the correlation between Fped and Froh. However, when smaller thresholds were adopted (≤6 Mb), the averages of Froh were between 12 % and 36 %

Table II. Mean (SD) of linkage disequilibrium for closely located syntenic pairs of markers, measured according to the statistics r^2 and D' and frequency (%) of pairs with r^2>0.30 (Média (DP) de desequilíbrio de ligação para pares de marcadores sintênicos separados por até 1Mb, de acordo com as estatísticas r^2 e D' e frequência (%) de pares com r^2>0,30).

Distance (Mb)	N	r^2		D'		r^2>0.30 (%)
0.0-0.1	21975	0.215	(0.275)	0.759	(0.325)	24.41 %
0.1-0.2	23877	0.143	(0.207)	0.680	(0.348)	14.61 %
0.2-0.3	23298	0.111	(0.167)	0.647	(0.355)	10.13 %
0.3-0.4	22892	0.094	(0.145)	0.626	(0.359)	7.57 %
0.4-0.5	22571	0.083	(0.128)	0.615	(0.360)	5.91 %
0.5-0.6	22256	0.076	(0.117)	0.601	(0.362)	4.81 %
0.6-0.7	22241	0.073	(0.113)	0.597	(0.364)	4.50 %
0.7-0.8	22094	0.070	(0.110)	0.585	(0.363)	4.01 %
0.8-0.9	21927	0.065	(0.101)	0.578	(0.363)	3.47 %
0.9-1.0	21812	0.063	(0.098)	0.572	(0.365)	3.36 %

greater than the average of Fped, possibly due to the consideration of more remote ancestors in the estimate of inbreeding according to Froh. However, ROH thresholds greater than 8 Mb resulted in Froh averages smaller than those obtained using Fped.

DISCUSSION

The Gyr cattle in Brazil are a dual purpose breed used for dairy and meat. Over the past few decades, the number of farmers who raise Gyr cattle for meat has been decreasing sharply, and the breed has been selected for dairy, with the sires tested by their progeny. Furthermore, the cows have been crossbred, primarily to Holstein bulls, to produce the F1 dairy heifers that are extensively used in dairy farms (Guimarães *et al.*, 2006). These situations have contributed to a decrease in the population of the animals, resulting in a small number of representative commercial bulls, which is reflected in the sample size used in our research.

Because we employed a relatively stringent criterion regarding the MAF threshold and the sample size, only approximately 44 % of the available marker data were considered in the present study. In such circumstances, only markers that were separated by up to 100 kb exhibited an average r^2 higher than 0.20-0.30, the range usually employed in previous studies to define the levels of LD from which genomic selection would work (Meuwissen *et al.*, 2001; Sargolzaei *et al.*, 2008; Hayes *et al.*, 2009).

The results for the decay of LD with physical distance were in reasonable agreement with the findings of previous studies (McKay *et al.*, 2007; Khatkar *et al.*, 2008). The study of McKay *et al.* (2007) reported the pattern of LD in eight breeds of cattle, including two *Bos indicus* meat-type breeds (Nellore and Brahman), and for all of them, the useful LD for association studies was not extended by more than 0.5 Mb. These authors suggested that *Bos indicus* breeds have substantially lower levels of LD at shorter inter-marker ranges than *Bos taurus* breeds. The figures found for Brazilian Gyr dairy cattle seems to confirm this trend,
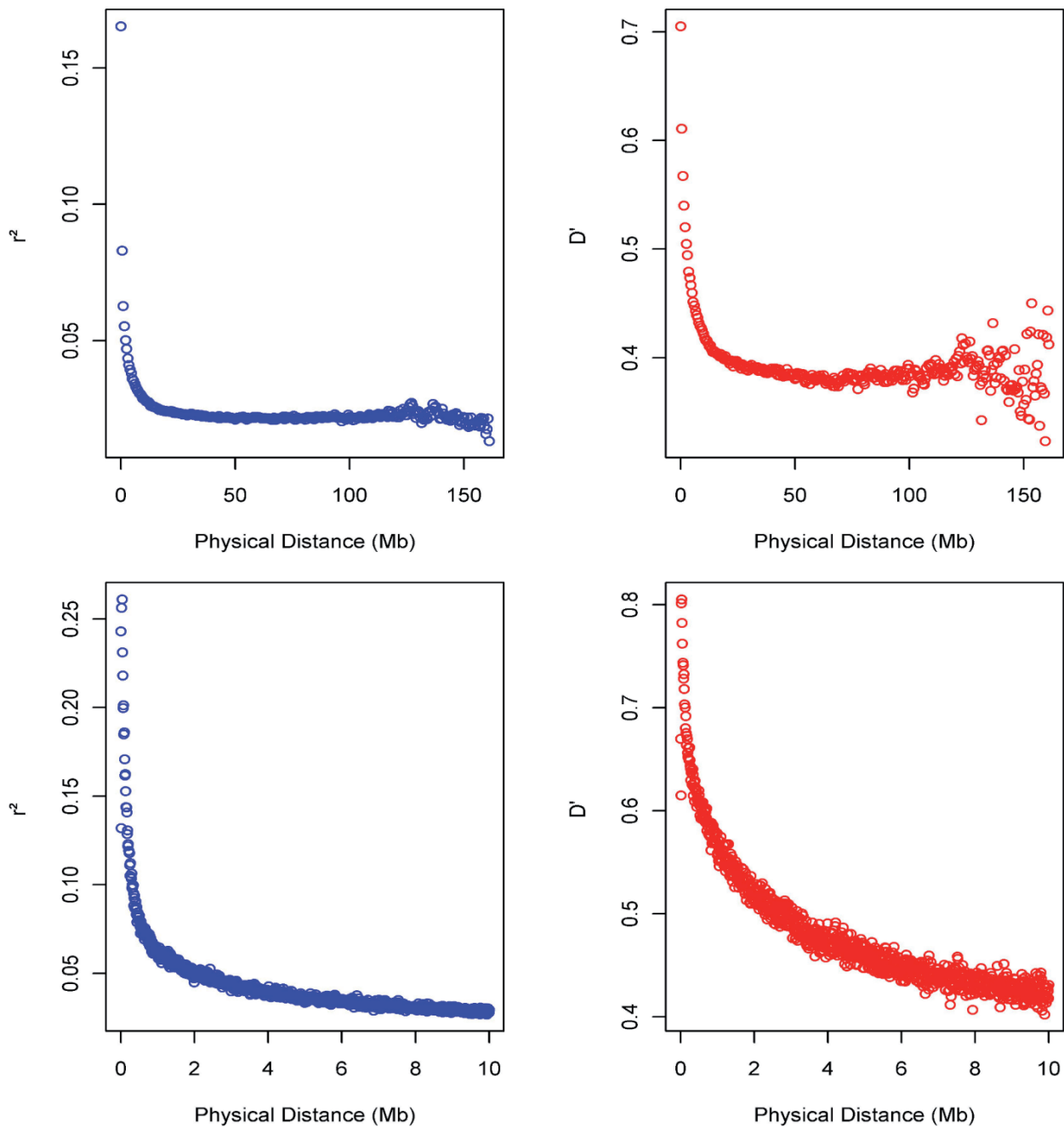
Figure 2. Relationship between linkage disequilibrium (LD) among SNP pairs and physical inter-marker distance (in Mb), pooled over all autosomes, in Brazilian Gyr dairy cattle. LD is plotted against the distance between syntenic pairs, averaged in intervals of 0.5 Mb (top) and 10 kb (bottom), based on the statistics r^2 (blue) and D' (red) (Relação entre o desequilíbrio de ligação (LD) entre pares de SNPs e distância física intermarcador (em Mb), agrupados sobre todos os autossomos, em bovinos leiteiros da raça Gir).

as the values found for markers separated by up to 20 kb were notably lower than the comparable estimates found in the literature for Holstein populations in North America (Sargolzaei *et al.*, 2008) and slightly lower than those reported for Australian Holsteins (Khatkar *et al.*, 2008). According to McKay *et al.* (2007), this result could be related to ascertainment bias (SNPs were detected because they were common among *Bos taurus* breeds) or could reflect the historically larger effective population sizes in *Bos indicus* breeds.

The trend of reduction in the effective population size in recent generations observed in this study was also observed in populations of Holstein (Sargolzaei *et al.*, 2008) and West African cattle (Thévenón *et al.*, 2007). This decay could be attributable to intense selection, which explains the lower effective population size in Gyr and Holstein populations compared with West African populations (raised in extensive pastoral systems).

The estimates of Ne in the present study must be considered a rough approximation. Nevertheless, the values calculated for Ne in recent generations (approximately 55 and 39, three and two generations ago) were in reasonable agreement with the estimates of below
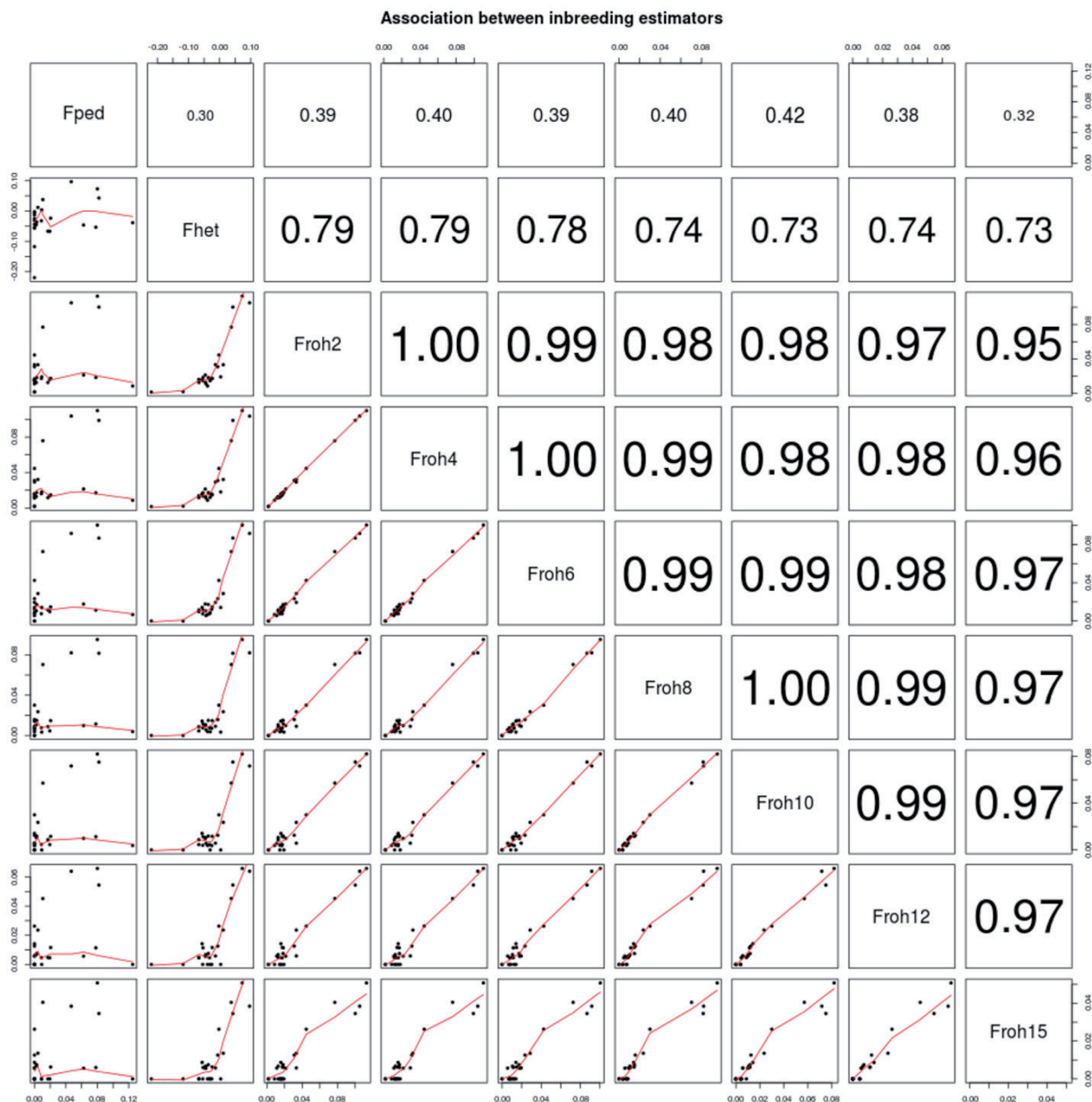
Figure 3. Scatter plots comparing different estimators of inbreeding (below diagonal) and the corresponding Pearson's correlation coefficients (above diagonal). *Fped= pedigree-based estimate of individual inbreeding (F); Fhet= F based on the difference between observed and expected homozygosity in SNP markers; Froh= estimated proportion of the genome located in runs of homozygosity (ROH), considering different minimum length thresholds to define a ROH (2 Mb, 4 Mb, 6 Mb, 8 Mb, 10 Mb, 12 Mb and 15 Mb: Froh2 to Froh15, respectively) (Gráficos de dispersão comparando diferentes estimadores de endogamia (abaixo da diagonal) e os correspondentes coeficientes de correlação de Pearson (acima da diagonal)).

50 individuals that were obtained using the average inbreeding coefficients reported previously for this breed (Queiroz *et al.*, 2000; Schenkel *et al.*, 2002; Faria *et al.*, 2009).

While the estimates of inbreeding in the present study can be considered small, these figures should be interpreted with caution. For instance, the averages of inbreeding using either marker or pedigree information were even smaller than the corresponding figures reported for the Fleckvieh breed, a breed with a larger effective population size that was studied by Ferencakovic *et al.* (2012). Conversely, the small effec-

tive population size in Gyr dairy cattle indicates that inbreeding must be considered in breeding and mating decisions to maintain long-term genetic diversity in this breed. The discrepancy between the estimated effective population size and estimates of inbreeding levels in the present study could be related to the small sample size and to sampling, while different criteria employed to prune marker data and to define a ROH can also influence the estimates of Froh (Ferencakovic *et al.*, 2013).

Genomic selection could also be useful to maintain genetic diversity in Gyr dairy cattle, e.g., by screening

a larger number of selection candidates than in conventional progeny tests and by capturing the Mendelian sampling term when estimating breeding values (Hayes *et al.*, 2009).

The significant effects of chromosome and log-transformed physical distance on the r^2 values found in our research were also observed in the Holstein population of North America studied by Sargolzaei *et al.* (2008). These authors suggested that inter-chromosome heterogeneity could be a result of intense selection, which also could explain the heterogeneity observed for Brazilian Gyr dairy cattle but at lower extent, as the generation interval for this breed is around eight years (Faria *et al.*, 2009) and is thus considerably higher than the values reported for Holstein populations (Stachowicz *et al.*, 2009).

The results regarding the comparison between r^2 and D′ at different distances confirm the data from previous studies with respect to the overestimation of LD by D′, especially in cases of low MAF and small samples sizes (Zhao *et al.*, 2007; Sargolzaei *et al.*, 2008). In this study, higher values of D′ were estimated at larger distances compared with those reported in Khatkar *et al.* (2008), although the results regarding r^2 were considerably closer.

If the initial results presented here are confirmed, the amount of LD estimated in Brazilian Gyr dairy cattle is expected to allow estimating genomic breeding values (GEBV) with accuracies of up to 0.8 (Calus *et al.*, 2008; Hayes *et al.*, 2009), although this obviously depends on the size of the training population, the availability of phenotypic records and the heritability of the traits. In addition, because the statistical power in association studies is directly related to r^2 (Khatkar *et al.*, 2008; Sargolzaei *et al.*, 2008), using larger dense SNP maps would achieve higher power and would capture QTL in regions that remained uncovered at the current density, for which the useful LD was estimated for markers separated by up to 100 kb.

A sharp trend of decay in the effective population size in recent generations was observed for Brazilian Gyr Dairy cattle. The current estimates for Ne can be considered low, indicating that inbreeding must be a matter of concern in breeding programs for this breed.

Pearson's correlations between Fped and Froh were weaker than those previously reported by Ferencakovic *et al.* (2012) after studying four taurine dairy cattle breeds. These authors verified moderate to strong correlations between Fped and Froh and suggested that Froh would provide good estimates of individual inbreeding coefficients.

As a general rule, Fped was slightly more closely correlated with Froh than it was with Fhet. While the traditional pedigree-based inbreeding coefficients are able to estimate the levels of autozygosity relative to an arbitrarily defined founder population, which is generally limited by the depth and completeness of pedigree information, the estimators based on the concept of ROH are able to estimate inbreeding levels due to both recent and more remote ancestors (Ferencakovic *et al.*, 2012).

It must be emphasized that the correlations reported in the present study are only an approximate proxy for the association between the different estimators of inbreeding investigated, especially after considering that the standard error (SE) of the correlations reported here is high (e.g., for a sample size= 25 and true correlation coefficient= 0.50, the approximate SE would be 0.15). Additionally, for three animals, a large discrepancy between Fped and marker-based estimates of inbreeding was verified **(figure 3)**, which had a large influence on the estimated correlations and could be related to, e.g., inconsistencies in the genealogical records. When the records of such animals were disregarded, the correlations between Froh and Fped were much stronger (varied from 0.71 to 0.80, depending on the length threshold to define a ROH, data not shown). Thus, another plausible explanation for the discrepancy between Fped and marker-based estimators of inbreeding is related to the possibility of obtaining more precise estimates when there is inconsistency and/or a lack of pedigree information, which has been often employed as a justification for using genomic information to estimate inbreeding levels (Keller *et al.*, 2011).

Due to the importance of the Gyr dairy cattle population to Brazilian producers, these preliminary findings on genomics, using a small but representative sample of AI commercial sires, will help in the design of future studies, which are needed to investigate the reasons for the large discrepancy between estimates of Ne and inbreeding levels in more detail, and will provide more accurate estimates of the extent of LD and related parameters in this population.

BIBLIOGRAPHY

Abecasis, G.R. and Cookson, W.O.C. 2000. GOLD – Graphical overview of linkage disequilibrium. *Bioinformatics*, 16: 182-183.

Calus, M.P.; Meuwissen, T.H.; de Roos, A.P. and R.F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 178: 553-561.

De Koning, D.J.; Archibald, A. and Haley, C.S. 2007. Livestock genomics: bridging the gap between mice and men. *Trends Biotechnol*, 25: 483-489.

Du, F.X.; Clutter, A.C. and Lohuis, M.M. 2007. Characterizing linkage disequilibrium in pig populations. *Int J Biol Sci*, 3: 166-178.

Faria, F.J.C.; Vercesi Filho, A.E.; Madalena, F.E. and Josahkian, L.A. 2009. Pedigree analysis in the Brazilian Zebu breeds. *J Anim Breed Genet*, 126: 148-153.

Ferencakovic, M.; Hamzic, E.; Gredler, B.; Solberg, T.R.; Klemetsdal, G.; Curik, I. and Sölkner, J. 2012. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J Anim Breed Genet,* 130: 286-293.

Ferencakovic, M.; Sölkner, J. and Curik, I. 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol,* 45: 42.

Guimarães, P.H.S.; Madalena, F.E. and Cezar, I.M. 2006. Comparative economics of Holstein/Gir F1 dairy female production and conventional beef cattle suckler herds: a simulation study. *Agri Sys,* 88: 111-124.

Hayes, B.J.; Bowman, P.; Chamberlain, A. and Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci,* 92: 433-443.

Hayes, B.J.; Visscher, P.M.; McPartlan, H.C. and Goddard, M.E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res,* 13: 635-643.

Hill, W.G. and Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor Appl Genet,* 38: 226-231.

Jannink, J.L.; Lorenz, A.J. and Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics,* 9: 166-177.

Keller, M.C.; Visscher, P.M. and Goddard, M.E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics,* 89: 237-249.

Khatkar, M.S.; Nicholas, F.W.; Collins, A.R.; Zenger, K.R.; Cavanagh, J.A.L.; Barris, W.; Schnabel, R.D.; Taylor, J.F. and Raadsma, H.W. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics,* 9: 187.

Lewontin, R.C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics,* 49: 49-67.

Madalena, F.E.; Teodoro, R.L.; Lemos, A.M.; Monteiro, J.B.N. and Barbosa, R.T. 1990. Evaluation of strategies for crossbreeding of dairy cattle in Brazil. *J Dairy Sci,* 73: 1887-1901.

McKay, S.D.; Schnabel, R.D.; Murdoch, B.M.; Matukumalli, L.K.; Aerts, J.; Coppieters, W.; Crews, D.; Dias Neto, E.; Gill, C.A.; Gao, C.; Mannen, H.; Stothard, P.; Wang, Z.; Van Tassel, C.P.; Williams, J.L.; Taylor, J.F. and Moore, S.S. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC Genetics,* 8: 74.

McQuillan, R.; Leutenegger, A.L.; Abdel-Rahman, R.; Franklin, C.S.; Pericic, M.; Barac-Lauc, L.; Smolej-Narancic, N.; Janicijevic, B.; Polasek, O.; Tenesa, A.; MacLeod, A.K.; Farrington, S.M.; Rudan, P.; Hayward, C.; Vitart, V.; Rudan, I.; Wild, S.H.; Dunlop, M.G.; Wright, A.F.; Campbell, H. and Wilson, J.F. 2008. Runs of homozygosity in European populations. *Am J Hum Genet,* 83: 359-372.

Melka, M.G. and Schenkel, F. 2010. Analysis of genetic diversity in four Canadian swine breeds using pedigree data. *Can J Anim Sci,* 90: 331-340.

Meuwissen, T.H.E.; Hayes, B.J. and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157: 1819-1829.

Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J. and Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81: 559-575.

Queiroz, S.A.; Albuquerque, L.G. and Lanzoni, N.A. 2000. Efeito da endogamia sobre características de crescimento de bovinos da raça Gir no Brasil. *Rev Bras Zootecn,* 29: 1014-1019.

Sargolzaei, M.; Schenkel, F.S.; Jansen, G.B. and Schaeffer, L.R. 2008. Extent of linkage disequilibrium in holstein cattle in North America. *J Dairy Sci,* 91: 2106-2117.

SAS Institute Inc. 2002. Statistical analysis software. 9. SAS Institute Inc. Cary. NC. USA.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet,* 123: 218-223.

Scheet, P. and Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet,* 78: 629-644.

Schenkel, F.S.; La Gioia, D.R. and Riboldi, J. 2002. Níveis de endogamia e depressão endogâmica no ganho de peso de raças zebuínas no Brasil. IV Simpósio Brasileiro de Melhoramento Animal (SBMA), Campo Grande, Brazil. http://sbmaonline.org.br/anais/iv/trabalhos/pdfs/ivt06bc.pdf (04/08/2014).

Slatkin, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet,* 9: 477-485.

Stachowicz, K.; Sargolzaei, M.; Miglior, F. and Schenkel, F.S. 2009. Rates of inbreeding and genetic diversity in canadian Holstein cattle. Technical report to the dairy cattle breeding and genetics committee. http://cgil.uoguelph.ca/dcbgc/Agenda0910/InbreedingHolsteins.pdf (15/08/2014).

Thévenon, S.; Dayo, G.K.; Sylla, S.; Sidibe, I.; Berthier, D.; Legros, H.; Boichard, D.; Eggen, A. and 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.

Yamazaki, T. 1977. The effects of overdominance on linkage in a multi-locus system. *Genetics,* 86: 227-236.

Wright, S. 1923. Mendelian analysis of the pure breeds of livestock. I. The measurement of inbreeding and relationship. *J Heredity,* 14: 339-348.

Zhao, H.; Nettleton, D. and Dekkers, J.C.M. 2007. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genet Res,* 89: 1-6.