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


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 (r² and D') for all pairs of markers in each autosomal 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 r² 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.

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’ (Levontin, 1964), r² (Hill and Robertson, 1968) and standardized χ² (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 r² 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 allelic 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 (r² 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 allele frequencies at a given locus $B$, and thus the frequencies of the possible haplotypes are represented by $p(A_1B_1)$, $p(A_1B_2)$, $p(A_2B_1)$ and $p(A_2B_2)$.

Following this notation, a quantity $D$ was defined as

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

Hence, $D'$ was calculated as

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

where $D_{max}$ was calculated as

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

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

The measure $r^2$ was calculated as

$$r^2 = \frac{(D)^2}{p(A_1)p(B_1)p(A_2)p(B_2)}.$$  

Within and inter-chromosomal heterogeneity in LD were investigated by fitting a general linear model to analyze $r^2$ data for all syntenic pairs of markers. The following model was fitted using the SAS GLM procedure (SAS Institute Inc 2002):

$$r^2 = \mu + C_i + \beta_1 (l_d)^2 + \beta_2 (l_d)^{2.5} + e_{ij},$$

where $r^2$ is the measure of $r^2$ for the $i$th pair of markers in the $i$th chromosome, $\mu$ is the overall mean for $r^2$ among syntenic pairs of markers, $C_i$ is the effect of $i$th chromosome, $\beta_1$ and $\beta_2$ are the linear and quadratic coefficients of the regression of $r^2$ on the log transformed physical distance ($l_d$) 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 ($N_e$) in the Gyr dairy cattle breed was also estimated at different generations by assuming that LD in any range $c$ (in Morgans) is 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 1. 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 $r^2$ 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).

<table>
<thead>
<tr>
<th>BTA</th>
<th>Length (Mb)</th>
<th>SNP (n)</th>
<th>Longest gap (Mb)</th>
<th>Mean distance ± SD</th>
<th>Mean r² ± SD</th>
<th>Mean D’ ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>161.02</td>
<td>1483</td>
<td>1.2044</td>
<td>0.109 ± 0.117</td>
<td>0.193 ± 0.263</td>
<td>0.755 ± 0.330</td>
</tr>
<tr>
<td>2</td>
<td>140.55</td>
<td>1192</td>
<td>1.7384</td>
<td>0.112 ± 0.127</td>
<td>0.196 ± 0.259</td>
<td>0.755 ± 0.331</td>
</tr>
<tr>
<td>3</td>
<td>127.91</td>
<td>1139</td>
<td>1.0998</td>
<td>0.112 ± 0.122</td>
<td>0.197 ± 0.260</td>
<td>0.764 ± 0.327</td>
</tr>
<tr>
<td>4</td>
<td>124.04</td>
<td>1110</td>
<td>1.5914</td>
<td>0.140 ± 0.177</td>
<td>0.215 ± 0.281</td>
<td>0.750 ± 0.334</td>
</tr>
<tr>
<td>5</td>
<td>125.80</td>
<td>898</td>
<td>0.8262</td>
<td>0.094 ± 0.097</td>
<td>0.222 ± 0.275</td>
<td>0.761 ± 0.323</td>
</tr>
<tr>
<td>6</td>
<td>112.06</td>
<td>1307</td>
<td>1.0143</td>
<td>0.124 ± 0.146</td>
<td>0.210 ± 0.281</td>
<td>0.749 ± 0.331</td>
</tr>
<tr>
<td>7</td>
<td>116.91</td>
<td>1078</td>
<td>1.5130</td>
<td>0.109 ± 0.129</td>
<td>0.207 ± 0.275</td>
<td>0.764 ± 0.321</td>
</tr>
<tr>
<td>8</td>
<td>107.62</td>
<td>975</td>
<td>0.8887</td>
<td>0.110 ± 0.116</td>
<td>0.199 ± 0.269</td>
<td>0.753 ± 0.319</td>
</tr>
<tr>
<td>9</td>
<td>106.20</td>
<td>998</td>
<td>0.7978</td>
<td>0.103 ± 0.119</td>
<td>0.201 ± 0.268</td>
<td>0.718 ± 0.334</td>
</tr>
<tr>
<td>10</td>
<td>85.09</td>
<td>888</td>
<td>1.9673</td>
<td>0.124 ± 0.159</td>
<td>0.183 ± 0.261</td>
<td>0.773 ± 0.325</td>
</tr>
<tr>
<td>11</td>
<td>127.94</td>
<td>743</td>
<td>1.0540</td>
<td>0.114 ± 0.119</td>
<td>0.221 ± 0.288</td>
<td>0.763 ± 0.332</td>
</tr>
<tr>
<td>12</td>
<td>81.32</td>
<td>788</td>
<td>0.7978</td>
<td>0.103 ± 0.119</td>
<td>0.201 ± 0.268</td>
<td>0.718 ± 0.334</td>
</tr>
<tr>
<td>13</td>
<td>84.54</td>
<td>786</td>
<td>0.8492</td>
<td>0.108 ± 0.111</td>
<td>0.197 ± 0.256</td>
<td>0.743 ± 0.324</td>
</tr>
<tr>
<td>14</td>
<td>77.74</td>
<td>760</td>
<td>1.3020</td>
<td>0.102 ± 0.114</td>
<td>0.196 ± 0.268</td>
<td>0.730 ± 0.341</td>
</tr>
<tr>
<td>15</td>
<td>76.32</td>
<td>730</td>
<td>1.5503</td>
<td>0.105 ± 0.113</td>
<td>0.175 ± 0.242</td>
<td>0.728 ± 0.338</td>
</tr>
<tr>
<td>16</td>
<td>65.96</td>
<td>595</td>
<td>1.0907</td>
<td>0.111 ± 0.117</td>
<td>0.188 ± 0.252</td>
<td>0.712 ± 0.336</td>
</tr>
<tr>
<td>17</td>
<td>65.18</td>
<td>544</td>
<td>1.6434</td>
<td>0.120 ± 0.155</td>
<td>0.190 ± 0.250</td>
<td>0.773 ± 0.319</td>
</tr>
<tr>
<td>18</td>
<td>75.56</td>
<td>728</td>
<td>1.5007</td>
<td>0.104 ± 0.117</td>
<td>0.206 ± 0.268</td>
<td>0.747 ± 0.320</td>
</tr>
<tr>
<td>19</td>
<td>69.17</td>
<td>620</td>
<td>0.8494</td>
<td>0.112 ± 0.116</td>
<td>0.181 ± 0.246</td>
<td>0.736 ± 0.340</td>
</tr>
<tr>
<td>20</td>
<td>61.83</td>
<td>601</td>
<td>1.5310</td>
<td>0.103 ± 0.120</td>
<td>0.209 ± 0.261</td>
<td>0.740 ± 0.333</td>
</tr>
<tr>
<td>21</td>
<td>53.27</td>
<td>510</td>
<td>0.8726</td>
<td>0.105 ± 0.113</td>
<td>0.177 ± 0.248</td>
<td>0.714 ± 0.346</td>
</tr>
<tr>
<td>22</td>
<td>64.92</td>
<td>567</td>
<td>1.0297</td>
<td>0.114 ± 0.116</td>
<td>0.189 ± 0.265</td>
<td>0.731 ± 0.342</td>
</tr>
<tr>
<td>23</td>
<td>43.13</td>
<td>394</td>
<td>0.6059</td>
<td>0.110 ± 0.109</td>
<td>0.168 ± 0.236</td>
<td>0.698 ± 0.351</td>
</tr>
<tr>
<td>24</td>
<td>51.67</td>
<td>486</td>
<td>1.0018</td>
<td>0.107 ± 0.115</td>
<td>0.201 ± 0.279</td>
<td>0.754 ± 0.333</td>
</tr>
<tr>
<td>25</td>
<td>48.56</td>
<td>462</td>
<td>1.6778</td>
<td>0.105 ± 0.128</td>
<td>0.167 ± 0.230</td>
<td>0.714 ± 0.325</td>
</tr>
<tr>
<td>26</td>
<td>46.00</td>
<td>445</td>
<td>1.0455</td>
<td>0.104 ± 0.124</td>
<td>0.164 ± 0.234</td>
<td>0.703 ± 0.345</td>
</tr>
<tr>
<td>27</td>
<td>51.58</td>
<td>461</td>
<td>2.0018</td>
<td>0.112 ± 0.145</td>
<td>0.219 ± 0.282</td>
<td>0.746 ± 0.330</td>
</tr>
<tr>
<td>28</td>
<td>45.00</td>
<td>462</td>
<td>1.5914</td>
<td>0.109 ± 0.117</td>
<td>0.198 ± 0.264</td>
<td>0.747 ± 0.330</td>
</tr>
<tr>
<td>Overall*</td>
<td>2584.55</td>
<td>23063</td>
<td>2.8671</td>
<td>0.110 ± 0.127</td>
<td>0.198 ± 0.264</td>
<td>0.747 ± 0.330</td>
</tr>
</tbody>
</table>

1Between adjacent SNP. *Overall mean. SNP= single nucleotide polymorphism. ('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).
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 desequilibrio de ligação para pares de marcadores sintéticos separados por até 1 Mb, de acordo com as estatísticas $r^2$ e D' e frequência (%) de pares com $r^2$>0.30).

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

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 be useful. Values of LD (above 0.20, according to $r^2$) 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 % 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.
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 (Thévenon 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.
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 effective 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 Khatakar 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 (GBEV) 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 $r^2$ (Khatakar 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 considered low, indicating that inbreeding must be a matter of concern in breeding programs for this breed.

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 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 density 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.

ACKNOWLEDGMENTS

Haroldo H. R. Neves and Daiane C. B. Scalez were supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) fellowships. Sandra A. Queiroz was granted a research fellowship from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Genotype data were generated within the project FUGATO-plus GenoTrack, which was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. We are thankful to Mr. Luiz Antônio Josakian, from Associação Brasileira de Criadores de Zebu, who provided the pedigree data.

BIBLIOGRAPHY


