Extent Of Linkage Disequilibrium In Brazilian Gyr Dairy Cattle Based On Genotypes Of AI Sires For Dense SNP Markers

C.R. Silva[*], H. H. R. Neves[*], *S. A. Queiroz*[*], J. A. D. Sena[§], E. C. G. Pimentel[‡],

Introduction

Gyr dairy cattle breed has been widely used for dairying in tropical region of Brazil, mainly in crossbreeding with breeds specialized for milk production (predominantly Holstein). These *Bos taurus x B. indicus* cows have shown to excel in profit in the low to medium input systems prevailing in the country (Guimarães et al. (2005)), mainly due to heterotic gains in traits like milk production, reproductive efficiency, productive longevity and survivability.

Genomic selection (GS) is expected to increase noticeably the rates of genetic gain in dairy gain mostly due to possibility of reduction in generation interval maintaining a high accuracy of selection. Hayes et al. (2009) argue that this strategy should double the rate of genetic gain in dairy industrial and Gyr dairy production could also be benefit from this technology. The success of fine-scale mapping and genomic selection depends mainly on the strength of linkage disequilibrium (LD) between markers and causal mutations (Sargolzaei et al. (2008)), in a way that the extent of LD within a population determines marker density required for successful association studies and subsequent implementation of GS. Amount of LD is equally important as a source of information about historical events of recombination, allowing inferences about genetic diversity and genomic regions that have undergone selection.

This study was carried out to assess the extent of linkage disequilibrium in Brazilian Gyr dairy cattle using genotypes of AI sires for dense SNP markers.

Material and methods

Genotypes of 25 AI sires of Brazilian Gyr dairy cattle breed, whose semen is currently marketed in Brazil, were obtained with Illumina's BovineSNP50 beadchip. Of a total of 54001 SNPs genotyped, 18098 markers had minor allele frequency (MAF) greater than 10% and were considered in LD measures. For each autosome (BTA), the genotypes were submitted to fastPHASE (Scheet & Stephens (2006)). Based on phased genotypes, two measures of linkage disequilibrium (r^2 and D') were calculated for all pairs of markers in each chromosome using GOLD software (Abecasis & Cookson (2000)). Interchromosomal

[*]DZ-FCAV-UNESP - Jaboticabal, SP. CEP:14884-900. Brazil.
[§]DBAA-FCAV-UNESP - Jaboticabal, SP. CEP:14884-900. Brazil.
[‡]Institute of Animal Breeding and Genetics, University of Göttingen, Göttingen, 37075, Germany

heterogeneity in LD was investigated by fitting a general linear model to analyze r^2 data for all syntenic pairs of markers. This model included effects of chromosome as classification variable and log transformed physical distance.
Similar approach to that used by Sargolzaei et al. (2008) was employed to estimate effective population size (Ne) based on LD measures. At different genetic distances, effective size was calculated as $Ne=((1/r^{2*})-1)/(4c)$, where c is mean recombination distance (in morgans) and r^{2*} is r^2 averaged across all SNP pairs in the given c range. It was assumed that 1 centimorgan(cM) equals 1 Mb. The estimate of Ne based on mean LD in any range c is expect to reflect effective population size at approximately (1/2c) generations ago. In this way, average r^2 was calculated for all SNP pairs in each of 24 distance classes (from 0.1 to 34 cM), in order to estimate Ne from 2 to 500 generations back.

Results and discussion

Descriptive statistics regarding to SNPs markers and linkage disequilibrium measures are presented in Table 1. Autosomes differed in length such that BTA25 was the shortest (43.08 Mb) and BTA1 was the longest (160.88 Mb) and the whole autosomal genome comprised 2696.83 Mb. SNP loci density varied among chromosomes, ranging from 8.5 SNPs / Mb (BTA6) to 5.8 SNPs / Mb(BTA5). Even after filtering SNP data, a considerable proportion of SNPs had MAF below 20% (Figure 1a). Overall mean of r^2 for syntenic pairs was 0.028.

Mean LD between adjacent markers averaged across all autosomes was about 0.21 and 0.68, measured according r^2 and D', respectively, confirming results of previous studies that pointed out that D' overestimated extent of LD, especially in case of low MAF. LD showed a clear exponential trend of decay with physical distance. At ranges 0 to 0.1 Mb, 0 to 0.2 Mb, 0 to 0.5 Mb and 0 to 1Mb, average r^2 was 0.202, 0.180, 0.142 and 0.112, respectively. In these same ranges, 22.91%, 19.69%, 14.05% and 9.54% of SNP pairs exhibited r^2 larger than 0.3, respectively. These results agree reasonably with findings of Mackay et al. (2007) with eight breeds, for which useful LD for association studies does not extended by more than 0.5 Mb.

Estimated Ne showed a trend of decay along time (Figure 1b), as also observed in populations of Holstein (Sargolzaei et al. (2008)) and West African Cattle (Thenevón et al. (2007)). This decay could be attributable to intense selection, what explains the lower Ne in Gyr and Holstein population compared to West African (raised in pastoral extensive systems). Estimates of Ne in the present study must be considered as a rough approximation. Even so, values calculated for Ne in recent generations (about 54 and 39, three and two generations ago) are in reasonable agreement with estimates below 50 individuals that can be calculated using the average inbreeding coefficients reported by Queiroz et al. (2000) and Schenkel et al. (2002) for this population. The low Ne in Gyr dairy cattle indicates that inbreeding must be considered in breeding and mating decisions to maintain long-term genetic diversity in this breed. Genomic selection could also be useful for this objective, for example by screening a larger number of selection candidates than in conventional progeny test and by capturing Mendelian sampling term when estimating breeding values (Hayes et al. (2009)). Significant effects of chromosome and log transformed physical distance influenced r^2 values, as also observed in Holstein (Sargolzaei et al. (2008)). This inter-chromosome heterogeneity could be result of intense selection in both breeds.

Table1: Summary of analyzed SNP markers for each autosome (BTA)

BTA	Length (Mb)	SNP (n)	Mean distance ± SD (Mb)[1]	Longest gap (Mb)	Mean r^2 ± SD[1]	Mean D' ± SD[1]
1	160.88	1158	0.14±0.16	1.460	0.21±0.26	0.69±0.34
2	140.55	923	0.16±0.19	2.002	0.23±0.27	0.71±0.35
3	127.85	867	0.15±0.18	1.637	0.21±0.25	0.68±0.35
4	123.94	874	0.15±0.16	1.212	0.20±0.25	0.69±0.35
5	125.80	733	0.18±0.24	2.396	0.24±0.29	0.68±0.36
6	122.47	1037	0.12±0.13	0.826	0.24±0.27	0.70±0.34
7	112.06	781	0.15±0.19	2.976	0.22±0.27	0.71±0.34
8	116.91	810	0.15±0.17	1.620	0.21±0.25	0.69±0.34
9	107.31	755	0.15±0.17	1.362	0.22±0.27	0.69±0.34
10	106.03	735	0.15±0.20	2.665	0.20±0.24	0.68±0.35
11	109.83	814	0.14±0.16	1.236	0.21±0.26	0.68±0.35
12	85.07	501	0.18±0.23	2.495	0.17±0.24	0.67±0.35
13	84.19	597	0.15±0.16	1.180	0.23±0.28	0.70±0.35
14	81.30	626	0.14±0.16	1.299	0.22±0.26	0.67±0.34
15	84.46	624	0.14±0.15	1.027	0.23±0.26	0.69±0.33
16	77.54	607	0.13±0.15	1.541	0.19±0.25	0.65±0.36
17	76.09	567	0.14±0.16	1.550	0.19±0.24	0.66±0.35
18	65.93	466	0.15±0.19	2.233	0.19±0.22	0.63±0.35
19	64.94	446	0.15±0.19	1.643	0.22±0.26	0.72±0.34
20	75.53	565	0.14±0.16	1.501	0.23±0.27	0.69±0.34
21	68.91	476	0.15±0.16	1.157	0.20±0.24	0.67±0.36
22	61.82	481	0.13±0.15	1.531	0.22±0.25	0.68±0.34
23	52.96	407	0.14±0.16	0.997	0.19±0.24	0.65±0.35
24	64.32	446	0.15±0.15	1.129	0.20±0.26	0.67±0.34
25	43.08	318	0.14±0.15	0.869	0.18±0.24	0.60±0.36
26	51.51	375	0.14±0.18	2.215	0.23±0.29	0.68±0.34
27	48.36	381	0.13±0.16	1.889	0.18±0.23	0.64±0.34
28	45.91	356	0.13±0.16	1.156	0.18±0.22	0.63±0.35
29	50.39	372	0.14±0.18	2.002	0.24±0.29	0.71±0.34
Overall	2696.83	18098	0.14±0.17	1.614	0.21±0.26	0.68±0.35

[1]Between adjacent SNP. [2]Overall mean. SNP: single nucleotide polymorphism

Sample size of current study is lower than minimum suggested by Khatkar et al. (2008) for this type of study and is expected that LD measures may be biased at some degree. Further studies considering a larger sample are recommended to provide more accurate estimates of extent of LD in this population. The amount of LD estimated in present study is expected to allow estimating genomic breeding values (GEBV) with accuracy about 0.8 (Hayes et al. (2009)), what is crucial for successful GS programs in Gyr dairy cattle, but using larger dense SNP maps would allow to achieve higher power in association studies and also to capture QTL in regions that remained uncovered at current density.

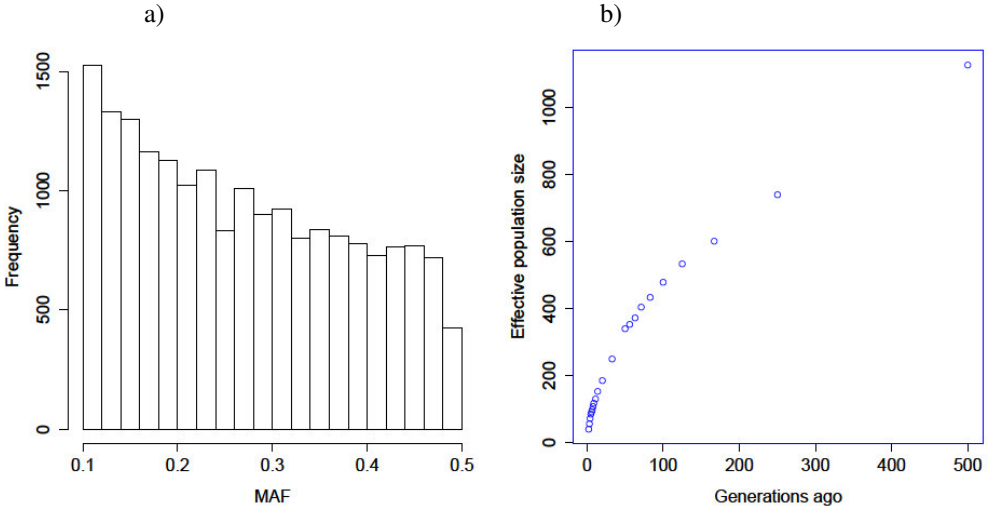

Figure 1: a) Histogram of minor allele frequency of SNPs. b) Estimated effective population size over time in Brazilian Gir dairy cattle.

Conclusion

Useful LD was estimated for markers until 100 kb apart when screening a small sample of Gyr dairy cattle. Denser SNP maps could allow achieving higher power in association studies. Trend of decay in effective population size over time was observed for this breed and estimates of recent Ne are low. Future studies about extent of LD using a larger sample of animals in this population are needed to confirm estimates in this study.

References

Abecasis, G., Cookson, W. (2000). *Bioinformatics*, 16**:**182-183.

Guimarães, H., Madalena, F. and Cezar, I. (2006). *Agri.Sys.*, 88:111-124.

Hayes, B., Bowman, P., Chamberlain, A. *et al.* (2009). *J. Dairy Sci.*, 92:433–443

Khatkar, M. Nicholas, F., Collins, A. *et al.* (2008). *BMC Genomics*, 9:187.

Mackay, S., Schnabel, R., Murdoch, B. *et al.* (2007). *BMC Genetics,* 8:74.

Queiroz, S., Albuquerque, L. and Lanzoni, N. *Rev. Bras. Zoot.*, 29:1014-1019.

Sargolzaei, M., Schenkel, F., Jansen, G. *et al.* (2008). *J. Dairy Sci.*, 91:2106-2117.

Scheet, P., Stephens, M. (2006). Am J Hum Genet 78: 629–644.

Schenkel, F., LaGioia, D. and Riboldi, J. (2002). In *Anais IV SBMA,* CD-ROM.

Thévenón, S., Dayo, G., Sylla, S. *et al.* (2007). *Anim. Genet.*, 38:277-286.