Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Estimation of Linkage Disequilibrium, Persistence of Phase and Effective Population Size of Brazilian Hereford and Braford Breeds

P. Biegelmeyer[*], M. M. Oliveira[†], *L.L. Cardoso*[†], C. C. G. Gomes[†], R. H. Higa[‡], N.J.L. Dionello[*], A. R. Caetano[§], J. P. Steibel[#], and F.F. Cardoso[†].
[*]Federal University of Pelotas, Pelotas, [†]Embrapa Southern Region Animal Husbandry, Bagé,
[‡]Embrapa Agriculture Informatics, [§]Embrapa Genetic Resources & Biotechnology, Brasília, Brazil,
[#]Michigan State University, East Lansing, USA.

ABSTRACT: A set of 41,241 SNP genotypes from 2,435 Hereford (HH) and Braford (BO) bovines were analyzed to estimate linkage disequilibrium (LD) levels, persistence of phase and effective population size of these populations. LD levels were estimated using the squared correlation of alleles at two loci (r^2) at varying distances. Average r^2 between adjacent SNP was 0.21 for HH and 0.16 for BO. Average inter-marker distance was 61 kb in both breeds. Useful LD values (r^2>0.2) were observed at 0-60 kb bins in HH and 0-20 kb bins in BO. Breeds demonstrated moderate to strong persistence of phase at all distances (range=0.53-0.97). The greatest phase correlations (r>0.9) were found in 0-50 kb bins. LD estimates decreased rapidly with increasing distances between SNPs, however, useful LD was observed in genomic regions spanning up to ~50 kb.
Keywords: beef cattle; effective population size; linkage disequilibrium; persistence of phase

Introduction

The occurrence of linkage disequilibrium (LD) between genetic markers and quantitative trait loci (QTL) is fundamental for performing genome-wide association (GWAS) and genomic selection (GS) studies since both rely on the non-random association between markers and functional mutations affecting the trait of interest (Goddard and Hayes (2009); Hayes et al. (2013)). Levels of LD are influenced by genetic and non-genetic processes affecting natural and domesticated populations, such as selection, mutation, genetic drift, non-random mating and population structure (Pritchard and Rosenberg (1999); Sabeti et al. (2007); Smith and Kuhner (2009)). LD also reflects the historical rates of recombination between loci, which are the basis for many selection studies (Nielsen, 2005).

Association studies between genetic markers and phenotypes can improve the understanding of genetic mechanisms influencing complex traits and eventually lead to the identification of genes responsible for the expression of economically important phenotypes. If genetic markers are in LD with genes involved with the expression of a trait, it is expected that their effects will be the same in the whole population (Espigolan et al. (2013)). So, the extent of LD between the genetic markers and QTL is a key factor affecting the accuracy and precision provided by association studies (Yan et al. (2009)), and can be helpful to indicate the density of SNPs that will be need to detect a QTL using LD analysis (McKay et al. (2007)).

In genomic selection (Meuwissen et al. (2001)), the genetic merit of animals is predicted using information from markers covering the whole genome, without a precise knowledge about QTL location or identification of the underlying polymorphisms. Nevertheless, sufficient LD between markers and QTL is needed, such that the marker allele-QTL phase persists across generations (de Roos et al. (2008)). The use of sufficiently dense marker map is fundamental to increase the accuracy of genomic estimated breeding values, since higher densities of marker panels are related to higher LD levels (Pimentel et al. (2013)).

The objective of this study was to estimate LD levels, persistence of phase and the effective population size in Brazilian Hereford (HH) and Braford (BO) cattle.

Materials and Methods

Genotypic data. A sample of 2,470 HH and BO cattle (with breed composition between ½ HH + ½ Zebu and ¾ HH + ¼ Zebu) born between 2008 and 2010 in commercial herds associated with the Brazilian Delta G Connection breeding program was used in this study. Genotyping of most of the samples was performed using the Illumina BovineSNP50 BeadChip (Illumina, USA), while 40 sires were genotyped with the Illumina High-Density (HD) Bovine BeadChip Array (Illumina, USA).

Quality control. Quality control was performed using the R snpStats package to remove samples with call rate <0.90, heterozygosity deviation >3.0 standard deviations, with mismatching sex assignment and misidentified duplications. Only autosomal markers were considered and quality control criteria excluded SNPs with call rates <0.98, minor allele frequencies (MAF) <0.03 and highly significant deviation from Hardy-Weinberg equilibrium ($P<10^{-6}$). The HD panel was filtered to select only the SNPs also present in the 50K panel. The final dataset contained 41,241 SNP from 2,435 samples. Sporadically missing genotypes were imputed during the phasing procedure using FImpute software (Sargolzaei et al. (2011)).

Statistical analyses. Linkage disequilibrium was calculated as pairwise r^2 (Weir (1996)), according to Badke et al. (2012), using ld_estimate R scripts. For each population, LD values between all pairs of SNPs of all chromosomes were binned according to the physical distance separating the loci. Average values of r^2 were calculated for each bin. The correlation of linkage phase (r) for pairs of SNPs between the two breeds was calculated according to Badke et al. (2012). SNPs pairs were binned according to the inter-marker distances, and average values of r were calculated for each bin, using markers common to both breeds. The decay of LD measures in relation to different inter-SNP distances was used to infer the effective population size (Ne).

Results and Discussion

Linkage disequilibrium. Average r^2 ± SD between adjacent SNPs across all chromosomes was 0.21 ± 0.27 for Hereford and 0.16 ± 0.22 for Braford. The analyses revealed a rapid decrease in the LD with increasing physical distances in both breeds (Figure 1). LD was greater than 0.2 and 0.3, respectively, for 34% and 25% of adjacent markers in Hereford, and 26% and 17% in Braford. Useful LD values ($r^2 > 0.2$) were observed in 0-60 kb bins in Hereford (range = 0.20 – 0.49), and in 0-20 kb bins in Braford (range = 0.21 – 0.43). Average $r^2 > 0.3$ was verified in 0-1 kb bins in Hereford (0.49) and Braford (0.43).

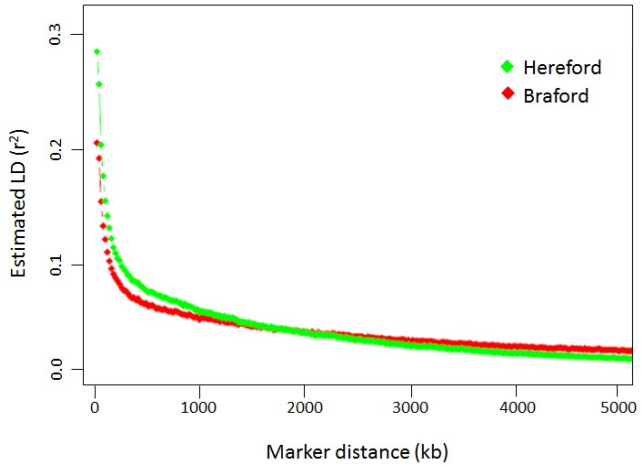

Figure 1: Extent of linkage disequilibrium (estimated r^2) as a function of inter-marker distance in Hereford and Braford populations

Moderate LD levels observed in this study were similar to the results presented by other authors. Average r^2 estimates of 0.205 and 0.20 were reported for Blonde d'Aquitaine (Berghain et al. (2012)) and North American Holstein (Bohmanova et al. (2010)) bulls, respectively. Using a high density SNP panel and 795 genotyped Nelore steers, Espingolan et al. (2012) observed an overall average r^2 of 0.17, and reasoned that lower levels of LD observed in *B. indicus* in comparison to *B. taurus*, especially at shorter distances, can be a consequence of the lower MAF generally observed in indicine breeds. Larger historical Ne in the *B. indicus* breeds as consequence of differences occurred during the *B. indicus* and *B. taurus* domestication and selection processes, is an alternative explanation for the lower LD detected in indicine breeds (Tenesa et al. (2007); McKay et al. (2007)). The lower average r^2 observed in this BO population in relation to HH can be a consequence of the mentioned differences between *B. indicus* and *B. taurus* breeds, considering the BO breed should have an average of 3/8 zebu.

Effective population size. The HH population showed higher LD than BO up to a distance of 1.78 Mb, while for larger distances BO showed a mean r^2 slightly higher than HH. These results suggest a smaller Ne for BO in recent generations, while HH has faced smaller Ne in past generations. Ne estimates for both breeds across generations are presented in Figure 2, indicating that current Ne for BO is ~220 while HH is ~153. The results indicate that two generations ago Ne in HH (Ne = 209) was higher than in BO (Ne = 161). Despite these drastic contrasts in recent generations, Ne in both populations show a general decline from older to present generations. From about thirty generations ago and before, BO showed again a larger effective size (Ne = 315) than HH (Ne = 309).

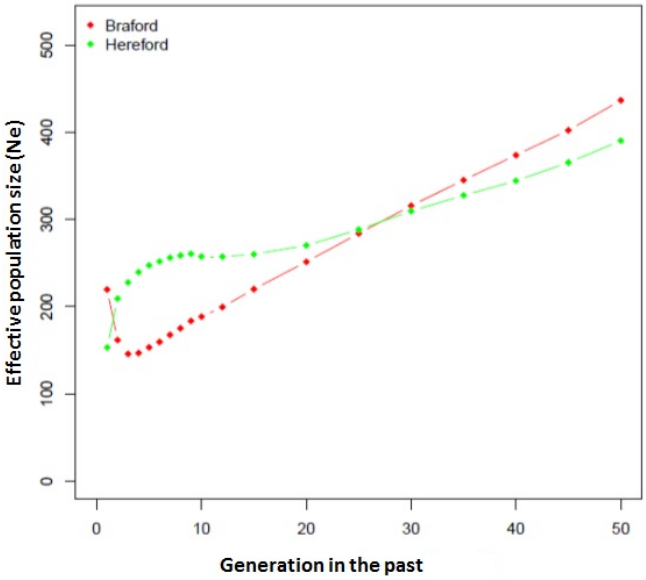

Figure 2: Effective population size as a function of past generation in Hereford and Braford populations

As we demonstrated in Figure 2, Ne for both breeds has declined over time, which can be a consequence of bottlenecks associated with the historical process of domestication and breed formation (Bovine Hap-Map Consortium (2009)). After the development and wide adoption of artificial insemination, the intense use of some animals for reproduction and high selection pressures applied for specific traits contributed to significant drops in Ne in the last ~50 years (Hayes et al. (2008)).

Persistence of phase. Moderate to strong persistence of phase at all distances (range = 0.53 – 0.97) were observed in both breeds (Figure 3). The greatest phase correlations (r > 0.9) were found in 0 to 50 kb bins (range = 0.92 – 0.97). Phase correlation decreased rapidly with increasing distance between SNPs, similarly as observed in average r^2. High marker phase correlation values are observed when gametic phase of two markers in LD are similar in two distinct populations (Uimari and Tapio (2011)), indicating that phase estimates from one population can be inferred from studies in another population. The observed proportion of estimated r values with reversed signs when considering BO and HH data was low, ranging from 5% to 34%.

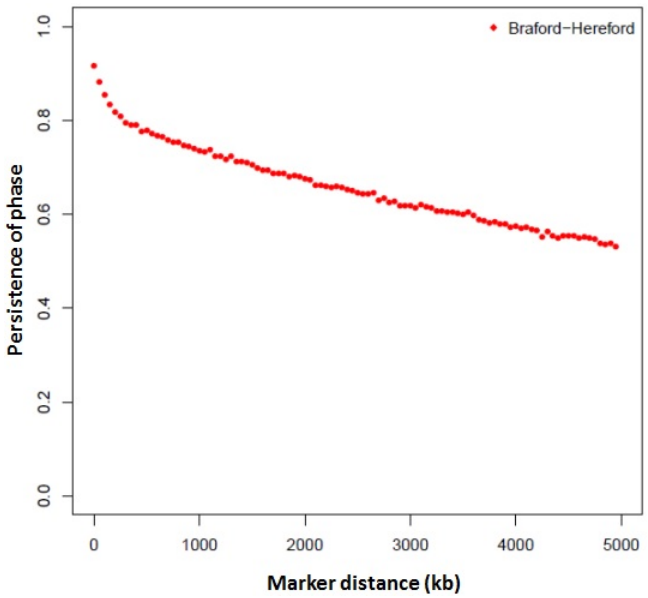

Figure 3: Phase correlation between Brazilian Hereford and Braford populations for SNPs pairs at varying marker distances

Conclusion

Our observations indicate that at least 50K equally spaced SNPs are necessary to delineate effective association and genomic selection studies in Brazilian BO and HH cattle. With SNP distances lower than 50 kb, QTL alleles are expected to have similar effects in both breeds due to high persistence of phase. Our results also indicate that at 50K SNP density pooling genotypes and phenotypic records from HH and BO will increase power in GWAS and genomic selection studies with these breeds.

Acknowledgments

Research supported by CNPq - National Council for Scientific and Technological Development grant 478992/2012-2, Embrapa - Brazilian Agricultural Research Corporation grants 02.09.07.004 and 01.11.07.002.07, and CAPES - Coordination for the Improvement of Higher Level Personnel grant PNPD 02645/09-2. Authors acknowledge the Delta G Connection for providing animals and data for this research.

Literature Cited

Badke, Y. M., Bates, R. O., Ernst, C. W., et al. (2012). BMC Genomics, 13:24

Beghain, J., Boitard, S., Weiss, B., et al. (2013). J. Anim. Breed. Genet., 130:294-302

Bohmanova, J., Sargolzaei, M., and Schenkel, F. S. (2010). BMC Genomics, 11:421

de Roos, A. P. W., Hayes, B. J., Spelman, R. J., et al. (2008). Genetics, 179(3):1503-1512

Espigolan, R., Baldi, F., Boligon, A. A., et al. (2013). BMC Genomics, 14:305

Goddard, M. E., and Hayes, B. J. (2009). Proc 18th AAABG, 26-29

Hayes, B. J., Lewin, H. A., and Goddard, M. E. (2013). Trends Genet., 29(4): 206-214

Hayes, B. J., Lien, S., Nilsen H., et al. (2008). Anim. Genet. 39(2):105-111

McKay, S. D., Schnabel, R. D., Murdoch, B.M., et al. (2007). BMC Genetics, 8:74

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Genetics, 157:1819-1829

Nielsen, R. (2005). Annu. Rev. Genet., 39:197-218

Pimentel, E. C. G., Wensch-Dorendorf, M., König, S., et al. (2013). Genet. Sel. Evol., 45:12

Pritchard, J. K., and Rosenberg, N. A. (1999). Am. J. Hum. Genet., 65:220-228

Sabeti, P. C., Varilly, P., Fry, B., et al. (2007). Nature, 449: 913-918

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2011). J. Dairy Sci. 94, E-Suppl. 1: 421 (333).

Smith, L.P., and Kuhner, M. K. (2009). Genet. Epidemiol., 33:344-356

Tenesa, A., Navarro, P., Hayes, B.J., et al. (2007). Genome Res., 17(4):520-526

The Bovine HapMap Consortium. (2009). Science, 324(5926):528-532

Uimari, P., and Tapio, M. (2011). J. Anim. Sci., 89(3):609-614

Yan, J., Shah, T., Warburton, M. L. et al. (2009). PLoS ONE, 4(12)