Characterization of linkage disequilibrium and consistency of gametic phase in Canadian goats

***L. F. Brito*[1], M. Jafarikia[2], D. A. Grossi[1], L. Maignel[2], M. Sargolzaei[1, 3] and F. S. Schenkel[1]**
[1]Centre for Genetic Improvement of Livestock, University of Guelph, Canada, [2]Canadian Centre for Swine Improvement Inc, Canada, [3]The Semex Alliance, Guelph, Canada.

ABSTRACT: The objective of this study was to estimate genome-wide levels of linkage disequilibrium (LD) in six Canadian goat breeds and to evaluate the persistence of gametic phase using the Illumina goat SNP50 BeadChip. A total of 976 animals were included in the analyses. The average values of LD between adjacent SNPs were 0.14, 0.15, 0.24, 0.19, 0.27 and 0.28 for Alpine, Saanen, Toggenburg, LaMancha, Nubian and Boer breeds, respectively. This indicates that, with a large enough training population, genomic selection could potentially be implemented within breed with the current 50k panel, but the Alpine and Saanen breeds might benefit from a denser panel. The gametic phase consistency (correlation between the signed square root of r^2) between adjacent markers was not high enough (~0.50) to support the pooling of breeds in a training population for genomic selection. For multi-breed genomic evaluation, a denser SNP panels seems to be required.
Keywords:
Genomic selection
Goat 50k panel
LD
Multi-breed population

Introduction

In recent years, genomic selection has been successfully implemented in dairy cattle breeding programs and has been presented as a promising tool to increase productivity in other species. One important parameter to evaluate before implementing genomic selection in a breeding program is the extent of linkage disequilibrium in the population. Linkage disequilibrium (LD) is defined as a nonrandom association of alleles at two or more loci and is influenced by population history, breeding system and the pattern of geographic subdivision (Slatkin, (2008)). A successful association analysis and, consequently, genomic selection depend on the marker density required, which is determined by the LD extent across the genome (Khatkar, et al. (2008)). A low LD level would require a higher marker density to enable markers to capture most of the genetic variation in the population.

The accuracy of genomic selection also depends on the number of records available to estimate marker effects (training population). It can be increased by combining data from multi-breed populations. However, the estimation of the effect of each marker across breeds requires not only high LD between the markers and the QTL in each breed, but also high persistence of gametic phase between the markers and the QTL across breeds. Persistence of gametic phase is a measure of the degree of agreement of gametic phase for pairs of markers between two populations (Badke, (2012)).

However, basic understanding of LD, as well as the persistency of phase across breeds, is limited in Canadian goats and has lagged behind when compared to the other species. Therefore, the objective of this study was to estimate genome-wide levels of linkage disequilibrium in six Canadian goat breeds and to evaluate the persistence of gametic phase between different breeds in order to evaluate the possible use a multi-breed training population for genomic selection.

Material and Methods

Animals and genotypes

A set of 976 goats from six breeds were genotyped using the Illumina goat SNP50 BeadChip containing 53,347 SNPs. The total number of genotyped animals for each breed by sex is described in Table 1. The animals were from 25 Ontario and Quebec herds, 2 artificial insemination (AI) centres and the Agriculture and Agri-Food Canada (AAFC) centre for animal genetic resources (Saskatoon, Saskatchewan).

For genotyping quality control, single nucleotide polymorphisms (SNPs) with minor allele frequencies (MAF) lower than 5% (for Alpine and Saanen breeds) or 15% (for other breeds, which have a much smaller number of genotyped animals) were filtered out. This procedure aimed to select only SNPs that were segregating in a significant number of animals in the sample. All the animals had less than 10% of missing genotypes, so no animal was excluded due to low call rate. Only autosomal SNPs were considered in this study and SNPs without chromosome number and/or position information were also excluded. After applying the described filtering criteria, a total of 45,338; 45,010; 33,761; 38,836; 33,125 and 35,858 SNPs were kept for further analyses for Alpine, Saanen, Toggenburg, LaMancha, Nubian and Boer breeds, respectively.

Determining the extent of linkage disequilibrium

The extent of LD between markers was measured by using r^2, as proposed by Hill and Robertson (1968), which is the squared correlation between alleles at 2 loci. r^2 was calculated for each pair of loci on each chromosome to determine the LD between adjacent SNPs, as well as the LD decay over different distances. To evaluate the pattern of LD along chromosomes, data was sorted into groups based on pair-wise marker distance and the average of each group was calculated.

Consistency of phase

The consistency of gametic phase was defined by the Pearson correlation of signed r values between two breeds. For each marker pair with a measure of r^2, the signed r value was determined by taking the square root of the r^2 value and assigning the appropriate sign based on the calculated disequilibrium (D) value. Data was sorted into groups based on pair-wise marker distance to determine the breakdown in gametic phase consistency across distances, as well as to be able to assess the phase consistency at the smallest distances possible, given the number of genotyped SNPs. For each distance group, the signed r values were then correlated between all six breeds using the CORR procedure in SAS (SAS Institute Inc., Cary, USA).

Results and Discussion

Levels of LD. The average of LD and distance observed between the adjacent SNPs in the six breeds are presented in Table 2. After the quality control, the 50k panel showed good coverage of the genome with an average gap size varying from 0.05 to 0.07 Mb. The breeds with a smaller number of animals had the largest average gap size between adjacent SNPs due to the exclusion of SNPs with minor allele frequency (MAF) lower than 0.15, while for Alpine and Saanen breeds was used a MAF threshold of 0.05.

Figure 1 displays the average r^2 values at given distance ranges for each breed. High LD values (r^2) were observed only at small distances between pairs of SNPs. For all the breeds, the average r^2 decreased with increasing marker distance. The average r^2 estimates, across all distances, for Alpine and Saanen breeds were the lowest values and they were similar for both breeds, which could be explained by their common ancestral origin.

Boer and Nubian had the highest levels of LD across all distances. LD levels were smaller than 0.05 at distances greater than 1.2 Mb for all breeds. Trends across distances were very similar (Figure 1), for all breeds and the LD level decayed at a very similar rate. The extent of LD decreased substantially from the first (up to 0.02 Mb) to the second range of distances (between 0.02 and 0.03 Mb). The number of SNP pairs at distances < 0.02 Mb was quite small (452 pairs, on average across breeds) though. The low level of long range LD may indicate that these breeds have not been under intense selection.

Alpine and Saanen were the breeds with the largest sample sizes. The higher observed levels of LD in the other 4 breeds could be due to sampling, but they are more likely due to smaller effective population size in those breeds. Therefore, it would be interesting to confirm the LD results obtained in this investigation using a larger number of genotyped animals from these 4 breeds. A higher level of LD is related to a higher accuracy of genomic estimated breeding values. Some studies (e.g. Calus (2008); Meuwissen et al. (2001)) have concluded that an r^2 value greater than 0.2 would be sufficient for genomic selection. At the average distance between adjacent SNPs in the goat 50k SNP panel (~0.06 Mb) most of the breeds exceeded or approached this value, except for Alpine and Saanen (r^2 ~ 0.12 - 0.13). This indicates that, with a large enough training population, genomic selection could potentially be implemented with the current 50k panel within breed, but Alpine and Saanen might benefit from a denser panel.

Carillier et al. (2013) found average r^2 value of 0.17 between adjacent SNPs using 50k panel in French Alpine and Saanen breeds. These authors also evaluated the contribution of genomic information to the accuracy of genetic evaluation and reported some potential gains. However, the expected gains were slightly lower when compared to gains obtained in other species. According to them, the lower gains were due to the structure and size of the training population.

Linkage phase. The consistency of phase between breeds indicates whether or not different breeds could potentially be pooled into one common training population to better estimate SNP effects. For goat genomic evaluation this would be very important due to the fact that there is a small number of genotyped animals in the small population-sized breeds. In Figure 2 is presented the consistency of gametic phase (Pearson correlation between signed r values) between all pairs of breeds. The highest consistency of phase was found between Alpine and Saanen, suggesting a greater level of relatedness between these breeds. However, even for these two breeds, the consistency of phase between adjacent markers was not high enough (~0.50) to support the pooling of breeds in a training population for genomic selection. In general the estimated correlations were low for all breeds even for short distances.

In dairy cattle, De Roos et al. (2009) evaluated the effect of combining multiple populations on the reliability of genomic predictions and concluded that the benefits of combining populations in a training set were higher when the populations have diverged for only a few generations ago, when the marker density was high, and when heritability was low. From the simulation studies reported by these authors, populations that had diverged 6 generations ago presented a correlation of phase higher than 0.8 for distances up to 0.45 Mb. Therefore, for multi-breed genomic evaluation in goats, a denser SNP panel seems to be required. For implementing genomic selection using the 50k panel in goat breeds, other ways to increase the training population should be sought, such as genotyping more animals in each breed or collaborate with other countries and share genotypes and phenotypes (EBVs) for genomic selection.

Conclusion

At the average distance between adjacent SNPs in the current 50k SNP panel (~0.06 Mb), most of the 6 Canadian breeds examined exceeded or approached the level of linkage disequilibrium that is useful for genomic prediction, except for Alpine and Saanen. This indicates that, with a large enough training population, genomic selection could potentially be implemented within breed with the current 50k panel, but the Alpine and Saanen breeds might benefit from a denser panel.

The highest phase consistency was found between Alpine and Saanen, indicating a greater level of relatedness between these two breeds. However, even for these two

breeds, the phase consistency between adjacent markers is not high enough to encourage the pooling of breeds in a single training population for genomic selection. For multi-breed genomic evaluation, a denser SNP panel seems to be required. Therefore, other ways to increase the training population for genomic selection using the 50k panel should be sought, such as genotyping more animals in each breed and/or collaborating with other countries for sharing genotypes and phenotypes (EBVs).

Acknowledgements

The authors thank the following organizations for providing funds and collaborating within the project: the sector councils of Quebec, Ontario and British-Columbia, who administer the Canadian Agricultural Adaptation Program (CAAP) for Agriculture and Agri-Food Canada; Ontario Goat; Société des éleveurs de chèvres laitières de race du Quebec; GoatGenetics.Ca; and the Brazilian Government through the Science without Borders Program that provides graduate fellowship for the first author.

Literature Cited

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

Calus, M. (2008). *Genetics*, *178*, 553.

Carillier, C., Larroque, H. Palhière, I. et al. (2013) J. Dairy Sc., 96: 7294-7305.

De Roos, A. P. W., Hayes, B. J., Goddard, M. E. (2009) *Genetics*,183:1545–1553.

Hill, W.G.G., Robertson, A.: (1968). Theor. Appl. Genet., 38:226–231.

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

Meuwissen, T., Hayes, B., and Goddard, M. (2001). Genetics, 157(4), 1819-1829

Slaktin, M. (2008). Nat. Rev. Genet. 9:477-485.

Table 1: Frequency of genotyped animals by breed and sex.

Breed	Males	Females	Total
Alpine	51	352	403
Saanen	51	267	318
LaMancha	11	70	81
Nubian	21	33	54
Toggenburg	7	46	53
Boer	17	50	67
Total	158	818	976

Table 2: Average linkage disequilibrium (r^2) between adjacent SNPs by breed and average distance between adjacent SNPs (Mb) in the six breeds.

Breed	r^2	Average distance between adjacent SNP (Mb)
Alpine	0.1445	0.05296
Saanen	0.1534	0.05336
LaMancha	0.1934	0.06181
Nubian	0.2721	0.07247
Toggenburg	0.2431	0.07132
Boer	0.2860	0.06701

Figure 1: Average r^2 values at given distances for six goat breeds.

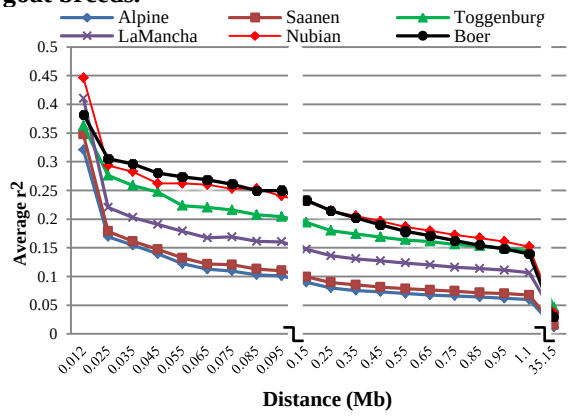

Figure 2: Consistency of gametic phase (Pearson correlations of signed r values) at given distances for 15 goat breed pairs.