Linkage Disequilibrium in Canadian Swine Breeds

M. Jafarikia[*], L. Maignel, S. Wyss, B. Sullivan

Introduction

Knowledge about linkage disequilibrium (LD) or non-random association of the genetic markers at two or more loci is important for carrying out genome wide selection (Meuwissen *et al.*, 2001) and mapping quantitative trait loci (QTL) (Meuwissen and Goddard, 2000). The availability of high density porcine single nucleotide polymorphism (SNP) panels to the swine industry in the past year has contributed to a better understanding of LD on a much larger scale than previously available. The application of high density porcine SNP panels in genomic selection within or across swine breeds is still under development around the world. Having thousands of SNPs on a single panel is very beneficial for the success of genomic selection. The number of genotyped animals and technical approaches required for successful implementation of genomic selection are highly dependent on LD. Therefore, the objective of this study was to investigate LD patterns within and across the major Canadian swine breeds.

Material and methods

Animal genotyping. Animals were sampled from herds across Canada and included 215 Duroc, 145 Landrace and 283 Yorkshire pigs. Genotyping was performed by DNA Landmarks (dnalandmarks.com) using the Illumina 60K (64,232) SNP array.

Filtering SNPs. All SNPs were tested for possible inconsistencies in Mendelian segregation. Du *et al.* (2007) reported that biases in LD could be influenced by low (less than 10%) minor allele frequency (MAF). Therefore, SNPs with less than 10% MAF were excluded from the LD analysis. All SNPs were also tested for the Hardy-Weinberg Equilibrium (HWE) and SNPs found to have significant disequilibrium ($p<0.01$) within a breed were excluded. SNPs with call frequency of less than 0.30 and SNPs on the sex chromosomes were also excluded.

Haplotype phasing. The study of LD and the consistency of linkage phases required phased haplotypes and fastPHASE software (Scheet and Stephens, 2006) was used for this purpose.

Linkage disequilibrium. The two preferred methods to measure LD are D´ and r^2 (Ardlie *et al.* 2002; Zhao *et al.*, 2005; Du *et al.* 2007) where r^2 has been reported as being a more robust measure of LD and less dependent on allele frequencies than D´ (Du *et al.* 2007). For this study, r^2 was used to measure the LD. It is defined by Hill and Robertson (1968) as the squared correlation of alleles at two sites and is defined as:

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

[*] Canadian Centre for Swine Improvement (CCSI), Building #54, 960 Carling Avenue, Central Experimental Farm, K1A 0C6, Ottawa, ON, Canada.

where, r^2 is a measure of LD ranging from 0 to 1, D=f(AB)-f(A)f(B), f(AB) is the frequency of the AB haplotype and f(A), f(a), f(B) and f(b) are the frequencies of alleles A, a, B and b, respectively. The negative or positive square root of r^2 (using the root with the same sign as D) was used as a measure of the LD phase.

Breed and Chromosome effects. Decay of LD was analyzed including effect of distance between SNPs (as a covariate), breed, chromosome and the interaction between chromosome and breed using the GLM procedure of SAS (2002-2003).

Results and discussion

Genotypes. The total number of SNPs used in the analysis was 37,109 after filtering out 27,123 SNPs for one or more of the following reasons. There were 12,585 SNPs that had MAF lower than 0.10 and 1,689 SNPs had call frequency less than 0.30. There were 3,445 SNPs that were not in HWE in at least one of the three breeds under study and 14,555 SNPs that did not have known chromosome information, of which most of these (12,592 SNPs) had unknown position on the chromosome. A total of 1,402 SNPs located on sex chromosomes were excluded, of which 1,381 were on X and 21 on Y.

Linkage disequilibrium. Table 1 illustrates the distribution of SNPs by chromosome and the rate of LD of adjacent SNPs within the Duroc, Landrace and Yorkshire breeds. Figure 1 shows the distribution of the average r^2 using the first three chromosomes of each of the breeds for illustration. In the third chromosome an unexpected increase in mean LD was observed in the range of 60 to 80 Mbp. One possible reason for this could be errors in SNP locations. Improvements in mapping of the pig genome will help if this is the case. To further investigate this phenomenon, the average LD for different distances within the three breeds was calculated (Table 2). The averages and standard deviations of LD were 0.31±0.38, 0.31±0.34 and 0.33±0.34 for Duroc, Landrace and Yorkshire, respectively. Du *et al.* (2007) reported r^2 of 0.3 to be sufficient for capturing the QTL. A total of 36, 38 and 40 percent of adjacent SNPs had r^2 of more than 0.3 in Duroc, Landrace and Yorkshire, respectively. To explore the potential application of genomic selection across breeds, the consistency of linkage phases within the three breeds were compared. Higher correlations of linkage phases between Yorkshire and Landrace breeds were observed than for these two breeds with Duroc. Using all SNP pairs, the correlation was 0.34 between Yorkshire and Landrace while the correlation of these two breeds with Duroc was 0.21 and 0.20, respectively. Correlations increased to 0.82, 0.65 and 0.66, respectively, when using only SNP pairs with distances of less than 50kb. The values increased to 0.99, 0.97 and 0.97, respectively, when using only SNP pairs with distances of less than 50kb which also had LD of more than 0.3.

Effect of breed and chromosome. Different levels of LD over various distances by breed are shown in table 2. Effects of breed, chromosome and interaction between breed and chromosome were highly significant on levels of LD. This would suggest caution on the application for genomic selection across the breeds, while high levels of LD and MAF frequency within each breed are promising for within breed application of genomic selection.

Table 1: Distribution of SNPs and rate of LD by chromosome and breed (average ± standard deviation)

SSC	SNPs	Average Interval	Average r^2 between adjacent SNPs		
		Mbp	Duroc	Landrace	Yorkshire
1	4931	0.06±0.18	0.35±0.40	0.37±0.36	0.39±0.35
2	2221	0.07±0.20	0.32±0.38	0.28±0.31	0.30±0.31
3	1694	0.09±0.25	0.26±0.35	0.26±0.31	0.27±0.31
4	2897	0.05±0.11	0.34±0.37	0.33±0.34	0.36±0.35
5	1686	0.06±0.21	0.31±0.38	0.29±0.33	0.31±0.33
6	1557	0.11±0.29	0.26±0.36	0.25±0.32	0.26±0.32
7	2784	0.05±0.06	0.35±0.38	0.33±0.34	0.33±0.33
8	1818	0.08±0.20	0.29±0.36	0.27±0.32	0.28±0.33
9	2122	0.07±0.23	0.27±0.36	0.29±0.32	0.32±0.34
10	1118	0.07±0.23	0.25±0.32	0.26±0.31	0.26±0.31
11	1510	0.06±0.13	0.28±0.34	0.30±0.34	0.28±0.32
12	896	0.08±0.21	0.28±0.35	0.24±0.29	0.27±0.32
13	2986	0.07±0.21	0.30±0.39	0.32±0.34	0.35±0.37
14	3249	0.05±0.05	0.38±0.40	0.42±0.37	0.41±0.36
15	2083	0.08±0.26	0.32±0.38	0.29±0.35	0.32±0.34
16	1328	0.07±0.20	0.28±0.35	0.29±0.33	0.29±0.32
17	1350	0.04±0.04	0.30±0.35	0.30±0.32	0.31±0.32
18	879	0.07±0.16	0.30±0.36	0.28±0.32	0.28±0.33

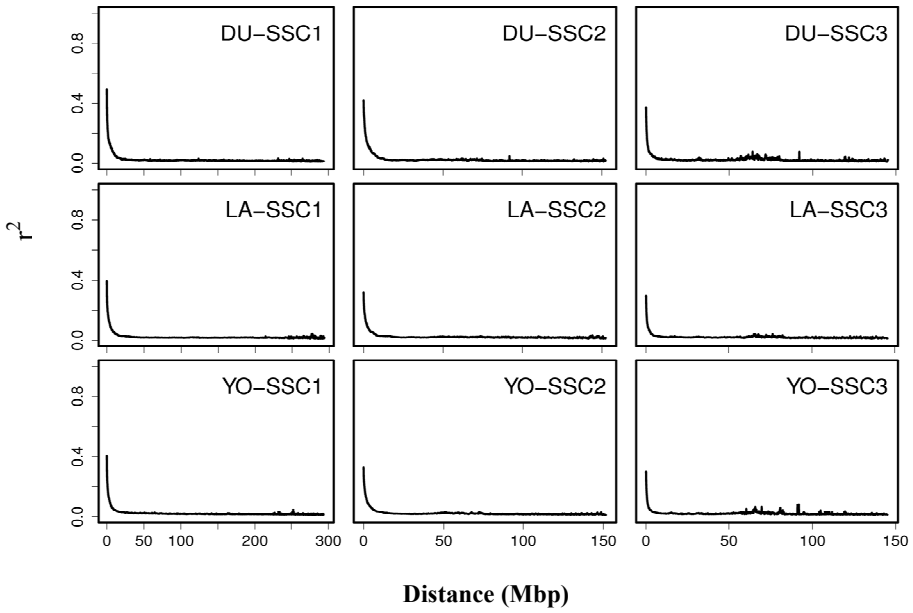

Figure 1: Estimation the average LD levels by distance between SNP pairs for Duroc (DU), Landrace (LA) and Yorkshire (YO) across the first three chromosomes (SSC1, SSC2 and SSC3).

Table 2: Distribution of LD by breed and distance (average ± standard deviation)

SNP pair distance (kb)	Breed	All adjacent SNP pairs		% adjacent SNPs with r^2> 0.3
		SNP pairs	r^2±SD	
0.0-50	Duroc	26904	0.33±0.38	38
	Landrace		0.33±0.35	40
	Yorkshire		0.34±0.35	42
50-100	Duroc	6541	0.29±0.36	34
	Landrace		0.29±0.33	35
	Yorkshire		0.31±0.33	37
100-200	Duroc	2200	0.26±0.35	30
	Landrace		0.25±0.31	30
	Yorkshire		0.27±0.32	33
200-1000	Duroc	1194	0.19±0.29	22
	Landrace		0.17±0.25	18
	Yorkshire		0.18±0.26	22
All distances	Duroc	37091	0.31±0.38	36
	Landrace		0.31±0.34	38
	Yorkshire		0.33±0.34	40

Conclusion

These findings would suggest good potential for application of the current SNP panel for genomic selection within breed. Across breed application may be feasible for the dam line breeds (Yorkshire and Landrace) using SNP pairs showing high levels of LD (>0.3) along very short distances (<50kb), although this would decrease the number of usable SNPs to less than half of the within breed numbers. Further research is underway to develop programs for the application of genomic selection in Canadian swine breeds.

Acknowledgements

The financial support of Agriculture and Agri-Food Canada (ACAAF program) and the contributions of DNA samples from Canadian Duroc, Yorkshire and Landrace breeders and of additional genotypes from Fast Genetics Inc. are gratefully acknowledged.

References

Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). *Nat. Rev. Genet.*, 3:299-309.

Du, F. X., Clutter, A. C., Lohuis, M. M. (2007). *Int. J. Biol. Sci.*, 3:166-78.

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

Meuwissen, T. H. E., and Goddard, M. E. (2000). *Genetics*, 155:421–430.

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

Scheet, P., and Stephens, M. (2006). *Am. J. Hum. Genet.*, 73(5):1162–1169.

Zhao, H., Nettleton, D., Soller, M. *et al.* (2005). *Genet. Res.*, 86(1):77–87.