

Mapping genes for quantitative traits using linkage disequilibrium

ME Goddard *

*Centre for Genetic Improvement of Livestock, Animal and Poultry Science,
University of Guelph, Guelph, Ontario, Canada N1G 2W1*

(Proceedings of the 9th European Colloquium on Cytogenetics of Domestic Animals;
Toulouse-Auzeville, 10-13 July 1990)

quantitative trait loci / marker-assisted selection / linkage disequilibrium

INTRODUCTION

Many economically important traits of livestock are quantitative in nature and controlled by many genes (quantitative trait loci or QTLs) and environmental factors. If we could find genetic markers linked to QTLs, this information would be useful for marker-assisted selection (MAS) and eventually might help us to identify and clone the QTL itself. Markers can be shown to be linked to a QTL for the trait under study by within family linkage studies. However, large numbers of offspring per family are required and the linkage phase must be determined for each family. This reduces the value of the information for MAS. The presence of other loci and environmental factors influencing the trait also makes it difficult to map the QTLs with any precision.

Linkage disequilibrium can cause associations between alleles at marker and QTL which are consistent across the population, rather than being restricted to a family. Such associations would be very useful for MAS as discussed by Lande and Thompson (1990). Because linkage disequilibrium usually occurs only with close linkage, it may be useful for more precise mapping of QTLs.

The proportion of the genetic variance controlled by the QTL which, due to linkage disequilibrium, can be explained by genetic markers is an important parameter. It influences both the value of the marker for MAS and the size of the experiment necessary to detect the effect (Lande and Thompson, 1990). The proportion of variance explained might be increased by using marker haplotypes at a number of closely linked markers instead of a single marker locus.

This paper describes the factors controlling the proportion of QTL variance explained by markers, including marker haplotypes, and the use of this information for mapping the QTL. Linkage disequilibrium can occur through crossbreeding,

* Currently on leave from Department of Agriculture and Rural Affairs, Victoria, Australia.

epistasis or finite population size. Here I will consider mainly the effect of finite population size and briefly mention crossbreeding.

MATERIALS AND METHODS

The following model was investigated by simulation. Four linked loci were equally spaced on the chromosome with a recombination rate between the outside loci of 'c'. The second locus was a QTL and the other 3 were genetic markers. There were 2 alleles per locus and initially all gene frequencies were 0.5 and there was linkage equilibrium. Finite population size (n) caused some inbreeding and linkage disequilibrium to accumulate. Each generation, $2n$ gametes were sampled after allowing for recombination in the parents.

The proportion of QTL variance explained by the markers (P) was calculated in 3 ways according to the marker information used: 1) for individual marker loci (P_1); 2) using a multiple regression based on the 3 loci (P_2); and 3) using the 8 marker haplotypes defined by the 3 loci (P_3). (This is equivalent to including the interactions between marker loci in the multiple regression.)

The number of replicates of each simulation was 400 ($n = 32$) or 200 ($n = 128$) or 100 ($n = 512$), so that the mean P values presented in tables I and II have standard errors of approximately 0.01.

RESULTS AND DISCUSSION

The results of the simulation depend upon 3 parameters: the population size (n), the recombination rate (c) and the number of generations (t). However, Hill and Robertson (1968) found for disequilibrium between two loci that if nc was held constant and t was scaled in proportion to n , the results were approximately constant. This proposition was investigated for the current 4 locus model and the results are presented in table I. The results show that if nc and t/n were held constant, P_3 was approximately the same. Similar results apply to P_1 and P_2 provided n is not too low. Consequently, the 3 parameters can be reduced to 2. Constant t/n implies almost equal inbreeding coefficient (F) and this is also given in table I.

Table I. Proportion of variance explained by marker haplotypes (P_3) as a function of population size (n), generations of inbreeding (t), recombination rate (c) and inbreeding (F).

F	n :	32		128		512	
	c :	0.1		0.025		0.00625	
		t	P_3	t	P_3	t	P_3
0.02		1	0.11	4	0.10	16	0.10
0.05		3	0.24	12	0.26	48	0.22
0.10		7	0.35	28	0.38	112	0.35
0.20		14	0.43	56	0.43		

The effect of nc and F on proportion of variance explained are shown in table II. P increased with F but reached a maximum and, although not shown in table II, declined at very high F values. The maximum for P_1 was much lower than the equilibrium expression ($1/(4nc/3 + 1)$) given by Sved (1971) because replicates were included where the marker locus became fixed for one allele. Marker haplotypes explain a much higher proportion of variance than single markers, especially at low F values. A linear combination of the markers (P_2) was intermediate between P_1 and P_3 . The third marker added significantly to both P_2 and P_3 showing that markers, other than those bracketing the QTL, are useful. Decreasing nc (*ie*, closer linkage) had little effect on P at very low F values, but caused P to approach a higher maximum as F increased.

Table II. Proportion of variance explained by markers as a function of inbreeding coefficient (F) and nc .

F	nc								
	0.8			3.2			6.4		
	P_1^a	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
0.02	0.01	0.05	0.11	0.01	0.05	0.11	0.01	0.05	0.11
0.05	0.05	0.14	0.29	0.05	0.13	0.24	0.04	0.12	0.24
0.10	0.10	0.28	0.47	0.09	0.22	0.35	0.07	0.19	0.31
0.20	0.17		0.59	0.14	0.31	0.43	0.10	0.22	0.32
0.30	0.24		0.64	0.14		0.43	0.10	0.22	0.30

^a P_1 : a single marker at recombination distance $c/3$ from QTL; P_2 : multiple regression equation based on 3 markers; P_3 : marker haplotypes based on 3 markers.

At levels of F accumulated over time in livestock breeds (0.10–0.30), the amount of variance explained by marker haplotypes would make them very valuable for MAS. Intense selection of males within some breeds causes the effective population size to be quite low. Young *et al* (1988) found the rate of inbreeding in the USA *Holstein* population to be 0.2%/yr corresponding to a generation effective population size of about 50. If $n = 100$ and $c = 0.032$, so that $nc = 3.2$, the average distance between markers is 1.6 cM. To cover the whole genome with markers at this density would require about 2000 markers. A more practical approach would be to map an important QTL to a 10–20 cM region using family linkage studies and then to find additional markers within this region for the study of linkage disequilibrium.

Lande and Thompson (1990) conclude that past crossbreeding will be a more important source of linkage disequilibrium than finite population size. However, they appear to be thinking of crosses between highly inbred lines. For lines which diverge due to genetic drift, the expectation of P_1 can be shown to be proportional to F_{ST}^2 , where F_{ST} is Wright's coefficient of divergence between populations. Consequently, at low F_{ST} values typical of livestock breeds, P_1 is very small. In the simulations reported above, crossbreeding could be simulated by crossing two independently derived replicates. When this was done, P values in the synthetic population were lower than in the replicates prior to crossing.

On the average, markers more tightly linked to a QTL show higher P values than less tightly linked markers, implying that P values would be useful in establishing the map position of the QTL. However, the variation in P between replicates is very large. The coefficient of variation is approximately 50% for P_3 , 70% for P_2 and over 100% for P_1 . Consequently, differences in P values between markers will seldom be significant. Hill and Weir (1988) previously pointed out the high variability of disequilibrium coefficients. Perhaps the only conclusion that can be drawn is that, if a high P value is found in a large population, the marker and QTL must be closely linked.

The effect of initial gene frequencies other than 0.5, multiallelic loci, selection and mutation on these results need further investigation.

CONCLUSIONS

Linkage disequilibrium between genetic markers and QTLs should cause relationships between the markers and quantitative traits which apply to an entire population or breed. The use of haplotypes based on several closely linked markers would make these relationships useful for marker-assisted selection.

ACKNOWLEDGMENTS

This research was supported by funds from the Natural Sciences and Engineering Research Council, Semex, Canada, and the Victorian Department of Agriculture and Rural Affairs.

REFERENCES

- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* 38, 226-231
- Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol* 33, 54-78
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol* 2, 125-141
- Young CW, Bonczek RR, Johnson DG (1988) Inbreeding of and relationship among registered Holsteins. *J Dairy Sci* 71, 1659