EFFECTS OF SELECTIVE GENOTYPING STRATEGIES ON THE PRECISION OF FINE MAPPING

Al. Stella[1], P. Boettcher[2] and P. Leone[2]

[1] CERSA and [2] IDVGA-CNR, 20090 Segrate, Italy

INTRODUCTION

Linkage mapping studies in livestock have been used to detect the presence of numerous apparent quantitative trait loci (QTL) (e.g. Andersson *et al.,* 1994 and Georges *et al.,* 1995). In general, these studies have positioned QTL in chromosomal regions spanning from 10 to 40 cM. While this level of precision is sufficient for some applications of marker assisted selection, additional fine mapping must be done to identify the genes responsible for these effects. In addition, results from linkage studies may be only confidently applied for selection within the families in which linkage phase between markers and QTL has been established. For these reasons, efforts are being made to detect linkage disequilibrium (LD) between markers and QTL that exists across populations. The power (per genotype) of linkage analysis for the detection of quantitative trait loci can be increased markedly by the selective genotyping of particular individuals, usually those with extreme phenotypes (Lander and Botstein, 1989). Selective genotyping has been used successfully in livestock (e.g. Kirkpatrick *et al.,* 2000) and is often used in linkage analysis of human disorders (e.g. Cardon *et al.,* 1994). In human studies, selection is typically based on strategies that also consider family structure, selecting multiple sibs with similar or dissimilar phenotypes, depending on the strategy applied (Cardon and Fulker, 1994). The properties of selective genotyping that improve the power of linkage mapping are likely to improve the precision of using LD for fine mapping, but the optimal strategy will likely depend on the trait, population and QTL of interest. Gonçalo *et al*. (2001) compared different strategies of selective genotyping for LD mapping in humans. The objective of this study was to compare the precision of fine mapping associated with several different selective genotyping strategies that may be applicable in livestock populations, in particular populations in which artificial insemination is employed to increase family size.

MATERIALS AND METHODS

Data. Simulation was used to generate data for the comparison of selective genotyping strategies. The design of the simulation was roughly based on the approach used by Meuwissen and Goddard (2000). It was assumed that previous linkage analysis studies had positioned a segregating QTL to lie within a small (20 cM) chromosomal region. The current study focused exclusively on this segment of the genome. New markers were placed within this region and associations with phenotypes were analyzed to estimate to which marker the QTL was most closely located.

The population consisted of 50,000 females and 125 males (N = 50,125 and $N_e \sim 500$). Genotypes for each animal consisted of two haplotypes for eleven evenly-spaced biallelic markers and an associated QTL and a normally distributed polygenic effect. The QTL was positioned near (recombination = 0.001) to the central marker. In the base generation,

individuals were randomly assigned two alleles at each of the eleven marker loci. Frequencies were set at 0.50 for all marker alleles at all loci. For the QTL locus, each animal was assigned 2 unique alleles. Polygenic effects were drawn from $N(0, \sigma^2_A)$. In future generations, marker and QTL genotypes were assigned according to rules of Mendelian inheritance and allowed for recominbation within the region. Polygenic effects were equal to means of parental effects plus a Mendelian sampling component. Random mating was simulated for 100 discrete generations. Analyses were based on the final generation. The QTL values and phenotypes were assigned only in this generation.

In effect, a biallelic QTL was simulated. At the end of 100 generations, most of the original 2N alleles had been lost due to random drift. The single existing allele that had a frequency closest to a predefined value was identified and assigned an effect to assure that the QTL accounted for a specified proportion of the genetic variance, and all other QTL alleles had effect zero. These predefined levels of QTL allelic frequency and variance were varied systematically and are shown in Table 1. Phenotypes were assigned by summing polygenic, QTL, and a random residual effect. Heritability varied and is shown in Table 1. Two types of phenotypes were simulated : 1) the resulting continuous values and 2) a binomial phenotype, defined by imposing a threshold such that highest 10% of the population on the continuous scale exhibited one phenotype and the remainder of the population had another. All 16 combinations of these variable parameters (Table 1 plus type of phenotype) were simulated.

Table 1. Parameters systemically varied in the simulation

	Levels	
Parameter	Low	High
Frequency of positive QTL allele	0.10	0.30
QTL variance (proportion of genetic variance)	0.10	0.30
Heritability	0.05	0.30

Selection strategies. For this study, genotyping of 1000 individuals (2%) was simulated. Five basic selection strategies were defined. The first was simple random selection (**RAN**) of individuals for genotyping. The second was "standard" selective genotyping (**STD**) in which the highest and lowest 1% were selected, regardless of family. The third strategy balanced genotyping across sires (**BAL**), choosing the 4 highest and lowest animals from each sire. The fourth strategy was similar to the *discordant selection* strategy applied in human studies (Gonçalo *et al.,* 2001). The 10 (20) sires with the most variability among daughters were identified and the highest and lowest 50 (25) daughters from each sire were genotyped (**DIS10**) (and **DIS20**). The final strategy was similar to *concordant selection* (Gonçalo *et al.,* 2001), in which siblings with similar extreme phenotypes are chosen. For this strategy, highest (lowest) daughters of the 5 highest (5 lowest) sires (based on daughter mean) were genotyped (**CON5**). This was repeated for the 10 highest (lowest) bulls (**CON10**). All strategies were applied in two approaches. Initially only the simple phenotypes of daughters was considered. Then all strategies were repeated based on daughter phenotypes adjusted for respective sire means. To test for the most likely location of the QTL, simple ANOVA of marker effects was performed for each of the eleven marker loci. The locus with the highest resulting F-test was considered

the most likely location of the QTL. The simulation was repeated 500 times for each combination of population parameters.

RESULTS AND DISCUSSION

Continuous phenotypes. Table 1 shows the proportion of times the correct location was identified for each strategy for selective genotyping based on unadjusted continuous phenotypes. Results for adjusted (by sire mean) were little different. Little advantage was obtained in considering family structure. In general, the best strategy was BAL, but performance of STD without regard for the distribution of daughters per sire was similar. In situations of high heritability (0.30), DIS20 also performed at a level similar to that of BAL. The strategies employing concordant selection (CON5 and CON10) were of little value for continuously distributed phenotypes, in some situations yielding precision of QTL position lower than that of RAN.

Table 2. Proportion (%) of times the correct location of the QTL was identified for various selection strategies and population parameters for continuous phenotypes

	QTL frequency = 0.10				QTL frequency = 0.30			
	h^2=0.05		h^2=0.30		h^2=0.05		h^2=0.30	
	QTL variance		QTL variance		QTL variance		QTL variance	
Strategy[A]	0.05	0.30	0.05	0.30	0.05	0.30	0.05	0.30
	----------(%)----------							
RAN	20.8	35.2	48.2	67.6	15.2	23.6	32.8	53.8
STD	**49.2***	**71.8***	74.4	86.0*	**38.4***	**60.6***	61.8	77.2*
BAL	47.4*	71.6*	**80.8***	**87.6***	37.8*	59.2*	**69.0**	**78.6***
DIS10	30.4	52.4	63.4	73.2	24.8	40.8	50.0	69.0
DIS20	37.8	60.4	76.0*	83.8*	28.8	50.6	62.4	77.0*
CON5	15.0	27.8	19.8	42.2	13.8	24.6	17.6	35.8
CON10	25.6	49.2	40.2	66.6	19.4	41.6	32.6	56.4

[A]RAN = random selection of individuals for genotyping, STD = selective genotyping, BAL = selective genotyping with balanced numbers of daughters per sire, DISn = discordant selection with n=10 or 20 sires, CONn = concordant selection with 5 or 10 high and low sires.
[B]Highest frequency is in bold and those not significantly ($P > 0.05$) different from the maximum are marked with '*'.

Binomial phenotypes. Results for the binomially expressed trait are in Table 3 and are notably different in several respects from the continuous trait (Table 2). First, major advantages were obtained by adjusting for sire mean, therefore, results from both approaches are shown. The best strategy was consistently one that selected based upon adjusted phenotypes (although differences were not always significant). Concordant selection (CON10) was advantageous when QTL effects were small. Discordant selection with a larger pool of sires (DIS20) was the best strategy overall, yielding either the highest precision or precision that was not significantly different from the optimal strategy in all situations except one. The BAL strategy also performed well in most situations, except for when a QTL was responsible for a large proportion (30%) of the genetic variance for a lowly heritable (h^2 = 0.05) trait. The STD approach was only competitive for QTL for highly heritable traits. Applying STD with adjusted phenotypes was useless, providing no indication of the position of the QTL.

Table 3. Proportion of times the correct location of the QTL was identified for various selection strategies and population parameters for binomial phenotypes

	QTL frequency = 0.10				QTL frequency = 0.30			
	h^2=0.05		h^2=0.30		h^2=0.05		h^2=0.30	
	QTL variance		QTL variance		QTL variance		QTL variance	
Strategy[A]	0.05	0.30	0.05	0.30	0.05	0.30	0.05	0.30
Unadjusted	----------(%)----------							
RAN	12.4	22.4	33.4	63.6	11.0	18.2	25.6	50.6
STD	16.2	35.4	53.0	74.4*	15.4	30.4	40.8*	65.0*
BAL	17.0	28.4	49.8	72.4	13.6	19.4	33.0	54.2
DIS10	18.4	35.4	47.2	70.6	15.2	36.6*	38.2	59.8
DIS20	20.0*[B]	40.8	47.0	73.0*	14.2	33.8	42.4*	63.2*
CON5	13.6	22.4	17.0	30.6	11.0	15.4	14.4	22.8
CON10	**24.0**	40.8	30.0	54.8	20.2*	36.2*	27.6	51.2
Adjusted								
RAN	11.6	20.6	30.3	60.0	11.8	16.4	23.6	48.4
STD	9.2	10.2	9.4	10.4	9.8	11.8	7.6	9.2
BAL	20.6*	37.4	**60.4***	**79.0***	16.0*	26.8	41.2*	63.6*
DIS10	21.2*	**49.4***	53.0	72.6	16.2	36.6*	43.0*	40.6
DIS20	21.8*	49.2*	53.8	76.8*	16.0*	**40.4***	**45.6***	**66.8***
CON5	13.8	21.8	17.2	31.6	11.2	16.2	14.0	23.2
CON10	**24.0**	40.0	30.2	54.4	**20.8***	35.0*	27.8	51.4

[A]RAN = random selection of individuals for genotyping, STD = selective genotyping, BAL = selective genotyping with balanced numbers of daughters per sire, DISn = discordant selection with n=10 or 20 sires, CONn = concordant selection with 5 or 10 high and low sires.
[B]Highest frequency is in bold and those not significantly ($P > 0.05$) different from the maximum are marked with '*'.

CONCLUSION

Selective genotyping increases the precision of fine mapping with LD. The best strategy for selection depends on the characteristics of the QTL and the phenotype. These results highlight the importance of experimental design for LD mapping studies of complex traits.

REFERENCES

Andersson, L., Haley, C.S., Ellegren, H., Knott, S.A., Johansson, M., Andersson, K. Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Håkansson, J. and Lundström, K. (1994) *Science* **263** : 1771-1774.

Cardon, L.R. and Fulker, D.W. (1994) *Am. J. Hum. Genet*. **55** : 825-833.

Cardon, L.R., Smith, S.D, Fulker, D.W., Kimberling, W.J., Pennington, B.F. and De Fries, J.C. (1994) *Science* **266** : 276-279.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, V., Steele, M.R., Zao, X., Womack, J.E. and Hoeschele, I. (1995) *Genetics* **139** : 907-920.

Gonçalo, R.A., Cookson, O.C. and Cardon., L.R. (2001) *Am. J. Hum. Genet.* **68** : 1463-1474.

Kirkpatrick, B.W., Byla, B.M. and Gregory, K.E. (2000) *Mamm. Genome* **11** : 136–139.

Lander, E. and Bostein, D. (1989) *Genetics* **121** : 185-199.

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