QTL MAPPING USING SELECTIVE DNA POOLING DATA

J. C. M. Dekkers, J. Wang, and R. Fernando

Department of Animal Science, Iowa State University, Ames, IA, 50011, USA

SUMMARY

A method to map QTL based on selective DNA pooling in half-sib family designs was developed. QTL position was estimated from marker frequency differences between progeny with high and low phenotypes and did not depend on differences in phenotypic means. The QTL effect was estimated by relating differences in phenotypic means to differences in QTL frequencies, which were estimated from QTL position and marker frequencies. Methods were first developed for a single family with two markers and then extended to multiple families and multiple markers. Simulation of a single half-sib family of 2,000 individuals with two markers showed that close to unbiased results were obtained with high power. Biases increased when measurement errors on marker frequencies increased and the QTL effect was small. Biases were minimal when multiple families and multiple markers were used.

Keywords: QTL mapping, genetic markers, selective DNA pooling

INTRODUCTION

Darvasi and Soller (1992) developed methods to identify markers linked to QTL based on differences in marker allele frequencies between individuals with high and low phenotypes. Darvasi and Soller (1994) showed that DNA pooling can be used to determine the required marker frequencies. Selective DNA pooling allows for a significant reduction in genotypings and was recently used to detect markers associated with milk protein content in dairy cattle by Lipkin *et al* (1998). Methods of Darvasi and Soller (1992, 1994) are, however, based on single marker analyses and do not allow separation of QTL position and effect. The objective of this study was, therefore, to develop methods to map QTL using selective DNA pooling data on flanking markers. Methods are initially derived for a single half-sib family with two markers and then extended to multiple families and markers.

MATERIALS AND METHODS

Single half-sib family. Consider an additive trait affected by a QTL and polygenes. The QTL is bracketed by two markers, M and N, with recombination rates r_1, r_2, and θ between M-Q, N-Q, and M-N. Fractions s of half-sib progeny with the highest and lowest phenotypes are identified. The common sire is heterozygous for all loci (MQN/mqn) and marker haplotypes are known. Let f_{LM}, f_{LN}, f_{UM}, and f_{UN} be the observed frequencies of M and N in the lower and upper tails, respectively, and p_{LM}, p_{LN}, p_{UM}, and p_{UN} the frequencies of progeny that received M and N from the sire. The latter can be obtained by adjusting observed marker frequencies by estimates of frequencies among dams (h): $p_{UM} = 2f_{UM} - h_M$.

Let p_{UQ} denote the (unobserved) upper tail frequency of progeny that received Q from the sire. Given p_{UQ}, expected frequencies of progeny in the upper tail that received M and N from the sire are:

$$E(p_{UM}) = (1-r_1)p_{UQ} + r_1 p_{Uq} = r_1 + (1-2r_1) p_{UQ} \quad [1]$$

$$E(p_{UN}) = (1-r_2)p_{UQ} + r_2 p_{Uq} = r_2 + (1-2r_2) p_{UQ} \quad [2]$$

Similar equations can be derived for the lower tail by replacing U with L.

Expected differences in paternal marker allele frequencies between the upper and lower tail are:

$$p_{UM} - p_{LM} = (1-2r_1)(p_{UQ}-p_{LQ}) = (1-2r_1)(2p_{UQ}-1) \quad [3]$$

and, assuming no interference, $p_{UN} - p_{LN} = (1-2\theta)(p_{UQ}-p_{LQ})/(1-2r_1) = (1-2\theta)(2p_{UQ}-1)/(1-2r_1)$ [4]

Solving equations [3] and [4] for r_1 and requiring r_1 to be less than _ results in:

$$r_1 = \_ - \_ V(1-2\theta)(p_{UM} - p_{LM})/(p_{UN} - p_{LN}) \quad [5]$$

Note that a solution to [5] does not exist if $(p_{UM}-p_{LM})$ and $(p_{UN}-p_{LN})$ differ in signs.

Estimates of p_{UQ} and p_{LQ} are obtained by substituting [5] in [3] and solving for p_{UQ}:

$$p_{UQ} = \_ + \_ (p_{UM}-p_{LM})/(1-2r_1) = \_ + \_ V(1-2\theta)(p_{UM} - p_{LM})(p_{UN} - p_{LN}) \quad [6a]$$

or

$$p_{UQ} = \_ - \_ (p_{UM}-p_{LM})/(1-2r_1) = \_ - \_ V(1-2\theta)(p_{UM} - p_{LM})(p_{UN} - p_{LN}) \quad [6b]$$

Once estimates of QTL frequencies are obtained, the QTL substitution effect _ can be estimated by relating QTL frequency differences to differences in trait means using methods developed by Darvasi and Soller (1994) for marker-associations. Let $\__L$, $\__U$, and _ denote the mean phenotype of progeny in the lower tail, the upper tail, and all progeny, respectively. Considering progeny in the upper and lower tail, the mean phenotype of progeny that received allele Q from the sire, deviated from _, is:

$$\mu_{UL.Q} = [p_{UQ}(\mu_U-\mu) + p_{LQ}(\mu_L-\mu)]/(p_{UQ}+p_{LQ}) \quad [7]$$

Similarly, the mean phenotype of progeny that received allele q from the sire is:

$$\mu_{UL.q} = [(1-p_{UQ})(\mu_U-\mu) + (1-p_{LQ})(\mu_L-\mu)]/(2-p_{UQ}-p_{LQ}) \quad [8]$$

Correcting for differences in mean phenotype due to factors other than the QTL under study based on selection intensity i_s, α can be estimated as (Darvasi and Soller 1994): $\alpha = (\mu_{UL.Q}-\mu_{UL.q})/i_s^2$ [9]

Multiple markers and multiple half-sib families. Methods described above can be extended to multiple markers and multiple families. Using [1], the observed frequency for marker k in tail j from sire i can be represented by the following mixed model:

$$p_{ijk} = r_k + (1-2r_k)p_{ijQ} + u_{ijk} + e_{ijk} \quad [10]$$

where u_{ijk} is a binomial sampling effect and e_{ijk} is a random measurement error in marker frequencies from DNA pooling. The model for a vector of frequencies on linked markers for sire i in tail j is:

$$\mathbf{p}_{ij} = \mathbf{r} + (\mathbf{1}-2\mathbf{r})\mathbf{p}_{ijQ} + \mathbf{u}_{ij} + \mathbf{e}_{ij} \quad [11]$$

where $Var(\mathbf{e}_{ij}) = \mathbf{I}\sigma_e^2$ and $Var(\mathbf{u}_{ij})$ is a matrix of covariances between frequencies of marker-QTL recombinants. For a given position of the QTL, $Var(\mathbf{u}_{ij})$ is known and has diagonal elements $(1-r_k)r_k/n$ (n is the number of progeny in the sample) and off-diagonals equal to $(1-r_k)r_k(1-2\theta_{kk'})/n$ for marker- QTL orders k'-k-QTL, with $\theta_{kk'}$ the (known) recombination rate between markers k and k', and off-diagonals equal to zero for marker-QTL orders k'-QTL-k. Similar sets of equations can be set up for other tails and families. Ignoring relationships, observations are independent among tails and families.

For a given position of the QTL, the vector of recombination rates between markers and the QTL, **r**, is known and, hence, model [11] can be fitted to observed marker frequencies using Henderson's mixed model equations, with residual variance, σ_e^2, estimated iteratively. Similar to regression interval mapping, the model is fitted with the QTL at positions 1 cM apart on the chromosome. The position with the lowest residual sum of squares gives the estimate of QTL position. Estimates of QTL frequencies p_{ijQ} at this position can be used to estimate QTL substitution effects for each sire using [9].

RESULTS

Methods were validated by stochastic simulation. A random measurement error was added to simulated marker frequencies. Results for a single family are in Table 1. Only replicates for which the estimate of r_l existed and fell between –0.5 and +0.5 were considered. In addition, the estimate did not always place the QTL inside the marker bracket. In these cases, the QTL was positioned at the nearest marker, which will be referred to as restricting r_l to the marker bracket. When the sire's marker alleles were absent among dams and measurement error was zero, estimates of QTL position and effect were nearly unbiased when QTL position was restricted to the bracket. Measurement errors caused a bias in position toward the center and increased the standard deviation of estimates. A substantial percentage of replicates had estimates that were outside the bracket or invalid (Tables 1), which increased when power decreased. Measurement errors caused a slight bias in estimates of the QTL effect. When the sire's marker alleles were also present among dams and observed frequencies were adjusted based on an estimate of marker frequencies among dams (true frequencies were used here), biases in estimates of QTL effect and position increased slightly (Table 1).

Initial results for mixed model analysis of a design with 7 families and 6 markers are in Table 2 for a QTL that explains 50 % (α=.50σ_p) or 12.5 % (α=.25σ_p) of genetic variance. Marker-QTL linkage phase differed by sire. Also, 50 % of sires were homozygous at the QTL and did, therefore, not contribute information. QTL position was estimated with little bias (Table 2) but the standard deviation of estimates was substantial. Estimates of QTL frequencies were unbiased within QTL genotype. In ongoing research, these estimates will be used to estimate sire QTL substitution effects.

DISCUSSION

An important result from this study is that the estimate of QTL position from selective genotyping depends entirely on differences in marker frequencies in selected groups (equation [5]) and not on differences in phenotypic means. With a single family and only two markers, the ability to map the QTL is limited, with estimates being invalid or outside the bracket for many replicates. These problems can be overcome with the mixed model analysis with multiple families and markers.

Estimates of frequencies of marker alleles from DNA pools are subject to measurement errors (Khatib *et al.* 1994). Additional errors are introduced by adjustments for marker frequencies among dams (Lipkin *et al.* 1998). These errors are expected to be independent between markers and, therefore, their impact can be reduced by considering markers that are external to the flanking markers when mapping a QTL within a given marker bracket. This is capitalized on in the mixed model by modeling covariances between linked markers.

Table 1. Means and standard deviations (in brackets) based on 5,000 replicates of estimates of QTL position and effect from selective DNA pooling data.

	Measurement error variance=0		Measurement error variance=.00278[1]	
	Dam freq[2]=0	Dam freq>0	Dam freq=0	Dam freq>0
		0=0.2 r_f=0.05 α=0.50σ_p		
r_f (x100)	4.7 (3.5)	4. 1 (6.7)	3.8 (9.8)	3.4 (11.3)
r_f restricted to bracket[3]	4.9 (3.0)	5.3 (4.4)	6.1 (5.7)	6.3 (6.1)
% outside[4] (% invalid[5])	8.8 (0.0)	22.3 (0.1)	29.5 (0.7)	33.5 (1.4)
α (x100)	49.3 (6.3)	49.9 (7.9)	50.5 (10.0)	51.0 (11.3)
		0=0.2 r_f=0.05 α=0.25σ_p		
r_f (x100)	3.8 (7.8)	3.2 (12.9)	4.2 (16.0)	4.7 16.8)
r_f restricted to bracket	5.4 (4.7)	6.7 (6.5)	7.9 (7.5)	8.3 (7.8)
% outside[4] (% invalid[5])	24.0 (0.5)	38.8 (4.0)	50.2 (9.5)	54.9 (13.2)
α (x100)	25.3 (6.0)	26.1 (7.6)	26.7 (9.9)	27.3 (10.9)

[1] Based on Lipkin *et al* (1998) for microsatellite markers.
[2] Frequency of the sire's marker alleles among dams, with adjustment based on true dam frequencies.
[3] Estimates of position outside the bracket were restricted to the nearest marker
[4] The percentage of replicates for which the QTL was mapped outside the marker bracket.
[5] invalid refers to the percentage of replicates for which the estimate of QTL position was invalid.

Table 2. Average estimates over 1,000 replicates of QTL position and frequencies based on a mixed model analysis of selective DNA pooling data.

QTL effect (α)	QTL position (cM) (True position = 46)		Difference in QTL allele frequencies between upper and lower tail					
			Qq sires		qQ sires		QQ or qq sires	
	Average	St.dev.	True	Estimate	True	Estimate	True	Estimate
0.50 σ_p	45.67	6.98	0.44	0.44	-0.44	-0.44	0.00	0.00
0.25 σ_p	45.87	16.10	0.23	0.23	-0.22	-0.23	0.00	0.00

This is in contrast to QTL mapping based on individual genotypes (e.g. Haley and Knott, 1992) where, assuming no genotyping errors and no interference, external markers provide no additional information to map a QTL in an interval. Including external markers as covariates can, however, reduce the impact of QTL outside the interval (Jansen 1993).

QTL mapping based on flanking markers in a half-sib design depends on knowing marker linkage phases in the common parent. With selective DNA pooling, frequencies at linked markers in pools selected based on phenotype give insufficient power to derive sire marker haplotypes; therefore, sire marker haplotypes must be obtained from other means. Finally, hypothesis testing and power to detect QTL based on selective DNA pooling will be addressed in subsequent research.

REFERENCES

Darvasi, A. and Soller, M. (1992) *Theor. Appl. Genet.* **85**:353
Darvasi, A. and Soller, M. (1994) *Genetics* **138**:1365
Haley, C. S. and Knott, S. A. (1992) *Heredity* **69**:315
Jansen, R. C. (1993) *Genetics* **135**:205
Khatib, H., Darvasi, A., Plotsky,Y. and Soller, M. (1994) *PCR Methods Appl.* **4**:13
Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A. and Soller, M. (1998) *Genetics* **149**:1557