Multiple-Marker Mapping for Selective DNA Pooling Within Large Families

M. Dolezal,*†¹ H. Schwarzenbacher,*†‡ M. Soller,§ J. Sölkner,* and P. M. Visscher†#

*Department of Sustainable Agricultural Systems, University of Natural Resources and Applied Life Sciences, Vienna, Gregor Mendel Str.33, 1180 Vienna, Austria

†Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh, UK

‡Chair of Animal Breeding at Technische Universität München, Hochfeldweg 1, 85354 Freising, Germany

§Department of Genetics, Hebrew University of Jerusalem, Jerusalem 91904, Israel

#Queensland Institute of Medical Research, Brisbane, Australia

ABSTRACT

Selective DNA pooling is a very powerful method for quantitative trait loci (QTL) mapping. It considerably reduces genotyping costs while maintaining high statistical power. Applied to a daughter design, milk samples of offspring with extreme phenotypic values for a trait of interest are assigned to high and low groups, respectively, and within each group the pooled DNA is used for densitometric estimation of allele frequencies in the 2 groups. A single-marker test for linkage between marker and QTL considers marker allele frequency differences between high and low groups. Single-marker across-sire test statistics are strongly affected by the number of sires that are heterozygous for a given marker and the QTL status (homozygous or heterozygous) of these sires, which decreases the accuracy of QTL mapping. Here we propose a simple method to deal with this problem by taking information from multiple linked markers into account. In particular, given the single-marker test statistics, a multiple-marker method was developed to predict test statistics for markers for which a sire was homozygous (or at any other location on the chromosome). Power and map resolution of the proposed method were assessed by simulation, and we show that for the same data set, multiple-marker mapping performed better than the commonly used single-marker analyses.

Key words: quantitative trait loci mapping, selective DNA pooling

INTRODUCTION

Selective DNA pooling in a daughter design has proven to be a powerful approach for detecting marker-

QTL linkage in cattle populations in which the commonly used granddaughter design is not feasible because of the undersized population structure on the male side. This approach has been used to map QTL for milk production traits (Lipkin et al., 1998; Mosig et al., 2001; Fisher and Spelman, 2004; Mariasegaram et al., 2002), ovulation rate (Gonda et al., 2004), twinning rate (Cobanoglu et al., 2005), clinical mastitis (Sharma et al., 2006), and beef marbling (Xiao et al., 2007). However, single-marker mapping for selective DNA pooling (SMM) as performed in these studies is limited by the fact that a different subset of tested sires will be heterozygous at each marker along a chromosome. Thus, SMM test statistics for across-sire analyses will be strongly affected by the number and QTL status of the specific sires that are heterozygous at a given marker. In selective DNA pooling experiments with limited numbers of sires, this frequently gives rise to a marker with a low test statistic flanked by markers with high test statistics. This constellation can be erroneously interpreted as 2 separate QTL, even though it is simply the result of lack of heterozygous sires at the intermediate marker.

Dekkers (2000) developed a least squares interval-mapping approach for selective DNA pooling based on flanking markers for single-sire analyses, which was advanced by Wang et al. (2007) to allow for multiple sires and multiple markers. Additionally, Wang et al. (2007) developed an approximate maximum likelihood approach for interval mapping with selective DNA pooling data. These methods deal effectively with the problem raised above. In addition, they provide separate information on QTL location and effect. However, they require sire marker haplotype information, which is often not available. In the present study, a method was developed for mapping of QTL with selective DNA pooling data (multiple-marker mapping, MMM) that does not rely on sire marker haplotype information. In contrast to the interval mapping methods of Dekkers (2000) and Wang et al. (2007) MMM does not provide separate estimates for QTL effect and QTL location.

Received May 31, 2007.

Accepted March 10, 2008.

¹Corresponding author: marlies.dolezal@gmx.at

The MMM methodology does, however, enable estimates of marker effects for all markers tested in all sires or, indeed, at any other location on the chromosome. Hence, it reduces the effect of unequal marker information of sires attributable to differing marker heterozygosity and provides aesthetically pleasing chromosome QTL maps.

MATERIALS AND METHODS

In selective DNA pooling, significance of marker i for a single sire j , heterozygous at i , is determined by the single-sire test statistic (Darvasi and Soller, 1994):

$$Z_{ij} = D_{ij}/SE(D_{ij}),$$

where D_{ij} is the sire marker allele frequency difference between daughters from the high tail of the phenotypic distribution assigned to the high pool and daughters from the low tail of the phenotypic distribution assigned to the low pool, and $SE(D_{ij})$ is the binomial standard error of D_{ij} .

Let $T_{ij} = Z_{ij}^2$. Then, under the H_0 of no linkage, $T_{ij} \sim \chi_{(1)}^2$. Significance of a marker i across n sires heterozygous at i is obtained by the multisire test statistics, $TS_i = \sum_{j=1}^n T_{ij} \approx \chi_{(n)}^2$, where T_{ij} is summed across all n sires that are heterozygous at marker i (Weller et al., 1990).

Given the observed single-sire test statistics (T_{ij}) for a series of markers that span a chromosome and for which a given sire is heterozygous, a method was developed to predict test statistics for markers for which the sire is homozygous (or at any other location on the chromosome), based on the markers for which the sire is heterozygous. The methodology is analogous to that used in a simple selection index to obtain the EBV of a target individual from the phenotypes of its relatives. In the EBV procedure, phenotypic information from several individuals having different degrees of genetic relationships to one another and to the target individual is combined to estimate the EBV of the target individual. In the procedure proposed here, the estimated single-sire test statistic corresponds to the EBV, a specific point on the chromosome corresponds to the target individual, the single-sire test statistic of a specific marker corresponds to the phenotype of an individual related to the target individual, and the variance-covariance matrix of single-sire test statistics at different points on the chromosome corresponds to the variance-covariance relationship matrix.

We first consider the pooling results from a single sire. The prediction of T_l at location l from an observation on T_i at location i , is $\hat{T}_l \sim (1 - 2r)^2 T_i$ (Soller and

Genizi, 1978), with r being the recombination fraction between the 2 loci. Hence, the predicted single-sire test statistic at an arbitrary locus is only a function of the proportion of recombination between that locus and the locus for which a single-sire test statistic was calculated. Under the null hypothesis of no linkage, the variance of T_i is 2 because of the $\chi_{(1)}^2$ assumption. The covariance between the predicted T_l and observed T_i is

$$\text{cov}(\hat{T}_l, T_i) \approx (1 - 2r)^2 \text{var}(T_i) = 2(1 - 2r)^2.$$

If the single-sire test statistic is observed at both locations i and l , then the covariance is equivalent to the previous expression:

$$\text{cov}(T_l, T_i) \approx (1 - 2r)^2 \text{var}(T_i) = 2(1 - 2r)^2.$$

Combining all the above results, a multipoint prediction of the test statistic at any location on the chromosome can be made by using a simple selection index analogy:

$$\hat{T}_l = \mathbf{b}'\mathbf{t},$$

where $\mathbf{t}$ is a vector of all observed test statistics (T_i), with $\mathbf{V} = \text{var}(\mathbf{t})$, the variance-covariance matrix of $\mathbf{t}$ with 2 on the diagonal and $(1 - 2r_{il})^2$ on off-diagonals, and $\mathbf{b}$ is the solution to $\mathbf{V}^{-1}\mathbf{c}$, with $\mathbf{c}$ being a vector of covariances between predicted (T_l) and observed (T_i) test statistics $[= 2(1 - 2r_{il})^2]$. The factor of 2 drops out of the equation and $\mathbf{V}^{-1}$ has a simple (tridiagonal) form when using the Haldane mapping function (Visscher, 1996).

Hence, for each sire separately, the single-sire test statistic at each location on the chromosome is predicted from the observed single-sire test statistics at the markers for which the sire is heterozygous. Subsequently, these estimates are combined across all sires, as if the single-sire test statistics were observed for all sires at each location. Significance of any position on the chromosome for across-sire analyses applying MMM is calculated analogously to across-sire SMM as the multisire

test statistic, $TS_i = \sum_{j=1}^n T_{ij} \approx \chi_{(n)}^2$, for each centimorgan along a chromosome. The degrees of freedom (df) at each centimorgan along the chromosome were calculated as the proportion of variance explained by the markers, namely, $df = (\mathbf{b}'\mathbf{V}\mathbf{b})/\text{var}(\mathbf{t})$. The ratio of these 2 variances is an estimate of the amount of information that we have to predict T_l (see Visscher, 1996). The df at each centimorgan are summed across sires as if the df were observed for all sires at each location as $df_l = \sum_{j=1}^n df_{ij}$.

Although df are not needed in the present study, where error rates are obtained by simulation under the null hypothesis, they will be useful when applying the MMM to actual data sets.

Although the MMM can be used to estimate single-sire tests statistics at the grid points between markers, the predicted MMM single-sire test statistics at these points are always less than the test statistic at the more significant of the 2 flanking markers. Let us consider the simplest case of 2 markers. At positions where single-sire SMM test statistics are available, single-sire MMM test statistics are identical to the SMM test statistics, because the solution vector of 2 neighboring markers is $\mathbf{b} = \{1, 0\}$ at the first marker and $\mathbf{b} = \{0, 1\}$ at the second marker. There is no influence of observed test statistics on each other in MMM. Consider now the grid points between 2 markers 2 cM apart; the solution vector for the position midway between them is $\mathbf{b} = \{0.5, 0.5\}$. Thus, when both markers have equally large SMM test statistics, the MMM test statistic for the marker between them is equal to the test statistics at the observed marker positions. If the 2 SMM marker test statistics are not equal, however, the predicted MMM test statistic in between is always lower than the higher of the 2 observed SMM test statistics. If markers are farther apart, the solution vector for the position midway and other grid points in between will not sum to 1, because there is not enough information coming from faraway markers (e.g., the solution vector for the midway position between 2 markers 10 cM apart is $\mathbf{b} = \{0.4901, 0.4901\}$). Thus, the grid points do not add any information to the single-marker analysis beyond that present at the marker points themselves. Extension of this argument to the multisire test statistics shows that there too, the test statistics of the grid points between markers will be less than that at the greater of the flanking marker positions. Hence, including these down-biased points in procedures for setting significance thresholds based on permutation tests, FDR (false discovery rate), or PFP (proportion of false positive) methods will increase multisire test-statistic significance thresholds, resulting in reduced power. Consequently, testing for the presence of a QTL and its location MMM will be applied only to the marker positions, and not to the intervening grid points. Nonetheless, application of MMM to the grid points results in an aesthetically pleasing and informative chromosome QTL map, and its use for this purpose is recommended.

Simulation Study

To validate the method, a mapping population was simulated to resemble a typical QTL mapping experiment in dairy cattle, applying selective DNA pooling in

a daughter design. The experiment takes advantage of the large number of female offspring per full-service AI sire and the phenotypes of these daughters generally available as EBV for a large number of production and functional traits. Ten paternal half-sib families with 2,000 progeny each were simulated. With regard to marker density, a 100-cM chromosome with 2, 6, or 16 evenly spaced markers was evaluated. In the 2-marker setting, the markers were at 33 and 67 cM; at the 6- and 16-marker spacing, 2 of the markers were at 0 and 100 cM, so that there were 4 and 14 evenly spaced intermediate markers along the chromosome, respectively. The 6-marker setting illustrates a planned marker coverage for a whole genome scan; the 2-marker setting illustrates a situation in which markers have to be omitted from analyses because of problems related to the pooling procedure; and the 16-marker setting illustrates a situation in which additional markers are analyzed on the most promising chromosomes for fine mapping.

With regard to marker information (defined as marker heterozygosity), 3 situations were evaluated, –100, 70, and 30% heterozygosity at each of the individual markers. A marker was simulated as heterozygous in a sire whenever the number drawn from a uniform distribution was lower than the set thresholds at either 0.3, 0.7, or 1 and as homozygous otherwise. These values represent a realistic range of heterozygosity when using microsatellite markers (Lipkin et al., 2004). Sire marker alleles at a given marker locus were simulated to be different from the dam alleles at that marker to circumvent having to make a correction for dam marker alleles. Crossovers were generated according to the Haldane mapping function (Haldane, 1919).

With regard to QTL alleles, 3 scenarios were simulated according to the standardized allele substitution effect (d), and according to frequency of the positive QTL allele in the dam population (f_D) and in the sire population (f_S). The 3 simulated scenarios were as follows: scenario 1: $d = 0.25$, $f_D = f_S = 0.5$; scenario 2: $d = 0.25$, $f_D = 0.2$, $f_S = 0.1$; scenario 3: $d = 0.125$, $f_D = f_S = 0.5$. Overall heritability of the trait was 0.25 in all 3 scenarios. The QTL was simulated at either 5, 50, or 55 cM. Table 1 gives an overview of the assessed scenarios.

Daughter phenotypes were affected by a QTL and by polygenic and environmental effects, all drawn from normal distributions with mean 0 and variances as given in Table 2, and simulated as

$$y_{ij} = \mu + g_{QTL_{ij}} + 1/2g_{sire_i} + 1/2g_{dam_j} + g_{M_{ij}} + \varepsilon_{ij},$$

where y_{ij} is the phenotypic value of daughter j of sire i , μ is the overall mean, $g_{QTL_{ij}}$ is the QTL effect based on the QTL alleles received from the sire and dam,

Table 1. Settings assessed by simulation

Markers ¹	Heterozygosity, ² %	QTL at cM ³	Scenario 1 ⁴	Scenario 2 ⁴	Scenario 3 ⁴
2	100	50	x		
	70	50	x		
	30	50	x	x	x
		5	x		
6	100	50	x	x	x
		55	x		
		5	x		
		50	x	x	x
	70	55	x		
		5	x		
		50	x	x	x
		55	x		
16	100	50	x		
		50	x	x	x
		50	x		
		50	x		

¹Number of markers on a chromosome.²Average marker heterozygosity.³Position of simulated single biallelic QTL in centimorgans.⁴Scenario 1: QTL allele substitution effect in phenotypic SD $d = 0.25$, frequency of positive QTL allele among dams f_D and frequency of positive QTL allele among sires $f_S = 0.5$. Scenario 2: $d = 0.25$, $f_D = 0.2$, $f_S = 0.1$. Scenario 3: $d = 0.125$, $f_D = f_S = 0.5$.

g_{sire_i} is the polygenic effect of sire i , g_{dam_j} is the polygenic effect of the dam of progeny j of sire i , $g_{M_{ij}}$ is the polygenic effect attributable to Mendelian sampling, and ε_{ij} is the environmental effect for progeny j of sire i .

Selective DNA pooling was performed with the 10% of daughters with the largest phenotype value and the 10% of daughters with the smallest phenotype value for each sire. Sire marker allele frequencies in the pools were obtained by counting the paternally inherited marker alleles, which were then scaled to 100%. Neither technical error variance nor different half-sib family size were considered, because the impact of both issues on selective DNA pooling was studied extensively by Baro et al. (2001). To set empirical chromosome-wide significance thresholds, for each simulation setting 1,000 replicates were simulated under the null hypothesis (QTL effect, $d = 0$). Each replicate consisted of 10 simulated chromosomes, 1 for each of 10 sires (a

total of 10,000 simulated chromosomes per simulation setting). Multisire SMM and MMM test statistics appropriate to each simulation setting were calculated for each of the markers on the chromosome. For each setting under the alternate hypothesis, we simulated 500 replicates. Each replicate involved 1 chromosome for each of 10 sires (a total of 5,000 simulated chromosomes for each setting). The multisire test statistic for each replicate was based on the single-sire SMM or MMM test statistics of these 10 sires. The markers with the greatest test statistic, under SMM or MMM, respectively, were chosen to represent the QTL, for purposes of testing significance against the 5 and 1% thresholds. The location at this marker was taken to represent the location of the QTL under SMM and MMM, respectively. Power, mean, and median of the absolute distance between simulated and estimated QTL locations were compared for SMM and MMM.

Table 2. Variances in the simulation study

Variance	H_0	Scenario 1 ¹	Scenario 2 ¹	Scenario 3 ¹
QTL	0	0.03125	0.015625	0.0078125
Polygenic paternal	0.0625	0.0546875	0.05859375	0.060546875
Polygenic maternal	0.0625	0.0546875	0.05859375	0.060546875
Mendelian sampling	0.125	0.109375	0.1171875	0.12109375
Error	0.75	0.75	0.75	0.75
Phenotypic	1.00	1.00	1.00	1.00

¹Scenario 1: QTL allele substitution effect in phenotypic SD $d = 0.25$, frequency of positive QTL allele among dams f_D and frequency of positive QTL allele among sires $f_S = 0.5$. Scenario 2: $d = 0.25$, $f_D = 0.2$, $f_S = 0.1$. Scenario 3: $d = 0.125$, $f_D = f_S = 0.5$.

Table 3. Empirical 1 and 5% significance thresholds for single-marker mapping (SMM) and multiple-marker mapping (MMM)

Markers ¹	Heterozygosity, ² %	5%		1%	
		SMM	MMM	SMM	MMM
2	100	20.13	20.13	24.78	24.78
	70	16.17	16.51	19.99	20.03
	30	10.24	10.57	12.96	13.99
6	100	22.69	22.69	27.19	27.19
	70	18.94	20.18	24.50	25.61
	30	13.01	15.14	17.77	19.69
16	100	24.17	24.17	29.49	29.49
	70	21.37	23.44	24.73	27.66
	30	14.69	20.18	18.56	25.17

¹Number of evenly spaced markers on a chromosome.²Average sire marker heterozygosity on a chromosome.

RESULTS

Significance Thresholds

Empirically derived 1 and 5% significance thresholds are presented in Table 3. As noted above, there was no influence of MMM test statistics on each other at heterozygous marker positions in a sire. Therefore, whenever all sires were heterozygous at a given marker (settings with 100% marker heterozygosity), MMM single- and multisire test statistics were identical to SMM test statistics, resulting in identical SMM and MMM significance thresholds. When some sires were not heterozygous at all markers, MMM thresholds, as expected, were always greater, because MMM multisire test statistics would generally be somewhat greater than SMM test statistics. The lower the proportion of heterozygosity, the greater the difference between SMM and MMM thresholds.

Power Calculations

Tables 4 and 5 show the results of power calculations for SMM and MMM, applying a 5 and 1% empirical significance threshold, respectively. The power of MMM was equal to or greater than that of SMM in all settings, although the power of SMM with a QTL of a large effect segregating with intermediate QTL allele frequencies in the population (scenario 1) was already very large. The superiority of MMM when a QTL of smaller effect was segregating, as in scenario 3, was more distinct (Table 4 and 5). The advantage of MMM over SMM increased with decreasing marker heterozygosity of sires independently of the number of markers on a chromosome or the QTL location. The largest advantage was observed in the setting with 6 markers, with an average heterozygosity of 30%. The average increase in power over all settings within scenario 3 was 5 and

9.4% when applying an empirical significance threshold of 5 and 1%, respectively. The increase in power for MMM in populations with extreme QTL allele frequencies, such as in scenario 2, was not so pronounced. This can be explained by the fact that many sires will be homozygous at the QTL, and the MMM test statistics at the marker level will be mainly made up of test statistics under the null hypothesis.

Deviance of QTL Location Estimates

Accuracy of the QTL location estimates applying a 1% empirical significance threshold for scenario 1 is presented in Table 6 and that for scenarios 2 and 3 is presented in Table 7. Single-marker mapping and MMM placed the QTL at the marker with the largest multisire test statistic, and not at any intermediate point (which is in contrast to standard interval mapping, which places the QTL between a marker bracket). Consequently, the distance between the closest marker and the QTL was a major factor determining the mean absolute deviance of QTL location estimates. Indeed, arithmetic means of absolute deviation between simulated and estimated QTL location increased with increasing distance between the simulated QTL and the closest marker position. In situations with less than 100% marker heterozygosity, arithmetic means were smaller for MMM than SMM, with the exception of the 6 marker, 30% heterozygosity setting; medians however, were identical. Means were often greater than medians, implying that in some instances the marker with the largest test statistic was not the closest marker. This happened more often for SMM than MMM and may have been due to the greater variance of multisire marker test statistics across a chromosome for SMM because of the large effect of number of sires heterozygous for a given marker on the multisire test statistic

Table 4. Power and relative power (in %) for single-marker mapping (SMM) and multiple-marker mapping (MMM) applying an empirical 5% significance threshold

Markers ¹	Heterozygosity, ² %	QTL at cM ³	Scenario 1 ⁴			Scenario 2 ⁴			Scenario 3 ⁴		
			MMM ⁵	SMM ⁶	MMM/SMM ⁷	MMM ⁵	SMM ⁶	MMM/SMM ⁷	MMM ⁵	SMM ⁶	MMM/SMM ⁷
2	100	50	99	99	100						
	70	50	98	98	100	73	73	100	63	61	102
	30	50	87	86	101						
6	100	5	100	100	100						
		50	100	100	100	81	81	100	79	79	100
		55	100	100	100						
	70	5	99	98	101						
		50	100	100	100	78	77	101	75	70	107
		55	99	99	100						
	30	5	90	89	101						
		50	94	93	101	64	63	101	48	43	112
		55	92	90	102						
16	100	50	100	100	100						
	70	50	100	99	100	84	83	101	79	76	104
	30	50	99	98	101						

¹Number of evenly spaced markers on a chromosome.²Average marker heterozygosity on a chromosome.³Position of simulated biallelic QTL in centimorgans.⁴Scenario 1: QTL allele substitution effect in phenotypic SD $d = 0.25$, frequency of positive QTL allele among dams f_D , and frequency of positive QTL allele among sires $f_S = 0.5$. Scenario 2: $d = 0.25$, $f_D = 0.2$, $f_S = 0.1$. Scenario 3: $d = 0.125$, $f_D = f_S = 0.5$.⁵Power of multiple-marker mapping (MMM).⁶Power of single-marker mapping (SMM).⁷Relative power of MMM/SMM.**Table 5.** Power and relative power (in %) for single marker mapping (SMM) and multiple-marker mapping (MMM) applying an empirical 1% significance threshold

Markers ¹	Heterozygosity, ² %	QTL at cM ³	Scenario 1 ⁴			Scenario 2 ⁴			Scenario 3 ⁴		
			MMM ⁵	SMM ⁶	MMM/SMM ⁷	MMM ⁵	SMM ⁶	MMM/SMM ⁷	MMM ⁵	SMM ⁶	MMM/SMM ⁷
2	100	50	98	98	100						
	70	50	97	97	100	64	62	103	47	43	109
	30	50	78	78	100						
6	100	5	99	99	100						
		50	99	99	100	70	70	100	62	62	100
		55	99	99	100						
	70	5	97	96	101						
		50	99	98	101	67	66	102	53	48	110
		55	98	96	102						
	30	5	83	78	106						
		50	90	87	103	50	47	106	27	22	123
		55	84	79	106						
16	100	50	100	100	100						
	70	50	99	99	100	78	78	100	65	62	105
	30	50	98	97	101						

¹Number of evenly spaced markers on a chromosome.²Average marker heterozygosity on a chromosome.³Position of simulated biallelic QTL in centimorgans.⁴Scenario 1: QTL allele substitution effect in phenotypic SD $d = 0.25$, frequency of positive QTL allele among dams f_D , and frequency of positive QTL allele among sires $f_S = 0.5$. Scenario 2: $d = 0.25$, $f_D = 0.2$, $f_S = 0.1$. Scenario 3: $d = 0.125$, $f_D = f_S = 0.5$.⁵Power of multiple-marker mapping (MMM).⁶Power of single-marker mapping (SMM).⁷Relative power of MMM/SMM.

Table 6. Deviation of QTL location estimate for scenario 1¹

Markers ²	Heterozygosity, ³ %	QTL at cM ⁴	MMM ⁵			SMM ⁶		
			Mean	SD	Median	Mean	SD	Median
2	100	50	17.0	0.0	17.0	17.0	0.0	17.0
	70	50	17.0	0.0	17.0	17.0	0.0	17.0
	30	50	17.0	0.0	17.0	17.0	0.0	17.0
6	100	5	5.3	1.7	5.0	5.3	1.7	5.0
		50	10.0	0.9	10.0	10.0	0.9	10.0
		55	5.3	2.0	5.0	5.3	2.0	5.0
	70	5	7.3	4.8	5.0	7.7	5.3	5.0
		50	10.5	3.2	10.0	11.0	4.5	10.0
		55	7.3	5.3	5.0	8.8	6.8	5.0
	30	5	10.6	8.8	5.0	10.4	8.6	5.0
		50	12.9	8.4	10.0	14.3	10.0	10.0
		55	10.6	8.6	5.0	12.2	10.5	5.0
16	100	50	3.4	2.5	3.0	3.4	2.5	3.0
	70	50	3.4	2.0	3.0	5.1	4.0	3.0
	30	50	5.2	4.3	3.0	8.7	7.8	3.0

¹Scenario 1: $d = 0.25$, $f_D = f_S = 0.5$; QTL allele substitution effect in phenotypic SD (d); frequency in the dam population (f_D) and frequency in the sire population (f_S).

²Number of evenly spaced markers on a chromosome.

³Average marker heterozygosity on a chromosome.

⁴Position of simulated QTL in centimorgans.

⁵Multiple-marker mapping.

⁶Single-marker mapping.

for this method. This was also reflected in a somewhat larger standard deviation of estimated QTL location for SMM. Within settings with the same number of markers, means and standard deviations increased with decreasing marker heterozygosity, reflecting the increasing importance under these settings of the num-

ber of sires heterozygous at a given marker in determining the multisire test statistic for that marker.

DISCUSSION

We proposed a simple but efficient method for performing MMM analyses for selective DNA pooling

Table 7. Deviation of QTL location estimate for scenario 2 and scenario 3

Scenario ¹	Markers ²	Heterozygosity, ³ %	MMM ⁴			SMM ⁵		
			Mean	SD	Median	Mean	SD	Median
2	2	70	17.0	0.0	17.0	17.0	0.0	17.0
		100	10.9	4.3	10.0	10.9	4.3	10.0
		70	11.5	5.8	10.0	12.0	6.4	10.0
	6	30	14.7	10.1	10.0	15.9	11.3	10.0
		70	5.2	5.2	3.0	6.3	5.9	3.0
	16	70	5.2	5.2	3.0	6.3	5.9	3.0
3	2	70	17.0	0.0	17.0	17.0	0.0	17.0
		100	11.9	6.5	10.0	11.9	6.5	10.0
		70	12.5	7.1	10.0	13.4	8.8	10.0
	6	30	14.1	9.4	10.0	13.3	8.4	10.0
		70	6.6	8.1	3.0	7.7	8.7	3.0
	16	70	6.6	8.1	3.0	7.7	8.7	3.0

¹Scenario 2: QTL allele substitution effect in phenotypic SD $d = 0.25$, frequency of positive QTL allele among dams $f_D = 0.2$ and frequency of positive QTL allele among sires $f_S = 0.1$. Scenario 3: $d = 0.125$, $f_D = f_S = 0.5$.

²Number of evenly spaced markers on a chromosome.

³Average marker heterozygosity on a chromosome.

⁴Multiple-marker mapping at single-marker mapping positions.

⁵Single-marker mapping.

within large half-sib families that also dealt with the problem of incomplete marker information. Although we derived the predicted test statistics for half-sib families, in principle the same method can be applied to other pedigree structures, for example, large full-sib families for mapping studies in swine or fish.

An implicit assumption of the MMM procedure is that there is a single QTL on the chromosome. However, in practice, $(1 - 2r)^2$ drops off rapidly with distance, so that if 2 QTL are well separated on the same chromosome, they should exert little effect on their respective observed or predicted test statistics.

In single-sire analyses, D_{ij} and Z_{ij} can be positive or negative, but T_{ij} is always positive. Thus, in contrast to other interval mapping procedures, the present procedure does not require prior knowledge of sire marker haplotypes. As a consequence, however, the present procedure should be less powerful than interval mapping having known haplotypes, because phase information is not used. If we consider the simplest possible situation, in which there is only a single marker, i with a significant test statistic T_{ij} for sire j , then the most likely location for the QTL is at T_i , and T_l can be estimated as $(1 - 2r_{il})^2 T_i$. If we extend this to a pair of significant markers, T_i and T_{i+1} , that differ in their marker-associated effect, then standard interval mapping methods will attempt to place the QTL with respect to the 2 markers in such a way that it will best explain the relative magnitude of effect observed at the 2 markers. The MMM approach, however, will simply take a combination of the 2 test statistics as predictors of the estimated test statistic at any given point. There is an interesting, nontrivial consequence of the present procedure. Consider a pair of markers flanking an interval and having equal test statistics. The interval mapping procedure will place the QTL between them and assign it a test statistic greater than either of the 2 flanking markers. Thus, the graph of test statistics between the markers will be convex. This is the usual convex pattern between adjacent markers shown by interval mapping of a chromosome, which locates the QTL in the interval between the markers. Consequently, whenever there is a homozygous marker equidistant between 2 heterozygous markers, it will be given the highest test statistic even though there is no actual information for this location on the chromosome. The MMM procedure will estimate the test statistic from the observed test statistic at the 2 flanking markers, and will assign a minimum test statistic to the midpoint between the markers, because at this point the predicted test statistic has the lowest correlation with the observed test statistics at the flanking markers. This has the conservative effect of assigning a smaller T_l to the missing datum than to the 2 flanking

markers. Thus, the graph of predicted test statistics between the markers will be concave, as shown in Figure 1.

Furthermore, for MMM the best estimate of QTL map location in a given chromosomal region will always be at the marker with the highest significant test statistic, and the estimated QTL effect will be the effect at this marker. This stands in contrast to standard interval mapping, which usually yields the best estimate for the QTL effect and location at some point intermediate between the 2 markers.

As mentioned above, a somewhat different subset of tested sires will be included in the set of heterozygous sires at each marker along a chromosome. Thus, under SMM, the multisire test statistic, TS_i , will be more or less strongly affected by the specific sires that are heterozygous for each marker. This can lead to mistaken inferences regarding the number and location of QTL on the chromosome. For example, if one of the sires shows a strong effect at markers i and $(i + 2)$, but is homozygous at an internal marker $(i + 1)$, then TS_i and $TS_{(i+2)}$ will be distinctly greater than $TS_{(i+1)}$, so that markers i and $(i + 2)$ show significant effects, whereas marker $(i + 1)$ will not be significant. This can lead to an inference of 2 QTL, at markers i and $(i + 2)$, respectively, separated by a region in which QTL are absent. If all sires were heterozygous at all markers, marker $(i + 1)$ would have been significant as well, and only a single QTL would be inferred. In contrast to SMM, MMM avoids this and therefore allows more accurate inferences about the number of QTL on a chromosome from the pattern of test statistics along the chromosome. As a result, the chromosomal graph of marker test statistics produced by the 1-cM grid search is much smoother and amenable to better interpretation.

The second problematic situation in SMM is depicted in Figure 1. The simulated QTL in this figure is at 50 cM, and flanking markers are at 33 and 67 cM. When the same number of sires is heterozygous at the 2 markers and at the QTL, multisire test statistics at both flanking markers are expected to be significant and yield test statistics of similar magnitude. As a result of chance sampling, however, the number of sires heterozygous for marker and QTL at the 33-cM marker might be larger than the corresponding number of 67 cM. Then, as shown in Figure 1 for SMM, the test statistic at 33 cM will be significant, but the test statistic at 67 cM does not reach the significance threshold. Consequently, if one decides about the QTL region to investigate further for fine mapping based on the SMM results, there is a risk of choosing too narrow a region on the chromosome, which might or might not include the QTL. In contrast, MMM will still approximate the con-

Figure 1. Example graph of a single replicate displaying across-sire single-marker mapping (SMM) and multiple-marker mapping (MMM) test statistics and empirically derived significance thresholds for both mapping methods. Single biallelic QTL at 50 cM and 2 markers at 33 and 67 cM were simulated. The simulated QTL allele substitution effect was 0.125 phenotypic standard deviations, and the QTL allele frequency in the whole population was 0.5.

dition of equal test statistics at the 2 markers, overcoming the problem of differing marker heterozygosity of the sires, and correctly identifying the interval containing the QTL.

Multiple-marker mapping does not solve the problem of separate estimation of the QTL effect and QTL position as interval mapping would. Nevertheless, it deals better with the problem of differing marker heterozygosity of sires than does SMM, so that for the same data set, the power and accuracy of QTL detection are increased without requiring additional genotyping work to infer sire-marker haplotypes. On the basis of the results of the simulations, we recommend the use of MMM for the analyses of selective DNA pooling when haplotype information is not available.

ACKNOWLEDGMENT

This work was supported by the FP5 program of the European Union (QLK5-CT-2001-02379). We thank 2 anonymous reviewers for their very valuable comments, which helped to improve this manuscript significantly.

REFERENCES

- Baro, J. A., C. Carleos, N. Corral, T. Lopez, and J. Canon. 2001. Power analysis of QTL detection in half-sib families using selective DNA pooling. *Genet. Sel. Evol.* 33:231–247.
- Cobanoglu, O., P. J. Berger, and B. W. Kirkpatrick. 2005. Genome screen for twinning rate QTL in four North American Holstein families. *Anim. Genet.* 36:303–308.
- Darvasi, A., and M. Soller. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* 138:1365–1373.
- Dekkers, J. C. M. 2000. Quantitative trait locus mapping based on selective DNA pooling. *J. Anim. Breed. Genet.* 117:1–16.
- Fisher, P. J., and R. J. Spelman. 2004. Verification of selective DNA pooling methodology through identification and estimation of the DGAT1 effect. *Anim. Genet.* 35:201–205.
- Gonda, M. G., J. A. Aria, G. E. Shook, and B. W. Kirkpatrick. 2004. Identification of an ovulation rate QTL in cattle on BTA14 using selective DNA pooling and interval mapping. *Anim. Genet.* 35:298–304.
- Haldane J. B. S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.* 8:299–309.
- Lipkin, E., M. O. Mosig, A. Darvasi, E. Ezra, A. Shalom, A. Friedmann, and M. Soller. 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: Analysis of milk protein percentage. *Genetics* 149:1557–1567.
- Lipkin, E., A. Nave, R. Tal, M. Soller, and A. Friedmann. 2004. Effect of incomplete informativity of markers on high resolution QTL mapping in half-sib family designs. *Plant Anim. Genomes XII Conf.*, San Diego, CA. http://www.intl-pag.org/12/abstracts/P3f_PAG12_316.html
- Mariasegaram, M., N. Robinson, T. Balasingham, A. Chamberlain, J. Lade, H. McPartlan, and M. Goddard. 2002. Mapping genes affecting milk composition in Australian dairy cattle using selective DNA pooling. No. 09-09 in *Proc. 7th World Congr. Genet. Appl. Livest. Prod.*, Montpellier, France. INRA, Castanet-Tolosan, France.
- Mosig, M. O., E. Lipkin, G. Khutoreskaya, E. Tchourzyna, M. Soller, and A. Friedmann. 2001. A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* 157:1683–1698.
- Sharma, B. S., G. B. Jansen, N. A. Karrow, D. Kelton, and Z. Jiang. 2006. Detection and characterization of amplified fragment length polymorphism markers for clinical mastitis in Canadian Holsteins. *J. Dairy Sci.* 89:3653–3663.

- Soller, M., and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* 34:47–55.
- Visscher, P. M. 1996. Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. *J. Hered.* 87:136–138.
- Wang, J., K. J. Koehler, and C. M. Dekkers. 2007. Interval mapping of quantitative trait loci with selective DNA pooling data. *Genet. Sel. Evol.* 39:685–709.
- Weller, J. I., Y. Kashi, and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* 73:2525–2537.
- Xiao, Q., T. A. Wibowo, X. L. Wu, J. J. Michal, J. J. Reeves, J. R. Busboom, G. H. Thorgaard, and Z. Jiang. 2007. A simplified QTL mapping approach for screening and mapping of novel AFLP markers associated with beef marbling. *J. Biotechnol.* 127:177–187.