Mapping Quantitative Trait Loci with Various Types of Progeny from Complex Pedigrees**

C. Lee* and X. L. Wu[1]
Lab. of Statistical Genetics, Institute of Environment and Life Science, Hallym University
Chuncheon, Kangwon-do 200-702, Korea

ABSTRACT : A method for mapping quantitative trait loci (QTL) was introduced incorporating the information of mixed progeny from complex pedigrees. The method consisted of two steps based on single marker analysis. The first step was to examine the marker-trait association with a mixed model considering common environmental effect and reversed QTL-marker linkage phase. The second step was to estimate QTL effects by a weighted least square analysis. A simulation study indicated that the method incorporating mixed progeny from multiple generations improved the accuracy of QTL detection. The influence of within-genotype variance and recombination rate on QTL analysis was further examined. Detecting a QTL with a large within-genotype variance was more difficult than with a small within-genotype variance. Most of the significant marker-QTL association was detectable when the recombination rate was less than 15%. *(Asian-Aust. J. Anim. Sci. 2001. Vol 14, No. 11 : 1505-1510)*

Key Words : Complex Pedigree, Mixed Models, QTL, Recombination Rate

INTRODUCTION

Inbred lines (Jeffrey et al., 1987; Duddley, 1993; Muehlbauer et al., 1998) and outbred populations (Haley et al., 1994; Zhang et al., 1998; de Knong et al., 2001) were widely used for QTL mapping in animal populations. In either case, developing methods for extracting genetic information from combined lines or complex pedigree was a concern to animal geneticists in recent years (Meuwissen and Goddard, 1997; George et al., 2000). This was because they often collected data from practical animal populations where there were various mating types and multiple generations. Furthermore, the power of QTL detection was substantially improved when more generations were included (Rebai and Goffinet, 1993; Darvasi and Soller, 1995). The objective of this study was to develop a method for QTL analysis with complex pedigrees based on single marker analysis.

MATERIALS AND METHODS

Single marker analysis of QTL with complex pedigree

The following mixed model was used to examine marker-trait association and to estimate marker means and variance.

$$y = X\beta + Z\theta + \varepsilon \quad (1)$$

where y was the vector of observations, β was the vector of unknown fixed sex and full-sib litter effects, θ was the vector of unknown random marker genotype effects grouped by full-sib litters, $\mathbf{X}$ and $\boldsymbol{Z}$ were the design matrices relating the elements in y to those in β and θ, and ε was the vector of residuals. Grouping marker means by full-sib litters took account of the common environmental effect and QTL-marker phases that possibly differed with litters. The θ was normally distributed with zero means and identical variances, and so was ε. Variance components and the solution of unknown vectors were obtained by SAS MIXED programs (SAS Institute Inc., 1990). An F ratio test and Wald Z test were performed to examine if the marker genotype variance was significant.

And the additive and dominance effects were then estimated by a weighted least square (WLS) analysis using the following regression model.

$$\hat{\theta} = P_1 a + P_2 d + r_g$$
$$= Pg + r_g \quad (2)$$

where $\hat{\theta}$ was the vector of the estimated marker genotype means grouped by full-sib litters, a and d were the additive and dominance marker effects associated with QTL, r_g was the residual genetic effect that were not explained by the QTL, P_1 was the matrix that related marker genotypes in θ to the additive estimate, taking 1, 0, and –1 for the marker genotypes AA, Aa, and aa, P_2 was the matrix that related marker genotypes in θ to the dominance estimate, taking -0.5, 1, and -0.5 for the marker

** This study was supported in parts by the Hallym Academy of Sciences, Hallym University, Korea, and Hunan Life Science Research Center, China.

* Address reprint request to C. Lee. Tel: +82-33-240-1794, Fax: +82-33-241-3422, E-mail: clee@sun.hallym.ac.kr
[1] Permanent Address: Lab. of Molecular Biology Applied to Animal Production, Hunan Institute of Animal and Veterinary Science, Changsha, Hunan 410131, P. R. China.

Received May 8, 2001; Accepted August 6, 2001

genotypes AA, Aa, and aa, $P=(P_1 \quad P_2)$, and $g=(a \quad d)'$.

Applying WLS to Equation (2) produced

$$\hat{g}=(P'WP)^{-}P'W\hat{\theta} \qquad (3)$$

where W was a diagonal matrix with diagonal element equaled to the number of corresponding observations for means in θ.

When both the QTL and marker locus had only two alleles, estimating QTL effect by the single marker analysis was straightforward. However, it was tricky when a marker locus with more than two alleles was involved. Population II reflected such a situation where a tri-allele marker locus was linked to a QTL of two alleles. When a multi-allele marker locus was involved, a putative multiple allele QTL model was used to determine the possible marker-QTL linkage phase before performing the final QTL analysis. The model assumed that each marker allele was linked to a QTL allele with different additive effect. Under the assumption, the marker genotype means were then partitioned into additive and dominance effects based on the following regression model under the restriction of $\sum a_i=0$ in a WLS analysis.

$$\mu_{ij}=0.5a_i+0.5a_j+d_{ij}+r_{ij} \qquad (4)$$

where μ_{ij} was the mean for the marker genotype with alleles i and j, a_i and a_j were the additive effect associated with the marker alleles i and j, respectively, d_{ij} was the dominance effect associated with these two marker alleles, and r_{ij} was the residual effect that was not explained by the QTL.

If all homozygous marker genotypes were observed, it was much easy and convenient to infer marker-QTL linkage phase by defining the marker allele substitution effect as below,

$$\hat{a}_i-\hat{a}_j=0.5(\hat{\mu}_{ii}-\hat{\mu}_{jj}) \qquad (5)$$

where $\hat{a}_i$ and $\hat{a}_j$ were the additive effect associated with the marker alleles i and j, respectively, $\hat{\mu}_{ii}$ and $\hat{\mu}_{jj}$ were the corresponding homozygous marker genotype means.

Simulation

The pedigree : A complex pedigree was simulated. The first 3 generations featured a typical backcross design for QTL mapping, and 4 more generations were generated from advanced backcross and sib-mating. A total of 220 animals were simulated for the pedigree.

Genotypes of markers and QTL and trait values : The marker genotypes and trait values were simulated using the SOLAR (Sequential Oligogenic Linkage Analysis Routines) software package for Linux system (Almasy and Blangero, 1998).

Three populations were produced in this study. In Population I, bi-allele marker and QTL were used. The two founder pigs from purebred stocks were homozygous at the marker locus and the QTL and had different genotypes, say $AAQQ$ (♂) and $aaqq$ (♀). The QTL allele (Q) with enhancing effect was linked to marker allele A and was carried by the male founder in the starting generation. In Population II, the marker had three alleles but QTL had only two alleles. Two founders shared one common marker allele, say C_1C_3Qq (♂) and C_1C_2qq (♀). The QTL allele (Q) with enhancing effect was linked to the third marker allele (C_3) and carried by the male founder. Population III was produced to examine the influence of within-genotype variation on QTL detection. Within-genotype variation in the Population III doubled the amount as in the Populations I and II. The number of alleles for the marker and QTL in the Population III was the same as that in the Population I, but the marker locus was fixed at the last generation. The founders were also homozygous at the marker locus and the QTL, but the genotypes by sex were opposite to the first population, say $bbqq$ (♂) and $BBQQ$ (♀). Therefore, the allele Q of QTL was carried by the female founder.

The input values of genetic parameters in the simulation were: $a=2$ for QTL additive effect, $d=1$ for QTL dominance effect, $\sigma_{wg}=0.5\sim1.0$ for within genotype standard deviation, $h_r^2=0.1\sim0.3$ for residual heritability (the fraction of trait variance after the QTL effect has been accounted for), $r=0.1$ for the recombination rate between the QTL and the marker. The overall mean was also applied, and the means of QTL genotypes were $\mu_{QQ}=72\sim74$, $\mu_{Qq}=71\sim73$ and $\mu_{qq}=68\sim70$. The simulated marker genotypes of animals in the complex pedigree were listed in table 1. The observed trait mean, marker heterozygosity, and inferred QTL allele (Q) frequency by generations were illustrated in figures 1, 2 and 3, respectively.

Additionally, the influence of recombination rate on QTL analysis was examined. The input values of the recombination rate ranged from 0 (complete linkage) to 50% (no linkage). The bi-allele marker locus model was used with the QTL allele, Q, consistently linked to allele A in the starting generation. The founder genotypes took any of the three possible combinations ($AAQQ\times aaqq$, $AaQq\times AaQq$ and $aaqq\times aaqq$), and it was decided by chance in the simulation.

Table 1. The simulated marker genotypes for animals in the complex pedigree

ID	Sex	Generation	Father ID	Mother ID	Marker genotypes		
					Locus I	Locus II	Locus III
001	m	0	0	0	AA	bb	C_1C_3
002	f	0	0	0	aa	BB	C_1C_2
101	f	1	001	002	Aa	Bb	C_2C_3
102	f	1	001	002	Aa	Bb	C_1C_1
103	f	1	001	002	Aa	Bb	C_1C_1
104	f	1	001	002	Aa	Bb	C_1C_2
116	m	2	001	101	Aa	Bb	C_2C_3
117	f	2	001	101	AA	bb	C_1C_2
122	m	2	001	102	AA	bb	C_1C_3
126	f	2	001	102	Aa	Bb	C_1C_1
156	m	2	001	104	AA	bb	C_1C_2
158	f	2	001	104	Aa	bb	C_2C_3
212	m	3	156	158	AA	bb	C_2C_3
309	f	3	122	103	AA	bb	C_1C_3
431	m	4	212	158	Aa	bb	C_2C_2
452	f	4	212	309	AA	bb	C_2C_3
464	f	4	212	103	Aa	bb	C_1C_3
466	f	4	212	103	AA	bb	C_1C_3
662	m	5	431	452	AA	bb	C_2C_3
698	f	5	431	452	AA	bb	C_2C_3

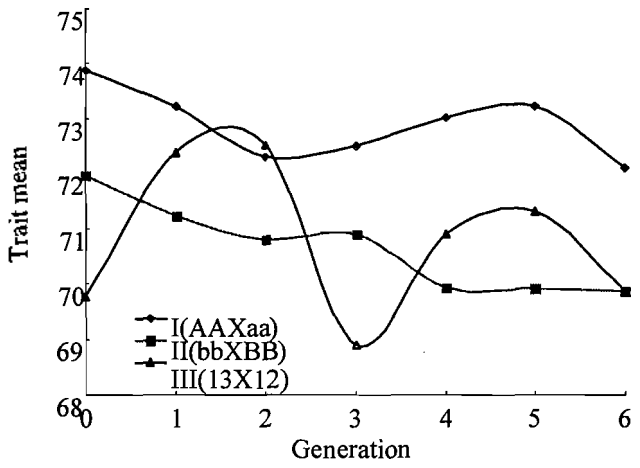

Figure 1. Trait mean by generations for Populations I, II and III

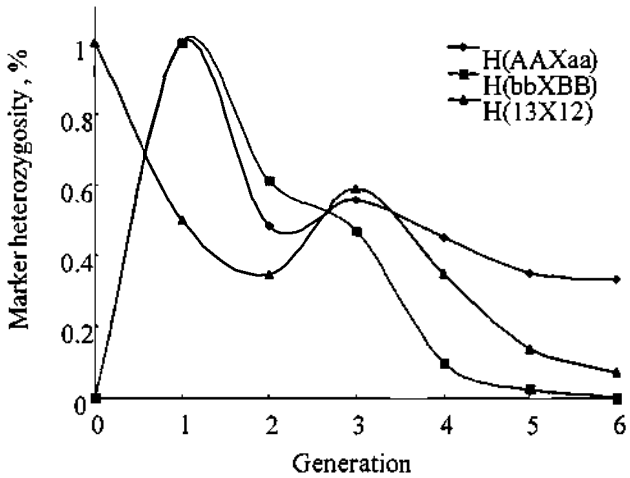

Figure 2. Marker heterozygosity by generations for Populations I, II and III

RESULTS

Marker-trait association and QTL variance

Marker-QTL association was significantly discovered in Populations I and II (table 2). Wald F ratios ranged from 7.81 to 13.60 in population I and were all highly significant ($p<0.001$). The F ratios in Population II were somewhat smaller, ranging from 7.63 to 9.07, but also significant ($p<0.01$). The F ratios showed a trend of decreased P-value as more generations were incorporated.

Wald Z test supported the improvement in QTL detection by incorporating more generations. In population I, the P value gradually decreased from 0.1480 with only the backcross progeny to 0.0058 with mixed progeny from all generations. In Population II, the P-value was also decreased from 0.1166 to less than 0.0001.

The estimates of marker heritability in populations I, II and III were 0.6614, 0.2721 and 0.4935, respectively. When the recombination rate was small, the estimate of marker heritability could be considered as an approximate estimate of the QTL heritability.

Table 2. Significance test of the marker-QTL association for Populations I and II

Generation	Wald F test		Wald Z test			
	F ratio	Pr>F	Var (MK)	StdDev	Z	Pr>|Z|
			----- Population I -----			
Backcross (BC)	10.52***	0.0001	1.3626	0.9393	1.45	0.1480
Mixed (BC+F3)	7.81***	0.0001	1.7856*	0.8840	2.02	0.0436
Mixed II (BC+F3+F4)	8.40***	0.0001	1.6244*	0.7125	2.28	0.0227
Mixed III (BC+F3+F4+F5)	10.88***	0.0001	1.6656**	0.6285	2.65	0.0079
Mixed IV(BC+F3+F4+F5+F6)	13.60***	0.0001	1.6916**	0.6129	2.79	0.0058
			----- Population II -----			
Backcross (BC)	7.63***	0.0005	1.4665	0.9344	1.57	0.1166
Mixed (BC+F3)	8.68***	0.0001	1.4616**	0.5612	2.60	0.0092
Mixed II (BC+F3+F4)	8.72***	0.0001	1.4258***	0.4790	6.95	0.0001
Mixed III (BC+F3+F4+F5)	9.76***	0.0001	1.4282***	0.4421	8.07	0.0001
Mixed IV(BC+F3+F4+F5+F6)	9.07***	0.0001	1.4014***	0.4216	9.01	0.0001

* p<0.05, ** p<0.01, *** p<0.001.

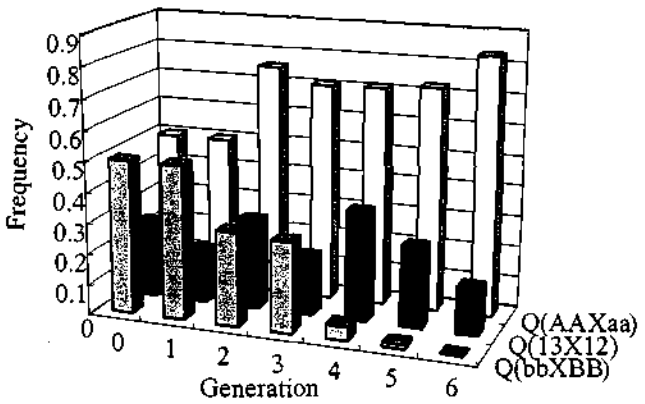

Figure 3. Frequency of dominant QTL allele (Q) by generations for Population I, II and III

Table 3. Estimate of allele substitution effect using a putative multi-allelic QTL model[1,2]

	Estimate	Std Err	T for H0	Pr>|T|	
				Raw	StepBon
a_1-a_2	0.1492	1.4132	1.06	0.2924	0.2924
a_1-a_3	-0.8155***	1.4852	-5.49	0.0001	0.0003
a_2-a_3	-0.9646***	1.6714	-5.77	0.0001	0.0003

*** p<0.001;

[1] Allele substitution effect was defined as half of the difference between the means of the respective homozygous genotypes.

[2] Raw = raw p-value; StepBon = adjusted p-values using the step-down Bonferroni method of Holm (1979).

Estimation of QTL Effect

The estimates of allele substitution effects were obtained in table 3. The estimate of allele substitution effect was significant between C_1 and C_3 and between C_2 and C_3 (p<0.01), but not between C_1 and C_2 (p>0.05). This indicated that allele Q at the QTL was closely linked to marker allele 3 comparing to the other marker alleles in this complex pedigree. On the other hand, both marker alleles 1 and 2 were linked to allele q of the QTL with larger frequency. From the results, marker allele C_3 was re-coded as "*C*" representing a dominant allele. The other two alleles C_1 and C_2 were accordingly re-coded as "*c*". Such re-coding made it easy to estimate QTL effects by following the principles for single marker analysis with a bi-allele marker locus.

The estimates of additive and dominance effects were all underestimated although they are not significant (p>0.05, table 4). The standard errors of the additive and dominance effects were gradually decreased by including more progeny. Furthermore, the estimate of dominance effect using BC+F3 or using BC+F3+F4 was not significant (p>0.05) in Population I as shown in table 4, but the estimate was significant using BC+F3+F4+F5 (p<0.05) or using BC+F3+F4+F5+F6 (p<0.01).

Influence of within-genotype variance and recombination Rate

The input value of within-genotype variance in Population 3 was two times larger than those in Populations I and II. In reality, a large within-genotype variance was likely to be observed when the expression of a QTL was affected by some other genes or modifiers. With this large within-genotype variance, including mixed progeny from multiple generations improved QTL detection (figure 4). The P-value for marker variance was larger than 0.05 using BC, BC+F3, or BC+F3+F4 while the marker variance was significant when using BC+F3+F4+F5 or BC+F3+F4+F5+F6. Detecting a QTL with a large within-genotype variance was more difficult than with a small within-genotype variance. Furthermore, the underestimation in the estimates of QTL was more serious, especially for dominance effects (table 5).

Table 4. Estimates of QTL additive and dominance effects for Populations I and II [1,2]

Generation	Additive		Wald Z test			
	$a\pm\sigma_a$	T	Pr>|T|	$d\pm\sigma_d$	T	Pr>|T|
			----- Population I -----			
Backcross (BC)						
Mixed (BC+F3)	1.7805***±0.3285	5.42	0.0001	0.9985⁺±0.5890	1.70	0.0933
Mixed II (BC+F3+F4)	1.7475***±0.2977	5.87	0.0001	0.9993⁺±0.5578	1.78	0.0722
Mixed III (BC+F3+F4+F5)	1.7491***±0.2674	6.54	0.0001	0.7803*±0.3320	2.35	0.0200
Mixed IV(BC+F3+F4+F5+F6)	1.7467***±0.2481	7.04	0.0001	0.8456**±0.3216	2.63	0.0093
			----- Population II -----			
Backcross (BC)						
Mixed (BC+F3)	1.7633**±0.6422	2.75	0.0077	0.7640*±0.3228	2.31	0.0239
Mixed II (BC+F3+F4)	1.6552***±0.3391	4.88	0.0001	0.7762***±0.1920	4.04	0.0001
Mixed III (BC+F3+F4+F5)	1.6412***±0.3370	4.87	0.0001	0.7480***±0.1862	4.02	0.0001
Mixed IV(BC+F3+F4+F5+F6)	1.6805***±0.2787	6.03	0.0001	0.6872**±0.1613	4.26	0.0001

⁺p<0.10, * p<0.05, **p<0.01, *** p<0.001.

[1] The input values of the additive and dominance effects for the QTL were 2.0 and 1.0, respectively, and the input value of within-genotype variance was 0.5.

[2] Separate estimates for the additive and dominance effects of QTL were not available with the backcross progeny because of lack of one homozygous genotype.

Table 5. Estimates of QTL additive and dominance effects for Population III[1,2]

Generation	Additive		Wald Z test			
	$a\pm\sigma_a$	T	Pr>|T|	$d\pm\sigma_d$	T	Pr>|T|
Backcross (BC)						
Mixed (BC+F3)	1.5686***±0.3991	3.93	0.0002	0.3913±0.5287	0.74	0.3936
Mixed II (BC+F3+F4)	1.5360***±0.3641	4.22	0.0001	0.3942±0.4749	0.83	0.4098
Mixed III (BC+F3+F4+F5)	1.4163***±0.3248	4.36	0.0001	0.2713±0.5718	0.38	0.7023
Mixed IV(BC+F3+F4+F5+F6)	1.4150***±0.3056	4.63	0.0001	0.2110±0.5410	0.39	0.2880

*** p < 0.001.

[1] The input values of the additive and dominance effects for the QTL were 2.0 and 1.0, and the input value of within-genotype variance was 1.0.

[2] Separate estimates for the additive and dominance effects of QTL were not available using only backcross progeny.

The results from the analyses with data simulated with various recombination rates were shown in figure 5. Most of the significant marker-QTL association between the marker and the QTL was detectable when the recombination rate was less than 15%. However, significant QTL-marker association was never observed with any recombination rate when the heritability was 0.05. Other failure to detect the QTL was observed when $h^2 = 0.1$ and $r = 0.1$. Such failures were so called "false negative" results in the QTL analysis when the heritability level was low. On the other hand, significant marker-trait association was observed even with a large recombination rate ($r = 0.4$) when the heritability was large ($h^2 = 0.5$). This was a "false positive" result when QTL heritability was large.

DISCUSSION

This paper presented an effort to make use of complex families for QTL mapping. The simulation study demonstrated that the method for incorporating information from various types of progeny and from multiple generations improved the accuracy in QTL detection. However, the QTL effects estimated by the single marker analysis were underestimated. The underestimation was intrinsic property associated with the linkage position (Liu, 1998). For example, the expected values of marker additive and dominance estimates in the F2 design were $(1-2r)a$ and $(1-2r)^2 d$. Therefore, methods that are capable of dealing with multiple markers are preferred (Zeng, 1993; Jansen, 1993; Kearsey and Hyne, 1994; Wu and Li, 1994; Kao et al., 1999). Theoretically, introducing these methods will offer a more powerful QTL detection and more precise estimates of QTL effects and position. Nevertheless, single marker analysis was still powerful in search for QTL, and it was easy and convenient in both theoretical and practical implementation.

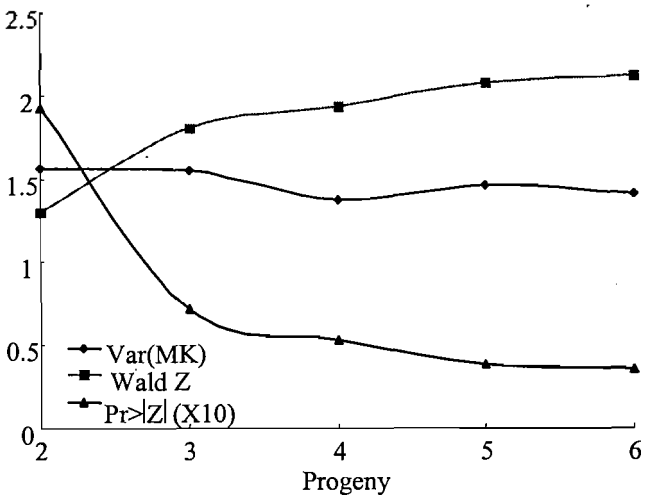

Figure 4. Marker variance estimate, Wald Z statistic, and the corresponding P-value in Population III. The X-axis numbers represented various progeny groups (2: the backcross progeny, 3: the mixed progeny from BC+F3, 4: the progeny from BC+F3+F4, 5: the progeny from BC+F3+F4+F5, and 6: the progeny from BC+F3+F4+F5+ F6). The P-value was multiplied by 10.

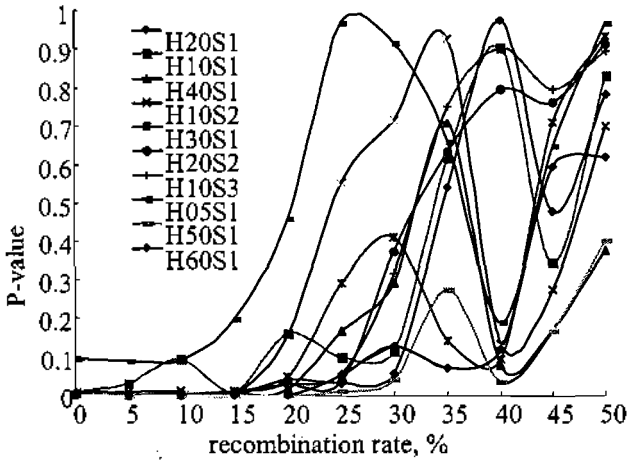

Figure 5. The P-value for marker genotype variance using the data simulated with various marker-QTL recombination rates and QTL heritabilities. The input values of the recombination rates ranged from 0 to 50% with an increment of 5%. The QTL heritability ranged from 0.05 to 0.60. HiSj indicated the replicate j with QTL heritability i.

REFERENCES

Almasy, L. and J. Blangero. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 62:1198-1211.

Darvasi, A. and M. Soller. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics. 141:1199-1207.

de Koning, D. J., N. F. Schulmant, K. Elo, S. Moisio, R. Kinos, J. Vilkki and A. Maki-Tanila. 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. J. Anim. Sci. 79:616-622.

Dudley, J. W. 1993. Molecular marker in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33:660-668.

George, A. W., P. M. Visscher and C. S. Haley. 2000. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081- 2092.

Haley, C. S., S. A. Knott and J. M. Elsen. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195-1207.

Holm, S. 1979. A simple sequentially rejective Bonferroni test procedure. Scand. J. Stat. 6:65-70.

Jasen, R. C. 1993. Interval mapping of multiple quantitative trait loci. Genetics. 135:205- 211.

Jeffreys, A. J., V. Wilson, R. Kelly, B. A. Taylor and G. Bulfield. 1987. Mouse DNA 'fingerprints': analysis of chromosome localization germ-line stability of hyprvariable loci in recombinant inbred strains. Nucleic Acid Res. 15:2823-2836.

Kao, C. H., Z. B. Zeng and R. D. Teasdale. 1999. Multiple interval mapping for quantitative trait loci. Genetics. 152:1203-1216.

Liu, B. H. 1998. Statistical Genomics: Linkage, Mapping, and QTL Analysis. CRC Press, Boca Raton, NY, USA.

Meuwissen, T. H. and M. E. Goddard. 1997. Estimation of effects of quantitative trait loci in large complex pedigrees. Genetics. 146:409-416.

Muehlbauer, G. J., J. E. Sepecht, M. A. Thomas-Compton, P. E. Staswick and R. L. Bernard. 1998. Near-isogenic lines --- a potential resource in the integration of conventional and molecular and molecular marker linkage maps. Crop Sci. 28:729-735.

Rebai, A. and B. Goffinet. 1993. Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor. Appl. Genet. 86:1014-1022.

SAS Institute Inc. 1990. SAS/STAT User's Guide, Version 6 (4th Ed.). Cary, NC, USA.

Wu, W. R. and W. M. Li. 1994. A new approach for mapping quantitative trait loci using complete genetic marker linkage maps. Theor. Appl. Genet. 89:535-539.

Zeng, Z. B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping of quantitative trait loci. Proc. Natl. Acad. Sci. USA 90:10972-10976.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller and M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics. 149:1959-1973.