COMPARISON OF MAXIMUM LIKELIHOOD ESTIMATION APPROACH AND REGRESSION APPROACH IN DETECTING QUANTITATIVE TRAIT LOCI USING RAPD MARKERS

Changren Weng[1], Thomas L. Kubisiak[2], C. Dana Nelson[3], James P. Geaghan[4], and Michael Stine[1]

Abstract:--Single marker regression and single marker maximum likelihood estimation were used to detect quantitative trait loci (QTLs) controlling the early height growth of longleaf pine and slash pine using a ((longleaf pine x slash pine) x slash pine) BC, population consisting of 83 progeny. Maximum likelihood estimation was found to be more power than regression and could also estimate the distance between markers and QTLs. Test statistic for the relationship between simple regression and maximum likelihood estimation is introduced. A total of four major QTLs linked to random amplified polymorphic DNA (RAPD) markers were detected explaining 19.7%, 10.7%, 12.8%, and 9.9% total variance of total height. Multiple regression analysis indicated that these four QTLs explained about 43.2% of the total variance of early height growth.

Keywords:--QTLs, random amplified polymorphic DNA, *Pinus palustris*, *Pinus elliottii*, maximum likelihood estimation, regression.

INTRODUCTION

RAPD (random amplified polymorphic DNA) markers are fragments of DNA amplified from genomes of organisms (Williams *et* al.1990, Welsch and McClelland 1990). The RAPD technique uses decamer nucleotides as primers to amplify regions of template DNA. Nucleotide mismatches at the priming sites such as those caused by insertions and deletions of one or more base pairs, or insertions, deletions, and translocations in the amplified regions, may lead to polymorphisms. RAPDs are dominant markers and are attractive because the procedure is simple, fast, and uses trace amounts of template DNA. The association between RAPD markers and QTLs can be detected using various methods. Markers that are tightly linked to QTLs may then be used for marker-assisted selection to guide breeding efforts.

Approaches developed for detecting QTLs using molecular markers can be classified into marker interval approaches and single marker approaches. For marker interval and single marker approaches, statistical methods, such as regression, maximum likelihood estimation, and moment can be applied. The single marker approach uses a single marker as the independent variable. This approach does not require any linkage information of markers. The marker interval approach uses an interval between two linked markers as the independent variable. This approach needs linkage information of markers for the analysis and cannot use unlinked markers. Regression and maximum likelihood estimation have been

[1] Louisiana Agricultural Experiment Station,Louisiana State University Agricultural Center, School of Forestry, Wildlife, and Fisheries; Baton Rouge, LA 70803, USA

[2] USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics, 23332 HWY 67, Saucier, MS 39574, USA

[3] International Paper Inc., 719 Southern Road, Bainbridge, GA 31717, USA

[4] Louisiana Agricultural Experiment Station, Louisiana State University Agricultural Center, Department of Experimental Statistics, Baton Rouge, LA 70803, USA

the two statistical methods most frequently used in detecting QTLs. Single marker simple regression is the most straightforward method for detecting the QTLs. This method compares the two or three phenotypic means of the trait for each marker genotype. The amount of variance explained by the marker is the basis for whether or not the marker is associated with a QTL (Tanksley *et al.* 1982). This approach has been successfully applied in detecting QTLs (Keim *et al.* 1990, Diers *et al.* 1992, Stuber *et al.* 1992, Young *et al.* 1993). In spite of its ease, single marker simple regression has three main problems. First, this method is more likely to give a biased estimation of QTL effect, underestimating the QTL effect. Second, the analysis cannot distinguish a tightly linked QTL of small-effect from a loosely linked QTL of major-effect. Third, the analysis cannot provide the genetic distance between the marker and the gene, which limits the value of the marker in practical application.

Single marker maximum likelihood estimation provides a way to overcome these two problems. The first use of likelihood statistic for linkage analysis was for analysis of human genetic linkage by Fisher (1935). Since then, the method of maximum likelihood has been commonly used in human genetic linkage analysis (Ott 1985). Weller (1986) applied maximum likelihood estimation techniques to analyze the association between a marker and a QTL in an F, family of two inbred lines. Lander *et al.* (1987) introduced QTL interval mapping strategy and the computer software for interval mapping using maximum likelihood estimation. Luo and Kearsey (1989) demonstrated the maximum likelihood estimation method to detect linkage between DNA markers and QTLs using RFLP markers. In this paper, the single marker maximum likelihood estimation approach will be compared with the single marker simple regression approach in detecting association between RAPD markers and QTLs controlling the early height growth in a ((longleaf pine x slash pine) x slash pine) BC, population.

MATERIALS AND METHODS

Plant population and field data

A ((longleaf pine x slash pine) x slash pine) BC, family was used for this study. The seeds from this backcross were sown in June 1996. Data from a total of 83 seedlings were used for this study. The total height in millimeters was measured for each of the 83 seedlings in January 1997.

RAPD markers and linkage

A total of 266 RAPD markers (150 were the F, parent-specific and heterozygous in the F, parent, and 116 were the recurrent slash pine parent-specific and heterozygous in the recurrent slash pine parent) were identified. 113 of the 150 F, parent-specific RAPD markers were mapped into 17 different linkage groups (pfl-pfl7), and 83 of the 116 recurrent slash pine parent-specific RAPD markers were mapped into 19 different linkage groups (pel-pel9). The remaining 70 RAPD markers remained unlinked (Weng *et al.* unpublished data). As no genotype information was available for the two grandparents, linkage maps can only be constructed for each of the two parents. Although marker genotypes of BC, trees and the linkage of mapped markers for each parent were available, there is no homologous information for the two parents. Therefore, when we searched for QTLs, we could only search each parent separately. Only those markers that were heterozygous in one of the parents and absent in the other parent were useful for detecting QTLs in this research. If there was an allelic QTL that was heterozygous in both parents, it would be detected as two non-allelic QTLs, one in each parent.

STATISTICAL METHODS

Assumptions and Distribution

If we consider the ((longleaf pine x slash pine) x slash pine) BC, population consisting of the 83 seedlings, each seedling can have one of the two genotypes for each QTL. The height trait was assumed to be controlled by several major-effect QTLs and many small-effect QTLs. Since there are so many QTLs, the accumulation of QTL effects can be considered as random effect and the distribution of total height can be approximated to be normal. The assumptions necessary for this research are:

1. total height trait is N(.1., 62),
2. gene action is additive,
3. there are no QTL-by- QTL interactions and no QTL-by-environment interactions, and
4. micro-environmental effects are random.

Taking into account a single QTL effect 26, the distribution will be N(I,t+6, 6_Q^2) or 1\1(1_Q, G_o^2) for the group of trees containing the QTL, and N(,t-6, a:), or N(1,1,,, 6_q^2) for the group of trees not containing the QTL.

Simple Regression

Suppose we have total height data of n individual trees with known marker genotypes,

_Y-{Y15 Y25 Yrj•

In the single-marker regression method, a RAPD marker genotype is used to represent a QTL genotype. The linkage between the marker and the putative QTL is assumed to be complete. Each RAPD marker is considered separately as the independent variable. The model will be:

$)\mathbb{L}^T =$ + (1)

where

: total height
: true mean
: QTL effect
ej : random error effect.

The marker effect can be analyzed by regression. Markers tested to be significant at a given significance level will considered to be a QTL. However, to consider a marker as a QTL is not accurate. If the QTL and the marker are not exactly at the same locus, the QTL effect will be underestimated. The farther the QTL is away from the marker, the larger will be the bias. Moreover, this model cannot tell how far the QTL is away from the marker. A more sophisticated model of regression can take the distance between the QTL and the marker into account.

For marker-present genotype, the model is:

$Y_{k,} = 11 + 041\text{-}28) + eu$ (2)

For marker-absent genotype, the model is:

= + a(2 8-1) + (3)

where

8: distance between the QTL and the marker in terms of recombination fraction.

By searching over the parameter space for various values of α and θ, the combinations that yield the smallest residual sum of squares can be determined. However, combinations will result in the same minimum residual sum of square as long as $\alpha(1-2\theta)$ or $\alpha(2\theta-1)$ is equal to the difference between the true mean and the sample mean for marker presence genotype group or marker absence genotype group. Therefore, this model cannot distinguish a tightly linked QTL of small effect from a loosely linked QTL of large effect. To obtain the accuracy, the maximum likelihood estimation was used.

Maximum likelihood estimation

For this model, we use "A" and "Q" to represent the presence of marker and the QTL, "a" and "q" to represent absence of marker and the QTL, and θ to represent the frequency of recombination between the marker and the QTL. If the parental genotypes are:

A θ Q
a θ q

the possible genotypes and their frequency for BC_1 progeny will be:

genotype	probability
A Q	$0.5(1-\theta)$
a q	$0.5(1-\theta)$
A q	0.5θ
a Q	0.5θ

When only genotypes for a single QTL are considered, each of the populations (with and without the QTL) can be considered as normally distributed. The distribution for population with and without the QTL will be $N(\mu_Q, \sigma_Q^2)$ and $N(\mu_q, \sigma_q^2)$, where

μ_Q : the true mean for population with the QTL,
σ_Q^2 : the true variance for population with the QTL,
μ_q : the true mean for population without the QTL, and
σ_q^2 : the true variance for population without the QTL.

When no linkage between the marker and QTL is considered, the distribution of Y will be a mixture of two normal distributions with same weight:

$$f(y_i \mid \mu_Q, \mu_q, \sigma_Q^2, \sigma_q^2) = 0.5\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_i-\mu_Q)^2}{2\sigma_Q^2}\right] + 0.5\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_i-\mu_q)^2}{2\sigma_q^2}\right]$$

When a marker is considered linked to a QTL, the distribution will be a mixture of two normal distributions that have different weights. The weights are related to the genetic distance between the marker and the QTL.

For individuals that have marker genotype "A," their distribution will be:

$$f(y_{Ai} \mid \mu_Q, \mu_q, \sigma_Q^2, \sigma_q^2, \theta) = (1-\theta)\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_{Ai}-\mu_Q)^2}{2\sigma_Q^2}\right] + \theta\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_{Ai}-\mu_q)^2}{2\sigma_q^2}\right]$$

For individuals that have marker genotype "a," their distribution will be:

$$f(y_{ai} \mid \mu_Q, \mu_q, \sigma_Q^2, \sigma_q^2, \theta) = (1-\theta)\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_{ai}-\mu_Q)^2}{2\sigma_Q^2}\right] + \theta\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_{ai}-\mu_q)^2}{2\sigma_q^2}\right]$$

The likelihood of the n individuals with known genotypes will be the product of all the distribution functions:

$L(\underline{Y}/, \mu_Q, \mu_q, \sigma_Q^2, \sigma_q^2, \theta) = [f(y_{A1})\ f(y_{A2})\ f(y_{A3}).....f(y_{AnA})][f(y_{a1})\ f(y_{a2})\ f(y_{a3}).....f(y_{ana})]$,

where

n_A: the total number of trees with marker genotype "A," and

n_a: the total number of trees with marker genotype "a."

To maximize the likelihood, we take the partial first derivative to each of the five parameters and set them equal to 0. The solution for θ, μ_Q, μ_q, σ_Q^2, and σ_q^2 will be the combination that maximizes the likelihood.

It can be very complicated to solve these equations. However, we could write a computer program to search the parameter space to find the combination of θ, μ_Q, μ_q, σ_Q^2, and σ_q^2 that maximizes the likelihood function. θ will be the distance between the marker and the QTL. μ_Q-μ_q will be the magnitude of the QTL effect.

The statistical inference (LOD)

The log of odds (LOD) is often used as the test statistic for mapping genomes and detecting QTLs. The LOD is the log of odds ratio of likelihood at alternative hypothesis θ (linkage) to the likelihood at null hypothesis θ_0=0.5 (no linkage).

H_0: $\theta=\theta_0=0.5$ vs H_1: $\theta=\theta_1$

$LOD = Log[L(\underline{Y}/\theta_1)/L(\underline{Y}/\theta_0)]$

$Log[L(\underline{Y}/\theta_1)/L(\underline{Y}/\theta_0)] > T$ (T: a LOD threshold)

$L(\underline{Y}/\theta_0)/L(\underline{Y}/\theta_1) < 1/10^T$

$L(\underline{Y}/\theta_1) = [f(y_{A1}/\theta_1)\ f(y_{A2}/\theta_1)\ f(y_{A3}/\theta_1)...f(y_{AnA}/\theta_1)][f(y_{a1}/\theta_1)\ f(y_{a2}/\theta_1)f(y_{a3}/\theta_1)....f(y_{ana}/\theta_1)]$

$L(\underline{Y}/\theta_0) = [f(y_{A1}/\theta_0)f(y_{A2}/\theta_0)\ f(y_{A3}/\theta_0).....f(y_{AnA}/\theta_0)][f(y_{a1}/\theta_0)\ f(y_{a2}/\theta_0)\ f(y_{a3}/\theta_0).....f(y_{ana}/\theta_0)]$

$$f(y_i \mid \theta_0) = 0.5\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_i-\mu_Q)^2}{2\sigma_Q^2}\right] + 0.5\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_i-\mu_q)^2}{2\sigma_q^2}\right]$$

$$f(y_{Ai} \mid \theta_1) = (1-\theta_1)\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_{Ai}-\mu_Q)^2}{2\sigma_Q^2}\right] + \theta_1\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_{Ai}-\mu_q)^2}{2\sigma_q^2}\right]$$

RESULTS AND DISCUSSION

The simple regression

$$f(y_{ai} \mid \theta_1) = (1-\theta_1)\left[\frac{1}{\sqrt{2\pi\sigma_Q^2}}\right]\exp\left[\frac{-(y_{ai}-\mu_Q)^2}{2\sigma_Q^2}\right] + \theta_1\left[\frac{1}{\sqrt{2\pi\sigma_q^2}}\right]\exp\left[\frac{-(y_{ai}-\mu_q)^2}{2\sigma_q^2}\right]$$

The LOD threshold is related to the Type I error rate in the following way:

$-2\ln[L(\underline{Y}/\theta_0)/L(\underline{Y}/\theta)]$ – Chi-square (df=1),

$P\{-2\ln[L(\underline{Y}/\theta_0)/L(\underline{Y}/\theta)]\} < 0.0032$,

$-2\ln[L(\underline{Y}/\theta_0)/L(\underline{Y}/\theta)] > 9.2$,

$\ln[L(\underline{Y}/\theta)/L(\underline{Y}/\theta_0)] > 9.2/2 = 4.6$,

$\log[L(\underline{Y}/\theta)/L(\underline{Y}/\theta_0)] > 2.0$.

This indicates that a Type I error rate of 0.0032 is equivalent to a LOD threshold of about 2.0.

The distribution of total height was tested and found to be normal (P=0.8674). This result satisfied our assumption that the total height is normally distributed. Subsequently, each marker was analyzed, one by one, for an association with the total height. For each marker, the tree with an unknown genotype was deleted from that test. Simple regression (model l) identified a total of 11 RAPD markers that were tested to have significant effect on the total height (Type I error of 0.0032, a p-value equivalent to a LOD of 2.0). Eight markers were located on linkage group pfl, one on group pf5, and two on group pf7. The R-square value for these 11 markers ranged from 0.139 to 0.197. One additional marker, B08_0790 on linkage group pe5, had an R-square value equal to 0.099 but was slightly above our significance threshold with a P value of 0.0041 (Table l).

Table l. The markers that were tested to be significant (P<0.0032) or to have a R-square equal to or greater than 9.9% using simple regression.

Marker	Degree of freedom	Sum of square for regression	Sum of square for error	F	R-square	group	P
110_1650	78	4441.5	22575.0	19.l	0.197	pfl	0.0001
200_0830	78	3697.6	23678.8	14.4	0.156	pfl	0.0003
299_1250	79	2535.5	22907.5	9.8	0.111	pfl	0.0024
324_1750	81	3981.2	23792.0	16.3	0.167	pfl	0.0001
384_1110	81	3690.6	23792.0	14.9	0.155	pfl	0.0002
384_1150	81	3360.4	23792.0	13.3	0.141	pfl	0.0005
G04_1250	76	3487.8	23193.4	13.5	0.150	pfl	0.0005
W02_1210	81	3297.4	23792.0	13.0	0.139	pfl	0.0005
Cl7_0670	80	2530.4	23693.0	9.6	0.107	pf5	0.0027
1810550	81	2739.4	23792.0	10.5	0.115	pf7	0.0017
E09_0810	77	3015.3	23643.0	11.3	0.128	p17	0.0012
B08_0790	79	2288.6	23004.5	8.7	0.099	pe5	0.0041

The maximum likelihood estimation using surface search

A computer program written in C language was used for calculating likelihood and LOD statistic. A range of values for μ_Q, μ_q, σ_Q^2, σ_q^2, and θ (Table 2) was searched to find the combination that maximized the LOD. The ranges were selected such that the maximum likelihood estimates of the parameters fell

into them. A total of 21 RAPD markers, belonging to eight different groups, were found to be significantly associated with total height at a LOD threshold > 2.0. Of the 21 significant markers, nine were located on linkage group pfl, five on pf5, two on pf7, one on pe5, two on pe8, one on pel8 , and one was unlinked.

Table 2. The ranges and steps for iteration of μ_Q, μ_q, σ_Q^2, σ_q^2, and θ.

Variable	Lower Bound	Upper Bound	Step Magnitude
μ_Q, μ_q	90 mm	130 mm	1 mm
σ_Q^2, σ_q^2	300	400	5
θ	0	0.36	0.02

The 11 RAPD markers found to be significantly associated with total height ($p<0.0032$) using simple regression were also significant with the maximum likelihood estimation approach (LOD>2.0). However, at the equivalent significance level, the number of significant RAPD markers using maximum likelihood estimation was about twice that using simple regression. This demonstrates the increased power obtained by employing a maximum likelihood-based approach. In addition, the maximum likelihood approach provides an estimate of the distance between the markers and their putatively linked QTLs.

A commonly used QTL genetic linkage program package, MapMaker/QTL, which employs a maximum likelihood-based interval approach, detected a total of 11 markers within the 36 linkage groups, whereas the single marker maximum likelihood estimation detected 20 (Table 3). Two possible reasons may have contributed to these results. First, the LOD may have been balanced by the other marker comprising the linked interval, possibly suggesting that MapMaker/QTL may be more conservative. Secondly, single marker maximum likelihood estimation may have absorbed some random variance, suggesting single marker maximum likelihood estimation may be more robust than a interval based approach. One last point, an unlinked marker (225_1300) was found to be significant using single marker maximum likelihood estimation but could not be detected with MapMaker/QTL, which requires information about linked markers only.

Table 3. The single marker maximum likelihood estimation test using surface search. The LOD values were compared with LOD obtained using MapMaker/TQL.

Marker	θ	μ_Q(mm)	μ_q(mm)	LOD	LOD (MapMaker)	group
110_1650	0.00	106.0	119.0	4.00	4.11	pf1
200_0830	0.00	113.0	100.0	3.17	2.75	pf1
274_0560	0.00	103.0	113.0	2.04	2.73	pf1
299_1250	0.00	116.0	106.0	2.26	2.00	pf1
324_1750	0.00	122.0	106.0	3.87	3.68	pf1
384_1110	0.00	103.0	116.0	3.47	3.05	pf1
384_1150	0.00	119.0	106.0	3.26	2.75	pf1
G04_1250	0.00	100.0	110.0	2.82	2.76	pf1
W02_1210	0.00	103.0	116.0	3.00	2.69	pf1

Table 3. Continued.

Marker	θ	μ_Q(mm)	μ_q(mm)	LOD	LOD (MapMaker)	group
141_0970	0.18	116.0	94.0	2.21	1.19	pf5
503_0400	0.26	94.0	122.0	2.21	1.59	pf5
A11_0920	0.00	103.0	113.0	2.43	0.98	pf5
C17_0670	0.16	110.0	94.0	2.82	1.39	pf5
J07_0860	0.16	94.0	110.0	2.04	1.35	pf5
181_0550	0.00	106.0	116.0	3.26	2.15	pf7
E09_0810	0.00	106.0	119.0	3.47	2.16	pf7
B08_0790	0.02	94.0	110.0	2.52	1.62	pe5
B08_1000	0.12	94.0	110.0	2.04	1.83	pe8
193_0400	0.10	110.0	94.0	2.36	1.53	pe8
D12_0820	0.00	103.0	110.0	2.00	0.97	pe18
225_1300	0.14	97.0	110.0	2.00	-	U*
175_0740	-	-	-	<2.0	1.10	pf5
U10_0680	-	-	-	<2.0	1.31	pf5
590_1220	-	-	-	<2.0	1.35	pf5
X04_0800	-	-	-	<2.0	1.21	pf5
384_0700	-	-	-	<2.0	1.44	pf7
B02_0700	-	-	-	<2.0	1.13	pf11
357_1000	-	-	-	<2.0	1.04	pf12

* U = unlinked to any groups

Multiple regression

Markers tested to be significant using maximum likelihood estimation belonged to six linkage groups. In order to avoid multicolinearity, one marker was chosen for each of these six linkage groups to do multiple regression. The multiple regression results indicated that the combination of four markers (110_1650, Cl7 0670, E09_0810, and B08_0790) from four linkage groups (pfl , pf5, pf7, and pe5) explained 43.2% of the total variance of the total height growth.

Table 4. The F tests for the four major effect linkage groups using multiple regression

Source	DF	Type III SS	Mean Square	F Value	Pr > F
110_1650	1	2660.44	2660.44	14.81	0.0003
Cl7_0670	1	2401.78	2401.78	13.37	0.0005
E09_0810	1	1164.03	1164.03	6.48	0.0132
B08_0790	1	1594.18	1594.18	8.88	0.0040

CONCLUSION

The single-marker maximum likelihood estimation approach was found to be more powerful than the simple regression approach in detecting markers linked to putative QTLs. It was also suggested that the single marker maximum likelihood estimation approach may be a more robust approach for detecting QTLs than interval based approach such as that used by MapMaker/QTL.

ACKNOWLEDGMENTS

The author thanks Kristel Davis, Mary Bowen, Glen Johnson, Wende Wu, Dr. Brian Marx, and Dr. Louis Escobar for their assistance and suggestions.

LITERATURE CITED

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics* 8:608-612.

Fisher, R. A. 1935. The detection of linkage with dominant abnormalities. *Ann. Engen.* 6:187-201.

Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. *Genetics* 126:735-742.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. MapMaker: An interactive computer pachage for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.

Luo, Z. W. and M. J. Kearsey. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity* 63:401-408.

Ott, J. 1985. Maximum likelihood estimation. In *Analysis of Human Genetic Linkage* p33-40 (John Hopkins University Press, Baltimore and London).

Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris, and E.S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132:823-839.

Tanksley, S.D., H. Medina-Filho, and C.M. Rick,. 1982. Use of naturally occuring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49:12-25.

Weller, J. I. 1986: Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42:627-640

Welsch, J., and M. McClelland. 1990: Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18:7213-7218.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18:6531-6535.

Young, N.D., D. Danesh, D. Menancio-Hautea, and L. Kumar. 1993. Mapping oligogenic resistance to powdery mildew in mungbean with RFLPs. *Theoretical and Applied Genetics* 87:243-249.