GENETIC MAPPING OF OLIGOGENIC TRAITS USING MAXIMUM LIKELIHOOD FUNCTIONS - BACKCROSSING POPULATIONS

Rodrigo Barros ROCHA[1]
Júpiter Israel MURO-ABAD[2]
Cosme Damião CRUZ[3]
Willian Silva BARROS[4]
Everaldo Gonçalves BARROS[3]
Elza Fernandes de ARAÚJO[5]

- *ABSTRACT: The qualitative nature and epistatic interaction underlying the oligogenic inheritance pattern should not be interpreted by traditional methods for QTL (Quantitative Trait Loci) detection. The objective of this study is to propose an alternative method for oligogenic loci detection– OTL (Oligogenic Trait Loci). In the Oligogenic Trait Mapping Method (O.T.M.M.), maximum likelihood probability functions were used to obtain "r" estimates, that express the distance among the molecular markers and the OTL loci, adjusted to the data using conditional probabilities of occurrence of the molecular markers and the OTL loci. Results show that the method was adequate for OTL detection even in relatively small backcrossing populations. Comparative analysis showed that the "r" O.T.M.M. estimates were more accurate than the single marker and simple interval mapping estimates. The O.T.M.M. does not need previous information on marker order in the linkage groups and is not affected by the presence of more than one QTL in a same linkage group.*

- *KEYWORDS: Genetic mapping; Q.T.L.; O.T.L.; O.T.M.M.*

1 Introduction

Oligogenic traits of discrete distribution and expression controlled by few genes of major effect are important for several crop breeding programs (Dekkers & Hospital, 2001; Rao & Li, 2000). For plant breeding proceedings, genetic disease resistance is probably one of the most important traits of oligogenic nature (Alzate-Marin et al., 2005). Plant resistance to Ceratocystis (Alfenas et al., 2004) and to rust (Junghans et al., 2003), are some examples of traits of economic interest for plant breeding, most likely of oligogenic inheritance.

[1] Centro de Pesquisas Agroflorestais de Rondônia, EMBRAPA, 78900-970, Porto Velho, RO – Brazil. E-mail: *rodrigo@cpafro.embrapa.br*
[2] Aracruz Celulose, 29197-900. Aracruz, ES – Brazil. E-mail: *jmuro@aracruz.com.br*
[3] Departamento de Biologia Geral, Universidade Federal de Viçosa – UFV, 36570-000. Viçosa, MG – Brazil.. E-mail: *cdcruz@ufv.br / ebarros@ufv.br*
[4] Universidade Federal de Viçosa – UFV, 36570-000. Viçosa, MG – Brazil. E-mail: *willian@pq.cnpq.br*
[5] Departamento de Microbiologia, Universidade Federal de Viçosa – UFV, 36570-000. Viçosa, MG – Brazil. E-mail: *ezfa@ufv.br*

The successful identification of appropriate genome regions for assisted selection procedures depends on an adequate experimental quality to ensure reliable estimates of genetic linkage among markers and QTL´s (Dias et al., 2005; Muro-Abad, et al., 2005; Badu et al., 2004; Yousef & Juvik, 2001). The number of evaluated individuals, the nature of the segregating population, and the map saturation are among the most important factors that affect the reliability of the genetic linkage estimates (Varshney et al., 2005; Young, 1999).

Methods for quantitative trait loci (QTL) identification are well documented, whereas the options for the analysis of control loci of oligogenic traits - OTL (Oligogenic Trait Loci) are sparser. The discrete trait expression and the epistatic interaction of multiple factors are not considered by the presumptions of the traditional quantitative methods that presuppose normality and inheritance controlled by many genes of small effect (Li, 2004; McIntyre et al., 2000; Xu et al., 1998; Zeng, 1994, Lander & Botstein, 1989). This limitation was also evaluated elsewhere. Visscher et al. (1996) compared methods using linear regression and generalized linear regression based on binary data and concluded that the methodologies do not differ in prediction accuracy and test power. Xu & Atchley (1996) extended the methodology of composite interval mapping to binary data using logistic regression models. Rao & Li (2000) presented strategies for the analysis of discrete variables using different allele models. Other authors propose Bayesian approximations for OTL detection (Yi et al., 2004).

Maximum likehood (M.L.) methods are quite popular in QTL mapping literature however none of the previous analyses strategies approaches the question as done in this work. In this method, defined as the Oligogenic Trait Mapping Method (O.T.M.M.), maximum likelihood probability functions were used to obtain "r" estimates that express the distance among the molecular markers and the OTL loci, based on the conditional probabilities of the molecular markers and OTL genotypes.

This study aimed to present an additional detection method of loci that control the expression of oligogenic traits (OTL), emphasizing epistatic segregations controlled by two and three genes in backcross populations.

2 Material and methods

2.1 OTL Detection – method of oligogenic trait mapping (O.T.M.M.)

The maximum likehood estimate (MLE) of the unknown parameters, $\hat{\Theta}$, is the value of Θ that corresponds to the maximum of $L(\Theta; x)$. The value of Θ that is most likely to have produced the observed data, x, according to the notation used by Schuster & Cruz, 2004 is:

$$L(\Theta; x) = \prod_{i=1}^{n} f(\Theta; x_i) \qquad (1)$$

Θ :vectors unknown parameters, x : observed data, n : number of observations.

In this application, the maximum likehood function represents the product of the individual's density probability functions of the oligogenic trait expression in N

observations; affected by the occurrence of the genotypic classes in the segregating populations. Liu, 1997 demonstrate that the markers segregation follows the multinomial distribution.

$$f(x) = \lambda p_1^{n1} p_2^{n2} ..... p_n^{n_n} \tag{2}$$

$$\lambda = \frac{N!}{n_1! n_2! ... n_n!} \tag{3}$$

$$N = \sum_{i=1}^{n} n_i \tag{4}$$

x : random variable, n_i : i´th event, p_i : occurrence probability of the i´th event

The oligogenic trait mapping method (O.T.M.M.) is based in the use of likelihood functions to estimate the distance among the markers and the OTL locus, based in the conditional probabilities of the oligogenic trait expression and the marker genotypes in the A, B and C loci:

$$P(Q / A_i B_j C_k) = \frac{P(Q \cap A_i B_J C_K)}{P(A_i B_j C_k)}, \tag{5}$$

in which:

$P(Q / A_i B_j C_k)$: conditional probability of manifest the binary phenotype considering the occurrence of the i´th genotype in the A loci, the j´th genotype in the B loci, k´th genotype in the C loci.

$P(Q \cap A_i B_J C_K)$: joint probability of manifest the binary phenotype considering the occurrence of the i´th genotype in the A loci, the j´th genotype in the B loci, k´th genotype in the C loci.

$P(A_i B_j C_k)$: marginal ocurrence probability of the genotype $A_i B_j C_k$ in the population.

In this way, the maximum likehood function can be expressed as (Schuster & Cruz, 2004):

$$L(y) = \prod_{i=1}^{n} f(y_i; M_i) \tag{6}$$

y : phenotypic vector of the trait binary expression.

y_i : binary phenotypic manifestation of the trait in the i´th individual.

M_i : marker genotype in the i´th individual.

$$f(y_i; A_i B_j C_k) = \lambda P(Q / A_i B_j C_k)^{n1} P(q / A_i B_j C_k)^{n2} \quad (7)$$

$P(Q / A_i B_j C_k)$: conditional probability of manifest the binary phenotype considering the occurrence of the i´th genotype in the A loci, the j´th genotype in the B loci, k´th genotype in the C loci.

$P(q / A_i B_j C_k)$: conditional probability of not manifest the binary phenotype considering the occurrence of the i´th genotype in the A loci, the j´th genotype in the B loci, k´th genotype in the C loci.

n_1 : number of manifested binary phenotypes observed in the population.

n_2 : number of not manifested binary phenotypes observed in the population.

The likelihood ratio was used to test the linkage between the marker and the oligogenic trait.

$$LOD = Log_{10}\left[L\frac{(y_i; N, r)}{(y_i; N, r = 0{,}5)} \right] \quad (8)$$

In which :

r : is the value of factorial linkage between marker locus and OTL.

The consistency, minimal variance, asymptotic normality, asymptotic efficiency, and the estimates of variance that can be predictable by the inverse index of information content are some desirable properties of the likelihood-function based methods. The M.L. functions are available for use in the software GQMOL (Cruz, 2008) - http://www.ufv.br/dbg/genes/gdown.htm.

2.2 Data simulation – reference genome, parental genotypes and mapping populations

The simulation mode of the software GQMOL (Cruz, 2008) was used to simulate backcross populations derived from contrasting parents. Seeking the assessment of populations with the most parsimonious genomic configuration for the QTL´s analysis, was simulated a genome with four linkage groups, 30 co-dominant equidistant markers and map saturation of 5 cM. The linkage groups 1 and 2 contain loci that affect the trait expression (OTL1 – located between markers C14 and C15 and OTL2 – located between markers C214-C215). The linkage group 3 has a locus that influences the expression only when the trait presents epistatic interactions among three genes (OTL3- located between markers C325 and C326). Linkage group 4 contains no loci that alter the trait expression in any of the cases studied.

2.3 Oligogenic trait simulation

Two oligogenic traits denominated C1 and C2 controlled by two and three genes of epistatic interaction, with expected ratios of 3:1 and 7:1 were simulated. The expression of these binary traits can be classified by the presence (1) or absence (0) governed by the

epistatic interaction among the oligogenic loci. Two markers of the original genome were related with the expression of the oligogenic trait C1 (OTL1 and OTL2) and three markers related with the expression of oligogenic trait C2 (OTL1, OTL2 and OTL3).

2.4 Genome analyses – analysis of individual loci and marker pairs.

The marker segregation ratio was verified using chi-square (χ^2). Recombination estimates were obtained by the graphic method in which the solution is the factorial linkage value in the interval of 0 to 0.5 cM with the highest "LOD score" value (Schuster & Cruz 2004). The linkage groups were clustered based on the maximal recombination frequency (r_{max} = 30%) and minimal LOD (LOD_{min} = 3). The best marker order was estimated by the SARF method (Sum of Adjacent Recombination Fractions).

2.5 Comparison of QTL detection methods

To compare the methods, 100 replications of backcross populations composed of 50, 100, 200, and 1000 plants and a single trait controlled by three loci were evaluated. Two loci that control the trait expression were located in the same linkage group (loci C15 and C120, linkage group 1) and a third locus in other linkage group (locus C215, linkage group 2). The candidate markers number indicated by the methodologies and the accuracy of the "r" estimates that express the distance among the molecular markers and the OTL loci were used in the comparisons.

3 Results and discussion

3.1 Data simulation

The simulation parameters were determined aiming the interpretation of results similar to those acquired in laboratory conditions, considering that: several mapping studies were conducted with approximately 200 individuals (Liu, 1997; Silva et al., 2007); diploid species are representative of the cultivated plant species; the most modern molecular marker techniques have co-dominant nature (Bunyamin et al., 2002, Charlesworth et al., 1994); four linkage groups are sufficient to exploit the epistasis interactions among two and three genes; a genetic map with saturation of 5 cM is considered moderately saturated (Mackay, 2001).

Backcrossing populations are especially important to breeding programs and to genetic studies, besides practical advantages for breeding and multiplication, these populations can be rapidly established in field. The contribution of only one parent to the progeny segregation facilitates the addictive gene effects isolation in breeding and QTL detection (Schuster & Cruz, 2004).

3.2 Estimates of conditional probability and likelihood functions simulation

The conditional probability estimates of the inheritance pattern needed to obtain the likehood functions are based on the following presuppositions: i) binary trait expression, ii) inheritance controlled by few genes and iii) multinomially distributed segregation

pattern. The likelihood functions for the epistatic proportions 3:1, controlled by 2 genes, and 7:1, controlled by 3 genes, for backcross populations are given below:

Backcross population, epistatic interaction between 2 genes (3:1)

$$L(p_i;n_i)=\lambda\left(\frac{2-r}{4}\right)^{n1}\left(\frac{r}{4}\right)^{n2}\left(\frac{1+r}{4}\right)^{n3}\left(\frac{1-r}{4}\right)^{n4} \quad (9)$$

Backcross population, Epistatic interaction among 3 genes (7:1)

$$L(p_i;n_i)=\lambda\left(\frac{4-r}{8}\right)^{n1}\left(\frac{r}{8}\right)^{n2}\left(\frac{3+r}{8}\right)^{n3}\left(\frac{1-r}{8}\right)^{n4} \quad (10)$$

p_i : occurrence probability of the i´th event; n_i : i´th event,

The use of other conditional probability values allows the obtaining of likelihood functions for others populations types and oligogenic traits inheritance pattern. The main limitation of this method is that M.L. is computationally intensive, requiring rather special programs to solve the likehood equations. All the models developed in this work are available in the software GENES.

3.3 Genetic mapping

According to Liu (1997) the development of a high quality genetic map is necessary to an accurate QTL detection directly related with the simulation parameters used. Prior to pair marker analysis, the individual marker segregation was tested by the chi-square test with Bonferroni correction for multiple tests (data not shown). No distorted segregation was detected, according to the chi-square test. Simulation studies allows the isolation of the random effect as the main source of interference in the gamete production and genotype recombination; and in this case, even if the segregation of some marker is biased, it should not be discarded (Liu, 1997).

Most used values of LOD score and maximum recombination rate were applied to determine the linkage groups (LOD = 3.0 and r = 0.30). The minimal LOD score value observed between the markers C311 and C312 ($LOD_{minimal}$ = 34.95) is a strong evidence of factorial linkage in the genetic map used for OTL detection. The deviations of the estimated values are comparable with those obtained by Silva et al., 2007. There were no observed unlinked markers. After estimating recombination between adjacent markers, the most likely marker order was determined by the SARF method.

Other studies also suggest that the observed results, in view of the practical nature of the mapping studies, were sufficient to QTL analysis. The evaluation of the mean information content and the confidence intervals of the "r" estimates indicate that the backcross and F2 populations mapped with co-dominant markers are among the breeding populations with the most precise recombination linkage estimates (Rocha et al., 2004; Liu, 1997). All the four original linkage groups with a mean saturation of 5.12 cM were recovered (Figure 1).

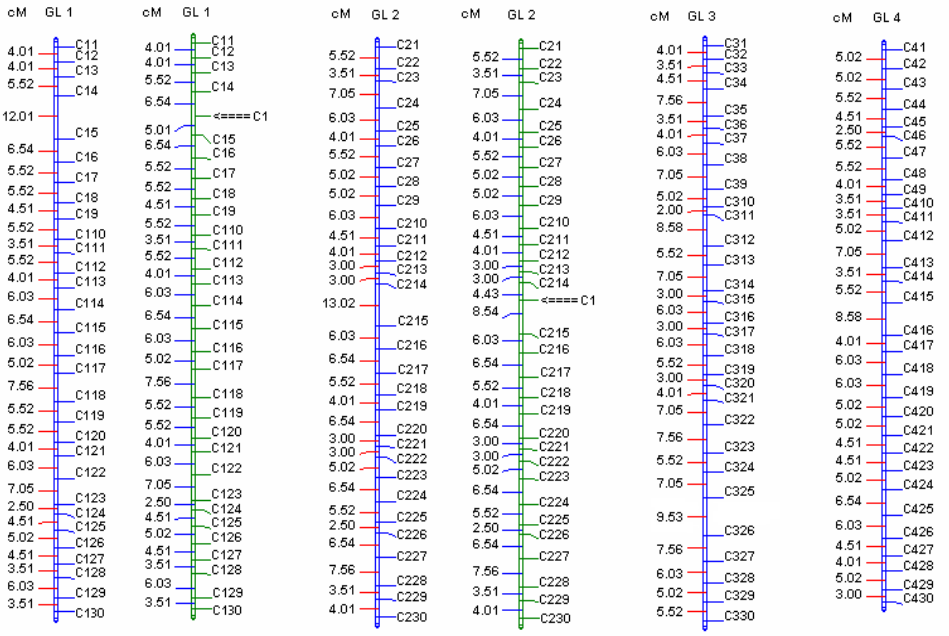

Figure 1 - Estimated genetic map of a backcross population composed by 200 individuals. The linkage groups (LOD = 3, r_{max} = 0.30 and Kosambi unit) were estimated by the analytical mapping method. The linkage groups highlighted in green contains the OTL loci position of the C1 estimated by O.T.M.M. method.

3.4 QTL detection

The O.T.M.M. allows an estimation of the factorial linkage among molecular markers and OTL associated to LOD scores values, that measures the reliability of the factorial linkage estimates (Table 1). The underlying data identified the OTL´s in the following intervals (C1 - intervals C14:C15 and C214:C215 C2 - intervals C14:C15, C214:C215 and C325 and C326) (Table 1).

The O.T.M.M method estimates the distance among the markers and the QTL locus individually and does not require previous knowledge on the marker order. The possibility of identifying OTL without knowing the marker order in the linkage groups is somewhat appealing, since numerous mapping studies have demonstrated the occurrence of unlinked markers for trait expression (Rocha et al., 2007, Novaes 2006).

After the identification of the putative OTL loci linkage group, the oligogenic loci distance estimates may be added to the other linkage group markers distance estimates, to test the factorial linkage and determine the best order. The inclusion of the OTL marker can slightly change the distance between adjacent markers considering that the factorial linkage values are not perfectly addictive (Figure 1).

Table 1 - O.T.M.M "r" estimates for the traits C1 and C2 in all genome positions with the associated LOD score values

Marker	GL	Dist(cM)	LOD	Dist(cM)	LOD	Marker	GL	Dist(cM)	LOD	Dist(cM)	LOD
		C 1		C 2				C 1		C 2	
C11	1	13.9	7.31	16,0	2,98	C31	3	48.2	0.02	48.3	0.01
C12	1	11,6	8,36	11,9	3,86	C32	3	48,2	0,02	48,3	0,01
C13	1	8,8	10,02	11,5	4,05	C33	3	48,8	0,01	44,4	0,07
C14	1	6,5	11,61	7,4	5,24	C34	3	46,0	0,08	41,8	0,16
C15	1	4,9	16,98	4,6	8,45	C35	3	46,5	0,06	35,6	0,47
C16	1	5,0	11,77	4,1	6,02	C36	3	46,0	0,08	32,6	0,68
C17	1	10,1	8,82	8,4	4,61	C37	3	48,2	0,02	35,1	0,51
C18	1	16,9	5,95	16,6	2,81	C38	3	48,2	0,01	37,4	0,37
C19	1	16,5	6,15	16,3	2,89	C39	3	49,4	0,00	42,8	0,12
C110	1	18,4	5,47	20,0	2,24	C310	3	49,4	0,00	38,4	0,31
C111	1	25,1	3,19	21,9	1,88	C311	3	48,2	0,01	37,4	0,37
C112	1	24,6	3,34	21,5	1,95	C312	3	44,2	0,17	29,2	1,01
C113	1	24,3	3,49	21,1	2,02	C313	3	51,6	0,01	34,1	0,59
C114	1	27,8	2,58	28,8	1,06	C314	3	52,5	0,07	42,8	0,12
C115	1	25,6	3,16	24,6	1,56	C315	3	52,5	0,10	46,1	0,03
C116	1	26,4	2,98	19,4	2,39	C316	3	52,5	0,13	38,4	0,31
C117	1	27,3	2,71	24,1	1,62	C317	3	52,5	0,07	34,1	0,59
C118	1	30,8	1,94	27,5	1,23	C318	3	52,5	0,19	40,5	0,20
C119	1	33,6	1,41	32,1	0,75	C319	3	52,5	0,17	44,4	0,07
C120	1	38,6	0,67	36,9	0,40	C320	3	52,5	0,20	45,5	0,04
C121	1	35,6	1,05	34,1	0,59	C321	3	52,5	0,23	44,4	0,07
C122	1	35,6	1,05	34,1	0,59	C322	3	52,5	0,17	40,1	0,22
C123	1	35,8	1,04	32,1	0,75	C323	3	52,5	0,04	28,8	1,06
C124	1	34,6	1,22	28,8	1,06	C324	3	45,4	0,11	14,0	3,12
C125	1	34,1	1,31	28,4	1,12	C325	3	46,5	0,06	9,0	4,30
C126	1	38,1	0,74	36,4	0,43	C326	3	43,1	0,24	6,8	7,70
C127	1	39,3	0,60	39,6	0,25	C327	3	37,9	0,74	4,6	5,45
C128	1	36,4	0,96	34,9	0,54	C328	3	41,3	0,38	14,3	3,03
C129	1	34,7	1,21	31,3	0,84	C329	3	43,6	0,20	21,9	1,88
C130	1	32,5	1,60	23,0	1,82	C330	3	47,6	0,03	25,8	1,37
C21	2	27,1	2,64	23,5	1,61	C41	4	47,1	0,04	42,8	0,12
C22	2	25,6	3,05	22,2	1,81	C42	4	48,2	0,02	41,5	0,16
C23	2	28,8	2,33	25,4	1,44	C43	4	52,5	0,04	50,6	0,00
C24	2	27,8	2,58	24,6	1,56	C44	4	48,8	0,01	44,4	0,07
C25	2	25,8	2,94	19,8	2,10	C45	4	49,4	0,00	44,9	0,06
C26	2	24,8	3,23	18,9	2,25	C46	4	51,1	0,01	46,6	0,03
C27	2	21,7	4,06	14,0	3,12	C47	4	48,2	0,02	43,8	0,08
C28	2	21,7	4,06	14,0	3,12	C48	4	47,6	0,03	43,2	0,10
C29	2	16,8	5,67	9,4	4,10	C49	4	45,4	0,11	45,5	0,04
C210	2	19,1	5,10	12,4	3,67	C410	4	42,5	0,28	42,8	0,12
C211	2	16,5	6,15	12,1	3,77	C411	4	40,8	0,43	38,9	0,28
C212	2	15,7	6,55	11,6	3,96	C412	4	40,8	0,43	43,3	0,10
C213	2	13,5	7,52	11,6	3,96	C413	4	42,0	0,33	42,2	0,14
C214	2	3,8	15,47	8,1	4,82	C414	4	44,8	0,14	42,8	0,12
C215	2	5,8	9,29	5,9	7,82	C415	4	47,6	0,03	47,7	0,01
C216	2	6,4	9,91	9,6	4,00	C416	4	51,6	0,01	49,4	0,00
C217	2	9,3	8,59	14,3	3,03	C417	4	47,1	0,04	40,6	0,20
C218	2	11,6	7,65	18,6	2,33	C418	4	47,1	0,04	40,6	0,20
C219	2	17,3	5,47	28,1	1,09	C419	4	43,6	0,20	35,1	0,51
C220	2	18,9	5,01	27,1	1,20	C420	4	44,8	0,14	40,5	0,20
C221	2	16,9	5,95	25,0	1,50	C421	4	50,6	0,00	50,6	0,00
C222	2	20,0	4,75	25,8	1,37	C422	4	45,4	0,11	45,5	0,04
C223	2	23,7	3,64	29,2	1,01	C423	4	43,6	0,20	46,1	0,03
C224	2	27,3	2,71	24,1	1,62	C424	4	45,9	0,08	46,1	0,03
C225	2	27,8	2,58	28,8	1,06	C425	4	48,8	0,01	46,6	0,03
C226	2	28,3	2,45	29,2	1,01	C426	4	52,5	0,03	50,0	0,00
C227	2	28,8	2,33	34,1	0,59	C427	4	48,2	0,02	43,7	0,08
C228	2	29,3	2,21	34,6	0,55	C428	4	45,9	0,08	46,1	0,03
C229	2	30,9	1,86	31,6	0,77	C429	4	44,7	0,14	40,4	0,20
C230	2	28,0	2,48	22,2	1,81	C430	4	45,9	0,08	39,2	0,25

G.L. – Linkage group; Dist (cM) - Distance between the marker and the OTL; LOD – Value of the LOD estimates.

The correct assumption of the trait segregation pattern is probably the chief constraint for the application of this method. The distinction between the segregation ratios 3:1 and 7:1 is relatively simple (an analysis of seven plants allows a distinction at a level of 5%), but can be complex in other types of genetic evaluations. Over 15,000 plants are necessary, for example, to discriminate the segregation ratios 9:7 from 27:37 in F2 populations, working with the same statistics and same error probability.

3.5 Comparison of QTL detection methods: candidate markers

If there were no experimental limitations of costs, time and labor, less precise methods could be equally used for the detection of oligogenic trait loci. The hypothesis tested in this work, is that a method that consider the presuppositions regarding loci number and the epistatic interaction nature results in a more accurate analysis with higher test power. The candidate markers number may be used as an indicator of the power test since the number of undetected QTL can be interpreted as an estimate of the occurrence of error type I and the identification of QTL in genome positions with distance over 50 map units as an estimate of the occurrence of error type II (Benjamini et al., 2005; Bernardo, 2004; Fernando et al., 2004).

The results further indicate that the number of S.M. candidate loci is nearly four times the number of candidate markers detected by the O.T.M.M. The higher number of candidate markers difficult the selection of few major-effect markers for assisted selection. In comparison with the S.M., the number of OTL loci undetected (error type I) by the likelihood functions was higher, but tended towards zero with the population increasing (Table 2). The S.M. method has clear constraints, such as the confounding the position and the QTL effect, and the occurrence of errors type II was relatively common even in large populations size (Table 2).

Table 2 - Proportion of type I error and type II error in the candidate markers indicated by the O.T.M.M. and S.M. methodologies in backcross populations composed by 50, 100, 200 and 1000 individuals

Method	Error type	Evaluated genotypes			
		50	100	200	1000
$S.M._{5\%}$	Type I error	0,00	0,00	0,00	0,00
	Type II error	0,65	0,72	0,75	0,60
$S.M._{1\%}$	Type I error	0,00	0,00	0,00	0,00
	Type II error	0,66	0,59	0,26	0,09
$O.T.M.M._{(LOD\geq 3)}$	Type I error	0,16	0,05	0,00	0,00
	Type II error	0,11	0,23	0,09	0,00

The occurrence of the errors type I and II have different implications to assisted selection. The error type II causes more prejudice to plant breeding considering that its results in erroneous candidate markers identification, hampering the assisted selection (Fernando et al., 2004). The joint interpretation of the likelihood functions after an exploratory single marker the markers analysis allows filter few more promissory

candidate markers from the initial marker set. The more restricted O.T.M.M. candidate marker group set is generally comprised in the major pool of possible markers indicated by the S.M. method, since the single marker F distribution values and the LOD of the likelihood functions have high correlation values (Liu, 1997).

3.6 Comparison of QTL detection methods: estimates of factorial linkage

The deviation between the parametric and the estimated "r" distributions allows the identification of the most accurate methodologies. The O.T.M.M. "r" distribution shows a better adjustment to the parametric values (Figure 2). The smaller deviations of the O.T.M.M. estimates allowed an accurate identification of the two major OTL loci that affect the expression of the trait C1 and the three major OTL loci that affect the expression of the trait C2 (Figure 2).

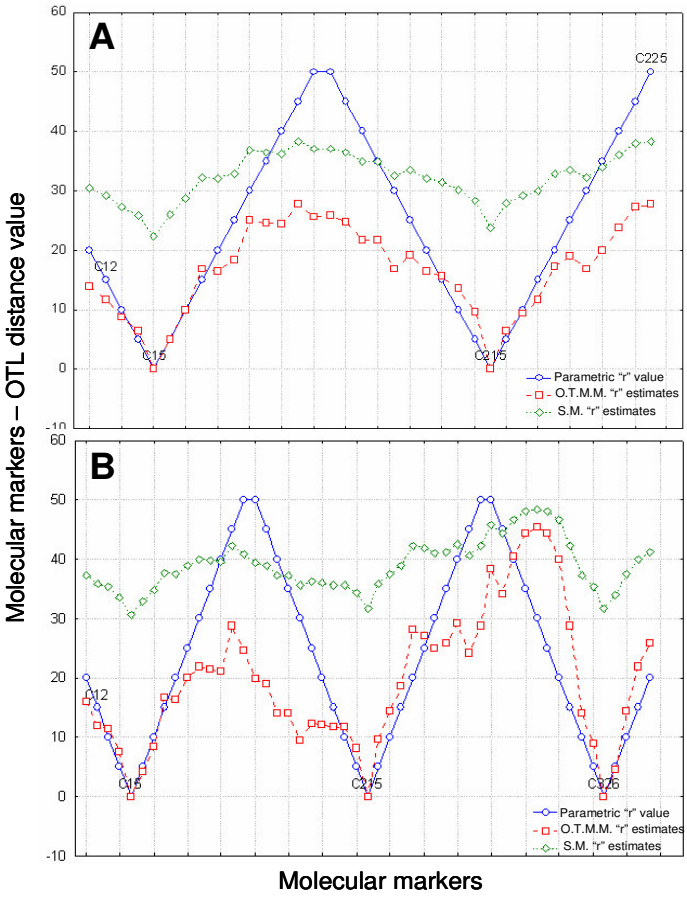

Figure 2 - O.T.M.M and S.M. "r" estimates distributions. A – "r" estimates of the C1 trait with expression controlled by two major oligogenic loci, B- "r" estimates of the C2 trait with expression controlled by three major oligogenic loci.

Unlike the S.M. method, in which the estimation of the distance values among OTL and marker locus depends on the magnitude of the mean contrast between the genotype groups that have and do not have the marker loci, the O.T.M.M. estimation is based on best fitting data estimates that consider the variation in all genotypic classes. In backcross populations, the O.T.M.M. hypotheses are therefore tested with three degrees of freedom and the S.M. hypothesis with only one.

The O.T.M.M. and S.M. "r" distribution were also compared to the simple interval mapping (I.M) LOD score distributions (Figure 3). As expected, an increase in the number of the evaluated genotypes raises the accuracy of all methodologies. The S.M. estimates of marker-OTL factorial linkage were the least accurate (Chen & Chen, 2005). I.M estimates identified the OTL even in experiments with a relatively low number of plants. However, the I.M results also indicated the occurrence of ghosts QTL in all replications and populations sizes.

Moreover, besides this well-known limitation (Carbonell et al. 1992), regions with LOD values higher than 3 and distance over than 50 map units from the OTL, were observed in 83% of the evaluated replications. This restriction hampered the OTL identification in the genome interval between OTL1 and OTL2; and LOD scores values higher than 3 were found in linkage group regions independent from the OTL segregation loci.

The O.T.M.M. identified quite accurate the OTL presence even in the experiments with low number of plants and with two OTL loci localized in the same linkage group (Figure 3). Although the literature offers a broader set of methods than evaluated here, there is no single method that applies well to all data. The use of the best-suited tools enables breeders to verify conclusions and evaluate their reliability in view of the error associated with the hypothesis test.

The O.T.M.M. properties are complementary to other methodologies and can be considered an auxiliary method for the characterization of the most important genome regions. The likelihood functions used in the O.T.M.M. are more accurate and restrictive than the single marker, do not require a previous marker order and are free of confounding adjacent OTL in the same linkage group.

Conclusions

The use of likehood functions to specifically foresee the oligogenic segregation pattern allows a unique interpretation with the following properties:

1. Allow the characterization of the oligogenic traits nature preventing the interpretation of a discrete distribution trait as a continuous trait.
2. Present higher test power than others exploratory QTL methodologies (simple marker and simple interval marker).
3. Does not need previous mapping information, it´s not affected by two QTL´s in the same linkage group.
4. Combine the properties of the M.L. estimates. The presumption of the correct oligogenic trait segregation pattern for analysis and the high number of calculate needed to solve the likehood equations can be considered it´s the main limitations.

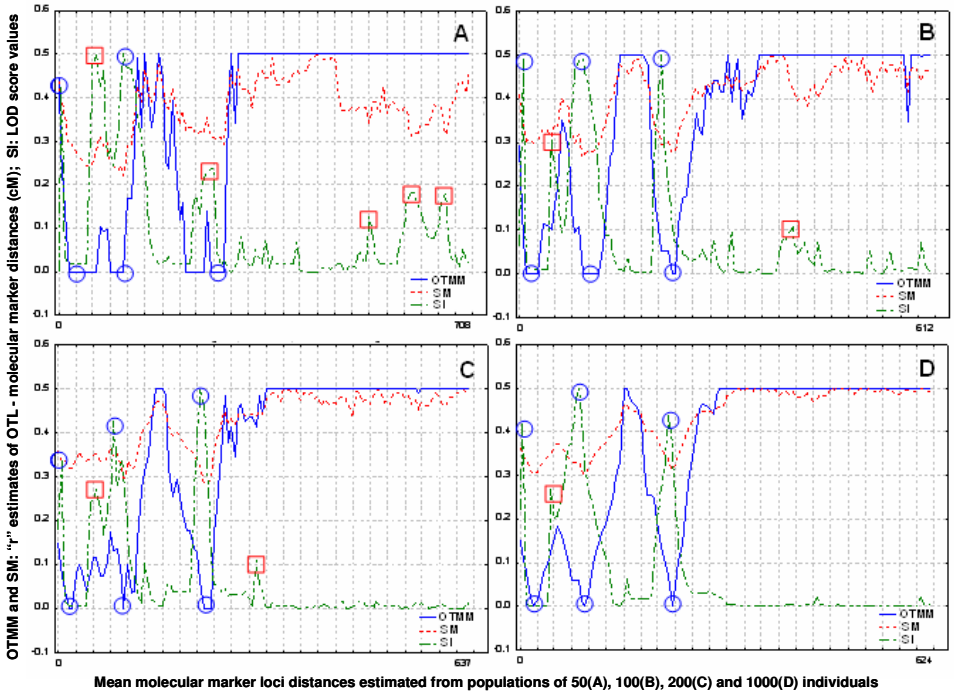

Figure 3 - O.T.M.M., S.M. "r" estimates distributions and I.M. LOD score distribution in all evaluated genome positions. The O.T.M.M. and S.M. estimates shows the distance among the markers and the OTL loci and the L.S estimates shows the LOD score values associated to the H_o hypothesis of QTL occurrence in the evaluated position. The circles highlight the correctly identification of the OTL loci and the squares the OTL located in genome regions not associated to the oligogenic trait. The estimates were based on A – 50 observations, B – 100 observations, C – 200 observations, D – 1000 observations.

Acknowledgements

The authors wish to thank the Brazilian agencies CAPES, CNPq and FAPEMIG for financial support.

ROCHA, R. B.; MURO-ABAD, J. I.; CRUZ, C. D.; BARROS, W. S.; BARROS, E. G.; ARAÚJO, E. F. Metodologia para mapeamento de locos controladores de características oligogênicas utilizando funções de máxima verossimilhança – populações de retrocruzamento. *Rev. Bras. Biom.*, São Paulo, v.26, n.3, p.7-21, 2008.

- RESUMO: A natureza discreta e a interação epistática destas características oligogênicas resultam em um padrão de herança que não deveria ser interpretado pelos métodos tradicionais de detecção de QTL. O objetivo deste trabalho é propor um método para mapeamento e detecção de locos controladores da expressão de características oligogênicas, OTLs ("Oligogenic Trait Loci").

Este método, definido como M.M.C.O (Método de Mapeamento de Características Oligogênicas), utiliza funções de verossimilhança para obtenção de estimativas de recombinação "r" de melhor ajuste aos dados, a partir das probabilidades condicionais de ocorrência entre o loco marcador e a herança da característica oligogênica. Análise comparativa mostra que as estimativas M.M.C.O. de ligação fatorial entre o loco marcador e os locos OTLs são mais acuradas do que as obtidas pelos métodos de marca simples e de intervalo simples. O M.M.C.O. não necessita de ordenamento prévio dos marcadores e não é influenciado pela ocorrência de OTL´s em um mesmo grupo de ligação.

- PALAVRAS-CHAVE: Mapeamento genético; Q.T.L.; O.T.L.; M.M.C.O.

References

ALFENAS, A. C.; ZAUZA, E. A. V.; MAFIA, R. G.; ASSIS, T. F**.** *Clonagem e doenças do eucalipto*. Viçosa: Editora UFV, 2004. 442p.

ALZATE-MARIN, A. L.; CERVIGNI, G. D. L.; MOREIRA, M. A.; BARROS, E. G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. *Fitopatol. Bras.*, Brasília**,** v.30, n.4, p.333-342, 2005.

BADU, R.; SUDHA, K. N.; PRASANNA, B. M.; GUPTA, H. S. Integrating marker assisted selection in crop breeding - prospects and challenges. *Curr. Sci.,* Bangalore, v.87, n.5, p.607-619, 2004.

BENJAMINI, Y.; YEKUTIELI, D. D. Quantitative trait loci analysis using the false discovery rate. *Genetics*, Baltimore, v.171, n.2, p.783-790, 2005.

BERNARDO, R. What proportion of declared QTL in plants are false? *Theor. Appl. Genet.,* Berlin, v.109, p.419-424, 2004.

BUNYAMIN, T.; THOMAS, E. M.; PAULS, K. P. Genetic mapping of agronomic traits in common bean. *Crop Sci.*, Madison, v.42, n.2, p.544-556, 2002.

CARBONELL, E. A.; GERIG, T. M.; BALANSARD, E.; ASIN, M. J. Interval mapping in the analysis of non-additive quantitative trait loci. *Biometrics*, Washington, v.48, p.305-315, 1992.

CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature,* London, v.371, n.6494, p.215-220, 1994.

CHEN, Z.; CHEN, H. On some statistical aspects of the interval mapping for QTL detection. *Stat. Sinica***,** Taipei, v.15, p.909-925, 2005.

CRUZ, C. D. *Programa para análises de dados moleculares e quantitativos - GQMOL***.** (2008) Viçosa: UFV. Software em desenvolvimento, disponível em "http://www.ufv.br/dbg/genes/gdown.htm". Acesso em fev. 2008.

DEKKERS, J. C. M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. *Nat. Genet.*, New York, v.3, n.1, p. 22-32, 2001.

DIAS, L. A .S.; ROCHA, R. B.; PICOLI, E. A. T. Distinctness of cacao cultivars using yield component data and RAPD markers. *Crop Breed. Appl. Biotechnol.*, Londrina, v.5, n.1, p.47-54, 2005.

FERNANDO, R. L.; NETTLETON, D.; SOUTHEY, B. R.; DEKKERS, J. C. M.; ROTHSCHILD M. F.; SOLLER, M. Controlling the proportion of false positives in multiple dependent tests. *Genetics*, Baltimore, v.166, n.1, p.611-619, 2004.

JUNGHANS, D. T.; ALFENAS, A. C.; BROMMONSHENKEL, S. H.; ODA, S.; MELLO, E. J.; GRATTAPAGLIA, D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus mode of inheritance and mapping of a major gene with RAPD markers. *Theor. App. Genet.*, Berlin, v.108, n.1, p.175-180, 2003.

LANDER, E. S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, Baltimore, v.121, n.1, p.185-199, 1989.

LI, J. *Development of multiple interval mapping for mapping QTL in ordinal traits*. 2004. 96f. Thesis (PhD) – North Carolina State University, Raleigh, 2004.

LIU, B. H. *Statistical genomics:* linkage mapping and QTL analysis. Boca Raton: CRC Press, 1997. 511p.

MACKAY, T. F. C. Genetic architecture of quantitative traits. *Ann. Rev. Genet.*, Palo Alto, v.35, n.1, 303-339, 2001.

MCINTYRE, L. M.; COFFMAN, C.; DOERGE, R. W. Detection and location of a single binary trait locus in experimental populations. *Genet. Res.*, London, v.78, p.79-92, 2000.

MURO-ABAD, J. I.; ROCHA, R. B.; CRUZ, C. D.; ARAÚJO, E. F. Crosses recommendation method for obtaining *Eucalyptus spp*. hybrids assisted by molecular markers. *Crop Breed. Appl. Biotechnol.*, Londrina, v.5, n.1, p.71-73, 2005.

NOVAES, E. *Mapeamento de QTLS para qualidade da madeira em Eucalyptus grandis x E. urophylla e ancoragem de clones BAC no mapa genético*. 2006. 171f. Thesis (Msc) – Universidade Federal de Viçosa, Viçosa, 2006.

RAO, S.; LI, X. Strategies for genetic mapping of quantitative traits. *Genetica*, Dordrecht, v.109, n.3, p.183-197, 2000.

ROCHA, R. B.; CRUZ, C. D.; ROSADO, A M.; BARROS, E. G.; ARAÚJO, E. F. Mapping of QTLS related with wood quality and development characteristics in hybrids (*Eucalyptus grandis x Eucalyptus urophylla*). *Rev. Árvore*, Viçosa, v.31, n.1, p.13-24, 2007.

ROCHA, R. B.; CRUZ, C. D.; BARROS, W. S.; FERREIRA, F. M.; ARAÚJO, E. F. Comparisons of segregating populations for genetic mapping. *Crop Breed. Appl. Biotechnol.*, Londrina, v.4, n.4, p.408 – 415, 2004.

SCHUSTER, I.; CRUZ, C. D. *Estatística genômica aplicada a populações derivadas de cruzamentos*. Viçosa: Editora UFV, 2004. 585p.

SILVA, L. C.; CRUZ C. D.; MOREIRA M. A.; BARROS E. G. Simulation of population size and genome saturation level for genetic mapping of recombinant inbred lines (RILs). *Genet. Mol. Biol.*, Ribeirão Preto, v.30, n.4, p.1101-1108, 2007.

VARSHNEY, R. K.; GRANER, A.; SORRELLS, M. E. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.*, Cambridge, v.23, n.1, p.48-55, 2005.

VISSCHER, P. M.; HALEY, C. S.; KNOTT, S. A. Mapping QTLs for binary traits in backcross and F_2 populations. *Genet. Res.,* London, v. 68, n.1, p.55-63, 1996.

XU, S.; ATCHLEY, W. R. Mapping quantitative trait loci for complex binary diseases using line crosses. *Genetics*, Baltimore, v.143, n.3, p.1417–1424, 1996.

XU S.; YONASH N.; VALLEJO R. L. Cheng H. H. Mapping quantitative trait loci for binary traits using a heterogeneous residual variance model: an application to Marek's disease susceptibility in chickens. *Genetica*, Dordrecht, v.104, n.2, p.171-178, 1998.

YI, N.; XU, S.; GEORGE, V.; ALLISON, D. B. Mapping multiple quantitative trait loci for ordinal trait. *Behav. Genet.*, New York, v.34, n.1, p.3-15, 2004.

YOUNG, N. D A cautiously optimistic vision for marker-assisted breeding. *Mol. Breed.*, Dordrecht, v.5, n.6, p.505-510, 1999.

YOUSEF, G. G.; JUVIK, J. A. Comparison of phenotypic and marker assisted selection for quantitative traits in sweet corn. *Crop Sci.*, Madison, v.41, n.3, p.645-655, 2001.

ZENG, Z. B. Precision mapping of quantitative trait loci. *Genetics*, Baltimore, v.136, n.4, p.1457-1461, 1994.

Received in 23.07.2008.

Approved after revised in 11.09.2008.