Crop Breeding and Applied Biotechnology 4:408-415, 2004
Brazilian Society of Plant Breeding. Printed in Brazil

Comparisons of segregating populations for genetic mapping

Rodrigo Barros Rocha[1], Cosme Damião Cruz[1*], Willian Silva Barros[1], Fabio Medeiros Ferreira[1], and Elza Fernandes de Araújo[1]

Received 16 January 2004

Accepted 7 October 2004

ABSTRACT - *The advent of molecular marker techniques allowed the construction of high informative genetic linkage maps for several vegetal species. Maximum likelihood has come into use for genetic mapping for estimation of several parameters, including the average information content (A.I.C.) and variance* $V(\hat{r})$ *of the recombination values "r". Our study presents the A.I.C. and* $V(\hat{r})$ *of various population types as additional criteria for the choice of the mapping population, since they define accuracy and trustworthiness of the estimated recombination frequencies. The populations with the most accurate recombination estimates are the* F_2 *population mapped with codominant markers and the RIL (recombinant inbred lines) populations for low recombination values. The employment of the A.I.C. and variance* $V(\hat{r})$ *of the recombination values may contribute to the quality of the molecular data and be an indicator for the cases where more robust analyses are necessary.*

Key words: Maximum likelihood, average information content, mapping populations.

INTRODUCTION

Only recently, linkage maps have been used to support the selection in breeding programs, even in well-studied species such as maize (*Zea mays*) or tomato (*Lycopersicum spp.*) (Yousef and Jurik 2001). Main obstacles that encumber the use of genetic maps in breeding programs are the kind of marker used for mapping and the lack of integrated maps that relate several marker types (Lander and Botstein 1989, Moreau et al. 2000).

The availability of neutral genetic markers, whose inheritance can be accompanied without influence of the environment, has led to the construction of molecular linkage maps for a number of plant species (Grattapaglia and Sederoff 1994, Harushima et al. 1998, Butruille et al. 1999). Due to its handiness and potential discrimination of innumerous markers, the molecular marker technique have boosted the importance of the genetic maps for assisted selection in breeding programs (Liu 1998).

Some authors consider that the genetic mapping for quantitative trait loci (QTL) detection and characterization is one of the more valuable applications of the molecular markers methodology integrated into breeding programs (Liu 1998). The trustworthiness of the distances values estimates and the right position of the markers along the linkage groups are directly related with the accuracy of the QTL´s localization and characterization.

[1]BIOAGRO, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brasil. *E-mail: cdcruz@ufv.br

The diversity of reproduction systems in vegetal species allows the use of different strategies for the development of mapping populations (Borem 1999). Autogamous species such as bean or soybean, for instance, allow the achievement of several population types for mapping: F_2 populations (Bunyamin et al. 2002), populations obtained by backcrossing (Matthews et al. 2001), or RIL (Recombinant Inbreed Lines) populations (Bouchez et al. 2002). Each population or progeny has its own distinctiveness, which must be taken into consideration by the researcher who determines the mapping population. Among the populations of widespread use for genetic mapping of plants, the most outstanding are F_2 populations derived from F_1 by coupling or repulsion (Harushima et al. 1998), populations obtained by backcrossing (Darvasi and Soller 1995), RIL (He et al. 2001, Bouchez et al. 2002), double-haploid (Weiguo et al. 2002, He et al. 2001), and F_1 populations of mixed segregation (Grattapaglia and Sederoff 1994).

Strategies that require few generations provide remarkable time gains for species with long life cycles, such as forest species (Grattapaglia et al. 1996). Grattapaglia and Sederoff (1994) suggest the use of F_1 populations derived from crosses of genetically contrasting genitors. This strategy makes use of the high genetic variability among genitors to obtain markers that are similar to those of a testcross. The mapping of the F_1 population can produce mixed segregation populations: 1:1, 3:1 if dominant molecular markers are used and 1:1, 1:2:1 for codominant molecular markers.

RIL populations are among the most commonly used progenies for the mapping of autogamous plants. Derived from F_2, they are advanced by self-polinization with the single seed descendent (SSD) method until they achieve a high homozygosity degree (Darvasi and Soller 1995). This is an appealing method, since it develops a permanent mapping population that allows continuous addition of markers into a preexisting map and also the implementation of more accurate experiments (Burr et al. 1998).

Double-haploid populations, developed more recently, are obtained artificially by duplication of haploid genomes from pollen grain and the achievement of diploid plants (Weiguo et al. 2002). Their advantages are, basically, similar to those of RIL populations (He et al. 2001), since they consist of homozygous individuals only.

When defining a mapping population, the average information content (A.I.C) and variance $V(\hat{r})$ associated to the recombination frequency estimates "r" should be considered as additional criteria, besides, time, cost limits and biological aspects of interest. That is especially important for species that permit the development of several types of mapping populations, allowing to the researcher to choose those that result in more precise maps with minor number of genotyped individuals.

Although other methodologies can be used to derivate the genetic linkage estimates: moments methods estimation (Darvasi and Weller 1992), Markov chain Monte Carlo method (Nielsen 2000, Laval et al. 2003), the maximum likelihood methods are the most used to genetic linkage analysis (Ritter et al. 1990, Luo and Kearsey 1991, Liu 1998). The maximum likelihood methodology allows the obtainment of consistent estimators of asymptotic efficiency and normal distribution, where the mean is given by the parameter and the variance by the inverse of the information content, considering a unique unknown parameter and large samples (Weir 1996).

Although these estimates are important indicators of the quality of information used in the construction of genetic maps, values of these for different mapping populations are not available in literature. Consequently, little use is made of these indices when choosing a population; Liu (1998) and Ritter et al. (1990) show the expressions for the F_2 and backcrossing populations.

This work presents the derivations and the estimates of the average information content and variance of the recombination values "r" for RIL, double-haploid and mixed segregation populations (1:1, 3:1 and 1:1, 1:2:1) and compares those values with other populations to, in a theoretical way, characterize the populations, and in a practical way, to aid in the a priori choice of the mapping populations.

METHODOLOGY

For large samples the variance of a maximum likehood estimator is (Weir 1996)

I) $$V(\hat{\theta}) \cong \frac{1}{E[I(\theta)]}$$

Where the information content is given by the negative second derivate of the maximum likehood function (Weir 1996, Liu 1998):

II) $$I(\theta) = -\left(\frac{\partial^2 L(\theta; x)}{\partial \theta}\right)$$

The obtainment of the information content estimators depends on the knowledge of the density probability function related with the independents observations x_1, x_2, ..., x_n. Considering the joint probability of these observations the likehood function is (Weir 1996, Liu 1998):

III) $$L(\theta; X) = p(X; \theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

Which is the maximum likehood function for the parameter or vector of parameter θ in function of the X values of vector of X values.

RIL Population

The density probability function associated with the genotypes distributions of the RIL population is obtained with the genotype frequency values at the F_6 generation:

Density probability function

IV) $$L(r;n_1) = \lambda\left[\frac{1}{2(1+2r)}\right]^{n1}\left[\frac{1}{2(1+2r)}\right]^{n2}\left[\frac{r}{(1+2r)}\right]^{n3}\left[\frac{r}{(1+2r)}\right]^{n4}$$

where,

V) $$\lambda = \frac{N!}{n_1!n_2!n_3!n_4!} \quad \text{and} \quad N = n_1 + n_2 + n_3 + n_4$$

Considering that the maximum likehood function and its logarithm have the same maximum point, the natural logarithm is usually applied in order to facilitate the derivation. The log likehood function can also be called of support function and its first derivate of score function.

Support Function

VI) $$\ln L(r;n_1) = \ln\lambda + (n_1 + n_2)\ln(1) + (n_3 + n_4)\ln r - N\ln(1+2r) - (n_1 + n_2)\ln 2$$

Score function

VII) $$\frac{\partial\ell(\theta;x)}{\partial\theta} = \frac{(n_3 + n_4)(1+2r) - 2Nr}{r(1+2r)}$$

Changing the n_3 and n_4 for their respective frequencies values and derivate again in function of the unknown parameter "r":

VIII) $$\frac{\partial^2\ell(\theta;x)}{\partial\theta} = \frac{-2}{r(1+2r)^2}$$

Where the average information content for the RIL population is given by

IX) $$I(\theta) = \frac{2}{r(1+2r)^2}$$,

and variance

X) $$V(\hat{\theta}) \cong \frac{r(1+2r)^2}{2}$$

Double-haploid population

The density probability, support and score functions are obtained by the same way showed for the RIL populations considering the genotypes frequency values

Density Probability Function

XI) $$L(r;n_1) = \lambda\left[\frac{1-r}{2}\right]^{n1+n2}\left[\frac{r}{2}\right]^{n3+n4}$$

Support Function

XII) $$\ln L(r;n_1) = \ln\lambda + (n_1 + n_2)\ln(1-r) + (n_3 + n_4)\ln r - N\ln 2$$

Score Function

XIII) $$\frac{\partial\ell(\theta;x)}{\partial\theta} = \frac{(n_3 + n_4)(1-r) - (n1 + n2)r}{r(1-r)}$$

Average Information Content (A.I.C.)

XIV) $$I(\theta) = -\left(\frac{\partial^2\ell(\theta;x)}{\partial\theta}\right) = \frac{1}{r(1-r)}$$

and variance,

XV) $$V(\hat{\theta}) \cong r(1-r)$$

Mixed population (1:1, 3:1).

Density Probability Function

XVI) $$L(r;n_1) = \lambda\left[\frac{2-r}{4}\right]^{n1}\left[\frac{1+r}{4}\right]^{n2}\left[\frac{r}{4}\right]^{n3}\left[\frac{1-r}{4}\right]^{n4}$$

Support Function

XVII) $$\ln L(r;n_1) = \ln\lambda + n1\ln(2-r) + n_2\ln(1+r) + n_3\ln r + n_4\ln(1-r) - N\ln 4$$

Score Function

XVIII) $$\frac{\partial\ell(\theta;x)}{\partial\theta} = \frac{Nr^3 - r^2(3n_2 + 2n_3 + n_4) - r(n_1 + n_3 + 2n_4 - 2n_2)}{(2-r)(1+r)r(1-r)}$$

Average Information Content (A.I.C.)

XIX) $$I(\theta) = -\left(\frac{\partial^2\ell(\theta;x)}{\partial\theta}\right) = \frac{r^2 - r - 0.5}{(2r - r^2)(1-r^2)}$$

and variance,

XX) $$V(\hat{\theta}) \cong \frac{(2r - r^2)(1-r^2)}{r^2 - r - 0.5}$$

Mixed population (1:2:1, 3:1)

Density Probability Function

XXI) $$L(r;n_i) = \lambda\left[\frac{(1-r)}{4}\right]^{n1+n2}\left[\frac{r}{4}\right]^{n3+n4}\left[\frac{1}{4}\right]^{n5+n6}$$

Support Function

XXII) $$\ln L(r;n_i) = \ln\lambda + (n_1 + n_2)\ln(1-r) + (n_3 + n_4)\ln(r) + (n_5 + n_6)\ln 1 - N\ln(4)$$

Score Function

XXIII) $$\frac{\partial\ell(\theta;x)}{\partial\theta} = \frac{-r(n_1 + n_2 + n_3 + n_4) + n_3 + n_4}{r(1-r)}$$

Average Information Content (A.I.C.)

XXIV) $$I(\theta) = -\left(\frac{\partial^2\ell(\theta;x)}{\partial\theta}\right) = \frac{1}{2r(1-r)}$$

and variance,

XXIV) $$V(\hat{\theta}) \cong 2r(1-r)$$

The information content values and the variances associated with the F_2 and backcrossing populations available at literature are showed at Table 1 (Ritter et al. 1990, Liu 1998).

The estimates of A.I.C. and $V(\hat{r})$ presented in this study were compared to those of other populations presented by Liu (1998). Figure 1 shows the A.I.C. distribution for the eight

evaluated mapping populations in function of the "r" values. The variance of the recombination fractions estimated using the inverse of expected A.I.C. is show in the Figure 2. To compare the populations, confidence intervals based on the normal distribution were also calculated (Table 1).

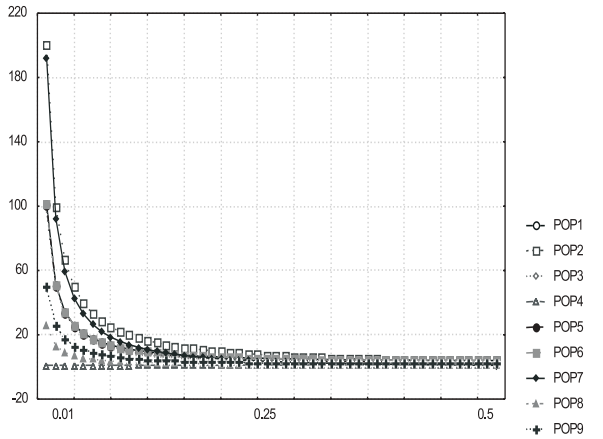

Figure 1. A.I.C. value distribution of the 9 evaluated populations for the recombination frequency "r" between 0.01 and 0.50

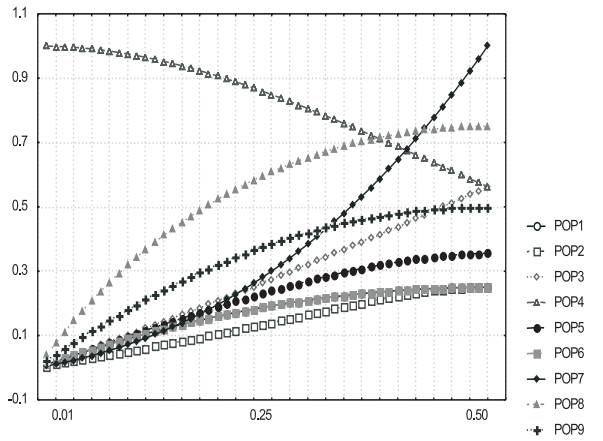

Figure 2. Variance value distribution of the 9 evaluated populations for the recombination value "r" between 0.01 and 0.50, using the inverse of expected A.I.C.

Table 1. Identification of the evaluated populations and the respective expression of the average information content (A.I.C.)

	Populations	A.I.C. Expression
POP 1	Backcrossing population, mapped with a dominant marker	$I(\hat{\theta}) \cong \frac{1}{r(1-r)}$
POP 2	F_2 population mapped with a codominant marker	$I(\hat{\theta}) \cong \frac{2(1-3r-3r^2)}{r(1-r)(1-2r+2r^2)}$
POP 3	F_2 population mapped with a dominant marker in coupling linkage phase	$I(\hat{\theta}) \cong \frac{2(3-4r+2r^2)}{r(2-r)(3-2r+r^2)}$
POP 4	F_2 population mapped with a dominant marker in repulsion linkage phase	$I(\hat{\theta}) \cong \frac{2(1+2r^2)}{(2+r^2)(1-r^2)}$
POP 5	F_2 population mapped with codominant markers and dominant markers	$I(\hat{\theta}) \cong \frac{2+(1-r)^2}{r(2-r)} + \frac{(1-r)^2}{2(1-r-r^2)} + \frac{r^2}{1-r^2} + \frac{(1-2r)^2}{2r(1-r)}$
POP 6	Double-haploid population	$I(\hat{\theta}) \cong \frac{1}{r(1-r)}$
POP 7	RIL population	$I(\hat{\theta}) \cong \frac{2}{r(1+2r)^2}$
POP 8	F_1 population mapped with a dominant marker of mixed segregation in (1:1, 3:1)	$I(\theta) = \frac{r^2-r-0.5}{(2r-r^2)(1-r^2)}$
POP 9	F_1 population mapped with codominant and dominant markers of mixed segregation in (1:2:1, 3:1)	$I(\hat{\theta}) \cong \frac{1}{2r(1-r)}$

RESULTS AND DISCUSSION

The statistical properties of an estimator are critical for interpreting estimates of recombination fractions. The A.I.C. and variance values associated with the recombination frequencies "r" allow the construction of confidence intervals based in the normal distribution for each population (Table 2). Studies on confidence intervals reveal that different population sizes are necessary for a same interval, depending on the analyzed progeny.

In order to compare populations in relation to the necessary number of individuals that obtain a same confidence interval, we established the values of "r" (0.05, 0.20 and 0.30) and probability (90%) (Table 2). Smaller confidence intervals than six map units are not expected when little over one hundred individuals are evaluated. Maps less accurate are obtained when the map saturation sinks and for the populations of mixed segregation and F_2 mapped by dominant markers in repulsion linkage phase; in general these require a higher number of individuals to map with the same precision of others populations.

As additional criteria to appraise the trustworthiness of genetic maps, confidence intervals should be calculated considering the number of individuals genotyped and the mean "r" value. Higher intervals values indicate that the position of some markers could be changed along the linkage groups. Errors in the markers order may result in problems to the QTL detection and positioning.

The evaluation by confidence intervals is interesting because they can be calculated considering different classes of segregating markers: coupling or repulsion, 3:1 or 1:1, or even considering separated linkage groups. In this way linkage groups, or regions inside these, more accurate than others may be detected. Considering that the QTL´s study seeks the identification of few markers related with the phenotype expression, this methodology might be especially useful in the characterization of the most accurate positions inside the linkage group.

F_2 populations, usually developed by selfing F_1 individuals derived from divergent crossings, are the most commonly used mapping populations. The reason for this is that they are easy to obtain and take part in most breeding programs, giving twice as many useful markers than a backcross population for the same number of DNA extractions and PCR assays. They also allow the estimation of both additive and dominant QTL effects. The major drawback of using F_2 and also backcross populations is that these populations cannot be indefinitely maintained, limiting more precise field tests and continuous mapping saturation. Vegetative propagations or large scale cloning techniques could be used to preserve the individuals of F_2 populations.

F_2 population mapped with codominant markers is the mapping population that presents the highest A.I.C. for any "r" value. As expected the discrimination of the heterozygotic

Table 2. Required population size for the construction of confidence intervals at a probability of 90% and "r" values of 0.05, 0.20 and 0.30. The inferior and superior limits of the confidence intervals are indicated

Confidence Intervals	$IC(r)_{1-\alpha} : \hat{r} - 0.01 \le r \le r + 0.01$			$IC(r)_{1-\alpha} : \hat{r} - 0.02 \le r \le r + 0.02$			$IC(r)_{1-\alpha} : \hat{r} - 0.03 \le r \le r + 0.03$			$IC(r)_{1-\alpha} : \hat{r} - 0.04 \le r \le r + 0.0$		
"r" values	**0.05**	**0.20**	**0.30**	**0.05**	**0.20**	**0.30**	**0.05**	**0.20**	**0.30**	**0.05**	**0.20**	**0.30**
POP 1	1285	4330	5683	321	1082	1421	143	481	631	80	271	355
POP 2	678	2831	4454	170	708	1113	75	315	495	42	177	278
POP 3	1365	5640	8678	341	1410	2169	152	627	964	85	352	542
POP 4	26892	24535	21808	6723	6134	5452	2988	2726	2423	1681	1533	1363
POP 5	1339	5223	7576	335	1306	1894	149	580	842	84	326	474
POP 6	1286	4330	5683	321	1082	1421	143	481	631	80	271	355
POP 7	819	5303	6494	205	1328	1623	91	589	721	51	331	405
POP 8	4807	14170	17688	1202	3542	4422	534	1574	1965	300	886	1105
POP 9	2570	8659	11365	642	2164	2841	286	962	1262	161	541	710

genotype by the codominant marker techniques provides a higher accuracy of the "r" estimates for this population.

The use of dominant markers in F_2 populations results in the detection of the two types of markers segregating: the couple and repulsion linkage phase markers. The increment in the variance values result from the repulsion configuration of the markers in the genitors is greater than that sourced from the use of dominant marker techniques. As it is well known, linkage statistics and recombination frequencies estimates using dominant markers are not properly estimated from repulsion F_2 matings. The mapping precision, in this case, is directly related with the proportion of the evaluated couple phase markers that should be the majority of markers used for mapping.

A study strategy with dominant markers that results in higher A.I.C. values are provided by the elimination of the heterozygotes from the population via autofecundation (as in RIL populations) or by artificial duplication of the haploid genome (as in double-haploid populations). Of course in these cases, the use of a codominant marker does not increase the "r" estimates accuracy.

The RIL population presents one of the lowest $V_{(\hat{r})}$ values when small "r" values are considered. However, as the map saturation sinks (r = 0.20), the RIL variance comes close to values of other populations: double-haploid, F_2 mapped with dominant markers in coupling linkage phase, F_2 mapped with dominant and codominant markers, backcross, and exceeds them when higher "r" values are considered ($r \geq 0.25$). When values are $r \geq 0.44$, the variance of the RIL population outstrips that of all other populations, including F_2 populations mapped with dominant marker in repulsion linkage phase and mixed segregation populations.

In RIL populations, the occurrence of several linkage events for the developing of the pure lines results in linkage equilibrium for most markers with higher "r" values making the development of less saturated maps difficult. Therefore, RIL populations are only appropriate for mapping in the construction of saturated maps, which the mean distance values are lower than 0.20. Burr et al. (1988) listed some advantages of RIL populations, highlighting the fact that RIL families constitute permanent populations in which the complete segregation allows more precise field tests and mapping of closely linked markers.

Different to RIL, double-haploid populations present one of the lowest A.I.C. variations as the map saturation decreases, although they also consist of homozygotic genotypes only. The artificial duplication of haploid genomes from pollen grains results in genotypes derived from a unique gamete, turning this population more appropriated to the development of less saturated maps. However, the occurrence of only one recombination event makes the mapping of closely linked markers difficult.

The A.I.C. distribution of double-haploid populations is identical to that of backcross populations mapped with dominant markers. These populations are usually good for mapping with any "r" value. The use of codominant molecular markers in backcross populations do not result in more accurate estimates of "r", considering that for this population, the dominant markers allow the identification of the two classes of segregating genotypes for mapping.

The population of mixed segregation presents one of the highest variance values for any "r" value. Only F_2 population mapped with a dominant marker in repulsion linkage phase for the interval of $r \leq 0.40$ is higher, nevertheless the mapping of mixed segregation populations using codominant markers results in more accurate "r" estimates. For these populations, and also for the F_2 populations mapped with codominant and dominant markers the genitors linkage phase doesn't change the A.I.C. expressions.

Population F_2 derived from F_1 by repulsion has an abnormal behavior with diminishing variance values as the map saturation decreases. This population is generally not recommended for mapping; besides the abnormal behavior, its variance alterations are small when the map saturation increases [$_{1(\theta)}$ range between 1.0 and 1.7]. To minimize this problem, while screening the genotypes the markers in couple linkage phase should be the majority of the total of markers evaluated.

Studies and recent theoretical developments have been allowing the overcome of many inherent limitations of the mapping procedure, as problems of lost data (Laval et al. 2003), unknowing of the linkage phase of the genitors (Lin et al. 2003), impossibility of obtaining more precise mapping populations, and others. Species of the genus Eucalyptus and Pinus for instance, their long life cycle and high genetic load hinder the development of populations that need more generations to develop. For these cases, where the researcher cannot choose others mapping populations, the number of evaluated individuals should be planned considering the molecular data quality wanted. However many crops allows the development of several mapping populations, that should be mapped considering also the precision of the "r" estimates.

Besides the applied molecular marker, the number of genotypes evaluated and the linkage phase of the markers; the genotype frequencies and the map saturation are determinant for the accuracy of the "r" estimates. The mean map saturation desired should be considered while choosing the mapping population and how many individuals to genotype, once the map saturation also increases the precision of the "r" estimates.

CONCLUSIONS

The definition of the mapping population and the number of individuals genotyped also determines the available

information quality for the genome analysis that affects the accuracy of recombinant frequency estimates, indispensable for analysis and detection of the QTL´s. If the biology of a species, as well as practical aspects of the process, allow the achievement of different segregation populations, A.I.C. and variance of recombinant frequency values must be taken into consideration.

Comparações entre populações segregantes para mapeamento genético

RESUMO - *Marcadores moleculares têm permitido a construção de mapas genéticos de ligação altamente informativos para várias espécies vegetais. A metodologia de máxima verossimilhança vem sendo utilizada no mapeamento genético para estimação de vários parâmetros, incluindo o conteúdo médio de informação (c.m.i) e a variância $V(\hat{r})$ dos valores de freqüência de recombinação "r". Nosso estudo apresenta o c.m.i e a $V(\hat{r})$ associados a vários tipos de populações como critério adicional a ser considerado na escolha da população para mapeamento, uma vez que definem a precisão e a confiabilidade das freqüências de recombinação estimadas. As populações com as estimativas de recombinação mais precisas são a F_2 mapeadas com marcadores codominantes e a RIL (linhagens recombinantes) para baixos valores de recombinação. O emprego do c.m.i e da $V(\hat{r})$ dos valores de freqüência de recombinação pode contribuir para a qualidade dos dados moleculares e ser um indicador para os casos onde análises mais robustas são necessárias.*

Palavras-chave: máxima verossimilhança, conteúdo médio de informação, populações de mapeamento.

REFERENCES

Borém A (1999) **Melhoramento de espécies cultivadas**. UFV, Viçosa, 817p.

Bouchez A, Hospital F, Causse M, Gallais A and Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. **Genetics 162**: 1945-1959.

Butruille DV, Guries RP and Osborn T (1999) Linkage analysis of molecular markers and quantitative trait loci in populations of inbred backcross lines of *Brassica napus* L. **Genetics 153**: 949-964.

Bunyamin T, Thomas EM and Pauls KP (2002) Genetic mapping of agronomic traits in common bean. **Crop Science 42**: 544-556.

Burr B, Burr FA, Thompson KH, Albertson MC and Stuber CW (1998) Gene mapping with recombinant inbreds in maize. **Genetics 118**: 519-526.

Darvasi A and Weller JI (1992) On the use of the moments method of estimation to obtain approximate maximum likehood estimates of linkage between a genetic marker and a quantitative locus. **Heredity 68**: 43-46.

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

Grattapaglia D and Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla u*sing a pseudo-testcross: mapping strategy and RAPD markers. **Genetics 137**: 1121-1137.

Grattapaglia D, Fernando LGB, Penchel R and Sederoff RR (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. **Genetics 144**: 1205-1214.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Baltazar AA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS and Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. **Genetics 148**: 479-494.

He P, Li JZ, Zheng XW, Shen LS, Lu CF, Chen Y and Zhu LH (2001) Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. **Crop Science 41**: 1240-1246.

Lander ES and Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics 121**: 185-199.

Laval G, SanCristobal M and Chevalet C (2003) Maximum-likelihood and Markov chain monte carlo approaches to estimate inbreeding and effective size from allele frequency changes. **Genetics 164**: 1189-1204.

Lin M, Lou XY, Chang M and Wu R (2003) A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. **Genetics 165**: 901-913.

Liu BH (1998) **Statistical genomics. Linkage, mapping and QTL analysis**. CRC Press, Boca Raton, 605p.

Luo ZW and Kearsey MJ (1991) Maximum likehood estimation of linkage between a marker gene and a quantitative trait locus. II. Application to backcross and double haploid populations. **Heredity 66**: 117-124.

Matthews FB, Devine TE, Weisemann JM, Beard HS, Lewers KS, MacDonald MH, Park Y, Maiti R, Lin J, Kuo J, Pedroni MJ, Cregan PB and Saunders JA (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. **Crop Science 41**: 516-521.

Moreau L, Lemarie S, Charcosset A and Gallais A (2000) Economic efficiency of one cycle marker-assisted selection. **Crop Science 40**: 329-337.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. **Genetics 154**: 931-942.

Ritter E, Gebhardt C and Salamini F. (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. **Genetics 125**: 645-654.

Weiguo L, Ming YZ, Enrique AP and Calvin FK (2002) Highly efficient doubled-haploid production in wheat (*Triticum aestivum* L.) via induced microspore embryogenesis. **Crop Science 42**: 682-692.

Weir BS (1996) **Genetic data analysis II**. Sinauer Associates, Sunderland, 445p.

Yousef GG and Jurik JA (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. **Crop Science 41**: 645-655.