Crop Breeding and Applied Biotechnology 5:310-317, 2005
Brazilian Society of Plant Breeding. Printed in Brazil

QTL mapping of Poisson traits: a simulation study

Alex de Oliveira Ribeiro[1*], Eduardo Bearzoti[2], and Thelma Sáfadi[2]

Received 6 June 2005

Accepted 30 July 2005

ABSTRACT - *QTL mapping consists of estimating the position and effects of genes (or linkage groups) that control quantitative traits. For traits that follow a normal distribution of probability, statistical approaches for QTL mapping are described in literature. However, for other traits the amount of research is little. The present work aimed to evaluate four alternative methods for QTL mapping of genes that control Poisson traits. The studied approaches were: standard interval mapping methodologies of Lander and Botstein and of Haley and Knott, a mixed model based on Poisson distributions, and a generalized linear model. Through computational simulation, the results showed great similarity for the different approaches under study. All approaches were highly effective for the tested simulations, especially when the abnormality of data in use was taken into account. Owing to the computational simplicity, Harley and Knott's methodology was considered most effective.*

Key words: QTL mapping, Poisson distribution, Simulation

INTRODUCTION

Most statistical methods for mapping quantitative trait loci (QTL) are based on models which assume normal distribution (Lander and Botstein 1989, Halley and Knott 1992, Jansen 1993, Zeng 1994, Jiang and Zeng 1995, Kao et al. 1999). In plant breeding however there are discrete traits of interest for which such an assumption does not hold, such as counts of number of ears or tillers, which may be modeled by a Poisson distribution, or disease status, scored as absent or present. As Jansen (1992) states, consideration of actual distribution should lead to an increase in statistical power, in relation to normality-based methods. The author developed a mixture model for mapping QTL with a general distribution, and verified the superiority of maximum likelihood if residuals are outcome of an exponential distribution. Visscher et al. (1996) developed QTL mapping for binary traits using a generalized linear model (McCullagh and Nelder 1989) and linear regression approach (Haley and Knott 1992). The methods were evaluated by stochastic simulation in a wide range of conditions. Very similar results were verified with regard to QTL position and effects and detection power. Consequently, the authors suggested that, owing to its

[1] Departamento de Matemática, Centro Universitário de Lavras, UNILAVRAS, Rua Padre José Poggel, 506, Centenário, 37.200-000, Lavras, MG, Brasil. *E-mail: analise@uai.com.br
[2] Departamento de Ciências Exatas, Universidade Federal de Lavras, UFLA, C. P. 3037, 37.200-000, Lavras, MG, Brasil.

simplicity and availability in standard software, simple linear regression as proposed by Haley and Knott (1992) would be more efficient for QTL mapping for binary traits. Another type of discrete data occurs when responses are categories of an ordinal scale, such as disease severity. In such cases, the proportional odds model, a kind of generalized linear model, can be used (Spyrides-Cunha et al. 2000).

For QTL mapping, it would be useful to verify whether normality-based methods as linear regression also show robustness with other discrete data, such as counts following a Poisson distribution. Linear regression does not use all information about trait distribution, even in presence of normality. The reason is that the distribution of phenotypes, given a marker genotype, is actually a mixture of distributions corresponding to the different QTL genotypes. Linear regression is essentially based on only the expectation of such distribution, while the method of maximum likelihood (Lander and Botstein 1989) takes the fact that it is a mixture into account. Nevertheless, with normality very similar results are found with both methods (Haley and Knott 1992). If the trait distribution is not normal, a general mixture model can be constructed for QTL mapping (Lynch and Walsh 1998), and the parameters estimated by maximum likelihood. Alternatively, a generalized linear model considering the actual distribution could be used, with a linear predictor corresponding to the expectation of that mixture. The linear regression approach of Haley and Knott (1992) is therefore a particular case of this model, considering identity as link function. Again, there is some loss of information in this generalized linear model, but it might also lead to no meaningful decrease of precision or QTL detection power.

The objective of this work was to investigate four different approaches for QTL mapping for Poisson traits using computer simulation. Two mixture models were used in association with maximum likelihood, assuming Poisson and normal distribution, respectively. The latter method corresponded to that of Lander and Botstein (1989). The generalized linear model mentioned above was also considered, in particular the Poisson distribution and the standard linear regression proposed by Haley and Knott (1992).

MATERIAL AND METHODS

Models for QTL mapping for Poisson traits

The theory of threshold characters (Falconer and Mackay 1996) was basis of the models developed by Visscher et al. (1996). Falconer and Mackay (1996) consider the cases of two (binary traits) and three classes (two thresholds). The generalization to an infinite number of classes, which is the case of a Poisson distribution, is not straightforward. Instead, we chose to use the general mixture model (Lynch and Walsh 1998) which is typical in QTL analysis, considering a mixture of two-parameter distributions from the exponential family. According to McCullagh and Nelder (1989), a probability density function belongs to the exponential family of distribution if it can be written under a general formula:

$$f(y;\theta,\phi) = \exp\left\{\frac{1}{a(\phi)}[y\theta - b(\theta)] + c(y;\phi)\right\}$$

We assume that genotypes of each QTL determine the value of the main parameter (denoted α), but the QTL does not affect the nuisance parameter ϕ. The number of distributions depends on the number of QTL genotypes in the population (e.g. two or three in a backcross or F_2 population, respectively). The weights or proportions of the mixture are determined by the probabilities of each QTL genotype, given the genotype of two adjacent and codominant marker loci flanking an interval where there may be a QTL. The distance between such markers is assumed to be known, obtained from a previously constructed molecular map. The following considerations refer to a backcross population, but generalization to other population types is straightforward.

Suppose plants from an inbred line 1 with QTL genotype designated *QQ* are crossed with plants from line 2 with QTL genotype *qq*, originating F_1 plants with genotype *Qq*. Such plants are backcrossed with line 1, and descendants are *QQ* or *Qq*. We further suppose that such genotypes determine values of q equal to q_Q and q_q, respectively. There is a fixed f value, regardless of the QTL genotype. The probability distribution of individuals of a given marker genotype is a mixture of two distributions, correspondent to genotypes *QQ* and *Qq*, with weights equal to the conditional probabilities of such QTL genotypes, on marker genotype. These probabilities are presented in Table 1 for a backcross. In this Table, markers 1 and 2 flanking an interval are represented each one by upper or lower case letters, referring to lines 1 and 2, respectively. For any position of the putative QTL in the interval, maximum likelihood expressions can be obtained to solve for functions of q_Q, q_q and f. These are not closed form expressions, but call for an iterative procedure, which is generally referred to as "expectation-conditional maximization (ECM) algorithm" (Meng and Rubin 1993, Zeng 1994). Across the marker interval, one should keep the estimates corresponding to the position that maximizes

likelihood. In the case of normal distribution, when convergence is attained, this process leads to the maximum likelihood estimates of Lander and Botstein.

In the case of Poisson distribution there is a single parameter, the mean, which is assumed to be affected by the putative QTL on the interval. Within each QTL genotype there is some variability due to environment effect, with variance equal to the mean. Despite the simplicity of this mixture model, this dependence of environment effect upon genotype may seem tantalizing, since it is rather unusual in quantitative genetics. This dependence is not unfounded. In the theory of threshold characters with two classes, for example, there is no such dependence in the liability scale, but it arises in the response scale if classes are designated by values 0 and 1, due to the different probabilities of value 1.

In either case (Normal or Poisson), detection of the putative QTL may be based on the likelihood ratio test statistic, or equivalently on the logarithm of the odds ratio (Lynch and Walsh 1998), which are generally displayed graphically across the marker interval. The overall significance level should be controlled, considering the whole set of intervals investigated, and decision rule could be based on the chi-square approximation or permutation tests (Doerge and Churchill 1996).

The regression model (generalized or not) for QTL mapping uses the following linear predictor:

$b_0 \quad b^*X^*$

where b_0 is the intercept, b^* accounts for the QTL effect, and X^* is equal to 1 if QTL genotype is *QQ*, and 0 if it is *Qq*. Since QTL genotype is unknown, h will refer to the expectation conditioned on marker genotype for actual data, and so X^* of individual i will be:

$X_i^* = 1 \times P(QQ \mid M_i) + 0 \times P(Qq \mid M_i) = P(QQ \mid M_i)$

where *P* is the probability of a given QTL genotype, given the marker genotype M_i of individual i. To every QTL position considered in the interval, a regression model is fitted with different values of X_i^* of each individual, since *P* depends on the recombination fractions between QTL and each flanking marker, and consequently on QTL position. Under normal distribution, is directly the expectation of Y_i, the phenotype of individual i. This approach corresponds basically to the model of Haley and Knott (1992). For non-normal distributions, is related to Y_i expectation through an appropriate link function (McCullagh and Nelder 1989). For Poisson it is usual to consider the natural log function. If $_i$ designates Y_i expectation, then $\ln \; _i \quad b_0 \quad b^*X^*$ and $_i \quad e^{b_0 \; b^*X^*}$

Table 1. Conditional probabilities of QTL genotypes given the genotype of flanking markers in an interval, considering a backcross population and assuming no interference*

	Marker genotype			
QTL genotype	M_1M_2/M_1M_2	M_1m_2/M_1M_2	m_1M_2/M_1M_2	m_1m_2/M_1M_2
QQ	$\frac{1-r_1-r_2+r_1r_2}{1-r_1-r_2+2r_1r_2}$	$\frac{r_2-r_1r_2}{r_1+r_2-2r_1r_2}$	$\frac{r_1-r_1r_2}{r_1+r_2-2r_1r_2}$	$\frac{r_1r_2}{1-r_1-r_2+2r_1r_2}$
Qq	$\frac{r_1r_2}{1-r_1-r_2+2r_1r_2}$	$\frac{r_1-r_1r_2}{r_1+r_2-2r_1r_2}$	$\frac{r_2-r_1r_2}{r_1+r_2-2r_1r_2}$	$\frac{1-r_1-r_2+r_1r_2}{1-r_1-r_2+2r_1r_2}$

* r_1, r_2: recombination fractions between QTL and marker locus 1 and 2, respectively

As usual in the generalized linear models theory, parameters b_0 and b^* can be estimated by maximum likelihood and the hypothesis that b^* is equal to zero tested by the likelihood ratio test using chi-square approximation (the so-called analysis of deviance). Rejection of such hypothesis would indicate the presence of a QTL on that position.

Data simulation

Individuals from backcross populations were simulated, receiving correspondent Boolean values "true" or "false" to mimic QTL and marker loci, according to the genotypes (that of the recurrent parent or F_1 population, respectively). Without loss of generality, a single interval was considered, flanked by two markers. Interval length was equal to 0.1 or 0.5 Morgan, to simulate more or less saturated molecular maps. Crossovers occurred according to Haldane's mapping function. Sample sizes corresponded to 50 or 200 individuals. Values of a fictitious discrete trait were generated for each individual, by randomly picking outcomes from Poisson distributions. In a given backcross, trait mean was equal to 1 or 10. The choice of such values aimed at the construction of distributions farther from or closer to normality, respectively, which could affect the similarity among the results of the mapping methods. A single interval was simulated considering the presence or absence of QTL. If there were no QTL on the segment, the

trait outcome of each individual of the same backcross was based on the same Poisson distribution. If a QTL were present, the Poisson parameter depended on the QTL genotype of the individual, but in a way that the backcross mean was always equal to 1 or 10. Designating the Poisson parameter of each QTL genotype as l_{QQ} and l_{Qq}, they can be expressed in terms of genetic effects as:

$_{QQ} = m + a$ $_{Qq} = m$

where *m* is the mean of the individuals *Qq*, and *a* is the difference between the means of the individuals of both QTL genotypes. Since the expected frequencies in all such genotypic classes are the same, the backcross mean is:

$$\bar{\lambda} = m + \frac{a}{2}$$

Parameter can also be interpreted as mean variance within QTL genotypic classes. If present on the segment, a QTL was responsible for explaining 0.02 or 0.2 of the proportion of phenotypic variation. Such magnitudes were chosen to consider a QTL of minor or major effect. If this proportion is represented by *p*, then:

$$p = \frac{V_g}{V_p}$$

where V_g is the genetic variance due to that QTL and equals to $a^2/4$, the correspondent value for a backcross. V_p, the phenotypic variance, is the sum of V_g and $^{-}$. From this expression it follows that, for fixed values of *p* and $^{-}$, it is possible to solve for *a* through:

$$a = 2\sqrt{\frac{p}{1-p}\bar{\lambda}}$$

and substituting it in the formula of and solving for *m*:

$$m = \bar{\lambda} - \frac{1}{2}a$$

With such values, the resulting parameters $_{QQ}$ and $_{Qq}$ were used to generate the trait values of the individuals according to the QTL genotype. It is interesting to note that, since $_{Qq}$ is necessarily non-negative for given and *p*, there is a maximum value allowed for *a*, equal to 2^{-}. The values used in this study did not violate the constraint of non-negativity.

Considering all combinations of different conditions regarding interval length, sample size, $^{-}$, QTL presence and *p* (if a QTL is present), 24 situations were studied. For each situation, 5000 stochastic simulations (independent random samples) were carried out. A same simulated data set was analyzed according to the four mapping methods mentioned previously: the mixture of Poisson distributions (PM), Lander and Botstein's mixture of normal distributions (LB), generalized linear regression (GLR), and Haley and Knott's standard linear regression (HK). The mean results across simulations allowed the evaluation of the detection power of QTL and estimation of bias and mean absolute error (MAE), which is a measure of precision. Under the null hypothesis (no QTL), actual type I error rates were comparable to the nominal significance level (0.05).

RESULTS AND DISCUSSION

Table 2 shows mean results of 5,000 simulations considering a sample size of 200 individuals, an interval length of 0.5 Morgan between markers and population mean $^{-}$ 1. One notes that under null hypothesis (no QTL), the actual type I error rates were below significance level (0.05) for all methods. The reason is that the patterns of the test statistic were intermediate between those of chi-square distributions with 1 and 2 degrees of freedom, with regard to 95 percentiles and distribution shape (data not shown). This agreed with previous results of Zeng (1994). Using simulation, the author verified that the statistical test distribution is similar to that of a chi-square for a fixed position in the interval, but it deviates towards the intermediate pattern when QTL position is estimated. Therefore, using chi-square distribution with an additional degree of freedom to take QTL position into account seems suitable, yet a little conservative. Anyway, type I error rates of the four methods were quite similar, ranging from 0.032 to 0.035. Comparing the mean estimates of *m* and *a* under null hypothesis to the corresponding parametric values (Table 2), no considerable biases can be seen. Among mapping methods, the mean estimates as well as MAE values were very close (those of Haley and Knott's method were actually a little lower), indicating that all methods have similar precision.

With null hypothesis false (a QTL present on the segment), detection power was strongly affected by the magnitude of QTL effect (Table 2). A QTL that explains 0.2 of the phenotypic variation will almost surely be detected under the considered conditions, while a QTL with $p = 0.02$ will have a chance to be detected close to 0.2 (20%). No appreciable differences in detection power were observed among all methods. The estimate of *m*, *a* and QTL position at $p = 0.02$ had mean values close to parametric values (Table 2). With $p = 0.2$, larger deviations were observed for methods GLR and LB, considering parameter *m*. Estimates of *a* with LB method showed a stronger downward bias. Mean QTL position estimates were similar, but the HK method tended to be more precise at $p = 0.02$.

Simulations under the same conditions of Table 2 but with $\bar{}$ 10 showed quite similar trends and are therefore not shown here. Rejection rates of null hypothesis (true or not) were very close to those with $\bar{}$ = 1, yet slightly higher. The same general tendency of similarity across methods was observed for detection power, bias and precision. This was not surprising, since with the increase of trait distribution approaches normality. The noticeable point is the occurrence of this similarity with low, that is, in cases where the trait nature is clearly discrete.

Table 3 shows identical conditions to those of Table 2 (sample size of 200 individuals, population mean $\bar{}$ 1), but with an interval length of 0.1 Morgan between markers. Under null hypothesis, the actual type I error rates were slightly inferior to those of Table 2. The reason for this may be that, with a smaller interval, the test would be more similar to that of fixed position, for which statistic test distribution is similar to a chi-square with 1 degree of freedom for backcrosses, according to Zeng (1994).

Table 2. Simulation results of four QTL mapping methods for a Poisson trait: mixture of Poisson (MP), generalized linear regression (GLR), method of Lander and Botstein (LB) and of Haley and Knott (HK). Values are means of 5000 runs, considering a backcross with mean = 1, sample size of 200 individuals and length interval of 0.5 Morgan between markers*

		p = 0			*p* = 0.02				*p* = 0.20			
Method		Rate of H_0 rejection	*m*	*a*	Rate of H_0 rejection	*m*	*a*	QTL Position†	Rate of H_0 rejection	*m*	*a*	QTL Position
	Nominal Value	0.05	1	0	1	0.8571	0.2857	0.25	1	0.5	1	0.25
PM	Mean	0.032	1.0011	-0.0034	0.226	0.8601	0.2791	0.2525	0.998	0.5081	0.9857	0.2497
	MAE	-	0.1016	0.1763	-	0.0858	0.1439	0.1629	-	0.0707	0.1399	0.0615
GLR	Mean	0.032	1.0022	-0.0033	0.225	0.8625	0.2834	0.2526	0.998	0.5547	1.0050	0.2498
	MAE	-	0.1018	0.1770	-	0.0842	0.1478	0.1587	-	0.0813	0.1646	0.0635
LB	Mean	0.034	1.0012	-0.0035	0.217	0.8602	0.2793	0.2521	0.996	0.5355	0.9486	0.2501
	MAE	-	0.1017	0.1767	-	0.0853	0.1437	0.1618	-	0.0734	0.1477	0.0705
HK	Mean	0.035	1.0011	-0.0031	0.219	0.8642	0.2831	0.2521	0.996	0.5029	0.9958	0.2498
	MAE	-	0.0950	0.1326	-	0.0923	0.1598	0.1192	-	0.0826	0.1604	0.0616

*p: proportion of phenotypic variance explained by QTL, H_0: null hypothesis (in the QTL), m: genotypic value of individuals Qq, a: difference between genotypic values of individuals QQ and Qq, MAE: mean absolute error. † Expressed in Morgan

The mean estimates of *m* and *a* under null hypothesis were again very close to parametric values, and they were similar among mapping methods, as well as MAE values, except for HK method, which was a little more precise. If the null hypothesis is false, there is an increase in detection power, if rates are compared to those of Table 2. This was a reflex of the fact that smaller intervals (dense maps) correspond to markers closer to putative QTL, leading to an increase in power, given large enough sample sizes for the occurrence of recombination. With p = 0.2, null hypothesis was rejected throughout in 5000 simulations by all methods. Again, detection power was strongly affected by QTL effect. Rejection rates of null hypothesis decreased to about 0.35 with p = 0.02. With a QTL of minor or major effect, estimates of *m*, *a* and QTL position were very similar among all methods, this was also observed regarding MAE. Since parameters *m* and *a* have the same real values in Tables 2 and 3, it is interesting to note that MAE values are lower in Table 3, indicating that the smaller interval allowed more precise estimates. Again, results with $\bar{}$ 10 showed very similar trends to those with $\bar{}$ 1, and are not shown here.

Tables 4 and 5 are analogous to Tables 2 and 3, respectively, but consider a sample size of 50 individuals. With regard to (true) null hypothesis rejection, it can be seen that rates were close under different sample sizes, especially for methods LB and HK. However, with false null hypothesis, QTL detection power was considerably decreased, especially when considering a QTL of minor effect (about 0.09 in both interval lengths). With p = 0.2, rates decreased more than 0.3 (Table 4) and 0.25 (Table 5) with the longer and the smaller interval, respectively. Under the former condition (Tables 2 and 3), methods PM and GLR were somewhat more powerful than LB and HK. Reduced sample sizes also affected precision. Comparing MAE values of Tables 2 and 3 to the corresponding ones in Tables 4 and 5 for parameters *m*, *a* and QTL position, it can be seen that they were higher throughout, sometimes more than doubled. Anyway, MAE values and mean estimates of such parameters were generally very similar across methods. For parameter *m*, methods PM and HK showed lower bias at p = 0.2 in Table 4 (larger interval), but they were a little less precise based on MAE. The QTL position was estimated with more precision by the HK method throughout.

Situations investigated in this study considered only

one population type, the backcross, which does not allow the discrimination of additive from dominance effects. In the strict sense, it is not possible to generalize the present results for other population types. However, Visscher et al. (1996) found very close trends considering backcross and F_2 populations for binary traits, in the sense that the normality-based method (HK) yielded quite similar results to those of the generalized linear model. It is therefore possible that this could be the case for the discrete distribution considered in this study.

This study also considered only one single interval investigated at a time. Although this implies in no loss of generality, in practice one would ideally remove QTL effects outside a given interval being tested. The aim is to remove genetic variation from residuals, increasing power and precision, and also to eventually eliminate biases caused by other QTL in the same linkage group. This idea led to the approach of combining multiple regression with interval mapping (Jansen 1992, Zeng 1994), the so-called composite interval mapping. The selection of regression

Table 3. Simulation results of four QTL mapping methods for a Poisson trait: mixture of Poisson (MP), generalized linear regression (GLR), method of Lander and Botstein (LB) and of Haley and Knott (HK). Values are means of 5000 runs, considering a backcross with mean = 1, sample size of 200 individuals and length interval of 0.1 Morgan between markers*

		p = 0			*p* = 0.02				*p* = 0.20			
		Rate of H_0 rejection	*m*	*a*	Rate of H_0 rejection	*m*	*a*	QTL Position†	Rate of H_0 rejection	*m*	*a*	QTL Position
Method	Nominal Value	0.05	1	0	1	0.8571	0.2857	0.05	1	0.5	1	0.05
PM	Mean	0.020	1.0019	-0.0007	0.355	0.8495	0.2991	0.0508	1	0.4990	0.9997	0.0507
	MAE	-	0.0867	0.1395	-	0.0751	0.1154	0.0361	-	0.0573	0.1154	0.0209
GLR	Mean	0.021	1.0020	-0.0008	0.356	0.8499	0.2996	0.0508	1	0.5079	1.0022	0.0506
	MAE	-	0.0868	0.1397	-	0.0750	0.1158	0.0354	-	0.0576	0.1185	0.0210
LB	Mean	0.021	1.0019	-0.0007	0.351	0.8495	0.2992	0.0508	1	0.5027	0.9962	0.0507
	MAE	-	0.0867	0.1396	-	0.0750	0.1154	0.0360	-	0.0565	0.1147	0.0217
HK	Mean	0.021	1.0024	-0.0019	0.351	0.8539	0.2904	0.0505	1	0.4981	1.0018	0.0505
	MAE	-	0.0806	0.0918	-	0.0792	0.1249	0.0301	-	0.0592	0.1184	0.0204

* *p*: proportion of phenotypic variance explained by QTL, H_0: null hypothesis (in the QTL), *m*: genotypic value of individuals *Qq*, *a*: difference between genotypic values of individuals *QQ* and *Qq*, MAE: mean absolute error. † Expressed in Morgan

Table 4. Simulation results of four QTL mapping methods for a Poisson trait: mixture of Poisson (MP), generalized linear regression (GLR), method of Lander and Botstein (LB) and of Haley and Knott (HK). Values are means of 5000 runs, considering a backcross with mean = 1, sample size of 50 individuals and length interval of 0.5 Morgan between markers*

		p = 0			*p* = 0.02				*p* = 0.20			
		Rate of H_0 rejection	*m*	*a*	Rate of H_0 rejection	*m*	*a*	QTL Position†	Rate of H_0 rejection	*m*	*a*	QTL Position
Method	Nominal Value	0.05	1	0	1	0.8571	0.2857	0.25	1	0.5	1	0.25
PM	Mean	0.029	1.0031	-0.0019	0.082	0.8603	0.2852	0.2531	0.675	0.5185	0.9584	0.2514
	MAE	-	0.2009	0.3441	-	0.1857	0.3099	0.1930	-	0.1425	0.2757	0.1300
GLR	Mean	0.028	1.0074	-0.0018	0.081	0.8633	0.2955	0.2536	0.654	0.5521	0.9962	0.2509
	MAE	-	0.2035	0.3484	-	0.1829	0.3187	0.1828	-	0.1392	0.3147	0.1228
LB	Mean	0.034	1.0035	-0.0011	0.085	0.8585	0.2919	0.2528	0.596	0.5345	0.9614	0.2504
	MAE	-	0.2030	0.3482	-	0.1864	0.3161	0.1868	-	0.1381	0.2916	0.1328
HK	Mean	0.033	1.0042	-0.0036	0.086	0.8678	0.2712	0.2516	0.594	0.5092	0.9798	0.2512
	MAE	-	0.1894	0.2617	-	0.1897	0.3379	0.1092	-	0.1615	0.3130	0.1149

* *p*: proportion of phenotypic variance explained by QTL, H_0: null hypothesis (in the QTL), *m*: genotypic value of individuals *Qq*, *a*: difference between genotypic values of individuals *QQ* and *Qq*, MAE: mean absolute error. † Expressed in Morgan

variables via some method as for example stepwise or backward elimination can account for the effects of other QTL. Under normality this is straightforward, by considering a regression model containing the effect(s) of a QTL in a given interval (for a fixed position) and parameters referring to markers other than those flanking that interval. Alternatively, such parameters could be incorporated in the regression model of Haley and Knott (1992). If the trait distribution is Poisson (or even other non-normal distribution), then a generalized linear regression should be considered. Using a linear predictor similar to the Zeng (1994) model of composite interval mapping, the approach would be similar to the PM method in this study, but eventually taking the effect of the remaining QTL into account. Another possibility would be the use of a linear predictor correspondent to the model of Haley and Knott (1992), but also with additional regression coefficients referent to marker loci not flanking the given interval.

The PM method is theoretically the one that can extract most information from the kind of data simulated in this study, for it considers their discrete nature and the mixture of distributions inherent in QTL analysis. Nevertheless, even under conditions of clearly discrete data (low), the four methods generally showed very similar results for precision of estimates and QTL detection power. This suggests that the choice of the method could be based on computational efficiency, which would prefer the HK method since it does not require an iterative process for parameter estimation, given a QTL position. Although this aspect of efficiency tends to be of minor importance since hardware is constantly being improved and speeded up, it may be of considerable relevance if one uses permutation tests or resampling techniques such as bootstrap. Our results as well as those of Zeng (1994) show that the use of chi-squares to determine critical regions for the likelihood ratio test is somewhat conservative, since this does not show the exact statistical distribution. Therefore, permutation tests could be useful when dealing with actual data.

ACKNOWLEDGEMENTS

This study was sponsored by CAPES.

Table 5. Simulation results of four QTL mapping methods for a Poisson trait: mixture of Poisson (MP), generalized linear regression (GLR), method of Lander and Botstein (LB) and of Haley and Knott (HK). Values are means of 5000 runs, considering a backcross with mean = 1, sample size of 50 individuals and length interval of 0.1 Morgan between markers*

		$p = 0$			$p = 0.02$				$p = 0.20$			
		Rate of H_0 rejection	m	a	Rate of H_0 rejection	m	a	QTL Position†	Rate of H_0 rejection	m	a	QTL Position
Method	Nominal Value	0.05	1	0	1	0.8571	0.2857	0.05	1	0.5	1	0.05
PM	Mean	0.024	0.9963	0.0043	0.090	0.8475	0.3011	0.0490	0.869	0.4913	1.0083	0.0500
	MAE	-	0.1774	0.2822	-	0.1569	0.2490	0.0395	-	0.1141	0.2353	0.0331
GLR	Mean	0.024	0.9968	0.0043	0.092	0.8475	0.3029	0.0492	0.868	0.4940	1.0183	0.0502
	MAE	-	0.1783	0.2841	-	0.1568	0.2506	0.0379	-	0.1139	0.2409	0.0308
LB	Mean	0.027	0.9963	0.0045	0.093	0.8474	0.3015	0.0491	0.859	0.4940	1.0085	0.0500
	MAE	-	0.1776	0.2826	-	0.1568	0.2493	0.0392	-	0.1126	0.2353	0.0331
HK	Mean	0.028	0.9962	0.0043	0.093	0.8622	0.2718	0.0495	0.858	0.4870	1.0171	0.0502
	MAE	-	0.1668	0.1919	-	0.1634	0.2808	0.0220	-	0.1174	0.2414	0.0300

*p: proportion of phenotypic variance explained by QTL, H_0: null hypothesis (in the QTL), m: genotypic value of individuals Qq, a: difference between genotypic values of individuals QQ and Qq, MAE: mean absolute error. † Expressed in Morgan

Mapeamento de QTL de características Poisson: um estudo de simulação

RESUMO - *O mapeamento de QTL (Locos de Caracteres Quantitativos) consiste na estimação da posição e dos efeitos de genes (ou grupos de ligação) controladores de características quantitativas. Para características que seguem uma distribuição normal de probabilidade existem metodologias estatísticas para o mapeamento de QTL já descritas na literatura, porém, para características discretas o volume de pesquisas é pequeno. Assim, este trabalho objetivou avaliar quatro métodos alternativos para o mapeamento de QTL controladores de características Poisson. As metodologias investigadas foram: mapeamento por intervalo padrão com os enfoques de Haley e Knott, e Lander e Botstein, misturas de distribuições Poisson e Modelos Lineares Generalizados. Os resultados, obtidos através de simulação computacional, mostraram muita similaridade entre as metodologias testadas, sendo que estas foram bastante eficazes, principalmente, considerando a falta de normalidade dos dados. Devido à simplicidade computacional, a metodologia de Haley e Knott foi considerada mais eficiente.*

Palavras-chave: Mapeamento de QTL, Distribuição de Poisson, simulação.

REFERENCES

Doerge RW and Churchill GA (1996) Permutation test for multiple loci affecting a quantitative character. **Genetics 142**: 284-294.

Falconer DS and Mackay TFC (1996) **Introduction to Quantitative Genetics**. 4th ed, Longman, New York, 464p.

Haley CS and Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity 69**: 315-324.

Jansen RC (1992) A general mixture model for mapping quantitative trait loci by using molecular markers. **Theoretical and Applied Genetics 85**: 252-260.

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. **Genetics 135**: 205-211.

Jiang C and Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics 140**: 1111-1127.

Kao CH, Zeng ZB and Teasdale RD (1999) Multiple interval mapping of quantitative trait loci. **Genetics 152**: 1203-1216.

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

Lynch M and Walsh B (1998) **Genetics and Analysis of Quantitative Traits**. Sinauer Associates, Sunderland, Massachusetts, 980p.

McCullagh P and Nelder JA (1989) **Generalized Linear Models.** 2nd ed, Chapman and Hall, London, 532p.

Meng XL and Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. **Biometrika 80**: 267-268.

Spyrides-Cunha MH, Demétrio CGB and Camargo LEA (2000) Proportional odds model applied to mapping of disease resistance genes in plants. **Genetics and Molecular Biology 23**: 223-227.

Visscher PM, Haley CS and Knott SA (1996) Mapping QTL for binary traits in backcross and F_2 populations. **Genetical Research 68**: 55-63.

Zeng ZB (1994) Precision mapping of quantitative trait loci. **Genetics 136**: 1457-1468.