A BAYESIAN APPROACH TO MAP QTL AND TO DETECT EPISTATIC EFFECTS IN A MAIZE POPULATION

Andréia da Silva MEYER[1]
Roseli Aparecida LEANDRO[2]
Antonio Augusto Franco GARCIA[3]
Anete Pereira de SOUZA[4]
Claudio Lopes de SOUZA JR[3]

▪ *ABSTRACT: The use of molecular markers has been the main tool to study the inheritance of quantitative traits, since it allows to estimate the position on the chromosomes and the effects of the QTL that controll these traits. The great difficulty of mapping QTL relates to the fact that the number of QTL is unknown and hence the dimension of the parametric space is also unknown. Bayesian approaches used with Markov Chain Monte Carlo method (MCMC) have been applied to infer QTL number, their positions in the genome and their genetic effects. The challenge is to obtain the sample from the joint posterior distribution of these parameters, since the number of QTL may be considered unknown and hence the dimension of the parametric space changes according to the number of QTL in the model. In this study, a Bayesian approach was applied, using a code implemented in the statistical program R, in order to map QTL for traits in a tropical maize population: plant height, ear height and grain yield. So, multiples QTL and epistatic effects were considered in the model and the number of QTL was considered as known. The MCMC methods were used to create a sample from the joint posterior distribution of the parameters. Models were adjusted with the crescent number of QTL and Bayes factor was used to select the most suitable model and, consequently, to estimate the number of QTL that control the traits assessed. The results from this study were compared to the results obtained using the frequentist approach for mapping QTL.*

[1]São Paulo State University – UNESP, Faculty of Agriculture and Veterinary – FCAV, Department of Exact Sciences, Jaboticabal, SP, Brazil. E-mail: *andreiameyer@fcav.unesp.br*
[2]University of São Paulo – USP, ESALQ, Department of Exact Sciences, Piracicaba, SP, Brazil. E-mail: *raleandr@usp.br*
[3]University of São Paulo – USP, ESALQ, Department of Genetics, Piracicaba, SP, Brazil. E-mail: *augusto.garcia@usp.br / clsouza@usp.br*
[4]University of Campinas – UNICAMP, Biology Institute and Molecular Biology Center and Genetic Engeneering, Department of Plant Biology, Campinas, SP, Brazil. E-mail: *anete@unicamp.br*

▪ *KEYWORDS: Bayes Factor; Bayesian inference; MCMC QTL mapping.*

1 Introduction

Many traits in plants and animals are quantitative in nature and are controlled by multiple genes and their interactions (epistasis). With the advent of molecular techniques it has become possible to map these loci that control these traits, named quantitative trait loci (QTL). This mapping entails identifying the QTL position on the genome and the estimates their genetic effect and interactions, thus shedding light on the quantitative traits. This is fundamental for breeding programs, because these loci could be associated with important traits, such as grain yield, plant height and volume, among others.

Various statistical models have been proposed for QTL mapping, ranging from simple linear regression models to sophisticated methods that simultaneously model the entire genome. The method widely used is composite interval mapping (CIM), which was proposed independently by Jansen & Stam (1994) and Zeng (1994). This model searches for QTLs in a given interval using outside markers as cofactors, for the purpose of control the variation caused by other QTL. However, the CIM method does not permit simultaneous mapping of multiple QTLs and hence cannot be used to estimate epistatic effects.

The two leading alternatives proposed in the literature for simultaneous mapping of multiple QTLs are multiple interval mapping (MIM), proposed by Kao & Zeng (1997) and Kao et al. (1999), and methods based on Bayesian inference by MCMC (Markov chain Monte Carlo) analysis, proposed by several authors (Satagopan et al., 1996; Heath, 1997; Satagopan & Yandell, 1998; Stephens & Fisch, 1998; Sillanpää & Arjas, 1998; Sen & Churchill, 2001; Gaffney, 2001; Wang et al., 2005; Yi, 2004; Yi et al., 2005, Yi et al., 2007). Multiple interval mapping combines the possibility of mapping and study of the genetic architecture of the quantitative traits, based on an algorithm to estimated the number, position, effects and interactions of QTL. In turn, Bayesian methods employ the joint posterior distribution to make inferences about the parameters of interest. This distribution is obtained by Bayes' Theorem and combines information from the parameters supplied by the data and expressed with the likelihood function, with uncertainty about the parameters expressed in the prior information. Hence, this approach has the advantage of allowing incorporation of additional information in the analysis based on previous experiments or knowledge of researchers, by attributing a prior distribution to each parameter of interest.

One of the pioneering works using this approach was presented by Satagopan et al. (1996). Since then, Bayesian methods have been used to form the basis of the new methods to map QTL. The challenge is to obtain a sample of the joint posterior distribution of these parameters, since usually the number of QTL is unknown and the dimension of the parametric space changes according to the number of QTL present in the model. Satagopan et al. (1996) considered the number of QTL

to be known, adjusted models to different numbers of QTLs and then selected the most suitable one using Bayes' factor as the selection criterion. Subsequent works (Satagopan & Yandell, 1998; Stephens & Fisch, 1998; Sillanpää & Arjas, 1998) overcame the problem of not knowing the number of QTl by considering it as unknown and using MCMC with reversible jumps (Green, 1995), which permits moving between models with different number of QTLs. The MCMC algorithm with reversible jumps is a very powerful and general technique to deal with posterior distributions when the parametric space does not have a fixed dimension.

Another challenge is the inclusion of epistatics interactions in the model, since this implies in considerable modification in the dimension of the parametric space, making the implementation of the algorithms used to build the samples of the joint posterior distribution of the parameters of interest much complex. Despite these challenges, the MCMC algorithm with reversible jumps has been employed by several authors to study epistasis (Yi & Xu, 2002, Yi et al., 2003, Narita & Sasaki, 2004). However, this simulation method converges very slowly and demands intense computational resources, making its use unwidely in many situations (Yi et al., 2005). Therefore, it is necessary to develop Bayesian methods to map QTL that are both efficient and easy to understand, particularly for more complex models, such as those that include epistasis. There is also a need to implement these algorithms in a way that is accessible to all researchers interested in the subject.

Based on the above, our aim here is to implement a Bayesian methodology to map QTLs in a case where the effects of epistasis are incorporated in the model. We compared the results obtained from this Bayesian approach to map QTLs against those obtained by the CIM method for three traits assessed in a tropical maize population.

2 Material and methods

2.1 Data Set

The data used in this work are presented in detail by Sibov et al. (2003a, 2003b) and Sabadin et al. (2008). To summarize, a maize population was developed from the cross between the inbred lines L-08-05F and L-14-4B, which have contrasting grain yield performance. The F_1 plants were selfed to gave rise to 400 F_2 plants, wich were selfed again to develope 400 $F_{2:3}$ progenies. These progenies were evaluated in five locations near the city of Piracicaba, SP, Brazil, in 1999/2000 (two locations) and 2000/2001 (three locations) growing seasons. The 400 progenies were allocated in four 10×10 lattices with two replications per location. Several traits were recorded in this experiment. Here we used grain yield (GY) in Mg per hectare, plant height (PH) in cm and ear height (EH) in cm.

The linkage map used to locate the QTLs contains 117 microsatellite marker loci, which are distributed in ten linkage groups, as described in Sibov et al. (2003a). The map length was approximately $1,634$ cM.

2.2 Statistical model

Assuming that the quantitative trait of interest is affected by S QTLs, the vector of observed phenotype values $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$, was described by the following linear regression model:

$$\mathbf{y} = \mathbf{1}\mu + X\boldsymbol{\beta} + \mathbf{e}, \tag{1}$$

where μ is a constant, $\boldsymbol{\beta}$ is the vector of genetic effects, X is the design matrix and e is the random errors vector, $e \sim N(0, I\sigma^2)$.

The matrix X and the vector of parameters $\boldsymbol{\beta}$ were defined using Cockerham's epistatic model (Kao & Zeng, 2002). The elements of the matrix X were defined by:

$$x_{ij1} = Q_{ij} - 1,$$

$$x_{ij2} = (1 + x_{ij1})(1 - x_{ij1}) - 0,5,$$

$$x_{ijj'k} = \begin{cases} x_{ij1}x_{ij'1} & k = 1 \\ x_{ij1}x_{ij'2} & k = 2 \\ x_{ij2}x_{ij'1} & k = 3 \\ x_{ij2}x_{ij'2} & k = 4 \end{cases}$$

where Q_{ij} is the number of dominant alleles of the genotype of the $j-$th QTL for the $i-$th individual, $j = 1, \ldots, S$ and $j \neq j'$. For Cockerham's model, $\beta_{j1} = \alpha_j$ and $\beta_{j2} = \delta_j$ correspond, respectively, to the additive and dominance effects of the j-th QTL and $\beta_{jj'1} = \alpha_j\alpha_{j'}$, $\beta_{jj'2} = \alpha_j\delta_{j'}$, $\beta_{jj'3} = \delta_j\alpha_{j'}$ e $\beta_{jj'4} = \delta_j\delta_{j'}$ are the epistatic effects between locus j and j': additive by additive, additive by dominance, dominance by additive and dominance by dominance (Kao & Zeng, 2002).

The inference about the parameters of interest, number of QTLs S, their locations on the chromosome $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_S)$ and their effects $\boldsymbol{\theta} = (\mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma^2)$, were carried out using the Bayesian approach, with S being considered as known, as proposed by Satagopan et al. (1996). We assumed a priori independence of the model's parameters and specified their prior distributions as follows: for the locations we assumed a uniform distribution in the interval $[0, L]$, with L being the total length of the chromosome; for the constant μ we assumed a distribution $N(0, \tau^2)$, with $\tau^2 = 10$, for the additive and dominance effects we used, respectively, $N(0, \sigma^2_\alpha)$ and $N(0, \sigma^2_\delta)$ where $\sigma^2_\alpha = 10$ and $\sigma^2_\delta = 10$. For the epistatic effects, it was assumed $N(0, \sigma^2_{\alpha\alpha})$, $N(0, \sigma^2_{\alpha\delta})$, $N(0, \sigma^2_{\delta\alpha})$ and $N(0, \sigma^2_{\delta\delta})$, where $\sigma^2_{\alpha\alpha} = \sigma^2_{\alpha\delta} = \sigma^2_{\delta\alpha} = \sigma^2_{\delta\delta} = 10$; for σ^2, a inverse-gamma distribution with hyperparameters $a = 8$ e $b = 8$, as proposed by Raftery et al. (2007), was adopted.

Thus, given model (1) and the prior distributions for the parameters of interest, the joint posterior distribution is

$$\pi(\boldsymbol{\lambda}, Q, \boldsymbol{\theta}|y) \propto \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-1}{2\sigma^2} \Big(y_i - \mu - \sum_{j=1}^{S} \alpha_j x_{ij1} - \sum_{j=1}^{S} \delta_j x_{ij2} - \right.\right.$$

$$\left.\left. \sum_{j=1}^{k} \alpha_j \alpha_{j'} x_{ijj'1} - \sum_{j=1}^{k} \alpha_j \delta_{j'} x_{ijj'2} - \sum_{j=1}^{k} \delta_j \alpha_{j'} x_{ijj'3} - \sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj'2} \Big)^2 \right\} \times \right.$$

$$\left. \prod_{j=1}^{S} \pi(Q_{ij}|\lambda_j, M) \right\} \times \left(\frac{1}{\text{L}}\right)^S \times \frac{1}{2\pi\tau^2} \exp\left\{ \frac{-1}{2\tau^2} (\mu)^2 \right\} \times$$

$$\prod_{j=1}^{S} \left(\frac{1}{2\pi\sigma_\alpha^2} \exp\left\{ \frac{-1}{2\sigma_\alpha^2} (\alpha_j)^2 \right\} \times \frac{1}{2\pi\sigma_\delta^2} \exp\left\{ \frac{-1}{2\sigma_\delta^2} (\delta_j)^2 \right\} \right) \times$$

$$\prod_{j=1}^{k} \left(\frac{1}{2\pi\sigma_{\alpha\alpha}^2} \exp\left\{ \frac{-1}{2\sigma_{\alpha\alpha}^2} (\alpha_j \alpha_{j'})^2 \right\} \times \frac{1}{2\pi\sigma_{\alpha\delta}^2} \exp\left\{ \frac{-1}{2\sigma_{\alpha\delta}^2} (\alpha_j \delta_{j'})^2 \right\} \times \right.$$

$$\left. \frac{1}{2\pi\sigma_{\delta\alpha}^2} \exp\left\{ \frac{-1}{2\sigma_{\delta\alpha}^2} (\delta_j \alpha_{j'})^2 \right\} \times \frac{1}{2\pi\sigma_{\delta\delta}^2} \exp\left\{ \frac{-1}{2\sigma_{\delta\delta}^2} (\delta_j \delta_{j'})^2 \right\} \right) \times$$

$$\frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left\{ \frac{-b}{\sigma^2} \right\}.$$

Obtaining the marginal posterior distribution for the parameters of interest is analytically complex, making it hard to obtain the posterior moments, such as the mean and variance, related to the relevant parameters. Therefore, we used an alternative computationally intensive method to obtain a sample of the joint posterior distribution for the parameters of interest. With the sample obtained it is easy to get the marginal moments (Casella & George, 1992). To construct the samples we utilized Gibbs sampler to update the mean, genetic effects, genotypes and variance parameters, and the Metropolis-Hastings algorithm to update the location of the QTLs. The description of the method to sample from the joint posterior distribution is presented in Appendixes A and B. We selected the model using Bayes' factor, with the harmonic means as estimators, as proposed by Newton & Raftery (1994) and the stabilized harmonic mean estimator, as proposed by Raftery et al. (2007) (Appendix C).

The program to analyze the data was implemented in the R statistical program (R Development Core Team, 2008) and codes are described in Meyer (2009). The joint posterior distribution was constructed using MCMC methods. The main steps for implementing the program, using the *Software* R, to construct a sample of the joint posterior distribution for the parameters of interest were:

1. Entry of phenotypic data, the molecular marker genotypes and the distance between the molecular markers;

2. Concatenation of chromosomes: So that all chromosomes can be considered simultaneously;

3. Assignment of initial values for all parameters; item [4.] Updating parameters:

4.1 Starts the counter of iterations at $t = 1$

4.2 Update $\sigma^2|\boldsymbol{\lambda}^{t-1}, Q^{t-1}, \mu^{t-1}, \boldsymbol{\alpha}^{t-1}, \boldsymbol{\delta}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, \boldsymbol{y}$, using the item d) of Appendix B;

4.3 Update $\lambda_j|\boldsymbol{\lambda}_{-j}^{t-1}, Q^{t-1}, \mu^{t-1}, \boldsymbol{\alpha}^{t-1}, \boldsymbol{\delta}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\gamma}3^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$, as described in item i) of Appendix A;

4.3 Update $Q_{ij}|\boldsymbol{\lambda}_{-j}^{t}, \mu^{t-1}, \boldsymbol{\alpha}^{t-1}, \boldsymbol{\delta}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$, as described in item ii) of Appendix A;

4.4 Update $\mu|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t-1}, \boldsymbol{\delta}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$ using the item a) of Appendix B;

4.5 Update $\alpha_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}_{-j}^{t-1}, \boldsymbol{\delta}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^1, \boldsymbol{y}$ using the item b) of Appendix B;

4.6 Update $\delta_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t}, \boldsymbol{\delta}_{-j}^{t-1}, \boldsymbol{\alpha\alpha}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$ using the item c) of Appendix B;

4.7 Update $\alpha\alpha_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t}, \boldsymbol{\delta}^{t}, \boldsymbol{\alpha\alpha}_{-j}^{t-1}, \boldsymbol{\alpha\delta}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$ using the item e) of Appendix B;

4.8 Update $\alpha\delta_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t}, \boldsymbol{\delta}^{t}, \boldsymbol{\alpha\alpha}^{t}, a\boldsymbol{\alpha\delta}_{-j}^{t-1}, \boldsymbol{\delta\alpha}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$, using the item f) of Appendix B;

4.9 Update $\delta\alpha_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t}, \boldsymbol{\delta}^{t}, \boldsymbol{\alpha\alpha}^{t}, \boldsymbol{\alpha\delta}_{2}^{t}, \boldsymbol{\delta\alpha}_{-j}^{t-1}, \boldsymbol{\delta\delta}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$ using the item g) of Appendix B;

4.10 Update $\delta\delta_j|\boldsymbol{\lambda}^{t}, Q^{t}, \boldsymbol{\alpha}^{t}, \boldsymbol{\delta}^{t}, \boldsymbol{\alpha\alpha}^{t}, \boldsymbol{\alpha\delta}_{2}^{t}, \boldsymbol{\delta\alpha}^{t}, \boldsymbol{\delta\delta}_{-j}^{t-1}, (\sigma^2)^t, \boldsymbol{y}$ using the item h) of Appendix B;

5. After update all parameters we calculated the density values of $L(\boldsymbol{\lambda}^t, Q^t, \boldsymbol{\theta}^t|\boldsymbol{y})$ and $L(\boldsymbol{\lambda}^t, Q^t, h(\boldsymbol{\theta}^t)|\boldsymbol{y})$, for calculation of the Bayes factor;

6. Made updates the parameters, increment the counter of iterations t and return to step 4.2;

7. Repeat step 6 M times, consider the period heating chain (burn-in) and spacing between points sampled called thin.

We generated chains with $120,000$ iterations, discarding the first 1,000 as burn-in, and used spacing of 50 iterations to reduce the correlation between the sampled values, resulting in a final sample size of 2,380. We visually checked the convergence of each chain by means of the Geweke criterion, implemented in the CODA package (Plummer et al., 2006). In the analysis of the three data sets were fitted models containing increasing numbers of QTL, starting with one and stopping at the number of QTLs indicated by the Bayes' factor. Finally, we interpreted the Bayes' factor according to the classification proposed by Jeffreys (1961).

3 Results

3.1 Grain yield

The analysis involved fitting models containing one, two, three and four QTLs. The acceptance rate in updating the QTL locations varied from 41% to 56%. Table 1 presents the results of Bayes' factor and their classification.

Table 1 - Bayes' factor for the QTL analysis of grain yield

$B^{1}_{(s+1)s}$	HM[2]	SHM[3]	Classification
B_{21}	1.4835E+05	3.5181E+04	very strong
B_{32}	3.4533E+02	1.6269E+02	very strong
B_{43}	4.6611	3.5168	weak

(1) Bayes' factor considering the adjusted model with s + 1 and s QTLs
(2) The estimates of Bayes' factor obtained considering the harmonic mean estimators
(3) The estimates of Bayes' factor obtained considering the stabilized harmonic mean estimators

A comparison using Bayes' factor of the model with one QTL with that containing two QTL (B_{21}) shows evidence that the latter best fits the data and a similar trend of adding a QTL was observed for B_{32} (strong evidence). But comparing the model with three QTL against that with four QTL (B_{43}) shows the three-QTL model fits the data best.

Table 2 presents a posterior summary of the parameters of the selected model. To enable simultaneous consideration of all the chromosomes, we concatenated the ten linkage groups. Hence, the locations of the QTLs ($\boldsymbol{\lambda}$) are presented considering the results of concatenating the linkage groups. Besides this, the locations are presented in centiMorgans. Note that the estimates of the additive and dominance effects and the additive by additive interactions are mostly negative, while the estimates of the additive by dominance interactions are all positive and the estimates for the dominance by dominance interactions are all negative.

Table 2 - Posterior summary and the 95% HPD region for the locations of the QTLs (λ), for the additive (α), dominance (δ) and epistatic ($\alpha\alpha$, $\alpha\delta$, $\delta\alpha$, $\delta\delta$) effects of the model with three QTLs

Parameters	Mean	s.d.	Median	HPD$(2,5\%)$	HPD$(97,5\%)$
μ	3.2896	0.0166	3.2900	3.2580	3.3215
σ^2	0.0793	0.0072	0.0800	0.0650	0.0929
α_1	-0.1451	0.0315	-0.1400	-0.2075	-0.0820
α_2	-0.1094	0.0281	-0.1100	-0.1611	-0.0527
α_3	0.1115	0.0315	0.1100	0.0475	0.1713
δ_1	0.1140	0.0495	0.1100	0.0195	0.2098
δ_2	-0.0613	0.0474	-0.0600	-0.1531	0.0307
δ_3	-0.0045	0.0496	0.0000	-0.0988	0.0973
λ_1	2.9549	0.0401	2.9600	2.8751	3.0309
λ_2	12.7577	0.0490	12.7600	12.6602	12.8494
λ_3	13.7621	0.0545	13.7600	13.6532	13.8678
$\alpha_1\alpha_2$	-0.0030	0.0419	0.0000	-0.0866	0.0782
$\alpha_1\alpha_3$	-0.0556	0.0466	-0.0600	-0.1428	0.0400
$\alpha_2\alpha_3$	0.0042	0.0460	0.0000	-0.0803	0.0960
$\alpha_1\delta_2$	0.0979	0.0649	0.1000	-0.0300	0.2323
$\alpha_1\delta_3$	0.0114	0.0683	0.0100	-0.1212	0.1539
$\alpha_2\delta_3$	0.0023	0.0681	0.0000	-0.1473	0.1259
$\delta_1\alpha_2$	-0.0208	0.0668	-0.0200	-0.1498	0.1132
$\delta_1\alpha_3$	0.0051	0.0651	0.0100	-0.1099	0.1400
$\delta_2\alpha_3$	0.0067	0.0651	0.0100	-0.1168	0.1362
$\delta_1\delta_2$	-0.0563	0.0907	-0.0600	-0.2481	0.1051
$\delta_1\delta_3$	-0.0294	0.0901	-0.0300	-0.1981	0.1538
$\delta_3\delta_3$	-0.0135	0.0916	-0.0100	-0.1936	0.1622

Table 3 presents the results for the locations and main genetic effects obtained by the CIM model presented in Sibov et al. (2003b) and those obtained here. To compare the two approaches, we transformed the unit of the genetic effects: additive and dominance presented in Sibov et al. (2003b) to the same unit of the effects used in this work. Note that the CIM model mapped four QTLs while the Bayesian approach mapped three QTLs. However, the QTL at 4/8 mapped by Sibov et al. (2003b) only explains 7.98% of the phenotypic variance, indicating it can have a small effect or is a false positive.

Figure 1 shows there is good agreement between the three QTLs mapped by the Bayesian approach and the QTLs mapped using the CIM model. The only disagreement is in chromosome 8, where the CIM model mapped one QTL more than did the Bayesian approach, as mentioned above.

Table 3 - Estimates of locations, additive and dominance effects of the QTLs mapped for grain yield for two mapping methods

methods	Qtl/Chrom.(1)	locations	Interval	effects	
				additive	dominance
CIM	1/2	52.88	umc1845-bnlg0166	0.0855	-0.0968
	2/7	125.20	dupssr13-umc1154	0.0804	0.0614
	3/8	57.94	phi0115-bnlg1176	-0.0776	-0.0065
	4/8	74.70	bnl1176-bnlg1607	-0.0915	0.0293
Bayesian	1/2	56.69	umc1845-bnlg0166	-0.1451	0.1140
	2/7	125.47	dupssr13-umc1154	-0.1094	-0.0613
	3/8	77.01	bnl1176-bnlg1607	0.1115	-0.0045

(1) Chromosome where the QTL is located

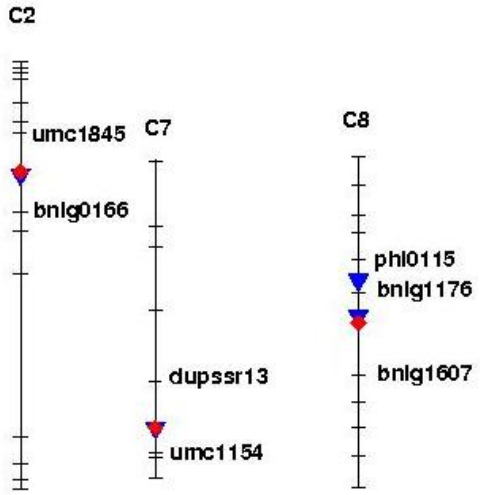

Figure 1 - Graph showing the location of all the QTLs mapped in the present work and in Sibov et al. (2003b) for the grain yield phenotype. The blue triangles show the QTLs mapped by Sibov et al. (2003b) and the red diamonds represent the QTLs mapped in the present work.

3.2 Ear height

In the joint analysis of the data on ear height, we adjusted models containing one, two, three, four and five QTLs. The acceptance rate used to update the QTL locations varied from 24% to 50%. Table 4 presents the results of Bayes' factor obtained considering the harmonic mean (HM) and stabilized harmonic mean (SHM) estimators and their classification.

Table 4 - Bayes' factor for the QTL analysis of Ear height

Model[1]	HM[2]	SHM[3]	Classification
B_{21}	30.7752	2.4868	Strong/ very weak
B_{32}	25.4537	7.4890	Strong/Weak
B_{43}	14464.6908	20.8394	very Strong/Strong
B_{54}	1.3795E-116	1.5798E-18	No evidence

(1) Bayes' factor considering the adjusted model with s + 1 and s QTLs
(2) The estimates of Bayes' factor obtained considering the harmonic mean estimators
(3) The estimates of Bayes' factor obtained considering the stabilized harmonic mean estimators

According to the result of the model choice procedure employing Bayes' factor, presented in Table 4, the model that best fits the data is that with four QTLs.

Table 5 shows a posterior summary for the parameters of the model with four QTLs, which was chosen by Bayes' factor. Note that all the estimates of the additive effects are positive and most of the estimates of the dominance effects are negative, while the estimates of the epistatic effects are balanced between positive and negative.

Table 6 presents the results for the locations and main genetic effects obtained by the CIM model presented in Sibov et al. (2003b) and those obtained by the Bayesian approach. Note once again that the Bayesian approach mapped one less QTL than the CIM model (five versus four). Also note the similarity between the three QTLs mapped by the Bayesian approach and the QTLs mapped using the CIM model. The exceptions are on chromosome 1, where the Bayesian approach did not map any QTL and the CIM model mapped one, and on chromosome 2, where the Bayesian technique mapped one QTL and the CIM model did not map any QTL (Figure 2). This can be due to the fact the QTLs have small effects or the inclusion of the epistatic effects in the Bayesian model.

Table 5 - Posterior summary and 95% HPD region for the locations of the QTLs (λ), for the additive (α), dominance (δ) and epistatic ($\alpha\alpha$, $\alpha\delta$, $\delta\alpha$, $\delta\delta$) effects of the model with four QTLs

Parameters	Mean	d.p.	Median	HPD(2.5%)	HPD(97.5%)
μ	0.8549	0.0037	0.8500	0.8476	0.8621
σ^2	0.0041	0.0004	0.0000	0.0034	0.0048
α_1	0.0223	0.0068	0.0200	0.0092	0.0358
α_2	0.0147	0.0066	0.0100	0.0025	0.0282
α_3	0.0184	0.0059	0.0200	0.0072	0.0303
α_4	0.0040	0.0063	0.0000	-0.0086	0.0158
δ_1	-0.0139	0.0097	-0.0100	-0.0335	0.0045
δ_2	-0.0061	0.0095	-0.0100	-0.0249	0.0123
δ_3	0.0001	0.0099	0.0000	-0.0198	0.0185
δ_4	-0.0105	0.0092	-0.0100	-0.0284	0.0067
λ_1	29.186	0.0647	29.200	27.918	30.434
λ_2	116.976	0.0418	117.000	116.127	117.745
λ_3	126.808	0.0724	126.800	125.548	128.351
λ_4	146.586	0.0499	146.600	145.621	147.580
$\alpha_1\alpha_2$	0.0009	0.0100	0.0000	-0.0195	0.0192
$\alpha_1\alpha_3$	-0.0049	0.0094	-0.0100	-0.0229	0.0135
$\alpha_1\alpha_4$	-0.0091	0.0098	-0.0100	-0.0290	0.0094
$\alpha_2\alpha_3$	-0.0001	0.0087	0.0000	-0.0175	0.0166
$\alpha_2\alpha_4$	-0.0041	0.0095	0.0000	-0.0227	0.0146
$\alpha_3\alpha_4$	-0.0041	0.0090	0.0100	-0.0122	0.0230
$\alpha_1\delta_2$	-0.0015	0.0142	0.0000	-0.0290	0.0265
$\alpha_1\delta_3$	-0.0097	0.0147	-0.0100	-0.0380	0.0182
$\alpha_1\delta_4$	-0.0164	0.0139	-0.0200	-0.0425	0.0106
$\alpha_2\delta_3$	-0.0042	0.0133	0.0000	-0.0302	0.0217
$\alpha_2\delta_4$	0.0145	0.0131	0.0100	-0.0103	0.0393
$\alpha_3\delta_4$	-0.0172	0.0128	-0.0200	-0.0431	0.0061
$\delta_1\alpha_2$	0.0121	0.0135	0.0100	-0.0125	0.0399
$\delta_1\alpha_3$	0.0085	0.0138	0.0100	-0.0188	0.0355
$\delta_1\alpha_4$	-0.0051	0.0142	-0.0100	-0.0342	0.0220
$\delta_2\alpha_3$	-0.0013	0.0119	0.0000	-0.0260	0.0199
$\delta_2\alpha_4$	-0.0018	0.0132	0.0000	-0.0286	0.0229
$\delta_3\alpha_4$	0.0310	0.0145	0.0300	0.0037	0.0606
$\delta_1\delta_2$	0.0055	0.0195	0.0100	-0.0338	0.0432
$\delta_1\delta_3$	0.0041	0.0192	0.0000	-0.0344	0.0413
$\delta_1\delta_4$	0.0067	0.0196	0.0100	-0.0315	0.0443
$\delta_2\delta_3$	0.0086	0.0181	0.0100	-0.0278	0.0429
$\delta_2\delta_4$	-0.0050	0.0175	-0.0100	-0.0404	0.0286
$\delta_3\delta_4$	-0.0050	0.0210	0.0300	-0.0123	0.0685

Table 6 - Estimates of locations, additive and dominance effects of the QTLs mapped for ear height for two mapping methods

methods	Qtl/Chrom.(1)	locations	Interval	effects	
				Aditivo	Dominância
CIM	1/1	150.38	bnlg0615-phi0037	0.0198	0.0056
	2/7	35.70	umc1632-umc1409	0.0169	-0.0111
	3/7	91.70	bnlg0434-dupssr13	0.0204	-0.0041
	4/7	119.20	dupssr13-umc1154	0.0221	0.0044
	5/9	1.01	umc1893-bnlg0430	-0.0084	-0.0242
Bayesian	1/2	49.06	umc11845-bnlg0166	0.0223	-0.0139
	2/7	19.46	umc1428-umc1632	0.0147	-0.0061
	3/7	117.78	dupssr13-umc1154	0.0184	0.0001
	4/9	11.06	umc1893-bnlg0430	0.0040	-0.0105

(1) Chromosome where the QTL is located

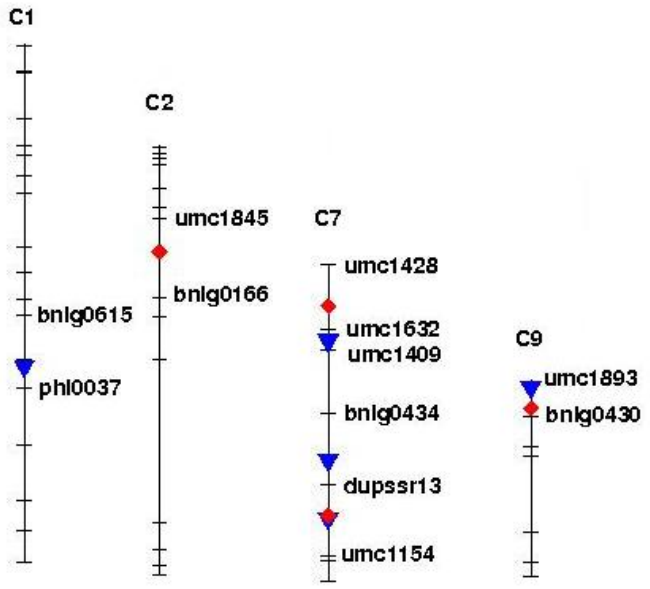

Figure 2 - Graph showing the location of all the QTLs mapped in the present work and in Sibov et al. (2003b) for the ear height phenotype. The blue triangles represent the QTLs mapped by Sibov et al. (2003b) and the red diamonds represent those mapped in the present work

3.3 Plant height

In the joint analysis of the data on plant height, we adjusted models containing one, two, three, four and five QTLs. The acceptance rate used to update the QTL locations varied from 29% to 66%. Table 7 presents the results of Bayes' factor.

Table 7 - Bayes' factor for the QTL analysis of Plant height

Model[1]	HM[2]	SHM[3]	Classification
B_{21}	164.3134	14.3657	very Strong/Strong
B_{32}	0.6183	4.1449	no evidence /weak
B_{43}	1.1565E+07	129.6249	very Strong
B_{54}	0.1221	0.7733	no evidence

(1) Bayes' factor considering the adjusted model with s + 1 and s QTLs
(2) The estimates of Bayes' factor obtained considering the harmonic mean estimators
(3) The estimates of Bayes' factor obtained considering the stabilized harmonic mean estimators

According to the Bayes' factor results in Table 8, the model that best fits the data is that containing four QTLs.

Table 9 presents a posterior summary for the parameters of the model with four QTLs, which was selected by Bayes' factor. Note that all the estimates of the additive effects are positive, most of the estimates of the dominance effects are negative and the estimates of the epistatic effects are mostly positive.

Table 9 presents the results for the locations and main genetic effects obtained by the CIM model presented in Sibov et al. (2003b) and those obtained by the Bayesian approach. Note that for the plant height characteristic, the Bayesian approach and the CIM model map the same number of QTLs.

Figure 3 shows that the results are similar for three QTLs mapped by the Bayesian approach and the QTLs mapped using the CIM model. The CIM model mapped one more QTL on chromosome 1, and the Bayesian approach mapped one more QTL on chromosome 7. These differences are probably due to the fact that the Bayesian method considers multiple QTLs simultaneously in the model, together with the epistatic effects.

Table 8 - Posterior summary and 95% HPD region for the locations of the QTLs (λ), for the additive (α), dominance (δ) and epistatic ($\alpha\alpha$, $\alpha\delta$, $\delta\alpha$, $\delta\delta$) effects of the model with four QTLs

Parameters	Mean	d.p.	Median	HPD(2.5%)	HPD(97.5%)
μ	1.7161	0.0054	1.7200	1.7052	1.7260
σ^2	0.0083	0.0008	0.0100	0.0069	0.0098
α_1	0.0306	0.0087	0.0300	0.0145	0.0488
α_2	0.0297	0.0101	0.0300	0.0104	0.049
α_3	0.0226	0.0088	0.0200	0.0055	0.0395
α_4	0.0258	0.0084	0.0300	0.0095	0.042
δ_1	-0.0052	0.0131	-0.0100	-0.031	0.0193
δ_2	-0.0263	0.0156	-0.0300	-0.0554	0.0053
δ_3	-0.0196	0.0129	-0.0200	-0.0441	0.0057
δ_4	0.0221	0.0149	0.0200	-0.005	0.0532
λ_1	15.368	0.0328	15.400	14.786	16.038
λ_2	29.509	0.0401	29.500	28.691	30.222
λ_3	117.493	0.0560	117.500	116.545	118.681
λ_4	127.207	0.0526	127.200	126.150	128.204
$\alpha_1\alpha_2$	0.0000	0.0149	0.0000	-0.0289	0.0288
$\alpha_1\alpha_3$	-0.0159	0.0132	-0.0200	-0.0417	0.0105
$\alpha_1\alpha_4$	0.0109	0.0125	0.0100	-0.0147	0.0346
$\alpha_2\alpha_3$	0.0033	0.0135	0.0000	-0.0251	0.0276
$\alpha_2\alpha_4$	0.0005	0.0130	0.0000	-0.0224	0.0288
$\alpha_3\alpha_4$	0.0005	0.0133	0.0000	-0.0291	0.0236
$\alpha_1\delta_2$	0.0013	0.0223	0.0000	-0.0417	0.0456
$\alpha_1\delta_3$	0.0307	0.0181	0.0300	-0.0051	0.0633
$\alpha_1\delta_4$	-0.0228	0.0198	-0.0200	-0.0614	0.0158
$\alpha_2\delta_3$	-0.0072	0.0193	-0.0100	-0.0447	0.0306
$\alpha_2\delta_4$	-0.0140	0.0207	-0.0100	-0.0552	0.0255
$\alpha_3\delta_4$	0.0070	0.0220	0.0100	-0.0355	0.0484
$\delta_1\alpha_2$	0.0087	0.0204	0.0100	-0.0322	0.0473
$\delta_1\alpha_3$	-0.0166	0.0186	-0.0200	-0.0503	0.0204
$\delta_1\alpha_4$	-0.0112	0.0175	-0.0100	-0.0448	0.0226
$\delta_2\alpha_3$	0.0096	0.0195	0.0100	-0.029	0.0478
$\delta_2\alpha_4$	0.0021	0.0180	0.0000	-0.032	0.0381
$\delta_3\alpha_4$	0.0073	0.0181	0.0100	-0.029	0.0423
$\delta_1\delta_2$	-0.0590	0.0323	-0.0600	-0.1213	0.0029
$\delta_1\delta_3$	0.0131	0.0263	0.0100	-0.0352	0.0672
$\delta_1\delta_4$	-0.0034	0.0286	0.0000	-0.0568	0.0537
$\delta_2\delta_3$	0.0082	0.0275	0.0100	-0.0445	0.0619
$\delta_2\delta_4$	0.0028	0.0266	0.0000	-0.0497	0.0557
$\delta_3\delta_4$	0.0028	0.0295	0.0300	-0.0336	0.0828

Table 9 - Estimates of locations, additive and dominant effects of the QTLs mapped for plant height for two different mapping methods

methods	Qtl/Chrom.(1)	locations	Interval	effects	
				Aditivo	Dominância
CIM	1/1	79.38	bnlg2238-umc2025	-0.0042	-0.0508
	2/1	153.38	bnlg0615-phi0037	0.0325	-0.0054
	3/2	53.88	umc11845-bnlg0166	0.0349	-0.0429
	4/7	117.20	dupssr13-umc1154	0.0286	0.0281
Bayesian	1/1	153.68	bnlg0615-phi0037	-0.0306	-0.0052
	2/2	53.29	umc11845-bnlg0166	0.0297	-0.0263
	3/7	24.63	umc1428-umc1632	0.0226	-0.0196
	4/7	121.77	dupssr13-umc1154	0.0258	0.0221

(1) Chromosome where the QTL is located

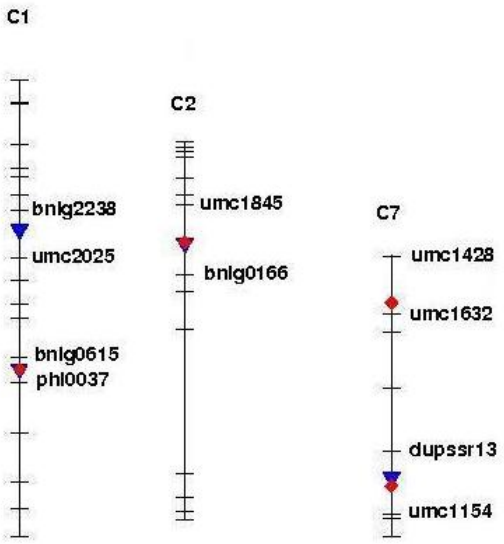

Figure 3 - Graph showing the location of all the QTLs mapped in the present work and those in Sibov et al. (2003b) for the plant height phenotype. The blue triangles represent the QTLs mapped by Sibov et al. (2003b) and the red diamonds represent those mapped in the present work.

4 Discussion

In the present work we used the Bayesian approach to map QTLs for three different corn traits: grain yield, plant height and ear height. In the Bayesian approach, the number of QTLs was considered known and the model was chosen by using Bayes' factor. We estimated with two different estimators: harmonic mean and stabilized harmonic mean. We concatenated the linkage groups so that all the chromosomes could be considered simultaneously, enabling inclusion of epistatic effects in the model.

For each of the three traits we successively adjusted models with a rising number of QTLs, increasing this number until Bayes' factor indicated the best model. Therefore, for all the phenotypes we fit models with one QTL more than the model that was selected according to Bayes' factor. In general the two estimators to obtain Bayes' factor produced very similar results.

We compared the mapping results obtained for the three traits by the Bayesian approach with those obtained by the CIM model, presented in Sibov et al. (2003b). There were some differences between the number of QTLs and positions on the genome. For grain yield, we mapped three QTLs while by the CIM model four were mapped. Probably the Bayesian approach mapped a smaller number of QTLs because epistasis is considered in the model or because the extra QTL mapped by CIM has a small effect. This can be expected to a certain extent because the models are different. However, there was agreement between the three QTLs mapped by the Bayesian approach and the CIM model. For the ear height phenotype, there were four QTLs mapped by the Bayesian approach versus five by the CIM model. Once again, there was one more mapped by CIM. In this case there were some differences between the results of the two methods. The CIM model mapped one QTL on chromosome 1, three on chromosome 7 and one on chromosome 9, while the Bayesian method mapped one QTL on chromosome 2, two on chromosome 7 and one on chromosome 9. For the plant height phenotype, both methods mapped the same number of QTLs, and only one difference was observed: the CIM model mapped one QTL on chromosome 1 that the Bayesian approach did not map and the Bayesian approach mapped one QTL on chromosome 7 that the CIM model did not map. We should point out that the Bayesian approach utilized here considered all the QTLs simultaneously in the model and also included epistasis terms. Hence some differences should be expected in comparison with the CIM results.

Acknowledgments

Received post-graduate fellowship from CAPES and CNPq.

MEYER, A. S.; LEANDRO, R. A.; GARCIA, A. A. F.; SOUZA JR. C. L.; SOUZA, A. P. Uma abordagem bayesiana para mapear QTL e detectar efeitos epistáticos em uma população de milho. *Rev. Bras. Biom.,* São Paulo, v.31, n.4, p.558-581, 2013.

▪ RESUMO: Muitos caracteres medidos em plantas e animais são de natureza quantitativa, ou seja, influenciados por múltiplos genes. Com o advento de novas técnicas moleculares tem sido possível mapear os locos que controlam tais caracteres, denominados QTLs (*Quantitative Trait Loci*). Mapear um QTL significa identificar sua posição no genoma, bem como, estimar seus efeitos genéticos e interações. A maior dificuldade para realizar o mapeamento de QTLs deve-se ao fato de que seu número é desconhecido e, conseqüentemente, a dimensão do espaço paramétrico também é desconhecida. Métodos Bayesianos juntamente com método Monte Carlo com Cadeias de Markov (MCMC), têm sido implementados para inferir conjuntamente o número de QTLs, suas posições no genoma e os seus efeitos genéticos . O desafio está em obter a amostra da distribuição conjunta a posteriori desses parâmetros, uma vez que o número de QTLs pode ser considerado desconhecido e a dimensão do espaço paramétrico muda de acordo com o número de QTLs presente no modelo. No presente trabalho foi utilizada uma abordagem bayesiana para mapear QTLs para caracteres numa população de milho tropical: altura de planta (AP), altura de espiga (AE) e produção de grãos. Para tanto, foram considerados múltiplos QTLs e efeitos de epistasia no modelo e o número de QTLs foi tratado como conhecido. Os métodos MCMC foram utilizados para gerar uma amostra da distribuição conjunta a posteriori dos parâmetros. Foram ajustados modelos com números crescentes de QTLs e o fator de Bayes foi utilizado para selecionar o modelo mais adequado e conseqüentemente, para estimar o número de QTLs que controlam os caracteres de interesse. Os resultados obtidos no presente trabalho foram comparados com os resultados obtidos pelo método frequentista para mapear QTLs.

▪ PALAVRAS-CHAVE: Fator de Bayes; inferência Bayesiana; MCMC; mapeamento de QTL.

References

BASTEN, C.J.; ZENG, Z.-B. *Windows QTL Cartographer 1.15*. Department of Statistics, North Carolina State University, Raleigh, NC, 1994.

CASELLA, G.; GEORGE, E.I. Explaining the Gibbs Sampler. *The American Statistician*, New York, v.46, p.167-174, 1992.

CHURCHILL, G.A; DOERGE, R.W. Empirical threshold values for quantitative trait mapping. *Genetics*, Austin, v.138, p.963-971, 1994.

GAFFNEY, P.J. *An efficient reversible jump Markov Chain Monte Carlo approach to detect multiple loci and their effects in inbred crosses.* 2001, 174p. Thesis (Philosophy in Statistics Doctor) - University of Wisconsin, Madison, 2001.

GREEN, P.J.; Reversible jump markov chaim Monte Carlo computation and Bayesian model determination. *Biometrika*, Lausanne, v.82, p.711-732, 1995.

HEATH, S.C. Markov chain Monte Carlo segregation and linkage analysis of oligogenic models. *The American Journal of Human Genetics*, Boston, v.61, p.748-760, 1997.

JEFFREY, H. *Theory of probability.* Oxford: Clarendon press, 1961. 447p.

KAO, C.H.; ZENG, Z-B.; TEASDALE, R.D. Multiple interval mapping for quantitative trait loci. *Genetics*, Austin, v.152, p.1203-1216, 1999.

KAO, C.H.; ZENG, Z-B. Modeling Epistasis of Quantitative Trait Loci Using Cockerham's Model. *Genetics*, Austin, v.160, p.1243-1261, 2002.

KOSAMBI, D.D. The estimation of map distances from recombination values. *Annual Eugenics*, New York, v.12, p.172-175, 1944.

NARITA, A.; SASAKI, Y. Detection of multiple QTL with epistatic effects under a mixed inheritance model in an outbred population. *Genetics Selection Evolution*, Versailles, v.36, p.415-433, 2004.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. V. *CODA: Convergence Diagnosis and Output Analysis for MCMC, R News*, v.6, p.7-11, 2006.

MEYER, A. S. *Uma abordagem bayesiana para mapeamento de QTLs em populações experimentais* 2009. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2009.

NEWTON, M.A.; RAFTERY, A.E. Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal Of Royal Statistical Society*, London, v.56, p.1-48, 1994.

R Development Core Team (2010). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

RAFTERY, A.E.; NEWTON, M.A.; SATAGOPAN, J. M.; KRIVITSKY, P.N. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: BERNARDO, J.M.; BAYARRI, M.J.; BERGER, J.O.; DAWID, A.P.; HECKERMAN, A.F.; SMITH, A.F.M; WEST, M. *Bayesian statistics.* Oxford: Oxford University Press, v.8, p.1-45, 2007.

SABADIN, P.K.; de SOUZA JÚNIOR, SOUZA, A.P. C.L.; GARCIA, A.A.F. QTL mapping for yeild components in tropical maize population using microsatellite markers. *Hereditas*, Lund, v.145, p.194-203, 2008.

SATAGOPAN, J. M.; YANDELL, B.S. Estimating the number of quantitative trait Loci via Bayesian model determination. *Special contributed paper session on genetic analysis of quantitative trait and complex disease, Biometric section, Joint Statistical Meetings*, Chicago, 1998.

SATAGOPAN, J. M.; YANDELL, B.S. NEWTON M.A.; OSBORN, T.C. A Bayesian approach to detect Quantitative trait Loci using Markov chain Monte Carlo. *Genetics*, Austin, v.144, p.805-816, 1996.

SEN, S.; CHURCHILL A statistical framework for quantitative trait mapping . *Genetics*, Austin, v.159, p.371-387, 2001.

SIBOV, S.T.; de SOUZA JÚNIOR, C.L.; GARCIA, A.A.F.; GARCIA, A.F.; SILVA, A.R.; MANGOLIM, C.A.; BENCHIMOL,L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays L.*) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. *Hereditas*, Lund, v.139, n.2,p.96-106, 2003a.

SIBOV, S.T.; de SOUZA JÚNIOR, C.L.; GARCIA, A.A.F.; GARCIA, A.F.; SILVA, A.R.; MANGOLIM, C.A.; BENCHIMOL,L.L.; SOUZA, A.P. Molecular mapping in tropical maize (*Zea mays L.*) using microsatellite markers. 2. Quantitative trait Loci (QTL) for grain yield, plant height, ear height and grain moisture . *Hereditas*, Lund, v.139, n.2, p.107-115, 2003b.

SILLANPÄÄ, M.J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data . *Genetics*, Austin, v.148, p.1373-1388, 1998.

STEPHENS, D.A.; FISCH, R.D. Bayesian analysis of quantitative trait locus data using reversible jump Markov Chain Monte Carlo *Biometrics*, Washington, v.54, p.1334-1347, 1998.

WANG, H.; ZHANG, Y.M.; MASINDE, G.L.; MOHAN, S; BAYLINK, D.J.; XU, S. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics*, Austin, v.170, p.465-480, 2005.

YI, N. A unified markov chain monte carlo framework for mapping multiple quantitative trait loci. *Genetics*, Austin, v.167, p.967-975, 2004.

YI, N.; ALLISON, D.B.; XU, S. Bayesian model choice and search strategies for mapping multiple epistatic quantitative trait loci. *Genetics*, Austin, v.165, p.867-883, 2003.

YI, N.; XU, S. Mapping quantitative trait loci with epistatic effects. *Genetics Research* , Cambridge, v.79, p.185-198, 2002.

YI, N.; YANDELL, B.S.; CHURCILL, G.A.; ALLISON, D.B.; EISEN, E.J.; POMP, D. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics*, Austin, v.170, p.1333-1344, 2005.

YI, N.; SHRINER, D.; BANERJEE, S.; MEHTA, T.; POMP, D YANDELL, B.S. An efficient Bayesian model selection approach for interacting QTL models with many effects. *Genetics*, Austin, v.176 , p.1865-1877, 2007.

ZENG, Z-B. Theoretical basis for separation of multiple linked genes effecs in mapping quantitative trait loci. *Procedure of the National Academic Science USA*, Washington, v.90, p.10972-10976, 1993.

ZENG, Z-B. Precision mapping of quantitative trait loci. *Genetics*, Austin, v.136, p.1457-1468, 1994.

Received in 19.06.2013.

Approved after revised in 21.02.2014.

Appendix A - Updating the parameters

The parameters were updated as follows:

i) Updating the positions ($\boldsymbol{\lambda}$): Each λ_j, $j = 1, \ldots, S$ was updated using the Metropolis-Hastings algorithm. The value proposed for the next chain state was generated according to Gaffney (2001): for the j-th QTL, a value was generated from a distribution $U[\lambda_j - d, \lambda_j + d]$, and the value proposed for the next chain state was given by $\lambda_j^* = min(|\lambda^{**}|, 2L - |\lambda^{**}|)$, where L is the total chromosome length. The proposal is denoted by $q(\lambda_j^*|\lambda_j)$ and the acceptance ratio is given by:

$$\alpha(\lambda_j, \lambda_j^*) = \frac{\pi(\lambda_j^*|\lambda_{-j}, Q, \boldsymbol{\theta}, \boldsymbol{y})}{\pi(\lambda_j|\lambda_{-j}, Q, \boldsymbol{\theta}, \boldsymbol{y})}$$

If the proposal for the next chain state, t, is accepted $\lambda_j^t = \lambda_j^*$, otherwise, $\lambda_j^t = \lambda_j^{t-1}$.

ii) Updating of Q_{ij}: the full conditional posterior distribution for each Q_{ij}, is a discrete distribution with the following probability:

$$p_{ij} = \pi(Q_{ij} = q|\lambda, Q_{i(-j)}, \boldsymbol{\theta}, \boldsymbol{y}) = \frac{\pi(Q_{ij} = q|\lambda_j)\pi(y_i|\boldsymbol{\theta}, Q_i, Q_{ij} = q)}{\sum_q \pi(Q_{ij} = q|\lambda_j)\pi(y_i|\boldsymbol{\theta}, Q_i, Q_{ij} = q)}$$

where $q = 0, 1, 2$. Thus Q_{ij} was updated using the Gibbs sampling algorithm.

iii) Updating of $\boldsymbol{\theta} = (\mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma^2)$:Since conjugate prior distributions were attributed to the parameters, their corresponding full conditional posterior distributions had closed forms, whose expressions are presented in Appendix B.

Appendix B - Full conditional posterior distributions

a) Full conditional posterior distribution for μ:

$$\mu|\boldsymbol{\lambda}, Q, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma_2, \boldsymbol{y} \sim N\left(\frac{\sum_{i=1}^{n}(mi)}{\sigma^2(\frac{1}{\tau^2}+\frac{n}{\sigma^2})}, \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right) \tag{1}$$

where:

$$mi = (y_i - \sum_{j=1}^{S}\alpha_j x_{ij1} - \sum_{j=1}^{S}\delta_j x_{ij2} - \sum_{j=1}^{k}\alpha_j\alpha_{j'} x_{ijj'1} - \sum_{j=1}^{k}\alpha_j\delta_{j'} x_{ijj'2} - \sum_{j=1}^{k}\delta_j\alpha_{j'} x_{ijj'3}$$

$$- \sum_{j=1}^{k}\delta_j\delta_{j'} x_{ijj'4})$$

b) Full conditional posterior distribution for α_{j^*}:

$$\alpha_{j^*}|\boldsymbol{\lambda}, Q, \mu, [\boldsymbol{\alpha}]_{j\neq j^*}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma_2, y \sim N\left(\frac{\sum_{i=1}^{n} x_{ij^*1} a_i}{\sigma^2\left(\frac{1}{\sigma_\alpha^2}+\frac{\sum_{i=1}^{n} x_{ij^*1}^2}{\sigma^2}\right)}, \frac{1}{\frac{1}{\sigma_\alpha^2}+\frac{\sum_{i=1}^{n} Q_{ij^*}^2}{\sigma^2}}\right) \tag{2}$$

where:

$$a_i = (y_i - \mu - \sum_{j\neq j^*}^{S}\alpha_j x_{ij1} - \sum_{j=1}^{S}\delta_j x_{ij2} - \sum_{j=1}^{k}\alpha_j\alpha_{j'} x_{ijj'1} - \sum_{j=1}^{k}\alpha_j\delta_{j'} x_{ijj'2} - \sum_{j=1}^{k}\delta_j\alpha_{j'} x_{ijj'3} -$$

$$\sum_{j=1}^{k}\delta_j\delta_{j'} x_{ijj'4})$$

c) Full conditional posterior distribution for δ_{j^*}:

$$\delta_{j^*}|\boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, [\delta]_{j\neq j^*}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma_2, y \sim N\left(\frac{\sum_{i=1}^{n} x_{ij^*2} d_i}{\sigma^2\left(\frac{1}{\sigma_\delta^2}+\frac{\sum_{i=1}^{n} x_{ij^*1}^2}{\sigma^2}\right)}, \frac{1}{\frac{1}{\sigma_\delta^2}+\frac{\sum_{i=1}^{n} x_{ij^*2}^2}{\sigma^2}}\right) \tag{3}$$

where:

$$d_i = (y_i - \mu - \sum_{j=1}^{S}\alpha_j x_{ij1} - \sum_{j\neq j^*}^{S}\delta_j x_{ij2} - \sum_{j=1}^{k}\alpha_j\alpha_{j'} x_{ijj'1} - \sum_{j=1}^{k}\alpha_j\delta_{j'} x_{ijj'2} - \sum_{j=1}^{k}\delta_j\alpha_{j'} x_{ijj'3} -$$

$$\sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj'l4})$$

d) Full conditional posterior distribution for σ^2:

$$\sigma^2 | \boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \boldsymbol{y} \sim IG\left(a + \frac{n}{2}, \frac{1}{\frac{\sum_{i=1}^{n} S_i}{2} + \frac{1}{b}}\right) \tag{4}$$

where:

$$S_i = (y_i - \mu - \sum_{j=1}^{S} \alpha_j x_{ijx} - \sum_{j=1}^{S} \delta_j x_{ij2} - \sum_{j=1}^{k} \alpha_j \alpha_{j'} x_{ijj'l1} - \sum_{j=1}^{k} \alpha_j \delta_{j'} x_{ijj'l2} - \sum_{j=1}^{k} \delta_j \alpha_{j'} x_{ijj'l3} -$$

$$\sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj'l4})^2$$

e) Full conditional posterior distribution for $\alpha\alpha_{j^*}$:

$$\alpha\alpha_{j^*} | \boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, [\boldsymbol{\alpha\alpha}]_{j \neq j^*}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma_2, y \sim N\Big(\frac{\sum_{i=1}^{n} x_{ij^*j'l1} E1_i}{\sigma^2\Big(\frac{1}{\sigma^2_{\gamma_1}} + \frac{\sum_{i=1}^{n} x^2_{ij^*j'l1}}{\sigma^2}\Big)}, \frac{1}{\frac{1}{\sigma^2_{\gamma_1}} + \frac{\sum_{i=1}^{n} x^2_{ij^*j'l1}}{\sigma^2}}\Big) \tag{5}$$

where:

$$E1_i = (y_i - \mu - \sum_{j}^{S} \alpha_j x_{ij1} - \sum_{j=1}^{S} \delta_j x_{ij2} - \sum_{j \neq j^*}^{k} \alpha_j \alpha_{j'} x_{ijj'l1} - \sum_{j=1}^{k} \alpha_j \delta_{j'} x_{ijj'l2} - \sum_{j=1}^{k} \delta_j \alpha_{j'} x_{ijj'l3} -$$

$$\sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj'l4})$$

f) Full conditional posterior distribution for $\alpha\delta_{j^*}$:

$$\alpha\delta_{j^*} | \boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, [\boldsymbol{\alpha\delta}]_{j \neq j^*}, \boldsymbol{\delta\alpha}, \boldsymbol{\delta\delta}, \sigma_2, \boldsymbol{y} \sim N\Big(\frac{\sum_{i=1}^{n} x_{ij^*j'l2} E2_i}{\sigma^2\Big(\frac{1}{\sigma^2_{\gamma_2}} + \frac{\sum_{i=1}^{n} x^2_{ij^*j'l2}}{\sigma^2}\Big)}, \frac{1}{\frac{1}{\sigma^2_{\gamma_2}} + \frac{\sum_{i=1}^{n} x^2_{ij^*j'l2}}{\sigma^2}}\Big) \tag{6}$$

where:

$$E2_i = (y_i - \mu - \sum_{j}^{S} \alpha_j x_{ij1} - \sum_{j=1}^{S} \delta_j x_{ij2} - \sum_{j=1j^*}^{k} \alpha_j \alpha_{j'} x_{ijj'l1} - \sum_{j \neq j^*}^{k} \alpha_j \delta_{j'} x_{ijj'l2} - \sum_{j=1}^{k} \delta_j \alpha_{j'} x_{ijj'l3} -$$

$$\sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj^l 4})$$

g) Full conditional posterior distribution for $\delta\alpha_{j^*}$:

$$\delta\alpha_{j^*} | \boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, [\boldsymbol{\delta\alpha}]_{3j\neq j^*}, \boldsymbol{\delta\delta}, \sigma_2, y \sim N\Big(\frac{\sum_{i=1}^{n} x_{ij^* j^l 3} E3_i}{\sigma^2 \Big(\frac{1}{\sigma^2_{\gamma_3}} + \frac{\sum_{i=1}^{n} x^2_{ij^* j^l 3}}{\sigma^2} \Big)}, \frac{1}{\frac{1}{\sigma^2_{\gamma_3}} + \frac{\sum_{i=1}^{n} x^2_{ij^* j^l 3}}{\sigma^2}} \Big) \tag{7}$$

where:

$$E3_i = (y_i - \mu - \sum_{j}^{S} \alpha_j x_{ij1} - \sum_{j=1}^{S} \delta_j x_{ij2} - \sum_{j=1 j^*}^{k} \alpha_j \alpha_{j'} x_{ijj^l 1} - \sum_{j=1}^{k} \alpha_j \delta_{j'} x_{ijj^l 2} - \sum_{j\neq j^*}^{k} \delta_j \alpha_{j'} x_{ijj^l 3} -$$

$$\sum_{j=1}^{k} \delta_j \delta_{j'} x_{ijj^l 4})$$

h) Full conditional posterior distribution for $\delta\delta_{j^*}$:

$$\delta\delta_{j^*} | \boldsymbol{\lambda}, Q, \mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha\alpha}, \boldsymbol{\alpha\delta}, \boldsymbol{\delta\alpha}, [\boldsymbol{\delta\delta}]_{j\neq j^*}, \sigma_2, y \sim N\Big(\frac{\sum_{i=1}^{n} x_{ij^* j^l 4} E4_i}{\sigma^2 \Big(\frac{1}{\sigma^2_{\gamma_4}} + \frac{\sum_{i=1}^{n} x^2_{ij^* j^l 4}}{\sigma^2} \Big)}, \frac{1}{\frac{1}{\sigma^2_{\gamma_4}} + \frac{\sum_{i=1}^{n} x^2_{ij^* j^l 4}}{\sigma^2}} \Big) \tag{8}$$

where:

$$E4_i = (y_i - \mu - \sum_{j}^{S} \alpha_j x_{ij1} - \sum_{j=1}^{S} \delta_j x_{ij2} - \sum_{j=1 j^*}^{k} \alpha_j \alpha_{j'} x_{ijj^l 1} - \sum_{j=1}^{k} \alpha_j \delta_{j'} x_{ijj^l 2} - \sum_{j=1}^{k} \delta_j \alpha_{j'} x_{ijj^l 3} -$$

$$\sum_{j\neq j^*}^{k} \delta_j \delta_{j'} x_{ijj^l 4})$$

Appendix C - Bayes Factor

The Bayes factor to compare model i and model j is a ratio of integrated likelihoods (also called normalizing constant or marginal likelihood) given by

$$f(y|M_j) = \int L(y|\lambda_j, \theta, Q_j)\pi(\lambda_j, \theta_j)d(\lambda_j, \theta_j). \tag{1}$$

The estimation of the integrated likelihood (1) is made using posterior simulation output $\boldsymbol{\theta}^t$, this artifice is used because in the majority of the cases is difficult to calculate the integrated likelihoods. We used the harmonic means as estimators, as proposed by Newton & Raftery (1994).

$$\hat{\pi}_{HM}(\boldsymbol{y}|M_j) = \Big[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{L(\boldsymbol{y}|M_j, \boldsymbol{\lambda}^t, Q^t, \boldsymbol{\theta}^t)}\Big]^{-1} \tag{2}$$

and the stabilized harmonic mean estimator, as proposed by Raftery et al. (2007)

$$\hat{\pi}_{SHM}(\boldsymbol{y}|M_j) = \Big[\frac{1}{T}\sum_{t=1}^{T}\frac{1}{L(\boldsymbol{y}|M_j, \boldsymbol{\lambda}^t, Q^t, h(\boldsymbol{\theta}^t))}\Big]^{-1}, \tag{3}$$

where $h(\boldsymbol{\theta}^t) = (\mu, \boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\alpha}\boldsymbol{\alpha}, \boldsymbol{\alpha}\boldsymbol{\delta}, \boldsymbol{\delta}\boldsymbol{\alpha}, \boldsymbol{\delta}\boldsymbol{\delta})$.