Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(1) 2010 11-18

Bayesian Hierarchical Models to Identify Quantitative Trait Loci using Replicated Lines

Susan J. Simmons[1*], Ann E. Stapleton[2], Fang Fang[3], Qijun Fang[3] and Karl Ricanek[4]
[1]28403 Department of Mathematics and Statistics, University of North Carolina Wilmington, NC
[2]Department of Biology and Marine Biology, University of North Carolina Wilmington, NC
[3]GIDP, University of Arizona, AZ
[4]Department of Computer Science, University of North Carolina Wilmington, NC

SUMMARY

The identification of locations on a genetic map that associate with quantitative traits is an important issue in plant breeding and gene identification in crops. Many of the available algorithms for quantitative trait loci (QTL) allow only one observation per genotype distribution. Information within plant lines is summarized into a single observation in order to utilize available programs. However, important variation information within lines is lost. We propose using a Bayesian hierarchical model that incorporates the multiple observations within plant lines. A Markov chain Monte Carlo model composition strategy is used to search and identify genetic markers associated with a quantitative trait. An extensive simulation study illustrates the effectiveness of this method. Results from applying this algorithm to Bay-0 × Shahdara Arabidopsis *thaliana* recombinant inbred line QTL experiment are discussed.

Keywords : Bayesian hierarchical model, Markov chain Monte Carlo model composition, Activation probability.

1. INTRODUCTION

Quantitative trait loci (QTL) experiments have yielded important biological and biochemical information necessary for understanding various traits (Lynch and Walsh 1997). Through QTL experiments, genetic markers responsible for weight gain in pigs (Jing-hu *et al.* 2007), pecking-related traits in chickens (Buitenhuis *et al*. 2005), and starch content and composition in maize (Séne *et al.* 2000) have been determined.

Over the last 20 years, there has been an abundance of algorithms proposed for QTL mapping such as interval mapping strategies (Lander and Botstein 1989, Luo and Kearsey 1992, Jansen 1993, Luo and Williams 1993), composite interval mapping strategies (Jansen and Stam 1994, Zeng 1994, Jiang and Zeng 1997, Gao and Yang 2006), multiple interval mapping (Haley and Knott 1992, Jensen 1996, Liu *et al.* 1996, Kao and Zeng 1997, Weber *et al.* 1999, Kao 2000, Zeng *et al.* 2000, Zeng *et al.* 2005), Bayesian interval mapping (Satagopan *et al*. 1996, Sillanpaa and Arjas 1998, Sen and Churchill 2001, Yandell *et al.* 2007) and model selection strategies (Ball 2001, Broman and Speed 2002, Sillanpaa and Corander 2002). However, most algorithms allow only one observation per genotype distribution. In situations such as plant QTL experiments where there can be cloned plants, observations within lines are summarized into a single observation to utilize these programs. The method of summarizing observations can vary; however, in doing so, important information about variation within the lines is lost (Pearson *et al.* 2007).

[*] *Corresponding author* : Susan Simmons
E-mail address : simmonssj@uncw.edu

Previous work (Pearson *et al.* 2007) investigated using a Bayesian hierarchical model with a systematic search. The search algorithm of Pearson *et al*. (2007) calculated the activation probability for each chromosome. Only those chromosomes identified as containing potential QTL information were retained in the analysis (chromosomes with activation probability greater than 0.5). Those chromosomes kept for further examination were each subdivided into two parts. For example, if only two chromosomes had a posterior probability greater than 0.5, these two chromosomes were divided in half, and the algorithm continued by finding the activation probability of the four regions. The algorithm persisted by further identifying important regions and then subdividing those regions into two halves. The algorithm stopped when regions were refined to only single markers. Through simulations, this methodology was shown to be superior to composite interval mapping when replicate information was inherent in the experiment. The simulations were created under the assumption that the additive error followed a normal distribution. In this paper, we further develop this methodology using a stochastic search in a Markov chain Monte Carlo model composition technique. This improved algorithm is validated through an extensive simulation study using marker information obtained from a QTL experiment on Bay-0 × Shahdara Arabidopsis *thaliana* recombinant inbred lines (Loudet *et al*. 2002). The Bayesian hierarchical model can efficiently incorporate replicate information and variations within each line.

2. METHODOLOGY

The measured trait in the QTL experiment is represented by y_{ij} where $i = 1, \ldots, L$ (L = number of lines) and $j = 1, \ldots, n_i$ (n_i = the number of replicates within line i). Plants within each line are genetically identical and can be considered clustered data. This clustered data is included in the model through a hierarchical approach in which plants or observations within each line are considered conditionally independent within each line. Under this supposition, the observed traits are assumed to come from independent normal distributions with means θ_i and variances σ_i^2 or in other words

$$y_{ij} \mid \theta_i,\ \sigma_i^2 \sim N(\theta_i, \sigma_i^2)$$

Due to the nature of most plant experiments, the number of observations within each line, n_i, could potentially be considered a random variable. The methodology developed herein assumes that the distribution of y_{ij}'s are conditioned on the sample size. For notational convenience, this dependency is suppressed in the equations throughout the manuscript. Furthermore, the means are assumed to be influenced by the marker information, which is denoted as $\mathbf{X}$ such that $\mathbf{X}$ is $M \times L$ where M is the number of markers and L is number of lines. The mean of each line is assumed to come from a normal distribution with mean $\mathbf{X}_i'\beta$ and variance τ^2 where $\mathbf{X}_i'$ is the transpose of i^{th} column of $\mathbf{X}$ or

$$\theta_i \mid \mathbf{X}, \beta, \tau^2 \sim N(\mathbf{X}_i'\beta, \tau^2)$$

Since no prior information is assumed to be known about which markers might be potential QTL, we assign a normal prior distribution with mean 0 and variance 100 on the coefficient for each marker ($\beta \sim N(0, 100)$). The prior distribution for the variance parameters σ_i^2 and τ^2 are assigned Inverse-$\chi^2(1)$ which has an infinite mean and variance.

These assumptions yield an implicit full posterior distribution; however, the full conditionals have a nice parametric form. The full conditional posterior distributions are represented below as $\mathbf{X} \mid \cdot$, which is the conditional distribution of the random variable $\mathbf{X}$ given all other quantities.

$$\theta_i \mid \cdot \sim N\left(\left(\frac{1}{\tau^2}\mathbf{I} + \gamma^{-1}\right)^{-1}\left(\frac{1}{\tau^2}\mathbf{X}'\beta + \gamma^{-1}Y\right), \left(\frac{1}{\tau^2}\mathbf{I} + \gamma^{-1}\right)^{-1}\right)$$

$$\sigma_i^2 \mid \cdot \sim \text{Inv-gamma}\left(\frac{n_i+1}{2}, \frac{1}{2}\left(1 + \sum_j (y_{ij} - \theta_i)^2\right)\right)$$

$$\beta_j \mid \cdot \sim \left(\left(\frac{1}{\tau^2}\mathbf{X}'\mathbf{X} + \frac{1}{100}\right)^{-1}\mathbf{X}'\theta, \left(\frac{1}{100} + \frac{1}{\tau^2}\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

$$\tau^2 \mid \cdot \sim \text{Inv-gamma}\left(\frac{L+1}{2}, \frac{1}{2}\left(1 + \sum_i (\theta_i - \mathbf{X}'\beta)^2\right)\right)$$

where,

$$\gamma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_L^2 \end{pmatrix}$$

Samples from the posterior distribution are used to estimate the likelihood of the data for a given model (t) as shown in equation (1)

$$p(D \mid M^{(t)}) = \int \ldots \int p(y \mid \theta, \sigma^2, \beta, \tau^2)\, p(\theta, \sigma^2, \tau^2, \beta)\, d\beta^{(t)}\, d\tau^{2(t)}\, d\theta^{(t)}\, d\sigma^{2(t)} \quad (1)$$

2.1 Stochastic Search

The Markov chain Monte Carlo model composition strategy (MC^3) (Boone *et al*. 2006) randomly generates an initial model selection vector. The model selection vector is a binary vector of length M. Along the model selection vector, locations with values of 1 indicate that the marker is included in the current model and locations with a value of 0 indicate that the marker is not included in the model. The likelihood of the data under the initial model selection vector is calculated (equation 1). A position along the model selection vector is randomly selected, and its value is arbitrarily switched such that if the original value is 1 then it becomes 0 and vice versa. Next the likelihood of the data under this new model is calculated (equation 1) and compared to the likelihood of the original model by

$$\alpha_{ij} = \min\left(1, \frac{p(D|M^{(j)})}{p(D|M^{(i)})}\right)$$

The value of α_{ij} is the transition probability from model i to model j. A Bernoulli random variable is generated with probability of success of α_{ij}. If the generated Bernoulli random variable is 1, then the new model (j) becomes the current state. If, however, the generated Bernoulli random variable is 0, the current model is maintained (i). Based on the model selected, a new randomly selected position on the model selection vector is chosen and its value is switched. The likelihood of this new model is calculated and used to determine the transition probability. This chain is continued until 2,000 states (or models) have been visited. A multiplicity of chains is simultaneously generated in the fashion described above to protect against premature convergence caused by a local minimum. This work uses ten parallel chains with randomly generated initial model selection vectors.

Upon completion of all chains, 20,000 states have been visited and the likelihoods of the model computed. The posterior model probability for the 20,000 models are calculated as

$$p(M^{(k)} \mid D) = \frac{p(D|M^{(k)})\, p(M^{(k)})}{\Sigma_k p(D|M^{(k)}) P(M^{(k)})} \quad (2)$$

We assume no prior model information, and therefore assign equal probability among models. Assuming $p(M^{(1)}) = p(M^{(2)}) = \ldots = p(M^{(k)})$ simplifies equation (2) to

$$p(M^{(k)} \mid D) = \frac{p(D|M^{(k)})}{\Sigma_k p(D|M^{(k)})} \quad (3)$$

Using this quantity, the activation probability for each marker is calculated as

$$p(\beta_j \neq 0 \mid D) = \Sigma_k p(\beta_j \neq 0 \mid M^{(k)}, D)\, p\,(M^{(k)} \mid D) \quad (4)$$

where $p(\beta_j \neq 0 \mid M^{(k)}, D) = 1$ if marker j is in the model and 0 if marker j is not in the model.

3. SIMULATIONS

In order to validate the current methodology, we performed 60 simulations on responses generated from the marker information in the Bay-0 and Shahdara Arabidopsis *thaliana* recombinant line (Loudet *et al*. 2002). This mapping population size and marker density is typical of populations used in plant breeding. The **X** matrix consisted of 38 markers shown in the genetic map in Fig. 1.

The marker state (in the biparental case, the state inherited from one parent or the other parent) is measured in each line. The resulting line distribution pattern is visualized in Fig. 2 for the 165 lines in this Bay × Sha population.

This **X** matrix was used to generate quantitative traits by the following model

$$y_{ij} = \sum_j a_j x_{ij} + \varepsilon_{ij}$$

where a_j is the effect of the j^{th} marker, x_{ij} is the marker information from the i^{th} line and j^{th} marker where x_{ij} equals –0.5 when the marker information is from the

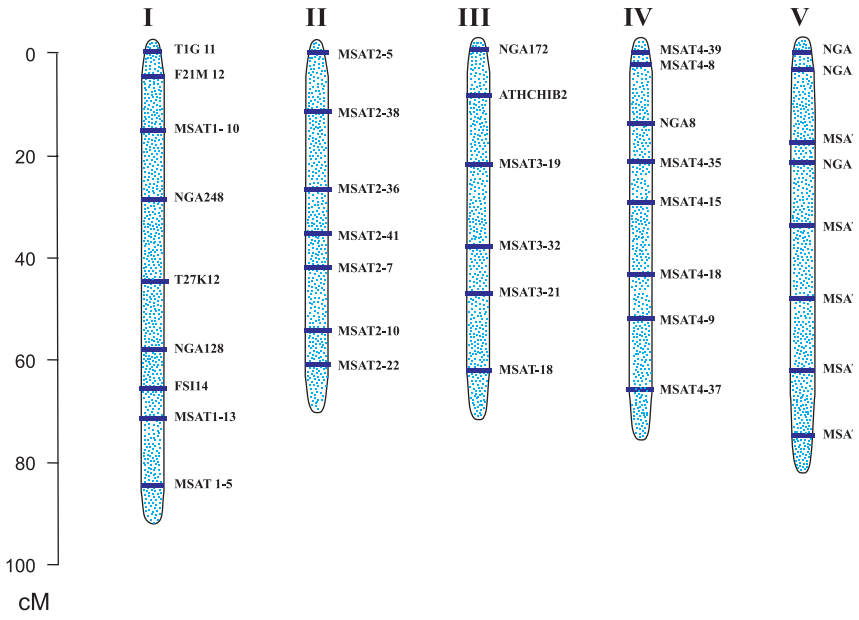

Fig. 1. Genetic map for the Bay-0 × Shahdara recombinant inbred line.

Bay parent and 0.5 when the marker information is from the Sha parent and ε_{ij} is the random noise. Since the true error distribution of quantitative traits is unknown, it is important to assess if the methodology is robust to the assumption of normally distributed errors. To investigate how well the method performs under nonnormal noise, the simulations in this study were generated using the gamma distribution for the error.

To incorporate different levels of variability, we used two different gamma distributions. The first gamma distribution had a shape parameter of 0.5 and a scale parameter of 1, giving an expectation for the variance within each line of 0.5. The second gamma distribution exhibited more variation with a shape parameter of 1 and a scale parameter of 3 giving an expected variance of 3 within each line. Effect sizes of

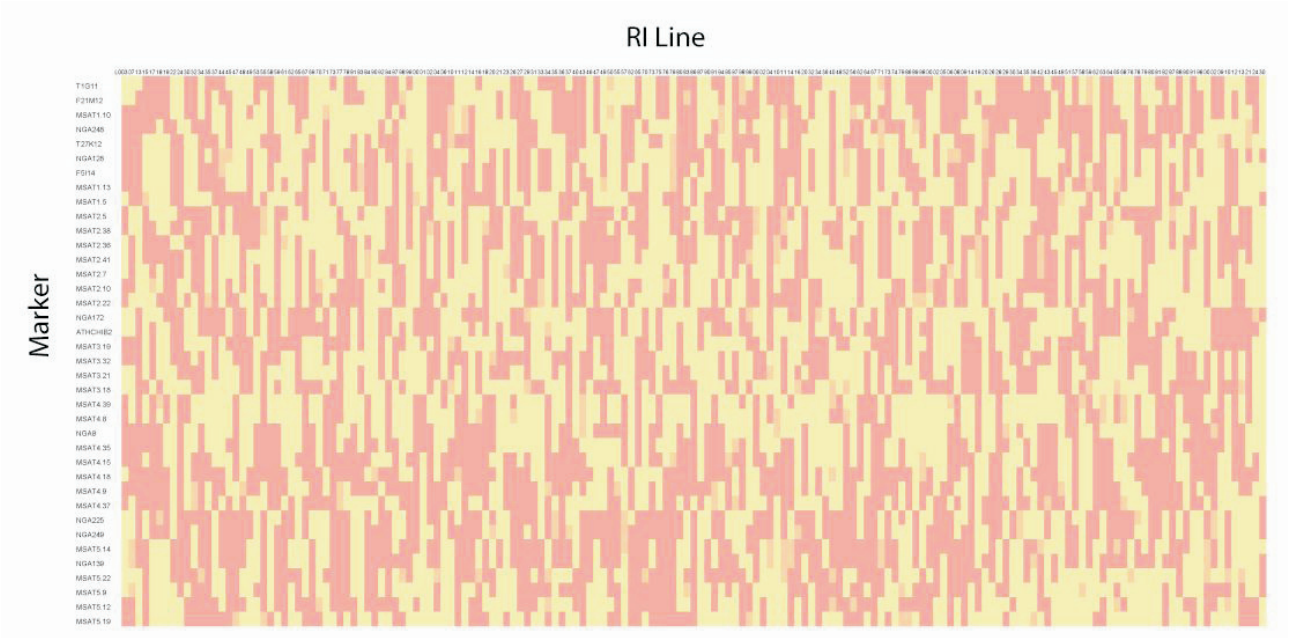

Fig. 2. Overview of marker distribution within lines. For each of the 38 markers (in rows) and 165 lines (in columns) the marker state is indicated by yellow for the Bay parent state and by red for the Sha parent state. Missing markers are indicated by medium beige coloration.

each QTL were selected in the range from 1 to 9, and each simulation had between one and six QTL in the model. There was a total of 60 different simulation scenarios developed for this study.

The new proposed approach was compared to the systematic approach outlined in Pearson *et al*. (2007). In Pearson *et al*. (2007), the authors investigated the effectiveness of the systematic search in a Bayesian hierarchical model to composite interval mapping when replicated information was involved in the experiment. The study simulated up to two QTL with varying effect sizes and a normal distribution for error. In this study, the systematic Bayesian hierarchical model (which will be referred to as the baseline approach) outperformed the composite interval mapping under the condition of replicates. However, false positives (markers adjacent to the QTL) were still evident in the baseline approach. Due to the low correlations among the markers of the Bay-0 × Shahdara populations, this should not occur. The conclusion drawn from this work was that the baseline approach successfully identified marker locations (and occasionally nearby marker locations) where composite interval mapping approach did not.

The current simulation study involves more complexity with regard to the number of QTL in each simulation (one to six QTL) and nonnormal noise. In the current study, the new proposed method did not identify any false positives, but was able to correctly identify all QTL in the model. The baseline approach of Pearson *et al*. (2007) still had a few false positives. Table 1 provides a snapshot of the results from the simulations created. The table illustrates the

Table 1. Summary results comparing proposed to baseline (Pearson *et al*. 2007), with variability in effect size and gamma noise.

Truth	Effect Size	Gamma Noise (α, β)	Baseline Results	Proposed Results
C1M2	1	0.5, 1	C1M2, **C1M5**	C1M2
		1, 3	C1M2	C1M2
C1M2	9	0.5, 1	**C1M1**, C1M2	C1M2
		1, 3	**C1M1**, C1M2	C1M2
C1M5, C2M15	2, 4	0.5, 1	**C1M4**, C1M5, **C2M14**, C2M15	C1M5, C2M15
		1, 3	**C1M4**, C1M5, **C2M14**, C2M15	C1M5, C2M15
C1M5, C2M15	6, 8	0.5, 1	**C1M4**, C1M5, **C2M14**, C2M15	C1M5, C2M15
		1, 3	C1M5, C2M15	C1M5, C2M15
C1M6, C2M15, C3M21	2, 4, 8	0.5, 1	C1M6, C2M15, C3M21	C1M6, C2M15, C3M21
		1, 3	C1M6, C2M15, **C2M16**, C3M21	C1M6, C2M15, C3M21
C1M2, C1M9, C2M15, C5M31	1, 3, 5, 9	0.5, 1	C1M2, **C1M3**, **C1M8**, C1M9, **C2M14**, C2M15, C5M31	C1M2, C1M9, C2M15, C5M31
		1, 3	C1M2, C1M9, C2M15, C5M31	C1M2, C1M9, C2M15, C5M31
C1M2, C1M5, C1M9, C2M15, C4M27, C5M33	1, 2, 5, 7, 8, 9	0.5, 1	C1M2, C1M5, C1M9, **C2M14**, C2M15, C4M27, C5M33	C1M2, C1M5, C1M9, C2M15, C4M27, C5M33
		1, 3	**C1M1**, C1M2, C1M5, C1M9, C2M15, C4M27, C5M33	C1M2, C1M5, C1M9, C2M15, C4M27, C5M33

effectiveness of the proposed approach in identifying QTL even when the noise is nonnormal and the model includes up to six QTL. An example can be found in the first row of results where the baseline finds two marker locations given gamma noise (0.5, 1) and effect size of 1, whereas the proposed correctly located the single marker.

A full disclosure of all experimental results is not provided due to size of the simulation study. However, the 60 simulations created similar results to the information provided in Table 1. Any interested reader may contact the corresponding author for the full detailed results.

4. COTYLEDON OPENING IN BAY-0 × SHAHDARA QTL EXPERIMENT

We apply the new methodology to a QTL experiment in Bay-0 × Shahdara Arabidopsis *thaliana* recombinant inbred line. The measured quantitative trait is the cotyledon opening, which is defined as the angle between the two preformed leaves (cotyledon) of Arabidopsis *thaliana* and is measured by a protractor. Values for the cotyledon opening can vary from 0° (no opening) to 180° (full opening). There are eight replicates from 165 recombinant lines placed on flats and exposed to ultraviolet radiation. Since not all seeds produce plants, the number of replicates within each line vary from zero to eight clones. The quantitative trait of cotyledon opening is the best available trait for the discovery of loci involved in ultraviolet radiation-B photoreceptor signal transduction and can aid in identifying the photoreceptor pathway.

Table 2. Activation probabilities for the top 10 genetic markers.

Marker	Activation Prob
ATHCHIB2	0.98302210
MSAT1.5	0.91681580
MSAT5.19	0.89501315
MSAT5.9	0.78504983
MSAT2.36	0.77669115
MSAT4.15	0.74650684
MSAT5.22	0.70556337
MSAT4.39	0.67210785
MSAT3.18	0.64656975
MSAT1.10	0.64496299

Table 2 presents the top ten activation probabilities from the Bayesian hierarchical model using the Markov chain Monte Carlo model composition search algorithm. Composite interval mapping for cotyledon opening (summarizing within lines using the median) in Bay × Sha QTL experiments found the following significant intervals F5I14-MSAT1.13, MSAT2.38-MSAT2.36, ATHCHIB2-MSAT3.19, MSAT4.9-MSAT4.37 and MSAT5.9-MSAT5.12. Similarities between the two methods include markers ATHCHIB2, MSAT5.9 and MSAT2.36, which are three of the five highest activation probabilities.

5. CONCLUSIONS

The identification of QTL can lead to many discoveries that impact society, e.g. identifying drought resistance for a plant or mass marker for cattle that can be used to breed meatier cattle. Data obtained through QTL experiments are complex and methodologies need to be able to handle the complexities in a robust manner; hence, we have presented an approach based on Bayesian hierarchical model that tolerates complexity well. The Bayesian hierarchical model has shown that it is a flexible model that can incorporate multi-levels of information, e.g. trait values for recombinant plant lines or environmental information or laboratory information, in an experiment. The proposed approach was able to correctly identify every QTL for each of the 60 simulations conducted. The algorithm appears to be robust to the assumption of normality, although further investigations might be needed.

The current methodology does not include potential epistasis among markers and/or environment in the model. Epistatic effects are of great interest and there is much work being done in this area. In addition, spatial dependencies among the plants in the experimental design are not taken into account in the current methodology. Future work will investigate expanding the hierarchical model to include epistasis among markers and environmental factors, as well as spatial dependencies among the plants.

REFERENCES

Ball, R.D. (2001). Bayesian methods for quantitative trait loci mapping based on model selection: Approximate analysis using the Bayesian information criterion. *Genetics*, **159**, 1351-1364.

Boone, E.L., Simmons, S.J., Ye, K. and Stapleton, A.E. (2006). Analyzing quantitative trait loci for the arabidopsis *thaliana* using Markov Chain Monte Carlo Model composition with restricted and unrestricted model spaces. *Statistical Methodology*, **3(1)**, 69-78.

Broman, K.W. and Speed, T.P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. Roy. Statist. Soc.,* **B64**, 641-656.

Buitenhuis, A.J., Rodenburg, T.B., Siwek, M., Cornelissen, S.J.B., Nieuwland, M.G.B., Crooijmans, R.P., Groenen, M.A., Koene, P., Bovenhuis, H. and van der Poel, J.J. (2005). Quantitative trait loci for behavioural traits in chickens. *Livestock Prod. Sci.*, **93(1)**, 95-103.

Gao, H. and Yang, R. (2006). Composite interval mapping of QTL for dynamic traits. *Chin. Sci. Bull.*, **51**, 1857-1862.

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

Jansen, R.C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics*, **135**, 205-211.

Jansen, R.C. and Stam, P. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, **136**, 1447-1455.

Jansen R.C. (1996). A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics*, **142**, 305-311.

Jiang, C. and Zeng, Z.B. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica*, **101**, 47-58.

Jing-hu, Z., Yuan-zhu, X., Bo, Z., Ming-gang, L., Feng-e, L. and Jia-lian, L. (2007). Detection of quantitative trait loci associated with live measurement traits in pigs. *Agric. Sci. China*, **6(7)**, 863-868.

Kao, C.H. and Zeng, Z.B. (1997). General formulae for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics*, **53**, 653-665.

Kao, C.H. (2000). On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics*, **156**, 855-865.

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

Liu, J., Mercer, J.M., Stam, L.F., Gibson, G.C., Zeng, Z.B. and Laurie, C.C. (1996). Genetic analysis of a morphological shape difference in the male genitalia of Drosophila simulans and D. mauritiana. *Genetics*, **142**, 1129-1145.

Loudet, Chaillou, Camilleri, Bouchez and Vedele (2002). Bay-0 × Shahdara recombinant inbred lines population: A powerful tool for the genetic dissection of complex traits in Arabidopsis. *Theo. Appl. Genet.*, **104(6-7)**, 1173-1184.

Luo, Z.W. and Kearsey, M.J. (1992). Interval mapping of quantitative trait loci in an F2 population. *Heredity*, **69**, 236¨C242.

Luo, Z.W. and Williams, J.A. (1993). Estimation of genetic parameters using linkage between a marker gene and a locus underlying a quantitative character in F2 populations. *Heredity*, **70**, 245-253.

Lynch M. and Walsh, B. (1997). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., Sunderland, MA.

Pearson, C., Simmons, S.J., Ricanek, K. and Boone, E.L. (2007). Comparative analysis of a hierarchical Bayesian method for quantitative trait loci analysis for the *Arabidopsis thaliana*. *Proc. Pattern Recognition in Bioinformatics*, *2007*, 60-70.

Satagopan, J.M., Yandell, B.S., Newton, M.A. and Osborn, T.C. (1996). Markov chain Monte Carlo approach to detect polygene loci for complex traits. *Genetics*, **144**, 805-816.

Sen, S. and Churchill, G.A. (2001). A statistical framework for quantitative trait mapping. *Genetics*, **159**, 371-387.

Séne, M., Causse, M., Damerval, C., Thévenot, C. and Prioul, J.L. (2000). Quantitative trait loci affecting amylose, amylopectin and starch content in maize recombinant inbred lines. *Plant Physiol. Biochem.*, **3(6)**, 459-472.

Sillanpaa, M. and Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics*, **148**, 1373-1388.

Sillanpaa, M.J. and Corander, J. (2002). Model choice in gene mapping: What and why. *Trends Genetics*, **18**, 301-307.

Weber, K., Eisman, R., Higgins, S., Kuhl, L., Patty, A., Sparks, J. and Zeng, Z.B. (1999). An analysis of polygenes affecting wing shape on chromosome three in Drosophila melanogaster. *Genetics*, **153**, 773-786.

Yandell, B.S., Mehta, T., Banerjee, S., Shriner, D., Venkataraman, R., Moon, J.Y., Neely, W.W., Wu, H., von Smith, R. and Yi, N. (2007). R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics*, **23**, 641-643.

Zeng, Z.B. (1994). Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457-1468.

Zeng, Z.B., Liu, J., Stam, L.F., Kao, C.H., Mercer, J.M. and Laurie, C.C. (2000). Genetic architecture of a morphological shape difference between two drosophila species. *Genetics*, **154**, 299-310.

Zeng, Z.B., Wang, T. and Zou, W. (2005). Modeling quantitative trait loci and interpretation of models. *Genetics*, **169**, 1711-1725.