# Mixed linear model approaches in mapping QTLs with epistatic effects by a simulation study

**Jan Bocianowski · Kamila Nowosad**

**Abstract** Epistasis refers to the phenotypic effects of interactions between alleles of different loci. Statistical detection of such interactions remains to be the subject of presented studies. A method based on mixed linear model was developed for mapping QTLs with digenic epistasis. Reliable estimates of QTL additive and epistasis effects can be obtained by the maximum-likelihood estimate methods. Likelihood ratio and $t$ statistics were combined for testing hypotheses about QTL effects (additive and epistasis effects). Monte Carlo simulations were conducted for evaluating the unbiasedness, accuracy and power for parameter estimation in QTL mapping. The results indicated that the mixed-model approaches could provide unbiased estimates for effects of QTLs. Additionally, the mixed-model approaches also showed high accuracy in mapping QTLs with epistasis effects. The information obtained in this study will be useful for manipulating the QTLs for plant breeding by marker assisted selection.

## Introduction

Quantitative traits are showing continuous variation in segregating populations. For a long time it has been assumed that quantitative traits are controlled by multiple genetic factors each having a small effect on the expression of the trait, known as the multiple factor hypothesis (East 1916). However, this hypothesis remained largely hypothetical for most of the last century as it was impossible to unravel the genetic basis of quantitative traits at the whole genome level using classical genetic methods. Recent advances in genome research involving a number of molecular-marker techniques and the availability of high-density molecular linkage maps, together with developments in analytical methods (Lander and Botstein 1989; Zeng 1994), facilitated the analysis of the genetic basis of quantitative traits at a single-locus level.

Main effects and epistatic effects of QTLs are important genetic components of quantitative traits. Epistasis refers to the phenotypic effects of interactions among alleles at multiple loci. Our current understanding of biochemical and physiological genetics, as well as the regulation of gene expression, strongly suggests the ubiquity of interactions among

J. Bocianowski (✉)
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland
e-mail: jboc@up.poznan.pl

K. Nowosad
Department of Genetics, Plant Breeding and Seed Production, Wrocław University of Environmental and Life Sciences, 53-363 Wrocław, Poland

gene products. There were also substantial interests in the classical quantitative genetics of epistasis, defined as the deviation from additivity of the effects between alleles of different loci (Cockerham 1954). Epistasis or the additive-by-additive interaction between loci on a genome that controls a quantitative trait is of great interest to geneticists (Eta-Ndu and Openshaw 1999; Gao and Zhu 2007; Kang et al. 2009; Zuo and Kang 2010; Binh et al. 2011; He et al. 2011). Strong interactions between QTLs have also been observed in maize (Lukens and Doebley 1999) and soybean (Lark et al. 1995). Recent genetic analyses using molecular markers in several plant species have clearly shown that, in addition to single locus QTLs, epistatic interactions play an important role on the genetic basis of quantitative traits (Lark et al. 1995; Maughan et al. 1996; Li et al. 1997; Yu et al. 1997; Poelwijk et al. 2011; Krajewski et al. 2012; Bocianowski 2012c, 2013, 2014).

In the previously paper (Bocianowski 2012a) was described the simulation study investigated a comparison between two estimation methods of the additive-by-additive (epistasis) effect: one, using only phenotypic data, and the other, which additionally takes into account the genotypic marker data. The main assumption was that the QTLs were located at marker positions. The shortcoming associated with the analyses of epistatic interactions published in previous studies (Bocianowski 2012a) is that the calculations were directly based on markers that are located at certain distances away from the QTLs involved in the epistases. The estimated effects are therefore biased depending on the distances between the marker loci and the QTLs. In this paper the method based on interval mapping was developed for mapping QTLs with additive and/or digenic epistasis effects. The analysis is based on a mixed linear-model approach, and put together the QTL main effects and digenic interactions that are possible with a two-locus data set in the same model (Wang et al. 1999).

A mixed model is a statistical model containing both fixed effects and random effects, that is mixed effects. These models are useful in a wide variety of disciplines in the biomedical, agricultural, physical, biological and social sciences (Parisseaux and Bernardo 2004; Yu et al. 2005b; Arbelbide et al. 2006; Aulchenko et al. 2007; Yang et al. 2007). They are particularly useful in settings where repeated measurements are made on the same statistical units, or where measurements are made on clusters of related statistical units. Because of their advantage to deal with missing values, mixed effects models are often preferred over more traditional approaches such as repeated measures analysis of variance. Mixed models can account for relationships among inbreds and for unbalanced data, and can incorporate marker data (Parisseaux and Bernardo 2004). A mixed-model procedure represents an in silico approach for gene mapping because it exploits phenotypic and genomic databases that are already available (Grupe et al. 2001).

The aim of the study reported in this paper was to compare two estimation methods for the parameters connected with the additive and additive-by-additive interaction gene action, i.e. the genotypic method, which is based on marker observations, and the phenotypic method, traditionally used in quantitative genetics, based on the only phenotypic observations. A mixed linear model approach was used to detect QTLs with main effects and QTLs involved in digenic interaction. The comparison was performed by the Monte Carlo simulation study.

## Materials and methods

### Plant material

For simplicity, we use a biparental homozygous (doubled haploid or recombinant inbred) lines population, from a cross between two homozygous lines. If in the experiment we observed $n$ significantly different plant lines, we obtain an $n$-vector of phenotypic mean observations $\mathbf{y} = [y_1 \ y_2 \ \ldots \ y_n]'$ and $q$ $n$-vectors of marker genotype observations $\mathbf{M}_l$, $l = 1, 2,\ldots, q$. The $i$-th element ($i = 1, 2,\ldots, n$) of vector $\mathbf{M}_l$ is equal to $-1$ or $1$, depending on the parent's genotype exhibited by the $i$-th line.

### Genetic models

In the first step of the selection, a mixed linear model for the simultaneous search of two interacting QTLs ($Q_i$ between flanking markers $M_{i-}$ and $M_{i+}$, and $Q_j$ between flanking markers $M_{j-}$ and $M_{j+}$) can be expressed as follows (Wang et al. 1999):

$$y_k = \mu + a_i x_{G_{ik}} + a_j x_{G_{jk}} + aa_{ij} x_{GG_{ijk}} + \sum_f u_{M_{fk}} e_{M_f}$$
$$+ \sum_l u_{MM_{lk}} e_{MM_l} + \varepsilon_k, \tag{1}$$

where $y_k$ denotes the phenotypic value of a quantitative trait measured on the $k$-th individual ($k = 1, 2,\ldots, n$); $\mu$ denotes the population mean; $a_i$ and $a_j$ denote the additive effects (fixed) of the two putative QTLs ($Q_i$ and $Q_j$) respectively; $aa_{ij}$ denotes the additive $\times$ additive epistatic effect (fixed) between $Q_i$ and $Q_j$; $x_{G_{ik}}, x_{G_{jk}}$ and $x_{GG_{ijk}}$ denote coefficients of QTL effects derived according to the observed genotypes of the markers ($M_{i-}, M_{i+}$ and $M_{j-}, M_{j+}$) and the test position (Wang et al. 1999); $e_{M_f} \sim N\left(0, \sigma_M^2\right)$ denotes the random effect of marker $f$ with indicator coefficient $u_{M_{fk}}$ ($-1$ for $m_f m_f$ and 1 for $M_f M_f$); $e_{MM_l} \sim N\left(0, \sigma_{MM}^2\right)$ denotes the random effect of the $l$-th marker interaction (between marker $K_l$ and marker $L_l$) with indicator coefficient $u_{MM_{lk}}$ ($-1$ for $M_K M_K m_L m_L$ or $m_K m_K M_L M_L$ and 1 for $M_K M_K M_L M_L$ or $m_K m_K m_L m_L$); and $\varepsilon_k \sim N\left(0, \sigma_\varepsilon^2\right)$ denotes the random residual effect.

The inclusion of $e_{M_f}$ and $e_{MM_l}$ in the model is intended to absorb the additive and epistatic effects of background QTLs (additional segregating QTLs other than the loci examined) for controlling the noise caused by these background QTLs.

Markers and marker pairs selected in the first step were subjected to backward stepwise selection in the second stage. The final model is as follows:

$$y_k = \mu + \sum_{t=1}^{p} a_t x_{G_{tk}} + \sum_{t=1}^{p-1} \sum_{\substack{t'=t+1 \\ t' \neq t}}^{p} aa_{tt'} x_{GG_{tt'k}}$$
$$+ \sum_{g=1}^{h} \sum_f u_{M_{fk}} e_{M_f} + \sum_{g'=1}^{h'} \sum_l u_{MM_{lk}} e_{MM_l} + \varepsilon_k, \tag{2}$$

where $p, g, g', h$ and $h' \in \{1,\ldots, q\}$. In the case of the large number of markers, selection of markers chosen for model (1) can be made independently inside all linkage groups; next, markers chosen in this way can be put in one group and subjected to the second step of selection. Model (2) can be written as a matrix form of the mixed linear model:

$$y = 1\mu + X\beta + Z\gamma + U_M e_M + U_{MM} e_{MM} + e_\varepsilon, \tag{3}$$

where $\mathbf{y}$ denotes $n$-dimensional vector of phenotypic values, $\mathbf{1}$ denotes the $n$-dimensional vector of ones, $\mu$ denotes the general mean, $\mathbf{X}$ denotes ($n \times p$)-dimensional matrix which columns are markers, $\boldsymbol{\beta}$ denotes the $p$-dimensional vector of unknown fixed effects of the form $\beta' = [\, a_1 a_2 \ldots a_p \,]$, $\mathbf{Z}$ denotes a matrix which columns are products of some columns of matrix $\mathbf{X}$, $\boldsymbol{\gamma}$ denotes the vector of unknown fixed effects of the form $\gamma' = [aa_{1,2}\, aa_{1,3} \ldots aa_{p-1,p}]$, $e_M \sim N\left(0, \sigma_M^2 R_M\right)$ denotes a random vector of marker effects, $e_{MM} \sim N\left(0, \sigma_{MM}^2 R_{MM}\right)$ denotes a random vector of interaction effects, $e_\varepsilon \sim N\left(0, \sigma_\varepsilon^2 I\right)$ denotes the $n$-dimensional vector of random variables such that $E(e_i) = 0$, $Var(e_i) = \sigma^2$, $Cov(e_i, e_j) = 0$ for $i \neq j$, $i, j = 1, 2,\ldots, n$, $\mathbf{U_M}$ and $\mathbf{U_{MM}}$ denote known incidence matrices, $\mathbf{R_M}$ and $\mathbf{R_{MM}}$ denote known symmetric matrices of incidence coefficients that can be obtained from the linkage relationships between the main-effect markers and between the pairs of interacting markers, respectively (Wang et al. 1999). The distribution of $\mathbf{y}$ is:

$$y \sim N\left(Gb, V\right), \tag{4}$$

where:

$$G = [\, 1 \quad X \quad Z \,], \tag{5}$$

$$b' = [\, \mu \quad \beta' \quad \gamma' \,], \tag{6}$$

$$V = \sigma_M^2 U_M R_M U_M' + \sigma_{MM}^2 U_{MM} R_{MM} U_{MM}' + \sigma_\varepsilon^2. \tag{7}$$

The likelihood function (L) for the parameters of effects $\mathbf{b}$ and variance components in model (1) is:

$$L\left(b, V\right) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y - Gb)' V^{-1}(y - Gb)\right], \tag{8}$$

When variance components of the model are known and if $\mathbf{G}$ is of full rank matrix, the estimate of $\boldsymbol{b}$ is given by (Searle 1982)

$$\hat{b} = \left(G' V^{-1} G\right)^{-1} G' V^{-1} y. \tag{9}$$

The total additive effect of genes influencing the trait ($a_g$) is defined as the sum of values of individual QTL effects. The total additive-by-additive epistasis effect of genes influencing the trait ($aa_g$) is defined as the sum of values of individual pairs' effects. The coefficient of determination were used to measure how

the model (2) fitted the data and, in this study, was the amount of the phenotypic variance explained by total QTLs with additive effects and QTL-pairs with epistatic effects ($R^2$).

## Phenotypic estimation

Estimation of the additive gene effect and additive-by-additive interaction of homozygous loci (epistasis) effect on the basis of phenotypic observations *y* requires identification of groups of extreme lines, i.e. lines with the minimal and maximal expression of the observed trait (Choo and Reinbergs 1982). The group of minimal lines consists of the lines which contain, theoretically, only alleles reducing the value of the trait. Analogously, the group of maximal lines contains the lines which have only alleles increasing the trait value. In this paper we identify the groups of extreme lines using the quantile method (Bocianowski et al. 1999), in which lines with the mean values smaller (bigger) than 0.03 (0.97) quantile of the empirical distribution of means are assumed as minimal (maximal) lines. The choose the quantiles 0.03 and 0.97 is results of previously study (Bocianowski et al. 1999). The total additive effect $a_p$ of all genes controlling the trait and the total additive-by-additive interaction effect $aa_p$ may be estimated by the formulas (Bocianowski and Krajewski 2009; Bocianowski 2012b)

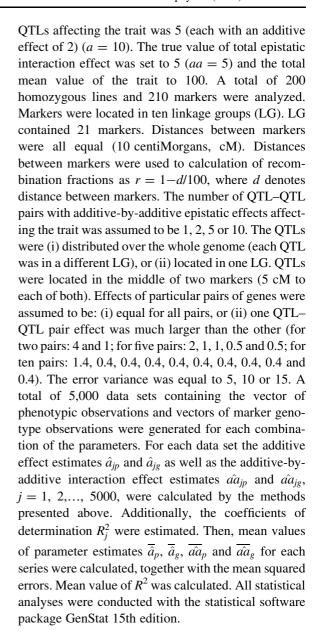$$\hat{a}_p = \frac{1}{2}\left(\overline{L}_{\max} - \overline{L}_{\min}\right) \qquad (10)$$

and

$$\hat{aa}_p = \frac{1}{2}\left(\overline{L}_{\max} + \overline{L}_{\min}\right) - \overline{L}, \qquad (11)$$

where $aa_p$ denotes total additive-by-additive interaction effect estimated on the basis of only phenotypes observations, $\overline{L}_{\min}$ and $\overline{L}_{\max}$ denote the means for the groups of minimal and maximal lines, respectively, and $\overline{L}$ denotes the mean for all lines.

## Simulation studies

In the Monte Carlo simulation studies comparing the "phenotypic" ($a_p$ and $aa_p$) and "genotypic" ($a_g$ and $aa_g$) estimates of the additive and additive-by-additive interaction of QTL effects the following variants of assumed parameters were adopted. The number of

QTLs affecting the trait was 5 (each with an additive effect of 2) ($a = 10$). The true value of total epistatic interaction effect was set to 5 ($aa = 5$) and the total mean value of the trait to 100. A total of 200 homozygous lines and 210 markers were analyzed. Markers were located in ten linkage groups (LG). LG contained 21 markers. Distances between markers were all equal (10 centiMorgans, cM). Distances between markers were used to calculation of recombination fractions as $r = 1 - d/100$, where $d$ denotes distance between markers. The number of QTL–QTL pairs with additive-by-additive epistatic effects affecting the trait was assumed to be 1, 2, 5 or 10. The QTLs were (i) distributed over the whole genome (each QTL was in a different LG), or (ii) located in one LG. QTLs were located in the middle of two markers (5 cM to each of both). Effects of particular pairs of genes were assumed to be: (i) equal for all pairs, or (ii) one QTL–QTL pair effect was much larger than the other (for two pairs: 4 and 1; for five pairs: 2, 1, 1, 0.5 and 0.5; for ten pairs: 1.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4, 0.4 and 0.4). The error variance was equal to 5, 10 or 15. A total of 5,000 data sets containing the vector of phenotypic observations and vectors of marker genotype observations were generated for each combination of the parameters. For each data set the additive effect estimates $\hat{a}_{jp}$ and $\hat{a}_{jg}$ as well as the additive-by-additive interaction effect estimates $\hat{aa}_{jp}$ and $\hat{aa}_{jg}$, $j = 1, 2, \ldots, 5000$, were calculated by the methods presented above. Additionally, the coefficients of determination $R_j^2$ were estimated. Then, mean values of parameter estimates $\overline{\hat{a}}_p$, $\overline{\hat{a}}_g$, $\overline{\hat{aa}}_p$ and $\overline{\hat{aa}}_g$ for each series were calculated, together with the mean squared errors. Mean value of $R^2$ was calculated. All statistical analyses were conducted with the statistical software package GenStat 15th edition.

## Results

Tables 1, 2 and 3 show results of simulation performed to compare the estimates of the additive and epistatic effects obtained by the genotypic and phenotypic methods with situations where error variances were equal to 5, 10 and 15, respectively. The phenotypic estimate of additive effect was less than 10, the true value, only for 10 QTL–QTL pairs. The genetic estimates of additive effects were bigger than

10 for 1, 2 and 5 pairs (except for five QTL–QTL pairs with the error variance equal to 5). The phenotypic estimates as well as the genotypic estimates of additive effects were bigger when the QTL–QTL pairs were located in one linkage group. The differences between the phenotypic estimate and the genotypic estimate of additive effects were always positive and were the smallest when the ten QTL–QTL pairs were assumed. Generally, differences between phenotypic and genotypic estimates were bigger when QTLs were located in many LG.

The phenotypic estimate of additive-by-additive effect was always bigger than 5, the true value (except for ten QTL–QTL pairs with unequal effects in many linkage groups). The genetic estimates of additive-by-additive epistatic effects were bigger than 5 for 1 and 2 pairs as well as for 5 pairs where error variance was equal to 10 and 15 (except for situation when five QTL–QTL pairs with unequal effects were located in the many linkage groups 5). The both (phenotypic and genotypic) estimates of additive-by-additive epistatic effects were bigger when the QTL–QTL pairs were located in one linkage group as well as when the QTL–QTL interaction effects were equal. The differences between the phenotypic estimate and the genotypic estimate of additive-by-additive effects were always positive (except only one situation: for two QTL–QTL

pairs with unequal effects located in the many linkage groups) and were the biggest when ten QTL–QTL pairs were assumed.

In general, a decrease of the estimates was accompanied by an increase of their mean squared error (Tables 1, 2, 3), because, larger estimates of additive effect as well as the additive-by-additive effect are more biased than smaller estimates. Variance explained by the QTLs with additive and epistatic effects ranged from 81 to 97 % (Tables 1, 2, 3).

# Discussion

There are several different strategies to map quantitative trait loci (Kearsey and Farquhar 1998), e.g., single-marker locus analysis (Liu 1998); simple interval mapping (Lander and Botstein 1989); composite interval mapping (Zeng 1993, 1994; Krajewski et al. 2012); marker regression (Kearsey and Hyne 1994; Wu and Li 1994; Bocianowski and Krajewski 2009); Bayesian methods (Sillanpää and Arjas 1998); and multiple interval mapping (Kao et al. 1999; Zeng et al. 1999). The latter methods have been shown to yield better power of QTL detection than interval mapping and single-marker locus analysis (Liu 1998; Piepho 2000). In this article we have demonstrated

**Table 1** Phenotypic and genotypic estimates of the total additive effect and the total additive-additive interaction effect obtained in the simulation study (error variance equal to 5)

| Number of QTL–QTL interaction effects | QTL–QTL interaction effects | Number of LG with QTL | Estimate | | | Mean squared error for | | Estimate | | | Mean squared error for | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\bar{\bar{a}}_p$ | $\bar{\bar{a}}_g$ | $\bar{\bar{a}}_p - \bar{\bar{a}}_g$ | $\hat{a}_p$ | $\hat{a}_g$ | $\overline{\overline{aa}}_p$ | $\overline{\overline{aa}}_g$ | $\overline{\overline{aa}}_p - \overline{\overline{aa}}_g$ | $\hat{aa}_p$ | $\hat{aa}_g$ | |
| 1 | 5 | 1 | 12.53 | 11.83 | 0.70 | 0.28 | 0.32 | 6.20 | 6.07 | 0.13 | 0.46 | 0.53 | 0.89 |
| 2 | Equal | 1 | 12.04 | 11.72 | 0.32 | 0.23 | 0.31 | 6.08 | 5.92 | 0.16 | 0.44 | 0.54 | 0.90 |
| 2 | Equal | Many | 11.93 | 11.60 | 0.33 | 0.42 | 0.57 | 5.93 | 5.76 | 0.17 | 0.48 | 0.52 | 0.87 |
| 2 | Unequal | 1 | 11.99 | 11.68 | 0.31 | 0.43 | 0.55 | 6.02 | 5.82 | 0.20 | 0.39 | 0.60 | 0.91 |
| 2 | Unequal | Many | 11.44 | 11.32 | 0.12 | 0.52 | 0.61 | 5.47 | 5.09 | 0.38 | 0.40 | 0.57 | 0.93 |
| 5 | Equal | 1 | 11.57 | 9.66 | 1.91 | 0.39 | 0.64 | 5.52 | 4.93 | 0.59 | 0.52 | 0.59 | 0.84 |
| 5 | Equal | Many | 11.24 | 9.39 | 1.85 | 0.42 | 0.58 | 5.29 | 4.85 | 0.44 | 0.61 | 0.59 | 0.90 |
| 5 | Unequal | 1 | 11.12 | 9.48 | 1.64 | 0.44 | 0.57 | 5.46 | 4.90 | 0.56 | 0.45 | 0.78 | 0.88 |
| 5 | Unequal | Many | 10.93 | 9.37 | 1.56 | 0.37 | 0.69 | 5.20 | 4.82 | 0.38 | 0.69 | 0.77 | 0.87 |
| 10 | Equal | 1 | 9.87 | 8.78 | 1.09 | 0.21 | 0.73 | 5.31 | 4.83 | 0.48 | 0.72 | 0.84 | 0.95 |
| 10 | Equal | Many | 9.70 | 8.69 | 1.01 | 0.42 | 0.80 | 5.09 | 4.69 | 0.40 | 0.79 | 0.90 | 0.94 |
| 10 | Unequal | 1 | 9.76 | 8.70 | 1.06 | 0.39 | 0.98 | 5.02 | 4.28 | 0.74 | 0.80 | 1.07 | 0.93 |
| 10 | Unequal | Many | 9.63 | 8.62 | 1.01 | 0.55 | 0.82 | 4.81 | 3.92 | 0.89 | 0.87 | 1.19 | 0.97 |

**Table 2** Phenotypic and genotypic estimates of the total additive effect and the total additive–additive interaction effect obtained in the simulation study (error variance equal to 10)

| Number of QTL–QTL interaction effects | QTL–QTL interaction effects | Number of LG with QTL | Estimate | | | Mean squared error for | | Estimate | | | Mean squared error for | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\bar{\hat{a}}_p$ | $\bar{\hat{a}}_g$ | $\bar{\hat{a}}_p - \bar{\hat{a}}_g$ | $\hat{a}_p$ | $\hat{a}_g$ | $\bar{\hat{a}\hat{a}}_p$ | $\bar{\hat{a}\hat{a}}_g$ | $\bar{\hat{a}\hat{a}}_p - \bar{\hat{a}\hat{a}}_g$ | $\hat{a}\hat{a}_p$ | $\hat{a}\hat{a}_g$ | |
| 1 | 5 | 1 | 12.66 | 11.93 | 0.73 | 0.26 | 0.32 | 6.33 | 6.11 | 0.22 | 0.39 | 0.52 | 0.94 |
| 2 | Equal | 1 | 12.38 | 11.80 | 0.58 | 0.31 | 0.41 | 6.13 | 6.02 | 0.11 | 0.40 | 0.63 | 0.86 |
| 2 | Equal | Many | 12.11 | 11.72 | 0.39 | 0.39 | 0.37 | 6.08 | 5.82 | 0.26 | 0.36 | 0.52 | 0.91 |
| 2 | Unequal | 1 | 12.08 | 11.71 | 0.37 | 0.49 | 0.35 | 6.05 | 5.90 | 0.15 | 0.41 | 0.49 | 0.93 |
| 2 | Unequal | Many | 11.91 | 10.97 | 0.94 | 0.58 | 0.62 | 5.61 | 5.70 | −0.09 | 0.53 | 0.57 | 0.84 |
| 5 | Equal | 1 | 11.92 | 10.29 | 1.63 | 0.60 | 0.69 | 5.88 | 5.56 | 0.32 | 0.52 | 0.59 | 0.86 |
| 5 | Equal | Many | 11.80 | 10.11 | 1.69 | 0.59 | 0.78 | 5.57 | 5.22 | 0.35 | 0.61 | 0.70 | 0.85 |
| 5 | Unequal | 1 | 11.63 | 10.02 | 1.61 | 0.53 | 0.62 | 5.70 | 5.08 | 0.62 | 0.49 | 0.71 | 0.86 |
| 5 | Unequal | Many | 11.09 | 9.67 | 1.42 | 0.53 | 0.70 | 5.46 | 4.92 | 0.54 | 0.52 | 0.66 | 0.88 |
| 10 | Equal | 1 | 10.01 | 9.71 | 0.30 | 0.72 | 0.88 | 5.61 | 4.97 | 0.64 | 0.68 | 0.79 | 0.90 |
| 10 | Equal | Many | 9.97 | 9.39 | 0.58 | 0.79 | 0.91 | 5.53 | 4.80 | 0.73 | 0.70 | 0.91 | 0.89 |
| 10 | Unequal | 1 | 9.98 | 9.50 | 0.48 | 0.98 | 1.07 | 5.29 | 4.32 | 0.97 | 0.59 | 0.98 | 0.95 |
| 10 | Unequal | Many | 9.73 | 9.01 | 0.72 | 0.92 | 1.11 | 4.71 | 4.01 | 0.70 | 0.91 | 1.15 | 0.91 |

**Table 3** Phenotypic and genotypic estimates of the total additive effect and the total additive–additive interaction effect obtained in the simulation study (error variance equal to 15)

| Number of QTL–QTL interaction effects | QTL–QTL interaction effects | Number of LG with QTL | Estimate | | | Mean squared error for | | Estimate | | | Mean squared error for | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\bar{\hat{a}}_p$ | $\bar{\hat{a}}_g$ | $\bar{\hat{a}}_p - \bar{\hat{a}}_g$ | $\hat{a}_p$ | $\hat{a}_g$ | $\bar{\hat{a}\hat{a}}_p$ | $\bar{\hat{a}\hat{a}}_g$ | $\bar{\hat{a}\hat{a}}_p - \bar{\hat{a}\hat{a}}_g$ | $\hat{a}\hat{a}_p$ | $\hat{a}\hat{a}_g$ | |
| 1 | 5 | 1 | 12.70 | 12.14 | 0.56 | 0.42 | 0.55 | 6.27 | 6.19 | 0.08 | 0.42 | 0.66 | 0.91 |
| 2 | Equal | 1 | 12.43 | 12.07 | 0.36 | 0.39 | 0.56 | 6.06 | 6.01 | 0.05 | 0.29 | 0.78 | 0.90 |
| 2 | Equal | Many | 12.15 | 11.82 | 0.33 | 0.43 | 0.48 | 6.18 | 5.93 | 0.25 | 0.38 | 0.44 | 0.86 |
| 2 | Unequal | 1 | 12.10 | 11.69 | 0.41 | 0.51 | 0.70 | 5.94 | 5.82 | 0.12 | 0.42 | 0.50 | 0.85 |
| 2 | Unequal | Many | 12.00 | 11.17 | 0.83 | 0.49 | 0.71 | 5.95 | 5.72 | 0.23 | 0.51 | 0.72 | 0.81 |
| 5 | Equal | 1 | 11.83 | 10.83 | 1.00 | 0.68 | 0.70 | 5.68 | 5.60 | 0.08 | 0.47 | 0.73 | 0.83 |
| 5 | Equal | Many | 11.70 | 10.55 | 1.15 | 0.40 | 0.80 | 5.53 | 5.38 | 0.15 | 0.46 | 0.82 | 0.91 |
| 5 | Unequal | 1 | 11.09 | 10.62 | 0.47 | 0.67 | 0.78 | 5.57 | 5.12 | 0.45 | 0.46 | 0.68 | 0.83 |
| 5 | Unequal | Many | 10.58 | 10.28 | 0.30 | 0.52 | 0.69 | 5.42 | 4.99 | 0.43 | 0.53 | 0.80 | 0.87 |
| 10 | Equal | 1 | 9.98 | 9.90 | 0.08 | 0.50 | 0.65 | 5.45 | 4.77 | 0.68 | 0.72 | 1.07 | 0.93 |
| 10 | Equal | Many | 9.90 | 9.72 | 0.18 | 0.89 | 0.72 | 5.24 | 4.81 | 0.43 | 0.82 | 0.93 | 0.95 |
| 10 | Unequal | 1 | 9.87 | 9.47 | 0.40 | 0.94 | 1.03 | 5.14 | 4.50 | 0.64 | 0.79 | 1.01 | 0.91 |
| 10 | Unequal | Many | 9.81 | 9.12 | 0.69 | 0.95 | 1.20 | 4.88 | 4.23 | 0.65 | 0.84 | 1.29 | 0.90 |

how to use mixed model for analysis of the main effects and epistatic effects of the QTLs. This study illustrates the ability of the analysis to assess the genetic components underlying the quantitative traits, and demonstrates the relative importance of the various components as the genetic basis of yield traits in aimed population. Understanding the genetic architecture of complex traits is a major challenge in the post-genomic era, especially for QTL-by-QTL interactions (Yang et al. 2007).

In the present study, a full-QTL model is proposed for modeling the genetic architecture of complex trait, which integrates the effects of multiple QTLs and epistasis into one mapping system. The most-

important results obtained from the simulation study show some stability of the properties of both methods of estimation over different types of genetic material. The lack of influence of error variance on estimation of additive as well as additive-by-additive gene action effects by both methods and on conclusions concerning the comparison of proposed methods of estimation, shows good prospect for application of our conclusions for different plant species. Moreover, the lack of influence of the number of linkage groups with QTL shows a possibility of using those methods for different genetic maps. In opposite, the number of QTL–QTL interaction effects influence on additive-by-additive gene action effects estimated by both methods and on conclusions concerning the comparison of proposed methods of estimation. Method presented in this paper may be preferred method of estimation of major and interacted QTLs for quantitative traits in bi-parental segregation population, because it provides results closer to the true values of total additive effects and total epistatic effects than previous methods based on fixed model (Bocianowski 2012b, c). The coefficients of determination of the proposed model are larger than obtained by using other methods: multiple interval mapping (Kao et al. 1999), penalized maximum likelihood (Zhang and Xu 2005).

Development of mixed linear model approaches and its application in quantitative genetics will create enormous challenges for quantitative geneticists in dealing with complicated genetic problems (Xu and Yi 2000). Applications of mixed models to association mapping and other genetic analyses in maize, wheat, *Arabidopsis*, and potato panels demonstrate that mixed models obtain fewer false positives and higher power than previous methods including genomic control, structured association, and principal component analysis (Yu et al. 2005a; Boer et al. 2007; Malosetti et al. 2007; Zhao et al. 2007; Zhang et al. 2008). Different mixed models have been proposed to map QTLs in complex pedigrees. Crepieux et al. (2004) proposed an identity by descent QTL mapping method using plant breeding data for self-pollinated crops. Crepieux et al. (2005) used this method to identify one QTL for kernel hardness and two QTLs for dough strength from data available in a wheat breeding program. The random model approach estimates a variance component associated with the QTL and identifies the marker interval that most likely

contains the QTL. This approach allows a better evaluation of the overall breeding value of an inbred and the identification of genomic regions associated with the trait (Arbelbide and Bernardo 2006).

A direct implication of epistasis, especially the involvement of QTLs in the epistatic interactions, is that the effects of the single-locus QTLs are mostly dependent on the genotypes of other loci, and, as can be seen from this analysis, the effect of a QTL can sometimes be negated by the genotypes of a second locus. Thus an attempt for utilization of the QTLs in the breeding programs has to taken into account for such epistatic effects. Epistatic effects have been considered to be important for complex traits by several researchers (Ma et al. 2005, 2007; Rebetzke et al. 2007; Krajewski et al. 2012). Determining the contribution of epistasis is important for understanding the genetic basis of complex traits. Hence, genetic models for QTL mapping assuming no epistasis can lead to a biased estimation of QTL parameters (Bocianowski 2013). A large number of epistatic effects have recently been detected in rice (*Oryza sativa* L.) using polymorphic markers in the whole genome (Hua et al. 2002; Mei et al. 2003, 2005). Epistatic effects have been found to be important in the expression of dough rheological properties in a wheat DH population (Ma et al. 2005).

Alternative for a mixed-model approach are: (1) Bayesian approach (Meuwissen et al. 2001; Xu 2003; Ter Braak et al. 2005), (2) penalized regression (Boer et al. 2002; Zhang and Xu 2005) and (3) the use of regularization paths (e.g., Hastie et al. 2001). However, results obtained on the basis of methods presented in this paper we show unbiased prediction of estimated parameters. We detected QTLs with additive effects and epistatic effects for quantitative trait using a homozygous lines population. The results showed that both additive effects and epistatic effects were important genetic bases of quantitative trait. The total QTLs with additive effects and epistatic effects explained more than 80 % of the phenotypic variation. The information obtained in this study will be useful for manipulating the QTLs for plant breeding by marker assisted selection.

# References

Arbelbide M, Bernardo R (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. Theor Appl Genet 112(5):885–890

Arbelbide M, Yu J, Bernardo R (2006) Power of mixed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. Theor Appl Genet 112(5):876–884

Aulchenko YS, de Koning DJ, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 177(1): 577–585

Binh TQ, Nakahori Y, Hien VTT, Khan NC, Lam NT, Mai LB, Yamamoto S (2011) Correlations between genetic variance and adiposity measures, and gene × gene interactions for obesity in postmenopausal Vietnamese women. J Genet 90:1–9

Bocianowski J (2012a) A comparison of two methods to estimate additive-by-additive interaction of QTL effects by a simulation study. J Theor Biol 308:20–24

Bocianowski J (2012b) Analytical and numerical comparisons of two methods of estimation of additive × additive interaction of QTL effects. Sci Agric 69(4):240–246

Bocianowski J (2012c) The use of weighted multiple linear regression to estimate QTL-by-QTL epistatic effects. Genet Mol Biol 35(4):802–809

Bocianowski J (2013) Epistasis interaction of QTL effects as a genetic parameter influencing estimation of the genetic additive effect. Genet Mol Biol 36(1):93–100

Bocianowski J (2014) Estimation of epistasis in doubled haploid barley populations considering interactions between all possible marker pairs. Euphytica 196:105–115

Bocianowski J, Krajewski P (2009) Comparison of the genetic additive effect estimators based on phenotypic observations and on molecular marker data. Euphytica 165: 113–122

Bocianowski J, Krajewski P, Kaczmarek Z (1999) Comparison of methods of choosing extreme doubled haploid lines for genetic parameter estimation. Colloq Biometryczne 29: 193–202

Boer MP, Ter Braak CJF, Jansen RC (2002) A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 162: 951–960

Boer MP, Wright D, Feng L, Podlich DW, Luo L, Cooper M, van Eeuwijk FA (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177(3): 1801–1813

Choo TM, Reinbergs E (1982) Estimation of the number of genes in doubled haploid populations of barley (Hordeum vulgare). Can J Genet Cytol 24:337–341

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. Genetics 39: 859–882

Crepieux S, Lebreton C, Servin B, Charmet G (2004) Quantitative trait loci (QTL) detection in multicross inbred designs: recovering QTL identical-by-descent status information from marker data. Genetics 168:1737–1749

Crepieux S, Lebreton C, Flament P, Charmet G (2005) Application of a new IBD-based mapping method to common wheat breeding population: analysis of kernel hardness and dough strength. Theor Appl Genet 111:1409–1419

East EM (1916) Studies on size inheritance in Nicotiana. Genetics 1(2):164–176

Eta-Ndu JT, Openshaw SJ (1999) Epistasis for grain field in two $F_2$ populations of maize. Crop Sci 39:346–352

Gao YM, Zhu J (2007) Mapping QTLs with digenic epistasis under multiple environmental and predicting heterosis based on QTL effects. Theor Appl Genet 115:325–333

Grupe A, Germer S, Usuka J, Aud D, Belknap JK, Klein RF, Ahluwalia MK, Higuchi R, Peltz G (2001) In silico mapping of complex disease-related traits in mice. Science 292:1915–1918

Hastie T, Tibshirani R, Friedman JH (2001) The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, Berlin

He XH, Qin H, Hu Z, Zhang T, Zhang YM (2011) Mapping of epistasis quantitative trait loci in four-way crosses. Theor Appl Genet 122:33–48

Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang QF (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885–1895

Kang ST, Kwak M, Kim HK, Choung MG, Han WY, Baek IY, Kim MY, Van K, Lee SH (2009) Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [Glycine max (L.) Merr.]. Euphytica 166:15–24

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152: 1203–1216

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Kearsey MJ, Hyne V (1994) QTL analysis: a simple 'marker regression' approach. Theor Appl Genet 89:698–702

Krajewski P, Bocianowski J, Gawłowska M, Kaczmarek Z, Pniewski T, Święcicki W, Wolko B (2012) QTL for field components and protein content: a multienvironmental study of two pea (Pisum sativum L.) populations. Euphytica 183:323–336

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

Lark KG, Chase K, Adelf F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Li ZK, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three grain yield components in rice. Genetics 145:453–465

Liu BH (1998) Statistical genomics. CRC Press, Boca Raton

Lukens L, Doebley J (1999) Molecular evolution of the teosinte branched gene among maize and related grasses. Mol Biol Evol 18(4):627–638

Ma W, Appels R, Bekes F, Larroque O, Morell MK, Gale KR (2005) Genetic characterisation of dough rheological properties in a wheat doubled haploid population: additive

genetic effects and epistatic interactions. Theor Appl Genet 111:410–422

Ma XQ, Tang JH, Teng WT, Yan JB, Meng YJ, Li JS (2007) Epistatic interaction is an important genetic basis of grain yield and its components in maize. Mol Breed 20:41–51

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. Genetics 175:879–889

Maughan PJ, Saghai Maroof MA, Buss GR (1996) Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor Appl Genet 93:574–579

Mei HW, Luo LJ, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li ZK (2003) Gene actions of QTL affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor Appl Genet 107:89–101

Mei HW, Luo LJ, Ying CS, Wang YP, Yu XQ, Guo LB, Paterson AH, Li ZK (2005) Gene actions of QTL affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110:649–659

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514

Piepho HP (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156(4):2043–2050

Poelwijk FJ, Tănase-Nicola S, Kiviet DJ, Tans SJ (2011) Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J Theor Biol 272:141–144

Rebetzke GJ, Ellis MH, Bonnett DG, Richards RA (2007) Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 114:1173–1183

Searle SP (1982) Matrix algebra useful for statistics. Wiley, New York

Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148:1373–1388

Ter Braak CJF, Boer MP, Bink MCAM (2005) Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics 170:1435–1438

Wang DL, Zhu J, Li ZK, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environmental interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Wu WR, Li WM (1994) A new approach for mapping quantitative trait loci using complete genetic marker linkage maps. Theor Appl Genet 89:535–539

Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801

Xu S, Yi N (2000) Mixed model analysis of quantitative trait loci. Proc Natl Acad Sci USA 97(26):14542–14547

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23(12):1527–1536

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Saghai Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

Yu J, Arbelbide M, Bernardo R (2005a) Power of in silico QTL mapping from phenotypic, pedigree and marker data in a hybrid breeding program. Theor Appl Genet 110:1061–1067

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005b) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

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

Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res 74:279–289

Zhang YM, Xu S (2005) A penalized maximum likelihood method for estimating epistatic effects of QTL. Heredity 95:96–104

Zhang K, Tian J, Zhao L, Wang S (2008) Mapping QTLs with epistatic effects and QTL × environment interactions for plant height using a doubled haploid population in cultivated wheat. J Genet Genomics 35(2):119–127

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3:e4

Zuo Y, Kang G (2010) A mixed two-stage methods for detecting interactions in genomewide association studies. J Theor Biol 262:576–583