



## The use of weighted multiple linear regression to estimate QTL-by-QTL epistatic effects

Jan Bocianowski

*Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego, Poznań, Poland.*

### Abstract

Knowledge of the nature and magnitude of gene effects, as well as their contribution to the control of metric traits, is important in formulating efficient breeding programs for the improvement of plant genetics. Information concerning a genetic parameter such as the additive-by-additive epistatic effect can be useful in traditional breeding. This report describes the results obtained by applying weighted multiple linear regression to estimate the parameter connected with an additive-by-additive epistatic interaction. Three weight variants were used: (1) standard weights based on estimated variances, (2) different weights for minimal, maximal and other lines, and (3) different weights for extreme and other lines. The approach described here combines two methods of estimation, one based on phenotypic observations and the other using molecular marker data. The comparison was done using Monte Carlo simulations. The results show that the application of weighted regression to the marker data yielded estimates similar to those obtained by phenotypic methods.

*Keywords:* biometrical genetics, epistatic interaction, homozygous lines, QTL, simulation study, weighted regression.

Received: February 2, 2012; Accepted: June 29, 2012.

### Introduction

Knowledge of the nature and magnitude of gene effects, as well as their contribution to the control of metric traits, is important in formulating efficient breeding programs for plant genetic improvement. The inheritance of quantitative traits has been described as a “moving target” since these traits are affected not only by the actions of multiple individual genes but also by the interaction between genes (Lukens and Doebley, 1999). In this context, information on additive-by-additive interaction epistatic effects can be useful in traditional breeding programs (Sharmila *et al.*, 2007; Bnejdi and El Gazzah, 2010; Bnejdi *et al.*, 2010; da Silva Guimarães *et al.*, 2010; Lau and Muniandy, 2012).

Epistasis is a phenomenon in which the effects of one gene are modified by one or several other genes. The gene whose phenotype is expressed is referred to as epistatic whereas the phenotype that is altered or suppressed is referred to as hypostatic. Epistasis can be contrasted with dominance, which involves an interaction between alleles at the same gene locus, or with an additive effect, which involves an interaction between alleles at the other gene locus.

The term epistasis was first described by Bateson (1907) to denote the suppression of gene expression at one

locus by a gene at another locus. Fisher (1918) used the term epistacy to refer to a statistical interaction in the sense of a deviation from additive effects in the statistical linear model. Cordell (2002, 2009) has argued that the statistical tests that are often used to assess interactions (Hahn *et al.*, 2003; Moore, 2004; Chung *et al.*, 2007; Zhang and Liu, 2007; Gayán *et al.*, 2008) are of limited use in elucidating the type of biological interaction that Bateson had originally conceived. Phillips (2008) discussed the ambiguity in the term epistasis and distinguished what he considered to be three forms of epistasis, *i.e.*, 1) statistical epistasis, which describes a departure from additive effects in a statistical model, 2) compositional epistasis, which corresponds to epistasis in Bateson’s original sense of the term, *i.e.*, the masking of the effect of an allele at one locus by an allele at another locus, and 3) functional epistasis, which describes physical molecular interactions between various proteins (and other genetic elements). Jannink and Jansen (2001) described a contrast approach for building linear models and a likelihood method to test for the presence of epistasis.

Previous reports (Bocianowski, 2008, 2012a,b) described results obtained for analytical, numerical and simulation comparisons of two methods for estimating the parameter connected with the additive-by-additive interaction effect (epistasis), namely, the phenotypic method, based on extreme groups of homozygous lines, and the genotypic

method, based on marker observations. One of the conclusions of these studies was that the estimate of the total additive-by-additive interaction effect based on the marker data was in most cases smaller than that based on the phenotype.

The explanation of this phenomenon may be simple. Phenotypic data can be used to estimate only the total epistatic effect of all hypothetical gene pairs that determine the trait. By using marker data, which can be more or less precisely mapped in the genome, we can estimate individual effects of gene-by-gene interactions while keeping the number of QTL-by-QTL interactions intentionally low for practical reasons. The sum of the obtained effects is lower than the phenotypic estimate and this difference may be enhanced by the absence of markers in the regions where genes are located.

In addition to the above explanation, other possible sources of the differences between calculated estimates should be considered. The results mentioned above relate to QTL-by-QTL interaction effects obtained by the simplest possible methods, *i.e.*, by multiple linear regression on the marker data. In the present paper, modification of this regression by using empirical weights is shown to provide better agreement between the phenotypic and genotypic estimates.

In the approach described here, estimation of the parameter connected with the additive-by-additive epistatic interaction effect was based on extreme groups of homozygous lines and on data for genotypic markers. Weighted multiple linear regression was used to estimate the QTL-by-QTL epistatic effects by employing: (1) standard weighted regression with weights based on estimated variances, (2) different weights for minimal, maximal and other lines, and (3) different weights for extreme lines and different weights for other lines. The effectiveness of the proposed method was investigated by using Monte Carlo simulations.

## Methods

### Estimation of the effect of additive-by-additive interactions on gene action

As a starting point, consider a population of  $n$  significantly different homozygous (doubled haploid - DH or recombinant inbred - RI) plant lines from a cross between two homozygous parents, as is currently practiced in the breeding of selfing species. For this population, an  $n$ -vector of phenotypic mean observations  $\mathbf{y} = [y_1, y_2, \dots, y_n]'$  and  $q$   $n$ -vectors of marker genotype observations  $\mathbf{m}_l$ , where  $l = 1, 2, \dots, q$ , are obtained. The  $i$ -th element ( $i = 1, 2, \dots, n$ ) of vector  $\mathbf{m}_l$  is equal to -1 or 1, depending on the parental genotype exhibited by the  $i$ -th line.

### Estimation based on the phenotype

Estimation of the additive-by-additive interaction of homozygous loci  $aa$  (epistasis) (Kearsey and Pooni, 1996)

based on phenotypic observations  $y$  requires the identification of groups of extreme lines, *i.e.*, lines with minimal and maximal expression of the observed trait (Choo and Reinbergs, 1982). The group of minimal lines consists of lines that theoretically contain only alleles that reduce the value of the trait. Analogously, the group of maximal lines contains lines connected only with alleles that increase the trait value. In this report, the groups of extreme lines are identified by using the quantile method (Bocianowski *et al.*, 1999) in which lines with mean values  $\leq 0.03$  or  $\geq 0.97$  of the quantile corresponding to the empirical distribution of means were assumed to represent minimal and maximal lines, respectively. The total additive-by-additive interaction effect of  $aa$  may be estimated by the following formula (Surma *et al.*, 1984)

$$\hat{aa}_f = \frac{1}{2} (\bar{L}_{\max} + \bar{L}_{\min}) - \bar{L} \quad (1)$$

where  $\bar{L}_{\max}$  and  $\bar{L}_{\min}$  denote the means for the groups of maximal and minimal lines, respectively, and  $\bar{L}$  denotes the mean for all lines.

### Estimation based on the genotype

The estimation of  $aa$  was based on the assumption that genes responsible for the trait were marked perfectly as observed molecular markers. By choosing from all the observed markers  $p$  the variability of the trait and model observations for the lines can be defined as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{e} \quad (2)$$

where  $\mathbf{1}$  is the  $n$ -dimensional vector of ones,  $\mu$  is the general mean,  $\mathbf{X}$  is the  $(n \times p)$ -dimensional matrix of the form  $\mathbf{X} = [\mathbf{m}_{l_1} \quad \mathbf{m}_{l_2} \quad \dots \quad \mathbf{m}_{l_p}]$ ,  $l_1, l_2, \dots, l_p \in \{1, 2, \dots, q\}$ ,  $\beta$  is the  $p$ -dimensional vector of unknown parameters of the form  $\beta' = [a_{l_1} \quad a_{l_2} \quad \dots \quad a_{l_p}]$ ,  $\mathbf{Z}$  is a matrix in which the columns are the products of some columns of matrix  $\mathbf{X}$ ,  $\gamma$  is the vector of unknown parameters of the form  $\gamma' = [aa_{l_1 l_2} \quad aa_{l_1 l_3} \quad \dots \quad aa_{l_{p-1} l_p}]$  and  $\mathbf{e}$  is the  $n$ -dimensional vector of random variables such that  $E(e_i) = 0$ ,  $Var(e_i) = \sigma^2$  and  $Cov(e_i, e_j) = 0$  for  $i \neq j$  with  $i, j = 1, 2, \dots, n$ . The parameters  $a_{l_1}, a_{l_2}, \dots, a_{l_p}$  are the additive effects of the genes controlling the trait and parameters  $aa_{l_1 l_2}, aa_{l_1 l_3}, \dots, aa_{l_{p-1} l_p}$  are the additive-by-additive interaction effects. The epistatic interaction effects are assumed to show only loci with significant additive gene action effects. This assumption significantly reduces the number of potential significant effects and improves the usefulness of the regression model.

Defining  $\alpha' = [\mu \quad \beta' \quad \gamma']$  and  $\mathbf{G} = [\mathbf{1} \quad \mathbf{X} \quad \mathbf{Z}]$  yields the model

$$\mathbf{y} = \mathbf{G}\alpha + \mathbf{e} \quad (3)$$

If  $\mathbf{G}$  is of full rank, the estimate of  $\alpha$  is given by (Searle, 1982):

$$\hat{\alpha} = (\mathbf{G}' \mathbf{W}^{-1} \mathbf{G})^{-1} \mathbf{G}' \mathbf{W}^{-1} \mathbf{y} \quad (4)$$

where  $\mathbf{W}^{-1}$  is a diagonal matrix of unknown variances of observations that may be found empirically by estimation. The total additive-by-additive epistatic effect of genes that influence the trait is defined as:

$$\hat{aa}_f = \sum_{k=1}^{p-1} \sum_{\substack{k'=2 \\ k' \neq k}}^p \hat{aa}_{l_k l_{k'}} \quad (5)$$

The markers for model (2) may be selected by a stepwise regression procedure (Charcosset *et al.*, 2001). For this, a two-stage algorithm was used in which: (i) selection was done by using a forward-backward stepwise, multiple linear regression with a probability into and out of the model of 0.01 independently within all linkage groups, and (ii) markers selected in this way were placed in one group and subjected to the second backward selection (see Jansen and Stam, 1994). The final genetic model incorporated significant additive and epistatic effects. A QTL was declared if the phenotype was associated with a marker locus at  $p < 0.001$ . For the weights, three forms of matrix  $\mathbf{W}$  were proposed, as follows:

Proposal 1 - Ordinary weighted multiple linear regression, with weights based on the estimated variance. The matrix  $\mathbf{W}$  is defined as:

$$\mathbf{W} = (w_i) \quad (6)$$

where  $w_i$  is the estimated variance for the  $i$ -th line,  $i = 1, 2, \dots, n$ .

Proposal 2 - Weighted regression with different weights for minimal, maximal and other lines. The elements of matrix  $\mathbf{W}$  are:

$$w_i = \begin{cases} w_{\min}, & \text{for minimal lines,} \\ w_{\max}, & \text{for maximal lines,} \\ w_{\text{other}}, & \text{for other lines,} \end{cases} \quad (7)$$

where  $w_{\min}$  is the estimated average of variances for minimal lines,  $w_{\max}$  is the average of variances for maximal lines and  $w_{\text{other}}$  is the average of variances for other lines.

Proposal 3 - Weighted multiple linear regression with different weights for extreme lines and other lines. The elements of matrix  $\mathbf{W}$  are:

$$w_i = \begin{cases} 1, & \text{for extreme lines,} \\ \frac{1}{c}, & \text{for other lines,} \end{cases} \quad (8)$$

where

$$c = \frac{w_{\min} + w_{\max}}{2 \cdot w_{\text{other}}} \quad (9)$$

In the case of unweighted regression all the elements of matrix  $\mathbf{W}$  are equal to 1.

## Simulation studies

In the Monte Carlo simulations that compared the phenotypic and genotypic (unweighted and three forms of weighted) estimates of the additive-by-additive interaction of QTL effects the following variants of the selected parameters were adopted. The true value of the parameter was assumed to be equal to five ( $aa = 5$ ) and the total mean value of the trait to 100. The number of QTLs affecting the trait was five (each with an additive effect of two). A QTL was marked perfectly so that the marker and QTLs were in complete linkage disequilibrium with no recombination. At least two markers without additive effects were located between two QTLs. In all, 200 homozygous lines and 210 markers were analyzed. Markers were located in five, seven or 10 linkage groups (LG) that contained 42, 30 or 21 markers, respectively. Distances between markers were all equal (10 cM) or unequal (for five LG: 8, 9, 10, 11 and 12 cM; for seven LG: 8, 9, 9.5, 10, 10.5, 11 and 12 cM; for 10 LG: 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12 and 12.5 cM). The number of QTL-QTL pairs with additive-by-additive epistatic effects that affected the trait was assumed to be one, two or five. The QTLs with additive-by-additive epistatic effects were located in one LG or distributed over the whole genome (each QTL was in a different LG). The effects of particular pairs of genes were assumed to be equal for all pairs or 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). The variance for lines was assumed to include four variants: V1 - equal for all lines (10), V2 - greater for extreme lines (10) than for other lines (5), V3 - smaller for extreme lines (5) than for other lines (10) and V4 - different for minimal lines (10), maximal lines (20) and other lines (30). The error variance was equal to five for all generated observations. For each combination of parameters, 5000 random generations were obtained for the data set containing the vector of phenotypic observations and vectors of marker genotype observations. For each data set the line variances were estimated and the total additive-by-additive interaction effect was estimated by the phenotypic method ( $\hat{aa}_f$ ) and by unweighted ( $\hat{aa}_{g_0}$ ) and weighted ( $\hat{aa}_{g_1}, \hat{aa}_{g_2}, \hat{aa}_{g_3}$ ) versions of the genotypic method. The results were summarized as mean values for a series of simulations and mean squared errors.

## Results

Different experimental situations were examined in this large-scale simulation. Tables 1, 2, 3 and 4 show the results of simulations done to compare the estimates of addi-

tive-by-additive epistatic interactions obtained by the five methods: the phenotypic method and four genotypic methods (one unweighted and three forms of weighted) in which the distances between markers were the same (10 cM) and the number of linkage groups was five.

The results of this simulation showed that in all the four variants of variance for lines [equal for all lines (Table 1), greater for extreme lines than for other lines (Table 2), smaller for extreme lines than for other lines (Table 3) and different for minimal lines, maximal lines and other lines (Table 4)] the mean phenotypic estimate was always greater than the genotypic estimates (regardless of the method used). Genotypic estimates based on weighted regression were always greater than the unweighted estimates. The mean values for  $\widehat{aa}_{g_2}$  and  $\widehat{aa}_{g_3}$  were always slightly larger than  $\widehat{aa}_{g_0}$ , whereas the mean values for  $\widehat{aa}_{g_1}$  were much larger than  $\widehat{aa}_{g_0}$  and closest (among  $\widehat{aa}_g$ ) to  $\widehat{aa}_f$ . The relationship  $\widehat{aa}_f > \widehat{aa}_{g_1} > \widehat{aa}_{g_2} > \widehat{aa}_{g_3} > \widehat{aa}_{g_0}$  was generally observed, except when five QTL-QTL epistatic pairs were assumed to be present in many linkage groups, in which case  $\widehat{aa}_{g_1} > \widehat{aa}_f$ . The weighted estimates were always closer to the phenotypic estimates than the unweighted estimates.

The last five columns of Tables 1-4 show that a decrease in the estimates was accompanied by an increase in their mean squared error. Simulations were also done for the groups containing seven and ten LG, and for unequal distances between markers. The results of these analyses (not shown) were similar to those described above.

### Discussion

Most plant traits are quantitative in nature and are influenced by many genes of quantitative trait loci (QTL). Quantitative traits are also influenced by the environment and tend to show varying degrees of genotype × environment interaction. Epistasis, or interaction between non-allelic genes, is an important factor that affects the phenotypic expression of genes and the genetic variation in populations (Li *et al.*, 1997). An increasing body of knowledge on biological pathways and gene networks implies that gene-gene interactions (epistasis) are important and several reports have recently argued that much genetic variance in populations is due to such interactions (Carlborg and Haley, 2004; Marchini *et al.*, 2005; Nadeem and Azhar, 2005; Bhatti *et al.*, 2006a,b; Evans *et al.*, 2006; Lin *et al.*, 2008; Ullah *et al.*, 2010; Shakoor *et al.*, 2010). Mapping and estimating epistatic quantitative trait loci is a difficult and serious challenge and most of the existing methods for the detection of epistasis use a higher dimensional search (Chase *et al.*, 1997; Holland, 1998; Kao *et al.*, 1999; Zeng *et al.*, 1999; Carlborg *et al.*, 2000; Sen and Churchill, 2001).

**Table 1** - Phenotypic and genotypic estimates of the total additive-by-additive interaction epistatic effect obtained in the simulation study (five linkage groups, all distances between markers of 10 cM, variance was equal for all lines - variant VI).

Number of QTL-QTL interaction effects	QTL-QTL interaction effects	Number of LG with QTL	Estimate					Mean squared error for				
			$\widehat{aa}_f$	$\widehat{aa}_{g_0}$	$\widehat{aa}_{g_1}$	$\widehat{aa}_{g_2}$	$\widehat{aa}_{g_3}$	$\widehat{aa}_f$	$\widehat{aa}_{g_0}$	$\widehat{aa}_{g_1}$	$\widehat{aa}_{g_2}$	$\widehat{aa}_{g_3}$
1	5	1	6.79	5.32	6.21	5.37	5.35	0.41	0.69	0.55	0.60	0.62
2	Equal	1	6.30	5.30	6.09	5.37	5.33	0.42	0.72	0.59	0.68	0.69
2	Equal	Many	5.51	4.93	5.30	5.06	5.01	0.52	0.85	0.66	0.71	0.77
2	Unequal	1	6.22	5.28	5.79	5.33	5.29	0.44	0.73	0.58	0.65	0.70
2	Unequal	Many	5.63	4.80	5.23	4.99	4.85	0.54	0.89	0.67	0.79	0.83
5	Equal	1	5.29	5.08	5.20	5.11	5.09	0.43	0.70	0.52	0.61	0.66
5	Equal	Many	4.71	4.14	4.99	4.32	4.27	0.55	0.92	0.69	0.83	0.90
5	Unequal	1	5.31	5.03	5.18	5.07	5.06	0.47	0.79	0.58	0.65	0.71
5	Unequal	Many	4.68	4.02	4.81	4.20	4.11	0.59	1.01	0.82	0.90	0.99

**Table 2** - Phenotypic and genotypic estimates of the total additive-by-additive interaction epistatic effect obtained in the simulation study (five linkage groups, all distances between markers of 10 cM, variance was greater for extreme lines than for other lines - variant V2).

Number of QTL-QTL interaction effects	QTL-QTL interaction effects	Number of LG with QTL	Estimate					Mean squared error for				
			$\widehat{aa_f}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_3}}$	$\widehat{aa_f}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_3}}$
1	5	1	6.77	5.33	6.60	5.36	5.35	0.42	0.63	0.49	0.57	0.59
2	Equal	1	6.32	5.28	6.01	5.39	5.30	0.41	0.69	0.53	0.60	0.66
2	Equal	Many	5.49	4.94	5.22	5.08	4.99	0.48	0.71	0.52	0.63	0.69
2	Unequal	1	6.17	5.28	6.06	5.44	5.31	0.47	0.72	0.50	0.68	0.70
2	Unequal	Many	5.60	4.81	5.43	4.97	4.88	0.52	0.93	0.72	0.85	0.89
5	Equal	1	5.29	5.02	5.20	5.09	5.04	0.44	0.79	0.62	0.67	0.71
5	Equal	Many	4.79	4.09	5.03	4.29	4.21	0.55	0.91	0.60	0.77	0.82
5	Unequal	1	5.30	5.11	5.22	5.18	5.12	0.48	0.80	0.55	0.69	0.73
5	Unequal	Many	4.71	4.00	4.98	4.44	4.37	0.61	0.97	0.77	0.85	0.90

**Table 3** - Phenotypic and genotypic estimates of the total additive-by-additive interaction epistatic effect obtained in the simulation study (five linkage groups, all distances between markers of 10 cM, variance was smaller for extreme lines than for other lines - variant V3).

Number of QTL-QTL interaction effects	QTL-QTL interaction effects	Number of LG with QTL	Estimate					Mean squared error for				
			$\widehat{aa_f}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_3}}$	$\widehat{aa_f}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_3}}$
1	5	1	6.76	5.27	6.61	5.34	5.30	0.39	0.55	0.42	0.49	0.51
2	Equal	1	6.30	5.26	6.17	5.37	5.27	0.43	0.62	0.49	0.57	0.60
2	Equal	Many	5.48	4.91	5.30	5.11	4.98	0.46	0.66	0.51	0.62	0.63
2	Unequal	1	6.19	5.25	5.98	5.37	5.29	0.45	0.82	0.50	0.76	0.81
2	Unequal	Many	5.61	4.94	5.37	5.08	5.01	0.49	0.65	0.53	0.61	0.62
5	Equal	1	5.22	5.07	5.14	5.11	5.08	0.40	0.59	0.49	0.54	0.57
5	Equal	Many	4.77	4.21	4.95	4.42	4.30	0.56	0.95	0.69	0.79	0.83
5	Unequal	1	5.21	5.09	5.17	5.13	5.12	0.48	0.75	0.52	0.60	0.63
5	Unequal	Many	4.88	4.19	4.92	4.25	4.22	0.57	0.99	0.60	0.77	0.80



**Table 4** - Phenotypic and genotypic estimates of the total additive-by-additive interaction epistasis effect obtained in the simulation study (five linkage groups, all distances between markers of 10 cM, variance was different for minimal lines, maximal lines and other lines - variant V4).

Number of QTL-QTL interaction effects	QTL-QTL interaction effects	Number of LG with QTL	Estimate				Mean squared error for					
			$\widehat{aa_f}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_0}}$	$\widehat{aa_{g_1}}$	$\widehat{aa_{g_2}}$	$\widehat{aa_{g_3}}$		
1	5	1	6.82	5.39	6.68	6.01	5.95	0.32	0.61	0.40	0.55	0.59
2	Equal	1	6.34	5.31	6.21	5.50	5.44	0.39	0.59	0.44	0.57	0.58
2	Equal	Many	5.48	4.93	5.29	5.01	4.97	0.42	0.73	0.53	0.61	0.69
2	Unequal	1	6.25	5.26	6.00	5.41	5.39	0.38	0.63	0.49	0.56	0.59
2	Unequal	Many	5.51	4.85	5.23	4.97	4.91	0.49	0.72	0.58	0.61	0.66
5	Equal	1	5.32	5.07	5.25	5.12	5.08	0.40	0.69	0.49	0.53	0.60
5	Equal	Many	4.70	4.15	4.91	4.25	4.20	0.50	0.91	0.54	0.62	0.71
5	Unequal	1	5.28	5.10	5.20	5.14	5.12	0.45	0.86	0.61	0.73	0.77
5	Unequal	Many	4.51	3.98	4.83	4.17	4.11	0.57	1.08	0.70	0.79	0.83

The present report is the first to compare a phenotypic estimator of additive-by-additive epistatic interaction effect with genotypic estimators obtained using weighted multiple linear regression based on a simulation study. The combination of parameters adopted in the Monte Carlo simulation study corresponded to cases often encountered in real studies.

Quantitative geneticists have long recognized the importance of gene-by-gene interactions and these have been documented for numerous crops and for various traits. Information on epistasis related to additive effects will be useful in traditional breeding. In breeding programs the use of a superior genotype in various environments when this superiority has been predicted based on QTL information obtained only in one environment (Lin *et al.*, 2008). Epistasis is a challenge to plant breeders and can reduce the progress of quantitative traits from selection. Hence, in genetic and breeding studies, it is very important to consider the non-allelic interaction effect as well as the precision of their estimation.

This paper describes the use of weighted multiple linear regression to estimate QTL-by-QTL epistatic effects, with three forms of weighting being proposed. The effectiveness of the proposed methods was validated by a series of simulations. The results confirmed the universality and stability of additive-by-additive epistatic interaction effects assessed by the weighted regression method. Similar relationships between phenotypic and genotypic estimators were obtained, irrespective of the number of linkage groups (number of chromosomes), distance between markers, the number and location of QTL-QTL epistatic pairs and the different variants of line variances.

**Conclusion**

Multiple linear weighted regression is useful for estimating additive-by-additive epistatic interactions. Irrespective of the variants of line variances, similar relationships were obtained between  $\widehat{aa_f}$ ,  $\widehat{aa_{g_0}}$ ,  $\widehat{aa_{g_1}}$ ,  $\widehat{aa_{g_2}}$  and  $\widehat{aa_{g_3}}$ . This finding indicates that these methods of estimation may be applied to different experimental situations. The ordinary weighted multiple linear regression method, with weights based on an estimated variance, is the preferred method because it provides results closer to the true values. Overall, the results described here indicate that the estimation of epistatic interaction effects by the weighted regression method may be applied to different plant species.

**References**

Bateson W (1907) Facts limiting the theory of heredity. Science 26:649-660.  
 Bhatti MA, Azhar FM and Alvi AW (2006a) Estimation of additive, dominance and epistatic components of genetic varia-

- tion in fibre quality characters of up-land cotton grown in salinized conditions. *Int J Agric Biol* 8:824-827.
- Bhatti MA, Azhar FM, Alvi AW and Ayub M (2006b) Triple test cross analysis of seed cotton (*Gossypium hirsutum* L.) yield and its components grown in salinized conditions. *Int J Agric Biol* 8:820-823.
- Bnejdi F and El Gazzah M (2010) Epistasis and genotype-by-environment interaction of grain protein content in durum wheat. *Genet Mol Biol* 33:125-130.
- Bnejdi F, Saadoun M, Allagiu MB, Hanbury C and El Gazzah M (2010) Relationship between epistasis and aggressiveness in resistance of pepper (*Capsicum annuum* L.) to *Phytophthora nicotianae*. *Genet Mol Biol* 33:279-284.
- Bocianowski J (2008) Comparison of two methods of estimation of nonallelic interaction of QTL effects on the basis of doubled haploid lines in barley. *Agric Conspec Sci* 73:183-187.
- Bocianowski J (2012a) Analytical and numerical comparisons of two methods of estimation of additive x additive interaction of QTL effects. *Sci Agric* 69:240-246.
- Bocianowski J (2012b) A comparison of two methods of estimation of additive-by-additive interaction of QTL effects by a simulation study. *J Theor Biol* 308:20-24.
- Bocianowski J, Krajewski P and Kaczmarek Z (1999) Comparison of methods of choosing extreme doubled haploid lines for genetic parameter estimation. *Colloq Biometr* 29:193-202.
- Carlborg O and Haley CS (2004) Epistasis: Too often neglected in complex trait studies? *Nat Rev Genet* 5:618-625.
- Carlborg Ö, Andersson L and Kinghorn B (2000) The use of a genetic algorithm for simultaneous mapping interacting quantitative trait loci. *Genetics* 155:2003-2010.
- Charcosset A, Mangin B, Moreau L, Combes L, Jourjon M-F and Gallais A (2001) Heterosis in maize investigated using connected RIL populations. In: Gallais A, Dillmann C and Goldringer I (eds) *Quantitative Genetics and Breeding Methods: The Way Ahead*. Proceedings of the Eleventh Meeting of the EUCARPIA, Section Biometrics in Plant Breeding. Institut National de la Recherche Agronomique, Paris, Les Colloques 96, pp 89-98.
- Chase K, Adler FR and Lark KG (1997) EPISTAT: A computer program for identifying and testing interaction between pairs of quantitative trait loci. *Theor Appl Genet* 94:724-730.
- Choo TM and Reinbergs E (1982) Estimation of the number of genes in doubled haploid populations of barley (*Hordeum vulgare*). *Can J Genet Cytol* 24:337-341.
- Chung Y, Lee SY, Elston RC and Park T (2007) Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. *Bioinformatics* 23:71-76.
- Cordell HJ (2002) Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463-2468.
- Cordell HJ (2009) Detecting gene-gene interaction that underlie human diseases. *Nat Rev Genet* 10:392-404.
- da Silva Guimarães LM, de Resende MDV, Lau D, Rosse LN, Alves AA and Alfenas AC (2010) Genetic control of *Eucalyptus urophylla* and *E. grandis* resistance to canker caused by *Chrysosporthe cubensis*. *Genet Mol Biol* 33:525-531.
- Evans DM, Marchini J, Morris AP and Cardon LR (2006) Two-stage two-locus model in genome-wide association. *PLoS Genet* 2:e157.
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edin* 52:399-433.
- Gayán J, González-Pérez A, Bermudo F, Sáez ME, Royo JL, Quintas A, Galan JJ, Morón FJ, Ramirez-Lorca R, Real LM, et al. (2008) A method for detecting epistasis in genome-wide studies using case-control multi-locus association analysis. *BMC Genomics* 9:e360.
- Hahn LW, Ritchie MD and Moore JH (2003) Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 19:376-382.
- Holland JB (1998) EPISTACY: A SAS program for detecting two-locus epistatic interaction using genetic marker information. *J Hered* 89:374-375.
- Jannink J-L and Jansen R (2001) Mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* 157:445-454.
- Jansen RC and Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447-1455.
- Kao C-H, Zeng Z-B and Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. *Genetics* 152:1203-1216.
- Kearsey MJ and Pooni HS (1996) *The Genetical Analysis of Quantitative Traits*. Chapman and Hall, London, 381 pp.
- Lau C-H and Muniandy S (2012) Lack of evidence for inter-molecular epistatic interactions between adiponectin and resistin gene polymorphisms in Malaysian male subjects. *Genet Mol Biol* 35:38-44.
- Li ZK, Pinson SRM, Park WD, Paterson AH and Stansel JW (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* 145:453-465.
- Lin HJ, Tan DF, Zhang ZM, Lan H, Gao SB, Rong TZ and Pan GT (2008) Analysis of digenic epistatic and QTL x environment interactions for resistance to banded leaf and sheath blight in maize (*Zea mays*). *Int J Agric Biol* 10:605-611.
- Lukens LN and Doebley J (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet Res* 74:291-302.
- Marchini J, Donnelly P and Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37:413-417.
- Moore JH (2004) Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn* 4:795-803.
- Nadeem K and Azhar FM (2005) Genetic analysis of fibre length and strength of *Gossypium hirsutum* L. *Int J Agric Biol* 7:263-265.
- Phillips PC (2008) Epistasis - The essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9:855-867.
- Searle SP (1982) *Matrix Algebra Useful for Statistics*. John Wiley and Sons, New York, 464 pp.
- Sen S and Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics* 159:371-387.
- Shakoore MS, Malik TA, Azhar FM and Saleem MF (2010) Genetics of agronomic and fiber traits in upland cotton under drought stress. *Int J Agric Biol* 12:495-500.
- Sharmila V, Ganesh SK and Gunasekaran M (2007) Generation mean analysis for quantitative traits in sesame (*Sesamum indicum* L.) crosses. *Genet Mol Biol* 30:80-84.

- Surma M, Adamski T and Kaczmarek Z (1984) The use of doubled haploid lines for estimation of genetic parameters. *Genet Polon* 25:27-32.
- Ullah S, Khan AS, Raza A and Sadique S (2010) Gene action analysis of yield and yield related traits in spring wheat (*Triticum aestivum*). *Int J Agric Biol* 12:125-128.
- Zeng Z-B, Kao C-H and Batsen CJ (1999) Estimating the genetic architecture of quantitative traits. *Genet Res* 74:279-289.
- Zhang Y and Liu JS (2007) Bayesian inference of epistatic interactions in case-control studies. *Nat Genet* 39:1167-1173.

*Associate Editor: Dario Grattapaglia*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.