Comparison of Two Methods of Estimation of Nonallelic Interaction of QTL Effects on the Basis of Doubled Haploid Lines in Barley

Jan BOCIANOWSKI (✉)

Summary

The paper presents numerical comparison of two methods of estimation of nonallelic interaction of QTL effects. In the first method we assume that we observe only the plant phenotype, while in the second method we have additional information from the molecular markers observations. In this paper we analysed phenotypic data on 120 barley doubled haploid lines, derived from cross Clipper × Sahara and data concerning 183 molecular markers. The analysed traits were beta-amylase activity, alpha-amylase activity, beta-glucanase activity and cyst nematode resistance. Results obtained for three from four traits show that by using molecular marker observations we obtain estimators that have smaller absolute values than estimators obtained by the phenotypic method.

Key words

nonallelic interaction, quantitative trait loci, doubled haploid lines

Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Wojska Polskiego 28, 60-637 Poznań, Poland
✉ e-mail: jboc@au.poznan.pl

Received: October 31, 2006 | Accepted: March 3, 2008

Introduction

The investigations regarding the inheritance of quantitative traits focus mainly on the characteristic of the way of genes action. This kind of information we can obtain, among the other things, on the basis of genetic parameters which estimate effect of action and interaction of loci in homozygous state. Results of these parameters testing are the base for the conclusion about additive gene action and nonallelic interaction (Surma, 1996). The estimation of the additive gene action effect and the nonallelic interaction (epistasis or additive × additive interaction) effect is possible on the basis of phenotypic observations only (Kaczmarek et al., 1984) or by quantitative trait loci (QTL) mapping using DNA markers (Doerge et al., 1997; Jansen, 1997).

The objective of this paper is to perform some numerical comparisons of these two estimation methods applied to the parameter connected with the additive gene action effect and the parameter connected with the epistasis effect.

Estimation of the additive and nonallelic gene action effects

If in the experiment we observe n homozygous (recombinant linbred, RI; doubled haploid, DH) plant lines, we get an n-vector of phenotypic mean observations $\mathbf{y} = [y_1\ y_2\ \ldots\ y_n]'$ and q n-vectors of markers 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 –1 or 1, depending on the parent's genotype exhibited by the i-th line.

Estimation based on the phenotype

Calculation of the additive gene effect *[a]* as well as nonallelic interaction of homozygous loci (epistasis) effect *[aa]* (Kearsey and Pooni, 1996) on the basis of phenotypic observations **y** requires identification of the groups of extreme lines, i.e., lines with minimal and maximal expression of the observed trait. The group of minimal lines consists of the lines which contain, theoretically, only alleles decreasing the value of the trait. Analogously, the group of maximal lines contains the lines which have only alleles increasing the trait value.

Genetic considerations show that the expected value of the trait for a maximal line is $E(y_i)=\mu+[a]+[aa]$, where μ is the general mean, and that the expected value for a minimal line is $E(y_i)=\mu-[a]+[aa]$. The values of *[a]* and *[aa]* can be estimated by the formulas (Choo and Reinbergs, 1982)

$$\hat{[a]}_f = \frac{1}{2}\left(\overline{L}_{\max} - \overline{L}_{\min}\right), \tag{1}$$

$$\hat{[aa]}_f = \frac{1}{2}\left(\overline{L}_{\max} + \overline{L}_{\min}\right) - \overline{L}, \tag{2}$$

where $\overline{L}_{\min}$ and $\overline{L}_{\max}$, denote the means for the groups of minimal and maximal lines, respectively, $\overline{L}$ denotes the mean for all lines. Some methods of identification of the minimal and maximal lines were considered by Bocianowski et al. (1999).

Estimation based on the genotypic observations

Estimation of *[a]* and *[aa]* is based on the assumption that the genes responsible for the trait are closely linked to observed molecular marker. By choosing from all observed markers p we can explain the variability of the trait, and model observations for the lines as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{3}$$

where **1** denotes the n-dimensional vector of ones, μ denotes the general mean, **X** denotes $(n\times p)$-dimensional matrix of the form $\mathbf{X} = \left[\begin{array}{cccc} \mathbf{m}_{l_1} & \mathbf{m}_{l_2} & \ldots & \mathbf{m}_{l_p} \end{array}\right]$, $l_1, l_2, \ldots, l_p \in \{1, 2, \ldots, q\}$, $\boldsymbol{\beta}$ denotes the p-dimensional vector of unknown parameters of the form $\boldsymbol{\beta}' = \left[\begin{array}{cccc} [a]_{l_1} & [a]_{l_2} & \ldots & [a]_{l_p} \end{array}\right]$, **Z** denotes matrix which columns are products of some columns of matrix **X**, $\boldsymbol{\gamma}$ denotes the vector of unknown parameters of the form $\boldsymbol{\gamma}' = \left[\begin{array}{cccc} [aa]_{l_1l_2} & [aa]_{l_1l_3} & \ldots & [aa]_{l_{p-1}l_p} \end{array}\right]$, **e** denotes the n-dimensional vector of random variables such that $E(e_i) = 0$, $Cov(e_i, e_j) = 0$ for $i\neq j$, $i, j = 1, 2, \ldots, n$. The parameters $[a]_{l_1}, [a]_{l_2}, \ldots, [a]_{l_p}$ are the additive effects of the genes controlling the trait and the parameters, $[aa]_{l_1l_2}$, $[aa]_{l_1l_3}$, $\ldots$, $[aa]_{l_{p-1}l_p}$, are the additive × addi tive interaction effects. We assume that the epistatic interaction effects show only loci with significant additive gene action effects. This assumption significantly decreases the number of potential significant effects and cause regression model more useful.

Denoting by $\boldsymbol{\alpha}' = \left[\mu \quad \boldsymbol{\beta}' \quad \boldsymbol{\gamma}'\right]$ and $\mathbf{G} = \left[\mathbf{1} \quad \mathbf{X} \quad \mathbf{Z}\right]$ we obtain the model

$$\mathbf{y} = \mathbf{G}\boldsymbol{\alpha} + \mathbf{e}. \tag{4}$$

When **G** is of full rank, the estimate of $\boldsymbol{\alpha}$ is given by

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{y} \tag{5}$$

(Searle 1982)

The total additive effect of genes influencing the trait:

$$\hat{[a]}_g = \sum_{k=1}^{p}\left|\hat{[a]}_{l_k}\right| \tag{6}$$

and the total epistasis effect:

$$\hat{[aa]}_g = \sum_{k=1}^{p-1} \sum_{\substack{k'=2 \\ k' \neq k}}^{p} \hat{[aa]}_{l_k l_{k'}}. \quad (7)$$

Comparison of estimators

The estimators of the total additive gene action, (1) and (6), and of the total epistasis effect, (2) and (7), can be analytically compared under some simplifying assumptions.

Model (3) treats the marker observations as fixed. In fact, the vectors $\mathbf{m}_l$, $l = 1, 2, \ldots, q$, constitute observations of some random variables. Then, the following two genetic assumptions can be taken into account:

(i) that the markers are unlinked, that is, for any two markers probability of observing (1, 1) or (–1, –1) is the same as observing (1, –1) or (–1, 1);

(ii) that the segregation of each marker is concordant with the genetic model appropriate for the analysed population, which in our case means that the probability of observing "–1" is the same as observing "1".

If the marker data satisfied exactly assumptions (i) and (ii) we would have

$$\hat{\mu} = \bar{y};$$

$$\hat{[a]}_g = \sum_{k=1}^{p} \left| \frac{1}{2} \left(\bar{y}^{(l_k,+)} - \bar{y}^{(l_k,-)} \right) \right|, \quad (8)$$

$$\hat{[aa]}_f = \sum_{k=1}^{p-1} \sum_{\substack{k'=2 \\ k' \neq k}}^{p} \left[\frac{1}{2} \left(\bar{y}^{(l_k l_{k'},+)} + \bar{y}^{(l_k l_{k'},-)} \right) - \bar{y} \right],$$

where $\bar{y}^{(l_k,+)}$ and $\bar{y}^{(l_k,-)}$ denote the means for lines with observations of k-th marker equal 1 and –1, respectively, $\bar{y}^{(l_k l_{k'},+)}$ and $\bar{y}^{(l_k l_{k'},-)}$ denote the means for lines with observations of k-th and k'-th markers equal 1 and –1, respectively.

Practically, the marker data do not fulfill exactly the conditions leading to (8). The assumption (i) is, however, approximately true if the markers chosen to model (3) are weakly linked, that is, if they are far from each other on the linkage map (possibly in different linkage groups). The assumption (ii) is usually tested by a χ^2 test before any linkage analysis is done.

Example

The data used for this example concern a set of 120 doubled haploid lines of barley, derived from the cross between the Australian barley variety Clipper and the Algerian landrace Sahara 3771 at Waite Agricultural Research Institute, University of Adelaide, Australia (Karakousis et al., 2003). The lines were investigated with respect to four phenotypic traits: beta-amylase activity, BA; alpha-amylase activity, AA; beta-glucanase activity, BG; cyst nematode resistance, CCN. We used observations of 183 molecular markers (SSR and RFLP).

In this paper groups of minimal and maximal lines were identified using the quantile method (Bocianowski et al., 1999). The number of genes (number of effective factors) obtained on the basis of phenotypic observations only was calculated using formula presented by Kaczmarek et al. (1988). To correct the missing marker observations we used the imputation method of Martinez and Curnow (1994). Selection of markers chosen for model (3) was made in this paper in two stages. First, selection was made by backward stepwise search independently inside all linkage groups. Then, markers chosen in this way were put in one group and subjected to the second backward selection (Jansen and Stam, 1994). At the both stages we used the critical significance level equal to $\alpha = 0.001$.

Results

Table 1 contains the results of the comparison of estimates of the parameters *[a]* and *[aa]* calculated on the basis of formulas (1) and (6) and formulas (2) and (7), re-

Table 1. Estimates of genetic parameters calculated by two methods for doubled haploid lines of barley

Trait	BA	AA	BG	CCN
Estimation based on the phenotype				
Number of genes	4.5	4.6	5.5	0.7
$[a]_f$	798.4 (121.3)[a]	233.5 (30.2)	276.5 (28.4)	16.7 (2.81)
$[aa]_f$	102.9 (15.8)	-12.4 (4.76)	-103.2 (9.9)	-4.1 (1.02)
Estimation based on the genotypic observations				
QTL number	4	2	3	3
$[a]_g$	773.1 (142.0)	92.8 (59.4)	451.1 (55.8)	49.02 (4.9)
$[aa]_g$	90.1 (17.6)	-22.0 (7.5)	0.0 (0.0)	1.38 (0.42)
R^2 [%]	35.4	36.1	27.5	71.30

[a] Standard error

spectively. The number of genes and QTLs obtained for beta-amylase activity were similar, 4.5 and 4, respectively. For the other traits difference in this number was about 2 (only for cyst nematode resistance the QTL number was larger than number of genes calculated on the basis of phenotypic observations only). For two traits (beta-amylase activity and alpha-amylase activity) we observed that the additive effect calculated on the basis of the marker observations is lower than the total additive effect obtained from phenotypic observations only. For beta-glucanase activity and cyst nematode resistance we observed that the value of parameter *d* calculated on the basis of formula (1) is lower than the value calculated on the basis of formula (6). The results obtained for three out of four analyzed traits (except alpha-amylase activity) show that by the use of molecular marker observations we obtained estimates which have smaller absolute values than estimates obtained by the use of phenotypic method. For beta-glucanase activity we did not find statistically significant nonallelic interaction effects. Epistasis effect with two opposite signs was obtained for cyst nematode resistance. Percentage of trait variation accounted by QTL and epistasis ranged from 27.5% (for beta-glucanase activity) to 71.3% (for cyst nematode resistance).

Discussion

The problem of epistasis was considered in some QTL mapping studies (e.g. Stuber et al., 1992; Cheverud and Routman, 1995; Cockerham and Zeng' 1996; Kao et al., 1999; Goodnight, 2000), but not sufficiently, and many theoretical and statistical issues involved in epistasis have not been discussed. In the papers aiming estimation of genetic parameters on the basis of marker observation the assumption of the lack of epistasis effect is very often a result of too large number of parameters in a model (Jansen and Stam, 1994).

The comparison of the two estimation methods of the additive gene action effect as well as the nonallelic interaction of homozygous loci effect shows some reasons for resemblance and differences between estimates based on the phenotypic observations and estimates using genetic observations. The example shows different degree of resemblance of estimates for different traits.

The additive effects obtained for beta-glucanase activity and cyst nematode resistance from phenotypic observations only are lower than the additive effects calculated on the basis of the marker observations because in both cases doubled haploid lines are diverse on the genetic level with simultaneously lack of phenotypic diversity. The effects of nonallelic interaction between homozygous loci with opposite signs for cyst nematode resistance, obtained herein, are the results of the estimation of too small number of genes on the basis of phenotypic observation only.

The presented results hint that we can use regression model with epistatic interaction for QTL characterization. Very important is assumption that additive × additive effects show only loci linked to markers with significant additive effects. The method presented in the paper was useful in the example of real data set.

Further studies are necessary with respect to epistasis effect conduct by simulation analysis that would make possible consideration of different experimental situations.

Conclusion

The additive effect obtained from phenotypic observations only is lower than the additive effect calculated on the basis of the marker observations because DH lines are diverse on the genetic level with simultaneously lack of phenotypic diversity.

The effects of nonallelic interaction between homozygous loci with opposite signs for the compared estimation methods are the results of the estimation of too small number of genes on the basis of phenotypic observation only.

The method of estimation of the additive × additive interaction effects was useful statistical tool for QTL characteristic.

References

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

Cheverud J. M., Routman E. J. (1995). Epistasis and its contribution to genetic variance components. Genetics 139: 1455-1461

Choo T. M., Reinbergs E. (1982). Estimaton of the number of genes in doubled haploid populations of barley (*Hordeum vulgare*). Canadian Journal of Genetics and Cytology 24: 337-341

Cockerham C. C., Zeng Z.-B. (1996). Design III with marker loci. Genetics 143: 1437-1456

Doerge R. W., Zeng Z.-B., Weir B. S. (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. Stat. Sci. 12: 195-219

Goodnight C. J. (2000). Quantitative trait loci and gene interaction: the quantitative genetics of metapopulation. Heredity 84: 587-598

Jansen R. C. (1997). Mapping QTLs in experimental and breeding populations. In: Krajewski P, Kaczmarek Z (eds) Advances in Biometrical Genetics, Proceedings of the Tenth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, Institute of Plant Genetics, Poznań, pp 27-34

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

Kaczmarek Z., Surma M., Adamski T. (1984). Some genetic parameters – their interpretation and estimation. Listy Biometryczne – Biometrical Letters 21: 3-20 (in Polish)

Kaczmarek Z., Surma M., Adamski T. (1988). Epistatic effects in estimation of the number of genes on the basis of doubled haploid lines. Genetica Polonica 29: 353-359

Kao C.-H., Zeng Z.-B., Teasdale R. D. (1999). Multiple interval mapping for quantitative trait loci. Genetics 152: 1203-1216

Karakousis A., Barr A. R., Kretschmer J. M., Manning S., Jefferies S. P., Chalmers K. J., Islam A. K. M., Langridge P. (2003). Mapping and QTL analysis of the barley population Clipper × Sahara. Australian Journal of Agricultural Research 54(12): 1137-1140

Kearsey M. J., Pooni H. S. (1996). The genetical analysis of quantitative traits. London, Chapman and Hall

Martinez O., Curnow R. N. (1994). Missing markers when estimating quantitative trait loci using regression mapping. Heredity 73: 198-206

Searle S. P. (1982). Matrix algebra useful for statistics. John Wiley & Sons, New York

Stuber C. W., Lincoln S. E., Woff D. W., Helentjaris T., Lander E. S. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular marker. Genetics 132: 832-839

Surma M. (1996). Biometrical-genetic analysis of quantitative traits of hybrids and doubled haploids in spring barley. Treatises and Monographe 3, Institute of Plant Genetics, Poznań (in Polish)

acs73_29