

Kansas State University Libraries

## New Prairie Press

---

Conference on Applied Statistics in Agriculture

2008 - 20th Annual Conference Proceedings

---


# MULTI-TRAIT QTL MAPPING USING A STRUCTURAL EQUATION MODEL

Xiaojuan Mi

Kent M. Eskridge

Dong Wang

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Mi, Xiaojuan; Eskridge, Kent M.; and Wang, Dong (2008). "MULTI-TRAIT QTL MAPPING USING A STRUCTURAL EQUATION MODEL," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1095>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

## MULTI-TRAIT QTL MAPPING USING A STRUCTURAL EQUATION MODEL

Xiaojuan Mi, Kent M. Eskridge and Dong Wang

Department of Statistics, Univ. of Nebraska, Lincoln, NE 68583-0963, USA

### Abstract

Research on mapping quantitative trait loci (QTL) often results in data on a number of traits that have well established causal relationships. Many multi-trait QTL mapping methods, taking into account the correlation among the multiple traits, have been developed to improve the statistical power of the test for QTL and the precision of parameter estimation. However none of these methods are capable of incorporating the causal structure among the traits with the consequence that genetic functions of the QTL may not be fully understood. Structural equation modeling (SEM) allows researchers to explicitly characterize the causal structure among the variables and to decompose the effects into direct, indirect, and total effects. In this paper, we developed a multi-trait SEM method of QTL mapping that takes into account the causal relationships among traits. The performance of the proposed method is evaluated by simulation study. Compared with single trait analysis and the multi-trait least-squares analysis, our proposed model (Multi-trait SEM) provides important insight into how QTLs regulate traits by investigating the direct, indirect, and total QTL effects, which is generally not possible with other methods. The approach also helps with building models that more realistically reflect complex relationships among QTL and traits, and is more precise and efficient in QTL mapping than single trait analysis.

**Key words:** QTL mapping; multiple traits; SEM; least squares.

### 1. Introduction

Research on quantitative trait loci (QTL) often results in data on a number of traits that have well established causal relationships. For example, in wheat genetics it is common to collect data on grain yield and yield components such as, kernel weight (TKW), spikes per square meter (SPSM), and kernels per spike (KPS) where the causal relationships among these traits is well-established (Dofing & Knight 1992). A QTL may affect SPSM, KPS, and TKW, which act as intermediate variables and ultimately affect yield. Common single trait QTL procedures can be used to estimate the total QTL effects but not the direct and indirect effects. However, the indirect and direct effects can help to answer important questions that are not addressed by examining the total effect. For instance, a pleiotropic QTL can have a positive direct effect on grain yield, but a negative effect on a yield component. Without knowing the causal relationship, a breeder might select against the QTL thinking it only affects the yield component detrimentally, not knowing it is actually beneficial on the important trait of grain yield. Thus the total effect can provide a misleading impression. To understand the genetic effects thoroughly, we should understand not only their total effects, but also their direct and indirect effects through other traits by taking advantage of those causal relationships among traits. Such a strategy of QTL mapping can provide additional insight into how QTLs regulate traits directly and indirectly through other traits. It should also improve the power to detect the QTL and the precision of the location estimate.

Although it is common to collect data observations on multiple traits, which are causally related or genetically correlated, frequently, QTLs are mapped for each trait separately using single trait analyses (Lander & Botstein 1989; Haley & Knott 1992; Jansen & Stam 1994; Zeng 1994). Alternatively, several multiple traits QTL analysis (joint analysis) methods have been developed taking into account the correlation among multiple traits. These methods have been shown to improve statistical power for QTL detection and precision of parameter estimate compared to single trait analysis. Among the most effective approach are multi-trait maximum-likelihood (ML) (Jiang & Zeng 1995 and Korol *et al* 1995, 1998), multi-trait least square (LS) (Korol *et al.* 1995, 1998; Calinski *et al.* 1999; Knott and Haley 2000; Hackett *et al.* 2001), principal component analysis (PCA) (Weller *et al.* 1996; Mangin *et al.* 1998; Calinski *et al.* 2000), and discriminate analysis (DA) (Gilbert & Le Rol 2003). Multi-trait ML, implemented with the ECM algorithm, extracts maximum information from the data, but could not be implemented in complex data structure because of the difficulty in computations. Multi-trait LS, which regresses the quantitative trait value on the conditional expected genotypic value, produces very similar result to ML and simplifies the computation (Haley & Knott 1992). The PCA method transforms multiple traits into canonical variables so that the single trait analyses can be carried out for each canonical variable. Similarly the DA is based on the linear combination of the traits, specific to each tested position and analyzed by a univariate method. But the results of PCA and DA may be difficult to interpret. However, none of the above methods take advantage of causal structure among the traits. Multi-trait QTL mapping can provide additional insight into the genetic functions of QTL to consider causal structure.

Structural equation modeling (SEM) is an important method in finding the appropriate model and estimating the causal relationships among variables. It allows researchers to decompose the effects of one variable on another into direct, indirect, and total effects. Direct effects are the influence of one variable on another that are not mediated by any other variable. Indirect effects are mediated by at least one other variable, and the total effect is the sum of direct and indirect effects. By explicitly accounting for the causal structure among traits, SEM can provide more insight into multiple trait QTL analysis.

In this paper, we developed a multi-trait SEM method of QTL mapping using a population of recombinant inbred lines (RILs). The method is sufficiently flexible in that it can handle a variety of genetic models, such as dominance effects, epistasis, and multiple interval mapping. The proposed model is compared with multi-trait LS composite interval mapping and single-trait LS composite interval mapping in terms of the statistical power of QTL detection, and the precision of parameter estimate. The performance of the proposed method is evaluated by simulation of a RIL population, which is derived by crossing two inbred parents and repeatedly mating the resulting siblings for 10 generations or more to ensure that they are inbred.

## 2. Statistical Method

In this section, we develop our model using a RIL population as an example applying the composite interval mapping method. Consider  $p$  causally related traits  $y_1, y_2 \dots$ , and  $y_p$  affected by a QTL located between two flanking markers ( $A$  and  $B$ ), and a number of genetic cofactor markers located on the same chromosome. There are four possible flanking marker genotypes ( $AABB$ ,  $AAbb$ ,  $aaBB$ , and  $aabb$ ) and two possible QTL genotypes ( $QQ$  and  $qq$ ) assuming no double crossover. Our approach is based on two steps.

**Step1:** Derive the conditional expected genotypic value at a given position from the observed flanking marker genotypes since the QTL genotype is not observed. The genotypic effects of the two QTL genotypes are set at  $u + a$  and  $u - a$  for  $QQ$  and  $qq$  respectively, where  $u$  is the mean and  $a$  is the genetic additive effect. Haldane's (1931) mapping function is used to transform genetic distance in Morgans into recombinant fraction. Then, the expected values for four marker genotypes can be derived in terms of the putative QTL (Table 1) (Knapp, *et al.*, 1990; Sari-Gorla *et al.*, 1997). The coefficients (denoted as  $x_{QTL}$ ) of  $a$  in terms of recombination fractions for each of the four flanking marker combinations can be expressed as:

$$x_{QTL} = \begin{cases} 1 & AABB \\ (R_B - R_A) / R & AAbb \\ (R_A - R_B) / R & aaBB \\ -1 & aabb \end{cases}$$

where  $R=R_A+R_B$ ;  $R_A$  is the recombination frequency between marker  $A$  and QTL  $Q$ ;  $R_B$  is the recombination frequency between marker  $B$  and QTL  $Q$ ;  $R$  is recombination frequency between marker  $A$  and marker  $B$  respectively.

**Step 2:** Construct the SEM by treating the coefficients  $x_{QTL}$  of the genetic effect as observed values. The term SEM denotes a class of analytic techniques that usually include estimation of the unobserved or latent factors and of the structure of the relationships among these latent factors (Loehlin 1992). We restrict this paper to the special case of a SEM known as path analysis (Marcoulides & Schumacker 1996) where every variable in the model is directly measured or observed. The statistical model for mapping a QTL for multiple traits is specified in matrix form (Bollen 1989):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} & \cdots & \beta_{1p} \\ 0 & 0 & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} x_{QTL} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1q} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pq} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix} \quad (1)$$

where  $y_k$  is the phenotypic value for trait  $k$ ;  $\beta_{kl}$  is the regression coefficient of trait  $l$  on trait  $k$ ;  $\alpha_k$  is the additive effect of the putative QTL on trait  $k$ ;  $x_{QTL}$  is the calculated coefficient of the additive QTL effect;  $\gamma_{kj}$  is the regression coefficient of cofactor marker  $j$  on trait  $k$ , assuming  $q$  markers are selected as cofactor markers to control the variation from these QTLs;  $x_j$  is the genotype of the  $j$ th cofactor marker, which takes values of 1 and -1 for marker genotype  $MM$  and  $mm$  respectively; and  $e_k$ , the residual effect on trait  $k$  which is assumed to be multivariate normally distributed with means zero and covariance matrix

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

Model (1) is more compactly written as

$$\mathbf{y} = \mathbf{B}\mathbf{y} + \mathbf{\Gamma}\mathbf{x} + \boldsymbol{\zeta}, \text{ where } \boldsymbol{\zeta} \sim N(0, \Psi) \quad (2)$$

Where  $\mathbf{y}$  is a  $p \times 1$  vector of  $y_k$ ,  $\mathbf{B}$  is the  $p \times p$  coefficient matrix (contains  $\beta$ s) that describes the causal relationships among the  $p$  traits, where  $(\mathbf{I} - \mathbf{B})^{-1}$  exists;  $\mathbf{\Gamma}$  is the  $p \times (q+1)$  coefficient

matrix (contains  $\alpha_{ks}$  and  $\gamma_{kj}$ s) that describes causal relationship between endogenous variables (traits); and exogenous variables (QTL and markers);  $x$  is a  $(q+1) \times 1$  vector of exogenous variables, which include  $x_{QTL}$  and  $q$  cofactor markers used for background control, and  $\zeta$ , a  $p \times 1$  vector of errors, is assumed to be multivariate normally distributed with mean zero and covariance matrix  $\Psi$  which is diagonal. Elements in  $\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Psi}$  are parameters that need to be estimated.

We consider only recursive models, which are characterized by the fact that the matrix  $\mathbf{B}$  is upper triangular. This assumption is reasonable in many domains, since it basically forbids feedback causation.

**Parameter estimation:** Under the SEM, the unknown parameters are estimated so that the model implied covariance matrix  $\Sigma(\theta)$  is close to the sample covariance matrix  $\mathbf{S}$ . The closeness requires a fitting function based on  $\Sigma(\theta)$  and  $\mathbf{S}$  to be minimized. The commonly used fitting functions for SEM are maximum likelihood (ML):

$$F_{ML} = \log |\Sigma(\theta)| + \text{tr}(\mathbf{S}\Sigma^{-1}(\theta)) - \log |\mathbf{S}| - (p + q + 1)$$

where  $\Sigma(\theta) = \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\Sigma_{xx}\mathbf{\Gamma}' + \mathbf{\Psi})[(\mathbf{I} - \mathbf{B})^{-1}]' & (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\Sigma_{xx} \\ \Sigma_{xx}\mathbf{\Gamma}'[(\mathbf{I} - \mathbf{B})^{-1}]' & \Sigma_{xx} \end{pmatrix}$  is the model implied

covariance matrix,  $\theta$  contains the parameters in  $\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Psi}$  that need to be estimated,

$\mathbf{S} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{xx} \end{pmatrix}$  is the observed sample covariance matrix among manifest variables,  $p$  is the

number of traits, and  $q + 1$  is the number of cofactor markers including one for the QTL. Although ML is based on the assumption of multivariate normality, its estimation procedures are robust to moderate violation of this assumption (Joreskog and Sorbom 1989; Bollen 1989).

**Hypothesis tests:** For QTL mapping, we are most interested in the existence of a QTL. The hypothesis test can be formulated as:

$H_0: \alpha_1 = \alpha_2, \dots, = \alpha_p = 0$  (restricted model, *i.e.*, the putative QTL does not exist)

$H_A$ : at least one of them is not zero (unrestricted model, *i.e.*, the putative QTL exists)

The Likelihood Ratio (LR) statistic is

$$LR = -2[\log L(\hat{\theta}_r) - \log L(\hat{\theta}_u)]$$

where  $\hat{\theta}_r$  is the estimator under the restricted model, and  $\hat{\theta}_u$  is the estimator under the unrestricted model. LR is approximately chi-square distributed when the restricted model is true. The degrees of freedom is the difference in the degrees freedom for the two models. In the SEM framework, the LR statistic is calculated as the difference in the usual chi-square estimators for the restricted and unrestricted models, with the degrees freedom equal to the difference in degrees of freedom between the two models (Bollen 1989). Because the test is performed for a number of intervals, the distribution of the maximum LR statistic is very complicated. Therefore, it is difficult to determine an exact significance critical value. Zeng (1994, 1995) suggested that the error rate of the test per interval,  $\alpha$ , can be approximated by using the Bonferroni correction, that is, use  $\chi^2_{\alpha/M, m+1}$  to approximate the critical value of the test where  $M$  is the number of intervals involved in the test,  $m$  is the number of traits including one for the position of the putative QTL.

Note that, the multi-trait SEM and the multi-trait LS models have similar structure, but are used to test different hypotheses of QTL effects, and estimate different parameters. With the multi-trait SEM, the existence of direct QTL effects is tested, and the total, indirect and direct QTL effects are estimated. While with the multi-trait LS approach, the existence of the total QTL effects are tested, and only the estimates of the total QTL effects are provided.

### 3. Simulation Study

We investigated the multi-trait SEM method by simulation experiments. The data were simulated for 100 replicates of 300 lines from a RIL population. On a single chromosome segment of length 100 cM, 11 evenly spaced markers were simulated. A single additive QTL was placed at 42 cM to affect three traits, which are causally related as in Figure 1, where  $Y_3$  is yield component 3,  $Y_2$  is yield component 2, and  $Y_1$  is yield.

The phenotypic values for each individual are determined by the equation (3), the causal relationship among traits, effects of QTL sampled (where QTL takes values of 2 and 0 for genotype  $QQ$  and  $qq$  respectively), and random residual effects were sampled from the multivariate normal distribution with mean zero and covariance matrix (4).

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{pmatrix} 0 & 0.5 & 0.25 \\ 0 & 0 & -0.5 \\ 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} -0.125 \\ 0.5 \\ 0.25 \end{bmatrix} QTL + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (3)$$

$$\Psi = \begin{pmatrix} 1.6 & 0 & 0 \\ 0 & 1.8 & 0 \\ 0 & 0 & 2.5 \end{pmatrix} \quad (4)$$

Prior to the analysis, a test of multivariate normality was performed for all observed variables. The results show that the Mardia and Henze-Zirkler multivariate tests reject the multivariate normality but the multivariate plot indicate approximate normality, suggesting a minor violation of multivariate normality. The parameter estimates from multi-trait SEM are obtained by minimizing the ML fitting function in this study.

A single QTL was tested sequentially at each 1-cM point along the chromosome. Since markers are evenly distributed and widely separated, all markers except flanking markers are fitted in the model to control the genetic background (Jiang & Zeng 1995). Multi-trait SEM, multi-trait LS, and single-trait LS were applied at 1cM segments of the chromosome. Means and standard deviations of all parameter estimates were calculated from 100 replications (Table 3). The statistical power was determined by the proportion of the number of runs over 100 replicates with test statistic values greater than a critical value. We used  $\chi^2_{0.005/4}=14.86$  (approximated by the Bonferroni correction (Jiang & Zeng 1995)) as the critical value for the multi-trait SEM and the multi-trait LS, and  $\chi^2_{0.005/2}=10.60$  for the single-trait LS. The overall power was calculated as the proportion of times the QTL was detected for at least one of the three traits.

### 4. Results

Table 2 shows the observed power of detection of the QTL over 100 replicates by three different mapping methods. The power of multi-trait SEM is very close to that of the multi-trait LS, because both methods use a multivariate approach with the same observed covariance

matrix. The QTL detection power for the two multi-trait analysis methods was much higher than that of single-trait analysis. This likely resulted because the pleiotropic QTL has a larger positive direct effect on  $Y_2$  but a negative indirect effect, which in turn reduces the total QTL effect on  $Y_2$ . The relatively small QTL effects associated with  $Y_1$ ,  $Y_2$  and  $Y_3$  may not be detected using single trait LS. Multi-trait SEM QTL analysis would be most effective when the direct and indirect effects of a QTL are in opposite directions. If the direct and indirect QTL effects are in the same direction, the power of the SEM multi-trait analysis may be less than the overall power of single-trait analysis since the single-trait method tests the total QTL effect. In such a situation, the total QTL effect should be larger than either of the direct or indirect effects tested with the multi-trait SEM approach. Based on our QTL detection power analysis, both multi-trait methods, multi-trait SEM and multi-trait LS, perform substantially better than single-trait LS analysis.

Table 3 shows the estimates (and standard deviations) of QTL effects and positions resulting from the three different mapping methods. All estimates are relatively unbiased with high precision except the QTL position estimates from the single-trait LS method, which display markedly higher standard deviation. In general, the precision and accuracy of estimating QTL positions and effects by multi-trait SEM and multi-trait LS are much greater than single-trait LS. However, the multi-trait SEM is favored over the multi-trait LS and single-trait LS analyses because direct and indirect QTL effects can be detected.

The likelihood ratio test statistic profiles were plotted against the chromosome position (Figure 2) to compare the three methods of QTL detection. The QTL profiles of the two multi-trait methods (multi-trait SEM and multi-trait LS) were nearly identical with clear peaks at 42 cM. The QTL profiles of the single-trait LS method identified peaks at the same chromosome position but did not display LR values as high as either multi-trait method. This indicates the single-trait LS method has a lower chance of detecting the QTL effects in this example when the total QTL effect is reduced due to compensating effects among traits.

## 5. Discussion and Conclusion

We have presented a multi-trait SEM method for QTL mapping, taking into account the causal relationships among multiple traits, extending the work of Knott and Haley (2000). The performance of the method was illustrated using simulated data. The advantages of the multi-trait SEM over the multi-trait LS and the single-trait LS are: it provides important insight into how QTLs regulate traits by allowing investigation of the direct, indirect, and total QTL effects, which is impossible with other methods. Knowledge of these effects can be very important for plant breeders who would like to (1) break the unfavorable indirect QTL effects; (2) obtain more precise and efficient estimates and tests in QTL mapping than possible in the single trait analysis; (3) and who would value statistical methods that can be simply performed using commonly available statistical software such as SAS, or Mplus.

A prerequisite of the proposed method is prior knowledge of the causal relationships among the multiple traits, since SEM is generally as a confirmatory rather than exploratory procedure. Theoretical insight and judgment by the researcher is very important in building a model. In practice, one can obtain some basic background about the key structure of the model either from knowledge of the related field or from preliminary data analysis. Other applications likely may require more model development based on procedures described elsewhere (eg see Bollen, 1989).

The model considered in this paper was illustrated using a RIL population to provide a general idea of the nature of QTLs affecting the traits, and did not include epistatic effect of the markers or QTL. However, the general formulas can be easily applied to obtaining the MLEs and evaluating the likelihoods for different population structures (such as F2, backcross), and genetic models by setting up the corresponding conditional QTL genotype probability and genetic  $\mathbf{x}$  matrix. In addition, it is also possible to test pleiotropic effects against close linkage and gene by environment interactions at a given genome position where the presence of a QTL is indicated by joint mapping. Although the proposed multi-trait SEM approach may not always be appropriate for every QTL mapping application, it does provide an attractive complementary method to understand complicated biological pathways and systems using available marker and phenotypic trait data.

### Acknowledgements

We would like to thank the anonymous referees whose comments helped to improve this paper.

### References

- Bollen, K.A., 1989 *Structural equations with latent variables*. New York: Wiley Interscience.
- Calinski, T., Kaczmarek, Z., Krajewski, P., Frova, C. and Sari-Gorla, M., 2000 A multivariate approach to the problem of QTL localization. *Heredity* **84**: 303-310
- Dofing, S.M., Knight, C.W., 1992 Alternative model for path analysis of small-grain yield. *Crop Sci.* **32**: 487-489.
- Gilbert, H. & Le Roy, P., 2003 Comparison of three multitrait methods for QTL detection. *Genet. Sel. Evol.* **35**: 281-304.
- Hackett, C. A., Meyer, R. C. and Thomas, W. T. B., 2001 Multi-trait QTL mapping in barley using multivariate regression. *Genetic Research* **77**: 95-106.
- Haldane, J.B.S., and C. Waddington, 1931 Inbreeding and linkage. *Genetics* **16**: 357-374.
- Haley, C. S., and Knott, S.A., 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315-324.
- Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447-1455.
- Jiang, C. J., and Zeng, Z. B., 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**: 1111-1127.
- Joreskog, K. G. and Sorbom, D., 1989 *LISREL 7: A Guide to the program and Applications, 2nd edition*. Chicago: SPSS Inc.
- Knapp, S. J. and Bridges, W. C., 1990 Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. *Genetics* **126**: 769-777.
- Knott, S. A., and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. *Genetics* **156**: 899-911.
- Korol A. B., Ronin Y. I. and Kirzhner V. M., 1995 Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* **140**: 1137-1147.
- Korol A. B., Ronin Y. I., Nevo E. and Hayes P. M., 1998 Multi-interval mapping of correlated trait complexes. *Heredity* **80**: 273-284.
- Lander E. S., and Botstein D., 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**:185-199.



- Loehlin, J. C., 1992 *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale, NJ: Erlbaum.
- Mangin, B., P. Thoquet and N. Grimsley, 1998 Pleiotropic QTL analysis. *Biometrics* **54**: 88–99.
- Marcoulides, G. A. & Schumacker, R. E. (eds.), 1996 *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sari-Gorla, M., T. Calinski, Z. Kaczmarek, and P. Krajewski. 1997 Detection of QTL x environment interaction in maize by a least squares interval mapping method. *Heredity* **78**:146–157.
- Weller, J. I., G. R. Wiggans, P. M. Van Raden and M. Ron, 1996 Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theor. Appl. Genet.* **92**: 998–1002.
- Zeng, Z. B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457-1468.

**Table 1.** Expected genotypic values for all possible flanking marker genotypes in a RIL population assuming no double crossover

Marker Genotype	Probability of QTL genotype		Expected genotypic value
	$QQ$	$qq$	
$AABB$	1	0	$u + a$
$AAbb$	$R_B/R$	$R_A/R$	$u + a(R_B - R_A)/R$
$aaBB$	$R_A/R$	$R_B/R$	$u + a(R_A - R_B)/R$
$aabb$	0	1	$u - a$

where  $R_A = 2r_A/(1+2r_A)$ ;  $R_B = 2r_B/(1+2r_B)$ ;  $R = r_A + r_B$ , where  $R_A$  and  $r_A$  are the recombination frequency and genetic distance (in cM) between marker  $A$  and QTL  $Q$ ,  $R_B$  and  $r_B$  are the recombination frequency and genetic distance (in cM) between marker  $B$  and QTL  $Q$ ,  $R$  is the recombination frequency between marker  $A$  and  $B$  respectively

**Table 2.** QTL detection power of multi-trait SEM, multi-trait LS, and single-trait LS methods of QTL analysis

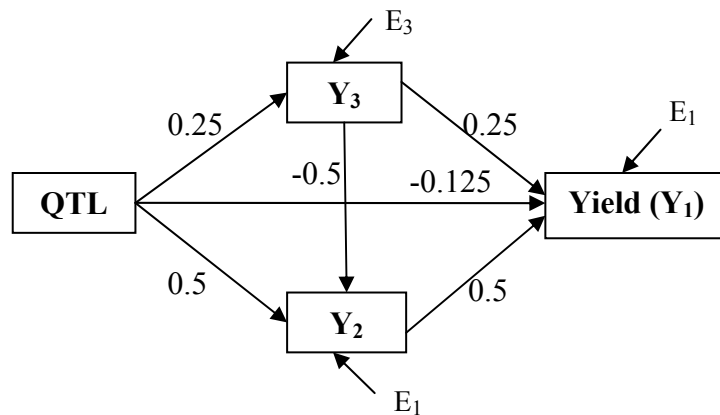
Methods	Trait	Power ( $\alpha = 0.05$ )	
		Single trait	Overall*
multi-trait SEM			91
Multi-trait LS			90
Single-trait LS	Y1	2	42
	Y2	39	
	Y3	4	

\*Power QTL was detected for at least one of the three traits

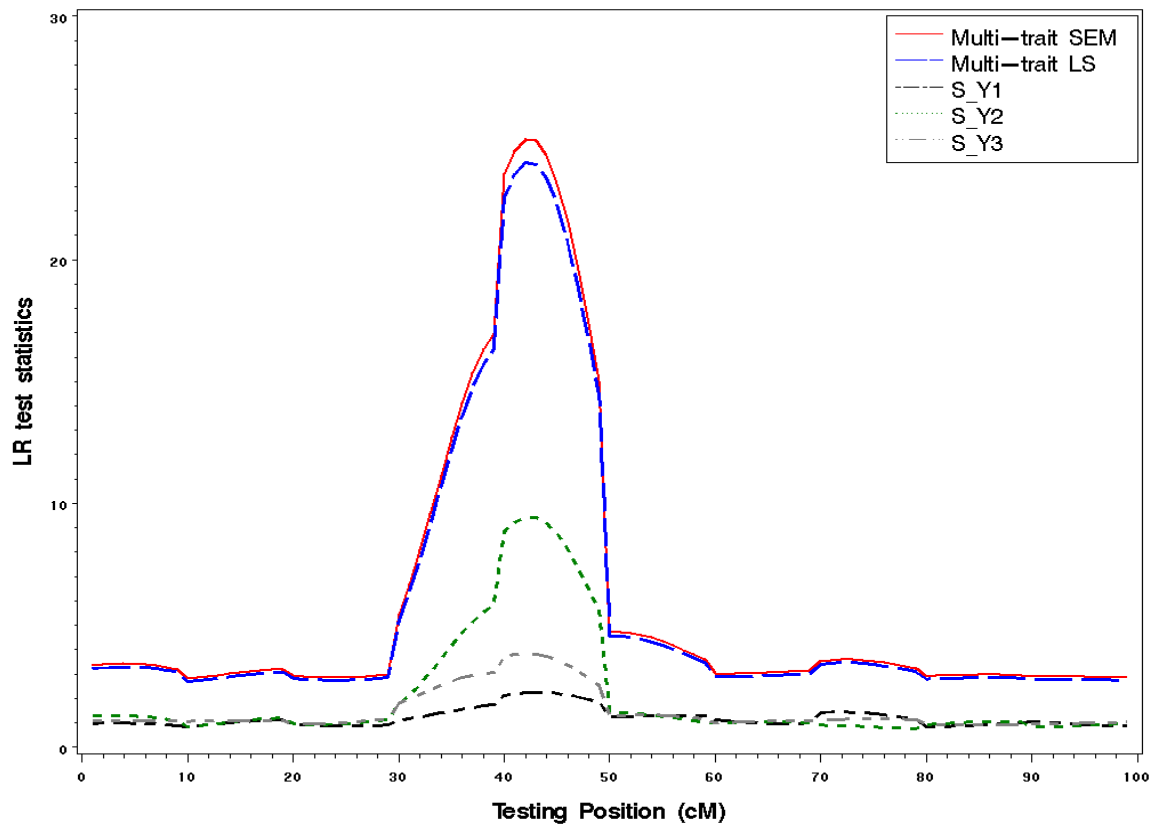
**Table 3.** Parameters and estimates of QTL Positions and effects in the Simulation

Methods	Trait	Position (CM)	Putative QTL Effect		
			Total	Direct	Indirect
Parameters	Y1		0.125	-0.125	0.250
	Y2	42	0.375	0.500	-0.125
	Y3		0.250	0.250	0.000
Multi-trait SEM	Y1		0.127 (0.147)	-0.130 (0.128)	0.258 (0.073)
	Y2	42.2 (1.84)	0.410 (0.151)	0.525 (0.119)	-0.115 (0.076)
	Y3		0.234 (0.152)	0.234 (0.151)	0
Multi-trait LS	Y1		0.125 (0.147)		
	Y2	42.2 (1.84)	0.409 (0.150)		
	Y3		0.234 (0.151)		
Single-trait LS	Y1	50.4 (24.02)	0.126 (0.145)		
	Y2	43.6 (13.75)	0.407 (0.149)		
	Y3	45.1 (21.71)	0.232 (0.151)		

Estimates are means over 100 replicates with standard deviation in parentheses, by the joint mapping on three traits (multi-trait SEM and multi-trait LS) and by the separate mapping on each trait (single-trait LS).



**Figure 1.** The path diagram of the simulation model



**Figure 2.** Results of QTL mapping on three traits from an RIL population using three mapping methods. The average Likelihood ratio test statistics over 100 replicates are plotted at every 1 cM position of a chromosome. Multi-trait SEM and multi-trait LS are the joint mapping on three trait using SEM and least-squares respectively, S\_Y1 is the separate mapping on trait Y1, S\_Y2 is the separate mapping on trait Y2, S\_Y3 is the separate mapping on trait Y3. The data were generated with a single additive QTL located at 42cM to affect three traits. Eleven markers on a 100 cM chromosome were simulated and used in the analyses.