

Analysis of Genotype-by-Environment Interaction in Wheat Using a Structural Equation Model and Chromosome Substitution Lines

P. Dhungana, K. M. Eskridge,* P. S. Baenziger, B. T. Campbell, K. S. Gill, and I. Dweikat

ABSTRACT

The presence of genotype-by-environment interaction (GEI) complicates selection of superior genotypes and an understanding of environmental and genotypic causes of significant GEI is important in all stages of plant breeding. We present a systematic approach for understanding GEI of complex interrelated traits by combining chromosome substitution lines that allowed us to study the effects of genes on a single chromosome with a structural equation model that approximated the complex processes involving genes, environmental conditions, and traits. We applied the approach to recombinant inbred chromosome wheat lines grown in multiple environments. The final model explained 74% of the yield GEI variation and we found that spikes per square meter (SPSM) GEI had the highest direct effect on yield GEI and that the genetic markers were mostly sensitive to temperature and precipitation during the vegetative and reproductive periods. In addition, we identified a number of direct and indirect causal relationships that described how genes interacted with environmental factors to affect GEI of several important agronomic traits which would not have been possible with previously used methods.

P. Dhungana, Monsanto Co., 800 N. Lindbergh Blvd., St. Louis, MO 63167; K.M. Eskridge, P.S. Baenziger, and I. Dweikat, Univ. of Nebraska, Lincoln, NE 68583-0940; B.T. Campbell, USDA-ARS, Florence, SC; K.S. Gill, Washington State Univ., Pullman, WA. Published as Univ. of Nebraska ARD Journal Series no. 14689. Received 26 June 2006. *Corresponding author (keskridge1@unl.edu).

Abbreviations: CNN, cv. Cheyenne; GEI, genotype-by-environment interaction; KPS, kernels per spike; ML, maximum likelihood; P, precipitation; QTL, quantitative trait locus; RICLs, recombinant inbred chromosome lines; RILs, recombinant inbred lines; SEM, structural equation modeling; SPSM, spikes per square meter; SR, solar irradiance; T, temperature; TKW, thousand kernel weight; YLD, yield.

UNDERSTANDING HOW genetic and environmental factors influence complex traits, such as grain yield, is a challenging problem in the plant sciences. Grain yield is dependent on a number of supporting traits influenced by many different genes and by a multitude of environmental conditions at different stages during plant development (Ashikari et al., 2005). In addition, the causal relationships between the environmental factors and the responses of the various traits depend on the genetic background of the plants. These differential relationships are defined as GEI which exists when two or more genotypes respond differently to environmental conditions. In plants, GEI is very important in that a genotype that performs well in one environment, relative to another genotype, may or may not perform well in another environment. However, understanding GEI is difficult since it is the result of many complex interrelationships among physiological, molecular, and environmental variables. An improved understanding of the relationship between environment and genes affecting quantitative traits (recognized as quantitative trait loci [QTL]) would provide considerable insight into how genetic, molecular, and physiological mechanisms are affected by various

Published in *Crop Sci* 47:477–484 (2007).

doi: 10.2135/cropsci2006.06.0425

© Crop Science Society of America

677 S. Segoe Rd., Madison, WI 53711 USA

environmental conditions, especially those that can be controlled in managed ecosystems.

The most commonly used approaches to understanding GEI involve several challenging steps and are limited in their capabilities of characterizing the complex relationships involved in GEI. The first step is to cross at least two genotypes, molecularly characterize hundreds of progeny representing segregants of the entire genome, and evaluate the progeny phenotypically under different environmental conditions (Tanksley et al., 1982; Edwards et al., 1987; Stuber et al., 1987; Lander and Botstein, 1989). Then, the GEI of a number of traits is evaluated for QTL-by-environment interaction where each trait is quantitatively assessed either by comparing results of QTL analyses conducted in individual environments or by using a univariate statistical model with QTL, environment, and QTL-by-environment variables as predictor variables (Stuber et al., 1992; Beavis, 1994; Beavis and Keim, 1996; Vargas et al., 1998, 1999; Crossa et al., 1999). Molecular characterization of hundreds and possibly thousands of progeny across the entire genome is likely not cost effective and using univariate statistical models is not conducive to understanding the complex relationships (epistasis, pleiotropism, etc.) among a number of traits and variables present when describing QTL-by-environment interaction. Moreover, a single dependent variable quantitative approach cannot describe the complicated relationships between traits, QTL, and environments where some traits function simultaneously as both dependent variables to be predicted by other genetic and environmental factors, and as independent predictor variables of other traits. Novel genetic and quantitative methodologies are needed for understanding how genes interact with environment in complex traits.

Here we use an approach for dissecting GEI of complex traits that combines two distinctive components: chromosome substitution lines and structural equation modeling (SEM). Chromosome substitution lines, though rare in animals (Mackay, 1980), are common in wheat (*Triticum aestivum* L., $2n = 6x = 42$) where they have a long history of use in genetic analysis (Sears, 1953). In a wheat chromosome substitution line, 20 of the 21 chromosome pairs come from one cultivar (recurrent parent) and the remaining chromosome pair comes from a second cultivar (donor parent). By comparing the recurrent parent to its chromosome substitution line, the contribution of genes on the substituted chromosome from the donor parent or the loss of genes from the recurrent parent can be identified. Substitution lines effectively dissect the large hexaploid wheat genome into 1/21 increments with a common background, thus greatly reducing background variability and effects that would be found in segregating progeny populations from two parents. The recurrent parent and chromosome substitution lines are true

breeding, hence can be tested in multiple environments. In previous research (Berke et al., 1992), chromosome 3A contained genes affecting anthesis date, grain yield, and yield components.

To dissect the location and relationship of genes on a chromosome, recombinant inbred chromosome lines (RICLs) can be developed (Law, 1966; Berke et al., 1992; Joppa et al., 1997; Shah et al., 1999) which are analogous to recombinant inbred lines (RILs) but in this case 20 chromosome pairs are from the recurrent parent and only the chromosome of interest is allowed to recombine, again reducing the background variability. Since RICL lines are true breeding and can be evaluated in numerous environments for the traits of interest, QTL-by-environment interaction can be precisely evaluated which can give considerable insight into complex gene-environment interactions when used in conjunction with SEM.

Genotype-by-environment interactions are very common in plants, yet understanding the underlying QTL-by-environment interaction is difficult. Most agronomically important traits are the result of a number of genetic, molecular, and physiological mechanisms that affect the trait of interest either directly or indirectly through other intermediate traits (Adams 1967; Thomas et al., 1970; Hamid and Grafius 1978; García del Moral et al., 1991; Dofing and Knight 1992). Depending on the trait, this network among variables can be quite complex with the relationships among variables depending on a number of factors, foremost of which is the stage of development and the environmental conditions during those stages. In a similar fashion, the GEI of each trait in the system will be influenced, either directly or indirectly, by a number of QTL-by-environment interactions and GEI of other traits, which may, in turn, be influenced by other factors (Campbell et al., 2003). Identifying these complex relationships among variables will likely require considerable precision and since RICLs are more powerful for QTL analysis than RILs (Kaeppler, 1997), RICLs will provide more precision for understanding QTL-by-environment interaction.

Structural equation modeling is a generalization of path analysis proposed by Wright (1921) and is used to quantitatively analyze the causal structure among a number of variables where each may function as a dependent variable in some equations and an independent variable in others (Bollen, 1989). Because of this, SEM is ideal for characterizing the complex relationships of GEI where many traits function as both dependent variables to be predicted by environmental and genetic factors, but also as independent predictor variables of other traits further downstream. To use SEM to analyze GEI, prior knowledge of the direction of the causal relationships is assumed and specified through a path diagram and the model is then algebraically specified by a system of regression-type equations where each variable is adjusted to contain only

GEI effects. A final model is then developed by fitting successive models and retaining significant QTL-by-environment variables which result in a better fitting model. The final model yields path coefficients and a path diagram that contains only significant paths thus giving insight into important relationships between the traits, QTL, and the environmental variables. The purpose of this work is to present a systematic approach for understanding GEI of complex traits by using chromosome substitution lines and a structural equation model that approximates the complicated relationships among QTL, environmental variables, and traits.

MATERIALS AND METHODS

Winter Wheat Recombinant Inbred Chromosomes Lines

Field trials with a population of 75 recombinant inbred chromosomes lines (RICLs-3A) were conducted at Lincoln, NE, in 1999 to 2001, and Mead and Sidney, NE, in 2000 and 2001 for a total of seven environments using four replicate randomized complete block field designs in 1999 and four replicate incomplete block field designs in the other years as described in Campbell et al. (2003). The RICLs-3A population was derived from a cross between cv. Cheyenne (CNN) and the chromosome substitution line CNN (Wichita 3A). The agronomic data used in this study were grain yield (YLD) and the yield-component traits 1000 kernel weight (TKW), kernels per spike (KPS), and SPSM. Linkage analysis for chromosome 3A detected nine QTL for the traits measured (Campbell et al., 2003). Six DNA markers (Xtam055, Xbarc86, Xbarc67, Xksua6, Xbcd1555, and Xbcd361), linked to QTL, were considered as genotypic predictors and proxies for previously identified QTL. Climatic data for daily minimum and maximum temperature ($^{\circ}\text{C}$), daily solar irradiance (W m^{-2}), and daily precipitation (mm) at each of the experimental sites in each year were obtained from the High Plains Regional Climate Center, University of Nebraska, Lincoln, NE. On the basis of the winter wheat development model developed by Streck et al. (2003), the winter wheat growing season was divided into three periods based on developmental stages: 1, seedling emergence to terminal spikelet initiation; 2, terminal spikelet initiation to anthesis; and 3, anthesis to physiological maturity. In addition to these three periods, total precipitation during 3 mo before sowing was also used to consider soil moisture before sowing in the analyses. Stored soil moisture is important in arid agriculture. A total of 10 environmental covariates were considered for this study: T1, T2, T3, SR1, SR2, SR3, P0, P1, P2 and P3, where T stands for mean daily temperature; SR stands for total solar irradiance; P stands for total precipitation; the suffix 0 stands for the 3 mo period before sowing; and suffixes 1, 2, and 3 for the first, second, and third development periods, respectively.

Environmental covariates for the first two developmental periods were used to model SPSM and KPS GEIs, since pre-anthesis environmental conditions were most critical for those traits (Donmez et al., 2001). In a similar way, environmental covariates for the last two developmental periods were used to model TKW GEI since this trait is mostly influenced by the

environmental conditions during reproductive and grain-filling periods (Donmez et al., 2001). Since YLD is affected by all yield components, which are associated with different stages of development, all environmental covariates were used to model YLD GEI. Total precipitation for 3 mo before sowing was also used for modeling all yield and yield component GEIs to consider pre-sowing soil moisture.

Definition of Variables

Structural equation modeling requires that variables be identified as either exogenous or endogenous. Exogenous variables are those determined outside of the system and are always used as independent variables (predictor variables) whereas endogenous variables are determined by the system and are used as dependent as well as independent variables in the system of equations. Since our objective was to model yield GEI, observed yield and yield component residuals (YLD_{GEI} , TKW_{GEI} , KPS_{GEI} , and SPSM_{GEI}) were used as the measured endogenous Y variables. These residuals were obtained by subtracting the estimated main effects of genotypes, environments and blocks within environments from observed values. As an example, for variable Y , $Y_{\text{GEI}} = Y - G - E - B(E)$ where G and E are genotype and environment main effects and $B(E)$ is the effect of block nested in environment. The exogenous variables (X_{ij}) were the cross-products of the i th genotypic covariate and j th environmental covariate. Since genotype and environment main effects were removed from the endogenous variables, the only feasible exogenous variables were those that were cross-products of genotypic and environmental covariates, and no environmental or genotypic covariates by themselves were considered in the model. In addition, because yield and yield component residuals were modeled, exogenous variables (X) were also adjusted for main effects. For example, let \mathbf{Y}_i be a column vector of the i th endogenous variable, and, \mathbf{Z} and \mathbf{X} be exogenous variables where \mathbf{Z} is the incidence matrix of the main effects for genotype, environment and block within environment. Then the full linear model is: $\mathbf{Y}_i = \mathbf{Z}\beta_1 + \mathbf{X}\beta_2 + e_i$. After fitting the main effects of genotype, environment, and block within environment (\mathbf{Z}), the vector of residuals is, $\mathbf{Y}_i - \mathbf{Z}\beta_1^0 = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}_i$, where β_1^0 is the estimate for β_1 which equals $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_i$. The resulting reduced model is $(\mathbf{I} - \mathbf{P}_z)\mathbf{Y}_i = (\mathbf{I} - \mathbf{P}_z)\mathbf{X}\beta_2 + d_i$, where $\mathbf{P}_z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and, $d_i = (\mathbf{I} - \mathbf{P}_z)e_i$. As our primary interest was to estimate β_2 in the reduced model, we adjusted variable \mathbf{X} with $\mathbf{I} - \mathbf{P}_z$ before using it as a predictor. This adjustment allowed us to get estimates equivalent to what we would get by fitting the full model and were based on standard linear model theory (Searle, 1971; Ravishanker and Dey, 2002; Dhugana, 2004). All plot observations for all \mathbf{Y} and \mathbf{X} variables used in this study were adjusted for $\mathbf{I} - \mathbf{P}_z$ to remove main effects of genotypes, environments, and blocks as described in the example.

Statistical Model Formulation

Prior knowledge of the directions of causal relationships among the variables as required by SEM, was based on the causal, unidirectional relationships among yield components for small grains proposed by Dofing and Knight (1992). They proposed these relationships among yield components because: (i) yield components develop sequentially, with later-developing components being controlled by earlier-developing ones (Adams,

Table 1. SAS PROC CALIS statements for the final model.†

```
PROC CALIS COVARIANCE METHOD = ML DATA = DSET1 RESIDUAL
TOTEFF MODIFICATION GCONV = .0001 MAXITER = 10000 MAXFUNC
= 10000;
```

```
LINEQS
```

```
SPSMGEI = b1 rmt11 + b2 rmp42 + b3 rmt32 + E1,
KPSGEI = a1 SPSMGEI + b4 rmp10 + b5 rmp41 + b6 rmt61 + E2,
TKWGEI = a2 SPSMGEI + a3 KPSGEI + b7 rms12 + b8 rmt22 + E3,
YLDGEI = a4 SPSMGEI + a5 KPSGEI + a6 TKWGEI + b9 rmt32 + b10
rmp41 + b11 rmp10 + E4;
```

```
STD
```

```
E4 = VARE4,
E3 = VARE3,
E2 = VARE2,
E1 = VARE1;
```

```
run;
```

†rmt11 = Tam055*T1; rmp42 = Bcd1555*P2; rmp32 = Xbarc67*T2; rmp10 = Tam055*P0; rmp41 = Bcd1555*P1; rmt61 = Ksua6*T1; rms12 = Tam055*SR2; rmt22 = Xbarc86*T2. ais and bis are path coefficients associated with yield and yield components residuals and cross-product covariates, respectively. Note: all cross-product covariates are adjusted for main effects. GEI, genotype-by-environment interaction; SPSM, spikes per square meter; KPS, kernels per spike; TKW, thousand kernel weight.

1967; Thomas et al., 1970; Hamid and Grafius, 1978; García del Moral et al., 1991; Dofing and Knight, 1992), and (ii) the unidirectional model more realistically reflects operative causal relationships among yield components than a bidirectional model. The unidirectional path model has been used in most of the recent yield component analysis for small grains (Donaldson et al., 2001; García del Moral et al., 2003; Maman et al., 2004).

The model was algebraically specified using SEM with observed variables (Model 1) (Bollen, 1989):

$$\mathbf{Y} = \mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} + \boldsymbol{\zeta} \quad [1]$$

where \mathbf{Y} is a 4×1 vector of endogenous residual variables: YLD_{GEI} , TKW_{GEI} , KPS_{GEI} , and $SPSM_{GEI}$; \mathbf{B} is the 4×4 coefficient matrix with parameters, much like regression coefficients, that quantify the causal relationships among endogenous variables:

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{34} & \beta_{24} & \beta_{14} \\ 0 & 0 & \beta_{23} & \beta_{13} \\ 0 & 0 & 0 & \beta_{12} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

\mathbf{X} is a $q \times 1$ vector of $(\mathbf{I} - \mathbf{P}_z)$ adjusted exogenous variables, where q = number of cross-products of genotypic covariates with environmental covariates (X_{ij}) retained in the model; $\mathbf{\Gamma}$ is the $4 \times q$ coefficient matrix that shows the causal relationship among endogenous and exogenous variables; $\boldsymbol{\zeta}$ is the 4×1 disturbance vector associated with endogenous variables and is assumed to have $\mathbf{E}(\boldsymbol{\zeta}) = 0$ and covariance matrix, $\mathbf{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}') = \boldsymbol{\Psi}$ and uncorrelated with the exogenous variables. It is also assumed that $(\mathbf{I} - \mathbf{B})$ is nonsingular so that $(\mathbf{I} - \mathbf{B})^{-1}$ exists. Generally, \mathbf{B} is a triangular matrix with zeroes on its diagonal.

It is important to note that Model 1 is a generalization of factorial regression. After simplification, Model 1 becomes $\mathbf{Y} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{X} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}$ and by letting $\mathbf{v} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}$ and $f = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}$, results in a multivariate factorial regression model, $\mathbf{Y} = \mathbf{v}\mathbf{X} + f$. Each row of this model represents a factorial regression model for each dependent variable (YLD_{GEI} , TKW_{GEI} , KPS_{GEI} , and $SPSM_{GEI}$) as a function of $(\mathbf{I} - \mathbf{P}_z)$ adjusted cross product covariates (X_{ij}). Here \mathbf{v} is a $4 \times q$ coefficient matrix showing the

effect of q cross product terms on yield and yield components GEIs. By using SEM, we are able to estimate \mathbf{v} in terms of \mathbf{B} and $\mathbf{\Gamma}$, which provide additional insight into how covariates contribute to yield GEI directly and indirectly via yield component GEIs, how yield component GEIs are interrelated, and which yield component GEI explained most of the yield GEI variance. With $\mathbf{\Gamma} = 0$, Model 1 becomes a path model for yield components residuals (YLD_{GEI} , TKW_{GEI} , KPS_{GEI} , and $SPSM_{GEI}$).

Estimation

Conceptually, SEM parameters are estimated by minimizing the difference between the observed covariance matrix ($\boldsymbol{\Sigma}$ in Eq. [2]) and the model-predicted covariance matrix $\{\boldsymbol{\Sigma}(\theta)$ in Eq. [3], where θ contains the elements of model parameters \mathbf{B} , $\mathbf{\Gamma}$, $\boldsymbol{\Psi}$ (Bollen 1989). The observed covariance matrix ($\boldsymbol{\Sigma}$) and the model-predicted covariance matrix $\{\boldsymbol{\Sigma}(\theta)\}$ were defined as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} \quad [2]$$

and

$$\boldsymbol{\Sigma}(\theta) = \begin{pmatrix} (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})[(\mathbf{I} - \mathbf{B})^{-1}]' & (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Sigma}_{xx} \\ \boldsymbol{\Sigma}_{xx}\boldsymbol{\Gamma}'[(\mathbf{I} - \mathbf{B})^{-1}]' & \boldsymbol{\Sigma}_{xx} \end{pmatrix} \quad [3]$$

where $\boldsymbol{\Sigma}_{yy}$, $\boldsymbol{\Sigma}_{xx}$, $\boldsymbol{\Sigma}_{yx}$ denote the covariance matrices of the observed \mathbf{Y} , \mathbf{X} and between Y and X variables with orders $p \times p$, $q \times q$ and $p \times q$, respectively. Note that the covariances among the exogenous variables $\boldsymbol{\Sigma}_{xx}$ were explicitly incorporated in the model as described in Eq. [3]. The covariance matrix of all variables, $\boldsymbol{\Sigma}$ was of order $(p + q) \times (p + q)$. The model implied covariance matrix, $\boldsymbol{\Sigma}(\theta)$ (Eq. [3]) was obtained by using Eq. [1] and its reduced form $\mathbf{Y} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{\Gamma}\mathbf{X} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}$. The total number of model parameters to be estimated was t = number of path coefficients + number of error variances for the endogenous variables (p) + number of variances and covariances between all exogenous variables $(1/2)[q(q+1)]$. Maximum likelihood (ML) was used to estimate θ where non-linear optimization was used to maximize the ML fit function (Eq. [4]) (Bollen, 1989; Schumacker and Lomax, 1996).

$$F_{ML} = \log|\boldsymbol{\Sigma}(\theta)| + \text{tr}\{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\theta)\} - \log|\boldsymbol{\Sigma}| - (p + q) \quad [4]$$

The analysis was conducted with all plot observations adjusted for genotype, environment, and block within environment effects using the SAS PROC CALIS procedure (SAS Institute, 1999) with the ML estimation method following the steps laid out by Hatcher (1998). The final PROC CALIS program is given in Table 1. Important exogenous variables (X_{ij}) were identified by using an SEM forward selection procedure where the final structural equation model was developed by fitting successive models and retaining significant exogenous variables which resulted in a better-fitting model.

Evaluation of Model Fit

The χ^2 statistic is $(n - 1)$ times the fit function (F_{ML}), where n is number of observations (Schumacker and Lomax, 1996). A nonsignificant chi-square value ($P > 0.05$) indicates that the two matrices $\{\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}(\theta)\}$ are not statistically different. The degrees of freedom (df) associated with χ^2 are deduced as: $df = (1/2)(p + q)(p + q + 1) - t$, where $p + q$ is the number of observed variables analyzed and t is the total number of

parameters estimated. The χ^2 test is sensitive to sample size (Schumacker and Lomax, 1996), however, it is the only statistical test available to test the overall fit of the model. Several goodness-of-fit indices (GFI, AGFI, and NFI) were used to evaluate the fit of the structural equation model. GFI is the percentage of observed covariances explained by the model implied covariances (Pugesek, 2003; Schumacker and Lomax, 1996). GFI = F_{ML}/F_0 , where F_0 is the fit function for the null model (i.e., when $\mathbf{B} = 0$ and $\mathbf{\Gamma} = 0$ in Eq. [1]). When the df is large relative to the sample size, GFI is biased downward (Kline, 1998). The AGFI is a variant of GFI which uses mean squares instead of total sums of squares. Equivalently, the AGFI adjusts the GFI for the df of a model relative to the number of variables (Schumacker and Lomax, 1996; Jöreskog and Sorbom, 1996). NFI reflects the proportion by which the proposed model improves the fit compared to the null model (Bentler and Bonnett, 1980; Schumacker and Lomax, 1996) and is given by $(\chi^2_{null} - \chi^2_{model})/\chi^2_{null}$. GFI, AGFI, and NFI values range from zero to one, with one for the best possible fit.

Nonsignificant χ^2 and GFI, AGFI, and NFI values greater than or equal to 0.90 were used to justify the final model. The squared multiple correlation, R^2 for each yield and yield component was reported. The results were interpreted using standardized coefficients and by calculating the total direct and indirect effects of the cross-product covariates (X_{ij}) and yield-component GEIs on yield GEI.

RESULTS AND DISCUSSION

Six selected DNA marker loci (Xtam055, Xbarc86, Xbarc67, Xksua6, Xbcd1555, and Xbcd361), each closely linked with a different QTL and 10 environmental covariates (T1, T2, T3, P0, P1, P2, P3, SR1, SR2, and SR3) were used to define all possible cross-products which were adjusted for the main effects of genotype and environment as described above and used as exogenous variables. Grain yield and yield components (YLD_{GEI} , TKW_{GEI} , KPS_{GEI} , and $SPSM_{GEI}$) were also adjusted for genotype and environment effects and used as endogenous variables. The final model fit well as demonstrated by a nonsignificant χ^2 ($\chi^2_{(21)} = 14.65$, $P = 0.84$) and the GFI, AGFI, and NFI goodness-of-fit indices all being considerably greater than 0.95. Exogenous variables retained in the final model had a significant path to at least one of the endogenous variables (Fig. 1). The correlations between the exogenous variables in the final model ranged from -0.64 to 0.84 with six coefficients being significant, $P < 0.05$ (data not shown).

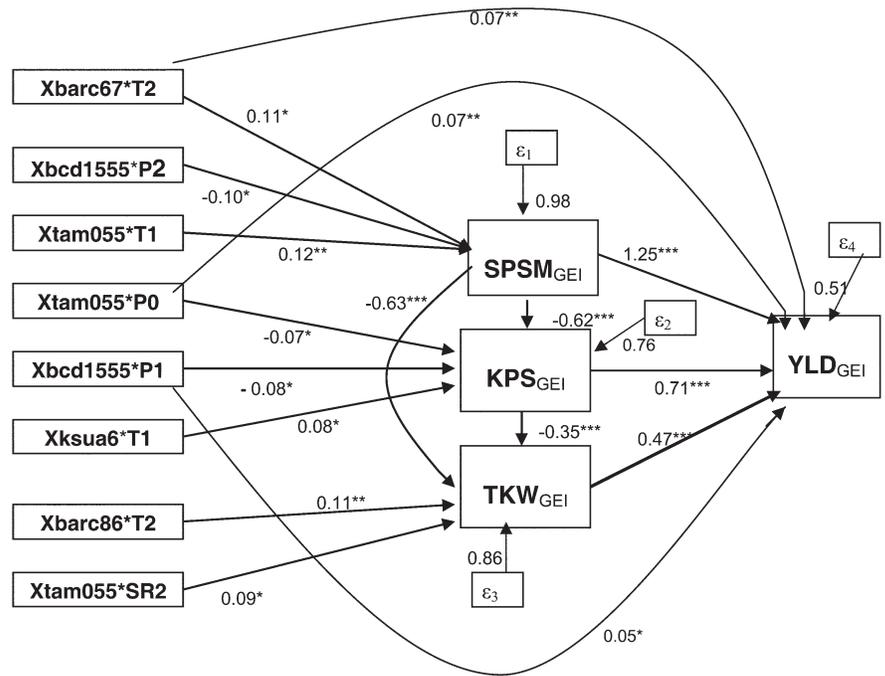


Figure 1. Path estimates of structural equation model of yield genotype-by-environment interaction (YLD_{GEI}), yield component GEIs (spikes per square meter [$SPSM_{GEI}$], kernels per spike [KPS_{GEI}], and thousand kernel weight [TKW_{GEI}]), and adjusted cross-products (marker × environmental covariate). T, mean daily temperature; SR, total solar irradiance; P, total precipitation; suffix 0, the 3-mo period before sowing; 1, the first developmental period; 2, the second developmental period; and 3, the third developmental period. Arrows represent the direction of influence of variables. The numbers by the arrow lines represent the estimated standardized coefficients. The terms ϵ_4 , ϵ_3 , ϵ_2 , and ϵ_1 are the disturbance terms for yield and yield components and represent the unexplained variances of those variables. Covariances among all cross-product variables were estimated as part of the model but arrows among these variables were excluded to simplify the figure. Significance level: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

The R^2 values for all endogenous variables are reported in Table 2. Approximately 74% of the variation in total GEI for grain yield (YLD_{GEI}) was accounted for by the model. Based on the standardized path coefficients, GEI for SPSM had a greater positive direct effect ($P < 0.001$) on yield GEI than KPS and TKW GEIs (Fig. 1). Since SPSM GEI had large negative direct effects on KPS and TKW GEIs, the net total effect of SPSM GEI on yield GEI was less than the direct effect (Table 2). This positive total effect revealed that higher SPSM than what we expect from the additive genotypic and environmental effects resulted in a relatively large positive increase in yield GEI which ultimately increased grain yield. However, only about 3% of the variation in GEI of SPSM was explained by the three marker × environmental covariate interactions retained in the model (Fig. 1). Significant direct positive effects ($P < 0.001$) of KPS and TKW GEIs on yield GEI were also observed but they were smaller than the effect of SPSM GEI.

Eight marker × environmental covariates which had statistically significant path coefficients ($P < 0.05$) were retained in the final model. Of those, Xbarc67 × T2 showed a significant positive direct effect on yield GEI and indirect

Table 2. Direct, indirect and total effects of yield component genotype-by-environment interactions (GEIs) and adjusted cross-product covariates on grain yield GEI ($R^2 = 0.74$).

| Variables [†] | Direct effect | Indirect effect | Total effect | R^{2*} |
|-----------------------------|---------------|-----------------|--------------|----------|
| Thousand kernel weight GEI | 0.47 | 0 | 0.47 | .26 |
| Kernels per spike GEI | 0.71 | -0.16 | 0.55 | .42 |
| Spikes per square meter GEI | 1.25 | -0.64 | 0.61 | .03 |
| Xbarc67 × T2 | 0.07 | 0.07 | 0.14 | |
| Xtam055 × T1 | 0 | 0.07 | 0.07 | |
| Xtam055 × SR2 | 0 | 0.04 | 0.04 | |
| Xtam055 × P0 | 0.07 | -0.04 | 0.03 | |
| Xbarc86 × T2 | 0 | 0.05 | 0.05 | |
| Xbcd1555 × P2 | 0 | -0.07 | -0.07 | |
| Xbcd1555 × P1 | 0.05 | -0.05 | 0 | |
| Xksua6 × T1 | 0 | 0.05 | 0.05 | |

[†]T, mean daily temperature; SR, total solar irradiance; P, total precipitation. The suffix 0 stands for the 3-mo period before sowing; 1, for the first developmental period; 2, for the second developmental period; and 3, for the third developmental period.

^{*} R^2 refers to percentage of variation explained on total GEI for given trait accounted for by the model.

effect through SPSM GEI on yield GEI (Fig. 1). Even after considering the negative indirect effects of Xbarc67 × T2 through SPSM GEI, the net effect of Xbarc67 × T2 on yield GEI was found to be larger than the effects of other covariates (Table 2). The difference between WI and CNN genotypes at Xbarc67, for both SPSM and yield, was sensitive to temperature from terminal spikelet initiation to anthesis (T2) and increased as temperature increased. Two other covariates, Xtam055 × P0 and Xbcd1555 × P1 had direct and indirect effects on yield GEI that were generally smaller than those of Xbarc67 × T2, however, the direct effects were positive, while the indirect effects through KPS_{GEI} were negative (Fig. 1). Campbell et al. (2004) also found a large Xbarc67 × T2 effect on yield GEI, however, they did not find effects of either Xtam055 × P0 or Xbcd1555 × P1. Campbell et al. (2004) likely did not find these covariates related to yield GEI since the total net effects were small and they used factorial regression which is only capable of detecting total effects.

Three marker × environmental covariate interactions (Xtam055 × T1, Xbarc67 × T2, and Xbcd1555 × P2) that were retained in the model were predictors of GEI in SPSM (Fig. 1). The large positive effect of Xtam055 × T1 on SPSM_{GEI} indicated that higher temperatures from seedling emergence to terminal spikelet initiation provided relatively more favorable conditions for WI genotypes at Xtam055 for spike development. Similarly, the positive coefficient between Xbarc67 × T2 and SPSM_{GEI} implied that higher temperatures in the second growth (T2) period were relatively more favorable for the WI genotype at Xbarc67 than CNN in terms of SPSM. The negative coefficient between Xbcd1555 × P2 and SPSM GEI showed that increased precipitation in the repro-

ductive period (P2) was relatively less favorable for the WI genotype at Xbarc67 than CNN in terms of SPSM. Campbell et al. (2004) also found Xtam055 × T1 to have the largest effect on SPSM_{GEI} and although they did not find a relationship between Xbarc67 × T2 and SPSM_{GEI}, they did find similar covariates containing markers near Xbarc67 (Xksua6 × P1, Xksua6 × T1, and Xbcd366 × T1) that were related to SPSM_{GEI}.

The negative relationship between Xbcd1555 × P1 and KPS GEI (KPS_{GEI}) suggested that increased precipitation during vegetative growth (P1) was relatively less favorable for the WI genotype for KPS. However, the small significant direct positive effect of Xbcd1555 × P1 on yield GEI nullified the indirect negative effect of this covariate through KPS_{GEI} (Table 2). KPS_{GEI} was also affected by Xksua6 × T1 and Xtam055 × P0 where the positive effect of Xksua6 × T1 on KPS_{GEI} was interpreted in a manner similar to Xtam055 × T1 on SPSM_{GEI}. The negative relationship between Xtam055 × P0 and KPS_{GEI} indicated that increased soil moisture during the pre-sowing period (P0) was relatively less favorable for WI genotypes compared to CNN at Xtam055. Similar results were reported in Campbell et al. (2004) in that the largest two covariates associated with KPS_{GEI} were Xksua6 × T1 and Xksua6 × P1, however, they did not find significant covariates containing either Xtam055 or Xbcd1555 which possibly could have been due to their factorial regression model failing to account for SPSM_{GEI} as a predictor variable of KPS_{GEI} as in the structural equation model used here.

Both Xbarc86 × T2 and Xtam055 × SR2 were positively related to TKW_{GEI}, while kernel weight GEI (TKW_{GEI}) was affected by Xbarc86 × T2 and Xtam055 × SR2. Campbell et al. (2004) found TKW_{GEI} was related to covariates with markers from similar areas on chromosome 3A (Xbarc86 × SR3, Xbcd366 × SR3, Xbarc67 × SR3). They did not find a covariate with markers close to Xtam055, however, which could have been due to their factorial regression model not including KSP_{GEI} or SPSM_{GEI} as other predictor variables.

These significant relationships between yield and yield component GEIs suggest that our understanding of yield GEI can be enhanced by identifying the genotype and environmental factors that contribute to the GEIs in TKW, KPS, and in particular, SPSM. Hence, in future GEI studies, more effort should be focused on finding factors that predict yield component GEIs to improve our understanding of the molecular character of genes that control important traits.

The positive coefficient between Xbarc67 × T2 and SPSM_{GEI} implied that higher temperatures in the second growth (T2) period were relatively more favorable for the WI genotype at Xbarc67 than CNN in terms of SPSM. The higher temperature may have provided favorable conditions to form more spikes or to prevent the formed spikes

from being aborted. These results are logical in that WI was developed in Kansas and CNN in Nebraska. The higher temperatures during development would be more similar to the “Kansas” environment where WI genotypes would be favored. Campbell et al. (2004) also identified Xbarc67 \times T2 as the largest marker \times environmental covariate interaction responsible for yield GEI using individual factorial regression. However, our results demonstrate a deeper understanding of the yield GEI by showing that the effect of Xbarc67 \times T2 on yield GEI was partly due to its direct effect and partly due to its indirect effect via KPS GEI and TKW GEI. The negative relationships of Xbcd1555 \times P1 with KPS_{GEI} and Xbcd1555 \times P2 with SPSM_{GEI} could mean that WI is more drought tolerant than CNN since it was developed in hotter Kansas environments.

In this study, we have demonstrated a flexible method to study GEI using quantitative traits. This approach targets genes on an individual chromosome via chromosome substitution lines and RICLs and utilizes a single, biologically relevant model incorporating causal relationships among interrelated traits and environmental factors via SEM. Unlike traditional QTL and GEI approaches, a biologically meaningful conclusion can be reached within the scope of this approach, since complex direct and indirect relationships among genetic and environmental variables are accounted for. An “all inclusive” model is increasingly important to consider when studying complex traits of interest (e.g., grain yield) in wheat and other higher plants. Grain yield in wheat is known to be controlled by yield-component traits that are interrelated, have compensatory effects, develop sequentially at different growth stages, and have large GEIs. The large wheat grain-yield GEI associated with genes on chromosome 3A has been explained by direct effects of TKW and direct and indirect effects of SPSM and KPS. Moreover, the known GEI associated with the expression of grain yield QTL previously detected on chromosome 3A can be explained by the direct and indirect effects of yield component QTL-by-environmental covariate interactions represented by molecular markers linked to those QTL.

Many of the most popular methods used to analyze GEI, such as the AMMI model, factorial regression, and stability analysis, are based on univariate models that have a single trait as the dependent variable where each trait is analyzed and interpreted separately (Kang and Gauch, 1996; Vargas et al., 1999; Crossa et al., 1999; Campbell et al., 2004). These univariate approaches can provide considerable insight but generally are not helpful in describing the complex relationships among traits and variables present in GEI. In most GEI studies, each trait is separately analyzed without including other traits as predictors with the consequence of not being able to detect the effects of other traits or to obtain a comprehensive understanding of GEI. Factorial regression and partial least squares models as proposed by Crossa et al. (1999) and Vargas et al. (1999) may be used to incorporate other traits as independent variables but even these models are

limited in that they are not conducive to depicting complex causal relationships among variables and they cannot account for indirect influences of other traits, QTL, or environmental variables acting through intermediate variables on GEI of the trait of interest. Such indirect influences are well known and expected in yield and yield components analysis and are important in understanding GEI of yield.

Structural equation modeling, unlike most univariate approaches, can be used to develop a succinct, graphic, and comprehensive view of how traits, genetic factors, and environmental variables all work together as a system to affect GEI. Structural equation modeling allows one to decompose relationships among traits, QTL, and environmental variables into direct and indirect causal relationships that aid with understanding how genes interact with environmental factors to affect GEI of important agronomic traits. With SEM, the direction of the causal relationships is described through a path diagram and the model is then algebraically specified by a system of regression-type equations meaning that any GEI method based on linear models, such as factorial regression, can be considered a special case of SEM. In general, SEM can provide insight into processes that can be modeled as systems of linear equations, such as the GEI of yield and yield components in wheat.

This study shows that GEI associated with a complex trait such as grain yield is affected by a multitude of genetic and environmental factors. Given the additional insight provided by our approach, using substitution lines in conjunction with SEM is a reasonable approach for studying GEI in other plants and organisms. As genes are identified that control complex and agronomically important traits such as grain yield (Ashikari et al., 2005), this approach can easily be extended to describe how the genes interact with the environment to create the needed phenotypes with higher and more stable grain yields.

Acknowledgments

We thank Drs. R. Morris, S.D. Kachman, and the anonymous referees for helpful reviews of the manuscript. Partial funding for this research was from USDA-IFAFS competitive grant 2001-04462 and USDA, NRICGP 00-353000-9266 and 2004-35300-1470.

References

- Adams, M.W. 1967. Basis of yield component compensation in crop plants, with special reference to the field bean. *Crop Sci.* 7:505–510.
- Ashikari, M., H. Sakakibara, S.Y. Lin, T. Yamamoto, T. Takashi, A. Nishimura, E.R. Angeles, Q. Qian, H. Kitano, and M. Matsuoka. 2005. Cytokinin oxidase regulates rice grain production. *Science* 309:741–745.
- Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 250–266. *In Proc. 49th Annu. Corn and Sorghum Int. Res. Conf.*, Chicago. 8–9 Dec. 1993. Am. Seed Trade Assoc., Chicago, IL.
- Beavis, W.D., and P. Keim. 1996. Identification of QTL that are affected by environment. *In M.S. Kang and H.G. Gauch,*

- Jr. (ed.) New perspectives on genotype by environment interactions. CRC Press, Boca Raton, FL.
- Bentler, P.M., and D.G. Bonnett. 1980. Significance tests and goodness-of-fit in the analysis of covariance structures. *Psych. Bull.* 88:588–606.
- Berke, T.G., P.S. Baenziger, and R. Morris. 1992. Chromosomal location of wheat quantitative trait loci affecting agronomic performance of seven traits using reciprocal chromosome substitutions. *Crop Sci.* 32:621–627.
- Bollen, K.A. 1989. Structural equations with latent variables. John Wiley & Sons, New York.
- Campbell, B.T., P.S. Baenziger, K.M. Eskridge, H. Budak, N.A. Streck, A. Weiss, K.S. Gill, and M. Erayman. 2004. Using environmental covariates to explain genotype \times environment and QTL \times environment interactions for agronomic traits on chromosome 3A of wheat. *Crop Sci.* 44:620–627.
- Campbell, B.T., P.S. Baenziger, K.S. Gill, K.M. Eskridge, H. Budak, M. Erayman, and Y. Yen. 2003. Identification of QTL and environmental interaction associated with agronomic traits on chromosomes 3A of wheat. *Crop Sci.* 43:1493–1505.
- Crossa, J., M. Vargas, F.A. van Eeuwijk, C. Jiang, G.O. Edmeades, and D. Hoisington. 1999. Interpreting genotype \times environment interaction in tropical maize using linked molecular markers and environmental covariables. *Theor. Appl. Genet.* 99:611–625.
- Dhungana, P. 2004. Structural equation modeling of genotype \times environment interaction. Ph.D. diss. Univ. of Nebraska, Lincoln, NE.
- Dofing, S.M., and C.W. Knight. 1992. Alternative model for path analysis of small-grain yield. *Crop Sci.* 32:487–489.
- Donaldson, E., W.F. Schillinger, and S.M. Dofing. 2001. Straw production and grain yield relationships in winter wheat. *Crop Sci.* 41:100–106.
- Donmez, E., R.G. Sears, J.P. Shroyer, and G.M. Paulsen. 2001. Genetic gain in yield attributes of winter wheat in the Great Plains. *Crop Sci.* 41:1412–1419.
- Edwards, M.D., C.W. Stuber, and J.F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116:113–125.
- García del Moral, L.F., J.M. Ramos, M.B. García del Moral, and P. Jimenez-Tejada. 1991. Ontogenetic approach to grain production in spring barley based on path coefficient analysis. *Crop Sci.* 31:1179–1185.
- García del Moral, L.F., Y. Rharrabti, D. Villegas, and C. Royo. 2003. Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: An ontogenetic approach. *Agron. J.* 95:266–274.
- Hamid, Z.A., and J.E. Grafius. 1978. Developmental allometry and its implication to grain yield in barley. *Crop Sci.* 18:83–86.
- Hatcher, L. 1998. A step-by-step approach to using the SAS system for factor analysis and structural equation modeling. SAS Institute Inc., Cary, NC.
- Joppa, L.R., D. Changheng, G.E. Hart, and G.A. Hareland. 1997. Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. *Crop Sci.* 37:1586–1589.
- Jöreskog, K.G., and D. Sorbom. 1996. LISTREL 8: User's reference guide. Scientific Software International, Chicago, IL.
- Kaepler, S.M. 1997. Quantitative trait locus mapping using sets of near-isogenic lines: Relative power comparisons and technical considerations. *Theor. Appl. Genet.* 95:384–392.
- Kang, M.S., and H.G. Gauch (ed.). 1996. Genotype-by-environment interaction. CRC Press, Boca Raton, FL.
- Kline, R.B. 1998. Principles and practice of structural equation modeling. The Guilford, New York.
- Lander, E., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199.
- Law, C.N. 1966. The location of genetic factors affecting a quantitative character in wheat. *Genetics* 53:487–498.
- Mackay, T.F.C. 1980. Genetic variance, fitness and homeostasis in varying environments— an experimental check of the theory. *Evolution Int. J. Org. Evolution* 34:1219–1222.
- Maman, N., S.C. Mason, D.J. Lyon, and P. Dhungana. 2004. Yield components of pearl millet and grain sorghum across environments in the central Great Plains. *Crop Sci.* 44:2138–2145.
- Pugesek, B.H. 2003. Concepts of structural equation modeling in biological research. p. 42–59. *In* B.H. Pugesek et al. (ed.) Structural equation modeling—application in ecological and evolutionary biology. Cambridge Univ. Press, Cambridge, UK.
- Ravishanker, N., and D.P. Dey. 2002. A first course in linear model theory. Chapman and Hall, New York.
- SAS Institute. 1999. SAS/STAT. Release 8.2. SAS Inst., Cary, NC.
- Schumacker, R.E., and R.G. Lomax. 1996. A beginner's guide to structural equation modeling. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Searle, S.R. 1971. Linear models. John Wiley & Sons, New York.
- Sears, E.R. 1953. Nullisomic analysis in common wheat. *Am. Nat.* 87:245–252.
- Shah, M.M., P.S. Baenziger, and Y. Yen. 1999. Genetic analyses of agronomic traits controlled by wheat chromosome 3A. *Crop Sci.* 39:1016–1021.
- Streck, N.A., A. Weiss, and P.S. Baenziger. 2003. A generalized vernalization response function for winter wheat. *Agron. J.* 95:155–159.
- Stuber, C.W., M.D. Edwards, and J.F. Wendel. 1987. Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Sci.* 27:639–648.
- Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris, and E.S. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from 2 elite maize inbred lines using molecular markers. *Genetics* 132:823–839.
- Tanksley, S.D., H. Medina-Filho, and C.M. Rick. 1982. Use of naturally-occurring enzyme variation to detect and map genes-controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49:11–25.
- Thomas, R.L., J.E. Grafius, and S.K. Kahn. 1970. Genetic analysis of correlated sequential characters. *Heredity* 26:177–188.
- Vargas, M., J. Crossa, K. Sayre, M. Reynolds, M.E. Ramirez, and M. Talbot. 1998. Interpreting genotype by environment interaction in wheat using partial least square regression. *Crop Sci.* 38:679–689.
- Vargas, M., J. Crossa, F.A. van Eeuwijk, M.E. Ramirez, and K. Sayre. 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype \times environment interaction. *Crop Sci.* 39:955–967.
- Wright, S. 1921. Correlation and causation. *J. Agric. Res.* 20(10):557–585.