Theor Appl Genet (1999) 99:611–625

J. Crossa · M. Vargas · F.A. van Eeuwijk · C. Jiang
G.O. Edmeades · D. Hoisington

Interpreting genotype × environment interaction in tropical maize using linked molecular markers and environmental covariables

Received: 10 January 1999 / Accepted: 12 March 1999

Abstract An understanding of the genetic and environmental basis of genotype×environment interaction (GEI) is of fundamental importance in plant breeding. In mapping quantitative trait loci (QTLs), suitable genetic populations are grown in different environments causing QTLs×environment interaction (QEI). The main objective of the present study is to show how Partial Least Squares (PLS) regression and Factorial Regression (FR) models using genetic markers and environmental covariables can be used for studying QEI related to GEI. Biomass data were analyzed from a multi-environment trial consisting of 161 lines from a $F_{3:4}$ maize segregating population originally created with the purpose of mapping QTLs loci and investigating adaptation differences between highland and lowland tropical maize. PLS and FR methods detected 30 genetic markers (out of 86) that explained a sizeable proportion of the interaction of maize lines over four contrasting environments involving two low-altitude sites, one intermediate-altitude site, and one high-altitude site for biomass production. Based on a previous study, most of the 30 markers were associated with QTLs for biomass and exhibited significant QEI. It was found that marker loci in lines with positive GEI for the highland environments contained more highland alleles, whereas marker loci in lines with positive GEI for intermediate and lowland environments contained more lowland alleles. In addition, PLS and FR models identified maximum temperature as the most-important environmental covariable for GEI. Using a stepwise variable selection procedure, a FR model was constructed for GEI and QEI that exclusively included cross products between genetic markers and environmental covariables. Higher maximum temperature in low- and intermediate-altitude sites affected the expression of some QTLs, while minimum temperature affected the expression of other QTLs.

Key words Biplot · Factorial regression · Genetic marker · Genotype×environment interaction · Quantitative trait loci · Quantitative trait loci × environment interaction · Partial least squares regression

Communicated by P.M.A. Tigerstedt

J. Crossa (✉) · C. Jiang · G.O. Edmeades · D. Hoisington
International Maize and Wheat Improvement Center (CIMMYT),
Lisboa 27, Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico
e-mail: jcrossa@cimmyt.mx
Fax: +52 5804 7558

M. Vargas
Universidad Autónoma Chapingo, CP 56230, Chapingo,
Mexico and CIMMYT

F.A. van Eeuwijk
Department of Agriculture,
Environmental and Systems Technology,
Wageningen Agricultural University, Dreijenlaan 4,
6703 HA Wageningen, The Netherlands

Introduction

Genotypes grown in multi-environment trials react differently to environmental changes such as maximum and minimum temperature, radiation, soil characteristics, and precipitation. This differential response of genotypes from one environment to another is called genotype×environment interaction (GEI). An understanding of the genetic basis of adaptation and its physiological and environmental causes is of fundamental importance for understanding GEI, for assessing the association between phenotypic and genotypic values, and for enhancing the selection of superior and stable genotypes.

In order to map specific genomic segments affecting quantitative traits [quantitative trait loci (QTLs)] with the aid of molecular markers, a set of families (or lines) from a suitable genetic population such as an F_2, backcross, a recombinant inbred, or doubled haploids are grown in different environments. Various statistical models and procedures are used to detect and estimate the effect and position of QTLs (Lander and Botstein 1989; Knapp et al. 1990; Martinez and Curnow 1992; Jansen and Stam 1994; Zeng 1994). In these mapping studies, QTLs with large effect in some environments and no ef-

fect in others are commonly found; their variable effects constitute QTL × environment interaction (QEI). Procedures developed by Jiang and Zeng (1995) for estimating the effect of QTLs for multiple traits can be used to test the significance of QEI. Another method for studying QEI relying on least-squares interval mapping based on multiple regression was presented by Sari-Gorla et al. (1997).

When a significant QEI is detected, the estimation of the position and effects of the QTLs should be made for each environment. Romagosa et al. (1996) defined phenotypic principal-component scores obtained from the interaction in an Additive Main effect and Multiplicative Interaction (AMMI) model (Gollob 1968; Mandel 1971; Kempton 1984; Gauch 1988) as the traits to be mapped for adaptation in barley (*Hordeum vulgare* L.), and pattern analysis for studying the differential genotypic expression across environments. Recently, Balfourier et al. (1997) used the Factorial Regression (FR) model (Denis 1988; van Eeuwijk 1996; van Eeuwijk et al. 1996) in an attempt to interpret GEI in ryegrass (*Lolium perenne* L.) populations using isozyme and environmental data as external covariables. However, because the study did not involve QTLs mapping, QEI could not be quantified.

In QTL mapping it is assumed that genetic markers associated with the phenotypic expression of a trait are likely to be closely linked to genomic regions (QTLs) that affect the trait. In other words, it is assumed that there is a correlation between segregating alleles at marker loci and their linked QTLs, and that these QTLs give rise to associations of specific genetic markers with phenotypic values for the trait. For quantitative traits, the aim is to estimate the effect and position of QTLs (linked to some genetic marker) by means of statistical models and procedures. Thus, a logical approach to studying and interpreting the QEI of a given trait would be to examine the influence of the linked genetic markers (proven to be associated with specific QTL) on the GEI of that trait to see if the genetic markers that explain a large proportion of the GEI are associated with QTLs having a large QEI. Furthermore, as Balfourier et al. (1997) showed for isozymes, it would be useful to examine how much of the GEI for the trait is explained by the interaction between specific genetic markers (associated with QTLs) with particular environmental variables. A more-detailed description of the GEI and QEI, using molecular markers linked to QTLs and environmental variables, would help to identify climatic variables that affect the expression of some QTLs in some environments.

As the number of known markers increases, the marker map becomes denser. This makes it reasonable to assume that soon we will be able to completely link some markers to some specific QTLs. At that time, modelling QTLs by direct regression on markers will be feasible. Of course, this approach will require special measures for dealing with (almost) collinear marker information. In this paper we investigate the behavior of regression models, with protection against collinearity, for prescreening large numbers of genetic markers for possible inclusion in QTLs models. The same models provide the means to look at GEI in terms of markers and environmental covariables. In contrast, because of the large number of correlated markers, standard QTLs models will, as a rule, not efficiently survey all possible sets and will not be appropriately protected against collinearity.

Genotype×environment interaction has been studied using several statistical models (Crossa 1990). Some of them, such as the analysis of variance, the regression on the environmental mean model (Yates and Cochran 1938; Finlay and Wilkinson 1963; Eberhart and Russell 1966), and multiplicative models (such as the AMMI model), use only the phenotypic-response variable of interest. The environmental and genotypic interaction scores of the bilinear terms are estimated using statistics derived from the observed phenotypic data. No external information (e.g., on genotypes or environments) can be directly incorporated into these models. A shortcoming of standard statistical methods for studying the GEI of quantitative traits is that they measure the average GEI across the entire genome and do not take into consideration the possibility that different segments of the chromosomes (that determine the quantitative trait) can react differently to changing environmental conditions, and therefore have differential contributions to the total GEI.

When information is available on external environmental and/or genotypic variables, such as climatic data, soil information, disease or genetic markers, other statistical models, including Factorial Regression (FR) models and Partial Least Squares (PLS) regression (Aastveit and Martens 1986; Talbot and Wheelwright 1989; Vargas et al. 1998, 1999), can be used to examine which of these external environmental or genotypic variables influence the GEI of the trait. The non-linear PLS regression model is useful when two different (multivariate) measurement systems must be related, and when the large number of variables measured in each system should be reduced to a smaller number of, hopefully, more interpretable latent variables (commonly called PLS factors). In the context of multi-environment trials and GEI, PLS relates the two-way table of genotypes and environments of a given trait (response variable) to external environmental or genotypic variables (explanatory variables) in a single estimation procedure. Any number of environemntal and/or genotypic variables (collinear or not) can be included in PLS; the explanatory variables are linear combinations of the complete set of measured environmental and/or genotypic covariables.

Factorial regression models are ordinary linear models that explain GEI by differential genotypic sensitivity to specific environmental factors. An advantage of these models is that hypotheses about the influence of the external variables on the GEI of the trait can be statistically tested. Recently, Vargas et al. (1999) compared results obtained from PLS and FR in two large multi-environment trials and showed that both methods identified similar environmental and genotypic covariables that explain a large proportion of the GEI. They also showed that when a large number of external covariables are

used, PLS and FR are useful for identifying subsets of the most relevant external covariables affecting GEI. These covariables can be further introduced in a multiple factorial regression that includes combinations of environmental and/or genotypic covariables and their cross products, to explain a large proportion of GEI with relatively few degrees of freedom (*df*).

The main objectives of this study are to show how regression methods such as the Partial Least Squares Regression and the Factorial Regression models, together with genetic markers and environmental covariables (such as maximum and minimum temperature and sun hours), can be used to: (1) detect relevant sets of correlated markers and environmental variables that explain a significant proportion of the total GEI, and (2) study the influence of environmental variables on the expression of QTLs with the objective of assessing and interpreting the QEI that accounts for GEI. Data used were from a multi-environment trial consisting of biomass measures of maize lines from a $F_{3:4}$ segregating population derived with the purpose of mapping QTLs associated with adaptation differences between highland and lowland tropical maize (Jiang et al. 1999). Large GEIs for biomass, grain yield and harvest index have been reported by Lafitte and Edmeades (1997) and Lafitte et al. (1997) for highland and lowland genotypes when they were evaluated at sites with average mean temperatures ranging from 17°C to 28°C.

Materials and methods

Theory

Multiplicative models for describing GEI, such as FR or AMMI, are useful because they most-often use fewer *df* than the analysis of variance and express the GEI as a string of product (bilinear) terms comprising line sensitivities to critical environmental factors. However, while AMMI does not use explicit environmental variables, FR does. A full description of the FR models and their applications for interpreting GEI using environmental and/or line covariables are given in van Eeuwijk (1996). Vargas et al. (1998) and Vargas et al. (1999) described the theory of PLS in the context of GEI and gave details of its algorithm. Here, FR and PLS models are briefly described using, for simplicity, the same notation as Vargas et al. (1999).

Factorial regression models

FR models have a multiplicative structure for the interaction term. The estimate for the classical analysis of variance GEI is the residual table consisting of the two-way table of means corrected for line and site main effects, $(GEI)_{ij}=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}$ (where $\bar{y}_{ij.}$ is the mean of the i^{th} line on the j^{th} environment and $\bar{y}_{i..}$, $\bar{y}_{.j.}$, and $\bar{y}_{...}$ are the means of the i^{th} line and the j^{th} environment, and the overall mean, respectively).

GEI is modelled directly in relation to environmental covariables (with the regression coefficient depending on the line), or in relation to line covariables (with the regression coefficient depending on the environment). A FR model for the mean of the i^{th} line in the j^{th} environment, for which the interaction includes G (centered) line covariables x_{i1} to x_{iG}, can be written in matrix notation as

$$E(\mathbf{Y})=\mu\mathbf{1}_I\mathbf{1}'_J+\boldsymbol{\tau}\mathbf{1}'_J+\mathbf{1}_I\boldsymbol{\beta}'+\mathbf{X}\boldsymbol{\Gamma}' \quad (1)$$

where $\mathbf{Y}=(y_{ij})$ is a I×J matrix that contains the response variable biomass of lines in environments; μ is a scalar representing the overall mean; $\mathbf{1}_I$ and $\mathbf{1}_J$ are I×1 and J×1 unit vectors, respectively; $\boldsymbol{\tau}=(\tau_i)$ is the line main-effect vector of size I×1; and $\boldsymbol{\beta}=(\beta_j)$ is a J×1 vector representing the main effect of sites. The GEI consists of the product of the known line covariables, x_{i1} to x_{iG} (G≤I-1), represented by the I×G matrix $\mathbf{X}=(x_{ig})$ multiplied by the unknown environmental effects (potentialities), γ_{j1} to γ_{jG}, denoted by the J×G matrix $\boldsymbol{\Gamma}=(\gamma_{jg})$. Convenient constraints on the parameters are sum-to-zero over i for the parameters τ_i and over j for β_j and γ_{jg}. The line covariables are known, but the environmental potentialities should be estimated.

A FR model in which the GEI term contains H (centered) environmental covariables, z_{j1} to z_{jH}, can be written as

$$E(\mathbf{Y})=\mu\mathbf{1}_I\mathbf{1}'_J+\boldsymbol{\tau}\mathbf{1}'_J+\mathbf{1}_I\boldsymbol{\beta}'+\boldsymbol{\zeta}\,\mathbf{Z}'. \quad (2)$$

The first three additive terms are the same as in equation 1. The GEI term consists of the product of lines having differential effects (sensitivity), ζ_{i1} to ζ_{iH} (H≤J-1), that are collected in the I×H matrix $\boldsymbol{\zeta}=(\zeta_{ih})$ multiplied by the values of the environmental covariables, that are collected in the J×H matrix $\mathbf{Z}=(z_{jh})$. The values of the environmental variables are known, but the line sensitivities need to be estimated.

The structure of the FR model, including both line and environmental covariables simultaneously, is a logical extension of equations 1 and 2 (Denis 1988; van Eeuwijk et al. 1996) and can be written as

$$E(\mathbf{Y})=\mu\mathbf{1}_I\mathbf{1}'_J+\boldsymbol{\tau}\mathbf{1}'_J+\mathbf{1}_I\boldsymbol{\beta}'+\mathbf{X}\boldsymbol{\nu}\,\mathbf{Z}'+\mathbf{X}\boldsymbol{\Gamma}'+\boldsymbol{\zeta}\,\mathbf{Z}', \quad (3)$$

where the G×H matrix $\nu=(\nu_{gh})$ denotes the regression coefficients to cross products of line covariable x_g with environmental covariable z_h. Additional constraints are $\boldsymbol{\zeta}'\mathbf{X}=\mathbf{Z}'\boldsymbol{\Gamma}=\mathbf{0}$ (where $\mathbf{0}$ is a H×G matrix of zeros). As shown by van Eeuwijk et al. (1996), environmental and line covariables may be quantitative and qualitative, and more complicated FR models are possible by combining quantitative and qualitative covariables. Note in equation 3 that because the line and environmental covariables are centered, elements of the matrices $\mathbf{X}$ (value of the line covariable g on the line i) and $\mathbf{Z}'$ (value of the environmental covariable h on the site j) are positive (above average) or negative (below average). These signs combine with the positive or negative regression coefficient of the cross product between the line covariable x_g with the environmental covariable z_h to produce a $\mathbf{X}\boldsymbol{\nu}\,\mathbf{Z}'$ interaction term that can be positive or negative.

Partial least squares regression

The main objective of the PLS method is to identify a linear combination of the explanatory variables that provides latent vectors that optimally predict the response variable using an iterative procedure. The number of PLS factors to be retained is determined by a cross-validation procedure (Stone 1974) and the *F*-test proposed by Osten (1988). For the multivariate PLS, the response variable (biomass) is represented by the matrix $\mathbf{Y}$ of line performance on environments and the matrix $\mathbf{Z}=(\mathbf{z}_1, ..., \mathbf{z}_S)$ represents S environmental explanatory variables, such as temperature, precipitation, etc. These matrices can be expressed in a bilinear form as

$$\mathbf{Z}=\mathbf{T}\mathbf{P}'+\mathbf{E} \text{ and} \quad (4)$$

$$\mathbf{Y}=\mathbf{T}\mathbf{Q}'+\mathbf{F} \quad (5)$$

where matrix $\mathbf{T}$ contains the Z-scores, matrix $\mathbf{P}$ has the Z-loadings, matrix $\mathbf{Q}$ contains the Y-loadings, and $\mathbf{E}$ and $\mathbf{F}$ are the residual matrices. It is clear from equations 4 and 5 that the relationship between $\mathbf{Z}$ and $\mathbf{Y}$ is transmitted through the latent variables of matrix $\mathbf{T}$.

Therefore, when GEI is explained using S environmental covariables ($\mathbf{Z}$), Vargas et al. (1999) described the above equations using the transposition of $\mathbf{Y}$ such that, for $\mathbf{T}=\mathbf{Z}\mathbf{W}$ and $\boldsymbol{\zeta}=\mathbf{Q}\mathbf{W}'$, $E(\mathbf{Y}')=(\mathbf{T}\mathbf{Q}')'=\mathbf{Q}\mathbf{W}'\mathbf{Z}'=\boldsymbol{\zeta}\mathbf{Z}'$ (the same as the last term of equation 2). The rows of matrix $\mathbf{T}$ contain the Z-scores indexed by environ-

ments; the rows of matrix **W** have the Z-weights indexed by the environmental covariables; the rows of the matrix **Q** include the Y-loadings indexed by lines; and matrix ζ has the PLS approximation to the regression coefficients of **Y** to the explanatory covariables **Z**. When the GEI is explained using the K-line covariables represented by matrix $\mathbf{X}=(\mathbf{x}_1, ..., \mathbf{x}_K)$, then $\mathbf{T}=\mathbf{XW}$ and $\mathbf{\Gamma}=\mathbf{QW'}$, so that $E(\mathbf{Y})=\mathbf{TQ'}=\mathbf{XWQ'}=\mathbf{X\Gamma}$" (the same as the last term of equation 1). The rows of **T** have X-scores for lines; the rows of **W** include X-weights indexed by line variables; the rows of **Q** contain the Y-loadings of the environments; and **Γ** has PLS approximation to the regression coefficients of **Y** to the explanatory covariables **X**.

Results of the bilinear decomposition obtained from PLS can be summarized in a graphical form (biplot) that includes the representation of lines, environments, and covariables, i.e., matrices **T**, **W**, and **Q** are shown in the same biplot. The PLS biplot approximates the interactions of lines on environments (projections of rows of **T** on the rows of **Q** or vice versa) and it also approximates the regression coefficients of lines (environments) on environmental (line) covariables (projection of rows of **W** on the rows of **Q** or vice versa). A perpendicular projection of the lines on one site vector, extended in either direction, gives the relative values of the lines for the GEI.

Experimental data

Phenotypic and genotypic data are the same as used by Jiang et al. (1999) for identifying QTLs associated with the adaptation differences of 161 lines of an $F_{3:4}$ population derived from the cross between lowland lines derived from Population 21 (Johnson et al. 1986) and highland lines derived from germplasm collected in Peru, Mexico, Colombia, Bolivia, and Ecuador (Eagles and Lothrop, 1994). The 161 lines were tested in four contrasting environments of Mexico: Poza Rica (60 masl) winter season (PR), Tlaltizapan (940 masl) summer season (TL), El Batan (2240 masl) summer season (BA), and Toluca (2650 masl) summer season (TO). Several traits were measured in each of the 161 lines, but only biomass, measured as the total weight above-ground (g m^{-2}), is included in this study. The 161 lines were arranged in an alpha-lattice design with two replicates at each site. Composite interval mapping (CIM) procedures (Jansen and Stam 1994; Zeng 1994) were used for QTL analysis. Full details of the marker genotype determination, linkage map construction, QTL analysis, critical values used for QTL detection, and test of QTL×site interaction are given in Jiang et al. (1999).

The 161 lines were genotyped at 86 RFLP (Restrictive Fraction Length Polymorphic) loci, and results coded by the number of alleles from the highland parent as 2, 1 and 0, which represent the homozygous highland, heterozygous and homozygous lowland lines, respectively. As some marker data were missing or dominant (so that two or three lines are possible) the expected value of the number of alleles from the highland parent was calculated (Jiang and Zeng 1997); this expected value lies between 0 and 2. When these coded values are used as the independent variables in regression, the estimated regression coefficients represent the additive effect of the marker locus, which is generally regarded as the replacement effect (allele substitution effect or additive effect) of the lowland allele by the corresponding highland allele at the marker locus in the relevant environment.

The 86 RFLP markers were used in PLS regression and in the FR model as line variables and were named as *a*(*b*) where "*a*" denotes the chromosome number and "*b*" is the marker number within chromosome "*a*". Numbers denoted by "*b*" are in sequential order in relation to their location on the chromosome [i.e., 5(1), 5(2),..., 5(10) indicate that markers 1 and 2 are located closer together in chromosome 5 than markers 1 and 10].

In each of the sites, the crop cycle was divided into three stages: the vegetative stage (from sowing to 2 weeks before flowering), the flowering stage (from 2 weeks before flowering to two weeks after flowering), and the grain-filling stage (from 2 weeks after flowering until harvest date). The nine climatic covariables used in the PLS regression and in the FR model were: mean daily maximum temperature during the vegetative stage, flowering stage, and grain-filling stage (MTV, MTF, and MTG, respectively); mean daily minimum temperature during the vegetative stage, flowering stage, and grain-filling stage (mTV, mTF, and mTG, respectively); and sun hours per day during the vegetative stage, flowering stage, and grain-filling stage (SHV, SHF, and SHG, respectively). Daily maximum and minimum temperatures and sun hours were recorded at a meteorological station 200 m or less from the field trials.

When line covariables are used in the PLS regression, the response matrix of biomass measurements, **Y**, has 161 rows (lines) and 4 columns (sites), whereas the matrix of explanatory line covariables (genetic markers), **X**, has 161 rows (lines) and 86 columns (genetic markers). When environmental covariables are considered in the PLS analysis, the response matrix of biomass measurements, **Y**, has 4 rows (sites) and 161 columns (lines), whereas the matrix of explanatory climatic variables, **X**, has 4 rows corresponding to sites and 9 columns corresponding to climatic variables. For the PLS approach, the **Y** variables correspond to the line × site interaction matrix (residual matrix after adjusting for line and site main effects).

Results and discussion

The analysis of variance showed that 61% of the total sum of squares is explained by site mean differences, 13% by site×line interaction, and 9% by differences among line means (all were highly significant). As expected, the large genetic variability of the lines of the $F_{3:4}$ population and the environmental differences of the four testing sites made the GEI of biomass large and complex. This result agreed with that of Lafitte et al. (1997) who showed a large GEI of maize genotypes for yield and yield components when tested in different thermal regimes. The following analyses will attempt to explain the GEI of biomass in terms of genetic markers linked to QTLs and environmental variables.

Explaining line×site interaction using partial least squares and individual factorial regression with genetic markers as explanatory variables

Individual FRs with each genetic marker covariable were performed to determine their relative contribution to the GEI sum of squares. Although 40 individual factorial regressions (each using 3 *df*) for the genetic markers' covariables were significant at the 5% level, the first 30 markers with the largest sum of squares (Table 1) are the most interesting.

The cross-validation assessment and Osten's *F*-test for the number of significant PLS factors indicated that the first two PLS factors (out of 86) were significant for prediction. The first and second factors explained 17% and 10% of the variability in the GEI matrix of biomass, respectively. Table 1 shows the loadings for the first and second PLS factors of the 30 most-important genetic markers ordered by the first PLS factor. From these 30 markers, 27 correspond to markers determined by individual FRs as the most important covariables for explaining the GEI of biomass (Table 1). The rank order of the genetic markers in relation to their contribution to

Table 1 The 30 most-important markers, as determined by individual FR, with the sum of squares arranged in decreasing order of importance with respect to their contribution in explaining site×line interaction, and **X** loadings of the first two PLS factors (PLS1 and PLS2) of the 30 most-important genetic markers ordered by decreasing loading values for the first PLS factor (PLS1)

Source	Individual FR			Marker	PLS loadings	
	df	Sum of squares ×10^{-5}	Prob >*F*		PLS1	PLS2
Site×line	480	149.271	0.0000			
2(1)	3	9.995	0.0000	2(1)	0.2756	0.0999
6(5)	3	6.538	0.0000	5(3)	0.2185	–0.0451
6(6)	3	5.964	0.0000	2(2)	0.2149	0.0448
2(2)	3	5.809	0.0000	5(5)	0.1916	0.0265
5(3)	3	5.704	0.0000	10(6)	0.1831	0.0510
6(4)[a]	3	5.284	0.0000	3(5)	0.1821	–0.0127
5(5)	3	4.653	0.0000	1(8)	0.1693	–0.1311
3(5)	3	4.558	0.0001	9(1)	–0.1657	0.0563
10(6)	3	4.521	0.0001	1(4)	0.1616	–0.0734
9(1)	3	3.978	0.0004	5(4)	0.1591	0.0177
10(7)	3	3.873	0.0005	10(9)	0.1577	0.0633
10(9)	3	3.866	0.0005	1(5)	0.1574	–0.0336
1(4)	3	3.753	0.0006	3(6)	0.1556	–0.0380
1(8)	3	3.726	0.0006	1(9)	0.1544	–0.1201
3(6)	3	3.613	0.0008	5(6)	0.1542	0.0391
1(7)	3	3.577	0.0009	6(5)	0.1489	0.2641
5(6)	3	3.520	0.0010	9(3)	–0.1482	0.1061
9(3)	3	3.496	0.0011	9(2)	–0.1438	0.1092
1(5)	3	3.475	0.0011	3(4)	0.1411	0.0310
5(4)	3	3.468	0.0012	6(6)	0.1377	0.2553
10(8)[a]	3	3.437	0.0012	5(8)	0.1361	–0.0973
9(2)	3	3.250	0.0018	1(10)	0.1347	–0.1561
1(2)	3	3.107	0.0024	4(3)[b]	0.1337	–0.1201
1(9)	3	3.073	0.0026	4(4)[b]	0.1224	–0.0875
3(4)	3	2.970	0.0033	1(3)	0.1222	–0.0681
5(8)	3	2.812	0.0046	10(5)[b]	0.1205	0.0130
1(10)	3	2.687	0.0060	1(7)	0.1196	–0.1931
3(7)	3	2.661	0.0063	3(7)	0.1171	0.0368
1(3)	3	2.598	0.0072	10(7)	0.1158	0.1694
8(6)[a]	3	2.596	0.0073	1(2)	0.1153	0.1155
Pooled error	536	114.628				

[a] Genetic markers 6(4), 10(8) 8(6) ranked 37th, 39th and 63th, respectively, by the first PLS factor

[b] Within the 40 significant genetic markers determined by individual FR

explaining line×site interaction of biomass was very similar for the PLS method (ranking on the absolute size loading on the first axis) and the FR model (ranking on the SS explained). Note that although markers 4(3), 4(4), and 10(5) from PLS (Table 1) were not within the first 30 significant markers determined by individual FR, they were included within the following ten significant ones. On the other hand, markers, 6(4), 10(8), and 8(6) from FR (Table 1) were ranked 37th, 39th, and 63th, respectively, by the first PLS factor. The 30 genetic markers identified by PLS were considered the most important line covariables for explaining the GEI of biomass and were included in the PLS biplot.

The PLS biplot with the X scores for the lines, the Y loadings for the sites, and enriched with the X weights for the 30 genetic marker covariables identified above, is depicted in Fig. 1. The first PLS factor contrasted the low-altitude site, Poza Rica (PR), in winter, and the mid-altitude site, Tlaltizapan (TL), in summer, with the high-altitude sites El Batan (BA) and Toluca (TO). The second PLS factor discriminated between the lowland winter site (PR) versus the mid-altitude summer site (TL), and separated the two highland sites (BA and TO). A line having a positive (orthogonal) projection on a site vector (i.e., the line point and environmental vector are in the same quadrant) has positive interaction at that site, whereas a line located in the opposite direction (opposite quadrant) has a negative interaction with that same site. For example, some lines with a positive projection on the TO vector were 178, 2, 189, 124, 141, 42, 61, 67, 139, 106, 19, 148, 136, 144, 157, 121, 57, 127, 194, 163, 17, 156, and 174. Note that the lines are ordered by decreasing magnitude on the basis of their projections on the TO vector, such that line 178 has the largest GEI followed by line 2, and so forth. These lines interacted positively with TO; most of them have positive interaction (residual) values for biomass with TO (mean GEI value of 138.545 g m^2), but negative interaction values with TL, opposite quadrant (mean GEI value of –106.254 g m^2) (Table 2). On the other hand, some lines located on the opposite side (176, 8, 187, 130, 4, 110, 50, 33, 54, 149, 166, 180, 56, 35, and 30) have negative biomass interaction (residual) values with TO (mean GEI value of –106.006 g m^2), but positive interaction values with TL (mean GEI value of 122.018 g m^2) (Table 2). Lines close to the origin of the PLS biplot (0,0) have small interaction values with sites, and thus were not considered.

When the PLS biplot is enriched with the X weights of the 30 most-important genetic markers, it is interesting to observe the association between the subset of

Table 2 Number (or expected number) of alleles from the highland parent of genetic markers for lines with a positive GEI of biomass with Toluca and Tlaltizapan, and residuals (g m^{-2}) (interaction) for biomass

Line	Genetic marker						Tlaltizapan residual	Toluca residual
	2(1)[a]	2(2)[a]	6(5)[a]	6(6)[a]	10(7)[a]	10(6)[a]		
Lines with a positive GEI of biomass in Toluca								
178	1	1.1	2	2	2	2	–194.307	83.535
2	2	1.8	1	1	2	2	–94.682	216.160
189	1	1.1	2	2	2	2	–94.682	280.160
124	1	1.1	2	2	2	2	–180.057	21.285
141	1	1.1	2	2	1	1	–143.057	184.285
42	1	1.1	1	1	1.8	1	4.693	39.035
61	0	1.1	2	2	1.8	1	–96.057	184.285
67	2	1.4	1	1	0.4	2	–275.682	180.660
139	1	1.6	1	1	1	1	–31.682	17.660
106	1	1.1	2	2	1	1	–64.557	146.285
19	1	1.1	1	1	1.1	1	–74.182	112.660
148	2	1.4	0	0	1	2	–93.057	240.285
136	1	1.1	1	1	1	1	–116.307	315.535
144	2	1.4	1	1	1	1	–340.432	31.410
157	1	1.6	2	2	1	1	–185.807	22.035
121	2	1.4	2	2	2	2	–225.932	268.410
57	2	1.2	1	1	0.2	1	56.818	–16.340
127	1	0.0	2	2	2	1.2	–217.682	145.660
194	2	1.8	2	2	0.8	1	20.318	89.660
163	0	0.0	2	2	1	1	–3.307	163.535
17	1	1.1	1	1	1.2	1.9	–101.432	208.410
156	1	1.1	1	1	1	1	201.068	–18.090
174	2	1.4	1	1	2	1	–264.807	252.035
Mean	1.26	1.17	1.43	1.43	1.31	1.35	–109.339	137.763
Lines with a positive GEI of biomass in Tlaltizapan								
176	0	0	0	0	0	0	229.318	–103.340
8	0	0	2	2	0	0	–113.557	–3.215
187	2	1.4	1	1	0	0	76.943	–89.215
130	1	1.1	2	2	0	0	148.693	–79.965
4	1	0	1	1	1.8	1.1	112.568	–41.590
110	1	1.6	1	1	0.2	1	65.318	–2.840
50	1	0	0	0	0.8	0.1	74.693	–89.465
33	0	0	1	1	1	1	99.568	–160.590
54	1	1.1	0	0	0	0	221.818	–145.840
149	1	1.6	0	0	0	0	–148.432	–258.590
166	0	0	0	0	0.8	0	212.443	–151.215
180	1	1.1	1	1	0	0	403.443	–65.715
56	0	1.1	1	1	0.8	0	24.068	–106.090
35	1	1.1	0	0	1.2	1.9	278.568	–63.090
30	1	1.1	0	0	0	0	144.818	–229.340
Mean	0.73	0.74	0.66	0.66	0.44	0.34	122.018	–106.006

[a] Markers associated with lines having a positive interaction in Toluca

markers with the subset of lines that showed positive (or negative) interaction with certain environments, and to examine the allele substitution of these lines for those markers in terms of homozygous highland (coded value=2), heterozygous (coded value=1), and homozygous lowland (coded value=0). It could be hypothetized that the different adaptation of specific lines to the contrasting environments, (PR and TL) vs (TO and BA) or (PR) vs (BA) and (TO) vs (TL) could be associated, at least in part, to a different allele substitution of the markers associated with those subsets of lines. Furthermore, if these markers are associated with QTLs, it would be expected that the different allele substitution of the marker should be associated with a different allele substitution of the linked QTL.

In general, the PLS biplot showed that the 30 most-relevant RFLP markers are separated into two major groups; one group having 27 markers with high positive loadings for the first PLS factor and associated with some lines with a positive interaction with the two highland sites, TO and BA. All markers (and lines) that are located between the acute angle of two site vectors have a positive interaction with both sites, whereas all markers (and lines) between the negatively extended vectors (on the other side of the origin) have a negative interaction with both environments. The area between the two represents a situation in which the markers (and lines) have a positive interaction with one environment and a negative interaction with another environment. As previously mentioned, the perpendicular projection of a genetic marker on a site approximates the regression coefficient of the site on the allele state of that marker. For example, markers 6(5), 6(6) and 10(7) have a positive interaction with TO, but are negative with respect to BA. Markers 1–3 of chromosome 9, with negative loadings for the first PLS factor, are associated with lines having a positive interaction with TL and PR.

More specific associations between lines, subsets of linked markers, and sites can be observed. Markers 7–10 of chromosome 1, markers 3–4 of chromosome 4, and marker 8 of chromosome 5 tended to be more associated with lines having a positive biomass interaction with BA; markers 5 and 6 of chromosome 6, markers 6, 7, and 9 of chromosome 10, and markers 1 and 2 of chromosome 2 tended to be more associated with lines having a positive biomass interaction with TO; whereas markers 1–3 of chromosome 9 are associated with lines having a positive interaction with PR. Furthermore, linked marker subsets [1(3), 1(4), 1(5)], [3(4), 3(5), 3(6), 3(7)], [5(3), 5(4), 5(5), 5(6)] and [10(5) to 10(9)] tended to be located around the limit between the upper and lower right quadrants of the PLS biplot and are associated with lines having a positive interaction with sites TO and BA. Thus, the PLS biplot seems to identify clusters of linked (correlated) markers that are also associated with sets of specific lines having adaptation to particular environments.

It is expected that marker subset [2(1), 2(2), 6(5), 6(6), 10(6), and 10(7)] (with the highest projections on the TO vector) associated with lines having a positive interaction with highland site TO, will tend to have more alleles from the highland parent (coded values=2 and 1) than from the lowland parent (coded value=0). Lines located on the opposite quadrant of the PLS biplot, meanwhile, have a positive biomass interaction with site TL (and negative with TO) and will tend to have more alleles from the lowland parent than from the highland parent for that subset of markers. As shown in Table 2, lines with a positive interaction with TO had more alleles (or expected number of alleles) from the highland parent associated with markers 2(1), 2(2), 6(5), 6(6), 10(6), and 10(7) (mean allele code ranged from 1.17 to 1.43) than those lines with a positive biomass interaction with TL (mean allele code ranged from 0.34 to 0.74). No highland marker alleles with important contributions to explaining GEI were associated with lines with positive interaction effects in TL.

Similarly, lines having a positive interaction with the highland site BA (as opposed to lowland site PR) had, on average, more highland alleles than lowland alleles for marker subset [1(7), 1(8), 1(9), 1(10), 4(3), 4(4), and 5(8)]. Conversely, lines with a negative interaction with BA (positive interaction with PR, opposite quadrant) had more alleles from the lowland parent for that marker subset (data not shown).

The PLS method attempts to reduce the variability of a large and complex multi-dimensional data set to a few dimensions. In this example, there were 86 PLS factors, but only two were found significant by Osten's test; they explained 27% of the complex GEI for biomass. It is to be expected that some distortions will occur when representing lines, sites, and genetic markers in a two-dimensional diagram such as the PLS biplot. For example, site PR in Fig. 1 had a high loading for the third PLS factor, and lines such as 13, 131, 175 and 185, with low scores for the third PLS factor, are not highly associated with PR. Thus, they could not be considered as lines showing a positive GEI with PR.

From the 30 most useful markers for explaining the GEI of biomass, 27 had a high positive loading for the first PLS factor and are associated with lines having a positive interaction in highland sites TO and BA. The majority of these markers in lines with a positive biomass GEI in TO and BA had one or two alleles from the highland parents. Only three linked markers (1, 2 and 3 in chromosome 9) are associated with lines having a positive interaction with mid-altitude site TL and lowland site PR (a negative GEI with TO and BA). However, markers in lines with a positive GEI of biomass in PR had more alleles from the highland parents than markers in lines with a positive GEI in BA. This result would indicate that some lines with specific adaptation to lowland sites contain alleles from the highland parent, such as those in markers 1–3 from chromosome 9, that provide adaptation to lowland environmental conditions. This result is not surprising and could also have occurred for some markers associated with lines having a positive GEI with the lowland site TL (as opposed to TO); however, no relevant markers were associated with lines adapted to TL environmental conditions (Fig. 1). This small effect of the highland alleles in lowland environments agreed with results found by Jiang et al. (1999).

Markers 2(1), 5(3), 2(2), 5(5), 10(6), 3(5), 1(8), 9(1), 1(4), and 5(4) had the ten highest absolute loading values for the first PLS factor (Table 1). These markers, except 9(1), are between the acute angle formed by TO and BA site-vectors and are associated with lines having a positive interaction with BA and TO such as 148, 67, 178, 2, 139, 144 and 19, etc. These lines had, for those markers, more alleles from the highland parent than those lines located on the opposite side of the PLS biplot, which showed a negative interaction with BA and TO, but a positive interaction to PR and TL. Lines more adapted to PR and TL were, on average, more homozygous and heterozygous for lowland alleles in those markers than lines more adapted to the highland sites TO and BA (data not shown). This result would indicate that lowland alleles tend to show a broader adaptation to the lowland site PR and to the mid-altitude site TL than highland alleles (adapted only to highland sites TO and BA), possibly because PR winter and TL summer have somewhat similar temperatures, though the radiation is much lower in PR winter at all growth stages.

The composite interval mapping method used by Jiang et al. (1999) to map multiple QTLs and to assess their interaction with environments found six QTLs with significant effects and one QTL with a marginally significant effect for biomass. The authors found some significant QEI. The main QTLs for biomass were located in chromosomes 2, 3, 5, 6, 7, 8, and 10. Significant QTLs for biomass were located around markers 2(1)-2(2), 5(3)-5(4), 6(5)-6(7), and 10(6)-10(9) with significant QEI. These results agreed with our findings concerning the most relevant genetic markers found by the PLS and FR models for explaining the GEI of biomass. The PLS and

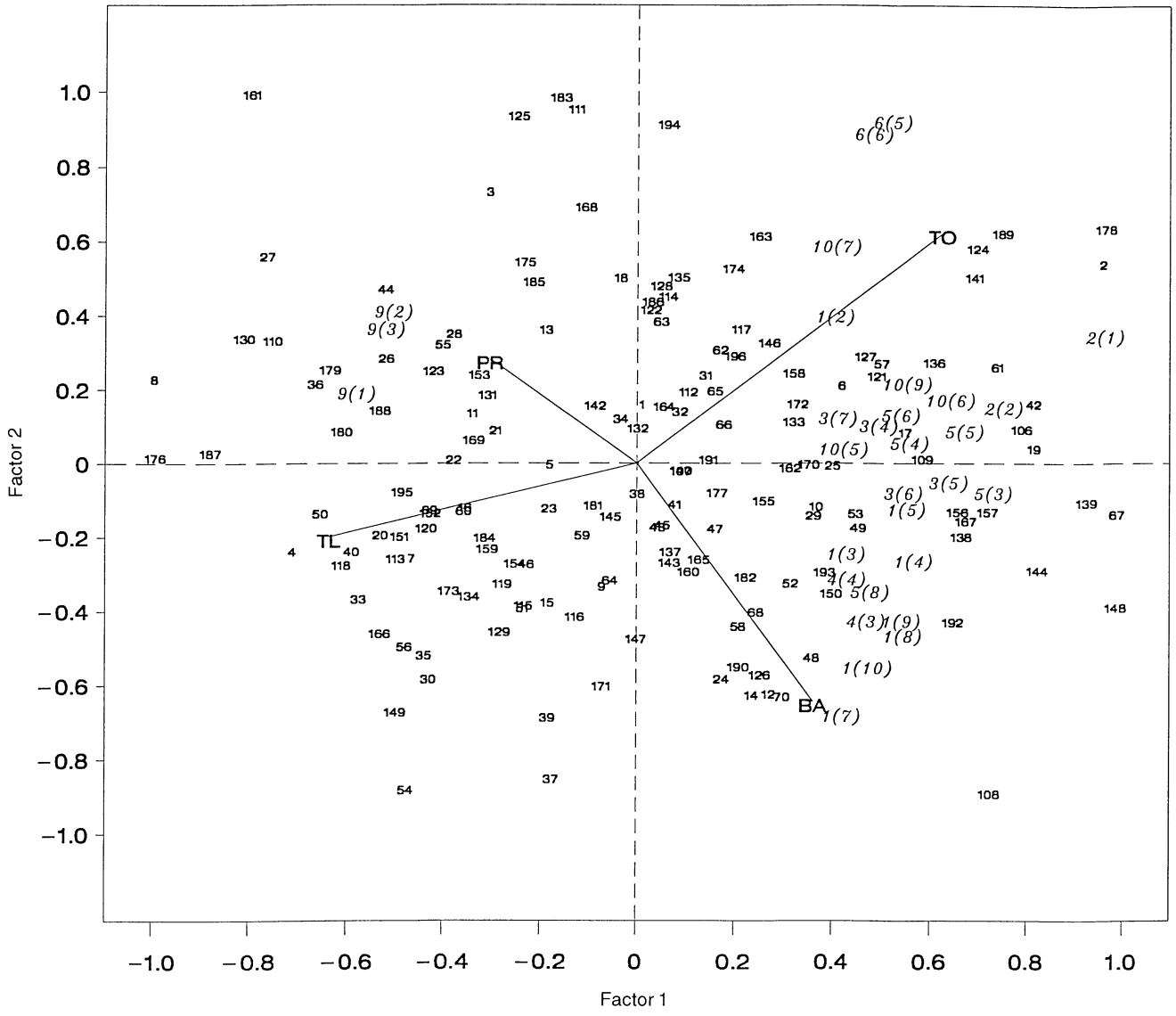

Fig. 1 Biplot of the first and second PLS factors representing the Y loadings of four sites (*PR*=Poza Rica; *TL*=Tlaltizapan; *BA*=El Batan; *TO*=Toluca); the X scores of 161 tropical maize lines (1–161) enriched with the X weights of 30 genetic markers

FR models found markers 2(1)-2(2), 5(3)-5(5), 6(5)-6(6), and 10(6)-10(7) to be important for explaining the GEI of biomass. These markers were linked to QTLs that had significant QEI. On the other hand, Jiang et al. (1999) found significant, or marginally significant, effects of QTLs located around markers 3(9)-3(10), 7(2)-7(3), and 8(4)-8(5) with no significant QEI. These results seem to agree with results from the PLS and FR models in which no markers from chromosomes 7 and 8, and only markers 4 to 7 of chromosome 3, had an important part in explaining the GEI of biomass. Jiang et al. (1999) reported several groups of QTLs, the most relevant being located at the end region of chromosome 10, showing large effects on biomass and grain yield favoring lowland alleles. This also agreed with the PLS and FR models that found markers 6, 7, and 9 of chromosome 10 are important genetic covariables for explaining the GEI of biomass in favor of lowland alleles for lines more adapted to TL and PR (right side of Fig. 1).

It is expected that lines with a positive interaction with a highland or lowland site should have more (or less) alleles from the highland parent (additive type of effect), which are responsible for that specific adaptation. For example, for biomass, significant QTLs in chromosome 2 for TO and TL were reported by Jiang et al. (1999). The PLS method identified markers 2(1) and 2(2) as important covariables for explaining GEI. The PLS biplot (Fig. 1) also suggests that markers in lines having a positive interaction with highland sites TO and BA tended to have more alleles from the highland parents than markers in lines more adapted to TL and PR. Conversely, Jiang et al. (1999) found that highland alleles had little detectable effect on biomass in lowland sites. This agrees with the findings of the present study where only three markers out of 30, 1–3 from chromosome 9, contain more alleles from the highland parent in those lines with specific adaptation to lowland sites.

Explaining line×site interaction using partial least squares and individual factorial regression with environmental covariables as explanatory variables

Characteristics of the four sites with respect to maximum and minimum temperatures and sun hours during vegetative, flowering, and grain-filling stages are given in Table 3. Clearly, the mid-altitude site (TL) and lowland site (PR) had higher maximum and minimum temperatures during the three growing stages, followed by the highland site BA. Toluca temperatures are at the margin for maize growth. These four sites showed similar sun hours during the entire growth cycle, except for TL with maximum sun hours during the flowering and grain-filling stages, and PR with minimum sun hours during all stages.

The cross-validation assesment and Osten's *F*-test for the number of significant PLS factors indicated that only the first PLS factor (out of four) was significant for prediction and explained 50% of the variability in the GEI matrix. The second factor was not significant and explained 24% of the GEI. MTV, MTF, and MTG had the highest loading values, followed by mTV, mTF, and mTG (Table 4). Results from the individual FR models gave the same results as PLS and ranked the environmental variables exactly the same as PLS (Table 4). Each individual FR uses 160 *df*.

The PLS biplot with the Z scores for sites and the Y loadings for the lines, and enriched with the Z weights for the nine environmental covariables, is depicted in Fig. 2. The first PLS factor contrasted Poza Rica and Tlaltizapan with El Batan and Toluca. Three sub-groups of environmental covariables are depicted in Fig. 2: (mTV, mTG, mTF), (MTG, MTV, MTF), and (SHG, SHF, SHV). The most sun hours during the three growing stages occurred in Tlaltizapan and the least in Poza Rica. As expected, these results showed that the maximum temperature during the entire growth stage is the main environmental factor causing the GEI of biomass.

Figure 2 showed that, in general, the distribution of lines across the upper and lower half of the PLS biplot followed a similar pattern to that observed in Fig. 1. However, distortions occurred because some lines and/or sites have large loadings for the third PLS factor. This is clear for TO, which had a high score for the third PLS factor, but appeared close to BA in Fig. 2.

Table 3 Mean daily maximum temperature (°C) during the vegetative stage (MTV), flowring stage (MTF) and grain-filling stage (MTG); mean daily minimum temperature (°C) during the vegetative stage (mTV), flowering stage (mTF) and grain-filling stage (mTG); sun hours per day during the vegetative stage (SHV), flowering stage (SHF), and grain-filling stage (SHG), for Toluca (TO), El Batan (BA), Tlaltizapan (TL), and Poza Rica (PR)

Variable	Site			
	TO	BA	TL	PR
MTV	22.78	25.38	32.02	26.94
MTF	20.81	24.40	32.94	25.38
MTG	20.73	24.03	30.48	27.87
mTV	3.22	9.79	16.67	17.29
mTF	7.06	8.58	17.64	15.62
mTG	3.74	7.65	17.50	16.67
SHV	6.33	7.12	6.64	4.58
SHF	5.54	5.13	9.60	3.66
SHG	5.54	6.10	7.96	5.33
Elevation (m)	2650	2240	940	60
Mean biomass (g m^{-2})	868.5	1013.9	868.1	408.2

Explaining line×site interaction using multiple factorial regression with genetic markers, environmental explanatory covariables, and their cross products

A strategy for selecting relevant sets of line and environmental covariables, proposed by Vargas et al. (1999), is to use multiple FR coupled to a stepwise selection procedure for multiple FR models. The authors found that PLS was effective in grouping correlated covariables and that the multiple FR with a stepwise procedure selected representative covariables from each of the sub-groups depicted in the PLS biplot. For example, if there were four subsets of covariables roughly delineated by each of the

Table 4 Z loadings of the first and second PLS factors for nine environmental covariables and partition of the total site×line interaction variance in nine individual factorial regressions (FRs)

Covariable	% Variance of site×line interaction explained by						
	Partial least squares		Factorial regression				
	PLS 1	PLS 2	Source	*df*	Mean square ×10^5	*F*	Prob >*F*
			Site×line	480	0.311	1.45	0.0001
MTV[a]	0.3938	–0.0682	MTV	160	0.511	2.39	0.0000
MTF	0.3902	–0.1297	MTF	160	0.504	2.35	0.0000
MTG	0.3836	0.1261	MTG	160	0.495	2.31	0.0000
mTF	0.3622	0.2159	mTF	160	0.470	2.19	0.0000
mTG	0.3540	0.2638	mTG	160	0.452	2.11	0.0000
mTV	0.3411	0.2795	mTV	160	0.429	2.01	0.0000
SHG	0.3207	–0.3976	SHG	160	0.398	1.86	0.0000
SHF	0.2641	–0.4393	SHF	160	0.337	1.57	0.0009
SHV	–0.0089	–0.6459	SHV	160	0.197	0.92	0.7233
			Pooled error	536	0.213		

[a] MT: maximum temperature; mT: minimum temperature; SH: sun hours per day; V: vegetative stage; F: flowering stage; G: grain-filling stage

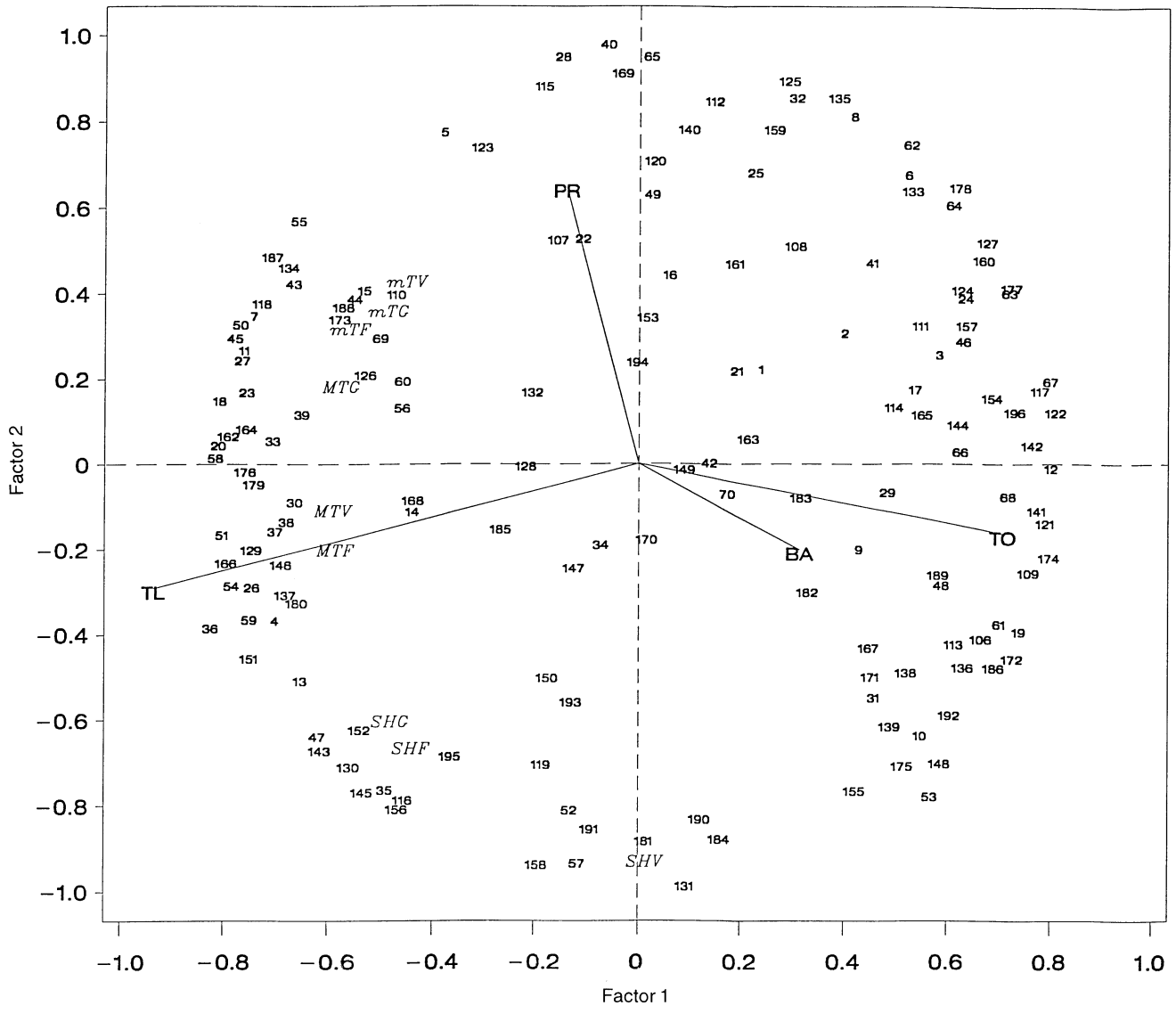

Fig. 2 Biplot of the first and second PLS factors representing the Z scores of four sites (*PR*=Poza Rica; *TL*=Tlaltizapan; *BA*=El Batan; *TO*=Toluca), the Y loadings of 161 tropical maize lines (1–161) enriched with the Z weights of nine environmental variables. Those variables are: mean daily maximum temperature during the vegetative stage (*MTV*); flowering stage (*MTF*), grain-filling stage (*MTG*); mean daily minimum temperature during the vegetative stage (*mTV*); flowering stage (*mTF*), grain-filling stage (*mTG*); sun hours per day during the vegetative stage (*SHV*); flowering stage (*SHF*), grain-filling stage (*SHG*)

four quadrants of the PLS biplot, the multiple FR with the stepwise procedure selects representative covariables from each subset. After identifying the most important subset of line and environmental covariables and their relevant representatives, Vargas et al. (1999) proposed fitting a multiple FR model with a stepwise procedure, including either the cross products among all the covariables or the cross product among the most-relevant covariables.

In the present study we started with the complete site × line term. Then the lines were replaced by the 30 markers (determined by PLS) (e.g., site×marker), the sites were substituted by the nine environmental covariables (e.g., environmental covariable × line), and finally the 30×9=270 cross-products terms were added. These were considered the three basic analyses, and in each case a multiple FR with a stepwise procedure for identifying the significant terms was used. In addition, other multiple FR models, including all 86 markers or other specific subsets of markers (e.g., marker blocks per chromosome), as well as specific subsets of environmental covariables and their corresponding cross products, were fitted and their results were discussed and compared with the three basic analyses.

Results of a stepwise variable search for a multiple FR, including the 30 terms of the type site×marker, led to a model with ten significant terms (Table 5) listed in the same order as in the model. The model explained 27% of the GEI with 30 *df* and left a non-significant deviation. Note that, as expected from results obtained by Vargas et al. (1999), the stepwise procedure selected markers from the subsets of linked markers suggested by

Table 5 Analysis of variance for stepwise multiple factorial regression models with line covariables, environmental covariables, and their cross products

Source	df	Sum of squares $\times 10^5$	Mean squares $\times 10^5$	*F*	Prob >*F*
Genetic marker covariables					
Site×line	480	149.271	0.311	1.46	0.0001
Site×2(1)	3	9.996	3.332	15.86	0.0000
Site×6(5)	3	7.013	2.338	11.13	0.0000
Site×5(3)	3	5.047	1.682	8.01	0.0000
Site×1(7)	3	4.515	1.505	7.16	0.0001
Site×10(6)	3	3.751	1.250	5.95	0.0005
Site×1(4)	3	2.421	0.807	3.84	0.0096
Site×5(8)	3	2.335	0.778	3.70	0.0116
Site×4(3)	3	2.007	0.669	3.18	0.0235
Site×9(1)	3	1.741	0.580	2.76	0.0414
Site×9(2)	3	1.803	0.600	2.86	0.0363
Deviation	450	108.641	0.241	1.14	0.0730
Pooled error	536	114.628	0.213		
Environmental covariables					
Site×line	480	149.271	0.311	1.46	0.0001
Line×MTV	160	81.800	5.113	2.39	0.0001
Deviation	320	67.470	0.210	0.99	0.5366
Pooled error	536	114.628	0.213		
Cross products					
Site×Line	480	149.271	0.311	1.46	0.0001
MTG×2(1)	1	9.890	9.890	47.09	0.0000
MTF×6(5)	1	5.222	5.222	24.86	0.0000
mTG×5(3)	1	5.002	5.002	23.81	0.0000
MTF×3(7)	1	5.109	5.109	24.32	0.0000
MTG×10(6)	1	3.546	3.546	16.88	0.0000
mTF×1(4)	1	2.017	2.017	9.60	0.0020
mTF×5(8)	1	1.305	1.305	6.21	0.0129
mTF×1(7)	1	1.280	1.280	6.09	0.0138
mTV×1(7)	1	1.287	1.287	6.12	0.0136
MTF×10(9)	1	0.996	0.996	4.74	0.0298
mTF×4(3)	1	1.248	1.248	5.94	0.0151
MTV×9(1)	1	0.971	0.971	4.62	0.0319
mTV×5(8)	1	0.834	0.834	3.97	0.0467
Deviation	467	110.650	0.236	1.12	0.1022
Pooled error	536	114.628	0.213		

the PLS biplot (Fig. 1). Furthermore, the terms site×2(1), site×6(5), site×5(3), site×1(7), and site×10(6) were the terms first selected and explained 20.3% of the GEI with 15 *df* (Table 5). In addition, multiple FR (using the stepwise procedure with blocks of markers included per chromosome) were performed for detecting the significant site×chromosome terms (data not shown). Then, all the significant site×marker terms from each chromosome were collected and simultaneously fitted in a multiple FR (analyses not shown). Using this procedure, 21 significant site×marker terms were identified and then introduced as the independent variable set in a stepwise variable search. In the final model, 13 site×marker terms were included. The first four terms included in the search were (in order of appearance) site×2(1), site×6(5), site×5(3), and site×10(6); they accounted for 17.4% of the GEI. These results confirmed those previously found by individual FR, by multiple FR using 30 markers, and the findings of Jiang et al. (1999).

Concerning the nine environmental covariables included in the multiple FR with a stepwise procedure, only the term line × MTV was significant; it explained 54.8% of the GEI of biomass with 160 *df* (Table 5) leaving a non-significant deviation. This result agreed with that of PLS and FR which found that maximum temperature during the vegetative stage (MTV) was the most important environmental covariable for explaining the GEI of biomass (Table 4).

The multiple FR with a stepwise procedure using 270 cross products from 30 genetic markers with nine environmental variables resulted in a model with 13 cross products explaining 26% of the GEI (Table 5, terms are listed in order of inclusion) and left a non-significant deviation sum of squares. Except for 9(2), all the markers found to be significant when fitting the model site×marker with 30 marker covariables, showed significant cross products with various environmental covariables. On the other hand, several environmental variables that were non-significant when fitting FR model line×environmental covariable with nine environmental covariables, showed significant cross products with some molecular markers. Although MTV seems to be the most important environmental covariable from the perspective of plant physiology, other maximum temperatures seem to have some importance on the cross products simply because they are highly correlated with MTV. It is interesting that cross products involving markers 2(1), 6(5), 5(3), 3(7), and 10(6) [MTG×2(1), MTF×6(5), mTG×5(3),

MTF×3(7), and MTG×10(6)] are the most important of the 13 significant cross products (Table 5) and explained 19.3% of the GEI of biomass with 5 *df*.

When the multiple FR model with a stepwise procedure, including the cross products among the genetic markers and the environmental covariables, is fitted to the data, the GEI is explained only by the term **Xν Z'** of equation 3 (see Materials and methods). Vargas et al. (1999) described the interpretation of the GEI interaction in terms of the FR cross products. The G×H matrix **ν** contains the regression coefficients (which can be positive or negative) of the cross products of line and environmental variables. The I×G matrix **X** and the H×J matrix **Z'** represent the line and environmental covariables, respectively, that have been previously centered to a mean of zero and variance equal to 1. Therefore, positive value of the i^{th} line (i^{th} row of **X**) for its g^{th} covariable (g^{th} column of **X**) indicates above-average performance, and negative value indicates below-average performance. Similarly, a positive value of the j^{th} site (j^{th} column of **Z**) for its h^{th} covariable (h^{th} row of **Z**) indicates an above-average performance, and negative value indicates a below-average performance. Thus, depending on the combination of signs of these three components, the predicted interaction term **Xν Z'** can be positive or negative.

The estimated regression coefficient of the 13 cross products were: MTG × 2(1) = –10.20; MTF × 6(5) = –6.50; mTG × 5(3) = –5.51; MTF × 3(7) = –4.63; MTG × 10(6) = –4.84; mTF × 1(4) = –4.31; mTF ×5(8) = –10.25; mTF × 1(7) = –13.40; mTV × 1(7) = 7.97; MTF × 10(9) = –3.81; mTF × 4(3) = –3.63; MTV × 9(1) = 3.71; and mTV × 5(8) = 6.07. In general, these estimates showed that highland alleles provide adaptation to highland environments and that lowland alleles confer adaptation to lowland environments. For example, lines with an above-average number of alleles of the highland parent for marker 2(1) do not perform well in sites TL or PR, which have an above-average MTG {[positive sign for 2(1) in **X**] ×[negative sign for the regression coefficient of MTG×2(1)]×[positive sign for MTG in TL or PR in matrix **Z'**]}. On the other hand, lines with a below-average number of alleles of the highland parent for marker 2(1) perform relatively well in site TL or PR, with an above-average MTG {[negative sign for 2(1) in **X**]×[negative sign for the regression coefficient of MTV×2(1)]×[positive sign for MTV in TL or PR in matrix **Z'**]}. Furthermore, lines with an above-average number of alleles of the highland parent from marker 5(3) produce relatively little biomass in TL (or PR) when higher minimum temperatures occur during the grain-filling stage (mTG).

Two additional multiple FR analyses were performed to compare results with those previously described. The first model considered a reduced number of cross products (12 cross products) resulting from genetic markers 2(1), 6(5), 5(3) and 10(6) with environmental covariables MTV, mTF and SHG (data not shown). These covariables were chosen based on the set of markers and environmental covariables suggested by the previous analyses and by the PLS biplots of Figs. 1 and 2. Results indicated that cross-products MTV×2(1), MTV×6(5), MTV×10(6), and mTF×5(3) were the only significant cross-products that explained 15.2% of the GEI of the biomass. The estimated regression coefficients of the four cross-products were all negative. These results clearly agreed with those of Jiang et al. (1999) and confirmed those previously found by PLS and individual and multiple-FR. The second multiple FR model comprises 602 cross products obtained from all 86 markers with seven environmental variables (MTV, MTF, MTG, mTV, mTF, mTG and SHG) (data not shown). Nineteen cross products, with 1 *df* each, were found significant and explained 33% of the GEI. From these, the six most important cross products were the same as the six most important cross-products of Table 5, which employed markers 2(1), 5(3), 6(5) and 10(6). This result clearly agreed with those obtained previously by PLS, individual and multiple FR, and those reported by Jiang et al. (1999).

Explaining line×site interaction by fitting individual markers to individual sites using multiple factorial regression

The model previously used to study the site×marker interaction fitted four parameters for each included marker, with a marker effect for each environment (3 *df*). There was no way to separate out the necessity of individual marker×environment terms. A model based on individual marker×environment terms is useful for determining the intensity of expression of individual markers (QTLs) in individual environments. Is QEI attributable to QTLs expressing themselves in a binary way (yes/no) in individual environments, so that different QTLs are expressed in different environments? Or, is QEI due to differences in the intensity of expression of specific QTLs over different environments? Or, is it a combination of both?

With 86 markers and four sites, there are 344 possible marker-site combinations. The regressors corresponding to these combinations were introduced as the set of independent variables in a multiple FR with a stepwise procedure. Table 6 showed the 32 significant marker-site combinations included in the final model arranged by chromosome (the order of inclusion of the terms in the model is shown in the 6^{th} column). Markers 2(1), 6(4), and 5(3) were significant in two sites (but with differing intensities); the rest were significant in only one site. Of the 29 significant markers, 17 of them are the same as the 30 markers determined by individual FR, and 15 are the same as the 30 markers found by the PLS method. The first four markers to be included were 2(1), 6(5), 5(5) and 10(6), with an intensity of expressions in the highland site TO of 61.4, 38.7, 29.9 and 33.2 g m^{-2} of biomass added per lowland allele subtituted by a highland allele, respectively. Markers 4(3) and 5(3) were in-

Table 6 Analysis of variance for stepwise multiple factorial regressions models when fitting individual genetic markers to individual sites

Source	*df*	Sum of squares $\times 10^5$	*F*	Prob >*F*	Order of inclusion	Regression coefficient
Site×line	480	149.271	1.46	0.0001		
1(2)-TL	1	0.923	4.32	0.0381	22	–33.4
1(5)-TL	1	1.918	8.97	0.0028	7	–35.7
1(7)-BA	1	2.652	12.40	0.0004	6	50.0
1(9)-PR	1	1.079	5.04	0.0251	30	–33.7
2(1)-TO	1	6.504	30.41	0.000	1	61.4
2(1)-TL	1	1.670	7.81	0.0053	11	–48.1
2(4)-TO	1	1.133	5.30	0.0217	28	37.0
3(7)-TO	1	0.969	4.53	0.0337	20	47.6
3(9)-TL	1	1.763	8.24	0.0042	12	–46.7
4(1)-BA	1	0.880	4.12	0.0428	24	–32.6
4(3)-BA	1	1.873	8.76	0.0032	8	132.4
4(5)-BA	1	1.915	8.95	0.0029	9	–78.5
4(8)-TO	1	0.898	4.20	0.0409	25	36.0
4(10)-TL	1	0.946	4.43	0.0357	23	30.4
5(2)-TO	1	1.377	6.44	0.0114	27	–68.7
5(3)-TO	1	0.866	4.05	0.0446	26	90.6
5(3)-BA	1	1.367	6.39	0.0117	14	55.4
5(5)-TO	1	3.539	16.55	0.0000	3	29.9
5(8)-TL	1	0.953	4.46	0.0351	21	–43.6
6(2)-TO	1	1.796	8.40	0.0039	13	39.6
6(4)-PR	1	1.825	8.53	0.0036	10	55.3
6(4)-TO	1	0.838	3.92	0.0482	32	67.4
6(5)-TO	1	5.393	25.22	0.0000	2	38.7
8(6)-BA	1	1.247	5.83	0.0160	16	–36.5
9(1)-TL	1	2.594	12.13	0.0005	5	30.7
9(2)-BA	1	1.116	5.22	0.0227	18	–111.0
9(4)-BA	1	1.007	4.71	0.0304	31	76.7
9(5)-TO	1	0.948	4.44	0.0355	29	43.9
10(3)-TL	1	1.299	6.07	0.0140	15	–61.0
10(6)-TO	1	3.227	15.09	0.0001	4	33.2
10(8)-BA	1	1.145	5.35	0.0211	19	–41.3
10(9)-TL	1	1.159	5.42	0.0202	17	–48.1
Deviation	448	92.447	0.97	0.6305		
Pooled error	536	114.628	0.213			

tensively expressed in BA and TO, respectively, when lowland alleles were substituted by highland alleles (132.4 g m^{-2} and 90.6 g m^{-2} of biomass per higland allele). On the other hand, allele substitution towards the highland parent on marker 9(2) in BA is expressed by a decrease in biomass of –111 g m^{-2} per highland allele.

Wright and Mowers (1944) and Whittaker et al. (1997) considered the multiple regression of phenotype trait value on marker type. The authors found that: 1) only marker flanking a QTL have non-zero regression coefficients of the same sign and 2) intervals with flanking markers with regression coefficients of opposite sign may be due to the presence of two QTLs with opposite sign or due to the presence of a ghost QTL or due to the presence of a QTL in adjoining intervals.

In general, results of this analysis indicate a decrease in biomass per lowland allele substituted by a highland allele in mid-altitude site TL and lowland site PR, but an increase in biomass production per lowland allele substituted by a highland allele is shown in highland sites TO and BA. However, this trend is expressed with different intensities in different markers. These results agree with the adaptation pattern of lines across sites and the association of genetic markers with lines and genetic markers with sites depicted in Fig. 1.

Conclusions

The results of this study show that PLS and FR could be useful tools for studying genetic differentiation associated with adaptation to specific environmental conditions. The FR model is a flexible and powerful technique that can be easily adapted to fit different partitions of the GEI using markers and environmental covariables disclosing GEI and QEI response patterns that exist on a large and complex tropical maize data set. PLS and FR identified 30 genetic marker covariables, most of them linked to QTLs with significant QEI which, in turn, account for a significant proportion of GEI for biomass. Based on the loadings of the first and second PLS factors, the PLS biplot showed major sets and subsets of linked markers as-

sociated which lines showing adaptation to specific environments. Furthermore, the majority of these markers were closely associated with QTLs detected by composite interval mapping and showed a significant QEI (Jiang et al. 1999). The PLS biplot helps to identify a set of lines with positive or negative interactions with specific environments and allows us to associate these sets of lines with particular sets of linked genetic markers. These markers showed, for specific lines, allele substitutions that favored the environment of their origin. Lowland alleles showed a bit broader adaptation than highland alleles, a finding that is consistent with the experience of CIMMYT breeders who found that, generally, highland germplasm is narrowly adapted (Eagles and Lothrop 1994).

Assuming that a particular allele composition of genetic markers is associated with a specific allele composition at the linked QTL, the different alleles of markers associated with lines having a positive or negative interaction with environments should be related to specific alleles at the linked QTL, which in turn results in specific genetic adaptation of those lines to those environments. Results showed a negative association between maximum temperatures and markers 2(1), 6(5), and 10(6) (negative regression coefficient for their cross products), indicating that high maximum temperatures, such as those occurring at TL and PR, negatively influenced the expression of QTLs associated with markers 2(1), 6(5), and 10(6) when these markers had a greater number of alleles from the highland parent. However, higher maximum temperatures positively influenced the expression of QTLs associated with those markers when they had a greater frequency of alleles from the lowland parent. High minimum temperature, the other environmental variable that has been identified as important, seems to positively affect QTLs associated with marker 5(3) when this marker has more alleles from the lowland parent, but has negative effects when it has more alleles from the highland parent. Thus, QTLs associated with markers 2(1), 6(5) and 10(6) are sensitive to maximum temperature, whereas QTLs associated with marker 5(3) are sensitive to minimum temperature. These results were in agreement with the QTL mapping study of Jiang et al. (1999) who found (in terms of the additive effects) a significant QTL: (1) on chromosome 2 [around genetic markers 2(1) and 2(2)] for biomass in TL in favor of the alleles from the lowland parent and in TO in favor of alleles from the highland parent, and (2) at the end region of chromosome 10 [around genetic markers 10(6) and 10(9)] for biomass in favor of lowland alleles.

Results from PLS and FR indicate the effect of genetic markers linked to important QTLs for adaptation to a site change with temperature. Additive effects of allelic substitution towards lowland alleles seem to provide a broader range of adaptation, within moderate ranges of temperature (lowland site PR and mid-altitude site TL), than those provided by allelic substitution in favor of highland alleles. Introducing lowland alleles into highland germplasm from chromosome regions such as those around markers 2(1)-2(2), 5(3)-5(4), 6(5)-6(6) and 10(6)-10(9) should make highland germplasm better adapted to a wider range of temperatures without losing other aspects of specific adaptation to highland environments. This strategy can be most efficiently carried out using a marker-assisted backcrossing breeding scheme.

In general, the results of this study do not contradict the findings of Ellis et al. (1992) and Lafitte et al. (1997) that adaptation differences between highland and lowland tropical maize germplasm are mainly due to differences in growth and development processes induced by temperature (i.e., cold temperatures in highland environments and warm temperatures in lowland environments). Adaptation to different temperatures in low-, mid- and high-altitude environments is (at least in part) the result of long-term natural and artificial selection with allelic substitution occurring at several loci. PLS and individual and multiple FR explained sizeable proportions of the GEI and led to meaningful biological interpretations.

References

Aastveit H, Martens H (1986) ANOVA interactions interpreted by partial least squares regression. Biometrics 42:829–844

Balfourier F, Oliveira JA, Charnet G, Arbones E (1997) Factorial regression analysis of genotype by environment interaction in ryegrass populations, using both isozyme and climatic data as covariables. Euphytica 98:37–46

Crossa J (1990) Statistical analyses of multilocation trials. Adv Agron 44:55–85

Denis J-B (1988) Two-way analysis using covariates. Statistics 19:123–132

Eagles HA, Lothrop JE (1994) Highland maize from Central Mexico – its origin, characteristics and use in breeding programs Crop Sci 34:11–19

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6:36–40

Ellis RH, Summerfield RJ, Edmeades GO, Roberts EH (1992) Photoperiod, temperature, and the interval from sowing to tassel initiation in diverse cultivars of maize. Crop Sci 32: 1225–1232

Eeuwijk FA van (1996) Between and beyond additivity and non-additivity; the statistical modelling of genotype by environment interaction in plant breeding. PhD thesis, Wageningen University, The Netherlands

Eeuwijk FA van, Denis J-B, Kang MS (1996) Incorporating additional information on gentoypes and environments in models for two-way genotype by environment tables. In: S Kang, Gauch HG (eds) Genotype by environment interaction. CRC Press, Boca Raton, Florida, pp 15–49

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant breeding programme. Aust J Agric Research 14:742–754

Gauch HG Jr (1988) Model selection and validation for yield trials with interaction. Biometrics 44:705–715

Gollob HF (1968) A statistical model which combines features of factor analysis and analysis of variance techniques. Psychometrika 33:73–115

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

Jiang C, Zeng Z-B (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetica 140:1111–1127

Jiang C, Zeng Z-B (1997) Mapping quantitative trait loci for dominant and missing markers in various crosses from two inbred lines. Genetics 101:47–58

Jiang C, Edmeades GO, Armstead I, Laffite HR, Hayward M, Hoisington D (1999) Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. Theor Appl Genet (in press)

Johnson EC, Fischer KS, Edmeades GO, Palmer AFE (1986) Recurrent selection for reduced plant height in lowland tropical maize. Crop Sci 26:253–260

Kempton RA (1984) The use of biplot in interpreting variety by environment interactions. J Agric Sci 103:123–135

Knapp SJ, Bridges W, Birked D (1990) Mapping quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79:583–592

Lafitte HR, Edmeades GO (1997) Temperature effects on radiation use and biomass partitioning in diverse tropical maize cultivars. Field Crops Res 49:231–247

Lafitte HR, Edmeades GO, Johnson EC (1997) Temperature response on tropical maize cultivars selected for broad adaptation. Field Crops Res 49:215–229

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

Mandel J (1971) A new analysis of variance model for non-additive data. Technometrics 13:1–18

Martinez O, Curnow RN (1992) Estimating the location and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–485

Osten DW (1988) Selection of optimal regression models via cross-validation. J Chemometrics 2:39–48

Romagosa I, Ullrich SE, Han F, Hates PM (1996) Use of additive main effects and multiplicative interaction model in QTLs mapping for adaptation of barley. Theor Appl Genet 93:30–37

Sari-Gorla M, Calinsky T, Kaczmarek Z, Krajewski P (1997) Detection of QTL×environment in maize by a least squares interval mapping method. Genetics 78:146–157

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J Roy Stat Soc, Ser B, 36:111–147

Talbot M, Wheelwright AV (1989) The analysis of genotype × environment interactions by partial least squares regression. Biuletyn Oceny Odmian, Zeszyt 21–22:19–25

Vargas M, Crossa J, Sayre K, Reynolds M, Ramírez ME, Talbot M (1998) Interpreting genotype×environment interaction in wheat using partial least squares regression. Crop Sci 38: 679–689

Vargas M, Crossa J, van Eeuwijk FA, Ramírez ME and Sayre K (1999) Using partial least squares, factorial regression and AMMI models for interpreting genotype×environment interaction. Crop Sci (in press)

Wright AJ, Mowers RP (1994) Multiple regression for molecular-marker, quantitative trait data from F_2 populations. Theor Appl Genet 89:305–312

Whittaker JC, Thompson R, Vissher PM (1997) On the mapping of QTL by regression of phenotype on marker-type. Heredity 77:23–32

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:556–580

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