CROP BREEDING, GENETICS & CYTOLOGY

A Multivariate Method for Classifying Cultivars and Studying Group × Environment × Trait Interaction

Jorge Franco, José Crossa,* Suketoshi Taba, and Henry Shands

ABSTRACT

Classification methods are used in genetic resource conservation and plant breeding. The two-stage Ward-Modified Location Model (Ward-MLM) clustering strategy defines initial groups by the Ward method and the MLM is then used to improve those groups. The three-way clustering strategy considers the same trait measured in different environments as different variables (environment–trait combination), so that the resulting clusters of cultivars have a small group × environment interaction (GEI) for all the traits included in the study. An important component of the GEI is the imperfect genotypic correlation or crossover interaction (COI). This study used the three-way Ward-MLM clustering strategy in three different data sets with the objectives of (i) evaluating the ability of the Ward-MLM methodology for clustering cultivars into groups with low COI, (ii) obtaining a graphical representation of the variables in a low dimensional scatter plot produced by the multidimensional scaling method (MDS) in which the GEI of continuous and categorical traits and environments can be visualized, and (iii) studying how the relationship between a pair of traits changes across environments. The three-way Ward-MLM strategy produced groups of cultivars with low levels of COI. The increment of the correlation coefficient values between groups with respect to the total correlation coefficients indicated that the groups formed by the three-way Ward-MLM strategy comprised subsets of individuals that had similar trait responses across environments.

I N GENETIC RESOURCES conservation and plant breeding research, classification methods are used for studying phenotypic and genetic diversity, formation of core subsets, and grouping cultivars into homogeneous clusters (Crossa et al., 1995; Franco et al., 1999). Continuous and categorical traits are measured in each of the gene bank accessions or cultivars evaluated in one or several environments, and hierarchical and statistical classification methods are employed to recover, as much as possible, the underlying subpopulations of individuals.

The method for clustering n individuals on the basis of p continuous and q categorical traits measured in r different environments is known as three-way (cultivar × environment × trait) cluster analysis. Basford and McLachlan (1985) proposed a classification method when all measured traits are continuous. The authors utilized the Gaussian Model (GM) that assumes that each observation with p traits measured in each environment is normally distributed. For r environments, the vector of each observation has dimensions $1 \times rp$, where $rp =$ trait–environment combinations (variables).

Lawrence and Krzanowski (1996) used the Location Model (LM), originally proposed by Olkin and Tate (1961), for classifying n individuals when p continuous traits and q categorical traits are measured in one environment. The LM combines the levels of all q categorical traits in one multinomial variable, W, with m levels ($w = 1,2,…,m$). Franco et al. (1998) modified the LM and proposed the Modified Location Model (MLM), by assuming that the m levels of the W variable and the p-multinormal variables for each subpopulation, are independent. The authors proposed using the MLM in a two-stage clustering strategy where initial groups are defined by the Ward (or any other hierarchical clustering strategy) method and the MLM is then used with those groups (Ward-MLM).

Franco et al. (1999) extended the Ward-MLM strategy to the case of clustering three-way data of cultivar × environment × trait. Thus, the vector of each observation for r environments is $1 \times (rp + 1)$. The three-way Ward-MLM clustering strategy considers the same trait measured in different environments as different variables (environment-trait combinations). Since the Ward-MLM strategy clusters individuals with consistent response across all variables in all the environments, it is reasonable to expect that the three-way cluster strategy should form groups of cultivars with negligible group × GEI for all the traits included in the study.

Group × environment interaction can be expressed as imperfect genotypic (or environmental) correlation or COI, or as heterogeneity of variance across environments or non-COI (Moll et al., 1978; Muir et al., 1992). The crossover GEI is the most important component of the GEI to be considered. Thus, the three-way Ward-MLM strategy should tend to form groups of cultivars with high genotypic correlations across environments

J. Franco, Facultad de Agronomía, Universidad de la República Oriental del Uruguay, Avd. Garzon 780 CP 12900. Montevideo, Uruguay; J. Crossa, Biometrics and Statistics Unit, CIMMYT, Apdo. Postal 6-641, 06600, México DF, México; S. Taba, Maize Genetic Resources Unit, CIMMYT, México; and H. Shands, National Seed Storage Laboratory, USDA, ARS, Fort Collins, CO 80523. Received 20 Sept. 2002. *Corresponding author (j.crossa@cgiar.org).

(low levels of COI) for most of the continuous and categorical traits. In addition, the three-way clustering method should form groups of cultivars with a consistent response with respect to the association of pairs of traits across environments.

Conceptually, the statistical models used for assessing, studying, and interpreting cultivar × environment interactions are univariate because they consider one trait at a time (Crossa, 1990). Multiplicative models for multienvironment cultivar trials that consider one trait at a time have been used for developing methods for clustering sites or cultivars into groups with statistically negligible COI (Cornelius et al., 1992, 1993; Crossa et al., 1993, 1996; Crossa and Cornelius, 1993, 1997; Abdalla et al., 1997). On the other hand, the three-way Ward-MLM clustering methodology is multivariate because it takes into consideration all the traits (continuous and categorical) simultaneously. Therefore, studying the multivariate GEI from the groups obtained by three-way Ward-MLM is complex. No formal studies have been conducted and published to demonstrate the ability of the three-way Ward-MLM clustering methodology to form clusters simultaneously with low levels of COI of the continuous and categorical traits. Furthermore, no studies have been published that assess the effect of GEI on the association between traits.

This study employed the three-way Ward-MLM clustering strategy in three different data sets with the objectives of (i) evaluating the ability of the Ward-MLM methodology for clustering cultivars into groups with negligible COI, (ii) obtaining a graphical representation of the variables (environment–trait combinations) in a low dimensional scatter plot produced by the multidimensional scaling method, where the COI and the non-COI of continuous and categorical traits and environments can be visualized, and (iii) studying the association between a pair of traits across environments, which is a multivariate assessment of GEI.

MATERIALS AND METHODS
Theory
Three-Way Modified Location Model (MLM)

Consider a random $n \times rp$ matrix of n observations, where, for each observation, p traits are measured in each of the r environments. This forms a matrix with n rows (observations) and rp columns. These rp environment–trait combinations will be named *variables*. By means of the two-stage Ward-MLM approach, it is possible to avoid the "independence across sites" assumption and therefore to estimate variance of the traits within sites and their covariances across sites, provided that n (if homogeneity for the variance–covariance matrix is assumed for each subpopulation), or n_i (if heterogeneity for the variance–covariance matrix is assumed for each subpopulation) is greater than $rp + 1$ (Mardia et al., 1979). The continuous and discrete traits are combined into an $rp + 1$ vector that contains the rp values of the continuous variables plus the values of the multinomial variable $W = s$, which combines the information from the categorical traits in all the environments.

Parameter estimation by maximum likelihood, the probability of membership for each observation in each subpopulation

to be used in the expectation maximization (EM) algorithm (Dempster et al., 1977), and other theoretical details of the three-way Ward-MLM are shown in Franco et al. (1999).

Two-Stage Ward-MLM Method

The two-stage Ward-MLM strategy was proposed by Franco et al. (1998), where the initial groups are generated by the Ward (1963) minimum variance within-groups hierarchical method. The number of groups are defined by the upper-tail approach, available in the software CLUSTAN (Wishart, 1987) combined with the likelihood profile associated with the likelihood-ratio test (Mardia et al., 1979). Then the MLM, using the Ward's groups as the starting (initial) point, is applied with the objective of improving the classification of the observations to those groups.

The Variance–Covariance and Correlation Matrices for Assessing GEI

In the context of the multivariate analysis of variance for a given classification of g groups, and using only the rp continuous variables, the square $rp \times rp$ matrices $\mathbf{V_T} = \mathbf{T}/n$, $\mathbf{V_B} = \mathbf{B}/g$ and $\mathbf{V_W} = \mathbf{W}/(n\text{-}g)$ represent the total, between, and pooled within-groups variance–covariance matrices, respectively ($\mathbf{T} = \mathbf{B} + \mathbf{W}$). Accordingly, the corresponding total ($\mathbf{R_T}$), between-groups ($\mathbf{R_B}$), and within-group ($\mathbf{R_W}$) correlations (Pearson product moment correlation) matrices can be calculated as $\mathbf{R_v} = \mathbf{D}^{-1} \times \mathbf{V} \times \mathbf{D}^{-1}$, where $\mathbf{V}$ is any of the three above-mentioned variance–covariance matrices. $\mathbf{D}$ is a diagonal matrix containing the square roots of the diagonal of $\mathbf{V}$ (that is the standard deviation for each variable). The elements of $\mathbf{R_T}$ represent the correlation of the variables across all the ungrouped cultivars before the classification; the elements of $\mathbf{R_B}$ contain the correlation between the centroid of the groups formed by the classification strategy and the elements of $\mathbf{R_W}$ had the pooled within-groups correlations.

When there are rp continuous and rq binary variables, the $(rp + rq) \times (rp + rq)$ matrices, $\mathbf{R_T}$, $\mathbf{R_B}$, and $\mathbf{R_W}$, contain the Pearson correlations between the continuous traits, the biserial correlation between binary and continuous traits, and the similarity measurement $1 - 0.5d^2$ (Anderberg, 1973) (where d^2 is the squared Euclidean distance or the simple matching distance) as the association between binary traits.

When the traits are p continuous and q ordinal, their values can be replaced by their ranks and the correlations can be computed. In this case, the $(rp + rq) \times (rp + rq)$ matrices, $\mathbf{R_T}$, $\mathbf{R_B}$, and $\mathbf{R_W}$, contain the Spearman rank correlations.

Because the variables in the three-way clustering analysis comprise the same traits measured in different environments, a good clustering strategy should have a between-group matrix, $\mathbf{R_B}$, with values larger than those of the total $\mathbf{R_T}$ matrix. This result indicates that the clustering strategy resulted in good control of the COI.

Once the three-way Ward-MLM classification strategy defines the final homogeneous groups, the GEI can be studied from different perspectives.

1. The diagonal elements of the within-group variance–covariance matrix are the variances estimated for each trait in each environment. Environments with large variance allow better expression of the cultivars than environments with small variance. In addition, heterogeneity of environmental variances for each trait reflects non-COI (Muir et al., 1992). This interaction is not relevant and can be ignored.
2. The correlation matrices, $\mathbf{R_T}$, $\mathbf{R_B}$, and $\mathbf{R_W}$, contain information about GEI. For example, consider the case of

two traits (T1 and T2) measured in two environments (E1 and E2). The three-way clustering strategy will form groups of cultivars with consistent responses across the four combinations T1E1, T1E2, T2E1, and T2E2. The correlations of interest are (i) T1E1 vs. T1E2 and T2E1 vs. T2E2, which measure the degree of similarity of the response between-groups ($\mathbf{R_B}$) or between cultivars within-groups ($\mathbf{R_W}$) with respect to the traits T1 and T2 in each environment. High and positive values of $\mathbf{R_B}$ (and/or $\mathbf{R_W}$) indicate non-COI, near zero or negative $\mathbf{R_B}$ (and/or $\mathbf{R_W}$) indicate COI; (ii) T1E1 vs. T2E1 and T1E2 vs. T2E2, which measure the degree of association between traits T1 and T2 in environments E1 and E2. If these correlations are similar, then there is no evidence of the effect of the environments on the association of the traits (no GEI effect on the association of the traits). If these correlations are different, then there is evidence of GEI in their association.

Correlations T1E2 vs. T2E1 and T2E1 vs. T1E2 are not of interest in this study.

Multidimensional Scaling for Assessing COI

The correlation across environments for studying non-COI and COI and the relationship between traits across environments can be better visualized in a two (or three) dimensional graph obtained from the multidimensional scaling method (MDS) (Mardia et al., 1979; Krzanowski and Marriot, 1994). The objective of the MDS method is to obtain a low-dimensional (two-dimensional, if possible) graphic representation of the $rp + rq$ variables (each variable corresponding to the combination of one trait and one environment) whose similarities have been measured by the correlation matrices. The MDS analysis requires a metric distance matrix; therefore, we used the optimal transformation (Mardia et al., 1979),

$$d_{ij} = [2(1 - r_{ij})]^{(1/2)}$$

where r_{ij} is the correlation coefficient between the ith and the jth variables. In this study, r_{ij} are the values of the matrices $\mathbf{R_B}$, $\mathbf{R_W}$, or $\mathbf{R_T}$.

The MDS method finds the geometric representation in two or three dimensions such that the sum of squares of the difference between the observed distances between two variables i,j (d_{ij}) in the $rp + rq$ dimensional space and the estimated distance in the two-dimensional space ($\hat{d}_{ij}$) is minimized (minimum standardized residual sum of squares, STRESS):

$$\text{STRESS} = [\Sigma_i \Sigma_j (d_{ij} - \hat{d}_{ij})^2 / \Sigma_i \Sigma_j d_{ij}^2]^{1/2}$$

In the geometric configuration of the MDS, two neighbor points indicate high positive correlation, whereas two distant points indicate negative correlation or absence of correlation. For example, the four variables representing trait T1 in environment E1, trait T2 in environment E1, trait T1 in environment E2, and trait T2 in environment E2, are T1E1, T2E1, T1E2, and T2E2, respectively. Assume that T1E1 and T2E1 are neighbors in the MDS graph and T1E2 is far away from T2E2 and the other two. Figure 1 is the MDS geometric representation of these four variables indicating that (i) traits T1 and T2 are highly correlated in E1 (variables T1E1 and T2E1 are together) but have small or negative correlation in E2 (variable T2E2 is near the center of the plot but variable T1E2 in the opposite quadrant), indicating an effect of the GEI in the relationship between traits T1 and T2; (ii) correlation between E1 and E2 for trait T1 (variables T1E1 and T1E2) is near zero or negative, that is, environments E1 and E2
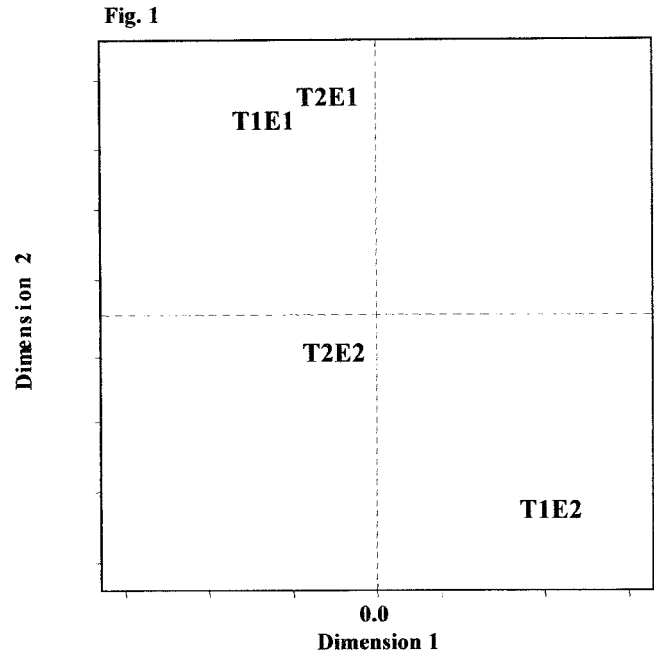

Fig. 1

Fig. 1. Hypothetical multidemensional scaling representation of the correlation coefficients among four variables (T1E1, T1E2, T2E1, and T2E2) representing two traits (T1 and T2) measured in two environments (E1 and E2).

ranked the groups or the cultivars within-groups differently. This indicates COI with respect to trait T1.

Experimental Data

Three distinct data sets were used to illustrate how the three-way Ward-MLM clustering strategy formed groups of cultivars with low COI for most of the traits measured. The data sets were selected (i) to include different types of crops and cultivars evaluated under very different environmental conditions, (ii) to have continuous as well as categorical traits, and (iii) to represent three multienvironment trials performed with different objectives. Data set 1 evaluated several maize (*Zea mays* L.) gene bank accessions with the purpose of forming core subsets. Data set 2 included maize inbred lines evaluated under different water regimens and levels of nitrogen. Data set 3 comprised a large number of bread wheat cultivars evaluated for various diseases important in the Southern Cone of South America.

Data Set 1

This data set came from 256 Caribbean maize accessions planted in three environments in México: Poza Rica (1992 and 1993) and Tlaltizapán (1994) (Taba et al., 1998; Franco et al., 1999). Means for the continuous traits, such as days to anthesis (DA), plant height (P), days to senescence (DS), and ear length (EL) in each environment, were used in the three-way Ward-MLM classification. The categorical trait was an agronomic scale (AS)(1 = poor, 2 = intermediate, 3 = good), which described the agronomic performance of the accessions in the field. After classification, the groups were characterized on the basis of the value of AS in each environment, as follows: **Good**—accessions with an AS value of 2 in one environment and 3 in the other environments (2,3,3 in any order) or accessions with values of 3 in the three environments (3,3,3); **Regular**—accessions with values of 2 for all environments (2,2,2) or 1 in one environment and 2 in the others (1,2,2 in any order) or 3 in one environment and 2 in the others (2,2,3 in any

Table 1. Means of six final groups (G1–G6) obtained from clustering 256 Caribbean maize accessions planted in three environments (data set 1) of the continuous traits, days to anthesis (DA), plant height (P), days to senescence (DS), and ear length (EL). Percentage of observations with different level of agronomic scale, and group size (*n*) (extracted from Table 4 of Franco et al., 1999).

Group	Continuous traits				Agronomic scale rating†				*n*
	DA	P	DS	EL	Good	Regular	Poor	Not defined	
	d	cm	d	cm			%		
G1	71	184	42	14.7	0	63	31	6	16
G2	84	246	42	15.9	4	77	17	2	48
G3	78	234	47	15.6	42	47	3	8	101
G4	79	223	45	15.9	0	77	9	14	22
G5	77	233	47	17.3	44	54	2	0	39
G6	90	246	39	14.2	0	10	80	10	30
Range	19	62	8	3.1	44	67	78	14	

† Good: agronomic scale value of 2 in one environment and 3 in the others or accessions with agronomic scale values of 3 in the three environments; Regular: agronomic scale values of 2 for all environments or 1 in one environment and 2 in the others or 3 in one environment and 2 in the others; Poor: agronomic scale values of 1 in all environments or 2 in one environment and 1 in the others; Not defined: agronomic scale value of 1,1,3 or 1,2,3 or 1,3,3 at the three environments.

order); **Poor**—accessions with values of 1 in all environments (1,1,1) or 2 in one environment and 1 in the others (2,1,1 in any order); **Not defined**—accessions with values of 1 and 3 in any two environments (1,1,3 in any order), (1,2,3 in any order), or (1,3,3 in any order). The correlations between the continuous and ordinal traits were computed as the Spearman rank correlations.

Data Set 2

The second data set comprised 211 maize lines tested in eight environments on the basis of different years (1992, 1994, 1996), seasons (A and B), and water and nitrogen stresses (in parentheses the number of identification of the environment): 1994-A-intermediate water stress (1); 1994-A-strong water stress (2); 1996-B-high nitrogen (3); 1996-B-low nitrogen; (4) 1996-A-low nitrogen (5); 1992-A-intermediate water stress (6); 1992-A-no water stress (7); and 1992-A-strong water stress (8). Means of the continuous trait anthesis-silking interval (A), number of rows per ear (E), days to silking (F), P, and grain yield (Y) were used. No categorical traits were measured. The correlations between the continuous traits were calculated as the Pearson correlations.

Data Set 3

The third data set is the 17th Vivero de Lineas Avanzadas de Trigo del Cono Sur (LACOS) (Kohli and Ulery, 1999), and it represents the evaluation of 205 bread wheat (*Triticum aestivum* L.) cultivars in seven sites. The continuous traits were the percentage of symptoms of leaf rust [LR, caused by *Puccinia triticina* Eriks. = *P. recondita* Roberge ex Desmaz. f. sp. *tritici* (Eriks. & E. Henn.) D.M. Henderson], powdery mildew (PM, *Erysiphe graminis* DC. f. sp. *tritici* Em. Marchal), Septoria leaf blotch (SP, caused by *Septoria tritici* Roberge in Desmaz.), and two morphological traits, P and DM. Because,

Table 2. Means of six final groups (G1–G6) obtained from clustering 211 maize lines planted in eight environments (data set 2) of the continuous traits anthesis silking interval (A), number of rows per ear (E), days to silking (F), plant height (P), grain yield (Y), and number of observations per group (*n*).

Group	A	E	F	P	Y	*n*
	d	Number	d	cm	Mg ha^{-1}	
G1	−0.156	8.9	77.7	165.3	4.23	57
G2	1.417	8.8	75.6	167.0	4.37	46
G3	0.229	9.4	78.1	181.5	5.19	20
G4	1.624	8.4	79.1	166.2	3.74	32
G5	1.032	9.9	76.1	179.5	5.81	24
G6	0.046	10.2	78.0	173.4	4.88	32
Mean	0.659	9.1	77.4	170.2	4.56	

in each site, cultivars may or may not be selected by the local breeder, the binary trait proportions of selected cultivars in each site were used (SE: 0 = no, 1 = yes). Each trait was measured in a different number of environments (6, 3, 3, 4, 7, and 6 environments, for LR, PM, SP, P, DM, and SE, respectively) to form a total of 29 variables (trait-environment combinations).

The correlations between continuous traits were computed as the Pearson correlations, those between binary traits were computed by the quantity, $1 - 0.5d^2$ and those between continuous and binary traits were calculated as the biserial correlations.

RESULTS AND DISCUSSION

For data set 1, the 256 accessions were classified into six final groups (G1–G6) (Table 1). Two groups, G3 and G5, with 101 and 39 observations, respectively, included plants that were early maturity, with large ears and good agronomic scale. Group G6 had 30 accessions with poor agronomic characteristics, late and tall plants with short ears. Franco et al. (1999) reported on the relationship between the groups and the geographical origin of the accessions.

For data set 2, the 211 cultivars were clustered into six groups (Table 2). Groups G3 and G5 had relatively high grain yield, plant height, and number of rows per ear, but exhibited differences in the anthesis-silking interval (A). Groups G2 and G4 showed low values for Y and E and high values for A. Group G1 had low values for A, E, P, and Y.

The 205 wheat cultivars of data set 3 were classified into six final groups (Table 3). Two groups, G4 and G6, had plants with low values of leaf rust, powdery mildew, and Septoria leaf blotch infection, and high values of the selected proportion. On the other hand, groups G1 and G5 showed high values for the disease symptoms, and the lowest values for the selected proportion.

In data set 1, GEI was evident for DS and AS, whereas the other traits did not show GEI (Fig. 2 of Franco et al., 1999). Table 4 shows that 91% (95 out of the total possible $15 \times 14/2 = 105$ pair-wise correlations) of the between-groups correlation coefficients were larger than the total correlation coefficients. This indicates that for data set 1, the three-way Ward-MLM strategy formed groups of cultivars with a similar response across

Table 3. Means of six final groups (G1–G6) obtained from clustering 205 bread wheat cultivars tested in seven sites (data set 3) of the continuous and binary traits, plant height (P), days to maturity (DM), symptoms of powdery mildew (PM), leaf rust (LR), and septoria leaf blotch (SP). Proportion of cultivars selected in Paso Fundo (PF), Ponta Grossa (PG), Barrow (BAR), Criadero (CRI), La Estanzuela (LE), and Marcos Juárez (MJ), mean of all environments (mean), and group size (n).

| Group | P | DM | Symptoms | | | Proportion of selected cultivars | | | | | | | n |
			PM	LR	SP	PF	PG	BAR	CRI	LE	MJ	Mean	
	cm	d		%									
G1	78.8	91.4	43.1	14.5	28.8	0.14	0.02	0.05	0.21	0.10	0.07	0.10	42
G2	82.0	109.8	33.8	14.3	22.0	0.00	0.11	0.22	0.00	0.44	0.22	0.17	9
G3	78.4	93.1	47.9	7.3	25.5	0.14	0.038	0.14	0.24	0.48	0.58	0.33	29
G4	79.1	94.9	36.4	10.3	15.4	0.12	0.10	0.02	0.49	0.53	0.27	0.26	49
G5	77.5	91.1	43.1	33.3	14.0	0.17	0.04	0.00	0.00	0.08	0.38	0.11	24
G6	76.5	94.4	37.3	7.3	13.9	0.04	0.19	0.52	0.31	0.23	0.58	0.31	52
Mean	78.2	94.0	40.3	12.8	19.3	0.11	0.14	0.18	0.27	0.30	0.36	0.23	

environments for all traits, that is, groups of cultivars with low COI.

For data set 1, in all cases (100%), the three pairwise correlation coefficients corresponding to each trait with respect to the three environments were larger for R_B as compared with R_T. For example, DA measured in environments 1, 2, and 3, DA1, DA2, and DA3, respectively, had between-groups correlation coefficients of 0.99 [(DA1 vs. DA2) = (DA1 vs. DA3) = (DA2 vs. DA3) = 0.99], whereas DA1, DA2, and DA3 had pair-wise total correlation coefficients of 0.95, 0.94, and 0.93 for (DA1 vs. DA2), (DA1 vs. DA3), and (DA2 vs. DA3), respectively (Table 4). These results indicate that for DA, the six final groups of cultivars identified by the three-way Ward-MLM method had non-COI GEI. However, for DS, the between-groups correlations coefficients were 0.69, 0.88, and 0.77 for (DS1 vs. DS2), (DS1 vs. DS3), and (DS2 vs. DS3), respectively (Table 4), indicating some COI GEI as shown in Fig. 2 of Franco et al. (1999).

For data set 2, Table 5 shows correlations among 16 variables (E1–E8, Y1–Y8), that is, two traits (number of rows per ear, E, and grain yield, Y) measured in eight environments. The three-way Ward-MLM method was effective in forming groups with low COI for some traits. Nearly 90% (701 out of 780) of the values in R_B were larger than the corresponding values in R_T. In 99% of the cases (221 out of 224), the 28 pair-wise correlation coefficients corresponding to each trait–environment combination were larger for the between groups than the total (data not shown). The values in R_W were smaller (and nearer to zero) than those in R_B for all the comparisons, showing the ability of the method for clustering cultivars with similar performance across environments (Table 5).

The three-way Ward-MLM applied to data set 3 formed groups with different levels of interaction with the environments, depending on the trait. Traits P, DM, and PM did not show any high level crossover GEI (Fig. 2a–2c); traits LR and SP showed some group × environment interaction (Fig. 2d and 2e). The proportion of selected cultivars (SE) is a trait that showed a high group × environment interaction (Fig. 2f), because in each environment the breeder selected the best cultivars and the infection pressure in each site was different. Therefore, this interaction indicates that cultivars selected in each environment depended on the cultivar ×

location × breeder interaction. Table 6 shows correlation of only two traits (LR vs. SE) measured in seven environments. In R_B, 87% of the correlation values (352 out of 406) were larger than those in R_T (data not shown). The relation between R_B and R_W matrices was similar to that observed for data set 2.

In summary, the results indicate that the three-way Ward-MLM strategy produced groups of cultivars with similar performance across environments (low levels of group × environment interaction). The increment of the correlation coefficients values of R_B with respect to the values of R_T shows that the groups formed by the three-way Ward-MLM strategy joined cultivars with similar trait response across environments.

The MDS representation of the rank correlation coefficients obtained from the between-groups correlation matrix (Fig. 3a) and the within-group correlation matrix (Fig. 3b) showed that the variables corresponding to the four continuous traits (DS, EL, DA, and P) across the three environments are located together for both correlation matrices. The categorical trait AS shared the third quadrant together with trait DS for between and within-groups correlation matrices. As previously mentioned, DA1, DA2, and DA3 had pair-wise between-groups correlation coefficients of 0.99 (Table 4), indicating a negligible COI.

The association between variables in opposite quadrants is negative or near zero. For example, the between-groups associations of P and EL in the three environments are in opposite quadrants of Fig. 3a and their between-group correlations are $r_{P1\ EL1} = 0.00$, $r_{P2\ EL2} = -0.16$, and $r_{P3\ EL3} = -0.01$ (Table 4). These indicate a small effect of the environments on the relationship of the traits P and EL. Traits DS and AS are in the same quadrant of Fig. 3a, and therefore, their correlations are positive; similar to the previous case, they do not change with the different environmental conditions, $r_{DS1\ AS1} = 0.79$, $r_{DS2\ AS2} = 0.89$, and $r_{DS3\ AS3} = 0.90$ (Table 4). Trait AS is positively associated with DS and EL and showed a negative or negligible association with P and DA. The ranges for these between-group correlations (shown in Table 4) are $-0.69 \leq r_{AS,DA} \leq -0.51$, $-0.33 \leq r_{AS,P} \leq 0.07$, $0.33 \leq r_{AS,EL} \leq 0.92$, and $0.78 \leq r_{AS,DS} \leq 0.90$. These associations were stable across environments.

From Table 4 and Fig. 3a and 3b, it can be concluded that both the between-group and the within-group anal-
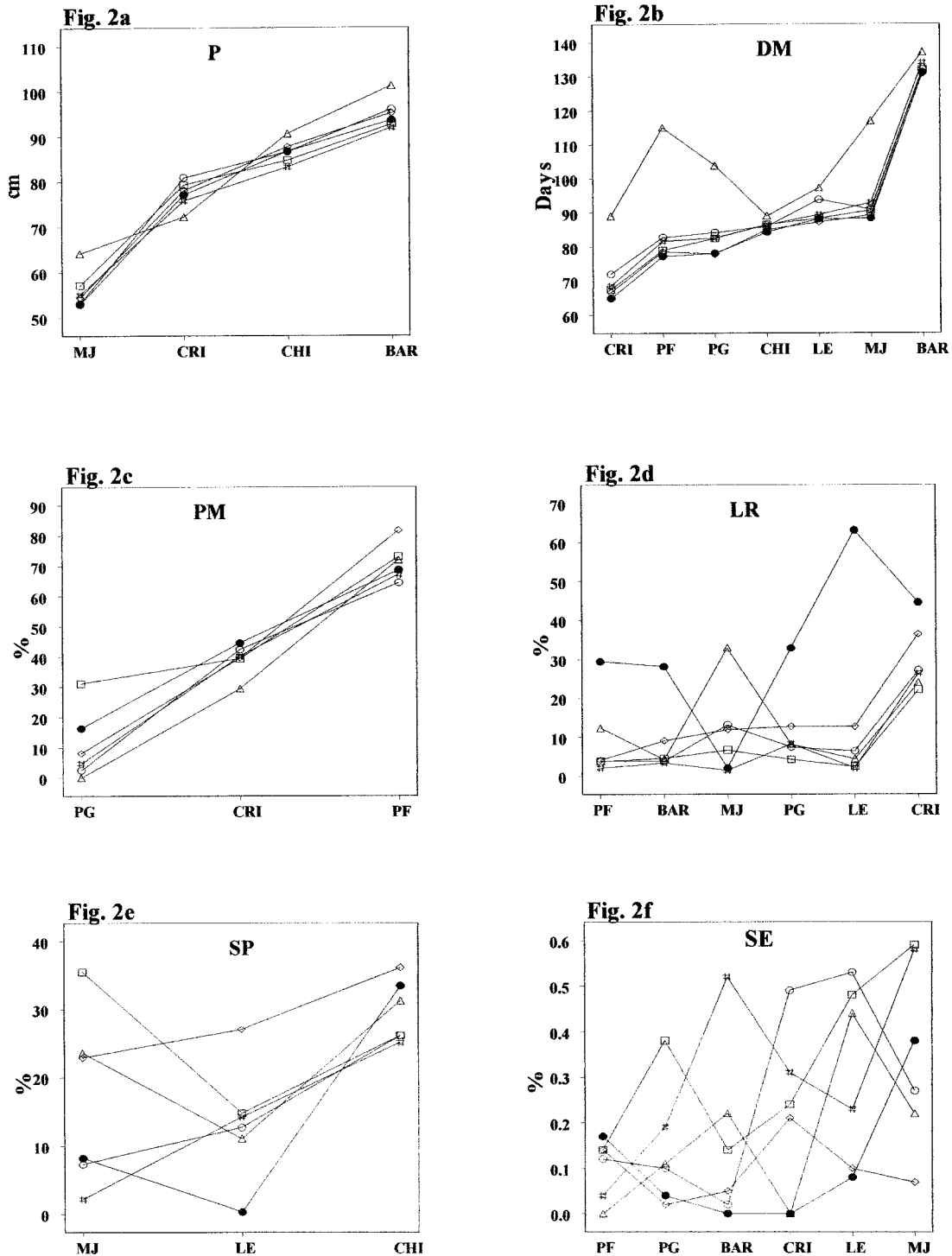
Fig. 2. Plot of the means of the groups obtained from the evaluation of 205 bread wheat cultivars in seven sites (data set 3). Traits are plant height (P), days to maturity (DM), symptoms of powdery mildew (PM), leaf rust (LR), Septoria leaf blotch (SP), and proportion of cultivars selected (SE). Environments are CRI (Paraguay), PF (Passo Fundo, Brasil), PG (Ponta Grossa, Brasil), CHI (Chillán, Chile), LE (Colonia, Uruguay), MJ (Marcos Juárez, Argentina), and BAR (Barrow, Argentina). Groups G1 (diamond), G2 (triangle), G3 (square), G4 (circle), G5 (dot), and G6 (hash).

yses showed negligible COI across environments for all five traits. Furthermore, the associations of the different traits do not seem to be highly affected by the differing environmental conditions. The three-way Ward-MLM clustering strategy formed homogenous and stable groups for most of the traits across the environments.

For data set 2, there are some differences on the MDS scatter plot of the between-group correlation matrix (Fig. 4a) as compared with the within-group correlation matrix (Fig. 4b). The MDS representation of the be-tween-group correlation matrix shows A and F together in all the environments and located in the first and

Table 4. Spearman rank correlation of the total matrix (R_T, upper diagonal) and the between-groups matrix (R_B, lower diagonal) for traits, days to anthesis (DA), plant height (P), days to senescence (DS), ear length (EL), and ordinal trait agronomy scale (AS) and environments 1, 2, and 3 for groups obtained from the evaluation of 256 Caribbean accessions planted in three environments (data set 1).

	DA1	DA2	DA3	P1	P2	P3	DS1	DS2	DS3	EL1	EL2	EL3	AS1	AS2	AS3
DA1	1	0.95	0.94	0.65	0.73	0.66	−0.86	−0.18	−0.27	−0.10	−0.16	−0.12	−0.16	−0.30	−0.19
DA2	0.99	1	0.93	0.64	0.72	0.65	−0.82	−0.23	−0.27	−0.11	−0.15	−0.10	−0.18	−0.31	−0.18
DA3	0.99	0.99	1	0.65	0.74	0.69	−0.83	−0.17	−0.32	0.12	−0.18	−0.15	−0.16	−0.30	−0.18
P1	0.89	0.87	0.86	1	0.78	0.76	−0.50	−0.04	−0.06	0.21	0.01	0.08	0.07	−0.09	−0.04
P2	0.86	0.84	0.86	0.94	1	0.81	−0.59	0.08	−0.02	0.19	0.00	0.06	0.09	−0.04	0.04
P3	0.79	0.76	0.77	0.96	0.98	1	−0.50	0.09	0.04	0.23	0.03	0.12	0.17	0.01	0.10
DS1	−0.91	−0.91	−0.92	−0.62	−0.63	−0.49	1	0.33	0.42	0.24	0.22	0.18	0.32	0.40	0.29
DS2	−0.44	−0.44	−0.42	−0.09	0.06	0.16	0.69	1	0.55	0.34	0.29	0.25	0.40	0.58	0.46
DS3	−0.59	−0.60	−0.64	−0.21	−0.26	−0.09	0.88	0.77	1	0.34	0.29	0.27	0.45	0.53	0.53
EL1	−0.35	−0.39	−0.40	0.00	−0.02	0.10	0.57	0.45	0.62	1	0.52	0.58	0.48	0.38	0.27
EL2	−0.48	−0.50	−0.53	−0.15	−0.16	−0.05	0.69	0.52	0.75	0.94	1	0.60	0.20	0.30	0.20
EL3	−0.35	−0.37	−0.40	−0.06	−0.09	0.01	0.52	0.31	0.55	0.98	0.95	1	0.27	0.28	0.16
AS1	−0.51	−0.53	−0.55	−0.09	−0.08	0.07	0.79	0.78	0.89	0.89	0.92	0.81	1	0.54	0.44
AS2	−0.67	−0.67	−0.69	−0.33	−0.27	−0.14	0.89	0.89	0.94	0.59	0.75	0.52	0.89	1	0.48
AS3	−0.57	−0.55	−0.57	−0.27	−0.18	−0.08	0.80	0.90	0.90	0.41	0.60	0.33	0.77	0.97	1

Table 5. Pearson correlation coefficients of the within-groups correlation matrix (R_W, upper diagonal) and the between-groups correlation matrix (R_B, lower diagonal) for traits number of rows per ear (E) and grain yield (Y) across eight environments (1–8) for groups obtained from the evaluation of 211 maize lines tested in eight environments (data set 2).

	E1	E2	E3	E4	E5	E6	E7	E8	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8
E1	1	0.34	0.25	0.11	0.24	0.31	0.31	0.33	0.66	0.34	0.15	0.13	0.13	0.31	0.23	0.21
E2	0.92	1	0.21	−0.09	0.14	0.10	0.16	0.12	0.22	0.49	0.09	0.04	0.08	0.15	0.07	0.15
E3	0.95	0.92	1	0.28	0.14	0.18	0.23	0.13	0.11	0.11	0.55	0.19	0.09	0.12	0.17	0.08
E4	0.79	0.86	0.74	1	0.27	0.08	0.06	0.07	0.10	−0.13	0.21	0.59	0.02	−0.11	−0.03	−0.04
E5	0.36	0.47	0.60	0.48	1	0.13	0.17	0.10	0.23	0.08	0.05	0.15	0.52	0.07	0.20	0.10
E6	0.54	0.62	0.70	0.54	0.74	1	0.45	0.33	0.17	0.00	−0.02	0.06	0.10	0.45	0.13	0.17
E7	0.28	0.34	0.50	0.22	0.72	0.93	1	0.40	0.21	0.11	0.05	0.06	0.14	0.23	0.48	0.22
E8	0.90	0.87	0.98	0.73	0.71	0.77	0.60	1	0.31	0.13	−0.11	0.07	0.09	0.25	0.24	0.59
Y1	0.90	0.72	0.80	0.55	0.07	0.49	0.30	0.74	1	0.44	0.16	0.14	0.24	0.40	0.34	0.39
Y2	0.96	0.97	0.95	0.80	0.42	0.69	0.43	0.90	0.86	1	0.26	0.01	0.24	0.32	0.29	0.25
Y3	0.84	0.75	0.89	0.48	0.40	0.76	0.65	0.87	0.90	0.88	1	0.31	0.19	0.13	0.24	0.00
Y4	0.69	0.76	0.69	0.97	0.61	0.53	0.25	0.71	0.41	0.69	0.38	1	0.17	−0.02	0.01	0.15
Y5	0.65	0.61	0.82	0.43	0.75	0.91	0.89	0.88	0.61	0.71	0.88	0.43	1	0.14	0.22	0.15
Y6	0.89	0.73	0.78	0.60	0.07	0.50	0.29	0.73	0.99	0.87	0.87	0.47	0.58	1	0.43	0.37
Y7	0.65	0.48	0.66	0.37	0.26	0.74	0.68	0.68	0.84	0.68	0.89	0.30	0.79	0.84	1	0.29
Y8	0.95	0.78	0.91	0.57	0.29	0.47	0.29	0.87	0.93	0.87	0.89	0.48	0.68	0.89	0.71	1

Table 6. Pearson correlations between leaf rust (LR) variables, biserial correlations between LR and proportion of selected plants (SE) variables, and similarities $1 − 05d^2$ (d^2 = Square Euclidean distance) between SE variables for groups obtained from the evaluation of 205 bread wheat cultivars tested in seven sites (data set 3). Measurements of the within-group correlation matrix (R_W, upper diagonal) and the between-groups correlation matrix (R_B, lower diagonal).

VARLOC	LR1	LR3	LR4	LR5	LR6	LR7	SE1	SE3	SE4	SE5	SE6	SE7
LR1	1	0.21	0.49	0.08	0.08	0.16	−0.09	−0.04	−0.05	−0.09	−0.04	−0.06
LR3	0.88	1	0.36	0.13	0.16	0.38	0.04	−0.20	−0.08	0.04	0.00	0.03
LR4	0.99	0.86	1	0.27	0.19	0.33	−0.04	−0.13	−0.08	0.00	0.04	−0.07
LR5	−0.25	−0.16	−0.24	1	0.11	0.14	−0.04	−0.06	0.01	−0.07	0.17	−0.04
LR6	0.95	0.73	0.96	−0.12	1	0.26	−0.06	−0.07	−0.12	0.07	−0.04	0.00
LR7	0.98	0.91	0.98	−0.27	0.93	1	0.02	−0.08	−0.12	−0.02	0.13	0.08
SE1	−0.43	−0.44	−0.43	−0.46	−0.38	−0.33	1	0.03	0.19	0.25	0.10	−0.05
SE3	−0.71	−0.56	−0.64	0.05	−0.71	−0.67	0.05	1	0.04	0.12	0.01	−0.05
SE4	−0.58	−0.76	−0.51	0.39	−0.38	−0.64	−0.14	0.64	1	0.13	−0.06	−0.10
SE5	−0.18	−0.51	−0.17	−0.64	−0.07	−0.20	0.69	0.02	0.21	1	−0.04	−0.01
SE6	0.55	0.55	0.53	−0.01	0.40	0.44	−0.87	−0.13	−0.09	−0.45	1	0.06
SE7	−0.44	−0.75	−0.47	−0.29	−0.35	−0.55	0.41	0.13	0.49	0.81	−0.24	1

fourth quadrants, respectively, whereas P, E, and Y are mixed in the second and third quadrants (Fig. 4a). Clearly, with respect to A and F, the groups formed by the three-way Ward-MLM strategy are very consistent across environments and showed low COI; for P, E, and Y the relationships are less consistent.

Within the clusters comprising P1 through P8, Y1 through Y8, and E1 through E8 a strong association can be observed of the traits E and Y in environments 1, 2, 3, 4, and 8 and a weaker association in environments 5, 6, and 7, that is, E1Y1 ($r = 0.90$), E2Y2 ($r = 0.97$), E3Y3 ($r = 0.89$), E4Y4 ($r = 0.97$), E8Y8 ($r = 0.89$),

and E5Y5 ($r = 0.75$), E6Y6 ($r = 0.50$), and E7Y7 ($r = 0.68$) (Table 5 and Fig. 4a).

The MDS analysis of the within-group correlation matrix shows traits A, F, and P clearly clustered in three separate groups. Traits E and Y form spread out pairs, for environments 1 to 5, E1 through Y1, E2 through Y2, E3 through Y3, E4 through Y4, and E5 through Y5 (Table 5 and Fig. 4b). In general, environmental effects do not seem to greatly influence the relationship between traits, and similar to traits A and F, groups for traits P, E, and Y showed low GEI.

In conclusion, for data set 2, with respect to A and
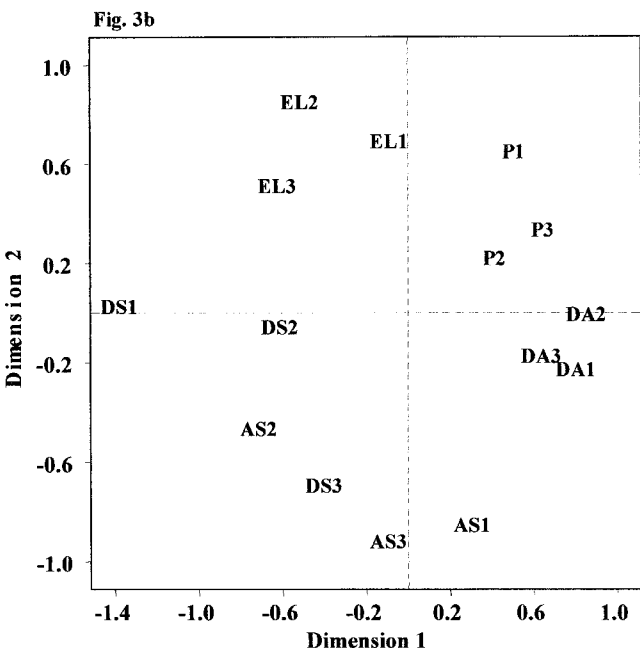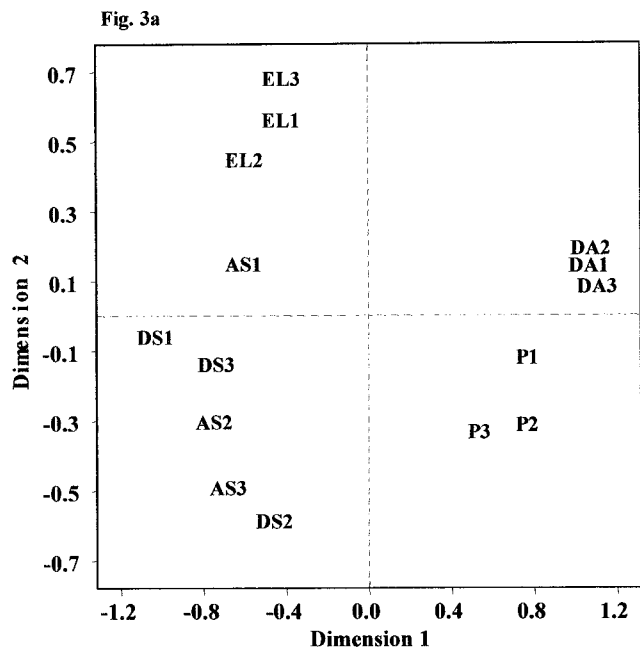
Fig. 3a

Fig. 3b

Fig. 3. Multidimensional scaling representation of the Spearman rank correlation coefficients obtained from the between-groups correlation matrix (Fig. 3a) and the within-group correlation matrix (Fig. 3b), for 15 variables (trait–environment) of the groups obtained from the evaluation of 256 Caribbean maize accessions planted in three environments (data set 1). Traits are ear length (EL), plant height (P), days to senescence (DS), days to anthesis (DA), and agronomic scale (AS). Environments are 1 through 3.
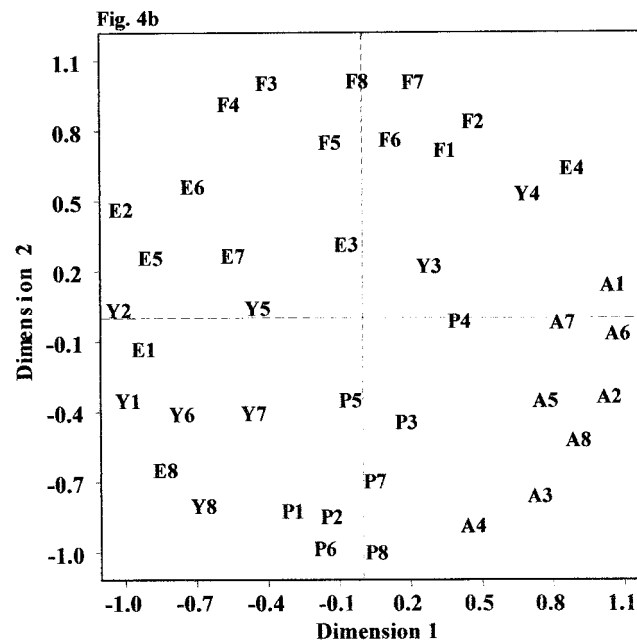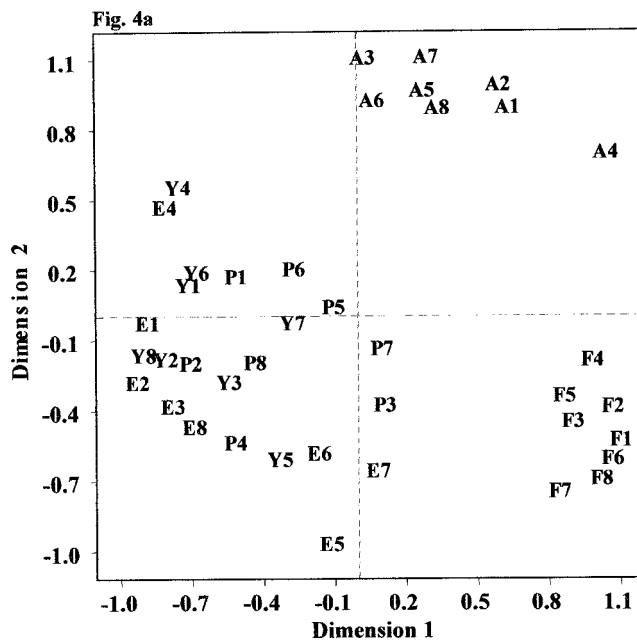
Fig. 4a

Fig. 4b

Fig. 4. Multidimensional scaling representation of the Pearson correlation coefficients obtained from the between-groups correlation matrix (Fig. 4a), and the within-group correlation matrix (Fig. 4b), for 40 variables (trait–environment) of the groups obtained from the evaluation of 211 maize lines evaluated in eight environments (data set 2). Traits are anthesis-silking interval (A), number of rows per ear (E), yield (Y), days to silking (F), and plant height (P). Environments are 1 through 8.

F, the groups formed by the three-way Ward-MLM strategy are very consistent across environments and showed low GEI. The MDS analysis of the within-group correlation matrix showed traits A, F, and P clearly clustered in three separate groups. Traits E and Y formed spread out pairs, for environments 1 to 5. More-over, environmental effects do not seem to greatly influence the relationship between traits, and similar to traits

A and F, groups for traits P, E, and Y showed low COI GEI.

For data set 3, Fig. 5a and 5b show that traits DM and SP are located in fairly compact groups in both analyses, indicating a low COI for groups × environment and for cultivars within-groups × environment. Results of the MDS analysis, using the between-groups correlation matrix (Fig. 5a), show that traits DM, SP, SE, and LR formed well-delineated groups, except for
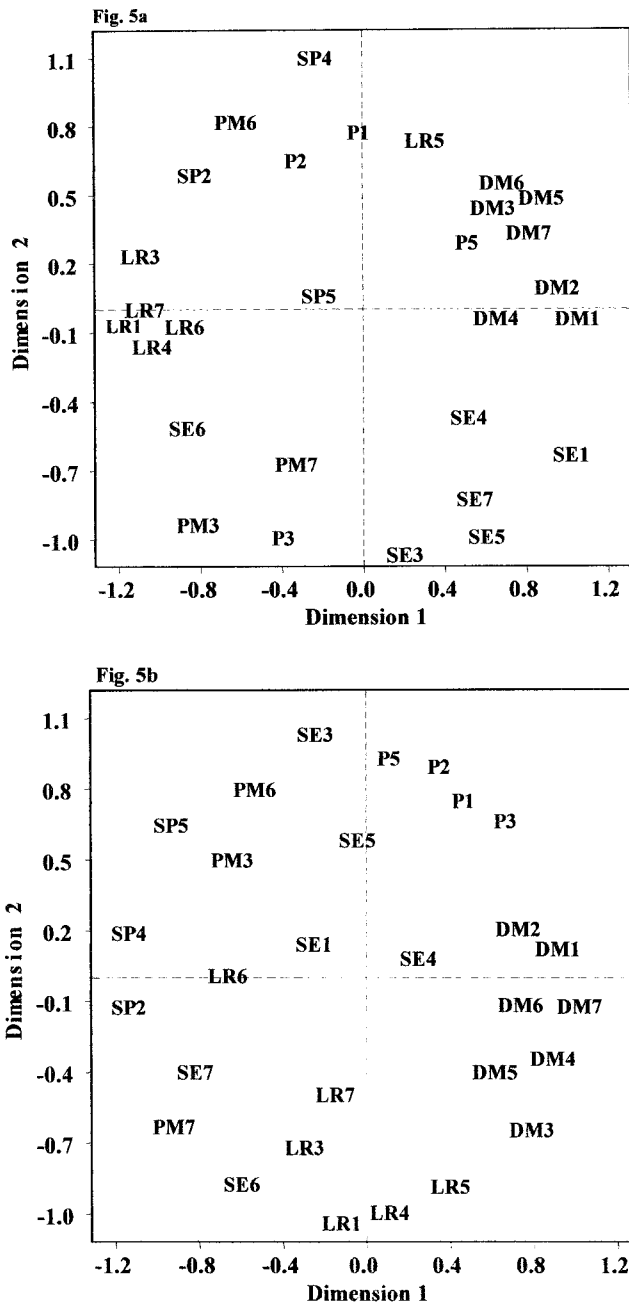
Fig. 5. Multidimensional scaling representation of the Pearson and biserial correlations and $(1 - 0.5d^2)$ similarity obtained from the between-groups correlation matrix (Fig. 5a), and the within-group correlation matrix (Fig. 5b) for 29 variables (trait–environment) variable of the groups obtained from the evaluation of 205 bread wheat cultivars evaluated in seven sites (data set 3). Traits are plant height (P), days to maturity (DM), powdery mildew (PM), septoria leaf blotch (SP), leaf rust (LR), and the proportion of selected cultivars (SE). Environments are 1 through 7.

LR5 and SE6. For the between-groups correlations, the Pearson, and biserial coefficients and the similarity measurements (Table 6) for variables LR1, LR3, LR4, LR6, and LR7 are all larger than 0.73, and the correlations with LR5 are, as expected, all negative. This indicates that environment 5, (Marcos Juarez, Argentina) generated a strong GEI, as indicated by the location of LR5 in the scatter plot of Fig. 5a. Similar behavior is detected

for trait SE in environment 6 (Passo Fundo, Brazil). Trait LR in environments 1, 3, 4, 6, and 7 is positively correlated with SE in environment 6.

The MDS analysis of the within-group correlation matrix of data set 3 showed traits P, DM, and sp. located in three well-separated groups (Fig. 5b). Variables from LR are located together in all environments, except environment 6 (LR6). Traits SE and PM are spread out in Fig. 5b, indicating that cultivars within-group × environment had COI and therefore near zero or low and negative correlations as shown in Table 6.

CONCLUSIONS

This study illustrated the use of the three-way Ward-MLM clustering strategy for grouping cultivars into homogeneous clusters with low GEI. The visual representation of the GEI of the continuous and categorical traits and environments after the grouping obtained by the three-way Ward-MLM clustering strategy is depicted in a scatter plot obtained by the multidimensional scaling method.

Results for the three data sets have shown that the three-way Ward-MLM strategy produced groups of cultivars with low levels of COI. The increment of the correlation coefficients values of between-groups with respect to the total correlation coefficients indicated that the groups formed by the three-way Ward-MLM strategy have similar trait response across environments. The use of MDS allows studying the effect of the GEI on the association between pairs of traits across environments and thus it represents an attempt to have a multivariate assessment of the GEI.

REFERENCES

Abdalla, O.S., J. Crossa, and P.L. Cornelius. 1997. Results and biological interpretation of shifted multiplicative model clustering of durum wheat cultivars and testing sites. Crop Sci. 37:88–97.

Anderberg, M.R. 1973. Cluster analysis for applications. Academic Press, New York.

Basford, K., and G.J. McLachlan. 1985. The mixture method of clustering applied to three-way data. J. Classif. 2:109–125.

Cornelius, P.L., M. Seyedsadr, and J. Crossa. 1992. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theor. Appl. Genet. 84:161–172.

Cornelius, P.L., D.A. Van Sanford, and M.S. Seyedsadr. 1993. Clustering cultivars into groups without rank-change interactions. Crop Sci. 33:1193–1200.

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

Crossa, J., and P.L. Cornelius. 1993. Recent developments in multiplicative models for cultivar trials. p. 571–577. In D.R. Buxton et al. (ed.) International crop science I. CSSA, Madison, WI.

Crossa, J., and P.L. Cornelius. 1997. Site regression and shifted multiplicative model clustering of cultivar trials site under heterogeneity of error variances. Crop Sci. 37:406–415.

Crossa, J., P.L. Cornelius, M. Seyedsadr, and P. Byrne. 1993. A shifted multiplicative model cluster analysis for grouping environments without genotypic rank change. Theor. Appl. Genet. 85:577–586.

Crossa, J., K. Basford, S. Taba, I. DeLacy, and E. Silva. 1995. Three-mode analyses of maize using morphological and agronomic attributes measured in multilocation trials. Crop Sci. 35:1483–1491.

Crossa, J., P.L. Cornelius, and M.S. Seyedsadr. 1996. Using the shifted multiplicative model cluster methods for crossover genotype-by-environment interaction. p. 175–198. In M.S. Kang and H.G. Gauch

(ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

Dempster, A.P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood for incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. B39:1–38.

Franco, J., J. Crossa, J. Villaseñor, S. Taba, and S.A. Eberhart. 1998. Classifying genetic resources by categorical and continuous variables. Crop Sci. 38:1688–1696.

Franco, J., J. Crossa, J. Villaseñor, S. Taba, and S.A. Eberhart. 1999. A two-stage, three-way method for classifying genetic resources in multiple environments. Crop Sci. 39:259–267.

Kohli, M.M., and A. Ulery. 1999. Resultados del decimoseptimo vivero de lineas avanzadas del Cono Sur. CIMMYT, Uruguay.

Krzanowski, W.J., and F.H.C. Marriot. 1994. Multivariate analysis. Part 1: Distributions, ordination and inference. E. Arnold, London.

Lawrence, C.J., and W.J. Krzanowski. 1996. Mixture separation for mixed-mode data. Statist. Comput. 6:85–92.

Mardia, K.V., J.T. Kent, and J.M. Bibby. 1979. Multivariate analysis. Academic Press, London.

Moll, R.H., C.C. Cockerham, C.W. Stuber, and W.P. Williams. 1978. Selection responses, genetic-environment interactions, and heterosis with reciprocal recurrent selection for yield in maize. Crop Sci. 18:641–645.

Muir, W., W.E. Nyquist, and S. Xu. 1992. Alternative partitioning of the genotype-by-environment interaction. Theor. Appl. Genet. 84:193–200.

Olkin, I., and R.F. Tate. 1961. Multivariate correlations models with mixed discrete and continuous variables. Ann. Math. Statist. 32:448–465.

Taba, S., J. Díaz, J. Franco, and J. Crossa. 1998. Evaluation of Caribbean maize accessions to develop a core subset. Crop Sci. 38:1378–1386.

Ward, J.H., Jr. 1963. Hierarchical grouping to optimize an objective function. J. Am. Statist. Assoc. 58:236–244.

Wishart, D. 1987. CLUSTAN user manual, 3rd Edition. Program Library Unit, University of Edinburgh, Edinburgh.