Biplot Analysis of Diallel Data

Weikai Yan* and L. A. Hunt

ABSTRACT

Diallel crosses have been used in genetic research to determine the inheritance of important traits among a set of genotypes and to identify superior parents for hybrid or cultivar development. Conventional diallel analysis is limited to partitioning the total variation of the data into general combining ability (GCA) of each genotype and specific combining ability (SCA) of each cross. In this paper we formulate a biplot approach for graphical diallel analysis. The biplot is constructed by the first two principal components (PCs) derived from subjecting the tester-centered diallel data to singular value decomposition. It displays the most important entry by tester patterns of the data and allows the following information to be extracted visually: (i) GCA of each genotype; (ii) SCA of each genotype; (iii) groups of parents with similar genetics; and (iv) superior hybrids. In addition, the biplot allows hypotheses to be formulated concerning the genetics of the genotypes. Three published diallel data sets of wheat (*Triticum aestivum* L.) and maize (*Zea mays* L.) were used to demonstrate the biplot approach and detailed procedures were provided for constructing and interpreting a biplot.

Diallel crosses have been used in genetic research to investigate the inheritance of important traits among a set of genotypes. Specifically, diallel crosses were devised to investigate GCA of the parents and to identify superior parents for use in hybrid and cultivar development. Analysis of diallel data is usually conducted according to the methods of Griffing (1956) that partition the total variation of the diallel data into GCA of the parents and SCA of the crosses. This paper introduces a biplot approach for visual analysis of diallel data.

The concept of biplot was developed by Gabriel (1971) to graphically display a rank-two matrix, which is a matrix resulting from multiplying a matrix with two columns by a matrix with two rows. The significance of this concept is that if a two-way dataset can be sufficiently approximated by a rank-two matrix, then it can be graphically displayed and investigated. Bradu and Gabriel (1978) explored the use of the biplot as a diagnostic tool for choosing an appropriate model for the analysis of two-way data. Since then, the biplot has been used in multi-environment trial (MET) data analysis.

In analyzing Ontario winter wheat performance trial data, Yan (1999) and Yan et al. (2000) proposed a GGE biplot, constructed from the first two principal components (PC1 and PC2) derived from PC analysis of environment-centered yield data. It was termed *GGE biplot* to emphasize that it displays both genotype main effect (G) and genotype × environment interaction (GE), which are two sources of yield variation that are relevant to, and must be considered simultaneously in, cultivar evaluation (Gauch and Zobel, 1996). Such a biplot has been used previously (Cooper et al., 1997), but the methodology formulated in Yan et al. (2000) allows the following issues to be visually addressed: (i) determination of the best performing cultivar in a given environment; (ii) identification of the most suitable environment for a given cultivar; (iii) comparison of any pair of cultivars in individual environments; and (iv) best cultivars for each environment and mega-environment differentiations. Later, Yan et al. (2001) developed an alternative GGE biplot that allows two additional questions to be explicitly addressed: (i) average yield and stability of the genotypes; and (ii) discriminating ability and representativeness of the environments, which are important for visual evaluation of cultivars and test environments for a given mega-environment.

Although the GGE biplot methodology was developed for MET data analysis, it should be applicable to all types of two-way data that assume an entry-by-tester data structure. In MET data, genotypes are entries and environments are testers. In diallel data, each genotype is both an entry and a tester. Our first objective is to demonstrate the use of biplot for diallel data interpretation. Another purpose of this paper is to provide a detailed description of the method for constructing and interpreting a GGE biplot.

MATERIALS AND METHODS

Three Diallel Datasets Used to Demonstrate the GGE Biplot Methodology

Wheat Resistance to Fusarium Head Blight

Buerstmayr et al. (1999) reported a diallel study of winter wheat resistance to Fusarium head blight (FHB). They used seven winter wheat genotypes of diverse origin and large differences in resistance to FHB. They presented data of areas under the disease progress curve and percentage of infected kernels for seven parents and their F1 hybrids; the two measures were highly correlated. Since no significant reciprocal effect was observed for this trait, averages of the reciprocals for individual crosses were reported. Here, only data for the percentage of infected kernels were used. The data were converted to percentage of kernels that were uninfected by FHB as a measure of resistance to the disease (Table 1).

Wheat Resistance to Stagonospora Nodorum Blotch

Du et al. (1999) reported a diallel study of six soft wheat genotypes, without reciprocals, evaluated for resistance to Stagonospora nodorum blotch (SNB), caused by *Stagonospora*

Crop Science Division, Dep. of Plant Agriculture, Univ. of Guelph, Guelph, ON, Canada N1G 2W1. Received 30 Nov. 2000. *Corresponding author (wyan@uoguelph.ca).

Published in Crop Sci. 42:21–30 (2002).

Abbreviations: ATC, average tester coordinate; FHB, Fusarium head blight; GCA, general combining ability; GGE, genotype main effect plus genotype × environment interaction; MET, multi-environment trial; PC, principal component; PC1, first principal component; PC2, second principle component; PSB, pink stem borer; SCA, specific combining ability; SNB, Stagonospora nodorum blotch; SREG, site regressions.

Table 1. Percentage of kernels unaffected by Fusarium head blight of seven winter wheat genotypes and their F1 hybrids. Data based on Buerstmayr et al. (1999). The genotype codes are A = Alidos, B = 81-F3-79, C = Arina, D = SVP-72017-17-5-10, E = SVP-C8718-5, F = UNG-136.1, and G = UNG-226.1.

	Testers							
Entries	A	B	C	D	E	F	G	Mean
	%							
A	27.5	35.7	46.4	53.7	33.3	64.9	43.3	43.5
B	35.7	37.5	46.2	40.8	51.9	45.6	57.5	45.0
C	46.4	46.2	38.7	49.1	50.4	55.6	69.4	50.8
D	53.7	40.8	49.1	51.2	49.4	48.1	57.5	50.0
E	33.3	51.9	50.4	49.4	42.5	63.1	68.9	51.4
F	64.9	45.6	55.6	48.1	63.1	60.0	63.1	57.2
G	43.3	57.5	69.4	57.5	68.9	63.1	43.7	57.6
Mean	43.5	45.0	50.8	50.0	51.4	57.2	57.6	50.7

nodorum (Berk.) Castellani & E.G. Germano. They reported data of incubation period, latent period, and percentage of necrosis of the parents and their F1 hybrids. Here, only the necrosis percentage data were taken and presented as percentage of healthy leaf areas as a measure of disease resistance (Table 2).

Maize Resistance to Pink Stem Borer

Butron et al. (1998) reported on percentage of yield loss in a 10-parent maize diallel cross following artificial infestation with corn pink stem borer (PSB, *Sesamia nonagrioides*). The yield loss data were converted into tolerance to PSB, measured by the yield of infested plants as percentage of that of non-infested plants (Table 3).

Mathematical Model for GGE biplot

Yan et al. (2001) compared two site regressions (SREG) models that can be used to generate GGE biplots, SREG2 and $SREG_{M+1}$. The SREG2 model consists of PC1 and PC2 derived from environment-centered data, referred to as primary and secondary effects, respectively; the $SREG_{M+1}$ model uses regressions of environment-centered data against genotype main effects as the primary effect, and PC1 derived from the residual of the regressions as the secondary effect. The SREG2 model had the advantage of explaining more variation, whereas the $SREG_{M+1}$ model had the advantage of explicitly indicating the average yield and stability of the genotypes and the representativeness and discriminating ability of the environments. However, through axis rotation, the SREG2 biplot can also approximately indicate the average yield and stability of the genotypes and the representativeness and discriminating ability of the environments. This allows the advantages of both models to be reasonably combined. Therefore, we use the SREG2 model for diallel data analysis.

Table 2. Percentage of healthy leaf area of six wheat genotypes and their F1 hybrids upon inoculation with Stagonospora nodorum Blotch. Data based on Du et al. (1999). The codes of the parents are: A = 18NT, B = Coker9543, C = Coker9803, D = L890682, E = TX82-11, and F = TX92D7374.

	Testers						
Entries	A	B	C	D	E	F	Mean
	%						
A	67.2	78.7	77.0	79.0	64.0	73.2	73.2
B	78.7	71.0	75.0	64.2	58.1	63.9	68.5
C	77.0	75.0	56.9	68.2	67.9	60.8	67.6
D	79.0	64.2	68.2	68.5	73.8	53.6	67.9
E	64.0	58.1	67.9	73.8	53.9	70.0	64.6
F	73.2	63.9	60.8	53.6	70.0	48.5	61.7

Table 3. Tolerance to infection by pink stem borer (PSB) of 10 corn inbreds (diagonal) and their F1 hybrids, as measured by the percentage of yield retained upon infection. Data based on Butron et al. (1998). The codes of the inbreds are: A = A509; B = A637; C = A661; D = CM105; E = EP28; F = EP31; G = EP42; H = F7; I = PB60; and J = Z77016.

	Testers										
Entries	A	B	C	D	E	F	G	H	I	J	Mean
	%										
A	85.8	89.1	86.3	82.0	86.6	92.4	82.9	88.1	84.2	86.0	86.3
B	89.1	88.3	79.7	72.4	89.6	78.5	97.6	84.8	83.0	89.8	85.3
C	86.3	79.7	89.9	88.3	97.2	86.0	72.2	92.5	74.9	82.6	85.0
D	82.0	72.4	88.3	(83.4)†	91.1	91.4	76.5	90.4	76.8	81.5	83.4
E	86.6	89.6	97.2	91.1	81.7	83.3	86.3	94.9	83.9	86.5	88.1
F	92.4	78.5	86.0	91.4	83.3	88.4	88.9	87.0	76.8	83.8	85.6
G	82.9	97.6	72.2	76.5	86.3	88.9	70.7	97.7	75.8	83.9	83.2
H	88.1	84.8	92.5	90.4	94.9	87.0	97.7	87.6	99.9	83.9	90.7
I	84.2	83.0	74.9	76.8	83.9	76.8	75.8	99.9	92.9	82.1	83.0
J	86.0	89.8	82.6	81.5	86.5	83.8	83.9	83.9	82.1	78.0	83.8

† Missing cell was replaced by the Column D average for completing the calculation.

When GGE biplot is applied to diallel data, the terms *average yield* and *stability* of the genotypes are correspondent to GCA and SCA, respectively, of the parents (Note that in conventional diallel analyses, SCA is associated with crosses rather then parents).

The SREG2 model is written as:

$$Y_{ij} - \beta_j = \lambda_1\xi_{i1}\eta_{j1} + \lambda_2\xi_{i2}\eta_{j2} + \in_{ij} \quad [1]$$

where Y_{ij} is the genotypic value of the combination (pureline parent or F1 hybrid) between Entry i and Tester j for the trait of interest; β_j is the mean of all combinations involving Tester j; λ_1 and λ_2 are the singular values for PC1 and PC2, respectively; ξ_{i1} and ξ_{i2} are the PC1 and PC2 eigenvectors, respectively, for Entry i; η_{j1} and η_{j2} are the PC1 and PC2 eigenvectors, respectively, for Tester j; and $\in_{ij}$ is the residual of the model associated with the combination of Entry i and Tester j. Since in diallel cross data each genotype is both an entry and a tester, i and j can refer to the same or different genotypes. When i = j, the combination is a pureline rather than a hybrid.

In some statistical software such as the Statistical Analysis System (SAS Institute, 1996), the singular values are usually combined with their respective row (entry) eigenvectors so that equation [1] looks like:

$$Y_{ij} - \beta_j = (\lambda_1\xi_{i1})\eta_{j1} + (\lambda_2\xi_{i2})\eta_{j2} + \in_{ij}. \quad [2]$$

To display PC1 and PC2 in a biplot, it is rearranged as

$$Y_{ij} - \beta_j = \xi^*_{i1}\eta^*_{j1} + \xi^*_{i2}\eta^*_{j2} + \in_{ij}, \quad [3]$$

where $\xi^*_{ik} = \lambda_k^{1/2}\xi_{ik}$ and $\eta^*_{jk} = \lambda_k^{1/2}\eta_{jk}$, with k = 1 or 2. This singular-value partitioning method is called *symmetrical scaling*.

Obtaining PC1 and PC2 Scores for Biplot Construction

Below we show how to obtain the scores for biplots using SAS, using the Fusarium study data as an example. When Table 1 is saved in space delimited text format, the following SAS statements will implement Eq. [2]:

```
DATA DS1;
INFILE 'FHB TABLE.PRN' FIRSTOBS=2;
INPUT ENTRY$ A B C D E F G;
PROC PRINCOMP COV OUT=PCAOUT N=2;
PROC PRINT;
ID ENTRY;
VAR PRIN1 PRIN2;
RUN;
```

Table 4. Principal component scores before scaling for the Fusarium head blight study.

	Entries		Testers	
	PC1†	PC2‡	PC1	PC2
	$\lambda_1\xi_{i1}$	$\lambda_2\xi_{i2}$	η_{j1}	η_{j2}
A	−25.178	15.626	0.579	−0.451
B	−8.965	−5.748	0.282	0.167
C	−3.396	−14.689	0.394	0.441
D	2.674	−8.695	0.071	0.252
E	−9.779	−0.112	0.652	0.036
F	22.020	−10.055	0.016	0.326
G	22.624	23.673	0.009	−0.635

† PC1 = first principal component.
‡ PC2 = second principal component.

Table 5. Principal component scores used for generating the biplot for the Fusarium head blight study.

	Entries		Testers	
	x-axis	*y*-axis	*x*-axis	*y*-axis
	ξ_{i1}^*	ξ_{i2}^*	η_{j1}^*	η_{j2}^*
A	−3.71	2.54	3.93	−2.78
B	−1.32	−0.93	1.92	1.03
C	−0.50	−2.39	2.68	2.71
D	0.39	−1.41	0.48	1.55
E	−1.44	−0.02	4.43	0.22
F	3.24	−1.63	0.11	2.00
G	3.33	3.85	0.06	−3.91

The SAS keyword *COV* specifies that the variance-covariance matrix calculated from the tester-centered diallel data be used in the principal component analysis. By default, without this specification, the correlation coefficient matrix would be used instead (SAS Institute, 1996). The SAS output of this program include, among others, (i) the eigenvectors of the first two PCs for each entry ($\lambda_1\xi_{i1}$ and $\lambda_2\xi_{i2}$) and each tester (η_{j1} and η_{j2}), which are listed in Table 4, (ii) the eigenvalues for PC1 and PC2, which are 304.227 and 205.019, respectively. The singular value for a PC is the square root of the sum of squares explained by the PC, which is the product of the eigenvalue multiplied by the number of entries. Therefore, the square root of the singular value for the kth PC is calculated as:

$$\lambda_k^{1/2} = (x_k n)^{1/4},$$

where x_k is the eigenvalue for the kth PC, and n is the number of entries. Symmetrical scaling is achieved by using $\lambda_k^{1/2}$ to divide the entry eigenvectors and to multiply the tester eigenvectors:

$$\xi_{ik}^* = \lambda_k^{1/2}\xi_{ik} = (\lambda_k\xi_{ik})/\lambda_k^{1/2}, \text{ and}$$
$$\eta_{jk}^* = \lambda_k^{1/2}\eta_{jk} = \eta_{jk}\lambda_k^{1/2}.$$

To illustrate, $\lambda_1^{1/2} = (304.227 \times 7)^{1/4} = 6.793$ and $\lambda_2^{1/2} = (205.019 \times 7)^{1/4} = 6.155$. These two values are used to divide the entry eigenvectors and to multiply the tester eigenvectors of Table 4 for PC1 and PC2, respectively, resulting in the symmetrically scaled PC1 and PC2 scores listed in Table 5. Values in Table 5 are used to construct the biplot (Yan et al., 2000, Fig. 1).

The biplot must be drawn to scale for meaningful visual analysis; the entry and tester names must be labeled precisely on the biplot; and various supplementary lines are necessary for effective visualization of the relationships among the entries, among the testers, and between the entries and the testers. The analyses reported here were conducted using the GGEbiplot software, a Windows application that fully automates biplot analysis of two-way data (Yan, 2001). The figures presented in this paper are direct outputs of the program.

RESULTS

Interpretation of the Wheat Fusarium Head Blight Data

General and Specific Combining Ability of the Entries

The biplot for the wheat FHB data explained 77% (46 and 31% by PC1 and PC2, respectively) of the total variation (Fig. 1). The GCA and SCA effects of the entries were examined by defining an average tester coordinate (ATC). An average tester is defined as a virtual tester whose PC1 and PC2 scores are equal to the average PC1 and PC2 scores, respectively, across all testers (indicated by a circle in Fig. 1A). The ATC is established with its abscissa passing through the origin and the average tester, and its ordinate passing through the origin and perpendicular to the abscissa (Fig. 1A). The positive end of the ATC abscissa is on the side of the biplot origin where the average tester is located.

The GCA effect of an entry may be defined by the value of its hybrid with the average tester. Thus, the GCA effects of the entries are approximated by their projections on to the ATC abscissa (i.e., vector of the average tester). The parallel lines perpendicular to the ATC abscissa help rank the entries in terms of GCA. Thus, Fig. 1A indicates that entries F and G (UNG-136.1 and UNG-226.1, both being derivatives of Sumai 3, a highly resistant wheat cultivar of Chinese origin) had the highest GCA effects, whereas Entry A (Alidos, the most susceptible genotype) had the lowest GCA effect. The GCA effects of the entries are in the order of: G > F > D > C > E ≈ B > A, which is roughly consistent with the order of G ≈ F > E ≈ D ≈ C > B > A, based on entry means (Table 1). Although there are variations, the entries with largest and smallest GCA effects were correctly identified by the biplot.

Because the biplot displays both GCA and SCA, and because GCA and SCA are orthogonal, if projections of the entries onto the ATC abscissa approximate their GCA effects, as just demonstrated, then projections of the entries onto the ATC ordinate must approximate their SCA effects, which represent the tendency of the entries to produce superior hybrids with specific testers. Entries G and A had the highest SCA effects (largest projections on to the ATC ordinate), whereas Entry E had the smallest SCA effect (smallest projection on to the ATC ordinate). Entries G and A had the highest SCA because they interacted positively with Testers B, C, D, and F, but negatively with themselves.

Two heterotic groups are suggested by Fig. 1A: Genotypes A and G as one group, and Genotypes B, C, D, and F as the other. Therefore, eight crosses, that is, [A, G] × [B, C, D, F] are expected to show heterosis defined as better than both parents. Entry E located near the ATC abscissa, and did not seem to belong to any of the groups.

Best Testers for General Combining Ability

An ideal tester should be highly discriminating of the entries and be highly representative of all testers. It is,

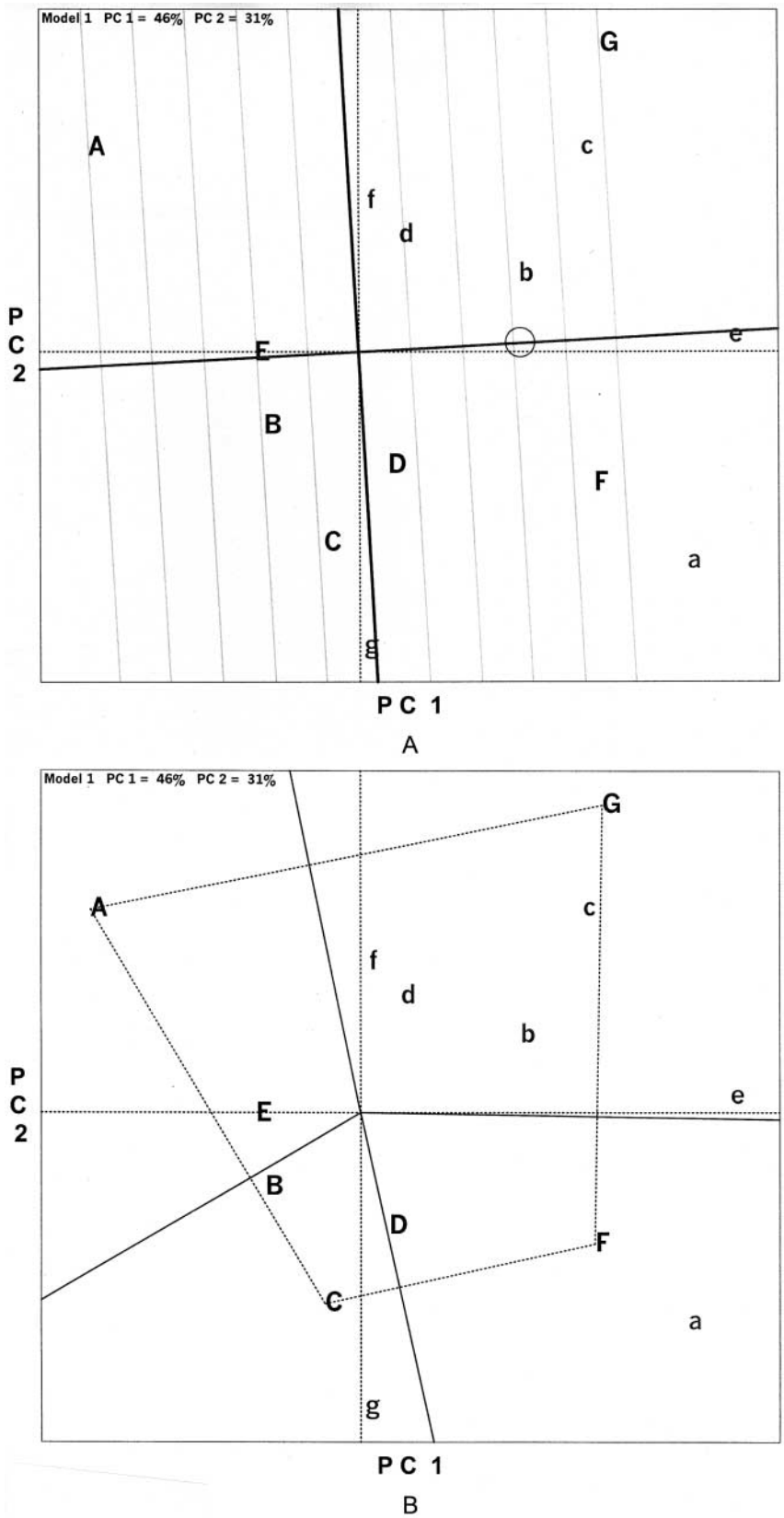

Fig. 1. Biplot based on diallel data of seven wheat genotypes with varying resistance to Fusarium head blight; (A) average tester ordination view, (B) polygon view. Codes of the genotypes are: A = Alidos, B = 81-F3-79, C = Arina, D = SVP-72017-17-5-10, E = SVP-C8718-5, F = UNG-136.1, and G = UNG-226.1. Genotypes are labeled with upper-case letters when viewed as entries and with lower-case letters when viewed as testers. Exact positions of the entries and testers are at the beginning of the labels. The circle indicates the average tester.

therefore, defined as a tester that has the longest vector of all testers (i.e., most discriminating) and zero projection onto the ATC ordinate (i.e., most representative of the testers). Therefore, the closer a tester's marker is to the ideal tester, the better it is as a tester. This ideal tester happens to coincide with the Tester E (Fig. 1A). Thus, Genotype E was the best tester in this dataset. The speculation is that the GCA effects of the entries should be reasonably assessed by the performance of their hybrid with Genotype E. Indeed, the entries are in the order of G > F > B ≈ C ≈ D > E > A, based on the actual values of the hybrids with Tester E, which is in rough agreement with the order of G ≈ F > E ≈ C ≈ D > B > A, based on the GCA effects, with the exception that Entry E per se was misplaced (Table 1).

Best Hybrids

The polygon view of a biplot provides the best way for visualizing the interaction patterns between entries and testers and to effectively interpret a biplot (Fig. 1B). It is drawn by connecting the entry markers positioned furthest from the plot origin using straight lines to form a polygon (or convex hull) such that all other entry markers are contained within the polygon. Lines perpendicular to each side or its extension of the polygon are drawn from the plot origin, which divide the biplot into several sectors, and each tester inevitably falls into one of the sectors. An interesting property of this polygon view of biplot is that testers falling into the same sector share the same best mating partner, which is the entry at the vertex of the polygon in that sector (Yan et al., 2000).

All interpretations of a biplot are based on the simple rule: the value of the hybrid between an entry and a tester is visualized by the projection of the entry's marker onto the vector of the tester or its extension (a vector of a tester is the line from the biplot origin towards the marker of the tester). Envision that a tester is located exactly on the perpendicular line to a side of the polygon; in other words, envision a tester whose vector coincides with the perpendicular line to a side of the polygon. Since the two vertex entries connected by this particular polygon side, and all entries located on this polygon side, have exactly the same projections onto the vector of the envisioned tester, they should be equally good mating partners with regard to the envisioned tester. Actually, they should be equally good with regard to all testers located on this perpendicular. They should be different from one another, however, with regard to all other testers. One entry will be better than the other with regard to testers located on its side of the perpendicular, and poorer than the other with regard to testers located on the other side of the perpendicular. Thus, the perpendicular lines divide the entries into groups.

Since the vertex entries have the largest distances from the origin, they are most responsive to the change of testers relative to other entries within respective groups. They are either the best or the poorest mating partners with some or all of the testers. It follows that the perpendiculars to the sides of the polygon also provide a way to group the testers based on their best mating partners. Thus, testers that fall in the same sector share the same best mating partner, which is the entry at the vertex of the polygon in that sector. Testers that fall in different sectors have different best mating partners. Entries located near the biplot origin are less responsive to the change of testers.

To illustrate, the biplot in Fig. 1B was divided into four sectors, with entries A, C, F, and G as the vertex entries, and are referred to as Sector A, Sector C, Sector F, and Sector G, respectively. No tester fell in Sector A, meaning that Entry A was not the best mating partner with any of the genotypes. Actually, Entry A produced the poorest combination or hybrid with itself and Tester E, which is located on the opposite side of the origin. A single tester, G, fell in Sector C, indicating that Entry C was the best mating partner with G. Moreover, since Genotype C, as a tester, was not in Sector C, the cross C × G must be better than both parents, and the term heterosis is used hereafter to refer to such situations. Had Tester C fallen in Sector C, the combination C × C (i.e., pureline C) would be the best among all crosses involving C, and therefore, heterosis between C and any other genotypes would not be possible. A single tester, Tester A, fell into Sector F, indicating that Entry F was the best mating partner for A, and the cross A × F was heterotic. Testers B, C, D, E, and F fell in sector G. Since G was not in this sector, all crosses between Genotype G and these genotypes should be heterotic.

To summarize, the biplot predicts the following F1 hybrids to be superior heterotic crosses: C × G in Sector C; F × A in Sector F; and G × B, G × C, G × D, G × E, and G × F in Sector G (Fig. 1B). In addition, Tester E was almost on the perpendicular line that separates Sectors F and G, meaning that Entries F and G were almost equally good as partners for E. Consequently, we have seven superior hybrids: F × [A, E] and G × [B, C, D, E, F]. Interestingly, in Sector G, G was predicted to be the best mating partner for C, and in sector C, C was predicted to be the best partner for G. C and G were, therefore, identified to be the best partners for each other, and C × G must be the best of all possible combinations. Other crosses were also predicted to be heterotic (Fig. 1A), such as A × C and A × D, but they were not predicted to be superior crosses (Fig. 1B).

Most of the above predictions can be verified from the original data (Table 1). Some are not consistent with the data, however. For example, the biplot predicts G to be the best and A the second best partner for F, whereas the data showed A to be the best and G the second best for F (Fig. 1B), even though the observed values for A × F (64.9%) and G × F (63.1%) were quite close. Also, based on Table 1, there were other heterotic crosses such as B × C, B × E, and C × E, which were not predicted to be heterotic by the biplot. (These crosses were apparently inferior to those that are identified to be superior crosses based on Fig. 1B). Such discrepancies are expected because the biplot explained 77 rather than 100% of the total variation. Since

all data contain some error, and since the biplot displays and makes predictions on the general pattern of the whole dataset, the predictions are probably more reliable than the individual observations.

Hypotheses Concerning the Genetic Relationships among the Genotypes

An immediate impression of Fig. 1B is that four vertex entries are apparently different from one another. Comparison of Entries A and G reveals the following information: (i) they were different in genetic responses; (ii) there was no heterosis between them; and (iii) as purelines, G was more resistant than A. This suggests that any dominant resistance genes (resistance genes are defined in this paper as all genes contributed to the apparent disease resistance regardless of their mechanisms) present in Entry A should also be present in Entry G. Similarly, any dominant resistance genes present in C should also present in F. Assuming that heterosis results from accumulation of different dominant gene loci, Entries A and C each would appear to carry at least one dominant resistance gene since heterosis was observed in their hybrids with F and G, respectively. Thus, F and G each would appear to carry at least two dominant resistance genes.

Because there was heterosis between F and G, they must each have carried a unique dominant resistance gene. These two genes are likely to be responsible also for the heterosis between A and F and that between C and G. Assuming that Genotypes B and D had similar genetics as C, as they were relatively close on the biplot, the heterosis observed between these genotypes and G can also be explained by the same two dominant genes.

Therefore, it would appear that at least three different resistance genes existed in the seven wheat genotypes: one shared by A and G, another shared by C (also B, D, and E) and F, and the third present in G but not in A, which may be the same as the one that is in F but not in B, C, D, or E. Were there no common dominant resistance genes in F and G, the hybrid F × G should be significantly better than A × F or C × G. Entry E may carry both genes from A and C, which predicts E to produce heterotic hybrids with both F and G but not with A or C. Following these reasoning, A × F, C × G, F × G, E × F, and E × G may each have combined all three dominant FHB resistance genes. However, since both the data and the biplot show that C × G was better than F × G and that E × G was better than E × F, there might be a recessive gene or a pair of epistatic genes present in C, E, and G. It follows that a cross combining Genotypes C (or E), F, and G may lead to breeding lines better than all of the parents. These analyses may help narrow down the crosses to be further investigated.

Interpreting the Wheat Stagonospora Nodorum Blotch Data

The biplot for the wheat Stagonospora nodorum blotch data explained 79% (59 and 20% by PC1 and PC2, respectively) of the total variation (Table 2, Fig. 2). Based on projections onto the ATC abscissa (Fig. 2A), Entry A showed the largest and Entry F the smallest GCA effects. The ranking of the genotypes for GCA was: A > B ≈ C ≈ D > E > F, which is consistent with the ranking based on entry means (Table 2). Fig. 2A suggests two heterotic groups: A and E as one and C, D, and F as the other. Therefore, hybrids [A, E] × [C, D, F] should show heterosis, as can be verified from Table 2. Entry B did not belong to any of the groups.

The biplot clearly shows why Entry A had the highest GCA effects: it was the vertex entry in a sector in which four of the other five testers, B, C, D, and F, fell (Fig. 2B). This means that Entry A was the best mating partner of Testers B, C, D, and F. Moreover, since Genotype A as a tester was not in this sector, heterosis was suggested in hybrids A × [B, C, D, F]. Fig. 2B also suggests that Entries C, D, and F (located on the same polygon side) were equally good for crossing with Testers A and E (located near the same perpendicular line). The crosses A × C, A × D, and A × F were indeed similar, and crosses E × C, E × D, and E × F were also similar (Table 2).

With regard to genetic constitutions, the six entries seemed to differ from one another, except that Entries C and D were apparently similar (Fig. 2B). Among the four vertex entries, A and E were different but the two did not produce heterosis, suggesting that any resistance gene in E must also be present in A (since A > E in GCA, Fig. 2A). Likewise, any resistance present in Entry F must also be present in Entry C (According to Table 2, C × F was slightly better than C, but this was not indicated by the biplot). E and F each carried at least one gene due to heterosis of some crosses. Thus A and C each must have carried at least two dominant resistance genes.

Heterosis occurred between A and C and between E and F, suggesting different dominant genes in A and E as one group and C, D, and F as another. Entry B, located near the plot origin, was intermediate between these two heterotic groups. Therefore, Entry B might carry two genes, one being the same as that in C and F, which caused heterosis when crossed with A; the other being the same as that in A and E. This explains the fact that B had better GCA effects than E and F and showed no large heterosis with any of the testers except A. Thus, at least four SNB resistance genes were involved in these six wheat genotypes. One shared by E, A, and B, another shared by F, C, D, and B, the third shared by A and C, and the fourth existed only in A. On the basis of these hypotheses, the crosses A × [B, C, D] should have carried all three genes and were equally resistant to SNB (Table 2).

Interpreting the Corn Pink Stem Borer Data

The biplot (Fig. 3) for the corn PSB study (Table 3) explained only 37% (PC1) + 26% (PC2) = 63% of the total variation. A large portion of the total variation was not accounted for by the biplot, reflecting the complexity of the genetics among the 10 corn inbreds in

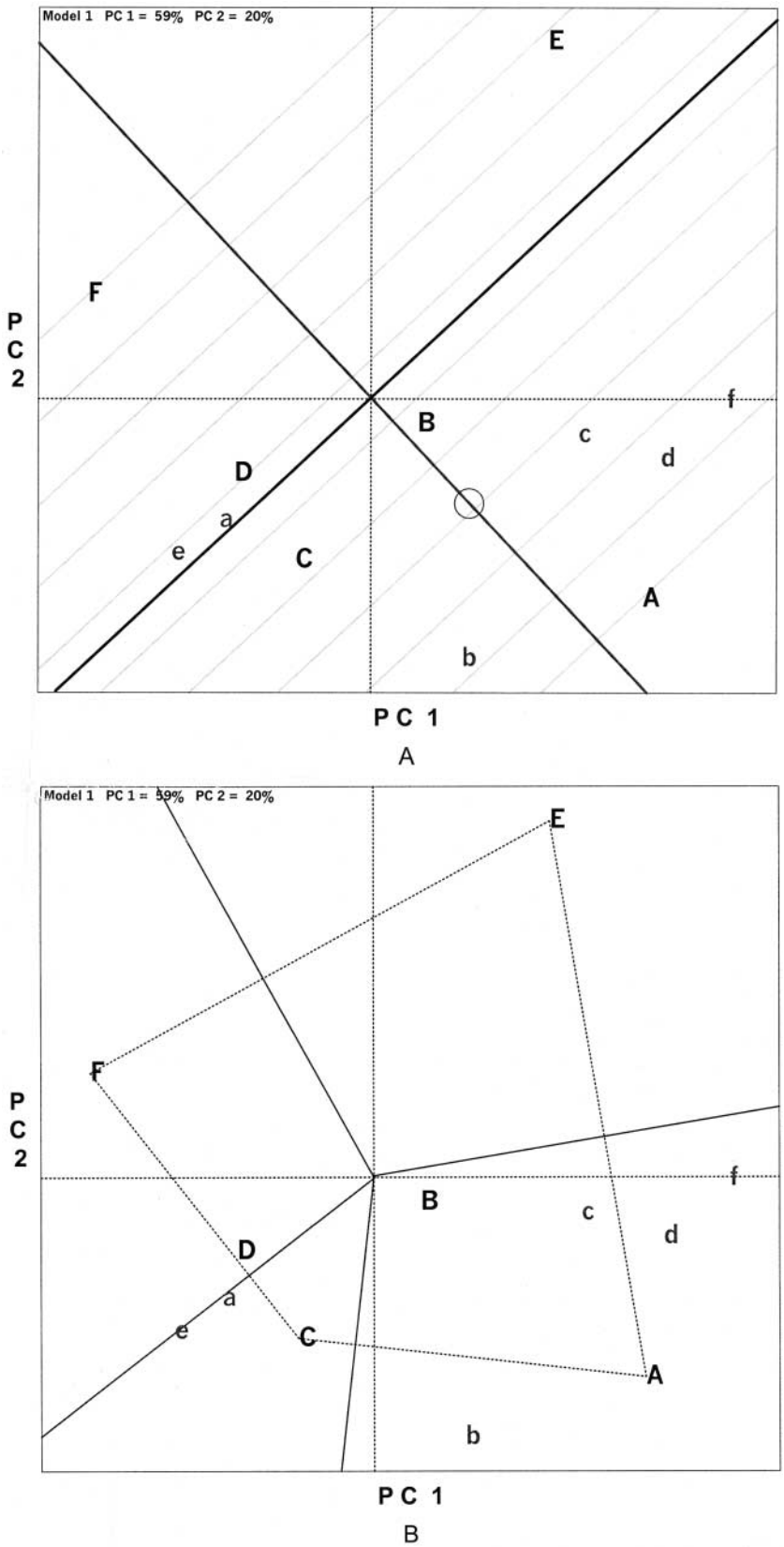

Fig. 2. Biplot based on diallel data of six wheat genotypes with varying resistance to Stagonospora nodorum blotch; (A) average tester ordination view, (B) polygon view. Codes of the genotypes are: A = 18NT, B = Coker9543, C = Coker9803, D = L890682, E = TX82-11, and F = TX92D7374. Genotypes are labeled with upper-case letters when viewed as entries and with lower-case letters when viewed as testers. Exact positions of the entries and testers are at the beginning of the labels. The circle indicates the average tester.

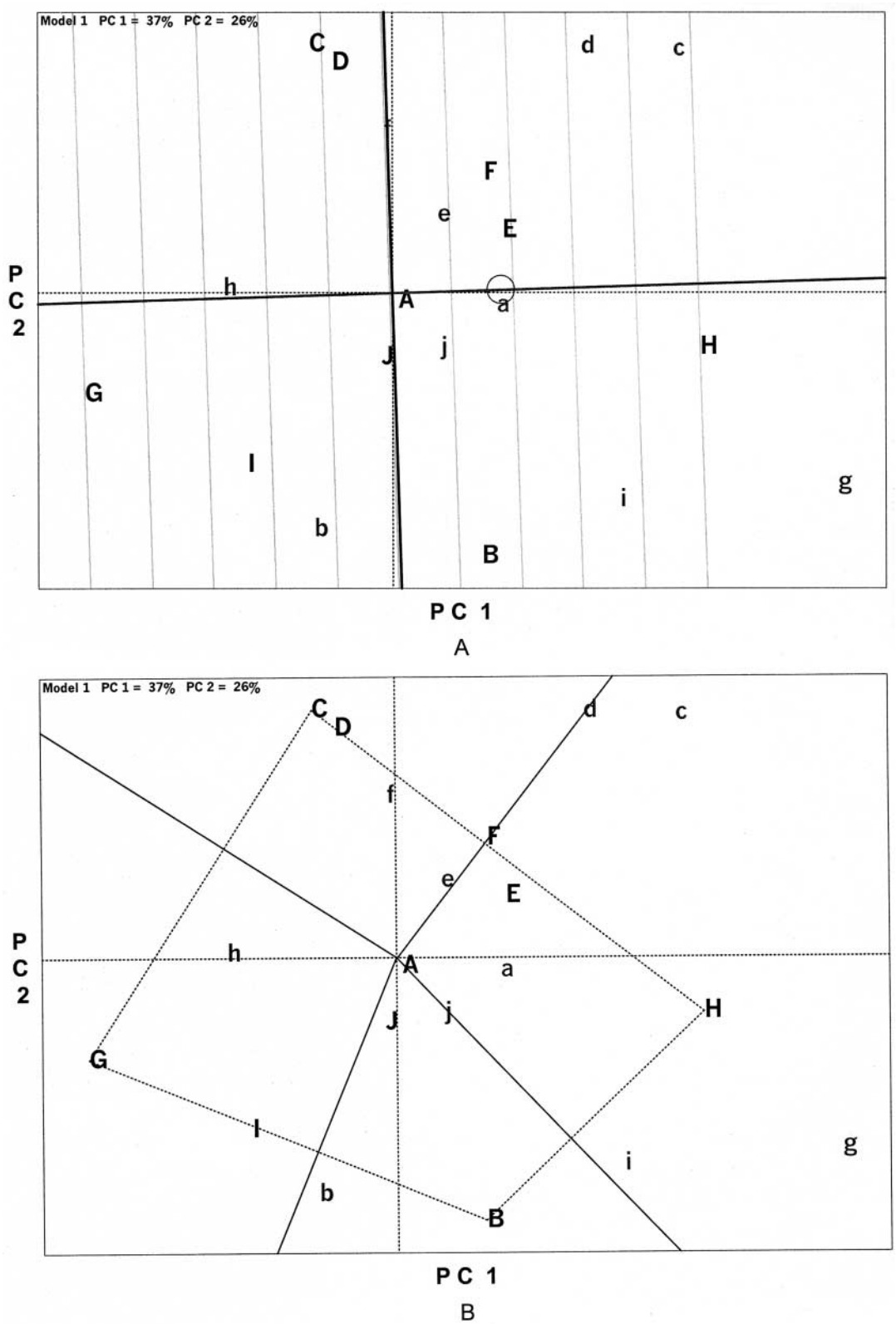

Fig. 3. Biplot based on diallel data of 10 corn inbreds with varying resistance to pink stem borer; (A) average tester ordination view, (B) polygon view. Codes of the genotypes are: A = A509; B = A637; C = A661; D = CM105; E = EP28; F = EP31; G = EP42; H = F7; I = PB60; and J = Z77016. Genotypes are labeled with upper-case letters when viewed as entries and with lower-case letters when viewed as testers. Exact positions of the entries and testers are at the beginning of the labels. The circle indicates the average tester.

PSB resistance. Nevertheless, the biplot still provides a useful tool for understanding the interactions among the inbreds.

First, the GCA and SCA effects of the parents can be visualized. On the basis of the projections onto the ATC abscissa (Fig. 3A), Entry H had the highest and G the smallest GCA. The ranking of the parents in GCA effects are: H > B ≈ E ≈ F > A ≈ J > C ≈ D > I > G, which is roughly consistent with the ranking of H > E > A ≈ B ≈ C ≈ F > D ≈ G ≈ I ≈ J based on the entry means (Table 3).

On the basis of projections onto the ATC ordinate (i.e., SCA) and along the ATC abscissa, the entries fell into two obvious heterotic groups: entries C and D as one group, and B and I as the other (Fig. 3A). These two groups interacted negatively, however, to give hybrids inferior to both parents, as can be verified from the original data (Table 3). *Negative heterosis* is predicted between the two groups because Entries C and D are located on the opposite of the ATC ordinate as Testers B and I, and vice versa (Fig. 3A). Negative heterosis may suggest involvement of recessive resistance genes, a phenomenon not observed in the two wheat data sets examined previously.

The polygon of Fig. 3B helps explain why Genotype H had the largest, and Genotype G the smallest, GCA effects. It indicates that Entry H was the best or near-best mating partner with seven of the 10 testers (except Genotypes B, F, and H per se). On the contrary, although Genotype G was the best mating partner with H and a good partner with B, it was a poor partner with all other testers. Fig. 3B also indicates that entries C and D were the best partners of Tester F, and Entry B was the best partner of B itself. The latter conflicts with the data and will be discussed further below.

Interestingly, Entries C, D, E, F, and H were all located on the same polygon side that connects Entries C and H, and Tester D located on the line perpendicular to this polygon side. This suggests that Genotypes C, D, E, F, and H were equally good in crossing with Genotype D. Examination of Table 3 indicates that this was indeed true. This may suggest that D carried epistatic effects with inbreds C, E, F, and H. Similarly, entries B, I, and G located on the same polygon side connecting B and G, and Tester B was almost on the perpendicular line of this polygon side. This suggests that Entries B, G, and I should produce similar hybrids with B. This suggestion was only partially true, however. The data indicate that hybrid G × B was much better than hybrid I × B, and B × B was intermediate (Table 3). The failure of the biplot to identify G as the best partner of B may have resulted from the major pattern that G had the lowest GCA (Table 3 and Fig. 3A). The biplot did indicate G to be the best partner of H, probably also because the major pattern that Genotype H had the largest GCA.

The 10 parents seemed to fall into seven groups, with C and D, E and F, and A and J being pairs of genotypes with apparently similar genetics (Fig. 3B). To understand the genetic relationships among the inbreds, we start from examining the vertex entries. As was previously discussed, Genotypes C and D may carry a recessive resistance gene, and Genotypes B and I may carry another recessive gene, which would explain the fact that hybrids between these two groups were inferior to both parents (Table 3). In contrast, the observed heterosis between Genotype H and Genotype G may suggest different dominant gene loci in H and G. Therefore, at least two recessive genes and two dominant genes may have been involved in the control of resistance to PSB in these four vertex inbreds.

Since there was no heterosis between G and C/D, C/D must carry the dominant gene present in G. Since there was heterosis in G × B, B must carry a dominant gene that is different from the one in G, which may be the same as the one in H due to lack of heterosis in B × H. Entries I, E/F and A/J, all located intermediate among the vertex genotypes, may be some types of the combinations of the dominant and recessive resistance genes. The pattern denoting H as the best mating partner with seven entries (all except F and B) suggests that H carries a dominant gene that is different from all dominant genes that are present in these seven entries.

DISCUSSION

Advantages of the Biplot Approach for Diallel Data Analysis

Compared with conventional methods of diallel analysis, the biplot approach has two advantages. The first advantage of the biplot is its graphical presentation of the data, which greatly enhances our ability to understand the patterns of the data. Assuming that the biplot sufficiently approximates the diallel data, the following information can be graphically visualized: (i) the GCA of the genotypes, (ii) the SCA of the genotypes, (iii) the best mating partner for each genotype, (iv) groups of similar genotypes, (v) the best crosses that are superior over their parents, and (vi) hypotheses to be formulated on the genetic relationships among the parents. This information can help the researcher focus on a few parents and crosses in further investigations.

The second advantage of the biplot approach is that it is more interpretative. While the conventional method of diallel analysis was designed to *describe* the phenotypic performance of the crosses, the biplot approach tries to *interpret* the phenotypic variation of the crosses by understanding the parents. In the conventional approach, although all variation is accounted for by GCA and SCA, the parents are evaluated only on their GCA effects. The term SCA is associated with crosses and has little impact on the understanding of the parents. Empirical evidence was provided in Yan et al. (2001), which demonstrated that entry PC1 scores had near-perfect correlation with entry main effects (i.e., the GCA effects, in terms diallel data) if the latter are larger than 35% of the total GGE variation; otherwise, the variation explained by PC1 would be considerably greater than the entry main effects. Because the PCs are least squares solutions, PC1 alone explains at least as much variation as, and typically more than, that by GCA. Thus, a biplot of PC1 vs. PC2 is generally more

powerful than the conventional approach in understanding the parents.

In our presentation of the results, the interpretations based on the biplots were frequently compared with the original data to indicate the validity of the biplot approach. The consistency between the biplot predictions and the original data should not be understood as indicating that the biplot approach is a redundant presentation of the data and, therefore, not needed. Rather, it indicates that the biplot is an excellent tool for revealing patterns that may not be noticed otherwise. For example, the biplot revealed for the corn PSB study that Genotype D tended to produce similar hybrids when crossed with Entries C, D, E, F, and H, and that Genotypes C and D had similar genetics (Fig. 3B). This result may be easily overlooked when using conventional methods.

Constraints of the Biplot Approach

A potential constraint of the biplot method is that it may fail to explain most of the variation and therefore fail to display all patterns of the data. This is most likely to occur with large datasets, small entry main effects, and complex entry-tester interactions. Even when this is the case, it can be assured that the biplot of PC1 vs. PC2, as least squares solutions, still displays the most important linear patterns of the data (Kroonenberg, 1995), as demonstrated by the corn PSB dataset. Nevertheless, other biplots, such as one consisting of PC3 vs. PC4, may be needed to fully understand the data. Such options are available in the GGEbiplot software (Yan, 2001). The method of Gauch and Zobel (1996) for estimating patterns vs. noise of the data may be adopted to determine if such biplots are needed. The pattern is estimated by the total SS of the tester-centered data minus the noise, which is estimated by the total treatment degrees of freedom multiplied by the error mean square and can be estimated from replicated data. A biplot of PC3 vs. PC4 is needed only when the pattern SS is considerably greater than that explained by the biplot of PC1 vs. PC2.

Another constraint of the biplot approach is lack of a measure of uncertainty. However, we suggest that the significance of the difference between two entries can be visually assessed from their plot distance relative to the plot size. In many cases, this visual assessment should be sufficient for a reasonable judgment; in other cases, the biplot patterns should be used to generate hypotheses rather than to make decisions.

The third constraint of the biplot approach is complexity of generating and interpreting biplots. This problem is solved, however, by the development of GGEbiplot software (Yan, 2001). GGEbiplot is a Windows application, which reads original data, generates biplots, and provides various perspectives of biplot visualization. It is available upon request with a charge.

REFERENCES

Bradu, D., and K.R. Gabriel. 1978. The biplot as a diagnostic tool for models of two-way tables. Technometrics 20:47–68.

Buerstmayr, H., M. Lemmens, S. Berlakovich, and P. Ruckenbauer. 1999. Combining ability of resistance to head blight caused by *Fusarium culmorum* (W.G. Smith) in the F1 of a seven parent diallel of winter wheat (*Triticum aestivum* L.). Euphytica 110: 199–206.

Butron, A., R.A. Malvar, P. Velasco, P. Revilla, and A. Ordas. 1998. Defense mechanisms of maize against pink stem borer. Crop Sci. 138:1159–1163.

Cooper, M., R.E. Stucker, I.H. DeLacy, and B.D. Harch. 1997. Wheat breeding nurseries, target environments, and indirect selection for grain yield. Crop Sci. 37:1168–1176.

Du, C.G., L.R. Nelson, and M.E. McDaniel. 1999. Diallel analysis of gene effects conditioning resistance to *Stagonospora nodorum* (Berk.) in wheat. Crop Sci. 39:686–690.

Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58:453–467.

Gauch, H.G., and R.W. Zobel. 1996. AMMI analysis of yield trials. p. 1–40. *In* M.S. Kang and H.G. Gauch (ed.) Genotype-by-environment interaction. CRC Press, Boca Raton, FL.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. Aust. J. Biol. Sci. 9:463–493.

Kroonenberg, P.M. 1995. Introduction to biplots for G×E tables. Dep. of Mathematics Res. Rep. no. 51. University of Queensland, St. Lucia, Australia.

SAS Institute. 1996. SAS/STAT user's guide. SAS Inst., Cary, NC.

Yan, W. 1999. Methodology of cultivar evaluation based on yield trial data—-with special reference to winter wheat in Ontario. Ph.D. diss. Univ. of Guelph, Guelph, ON, Canada.

Yan, W. 2001. GGEbiplot—a Windows application for graphical analysis of multi-environment trial data and other types of two-way data. Agron. J. 93:1111–1118.

Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavnics. 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci. 40:597–605.

Yan, W., P.L. Cornelius, J. Crossa, and L.A. Hunt. 2001. Two types of GGE Biplots for analyzing mulit-environment trial data. Crop Sci. 41:656–663.