

- by pedigree, F1 grain yield, grain yield heterosis, and RFLPs. *Theor. Appl. Genet.* 80:833–840.
- Stringfield, G.H. 1950. Heterozygosis and hybrid vigor in maize. *Agron. J.* 42:145–152.
- Troyer, A.F. 1990. A retrospective view of corn genetic resources. *J. Hered.* 81:17–24.
- Troyer, A.F. 1991. Breeding corn for the export market. p. 165–177. In D. Wilkinson (ed.) *Proc. Annu. Corn and Sorghum Ind. Res. Conf.*, 46th, Chicago, IL. 11–12 Dec. 1991. Am. Seed Trade Assn., Washington, DC.
- Troyer, A.F. 1994. Breeding early corn. p. 341–396. In A.R. Hallauer (ed.) *Specialty corns*. CRC Press, Boca Raton, FL.
- Troyer A.F. 1996. Breeding widely adapted, popular maize hybrids. *Euphytica* 92:163–174.
- Troyer, A.F. 1999. Background of U.S. hybrid corn. *Crop Sci.* 39:601–626.
- Troyer, A.F. 2000. Breeding temperate corn. p. 393–466. In A.R. Hallauer (ed.) *Specialty corns*. 2nd ed. CRC Press, Boca Raton, FL.
- Troyer, A.F., S.J. Openshaw, and K.H. Knittle. 1983. Measurement of genetic diversity among commercial corn hybrids. p. 83. In *Agronomy abstracts*. ASA, Madison, WI.
- Troyer, A.F., S.J. Openshaw, and K.H. Knittle. 1988. Measurement of genetic diversity among popular commercial hybrids. *Crop Sci.* 28:481–485.
- Troyer, A.F., and R.W. Rosenbrook. 1983. Utility of higher plant densities for corn performance testing. *Crop Sci.* 23:863–867.
- Zuber, M.S., and L.L. Darrah. 1980. 1979 U.S. corn germplasm base. p. 234–249. In D. Wilkinson (ed.) *Proc. Annu. Corn and Sorghum Ind. Res. Conf.*, 35th, Chicago, IL. 9–11 Dec. 1979. Am. Seed Trade Assn., Washington, DC.

## CROP BREEDING, GENETICS & CYTOLOGY

### Biplot Analysis of Test Sites and Trait Relations of Soybean in Ontario

Weikai Yan and Istvan Rajcan\*

#### ABSTRACT

Superior crop cultivars must be identified through multi-environment trials (MET) and on the basis of multiple traits. The objectives of this paper were to describe two types of biplots, the GGE biplot and the GT biplot, which graphically display genotype by environment data and genotype by trait data, respectively, and hence facilitate cultivar evaluation on the basis of MET data and multiple traits. Genotype main effect plus genotype by environment interaction effect (GGE) biplot analysis of the soybean [*Glycine max* (L.) Merr.] yield data for the 2800 crop heat unit area of Ontario for MET in the period 1994–1999 revealed yearly crossover genotype by site interactions. The eastern Ontario site Winchester showed a different genotype response pattern from the three southwestern Ontario sites in four of the six years. The interactions were not large enough to divide the area into different mega-environments as when analyzed over years, a single cultivar yielded the best in all four sites. The southwestern site, St. Pauls, was found to always group together with at least one of the other three sites; it did not provide unique information on genotype performance. Therefore, in future cultivar evaluations, Winchester should always be used but St. Pauls can be dismissed. Applying GT biplot to the 1994–1999 multiple trait data illustrated that GT biplots graphically displayed the interrelationships among seed yield, oil content, protein content, plant height, and days to maturity, among other traits, and facilitated visual cultivar comparisons and selection. It was found that selection for seed yield alone was not only the simplest, but also the most effective strategy in the early stages of soybean breeding.

SUPERIOR CULTIVARS must be evaluated on the basis of multi-environment trials (MET) and multiple traits to ensure that the selected cultivars have acceptable performance in variable environments within the target region and to meet the many-facets of the demand from the producers, processors, and the consumers. For

this reason, MET are conducted throughout the world for major crops every year in which multiple traits and characteristics are usually recorded. However, effective interpretation and utilization of the MET data in making selection decisions remain a major challenge to researchers. Effective analysis of MET data becomes an integral part of effective crop improvement. There are two major tasks for MET data analysis. The first is to determine whether the target region is homogeneous or should be divided into different mega-environments; the second is to select superior cultivars for a given mega-environment on the basis of multiple traits in addition to yield per se (Yan, 1999). The fulfillment of both tasks depends on an understanding of (i) the GE interaction pattern of the MET, which has been the focus of numerous studies (e.g., Kang, 1990; Kang and Gauch, 1996; Cooper and Hammer, 1996), and (ii) the interrelations among the breeding objectives (Yan and Wallace, 1995).

Recently, Yan (1999) and Yan et al. (2000) proposed a GGE biplot that allows visual examination of the GE interaction pattern of MET data. The GGE biplot emphasizes two concepts. First, although the measured yield is the combined effect of genotype (G), environment (E), and genotype by environment interaction (GE), only G and GE are relevant to, and must be considered simultaneously, in cultivar evaluation. Hence the term “GGE”. Second, the biplot technique developed by Gabriel (1971) was employed to approximate and display the GGE of a MET, hence the term GGE biplot. This GGE biplot was constructed by the first two principal

Crop Science Div., Dep. of Plant Agriculture, Univ. of Guelph, Guelph, ON, Canada N1G 2W1. Received 18 July 2000. \*Corresponding author (irajcan@uoguelph.ca).

**Abbreviations:** CHU, Corn Heat Unit; E, environment main effect; G, genotype main effect; GE, genotype by environment (or site) interaction; GGE, genotype main effect plus genotype by environment interaction effect; GT, genotype by trait interaction; MET, multi-environment trials; PC, principle component; SVD, singular value decomposition.

components (PC1 and PC2, also referred to as primary and secondary effects, respectively) derived from subjecting environment-centered yield data, i.e., the yield variation due to GGE, to singular value decomposition (SVD) (Yan, 1999; Yan et al., 2000). This GGE biplot was shown to effectively identify the GE interaction pattern of the data. It clearly shows which cultivar won in which environments, and thus facilitates mega-environment identification. A megaenvironment is defined as a group of locations that consistently share the same best cultivar(s). Another essential requirement for mega-environment differentiation is repeatability of the which-won-where pattern. Therefore, multi-site trials conducted over years are essential for addressing the mega-environment issue (Yan and Hunt, 1998; Yan, 1999; Yan et al., 2000).

In addition, the GGE biplot also has a usage in selecting superior cultivars and test environments for a given mega-environment. Provided that the genotypic PC1 scores have a near-perfect correlation with the genotype main effects, ideal cultivars should have a large PC1 score (high yielding ability) and a small (absolute) PC2 score (high stability). Similarly, ideal test environments should have a large PC1 score (more discriminating of the genotypes in terms of the genotypic main effect) and small (absolute) PC2 score (more representative of the overall environment) (Yan, 1999; Yan et al., 2000). The condition of near-perfect correlation between genotypic PC1 scores and genotype main effects is usually met for yearly METs (Yan and Hunt, 2001), but not for all datasets. To cover possible exceptions, an alternative GGE biplot was proposed, in which the PC1 is replaced by regressions of the environment-centered yield on the genotype main effects so that the primary scores for the genotypes are exactly the genotype main effects (Yan et al., 2001). This alternative model is thus more interpretable, but the greater interpretability is achieved at some expense of reduced variations explained by the biplots.

Soybean is the most important field crop in Ontario, grown for its protein and oil. The Ontario soybean-growing region is traditionally divided into several sub-regions on the basis of corn heat unit (CHU) (OMAFRA, 1993) and the breeding effort of the University of Guelph soybean-breeding project is focused primarily on the 2800 CHU sub-region, which primarily covers southwestern Ontario, but also a part of eastern Ontario. The routine cultivar evaluation was conducted in three southwestern Ontario sites (St. Pauls, Exeter, and Woodstock) and one eastern Ontario site (Winchester). Strong genotype by site interactions were noticed, though undocumented, from the yearly results, but it is not clear whether the four sites were different enough to constitute different mega-environments, and whether the four sites are sufficient or if any of them are redundant for cultivar evaluation for this area. Thus, the first objective of this study was to investigate the efficacy of the test sites using the GGE biplot technique (Yan et al., 2000). Cultivar evaluation on the basis of multiple traits is another important issue in plant breeding. The second objective of this study was to describe a genotype

by trait (GT) biplot, which is an application of the GGE biplot technique to study of the genotype by trait data, and to examine its usefulness in visualizing soybean trait relationships, and its application in cultivar evaluation, comparison, and selection.

## MATERIALS AND METHODS

### The Data

The Ontario soybean-growing region is conventionally divided into 2600, 2800, 3100, and 3400 CHU sub-regions. For each sub-region there is an independent cultivar evaluation project. The University of Guelph soybean-breeding program focuses on cultivar development and evaluation primarily for the 2800 CHU region. Data analyzed in this study were derived from the 1994 to 1999 soybean license trial data for the 2800 CHU region. Each year 90 to 125 adapted cultivars or breeding lines from public or private breeding programs were tested at four sites, namely Exeter, St. Pauls, Woodstock, and Winchester, representing the 2800 CHU region. The first three sites are located in southwestern Ontario with latitude around 42°N, while Winchester is located in eastern Ontario near 45°N. At each site, a lattice design (before 1998) or a nearest neighbor design (1998 or later) with four replicates was used. The experiments were planted according to local practice with planting rate about 50 seeds m<sup>-2</sup>. The harvested plot size was 8.25 m<sup>2</sup> (four 5.5 m rows with 37.5 cm apart). Days to maturity, plant height, and lodging score (1–5, with 1 being the best) were recorded in the field for each entry. Upon harvest, seed yield, protein concentration (%), oil concentration (%), 100-seed weight, and seed quality (1–5, with 1 being the best) were determined. For each entry at each test site, the average yield and average values of other traits were computed in accordance with the experimental design. The analysis reported here was based on these average values. In addition, protein yield and oil yield per hectare were computed by multiplying seed yield per hectare and protein concentration and oil concentration, respectively.

### The GGE Biplot

The GGE biplot method (Yan et al. 2000) was employed to study the genotype by site interaction of yield. It is based on the formula:

$$Y_{ij} - \bar{y}_j = \lambda_1 \xi_{i1} \eta_{j1} + \lambda_2 \xi_{i2} \eta_{j2} + \varepsilon_{ij} \quad [1]$$

where  $Y_{ij}$  is the average yield of genotype  $i$  in environment  $j$ ;  $\bar{y}_j$  is the average yield over all genotypes in environment  $j$ ; and  $\lambda_1 \xi_{i1} \eta_{j1}$  and  $\lambda_2 \xi_{i2} \eta_{j2}$  are collectively called the first principal component (PC1) and the second principal component (PC2);  $\lambda_1$  and  $\lambda_2$  are the singular values for the first and second principal components, PC1 and PC2, respectively;  $\xi_{i1}$  and  $\xi_{i2}$  are the PC1 and PC2 scores, respectively, for genotype  $i$ ;  $\eta_{j1}$  and  $\eta_{j2}$  are the PC1 and PC2 scores, respectively, for environment  $j$ ; and  $\varepsilon_{ij}$  is the residual of the model associated with the genotype  $i$  in environment  $j$ .

To display the PC1 and PC2 in a biplot, the  $\lambda$  values are absorbed into the genotype and environment scores so that the equation is written as:

$$Y_{ij} - \bar{y}_j = \xi_{i1}^* \eta_{j1}^* + \xi_{i2}^* \eta_{j2}^* + \varepsilon_{ij}$$

where  $\xi_{in}^* = \lambda_n^{1/2} \xi_{in}$  and  $\eta_{jn}^* = \lambda_n^{1/2} \eta_{jn}$ , with  $n = 1, 2$ . This scaling method has the advantage that PC1 and PC2 have the same unit (square root of original unit Mg ha<sup>-1</sup> in terms of yield), although other methods of scaling are equally valid.

A GGE biplot is generated by plotting  $\xi_{i1}^*$  and  $\xi_{i2}^*$  against

$\eta_{ji}^*$  and  $\eta_{j2}^*$ , respectively, so that each genotype or environment is represented by a marker in the biplot. The interpretation of a GGE biplot was first described in Yan (1999) and Yan et al. (2000).

### The Genotype by Trait Biplot

To display the genotype by trait two-way data in a biplot, the following formula is used:

$$\frac{T_{ij} - \bar{T}_j}{s_j} = \lambda_1 \zeta_{i1} \tau_{j1} + \lambda_2 \zeta_{i2} \tau_{j2} + \varepsilon_{ij} \quad [2]$$

where  $T_{ij}$  is the average value of genotype  $i$  for trait  $j$ ,  $\bar{T}_j$  is the average value of trait  $j$  over all genotypes,  $s_j$  is the standard deviation of trait  $j$  among the genotype averages;  $\zeta_{i1}$  and  $\zeta_{i2}$  are the PC1 and PC2 scores, respectively, for genotype  $i$ ;  $\tau_{j1}$  and  $\tau_{j2}$  are the PC1 and PC2 scores, respectively, for trait  $j$ ; and  $\varepsilon_{ij}$  is the residual of the model associated with the genotype  $i$  in trait  $j$ . Equation [2] is a principal component analysis of standardized data with two principal components. Because different traits use different units, the standardization is necessary to remove the units. PC1 and PC2 must be scaled as in Eq. [1] so that the 1 values are symmetrically distributed between the genotype scores and the trait scores. A genotype by trait (GT) biplot is constructed by plotting the PC1 scores against the PC2 scores for each genotype and each trait.

All biplots presented in this paper were generated by using the "GGEbiplot" software developed by the first author (yan@uoguelph.ca) of this paper, which fully automated the biplot analysis.

## RESULTS AND DISCUSSION

### Yearly Genotype by Site Relationships

Although the magnitude of G, E, and GE variances varied over years, GE is in general as important as G (Table 1). Although no error estimation was available due to the unavailability of replicated data, the magnitude of GE interaction relative to G justifies the consideration of GE in cultivar evaluation. Depending on the year, the GGE biplot explained 77% to 89% of the total yield variation due to G and GE (Fig. 1).

There are numerous ways to look at a GGE biplot, but the polygon view of a biplot is most relevant to the investigation of the mega-environment problem. The polygon was drawn on genotypes that located furthest from the plot origin (0,0) such that markers of all other cultivars are contained in the polygon. These vertex genotypes are the most responsive genotypes since they had the longest distance from the origin. The most responsive genotypes were those that were either the best or poorest performers at some or all of the locations. For each side of the polygon a perpendicular line is

drawn, starting from the origin of the biplot, such that the biplot is divided into sectors (quadrants) and markers of the sites fall into the same or different sectors. Because the positions of the genotypes on a biplot were unique, the perpendicular lines are also unique. The perpendicular lines actually serve as a facility to compare the two vertex genotypes connected by the respective polygon side. For example, in Fig. 1A, the perpendicular line to the side that connects genotypes T9310 and OAC9208 helps compare the two genotypes. Site Winchester (Fig. 1A) is near the perpendicular, meaning that the two genotypes yielded similarly at Winchester; site Exeter was on the side of genotype T9310, meaning that T9310 yielded better than OAC9208 at Exeter; and sites Woodstock and St Pauls were on the side of OAC9208, meaning that OAC9208 yielded better than T9310 at these two sites. Similarly, the perpendicular line to the polygon side connecting genotypes T9310 and K10116 helps compare these two genotypes and indicates that T9310 yielded better than K10116 at Exeter and all other sites. Thus, genotype T9310 yielded better than both OAC9208 and K10116. It also yielded better than all genotypes in the sector between the two perpendiculars flanking it, because it had the longer distance from the origin than any of these genotypes. Consequently, genotype T9310 yielded the best at Exeter in 1994. For the same reason, genotype OAC9208 yielded the best at St. Pauls and Woodstock in 1994.

As a rule, the vertex genotype in each sector is the best genotype at sites whose markers fall into the respective sector (Yan, 1999; Yan et al., 2000, 2001). Sites within the same sector share the same winning genotype, and sites in different sectors have different winning genotypes. Thus, the polygon view of a GGE biplot indicates the presence or absence of crossover GE interactions involving the most responsive genotypes, and is suggestive of the existence or absence of different mega-environments among the tested sites.

Figure 1A based on the 1994 data suggests two mega-environments, one represented by the Exeter, and the other by Woodstock and St. Pauls. This pattern, however, was not repeated in other years. In four of the six years, namely, 1995, 1997, 1998, and 1999, the eastern Ontario site Winchester was separated from the three southwestern sites. And in these four years, the three southwestern Ontario sites tended to group together. In all years except 1995, site St. Pauls grouped with at least one of the other three sites. These observations suggest two things. First, the eastern Ontario site Winchester may represent a mega-environment that is dif-

**Table 1. Variance components of soybean genotype main effect (G), site main effect (E), and genotype by site interaction (GE) in the 2800 CHU region license yield trials in 1994–1999.**

Year	Number of genotypes	Number of sites	Variance components			% of total variance		
			E	G	GE	E	G	GE
1994	90	4	0.018	0.043	0.027	20.2	49.1	30.3
1995	100	4	0.027	0.034	0.044	25.6	32.0	42.1
1996	90	3	0.042	0.039	0.045	33.1	30.8	36.0
1997	110	4	0.123	0.036	0.037	62.7	18.5	18.9
1998	125	3	0.737	0.041	0.039	90.3	5.0	4.8
1999	119	4	0.154	0.040	0.050	63.2	16.5	20.4

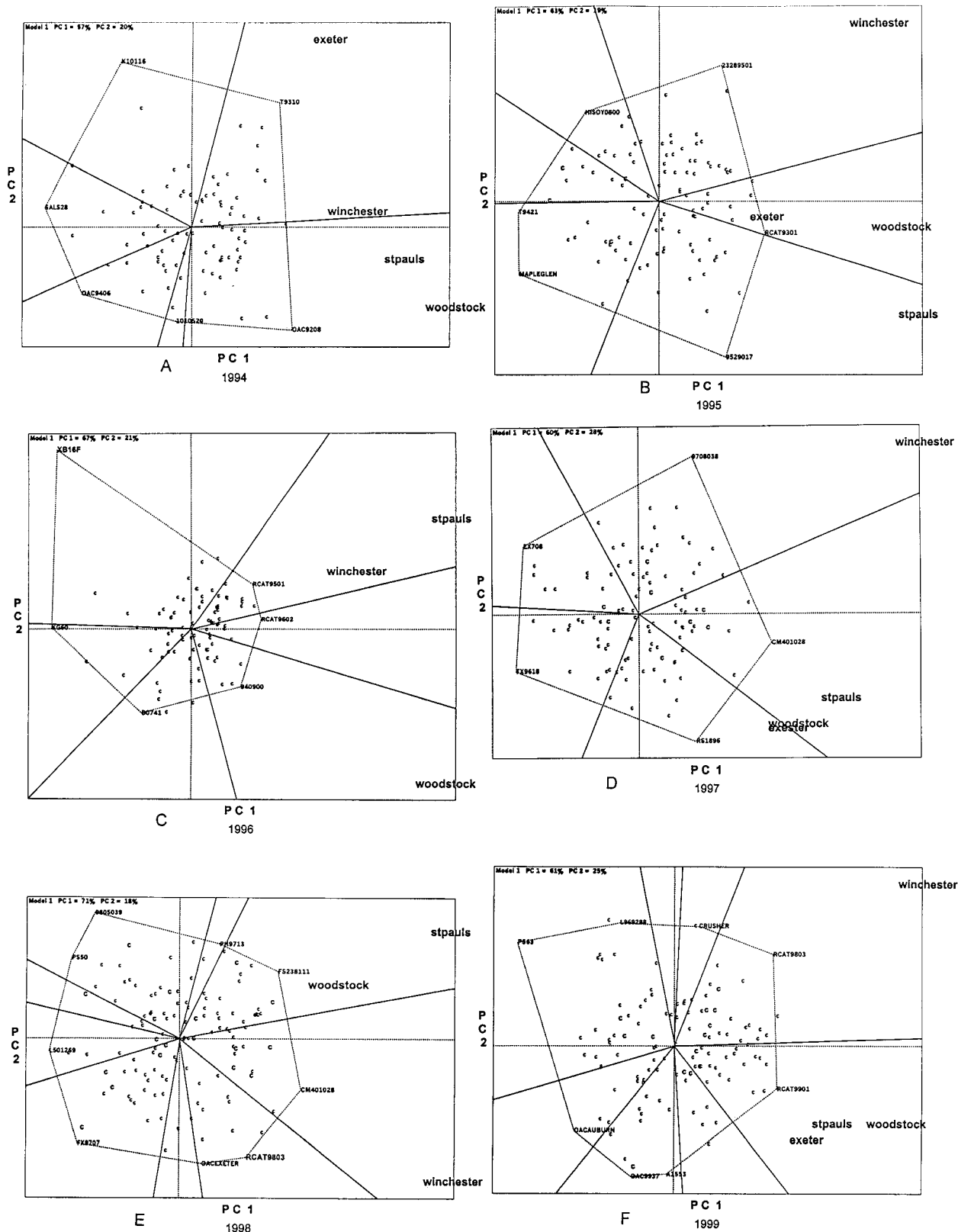
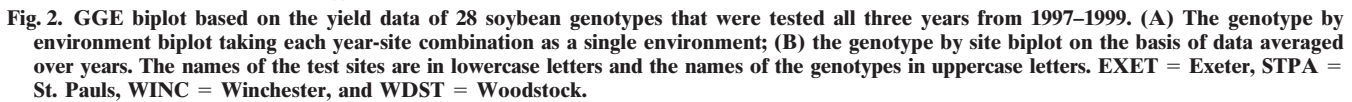


Fig. 1. Yearly GGE biplots for yield data from 1994–1999 soybean trials in the 2800 corn heat unit area of Ontario. The names of the test sites are spelled out in lowercase letters, with the first letter indicating the exact position of the site; the vertex genotypes are spelled out in uppercase letters, and all other genotypes are represented by “c”.





The GT biplot for each of the six years, based on Eq. [2], explained 52 to 63% the total variation of the standardized data (Fig. 3). This relatively low proportion reflects the complexity of the relationships among the measured traits. Nevertheless, the fundamental patterns among the traits should be captured by the biplots (Kroonenberg, 1995). In the GT biplot, a vector is drawn

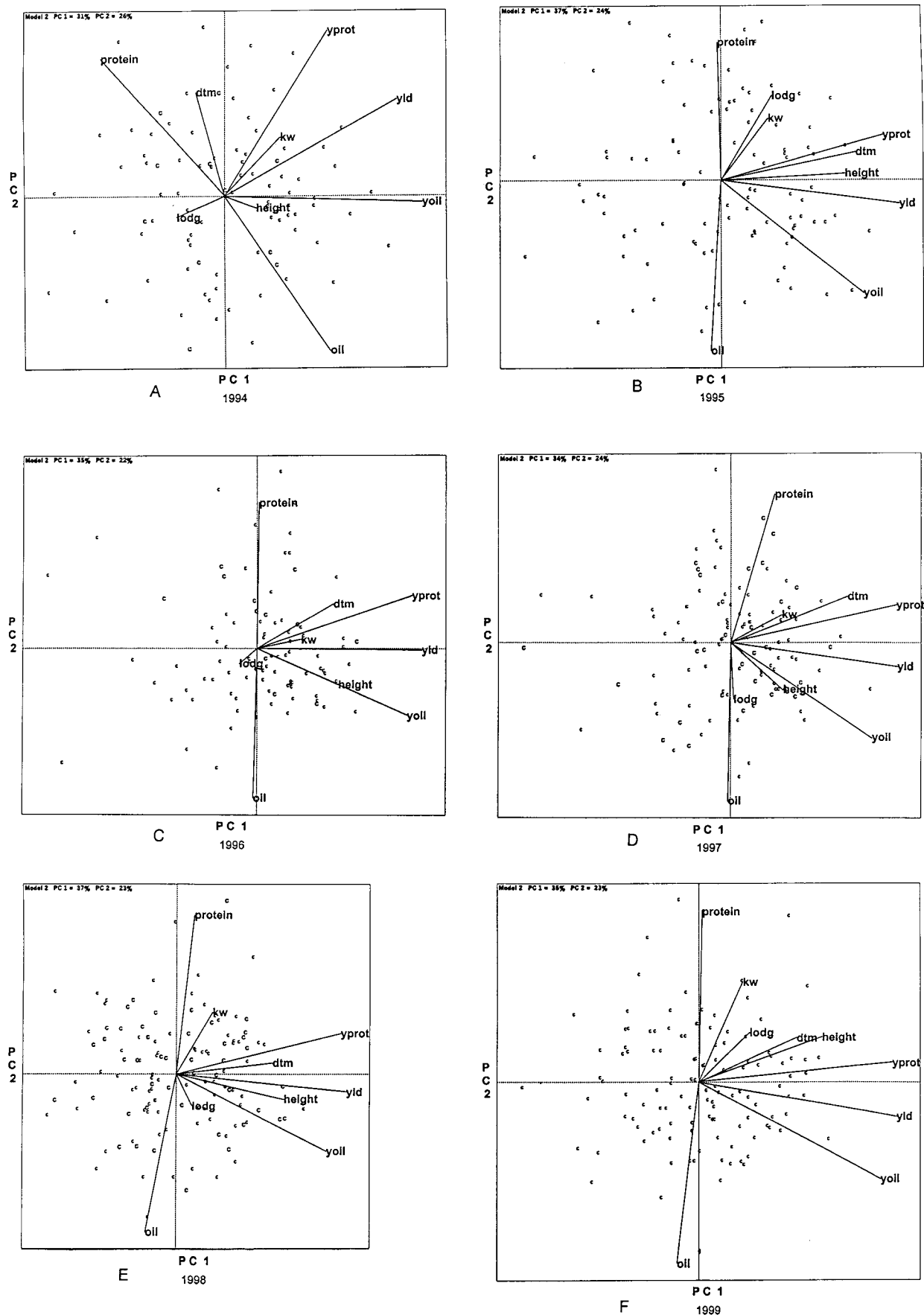


Fig. 3. The yearly soybean genotype by trait biplots of 1994–1999. The traits are spelled out in lowercase letters, and each genotype is represented by “c”. DTM = days to maturity, HEIGHT = Plant height, KW = 100 seed weight, LODG = lodging scores, OIL = percentage of oil, PROTEIN = percentage of protein, YLD = seed yield, YOIL = oil yield per unit land area, and YPROT = protein yield per unit land area.

**Table 2. Correlation coefficients among soybean traits based on the 1999 data.**

	% Oil	% Protein	Protein yield	Oil yield	Days to maturity	Lodging score	Height	Seed weight
Yield	-0.050	-0.077	0.949**	0.936**	0.260**	0.086	0.344**	0.080
% Oil		-0.690**	-0.265**	0.303**	-0.115	-0.084	-0.084	-0.274**
% Protein			0.240**	-0.316**	0.066	0.014	-0.005	0.177
Protein yield				0.812**	0.276**	0.084	0.333**	0.131
Oil yield					0.204*	0.053	0.293**	-0.019
Days to maturity						0.059	0.451**	0.181
Lodging score							0.519**	0.184*
Height								0.210*

\* Indicates significance at  $P < 0.05$ .\*\* Indicates significance at  $P < 0.01$ .

from the biplot origin to each marker of the traits to facilitate visualization of the relationships between and among the traits. Provided that the biplot explained a sufficient amount of the total variation, the correlation coefficient between any two traits is approximated by the cosine of the angle between their vectors. Thus,  $r = \cos 180^\circ = -1$ ,  $\cos 0^\circ = 1$ , and  $\cos 90^\circ = 0$ .

For all six years, the largest variation explained by the biplots came from seed yield, protein concentration, oil concentration, oil yield, and protein yield, as indicated by the relative length of their vectors. It is the interrelationships among these traits that are most relevant to soybean breeding. The most prominent relations revealed by these biplots are: (i) a strong negative association between protein concentration and oil concentration, as indicated by the large obtuse angles between their vectors, (ii) a near zero correlation between seed yield and protein concentration and between seed yield and oil concentration, as indicated by the near perpendicular vectors, and (iii) a positive association between protein yield and oil yield, both being closely correlated to seed yield, as indicated by the acute angles. Other relations revealed from the GT biplots include positive associations among seed yield, days to maturity, and plant height. Thus, the GT biplots graphically display the trait relations in soybean that are well-documented elsewhere (Burton, 1991). The correlation coefficients among the traits for 1999 indicate that the GT biplot correctly displays relationships among the traits that had relatively large loadings on either PC1 or PC2 (Table 2). Exact match is not to be expected however, because the biplot describes the interrelationships among all traits on the basis of overall pattern of the data whereas correlation coefficients only describe the relationships between two traits.

### Genotype by Trait Biplot as a Tool for Genotype Comparisons

Figure 4 is an enhanced version of Fig. 3F to demonstrate that the GT biplot can be used to compare genotypes on the basis of multiple traits and to identify genotypes that are particularly good in certain aspects and therefore can be candidates for parents in soybean breeding. Analogous to the analysis of the GGE biplots in Fig. 1 and 2, a polygon was drawn on the GT biplot. The perpendicular lines to the polygon sides facilitate comparison between neighboring vertex genotypes. Specifically, comparison between RCAT9901 and ps78 indi-

cates that RCAT9901 was better in oil concentration, oil yield, seed yield, and protein yield, whereas ps78 was better in protein concentration and seed weight, and was taller, later and more prone to lodge. Similarly, RCAT9901 had a greater value than RCAT9801 in all traits except oil concentration. ADV57151 was slightly better than PS78 in oil and protein content but had a lower value than PS78 on all other traits; PS63 had higher oil concentration but lower values on all other traits than ADV57151. Also analogous to the interpretation of the GGE biplot, Fig. 4 indicates that genotype RCAT9801 was highest in oil concentration, ADV57151 was highest in protein concentration, RCAT9901 was highest in seed yield, oil yield, and protein yield, whereas PS78 was the tallest, the latest, most prone to lodging, and had the largest seed weight. The measured trait values of these vertex genotypes are presented in Table 3 to validate the statements on the basis of the GT biplot. A GT biplot may not accurately reflect the means as it did not explain all variation of the data but it displays the most important patterns of the data.

The GT biplot can also be used to aid genotype selection on the basis of multiple traits. Fig. 5A demonstrates the selection results based on seed yield by culling genotypes that yielded below average. This was done by drawing a line that passes through the biplot origin and the marker of seed yield, followed by drawing a line that passes through the biplot origin and is perpendicular to the seed yield line. On the basis of biplot theory, if the biplot sufficiently approximates the data, then genotypes that fell on the same side of the perpendicular line as seed yield should have yielded above average, whereas genotypes on the other side of the perpendicular line should have yielded below average. Fig. 5B presents the results of culling genotypes that either had below-average oil concentration or had below-average protein concentration. This selection scheme is obviously not a viable selection strategy since only a single genotype was left and it had a below average yield. Fig. 5C presents the results from independent culling on oil concentration and seed yield, in which all genotypes that had relatively high protein concentration were discarded. Fig. 5D presents the results from independent culling on protein concentration and seed yield, in which all genotypes that had relatively high oil concentration were discarded. Fig. 5E presents the results from independent culling on oil yield and protein yield. Fig. 5E is almost equivalent to putting results of Fig. 5C and

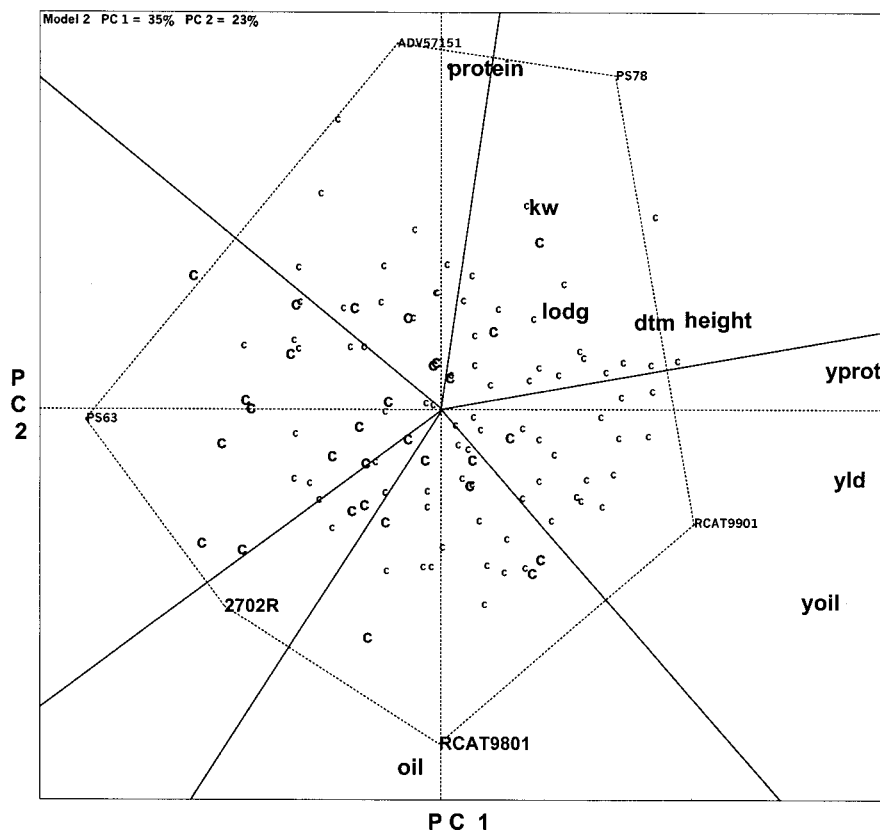


Fig. 4. The polygon view of the genotype by trait biplot for 1999, demonstrating soybean genotype comparison on the basis of a GT biplot. DTM = days to maturity, HEIGHT = Plant height, KW = 100 seed weight, LODG = lodging scores, OIL = percentage of oil, PROTEIN = percentage of protein, YLD = seed yield, YOIL = oil yield per unit land area, and YPROT = protein yield per unit land area.

Fig. 5D together and is almost identical to selecting on the basis of yield alone (Fig. 5A). Therefore, selection on the basis of seed yield alone is the simplest and most efficient strategy of selection, particularly in the early stages of soybean breeding. Although oil and protein concentration are important aspects of soybean quality, selection on these traits could be deferred until a later stage.

## CONCLUSIONS

Applying the GGE biplot technique to the 1994 to 1999 soybean trials in the 2800 CHU area of Ontario led to the following conclusions. First, the GE interaction was an important source of soybean yield variation and there were clear crossover GE interactions each year. Second, the eastern Ontario site Winchester was frequently different from the three southwestern Ontario sites in differentiating soybean genotypes. This difference, however, was not large enough to define a

different mega-environment. Third, crossover GE interactions were rare among the three southwestern sites. Particularly, site St. Pauls seldom provided unique information on genotype performance and seemed to be a redundant test site. On the basis of these findings, the 2800 CHU area is a single mega-environment with frequent crossover GE interactions. For reliable genotype evaluation, sites from both eastern (Winchester) and southwestern Ontario (Woodstock and Exeter) are needed, but the St. Pauls site appears to be unnecessary.

This study demonstrates that the GT biplot is an excellent tool for visualizing genotype by trait data. First, it effectively reveals the interrelationships among the soybean traits. Second, it provides a tool for visual comparison among genotypes on the basis of multiple traits. Third, it can be used in independent culling based on multiple traits and in comparing selection strategies. Based on the trait relationships, selection for seed yield alone is not only the simplest, but also the most effective strategy in the early stages of soybean breeding.

Table 3. Measured values of various traits for selected soybean genotypes tested in 1999.

Cultivars	Seed Yield	Protein yield	Oil yield	Days to maturity	Protein	Oil	Lodging score	Height	100 Seed weight
	kg ha <sup>-1</sup>				%			cm	g
RCAT9801	3338	1288	741	116	38.6	22.2	1.5	81	18.2
RCAT9901	3662	1472	787	123	40.2	21.5	2.2	94	20.0
ADV57151	3001	1323	594	121	44.1	19.8	2.2	85	21.6
PS78	3344	1461	655	128	43.7	19.6	2.0	102	19.5
PS63	2468	1000	523	120	40.5	21.2	1.5	82	17.2
2702R	2844	1146	631	116	40.3	22.2	1.6	77	17.2



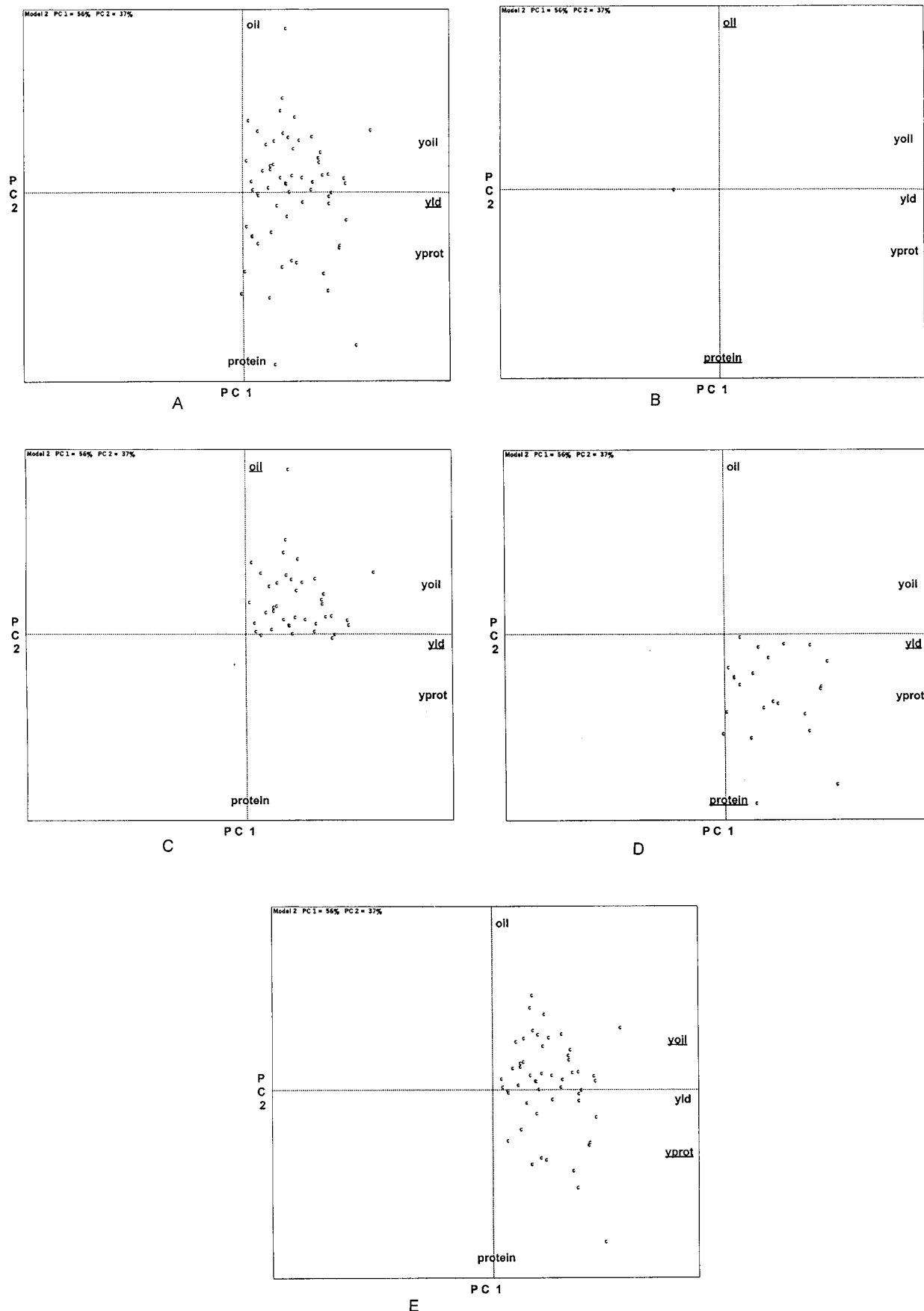


Fig. 5. Soybean Genotype selection on the basis of a GT biplot. (A) Selection on the basis of seed yield; (B) selection on the basis of oil and protein concentration; (C) selection on the basis of oil concentration and seed yield; (D) selection on the basis of protein concentration and seed yield; and (E) selection on the basis of oil yield and protein yield. DTM = days to maturity, HEIGHT = Plant height, KW = 100 seed weight, LODG = lodging scores, OIL = percentage of oil, PROTEIN = percentage of protein, YLD = seed yield, YOIL = oil yield per unit land area, and YPROT = protein yield per unit land area.

## ACKNOWLEDGMENTS

We would like to acknowledge Kemptville and Ridgetown Colleges at the University of Guelph, the Eastern Cereal and Oilseed Crop Research Centre of the Agriculture and Agri-Food Canada (AAFC) in Ottawa, ON, and the Greenhouse and Processing Crop Research Centre of AAFC at Harrow, ON, for providing the soybean trial data. We also acknowledge the financial contribution to this research, in part, by the Ontario Ministry of Agriculture, Food, and Rural Affairs and by the Ontario Soybean Growers.

## REFERENCES

- Burton, J.W. 1991. Development of high-yielding high-protein soybean germplasm. p. 109–117. *In* R.F. Wilson (ed.) *Designing value-added soybeans for the markets of the future*. American Oil Chemists Soc., Champaign, IL.
- Cooper, M., and G.L. Hammer. 1996. *Plant Adaptation and Crop Improvement*. CAB Int., Wallingford, Oxon, UK.
- Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58:453–467.
- Kang, M.S. 1990. Genotype-by-environment interaction and plant breeding. Dept. of Agronomy, Louisiana State Univ., Baton Rouge, LA.
- Kang, M.S., and H.G. Gauch. 1996. *Genotype-by-environment interaction*. CRC Press, Boca Raton, FL.
- Kroonenberg, P.M. 1995. Introduction to biplots for  $G \times E$  tables. Department of Mathematics, Research Report #51, University of Queensland.
- Ontario Ministry of Agriculture, Food and Rural Affairs. 1993. Crop Heat Units for Corn and Other Warm-Season Crops in Ontario [Online], available at: <http://www.gov.on.ca/OMAFRA/english/crops/facts/93-119.htm>. [modified 3 July 2001; verified 19 July, 2001]. Government of Ontario, ON, Canada.
- Yan, W. 1999. Methodology of cultivar evaluation based on yield trial data—with special reference to winter wheat in Ontario. Ph.D Thesis, University of Guelph, Guelph, ON, Canada.
- Yan, W., and D.H. Wallace. 1995. Breeding for negatively associated traits. *Plant Breed. Rev.* 13:141–177.
- Yan, W., and L.A. Hunt. 1998. Genotype by environment interaction and crop yield. *Plant Breed. Rev.* 16:135–178.
- Yan, W., and L.A. Hunt. 2001. Genetic and environmental causes of genotype by environment interaction for winter wheat yield in Ontario. *Crop Sci.* 41:19–25.
- Yan, W., L.A. Hunt, Q. Sheng, and Z. Szlavniks. 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 multi-environment trial data. *Crop Sci.* 41:656–663.