

Review

# Application of the GGE Biplot as a Statistical Tool in the Breeding and Testing of Early and Extra-Early Maturing Maize in Sub-Saharan Africa

**Baffour Badu-Apraku**<sup>1,\*</sup>, **Bamidele Fakorede**<sup>2</sup>, **Richard Akinwale**<sup>2</sup>,  
**Benjamin Annor**<sup>3</sup>, **Samuel Adewale**<sup>1</sup>, **Johnson Toyinbo**<sup>1</sup>, **Samson Akintibu**<sup>2</sup>

<sup>1</sup> International Institute of Tropical Agriculture (IITA), PMB 5320, Oyo Rd, Ibadan 200285, Nigeria

<sup>2</sup> Department of Crop Production and Protection, Obafemi Awolowo University, Ile-Ife 220005, Nigeria

<sup>3</sup> CSIR-Crops Research Institute, P. O. Box 3785, Fumesua, Kumasi AE-0607-0028, Ghana

\* Correspondence: Baffour Badu-Apraku, Email: b.badu-apraku@cgiar.org; Tel.: +234-2-7517472.

## ABSTRACT

In this paper is reviewed some aspects of the research conducted in sub-Saharan Africa in which the genotype main effect plus genotype by environment interaction (GGE) biplot was employed for the analysis and interpretation of the data. GGE biplot has been found quite effective in analyzing genotype  $\times$  environment interaction, genotype  $\times$  trait (GT) interaction, interpretation of diallel and line  $\times$  tester data, and evaluation of the efficiency of testers in hybrid production. Application of GGE biplot to genotype by environment data from several studies has helped to identify outstanding varieties, inbreds and hybrids of early and extra-early maize in terms of yield performance and stability under stress and non-stress environments. The use of GT biplot analysis has resulted in the identification of ear aspect (EASP), plant aspect (PASP), anthesis-silking interval (ASI), and number of ears per plant (EPP) as the most reliable traits for selection for yield under drought, low-N, high-N and well-watered environments. Studies comparing GT with path-coefficient analyses revealed that both methods identified EASP, plant height (PLHT), and ASI as the most important traits directly contributing to yield under drought stress. GT biplot identified EASP, EPP, and *Striga* damage as the most reliable traits for indirect selection for improved grain yield under *Striga* infestation. The biplot graphical analysis allowed visual display of the general combining ability (GCA) of the parental inbreds and specific combining ability (SCA) of the hybrids used in Griffings diallel mating design. In addition, information on the best mating partners, identification of proven testers and tester groups, and heterotic groups have been provided graphically. The disadvantages of the GGE biplot include limited number of entries, only two heterotic groups are handled by the method, and only fixed statistical model can be used. More attention needs to be focused on test of hypothesis and QTL analyses.

## Open Access

Received: 26 February 2020

Accepted: 01 June 2020

Published: 11 June 2020

Copyright © 2020 by the author(s). Licensee Hapres, London, United Kingdom. This is an open access article distributed under the terms and conditions of [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

**KEYWORDS:** biplot; genotype; genotype by environment interaction; maize; stability; sub-Saharan Africa

---

## INTRODUCTION

The climatic, edaphic, and management variability in sub-Saharan Africa (SSA) is great and too formidable to be dealt with ordinarily. For example, the soil variability goes, as it were, from foot-to-foot and it is necessary that crops be able to cope with the variation. Similarly, crop plants vary a great deal in their response to environmental conditions. Therefore, genotype  $\times$  environment interaction (GEI) has been defined as the degree of variation in response of a genotype across environments [1]. Genotype  $\times$  environment interaction has been measured more by variation in grain yield than any other trait in crop plants and the response of other plant parts is poorly understood. Generally, GEI is of two dimensions: magnitudinal and directional. In magnitudinal dimension, the relative response of one variety to the other remains the same despite the changes in the environment. In directional dimension, on the other hand, one variety may be higher yielding than the other in some environments and poorer yielding in others. It takes comparison of the two varieties to detect this, a definite limitation of most stability analytical systems.

The information obtained from multi-environment trials (METs) can be of high importance in Plant Breeding, Genetics and Agronomy by providing answers to three main objectives:

- a. accurately calculate and predict performance of a genotype from limited data;
- b. examine yield of genotypes and their stability patterns in response to the environments;
- c. provide guidance that are very reliable in selecting the best genotypes or agronomic treatments suitable for planting at new locations and in the coming years [2].

Observable uniqueness in ensuring the interaction between the environment and the genetic make-up is called phenotype. Phenotypes could be assessed, observed, estimated, and arranged in groups according to features that they have in common. Environmental factors may be regarded as locations, growing seasons, years, nitrogen levels, rainfall, temperature, all of which could have positive or negative effects on genotypes [3]. Wu and O'Malley [4] described two classes of environments with detailed differences in their information: micro-environmental differences that cannot clearly be forecast such as yearly differences in drought conditions, rainfall, and level of insect damage; and macro-environmental differences which can be forecast, such as category of soil, restricted temperatures and management put into

practice. According to the authors, the  $G \times E$  variance can only be projected for the macro-environmental state.

There are difficulties in determining the varietal performance evaluated in experiments containing genotypes (G), locations (L), and years (Y) due to the genotype  $\times$  location  $\times$  year ( $G \times L \times Y$ ) interactions not being easy to classify [5]. The complications resulting from  $G \times E$  interactions can best be avoided by identification of stable genotypes that are adapted across crop production environments. To ensure the maintenance of broader adaptation and yield stability, superior experimental varieties have been selected based on the performance across contrasting environments. For example, the Regional Drought Tolerant Early Trials are used as the vehicle for testing, identifying and exchanging drought tolerant early (90–95 days) and extra-early (80–85 days) cultivars and hybrids with broad adaptation to the agro-ecological zones of the four partner countries, Mali, Nigeria, Ghana and Benin in West Africa (WA). Based on the results of the Regional Trials, outstanding varieties are selected and evaluated in farmer participatory on-farm trials and demonstrations which are important vehicles for demonstrating the effectiveness of new technology to farmers and thus are instrumental in the identification, release and commercialization of stress tolerant maize varieties and hybrids in the partner countries. The trials also offer the National Agricultural Research Systems (NARS) partners the opportunity for selecting promising cultivars for introgression of favorable alleles into the breeding populations of national maize programs for broadening the genetic base.

Results of METs in WA have demonstrated the existence of GEI [6–10], emphasizing the need for extensive testing of cultivars in multiple environments over years before decisions are taken on cultivar recommendations. However, because of the limited resources of the national maize research programs of WA, there has been a need to conduct cultivar evaluations in a limited number of environments. To achieve this, we re-examined target testing environments in WA for their uniqueness as it was believed that some environments could never provide unique information, because of similarity to some other environments in separating and ranking genotypes without losing valuable information on genotypes. Furthermore, it was felt that stratification of maize evaluation environments could help improve heritability of measured traits, accelerate the rate of genetic gain from selection, and strengthen the potential competitiveness for seed production and maximize grain yields of farmers [11]. It was therefore very important to develop an in-depth understanding of the target agro-ecologies used for the evaluation of drought tolerant cultivars in WA and to determine if it could be subdivided into different mega-environments to facilitate a more meaningful cultivar evaluation and recommendation. It was believed that selected locations for METs should constitute a

sample of environments that adequately cover the range of environmental conditions of the target geographical region.

Furthermore, following the ratification of the protocol on seeds by the Heads of States of the Economic Community of West African States (ECOWAS) member countries in 2009, the West African Catalogue of Plant Species and Varieties (COAFEV) was made available in the sub-region [12]. The seed catalogue contains the list of varieties whose seeds could be produced and commercialized within the territories of the 17 member countries of ECOWAS and is an aggregate of the varieties registered in the national catalogues of the Member States. The catalogue offers a unique opportunity for the movement of good quality seeds of improved maize varieties and hybrids across the borders of the ECOWAS countries for production and marketing. As a result of these new developments and the implications of global warming, desertification, and recurrent drought in the sub-region, there was a need for a re-examination of the mega-environments in WA and the identification of core testing locations in each of the mega-environments in WA used for the evaluation of the three different regional trials in WA. A number of studies was therefore conducted to determine the representativeness, discriminating ability and the repeatability of the test locations used for the evaluation of the DT Regional Early Variety Trials and to identify core testing sites to facilitate testing, seed production and commercialization of drought tolerant cultivars in WA. Therefore, using the GGE Biplot statistical tool, Badu-Apraku et al. [13] examined the mega-environments in WA employed for testing the Regional extra-early maturing varieties. The test locations Zaria, Ilorin, Ikenne, Ejura, Kita, Babile, Ina, and Angaredebou were identified as the core testing sites of the three mega-environments for testing the Regional Uniform Variety Trials-Extra-early. In another study, involving the testing sites for the Regional Early Trials, test environments were classified into four mega-environments [14]. Four test locations were highly correlated in their ranking of the genotypes in group 1, suggesting that a promising early maturing cultivar selected in one of these locations in one country will also be suitable for production in the other locations within the same mega-environments in different countries [14]. Similarly, eight test locations were highly correlated in their rankings of the genotypes in group 2 and therefore, a promising cultivar identified in one of these locations were likely to be adapted to the other locations. It was concluded that selecting a cultivar out of these two locations would likely result in varieties adapted to other locations within the same mega-environment. The identification of the core testing sites was expected to facilitate the selection of high yielding and stable cultivars in the four different regional trials of WA [Regional Uniform Variety Trial (RUVT)-early, RUVT-extra-early, Drought Tolerant (DT) Regional Early and the DT Regional Extra-early variety Trials] and seed production and marketing across the countries of WA. The selection of suitable breeding and test locations are crucial to the success of a

maize breeding program. A test location must be discriminating so that genetic differences among genotypes can be easily detected. It is important therefore that the test locations are representative of the target environments so that selected genotypes would have the desired adaptation and would be representative of the target environment as well as repeatable so that genotypes selected from year to year will have superior performance. According to Yan et al. [15], only test locations with high discriminating ability were useful and only those that were also representative could be used in selecting superior genotypes. The repeatability of genotype ranking across years within test locations was also an essential aspect in test location evaluation. Using the GGE Biplot method, the GEI of the testing sites of the RUVT early and extra-early varieties in West and Central Africa (WCA) were studied and the test locations characterized and stratified into mega-environments and core testing sites to facilitate efficient and less costly testing of varieties [13,14]. On the other hand, the testing sites of the Regional Drought Tolerant Trials which were confined to the drought-prone locations in the four partner countries of the Drought Tolerant Maize for Africa (DTMA) project, namely, Nigeria, Ghana, Benin and Mali had not been studied. Therefore, it was believed that information on the representativeness, discriminating ability and repeatability of the testing sites of the DT Regional Variety Trials in WA would facilitate better understanding of the responses of drought tolerant maize genotypes in target drought environments and would be invaluable in designing an efficient and economic selection strategy for the International Institute of Tropical Agriculture (IITA) Maize Breeding Program. However, there was limited information on the representativeness, discriminating ability and repeatability of the testing sites of the Regional DT Trials which were largely in the drought prone locations in the four partner countries of the DTMA project, Nigeria, Ghana, Benin and Mali (Table 1). Therefore, twelve early maturing maize cultivars were evaluated for 3 years at 16 locations in WA to determine the representativeness, discriminating ability and the repeatability of the testing sites and to identify core testing sites using the GGE biplot method [16]. The results revealed that Zaria (Nigeria), Nyankpala (Ghana) and Ejura (Ghana) displayed the highest discriminating ability (Figure not shown). Two mega-environments were identified. Bagou, Nyankpala, Bagauda, Ikenne, and Mokwa constituted the first mega-environment (ME1) while Ejura, Ina and Sotuba represented the second (ME2). It was concluded that the first mega-environment would be more useful for evaluating early maize genotypes for tolerance to drought than the ME2 because locations in this mega-environment were more strongly correlated to Ikenne (managed drought stress site). Among the test locations, Bagou and Mokwa were found to be closely related to Ikenne in their ranking of the cultivars while Zaria was the exact opposite, indicating that Zaria was the least

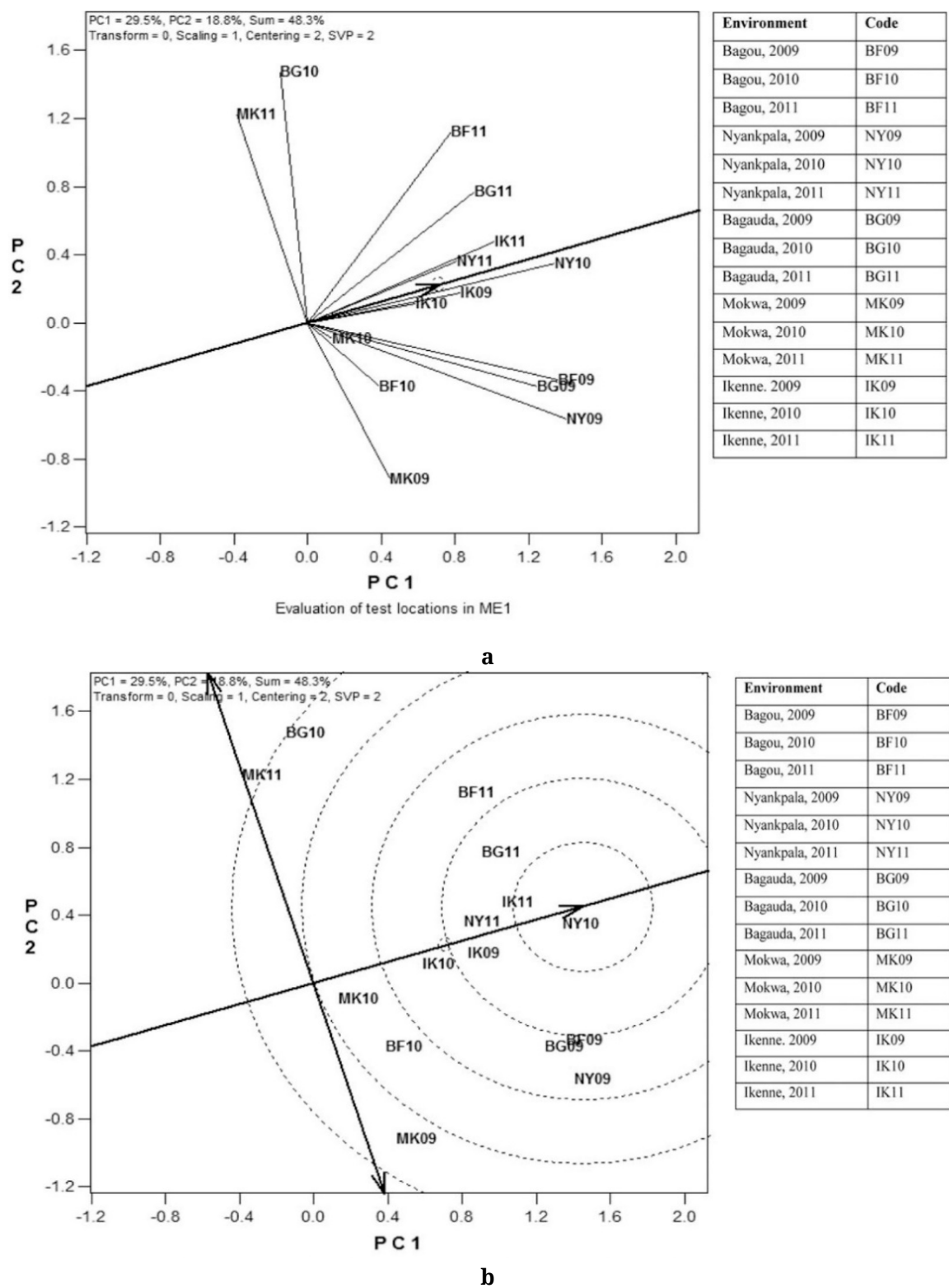
**Table 1.** Description of test locations used for evaluation of Regional Early-maturing drought tolerant (DT) maize in west Africa between 2009 and 2011.

Country	Location	Code	Agro-ecology <sup>†</sup>	Latitude	Longitude	Altitude (mm asl)	Rainfall during the growing season (mm)	Years when evaluation was conducted <sup>‡</sup>		
								2009	2010	2011
Benin	Angaradebou	AN	SS	11°32' N	3°05' W	297	1000	x	x	x
Benin	Bagou	BF	SS	11°28' N	2°23' W	303	1125	x	x	x
Benin	Ina	IN	NGS	9°58' N	2°44' W	358	900	x	x	x
Ghana	Ejura	EJ	FT	7°38' N	1°37' E	90	1460	x	x	
Ghana	Manga	MG	SS	11°01' N	0°16' W	270	718	x	x	x
Ghana	Nyankpala	NY	NGS	9°25' N	0°58' E	340	800	x	x	x
Nigeria	Badeggi	BD	SGS	9°05' N	6°15' E	118	1124	x	x	
Nigeria	Bagauda	BG	SS	12°01' N	8°19' E	520	840	x	x	x
Mali	Katibougou	KB	SS	12°50' N	8°09' W	285	700	x	x	x
Nigeria	Sabongari	SG	NGS	12°3' N	8°32' E	476	720	x	x	
Nigeria	Ikenne	IK	FT	6°53' N	3°42' E	60	1200	x	x	x
Mali	Sotuba	ST	NGS	12°39' N	7°55' W	337	740	x	x	x
Nigeria	Jibia	JB	NGS	13°05' N	7°13' E	607	600	x	x	
Nigeria	Mokwa	MK	SGS	9°18' N	5°4' E	457	1100	x	x	x
Nigeria	Samaru	SA	NGS	12°12' N	7°37' E	550	1040	x	x	
Nigeria	Zaria	ZA	NGS	12°00' N	8°22' E	640	1120	x	x	x

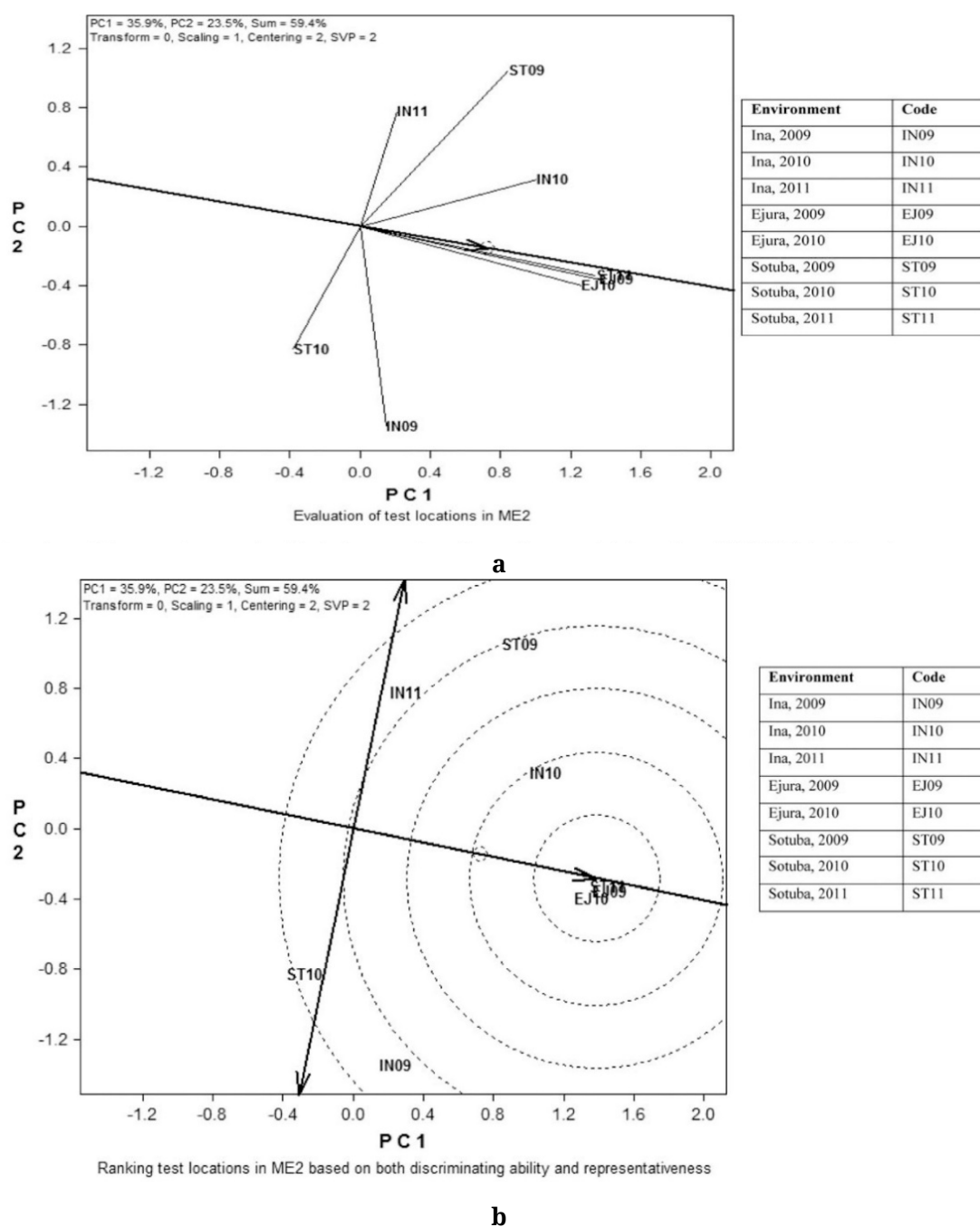
<sup>†</sup> SGS = southern Guinea savanna; NGS = northern Guinea savanna; FT = Forest-savanna Transitional zone; SS = Sudan savannah;

<sup>‡</sup> “x” marks the year when evaluation was carried out at the location.





**Figure 1.** Vector view of the GGE biplot showing (a) representativeness and repeatability of test locations in ME 1 (b) the ideal test environments in ME 1 based on their discriminating power and representativeness.



**Figure 2.** Vector view of the GGE biplot showing (a) representativeness and repeatability of test locations in ME 2 (b) the ideal test environments in ME2 based on their discriminating power and representativeness.

suitable for evaluating genotypes for drought tolerance (Figure not shown). Nyankpala and Ikenne were identified as the core testing site for ME1 (Figure 1b) and Ejura for ME2 (Figure 2b). Ikenne, Nyankpala and Ejura had moderately high repeatability and were closer to the average environment axis of each mega-environment (Figures 1a and 2a) and would be useful for culling unstable genotypes at the multilocation testing stage. It was concluded that other sites were less representative



and not repeatable and would not be useful for evaluating early maize cultivars for drought tolerance.

Beyond analysis of a MET data where significant GEI is singularly partitioned into eigen values in principal component analysis to obtain information on stability and adaptability of genotypes, as well as discriminativeness and representativeness of the environments, it is interesting to note that GGE biplot is very appropriate for the analysis of any other data that can cast into a 2-way table. This facilitates the use of GGE biplot in graphical analysis of traits relationship and genetic data obtained from factorial mating designs as well as QTL studies [17].

In the rest of this paper, we address the issue of stability of performance by comparing varieties, using different statistical methods. Specifically, our objectives are to (i) compare GGE biplot with other statistical analytic methods, (ii) determine the effect of genotype  $\times$  trait interaction, (iii) test GGE biplot for analysis of genetic data using diallel and line  $\times$  tester designs, (iv) evaluate the efficiency of testers in hybrid production, (v) discuss the strengths and weaknesses of the GGE biplot statistical tool, and (vi) give future directions.

### GGE BILOT COMPARED WITH OTHER ANALYTICAL METHODS

Stability studies have allowed researchers to identify broadly adapted cultivars for use in breeding programs and have been helpful in recommending new varieties to farmers [18]. Different concepts leading to different definitions of stability have been proposed over the years [19,20]. Lin et al. [19] identified three types of stability concepts:

Type 1: A genotype is considered stable if it is characterized by small among-environment variance. Becker and Léon [20] referred to this as static, or biological concept of stability. A stable genotype possesses an unchanged performance regardless of any variation of the environmental conditions. This concept of stability is useful for qualitative traits, disease resistance, or for abiotic stress characters such as drought and winter hardiness. Parameters used to quantify this type of stability are  $CV_i$  [21] and genotypic variance across environments ( $S_i^2$ ).

Type 2: A genotype is considered stable if its response to environments is parallel to the mean response of all genotypes in the trial. Becker and Léon [20] referred to this as the dynamic or agronomic concept of stability. A stable genotype has no deviations from the general response to environments and thus permits a predictable response to environments. A regression coefficient ( $b_i$ ) [22] and [23] stability variance ( $\sigma^2_i$ ) can be used to measure Type 2 stability.

Type 3: A genotype is considered stable if the residual mean square from its regression model on the environmental index is small. The environmental index is the mean yield of all the genotypes in each location minus the grand mean of all the genotypes in all locations. Type 3 is closely related to the dynamic or agronomic stability concept [19].

The methods proposed by [24]) and [25] are examples of Type 3 concept. Becker and Léon [20] stated that all stability procedures based on quantifying GEI effects belong to the dynamic concept. This includes the procedures for partitioning the GEI of Wricke's [26] ecovalence and Shukla's [23] stability of variance, procedures using the regression approach such as those proposed by Finlay and Wilkinson [22], Eberhart and Russell [24], and Perkins and Jinks [25] as well as non-parametric stability analyses such as rank summation index.

Lin and Binns [19] proposed Type 4 stability concept based on predictable and unpredictable non-genetic variation. The predictable component relates to locations while the unpredictable component relates to years. These researchers suggested the use of a regression approach for the predictable portion and the mean squares for years  $\times$  location interaction for each genotype as a measure of the unpredictable variation.

The procedure involving combined analysis of variance is the earliest and the most used analysis method to measure the existence of GEI from METs with replicates. In recent times, however, a wide range of methods have been proposed to study GEI that were broadly divided into four groups: analysis of variance, stability or parametric, qualitative or non-parametric, and multivariate methods. We will consider three multiplicative methods here; that is, cluster analysis, additive main effect and multiplicative interaction (AMMI), and genotype and genotype by environment interaction (GGE) effects.

**Cluster Analysis.** Cluster analysis is a numerical classification technique that defines groups of similar individuals. There are two types of classification. The first is non-hierarchical classification, which assigns each item to a class. The second type is hierarchical classification, which groups the individuals into clusters and arranges these into hierarchies for the purpose of studying relationships in the data. Comprehensive reviews of the applications of cluster analysis to study GEI can be found in [19]. The report from cluster analyses by Shaibu et al. [27] revealed the genetic diversity among the genotypes and identified genotypes that can be selected for hybridization and improvement of maize.

**Additive Main Effects and Multiplicative Interaction (AMMI).** Stability methods have been used in both univariate and multivariate statistics [19]. Among the multivariate methods, the additive main effects and multiplicative interaction (AMMI) analysis are widely used for GEI investigations. This method has been effective because it captures a large portion of the GEI sum of squares, clearly separating the main and interaction effects, and often provides meaningful interpretation of data to support a breeding program [2]. The AMMI model combines ANOVA for the genotype and environment main effects with Principal Components Analysis of GEI [28,29]. Therefore, based on the AMMI model (IPCA1 and IPCA2) the AMMI stability value (ASV) has been used

[30]. The ASV is comparable with the methods used by Shukla [23] and Eberhart and Russell [24] for genotype stability [30].

The AMMI method can be used more effectively to analyze METs than ANOVA and PCA. According to Zobel et al. [28], ANOVA fails to detect a significant interaction component, PCA fails to identify and separate the significant genotype and environment main effects, while linear regression models account for only a small portion of the interaction sum of squares. The AMMI method takes care of the flaws in these methods and is used for three main purposes:

- a. The model diagnosis. AMMI is more appropriate in the initial statistical analysis of yield trials, because it provides an analytical tool of diagnosing other models as subcases when these are better for particular data sets [31].
- b. AMMI clarifies the GEI by summarizing patterns and relationships of genotypes and environments [2,28].
- c. It improves the accuracy of yield estimates. Gains have been obtained in the accuracy of yield estimates that are equivalent to increasing the number of replicates by a factor of two to five [28]. Such gains may be used to reduce testing cost by reducing the number of replications, increasing the number of treatments (e.g., varieties) in the experiments, or improving efficiency in selecting the best genotypes. It has proven useful for understanding complex GEI. The results can be graphed in a useful biplot that shows both main and interaction effects for both the genotypes and environments.

AMMI combines ANOVA into a single model with additive and multiplicative parameters. The model equation is:

$$Y_{ij} - Y_j = \lambda_1 E_{i1} Y_{j1} + \lambda_2 E_{i2} Y_{j2} + \varepsilon_{ij} \quad (1)$$

where  $Y_{ij}$  is the measured mean of  $i$ th genotype in  $j$ th environment;  $Y_j$  is the grand mean;  $\lambda_1$  and  $\lambda_2$  are the singular values for PC1 and PC2;  $E_{i1}$  and  $E_{i2}$  are the PC1 and PC2 scores for genotype  $i$ ;  $Y_{j1}$  and  $Y_{j2}$  are the PC1 and PC2 scores for environment  $j$  and  $\varepsilon_{ij}$  is the error term.

The combination of ANOVA and PCA in the AMMI model, along with prediction assessment, is a valuable approach for understanding GEI and obtaining better yield estimates. The interaction is explained in the form of a biplot display where PCA scores are plotted against each other thereby providing a visual inspection and interpretation of the GEI components. Integrating biplot display and genotypic stability statistics enables genotypes to be grouped based on similarity of performance across diverse environments. Yield-stability statistic (YSi) was also used to recommend varieties for commercialization [32]. Kang [32] proposed an improved superior stability index (I) that is free from all the aforesaid drawbacks. A new approach, known as genotype selection index (GSI), was used by taking into consideration the AMMI stability value and mean yield for quantification of stability [33].

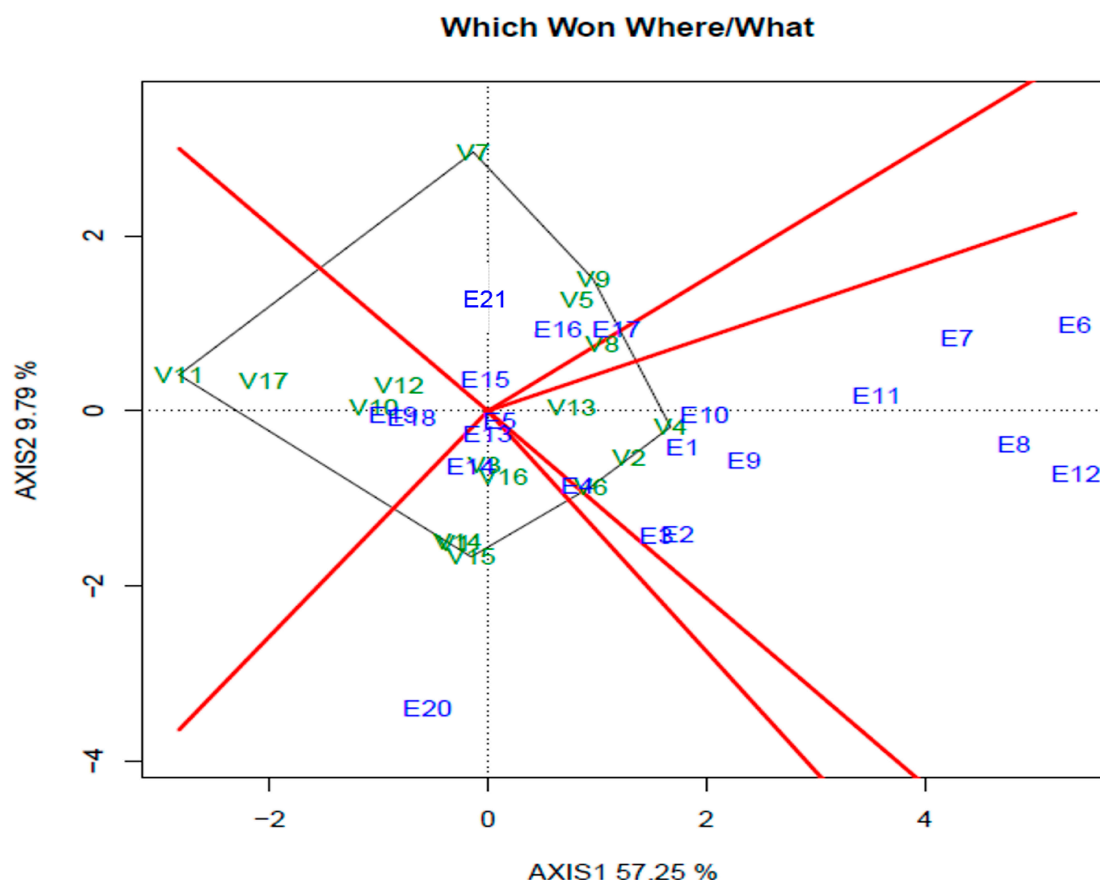
### GENOTYPE AND GENOTYPE BY ENVIRONMENT INTERACTION (GGE)

Yan et al. [34] proposed a methodology known as GGE biplot for graphical display of GEI patterns. It allows visual examination of the relationships among test environments, genotypes and GEI. It is an effective tool for: (i) mega-environment analysis (e.g., “which-won-where” pattern), where specific genotypes can be recommended to specific mega-environments [35,36]; (ii) genotype evaluation (the mean performance and stability); and (iii) environmental evaluation (the power to discriminate among genotypes in target environments) [37]. GGE biplot provides information on genotype main effects and GEI at the same time. In contrast to typical multivariate stability analysis methods where only GEI is considered, the GGE method considers genotype main effects as well. Different researchers [38,39] have used GGE biplot for the analysis of GEI and evaluation of maize genotypes. In general, results of the studies have shown that in most stability analysis experiments, the main effect of environment is high, while variations determined by the main effect of genotype and GEI that are recommendable and interpretable are low. Since the environment is not a controllable factor in the GGE biplot method, genotypic and GEI sources of variation are used to obtain more reliable results [34,40]. Because it graphically displays GEI effects, the GGE biplot method helps plant breeders to easily assess genotypic stability and combinations of genotypic stability and yield in different environments. It also allows assessment of the relationship between environments and facilitates the re-arrangement of target environments in plant breeding programs. GGE biplot weaknesses were reported by Akinwale et al. [17] as follows:

- i) it does not put into consideration the complexity of the genetics of traits coupled with confounding environmental effects;
- ii) it can only identify two distinct heterotic groups in a genetic study where even more exist;
- iii) it cannot estimate genetic variances, covariances, and heritability; and
- iv) there is limited literature on its application to molecular data.

One recent study compared 15 methods of stability analysis using 17 varieties of maize evaluated in four years with several locations within the year for a total of 21 environments [41]. Spearman's rank correlation coefficient was used to rank the varieties (Table 2). Many of the methods had no significant correlation with each other (Table 2). For example,  $CV_i$  had rather low significant correlation coefficients with four of the 15 cases, two positive and two negative. Similarly, most other parameters had 5 to 7 significant correlation coefficients with others in the study, thereby displaying the specificity in the methods. In the study, however, the GGE showed rather consistent pattern. The varieties which were farthest from the biplot origin were positioned at the vertexes and were the most responsive to environments contained in the sector of each

vertex. The polygon of the 17 varieties were under five vertexes, which were V4 (TZEE-SRBC<sub>5</sub>), V9 (TZEE-SRBC<sub>5</sub>), V15 (EV 8435-SR), V11 (TZECOMP<sub>4</sub>C<sub>2</sub>) and V7 (ACR 95TZE COMP<sub>4</sub>C<sub>3</sub>). Varieties and environments positioned in the same sector on the graph were not significantly different from each other. In this analysis, varieties and environments were positioned as follows (Figure 3):



**Figure 3.** A polygon view of GGE biplot analysis of 17 maize OPVs (V) and 21 environments (E) included in a 4-year trial at the Teaching and Research Farm of Obafemi Awolowo University, Ile-Ife, Nigeria.

V7 (ACR 95TZE COMP<sub>4</sub>C<sub>3</sub>), was the highest yielding variety in E22, E16, E17 and E15; V11 (TZECOMP<sub>4</sub>C<sub>2</sub>), was positioned as the best yielding variety in E19 and E18; V15 (EV 8435-SR), was the winning variety in E20, E5, E13 and E14; while V4 (TZEE-SRBC<sub>5</sub>), was positioned as the highest yielding variety in environments E1, E10, E9, E2, E3, E9, E11, E7, E6, E8 and E12 (Figure 3). In the same study, the vector view of the GGE biplot revealed that V4 (TZEE-SRBC<sub>5</sub>), V2 (ACR 90 Pool i6-DT) and V13 (BAG 97 TZECOMP<sub>3</sub>x<sub>4</sub>) were the most stable and high yielding varieties (Figure not shown). In addition, the GGE biplot identified the hypothetical ideal variety as a small circle on average environment axis and points to higher average grain yield variety. Variety V4 (TZEE-SRBC<sub>5</sub>) occupied that position and was, therefore, the ideal and best variety in this study.

**Table 2.** Spearman's rank correlation coefficients among stability parameters computed by 15 analytical methods for 17 OPV maize varieties evaluated in 21 environments in a 4-year period at the Teaching and Research Farm of Obafemi Awolowo University, Ile-Ife.

	Mean	$S^2_i$	CV(%)	$\omega_i$	$\sigma_i^2$	$b_i$	$S^2_{di}$	$R^2$	$\delta_i^2$	$\beta_i$	Di	Pi	RSI	Si(1)	Si2	YSI	ASV
Mean																	
Variance	0.50 *																
CV(%)	-0.32	0.59 *															
$\omega_i$	0.12	0.51 *	0.47														
$\sigma_i^2$	0.10	0.49 *	0.44	0.97 **													
$b_i$	0.71 **	0.81 **	0.29	0.06	0.03												
$S^2_{di}$	0.10	0.49 *	0.44	0.97 **	1.00 **	0.03											
$R^2$	0.31	0.06	-0.22	-0.78 **	-0.79 **	0.50 *	-0.79 **										
$\delta_i^2$	0.12	0.51 *	0.47	1.00 **	0.97 **	0.06	0.97 **	-0.78 **									
$\beta_i$	0.71 **	0.81 **	0.29	0.06	0.03	1.00 **	0.03	0.50 *	0.06								
Di	0.99 **	0.56 *	-0.23	0.10	0.08	0.78 **	0.08	0.37	0.10	0.78 **							
Pi	-0.98 **	-0.54 *	0.26	-0.06	-0.03	-0.78 **	-0.03	-0.40	-0.06	-0.78 **	-0.99 **						
RSI	-0.88 **	-0.20	0.62 **	0.05	0.06	-0.42	0.06	-0.33	0.05	-0.42	-0.84 **	0.83 **					

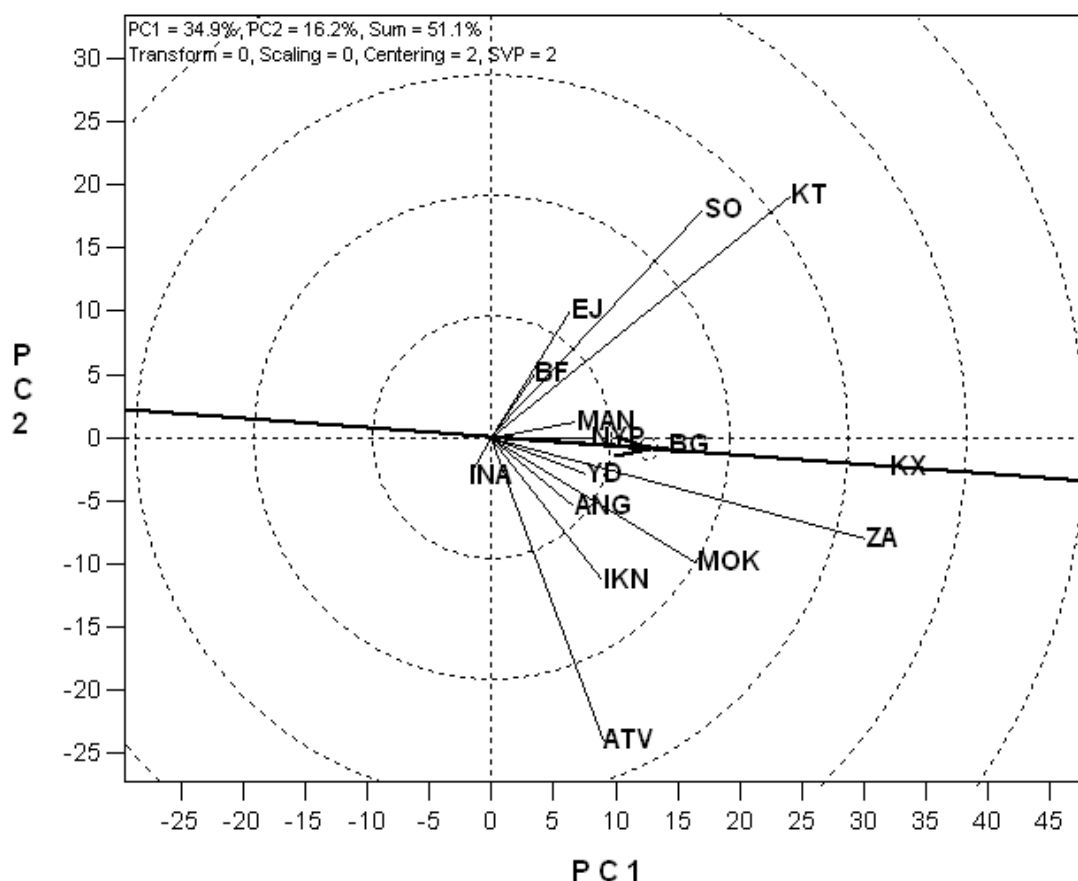


**Table 2.** *Cont.*

	Mean	$S^2_i$	CV(%)	$\omega_i$	$\sigma_i^2$	$b_i$	$S^2_{di}$	$R^2$	$\delta_i^2$	$\beta_i$	Di	Pi	RSI	Si(1)	Si2	YSI	ASV
Si(1)	0.17	0.46	0.35	0.28	0.20	0.40	0.20	-0.03	0.28	0.40	0.24	-0.19	0.00				
Si2	0.28	0.46	0.23	0.58 *	0.49 *	0.28	0.49 *	-0.36	0.58 *	0.28	0.29	-0.26	-0.16	0.74 **			
YSI	-0.58 *	-0.02	0.57 *	0.67 **	0.65 **	-0.41	0.65 **	-0.85 **	0.67 **	-0.41	-0.58 *	0.62 **	0.67 **	0.16	0.29		
ASV	0.10	0.44	0.48	0.78 **	0.71 **	0.06	0.71 **	-0.59 *	0.78 **	0.06	0.08	-0.10	0.08	0.02	0.34	0.46	
I	-0.62 **	-0.14	0.51 *	0.36	0.34	-0.44	0.34	-0.63 **	0.36	-0.44	-0.61 **	0.59 *	0.72 **	-0.16	-0.01	0.74 **	0.59 *

\*,\*\* = Significantly different from zero at the 0.05 and 0.01 probability levels, respectively.

$S^2_i$  = Genotype variance; CV = Coefficient of variation;  $\omega_i$  = Ecovalence;  $\sigma_i^2$  = Shukla's stability variance;  $b_i$  = Regression coefficient;  $S^2_{di}$  = Deviation from regression;  $R^2$  = Coefficient of determination;  $\delta_i^2$  = Deviation from adjusted regression;  $\beta_i$  = Adjusted regression coefficient; Di = Desirability index; Pi = Superiority index; Si(1) = Mean absolute rank differences; Si2 = Mean rank variances, YSI = Yield stability index; ASV = AMMI stability value; and SSI = Superior stability index.



**Figure 4.** The “discriminating power and representativeness” view of GGE biplot based on a genotype  $\times$  environment yield data of 18 early maturing maize cultivars evaluated in 15 locations across West Africa between 2006 and 2008.

Another study was conducted to examine the effect of  $G \times E$  on the performance and stability of 18 early maize cultivars and to identify core test sites and mega-environments at 15 locations in five countries of WA [14]. Results of the GGE biplot classified the locations into four mega-environments, regardless of their countries and Kita (KX, lat. 13°05' N, long. 09°25' W) in Mali was identified as the ideal location, and Zaria (lat. 13°05' N, long. 09°25' W) in Nigeria was close to the ideal location (Figure 4). In addition, variety 2004 TZE-W Pop STR C4 was identified in the study as the ideal cultivar because it had highest grain yield and was the most stable cultivar.

### Genotype $\times$ Trait Analysis

Analysis of trait relationship is an important aspect of maize improvement. Many important economic traits, such as grain yield, have low heritability because they are polygenic in nature. Thus, progress from direct selection for such traits is slow.

The scenario becomes compounded under stress conditions such as drought, low soil nitrogen and *Striga* parasitism which characterize maize production environments in sub-Saharan Africa. A better

alternative is to consider indirect selection through secondary traits that are relatively easier to measure and possess high heritability. Such traits can be identified based on analysis of relationship among traits. Correlation and regression analyses are the most popular statistics used for analysing trait relationships. However, in correlation and simple linear regression, relationship between only a pair of traits can be considered at a time. More recently, the use of other multivariate techniques, such as principal component analysis (PCA), principal coordinate analysis (PCoA), cluster and Additive and Multiplicative Model Interaction (AMMI) biplot, which analyse more than two traits at a time, has been in vogue. However, most of these statistics could not identify promising genotypes based on multiple traits. The genotype main effect plus genotype by environment interaction (GGE) biplot offers genotype by trait (GT) biplot analysis using robust statistical tools. Genotype-by-trait (GT) analysis presents the results of trait relationship by graphical display of the genetic relationships among traits [42]. It also provides information that helps to detect less important (redundant) traits and identify those that are appropriate for indirect selection for a target trait.

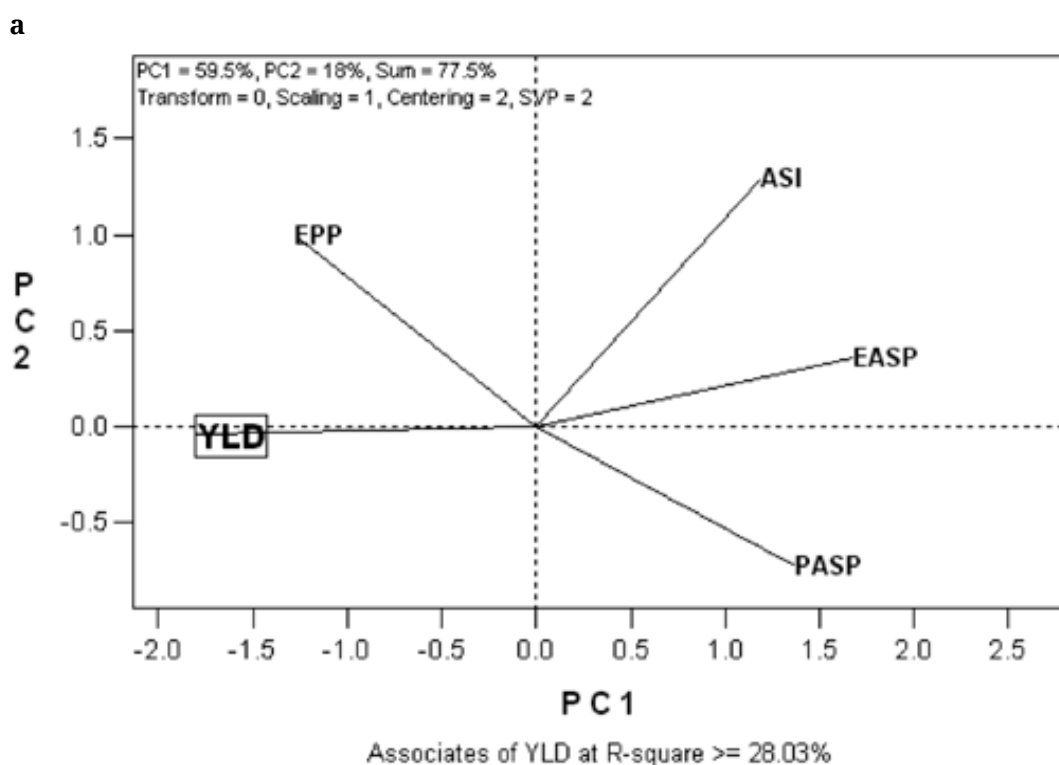
The GGE biplot model equation for the genotype-by-trait analysis is as follows:

$$(Y_{ij} - \mu - \beta_j) / d_j = \lambda_1 g_{i1} e_{1j} + \lambda_2 g_{i2} e_{2j} + \varepsilon_{ij} \quad (2)$$

Where  $Y_{ij}$  is the genetic value of the combination between inbred  $i$  and trait  $j$ ;  $\mu$  is the mean of all combinations involving trait  $j$ ;  $\beta_j$  is the main effect of trait  $j$ ;  $\lambda_1$  and  $\lambda_2$  are the singular values for PC1 and PC2;  $g_{i1}$  and  $g_{i2}$  are the PC1 and PC2 eigenvectors, respectively, for inbred  $i$ ;  $e_{1j}$  and  $e_{2j}$  are the PC1 and PC2 eigenvectors, respectively, for trait  $j$ ;  $d_j$  is the phenotypic standard deviation (with mean of zero and standard deviation of 1); and  $\varepsilon_{ij}$  is the residual of the model associated with the combination of Inbred  $i$  and trait  $j$ . For the GT biplot analysis, the data were not transformed ("Transform = 0") but were standard deviation-standardized ("Scale = 1"), and trait-centered ("centering = 2"). Therefore, the outputs are appropriate for visualizing the relationships among genotypes and traits.

Model II of the GGE biplot is considered the most appropriate for the analyses (Yan and Fregeau-Reid [43], Badu-Apraku et al. [44]). In the model, data were not transformed ("Transform = 0"), standard-deviation standardized ("Scale = 1") and were trait-centered ("Centering = 2"). The trait values were standard deviation-standardized in order to minimize the confounding effect that may result from the different units of measurement of the traits involved in the analysis. The polygon and vector views of the GT biplot were constructed using all measured traits and were based on genotype-focused singular value partitioning ("SVP = 2"), which rendered them appropriate for visualizing the relationships among traits. The entry/tester (mean vs stability) views were based on

trait-focused singular value partitioning (“SVP = 1”) which makes them appropriate for visualizing the relationships among genotypes. In order to identify ideal genotypes based on multiple traits graphically, data which are normally negatively correlated with target traits especially grain yield were redefined in reverse orientation. The re-orientations were necessary to ensure that the traits fell on one side of the average tester axis (ATC) coordinate, a requirement that had to be met in GGE biplot analysis to identify ideal entries across all testers from the entry/tester view of the biplots [15]. In addition, equal weights were assigned to each trait. Unlike other types of analyses in GGE biplot which lack test of significance, GT biplot analysis that assesses reliability of secondary traits for indirect selection of a target trait does the analysis at a specified probability level and at a stipulated R-square.



**Figure 5.** A vector view of the genotype  $\times$  trait biplot displaying most reliable traits for indirect selection for yield (inside box) under (a) managed moisture stress (b) well-watered environment (c) low-N environments (d) high N environments (e) *Striga* infestation. YLD/YD = Grain yield; EPP= number of ears per plant; ASI = anthesis-silking interval; EASP/EA = Ear aspect; PASP/PA = Plant aspect; EH = Ear height; PLHT/PH = Plant height; HC = Husk cover; LS = leaf senescence at 8 WAP; LS2 = leaf senescence at 10 WAP; DYS = Days to silking; DYA = Days to anthesis; LDTH = Leaf death score; STRA1 = *Striga* damage rating at 8 WAP; STRA2 = *Striga* damage rating at 10 WAP.

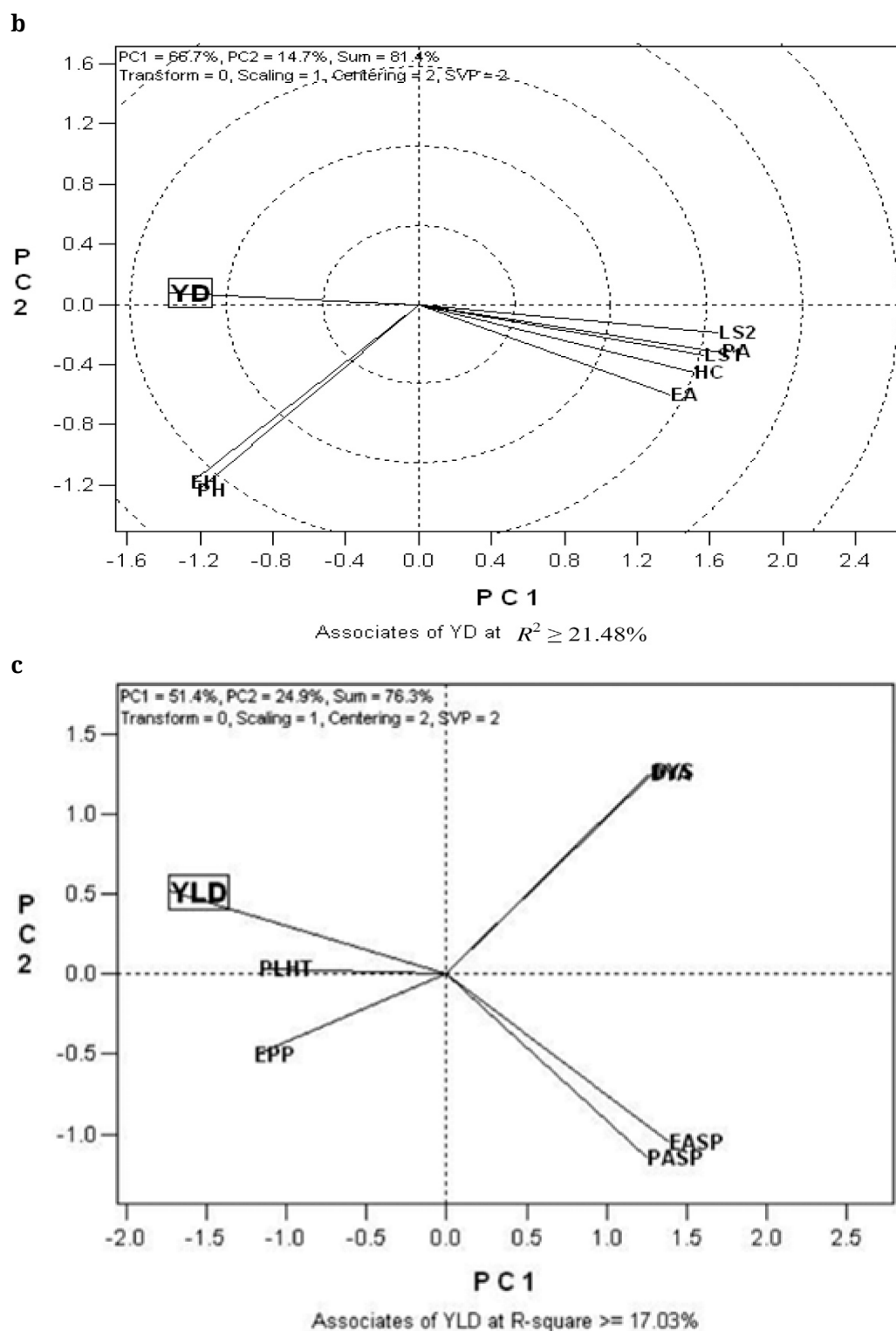
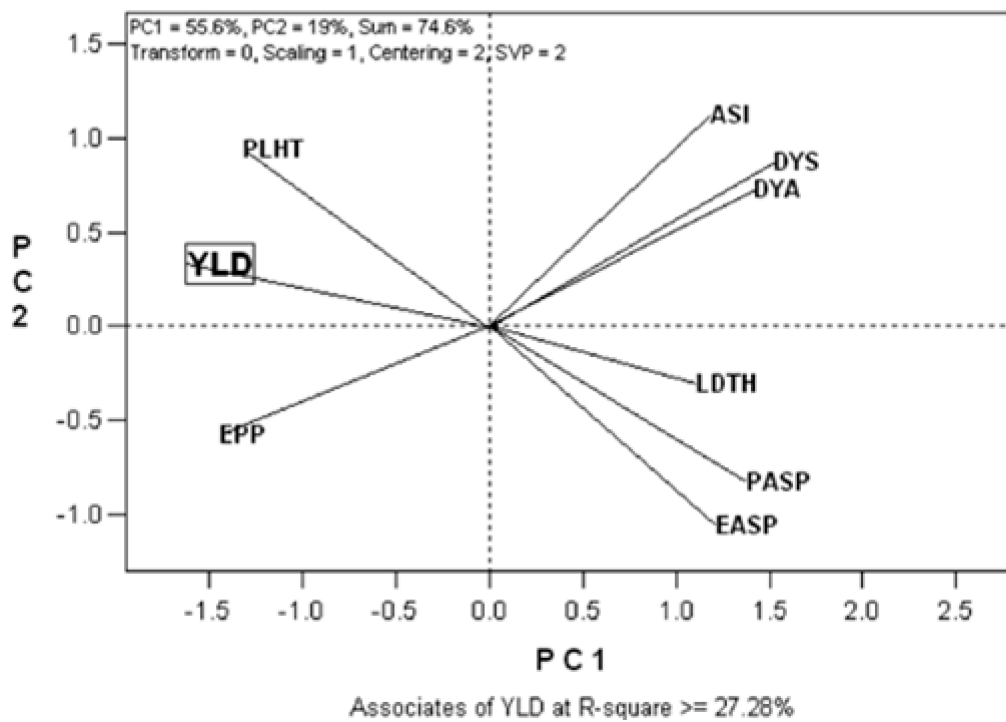


Figure 5. Cont.

d



e

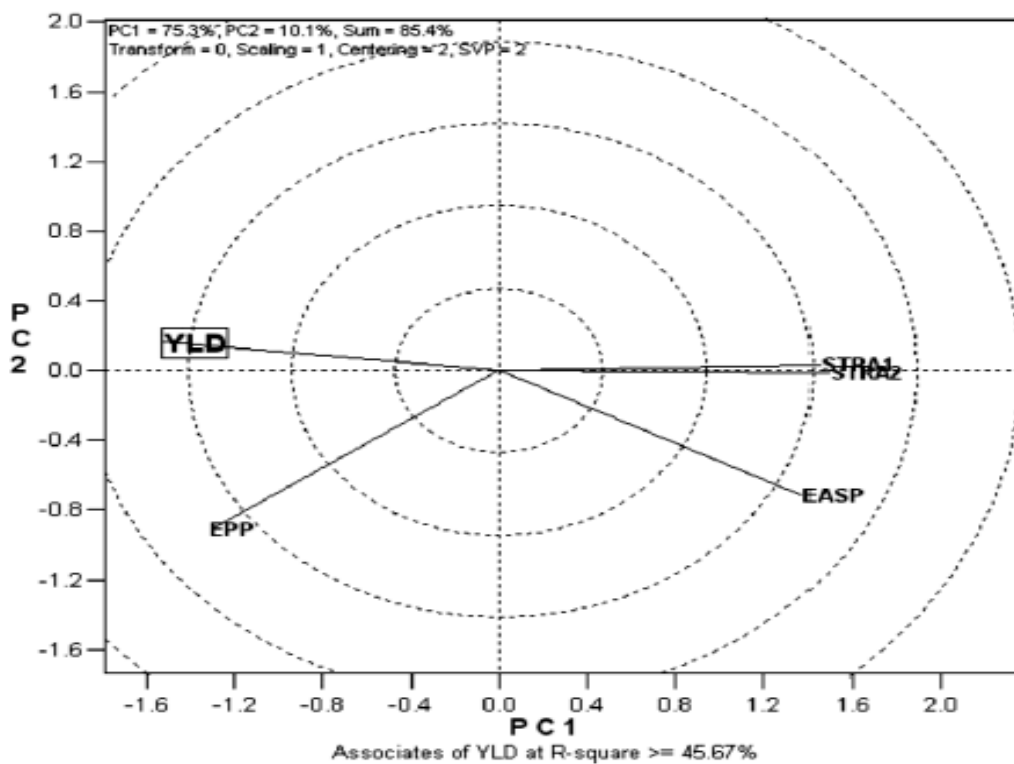


Figure 5. Cont.

Based on the use of GT biplot analysis, Badu-Apraku et al. [45] identified ear aspect (EASP), plant aspect (PASP), anthesis-silking interval (ASI), and number of ears per plant (EPP) as the most reliable traits for selection for yield under drought (Figure 5a). The most reliable traits

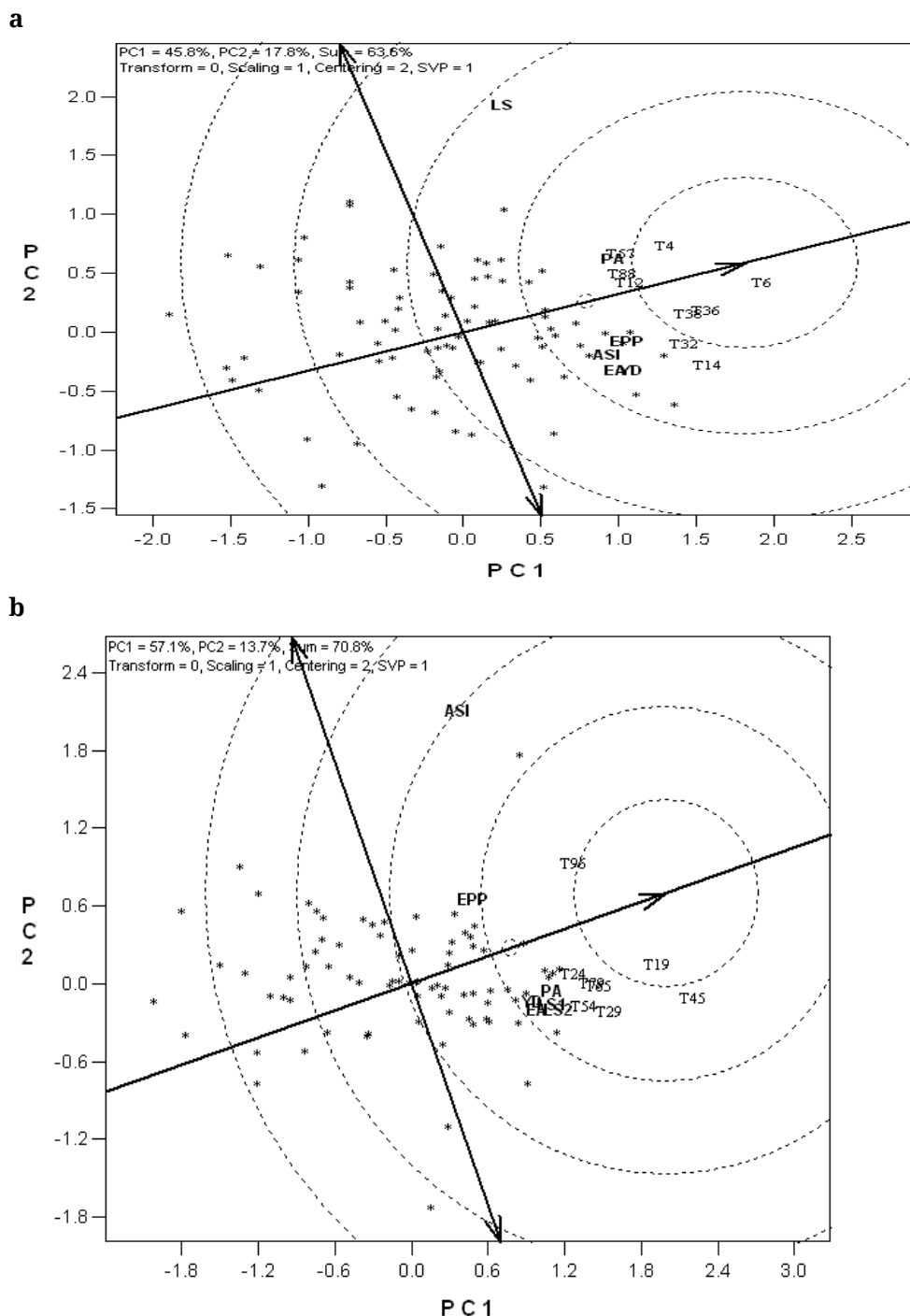


were days to anthesis (DYA), days to silking (DYS), stay green characteristic (LDTH), ASI, plant height (PLHT), EPP, EASP, under low-N (Figure 5b); PASP, DYA and DYS under well-watered environments (Figure 5c) and PLHT, EPP, PASP, and EASP under high-N environments (Figure 5d). In the same study, ASI, EPP, EASP, and PASP were identified as the most reliable traits for simultaneous selection of drought and low-N tolerant genotypes. Results reported in another study revealed that GT biplot identified EASP, EPP and *Striga* damage as the most reliable traits for indirect selection for improved grain yield under *Striga* infestation (Figure 5e) while EASP, PASP, PLHT and ear height (EHT) were identified under *Striga*-free conditions [46].

In order to validate consistency of the results of GT biplot with other multivariate techniques such as stepwise multiple regression analysis and path coefficient analysis, a study was conducted to compare the results of the GT biplot and path analyses by Badu-Apraku et al. [47]. Results revealed that both methods identified EASP, PLHT, and ASI as important traits directly contributing to yield under drought stress. Similarly, Oyekunle and Badu-Apraku [48] reported that the two methods identified EHT, EASP and PASP as important traits directly contributing to yield under drought. Furthermore, Badu-Apraku et al. [49] reported that GT biplot and sequential path analysis had consistent results as the two statistical analyses identified EASP as an important secondary trait with significant direct effect on yield under artificial *Striga* infestation, although in addition, GT biplot included EPP and *Striga* damage rating among the most reliable traits. On the basis of the results of GT biplot analysis of trait association under production stresses in WCA, recommendations have been made to reconstitute the traits to be used in the base index for selection of genotypes under each stress and combined base index for selection for multiple stress tolerance.

Apart from trait association analysis, GT biplot analysis of genotypes also provide graphical assessment, identification and selection of genotypes on the basis of single and multiple traits [35,42,50,51]. A comprehensive multi-trait selection procedure proposed by Yan and Fregeau-Reid [43] called “*Multi-trait-Selection Master*” of the GGE biplot software is a breeder’s tool that combines three selection strategies; independent selection, independent culling, and index selection, so that all the aspects of variety or parent line selection are taken into consideration. In a study with 23 tropical early-maturing maize inbreds, GT biplot identified 10 outstanding inbreds in terms of high grain yield and *Striga* resistance, which included TZEI 3, TZEI 2, TZEI 11, and TZEI 15 using the *Multi-trait-Selection Master*’ feature of the GGE biplot [46]. Graphical analysis of the performance of some early maize inbreds based on multiple traits identified TZEI 17, TZEI 3, TZEI 23, and TZEI 13 as the closest to the ideal genotype under drought stress, and TZEI 7, TZEI 2, and TZEI 11 under low-N conditions (Figure 6a) [45]. Similarly, TZE-W DT STR C<sub>4</sub> had the best performance based on multiple traits while TZE-W DT

STR C<sub>4</sub>, TZE-Y DT STR C<sub>4</sub>, Multicob Early DT, and TZE-W DT STR QPM C<sub>0</sub> were the closest to the ideal cultivar when *Striga* infested among 15 cultivars tested [39]. Similarly, results of selection of extra-early maize inbreds and hybrids under drought stress and low soil N using GGE biplot identified inbred TZEEI 6 as the ideal based on multiple traits under drought stress and inbreds TZEEI 96 and TZEEI 45 under low-N conditions (Figure 6b) [45].



**Figure 6.** A vector view of genotype-by-trait biplot showing the ranking of the 90 extra-early inbreds based on their mean performance **(a)** across six selected drought tolerance traits. **(b)** under low soil N conditions.

In summary, GT biplot revealed that the most reliable traits for selection for yield under drought were EASP, PASP, ASI, and EPP. DYA and DYS were the most reliable under well-watered environments. Under low-N, most reliable traits were DYA, DYS, LDTH, ASI, PLHT, EPP, EASP, and PASP whereas PLHT, EPP, PASP, and EASP were the most reliable traits under high-N environments.

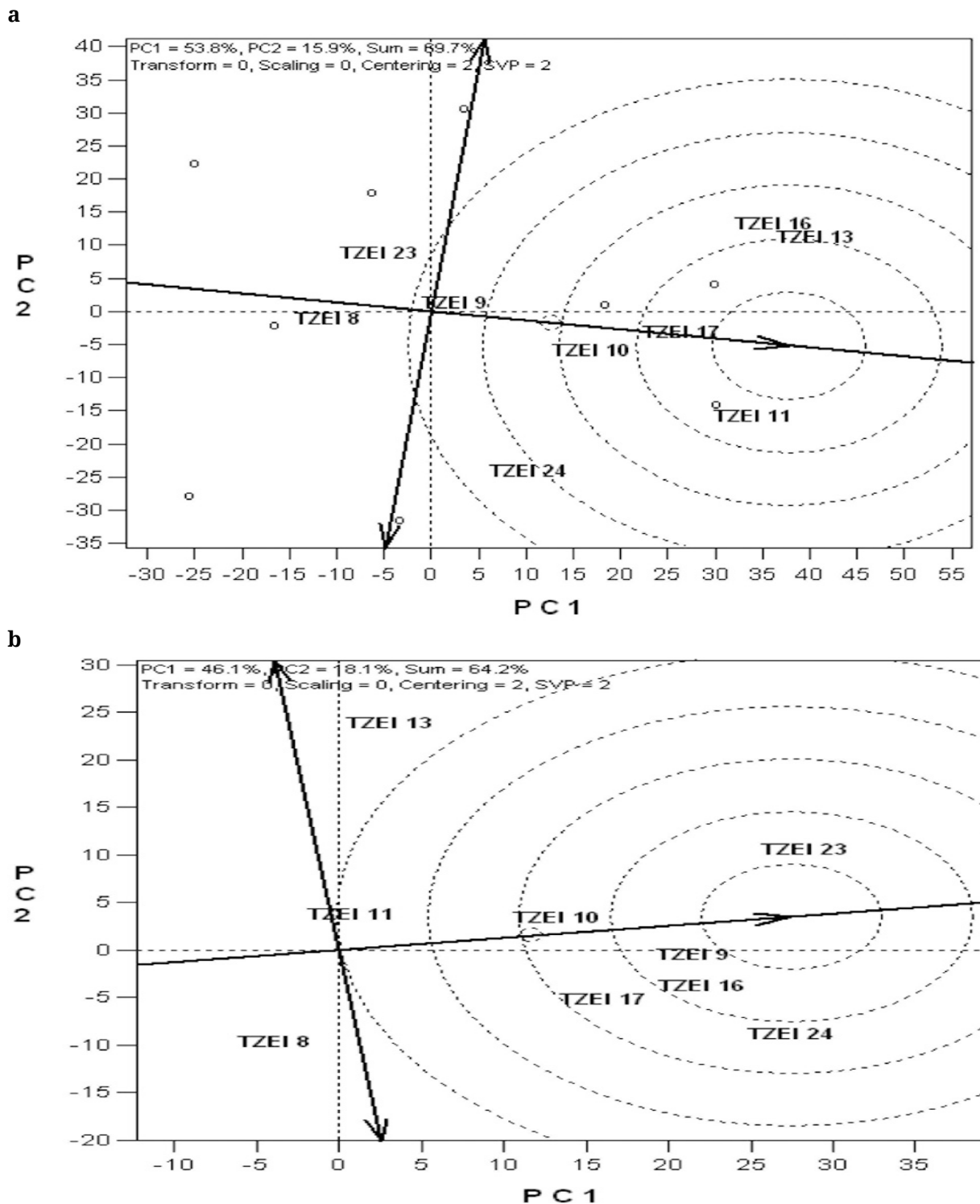
The ASI, EPP, EASP, and PASP were identified as most reliable traits for simultaneous selection of drought and low-N tolerant genotypes.

## Analysis of Genetic Data

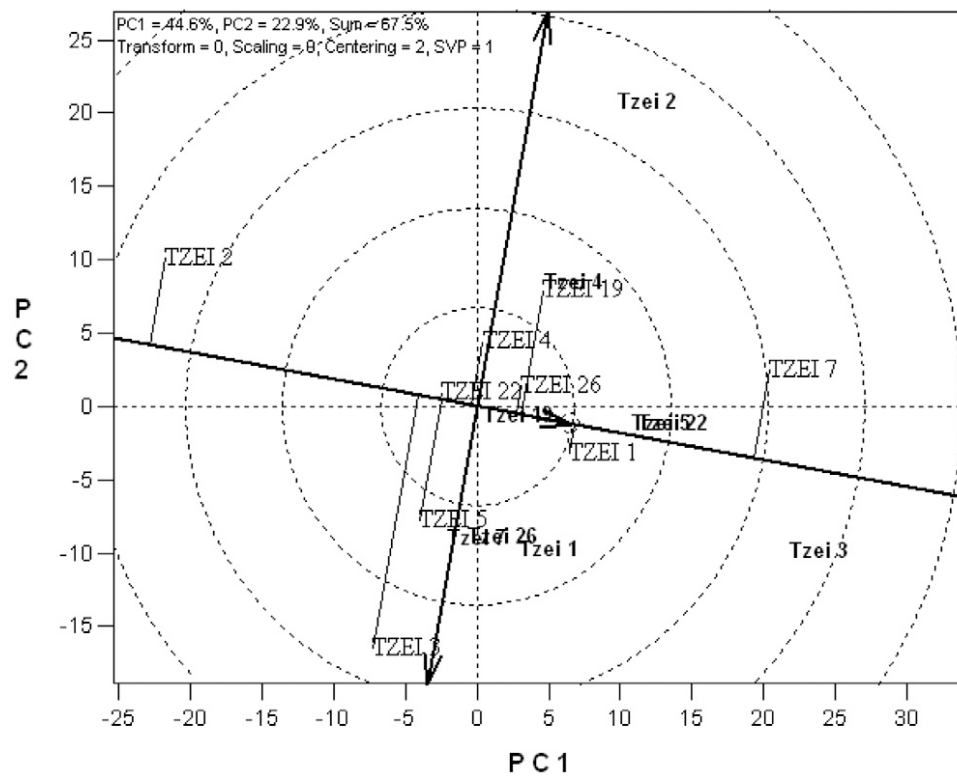
### *Diallel analysis*

Adequate knowledge and understanding of genetic variability, modes of inheritance and heterotic response in a germplasm are very crucial for determining appropriate methods to employ for improving the genetic resources. Backcrossing, inbreeding, hybridization, and the  $S_1$  recurrent selection method are among breeding strategies used by International Institute of Tropical Agriculture (IITA) Maize scientists to develop several early and extra-early maturing white and yellow-grained source populations, cultivars, and inbred lines. The choice of appropriate breeding strategy is dependent on the results of genetic analysis of the traits to improve in a set of parents. Genetic analyses are carried out using mating designs such as bi-parental progeny analysis, generation mean analysis, diallel, North Carolina designs, line  $\times$  tester, among others. The effectiveness of these mating designs in analysing the genetics of inheritance of important traits of crops has been tremendous but the rigour and complexity involved in the manual mathematical computations of genetic parameters and correct interpretations have been a great challenge especially to many new generation breeders. The GGE biplot graphical approach to analysis of genetic data makes computation and interpretation of results very fascinating. Yan and Hunt [52] proposed and demonstrated how GGE biplot could be used to analyse diallel data and give appropriate interpretations to its results.

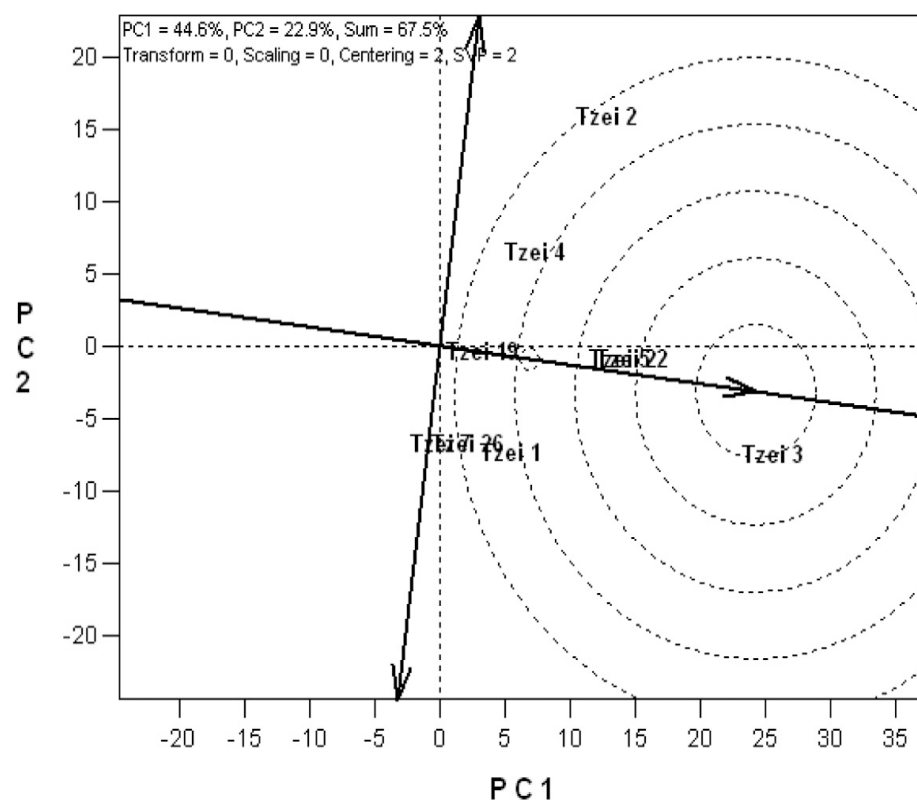
Generally, GGE biplot is suitable for analysis of any form of data that can be cast into two-way table. This makes GGE biplot super-adequate in analysing data from mating designs that are cross-classification or factorial in nature rather than nested design. Examples of such cross-classification designs are diallel, line  $\times$  tester and North Carolina Design II. The biplot graphical analysis allows visual display of the combining abilities (general (GCA) and specific combining ability (SCA)) of each of the parents and hybrids used based on Griffings diallel methods. In addition, information on the best mating partner, identification of proven testers and tester groups and heterotic groups are provided graphically.



**Figure 7.** Vector view of the genotype plus genotype  $\times$  environment biplot showing the ranking of the testers based on their discriminating ability and representativeness under (a) drought-stressed environments and (b) artificial *Striga* infestation. Exact positions of the entries and testers are at the beginning of the labels.



**Figure 8.** Biplot based on diallel data of nine early-maturing white inbred lines with varying performance across 12 environments; average tester ordination view.



**Figure 9.** A vector view of genotype plus genotype × environment biplot showing the ranking of testers based on their discriminating ability and representativeness across all research environments.

Badu-Apraku et al. [53] in a diallel study involving 9 yellow early maturing tropical maize inbreds under drought stress, *Striga* infestation, and optimal growing conditions as well as across research environments reported that the GCA effects of the inbred lines were of limited importance due to the over dominating effects of SCA. Thus, in the study, there were no inbreds that had significantly higher GCA than others and the lines could not be classified into distinct heterotic groups based on GCA and SCA. However, based on the relationship among testers, inbreds identified as the closest to the ideal tester were TZEI 17 and TZEI 11 under drought and TZEI 23 and TZEI 9 under *Striga* infestation (Figure 7a,b). In a similar study, thirty-six diallel crosses derived from nine tropical early maturing white maize inbreds were evaluated under drought, well-watered, *Striga*-infested, and *Striga*-free environments in Nigeria between 2007 and 2009 and the data collected were subjected to GGE biplot analysis [54]. Results revealed preponderance of additive gene action over the non-additive, for yield and most stress adaptive traits. The GGE biplot analysis revealed that TZEI 4 and TZEI 5 were closest to the ideal inbreds in yield performance and stability across the test environments. Two heterotic groups were identified; [TZEI 7, TZEI 19, TZEI 2, TZEI 4] and [TZEI 5, TZEI 3] (Figure 8). Tester TZEI 3 was the closest to the ideal tester while Entry TZEI 7 had the highest GCA effects across stress environments (Figure 9).

In summary, analysing diallel data using GGE biplot is very fascinating and it provides more genetic information beyond just the combining ability of the parents and hybrids. It gives additional information on the relationship among parents, identify testers, assess efficiency of testers, display relationships among testers, identify tester groups, reveals best mating partners, and most importantly, identify heterotic groups. These additional information are not readily available in conventional analysis of diallel data.

A major limitation to the use of diallel mating design is that there is a limit to the number of parents that can be involved. Results of only a few parents can be clearly displayed. As the number of parents to analyse increases, the results of GGE biplot become clustered and both entry and tester labels overlap and the biplot graphical views appear clumsy. In a breeding program where hundreds of inbred lines have to be analyzed, diallel analysis using GGE biplot becomes impracticable.

### **Line × Tester Analysis**

Because of the shortcomings of diallel design in handling large number of parents, line × tester analysis was proposed by Kempthorne [55]. This design requires availability of proven testers probably identified in previous studies. As appropriate as GGE biplot is in the analysis of line × tester data, original journal article wherein line × tester analysis is carried out using GGE biplot is very scanty if not unavailable possibly because authors and researchers are not aware of the enormous

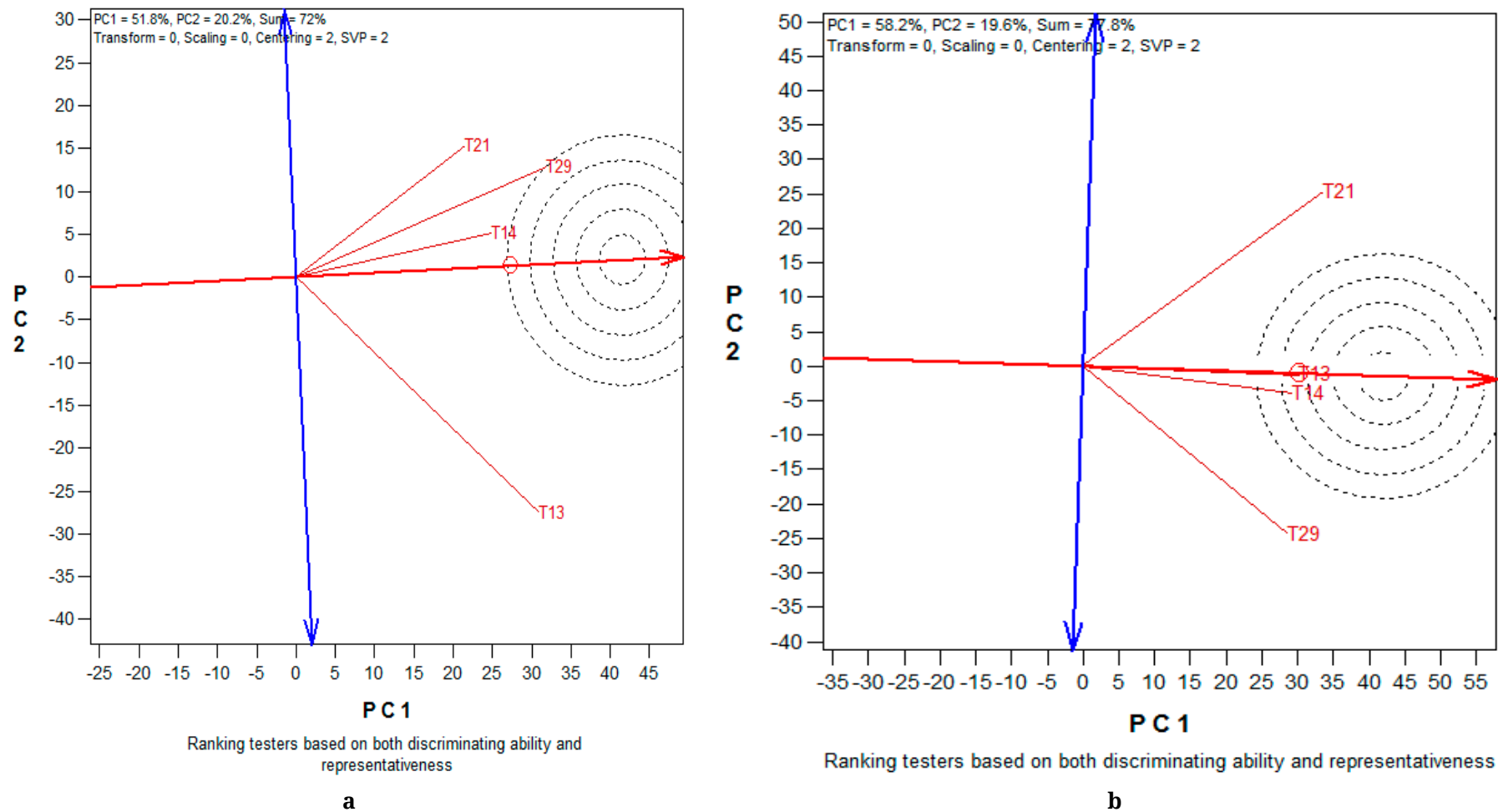


genetic information it provides. However, in the IITA Maize Program, GGE biplot has been used for analyzing data from a line  $\times$  tester study. Data were generated from 63 newly developed inbred lines crossed to four extra-early elite testers (TZEEI 13, TZEEI 14, TZEEI 21 and TZEEI 29) evaluated under multiple stress and stress-free environments. Using GGE biplot, an ideal tester could not be identified under stress environments. However, testers TZEEI 13 and TZEEI 14 were the closest to the ideal tester under nonstress environments (Figure 10) [56]. Inbred TZdEEI 34 was identified as outstanding in terms of GCA effects under both stress and nonstress environments. Testers TZEEI 13, TZEEI 21 and TZEEI 29 were found to be very efficient across stress environments based on their discriminating power while testers TZEEI 21 and TZEEI 29 were the best across nonstress environments (Figure 11).

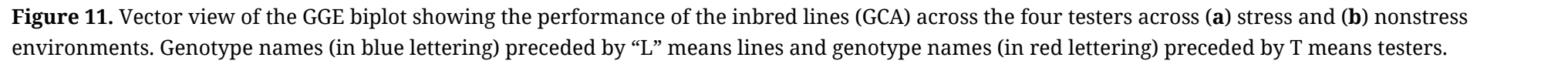
Challenges encountered with the application of GGE biplot for analysing data from line  $\times$  tester are similar to those of the diallel. However, because the number of testers used in line  $\times$  tester analysis is usually less than in diallel (where number of parents is considered as the number of testers), the graphical display of the results of line  $\times$  tester study is better than that of diallel. Interpretation of results of line  $\times$  tester is also easier and simpler than that of the diallel.

North Carolina design II (NCDII) is the third factorial mating design that could be analysed using GGE biplot analysis. However, there is no report in the literature where GGE biplot has been used for analysis of data from NCDII. One reason could be because larger number of parents can be accommodated in NCDII compared to the diallel and for better organization, males nested within set is considered as a factor in the statistical model rather than male factor.

We recommend that for GGE biplot to have a wider application in analysis of genetic as well as agronomic data, the proponents should consider incorporating features that will be appropriate for analyses of random and mixed model data and data from nested type of mating design.



**Figure 10.** A vector view of the GGE biplot showing the ideal testers based on the discriminating power and representativeness of the testers across (a) multiple stress and (b) nonstress environments. Genotype names (in red lettering) preceded by T means testers.

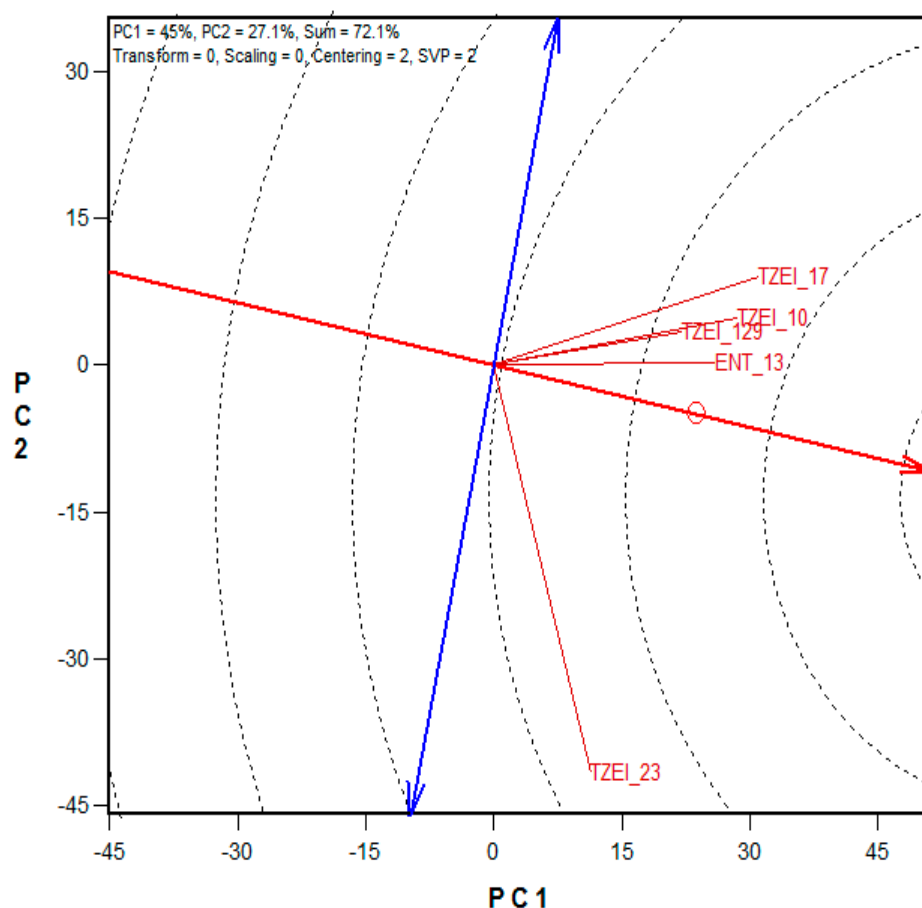


### Evaluation of the Efficiency of Testers in Hybrid Production

An important prerequisite for the development of high-yielding commercial hybrids is the availability of efficient testers, which could successfully discriminate, classify inbred lines into appropriate heterotic groups, and combine well with other inbred lines, open pollinated varieties or hybrids. An effective tester should be able to rank inbred lines correctly for performance in hybrid combinations and increase the differences between testcrosses for efficient discrimination [57]. Furthermore, such testers must have improved agronomic characteristics, resistance to diseases and tolerance/resistance to prevailing biotic and abiotic stresses such as drought, low-N and *Striga*. Assessment of the efficiency of testers is crucial for the identification of suitable genotypes for the development of productive hybrids and ensuring efficient utilization of resources. Therefore, the choice of proven testers is an important factor that determines progress made in maize hybrid development programs. At the initial stage of hybrid development in the early (90–95 days to physiological maturity) and extra-early (80–85 days to physiological maturity) maturing groups at IITA, proven testers were not available in the tropical germplasm. Therefore, the testers from the late maturity group (120 days to physiological maturity) in the temperate germplasm such as Mo17 and B73 were used. However, the results of crosses of tropical lines with temperate testers were not producing desirable results. After a while, testers were identified among the late/intermediate maize germplasm which were adopted for early and extra-early maize germplasm [58]. Over the years, several testers have been developed in the early and extra-early maturity group to facilitate the development of superior hybrids for SSA. This has necessitated identification of a few efficient testers for use in classifying the available inbred lines into heterotic groups as well as inbred lines for the development of outstanding commercial hybrids for production in SSA.

The GGE biplot tool has the potential for identifying efficient testers even though its use for such analysis has not been adequately explored. Several early maturing inbred lines, including TZEI 10, TZEI 17, TZE 23, TZEI 129 and ENT 13 have been identified as potential testers in the IITA Maize Improvement Program (MIP) using the GGE biplot statistical tool. The GGE biplot has been used to identify the most efficient testers among the five inbred lines. As described by Akinwale et al. [17] and Yan [59], the efficiency of a tester (testers were used to replace environments) is determined by the relationship among the testers and the length of the tester vector. The smaller the angle between any two testers, the more closely related the testers are while testers with longer vectors show high discriminating power or its ability to assess the grain yield of the crosses. Badu-Apraku et al. [53] evaluated nine tropical early maturing maize inbreds in diallel crosses under drought stress, *Striga* infestation, and in optimal growing conditions to examine the combining abilities and

heterotic patterns and identify appropriate testers using the GGE biplot tool. Inbred TZEI 17 was identified as the best tester under drought, TZEI 23 and TZEI 9 under artificial *Striga* infestation, and TZEI 13 across growing conditions. In another study, Annor et al. [60] identified inbreds TZEI 23 and TZEI 17 as the most efficient testers in grouping tropical early yellow maize inbreds (Figure 12) and were therefore recommended for use in hybrid development programs in SSA. In a more recent study, Badu-Apraku and Akinwale [56] identified extra-early inbred testers TZEEI 13, TZEEI 21 and TZEEI 29 as the most efficient across drought and *Striga*-infested environments and TZEEI 21 and TZEEI 29 across nonstress environments using the GGE biplot (Figure 10). Similarly, in studies conducted by Adewale (unpublished, 2019) using the GGE biplot, the inbred testers TZdEI 352 and TZEI 18 were identified as the most efficient across drought and *Striga*-infested environments and were therefore recommended for classifying early tropical white maize inbred lines which are yet to be field-tested and for the development of productive hybrids in the sub-region.



**Figure 12.** A vector view of genotype main effect plus genotype by environment (GGE) biplot showing the discriminating power and representativeness of the testers across 17 environments in Nigeria from 2014 to 2016.

### Strengths and Weaknesses of the GGE Biplot

The GGE biplot is a superior data-visualization tool widely used in several major areas of agronomy, plant breeding and for analysis in genetic studies involving GEI, test location evaluation, genotype evaluation, mega-environment investigation and identification of parental inbreds for hybrid development [61]. This tool allows researchers to graphically extract and utilize information from METs data and other types of two-way data [35]. However, the full potential and shortcomings of this powerful tool are not completely understood by breeders, geneticists, agronomists, ecologists, entomologists and pathologists. The limited use of this tool could be attributed to lack of understanding of its potential capability on the part of many researchers. Furthermore, the major weaknesses as well as potential useful areas of application of this tool are not easily realized by users. For instance, nearly all analysis from GGE biplot usually lack discrete statistical test of significance since its results are only graphically displayed. Consequently, Yan [62] indicated that the GGE biplot is preferably used in generating hypothesis rather than for making decisions while Yan and Ma [63] presented the theoretical and fundamental basis for appropriate use of GGE biplot. Thus, further evaluation of the theoretical basis and applied results of the software as well as critiquing of the results from the analysis and implications relative to other statistical tools is crucial. Yang et al. [64], in evaluating and critiquing the use of the GGE biplot, recommended using bootstrap procedure for testing significance of the  $G \times E$  biplot pattern. Contrarily, Yan et al. [65] reported that the complex bootstrap method is inadequate for the test of significance of biplot results.

GGE biplot was primarily developed for the analysis of complex GEI. After analysis of variance has showed significant mean squares for GEI, the GGE biplot complements the results, by graphically displaying the nature of the interactions. The introduction of the concept of crossover interaction and GGE biplot has increasingly been used in GEI data analysis in agriculture [34,46,66]. The GGE biplot statistical tool has been effectively used in identifying outstanding genotypes, assessing pairs of genotypes in specific environments and detecting the most appropriate test environments (those that possess the highest ability to discriminate among genotypes and are most representative of all test environments) [15,16,45,56,59]. Yan et al. [34] further emphasized that GGE biplot is more logical and biological than AMMI in explaining the PC1 score, which indicates genotypic effects rather than additive main effects. Using the GGE biplot tool, Badu-Apraku et al. [10] and Badu-Apraku and Lum [67] identified early maturing maize cultivars that were suitable for *Striga*-infested and *Striga*-free environments and determined their stability performance across environments. Similarly, Oyekunle et al. [68] used the GGE biplot analysis to assess the performance of early-maturing maize hybrids and identified ideal test locations in West Africa. Results of the study showed that Minjibir (Nigeria) and Nyankpala



(Ghana) were the most discriminating and representative locations and were declared as the ideal testing sites for the respective national maize programs of West African countries. Hybrids EWH-29, EWH-8, and EWH-30 were found to be the most stable in both countries, whereas EWH-26 and EWH-32 were the most stable hybrids only in Ghana.

The genotype-by-trait (GT) analysis component of the GGE biplot incorporates some features of AMMI, joint regression, as well as genetic correlation in data analysis, making it excellent compared with other multivariate approaches. Furthermore, multivariate analytical methods such as correlation, multiple regression, path analysis [69], factor analysis and cluster models have the demerit of not being efficient in identifying genotypes with certain desirable traits that could be used in a breeding program. GT analysis also gives information on the suitability of cultivars for production as well as information that help to identify redundant traits and detect those that are useful for indirect selection for a target trait.

Using GT biplot, Badu-Apraku and Akinwale [39] investigated inter-trait relationships under *Striga* infestation and identified EPP, *Striga* damage at 8 and 10 weeks after planting and ear aspect as useful selection indices for *Striga* resistance. In another study, sequential path analysis identified EASP as the only trait having significant direct effect on yield under *Striga* infestation [49] whereas GGE biplot revealed EASP, EPP and *Striga* damage as the most reliable secondary traits. The authors established that EASP should be incorporated in the base index for selecting for improved grain yield of extra-early maturing maize under *Striga* infestation, while the number of emerged *Striga* plants should be excluded.

Breeders focus more attention on estimating genetic parameters since they increase the effectiveness of predicting gains from selection for the genetic enhancement of crop cultivars. GGE biplot has been extensively employed in combining ability analysis and identification of heterotic patterns using diallel data [35,53] and line  $\times$  tester data [56]. Badu-Apraku and Akinwale [56] used the GGE biplot tool to determine the combining ability effects, identify distinct heterotic groups and efficient testers in a line  $\times$  tester study. The results obtained revealed close correspondence in the combining ability and patterns of grouping of the lines with that of the conventional line  $\times$  tester mating design, thus validating the GGE biplot as a reliable statistical tool for evaluating line  $\times$  tester data.

Despite the numerous strengths of GGE biplot analysis, some weaknesses still exist. The major shortcoming of GGE biplot in the analysis of trait relationships is its insensitivity to multi-collinearity and spurious correlations when used to identify reliable traits that could be employed for indirect selection for a target trait. This is because GGE biplot largely identifies reliable traits based on genetic correlations among traits, the traits that are highly correlated are displayed as

reliable, without considering which of the traits contribute indirectly to the target trait through other traits. This weakness was clearly revealed when the results of interrelationships among traits using GT biplot and sequential path analysis were compared [46].

In its application to analyze genetic data, classification of genotypes into heterotic groups has been based on the SCA effects only, which is represented by the projections of the entry vectors onto the ATC ordinate. In a situation where the GCA is preponderant over the SCA, classifying genotypes into heterotic groups based on SCA alone, as analysed by GGE biplot, will be grossly inefficient and the groups will not be distinct.

Another major challenge with diallel analysis using GGE biplot is that only two heterotic groups can be identified even when more groups are present. Other inbreds that cannot fit into the two groups become unclassified [17]. Furthermore, the proportion of parents classified are smaller relative to the total number of parents involved in the study. This is particularly of great concern especially in a standard breeding program that has committed considerable time, energy, efforts, land, funds and other resources to produce several inbred lines only to find out that just a few can be classified into heterotic groups for the purpose of hybrid development.

Another major shortcoming of GGE biplot analysis of genetic data such as diallel is that only fixed statistical model is applied. When genotype is considered as a random model where the experimenter is interested in computing genetic variances and heritability estimates, application of GGE biplot in the analysis of such data becomes limiting since the biplot has not been designed to display these parameter estimates graphically. Furthermore, GGE biplot tool has not been used in the analysis of data generated using North Carolina Designs I, II, and III and some other genetic designs. The use of GGE biplot in the analysis of the mating designs could facilitate a better understanding of the mating designs.

## CONCLUSIONS AND FUTURE DIRECTIONS

GGE biplot is the most widely used multivariate analytical tool in the analysis of plant breeding data. The interpretation of GGE biplot analysis of genetic data is more comprehensive with wider applicability than the conventional statistical methods. Nevertheless, the lack of discrete statistical test of significance in its analysis has sometimes made the reliability of its results debatable by researchers. However, its results have been found to be consistent with that of ANOVA, correlation, regression and multivariate statistical methods. GGE biplot has been extensively used in the tropical early and extra-early maize improvement program and its use has facilitated rapid progress in population improvement, variety and hybrid development in the IITA maize program. The statistical tool still possesses much more potentials

that are yet to be fully explored especially for the tropical maize germplasm.

#### **DATA AVAILABILITY**

The dataset of the study is available from the authors upon request.

#### **AUTHOR CONTRIBUTIONS**

BBA, BF and RA conceived and designed the reviewed experiments as well as drafted the manuscript. BBA, BF, RA, BA and SAK executed the experiments. SAD and JT assisted in drafting the manuscript. All authors critically reviewed the manuscript.

#### **CONFLICTS OF INTEREST**

The authors declare that there is no conflict of interest.

#### **ACKNOWLEDGEMENTS**

This research was supported by the Bill and Melinda Gates Foundation [OPP1134248]. The authors appreciate the staff of the IITA Maize Improvement Program and Obafemi Awolowo University Teaching and Research Farm for research facilities.

#### **REFERENCES**

1. Yau SK. Regression and AMMI analyses of genotype-environment interactions: an empirical comparison. *Agron J*. 1995;87:121-6.
2. Crossa J, Gauch HG, Zobel RW. Additive main effect and multiplicative interaction analysis of two international maize cultivar trials. *Crop Sci*. 1990;30:493-500.
3. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th edition. New York (US): Longman; 1965.
4. Wu RL, O'Malley DM. Non-linear genotypic response to macro- and micro-environments. *Theor Appl Genet*. 1998;96:669-75.
5. Lin CS, Binns MR. A superiority measure of cultivar performance for cultivar  $\times$  location data. *Can J Plant Sci*. 1988;68:193-8.
6. Fakorede MAB, Adeyemo MO. Genotype  $\times$  environment interaction components of variance for three types of maize varieties in the rain forest zone of south western Nigeria. *Niger J Agron*. 1986;1:43-6.
7. Badu-Apraku B, Fajemisin JM, Diallo AO. The performance of early and extra-early varieties across environments in West and Central Africa. In: Badu-Apraku B, Akoroda MO, Ouedraogo M, Quin FM, editors. Contributing to food self-sufficiency: Maize research and development in West and Central Africa: Proceedings of a Regional Maize Workshop; 1995 May 29-Jun 2; Cotonou, Benin Republic. Cotonou (Benin Republic): IITA; p. 149-59.
8. Badu-Apraku B, Abamu FJ, Menkir A, Fakorede MAB, Obeng-Antwi K, The C. Genotype by environment interactions in the original early vatriety trials in West and Central Africa. *Maydica*. 2003;8:93-104.

9. Badu-Apraku B, Fakorede MAB, Lum AF. Evaluation of experimental varieties from recurrent selection for Striga resistance in two extra-early maize populations in the savannas of West and Central Africa. *Exp Agric*. 2007;43:183-200.
10. Badu-Apraku B, Lum AF, Fakorede MAB, Menkir A, Chabi Y, The C, et al. Performance of early maize cultivars derived from recurrent selection for grain yield and *Striga* resistance. *Crop Sci*. 2008;48:99-112.
11. Gauch HG, Zobel RW. Identifying mega-environments and targeting genotypes. *Crop Sci*. 1997;37(2):311-26.
12. Food and Agriculture Organization (FAO). West African Catalogue of Plant Species and Varieties, Rome, Italy. 2008. Available from: <http://www.fao.org/3/i0062e/i0062e00.htm>. Accessed 2020 February 26.
13. Badu-Apraku B, Oyekunle M, Akinwale RO. Combining ability of early-maturing white maize inbreds under stress and nonstress environments. *Agron J*. 2011;103:544-57.
14. Badu-Apraku B, Akinwale RO, Menkir A, Obeng-Antwi K, Osuman AS, Coulibaly N, et al. Use of GGE biplot for targeting early maturing maize cultivars to mega-environments in West Africa. *Afr Crop Sci J*. 2011;19:79-96.
15. Yan W, Kang MS, Ma S, Woods S, Cornelius PL. GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Sci*. 2007;47:596-605.
16. Badu-Apraku B, Akinwale RO, Obeng-antwi K, Haruna A, Kanton R, Usman I, et al. Assessing the representativeness and repeatability of testing sites for drought-tolerant maize in West Africa. *Can J Plant Sci*. 2013;93:699-714.
17. Akinwale RO, Fakorede MAB, Badu-Apraku B, Oluwaranti A. Assessing the usefulness of GGE biplot as a statistical tool for plant breeders and agronomists. *Cereal Res Commun*. 2014;42(3):534-46. doi: 10.1556/crc.42.2014.3.16
18. Yayeh Z, Bosland WP. Evaluation of genotype, environment, and genotype-by-environment interaction for capsaicinoids in *Capsicum annuum* L. *Euphytica*. 2012;111:185-90.
19. Lin CS, Binns MR, Lefkovitch PL. Stability analysis: where do we stand? *Crop Sci*. 1986;26:894-9.
20. Becker HC, Leon J. Stability analysis in plant breeding. *Plant Breed*. 1988;101:1-23.
21. Francis TR, Kannenberg LW. Yield Stability Studies in short-season maize. I. A descriptive method for grouping genotypes. *Can J Plant Sci*. 1978;58:1029-34.
22. Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme. *Aust J Agric Res*. 1963;14:742-52.
23. Shukla CK. Some statistical aspects of partitioning genotype-environment components of variability. *Heredity*. 1972;29:237-45.
24. Eberhart SA, Russell WA. Stability parameters for comparing varieties. *Crop Sci*. 1966;6:36-40.
25. Perkins JM, Jinks JL. Environment and genotype-environment component of variability. II. Multiple lines and crosses. *Heredity*. 1968;23:18.

26. Wricke G. Uber eine method zur erassungderokologischen streubreite in feldversuchen. Z. Pflanzenzuchtg. 1962;47(1):92-6. German.
27. Shaibu AS, Muhammad IH, Adnan AA. Assessment of genetic diversity of maize (*Zea mays* L.) in Sudan Savannah. Appl Trop Agric. 2017;22(1):29-34.
28. Zobel RW, Wright MT, Gauch HG. Statistical analysis of a yield trial. Agron J. 1988;80:388-93.
29. Gauch HG, Zobel RW. AMMI analysis of yield trials. In: Kang MS, Gauch HG Jr, editors. Genotype by environment interactions. Boca Raton (FL, US): CRC Press; 1996. p. 85-122.
30. Purchase JL, Hatting H, Van Deventer CS. Genotype  $\times$  environment interaction of winter wheat (*Triticum aestivum* L.) in South Africa. II, Stability analysis of yield performance. S Afr J Plant Soil. 2000;17:101-7.
31. Gauch H. Model selection and validation for yield trials with interaction. Biometrics. 1988;44:705-15.
32. Kang MS. Simultaneous selection for yield and stability: Consequences for growers. Agron J. 1993;85:754-7.
33. Farshadfar E, Hosein SS, Zali H. Comparison of parametric and non-parametric stability statistics for selecting stable chickpea (*Cicer arietinum* L) genotypes under diverse environments. Aust J Crop Sci. 2012;6(3):1-17.
34. Yan W, Hunt LA, Sheng Q, Szlavniks Z. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci. 2000;40:597-605.
35. Yan W, Kang MS. GGE biplot analysis: A graphical tool for breeders. In: Kang MS, editor. Geneticists and Agronomist. Boca Raton (FL, US): CRC Press; 2003. p. 271.
36. Yan W, Tinker NA. Biplot analysis of multi-environment trial data: Principles and applications. Can J Plant Sci. 2006;86:623-45.
37. Ding M, Tier B, Yan W. Application of GGE biplot analysis to evaluate genotypes (G), environments (E), and  $G \times E$  interaction on *P. radiata*: A case study. In: Proceedings of the Australasian Forestry Genetic Conference Breeding for Wood Quality; 2007 Apr 11-14; Hobart, Australia.
38. Fan XM, Kang MS, Chen H, Zhang Y, Tan J, Xu C. Yield stability of maize hybrids evaluated in multi-environment trials in Yunnan, China. Agron J. 2007;99:220-8.
39. Badu-Apraku B, Akinwale RO. Cultivar evaluation and trait analysis of tropical early maturing maize under *Striga*-infested and *Striga*-free environments. Field Crops Res. 2011;121:186-94.
40. Yan W, Cornelius PL, Crossa J, Hunt LA. Two types of GGE biplots for analysing multi- environment trial data. Crop Sci. 2001;41:656-63.
41. Akintibu ST. Evaluation of stability analysis methods for maize improvement and production in south western Nigeria [Ph.D. dissertation]. Ile-Ife (Nigeria): Hezekiah Oluwasanmi Library, Obafemi Awolowo University; 2019.
42. Yan W, Rajcan I. Biplot evaluation of test sites and trait relations of soybean in Ontario. Crop Sci. 2002;42:11-20.

43. Yan W, Frégeau-Reid JA. Breeding Line Selection Based on Multiple Traits. *Crop Sci.* 2008;48:417-23.
44. Badu-Apraku B, Akinwale RO, Ajala SO, Menkir A, Fakorede MAB, Oyekunle M. Relationships among traits of tropical early maize cultivars in contrasting environments. *Agron J.* 2011;103:717-29.
45. Badu-Apraku B, Fakorede MAB, Oyekunle M, Akinwale RO. Selection of extra-early maize inbreds under low N and drought at flowering and grain-filling for hybrid production. *Maydica.* 2011;56:29-41.
46. Badu-Apraku B, Akinwale RO, Fakorede MAB. Selection of early maturing maize inbred lines for hybrid production using multiple traits under *Striga*-infested and *Striga*-free environments. *Maydica.* 2010;55:261-74.
47. Badu-Apraku B, Akinwale RO, Franco J, Oyekunle M. Assessment of reliability of secondary traits in selecting for improved grain yield in drought and low-nitrogen environments. *Crop Sci.* 2012;52:2050-62.
48. Oyekunle M, Badu-Apraku B. Assessment of interrelationships among grain yield and secondary traits of early-maturing maize inbred lines under drought and well-watered conditions. *Maydica.* 2018;63:1-10.
49. Badu-Apraku B, Akinwale RO, Oyekunle M. Efficiency of secondary traits in selecting for improved grain yield in extra-early maize under *Striga*-infested and *Striga*- free environments. *Plant Breed.* 2014;133(3):373-80. doi: 10.1111/pbr.12163
50. Morris CF, Campbell KG, King GE. Characterization of the end-use quality of soft wheat cultivars from the eastern and western US germplasm 'pools'. *Plant Genet Resour.* 2004;2:59-69. doi: 10.1079/PGR200435
51. Ober ES, Bloa ML, Clark CJA, Royal AK, Jaggard W, Pidgeon KW. Evaluation of physiological traits as indirect selection criteria for drought tolerance in sugar beet. *Field Crop Res.* 2005;91:231-49.
52. Yan W, Hunt LA. Biplot analysis of diallel data. *Crop Sci.* 2002;42:21-30.
53. Badu-Apraku B, Fontem LA, Akinwale RO, Oyekunle M. Biplot analysis of diallel crosses of early maturing tropical yellow maize inbreds in stress and nonstress environments. *Crop Sci.* 2011;51:173-88.
54. Badu-Apraku B, Oyekunle M, Akinwale RO, Lum AF. Combining ability and heterotic groups of early-maturing tropical white maize inbred lines under stress and nonstress environments. *Agron J.* 2011;103:544-57.
55. Kempthorne O. An Introduction to Genetic Statistics. New York (US): John Wiley & Sons; 1957.
56. Badu-Apraku B, Akinwale RO. Biplot analysis of line  $\times$  tester data of maize (*Zea mays* L.) inbred lines under stress and nonstress environments. *Cereal Res Commun.* 2019;47(3):518-30. doi: 10.1556/0806.47.2019.2a
57. Rawlings JC, Thompson DL. Performance level of criterion for the choice of maize testers. *Crop Sci.* 1962;2:217-20.
58. Badu-Apraku B, Fakorede MAB, Akinwale RO. Key challenges in maize breeding in sub-Saharan Africa. Cambridge (UK): Burleigh Dodds Science Publishing Limited; 2017.
59. Yan W. Crop variety trials: Data management and analysis. Hoboken (NJ, US): Wiley-Blackwell; 2014.

60. Annor B, Badu-Apraku B, Nyadanu D, Akromah R, Fakorede MAB. Identifying heterotic groups and testers for hybrid development in early maturing yellow maize (*Zea mays*) for sub-Saharan Africa. Plant Breed. 2020. <https://doi.org/10.1111/pbr.12822>
61. Malik WA, Forkman J, Piepho H. Testing multiplicative terms in AMMI and GGE models for multi-environment trials with replicates. Theor Appl Genet. 2019;32:2087-96.
62. Yan W. Singular-value partitioning in biplot analysis of multi-environment trial data. Agron J. 2002;94:990-6.
63. Yan W, Ma BL. Model diagnosis and GGE biplot analysis. Ottawa (Canada): Eastern cereal and oilseed research centre, agriculture and agri-food Canada; 2006.
64. Yang R, Crossa J, Cornelius PL, Burgueno J. Biplot analysis of genotype  $\times$  environment interaction: Proceed with caution. Crop Sci. 2009;49:1564-76.
65. Yan W, Glover KD, Kang MS. Letter to editor. Crop Sci. 2010;50:1121-3.
66. Crossa J, Cornelius C, Yan W. Biplots of linear-bilinear models for studying crossover genotype  $\times$  environment interaction. Crop Sci. 2002;42:136-44.
67. Badu-Apraku B, Lum AF. The pattern of grain yield response of normal and quality protein maize cultivars in stress and nonstress environments. Agron J. 2010;102:381-94. doi: 10.2134/agronj2009.0229
68. Oyekunle M, Haruna A, Badu-Apraku B, Usman IS, Mani H, Ado SG, et al. Assessment of early-maturing maize hybrids and testing sites using GGE biplot analysis. Crop Sci. 2017;57:2942-50.
69. Wright S. Correlation and causation. J Agric Res. 1921;20:557-85.

How to cite this article:

Badu-Apraku B, Fakorede B, Akinwale R, Annor B, Adewale S, Toyinbo J, et al. Application of the GGE Biplot as a Statistical Tool in the Breeding and Testing of Early and Extra-Early Maturing Maize in Sub-Saharan Africa. Crop Breed Genet Genom. 2020;2(3):e200012. <https://doi.org/10.20900/cbgg20200012>