

*Full Length Research Paper*

# **Evaluation of soybean (*Glycine max* (L.) Merr.) breeding lines and genotype-by-environment interaction across production environments in Southern Africa**

**Christabell Nachilima<sup>1</sup>, Henry Cordoba-Novoa<sup>2</sup>, Valerio Hoyos-Villegas<sup>2</sup>, Godfree Chigeza<sup>1</sup>, Noah Manda<sup>1</sup>, David Lungu<sup>3</sup> and Kelvin Kamfwa<sup>3\*</sup>**

<sup>1</sup>International Institute of Tropical Agriculture (IITA), Lusaka, Zambia.

<sup>2</sup>Department of Plant Science, McGill University, Montreal, Quebec, Canada.

<sup>3</sup>Department of Plant Science, University of Zambia, Lusaka, Zambia.

Received 29 September, 2021; Accepted 4 November, 2021

**Yield is a quantitative trait known to be influenced by changes in the environment in which the crop is grown, suggesting the need to evaluate soybean lines in different growing regions to assess their adaptability and stability. The objective of this study was to evaluate the performance, yield stability, and genotype × environment interaction (G×E) in soybean cultivars, and the informativeness of six environments in Southern Africa. A group of 62 soybean lines was grown and tested in six locations in Malawi and Zambia during the growing seasons in 2018 and 2019 using a 6 x 5 and 5 x 5 alpha lattice design. Genotype plus genotype × environment (GGE) biplots were used to reduce data complexity and analyze genotypes performance and stability; environments discriminativeness and representativeness; and traits associated with yield. The GGE analyses showed a significant effect of genotypes, environments, and G×E. Genotypes with local and wide adaptation and environments with desirable characteristics for testing cultivars were identified. A number of agronomic traits were identified to be positively and negatively associated with soybean yield. These results reinforce the utility of GGE analysis for ranking cultivars and environments in soybean breeding programs in Southern Africa.**

**Key words:** GGE, yield, adaptation, stability, heritability, flowering.

## **INTRODUCTION**

Soybean (*Glycine max* (L.) Merr.) is an important protein and oilseed crop throughout the world. Its high protein content (40 - 42%) makes it suitable and desirable in human diet (Patil et al., 2018). Soybean seeds have an oil content of approximately 18 to 22% (Patil et al., 2018)

making it a major source of both edible and inedible oil products. Its high-quality oil is used in food product manufacturing as well as in the production of inedible products such as caulks, mastics, plastics, and resins. Additionally, when integrated with a rotational system

\*Corresponding author. E-mail: [kelvinkamfwa@gmail.com](mailto:kelvinkamfwa@gmail.com).

with cereals, it enhances the level of nitrogen in the soil (Kumudini, 2010) playing an important role in soil fertility.

Soybean is grown in most parts of the world, with Brazil being the leading producer, followed by the USA and then Argentina. Africa produces 12,537 tons (about 2.2% of the global production) with South Africa being the leading soybean producer followed by Nigeria and then Zambia (FAOSTAT, 2019). Soybean production has gradually been increasing in Zambia making it the leading exporter to southern African countries such as Zimbabwe and Botswana (Abate et al., 2012). However, further growth of Zambia's soybean export market within the southern African region is constrained by its low yields. One of the contributing factors to low yields is the limited availability of varieties that are adapted to local and regional climatic conditions.

Soybean yield is controlled by multiple genes and it is affected by the interactions between the genotype and the environment, in previous studies soybean seed yield has been found to be associated with up to 17 QTL (Han et al., 2012). Varietal stability is a key factor affecting cultivar's response in varying environments. Stability analysis aims at examining the performance of a genotype relative to other genotypes in different environments (Bernardo, 2020). Breeders may require a variety that is stable across different production environments; however, there could be a genotype that may perform well in a specific environment but poorly in other environments. This lack of varietal stability may be challenging for breeding programs driven by the development of soybean varieties adapted to multiple environments in multiple countries such as the soybean breeding program of the International Institute of Tropical Agriculture (IITA). Therefore, it is important to determine the stability of a given genotype in multiple production environments to inform variety release decisions.

To fully express their genetic potential, cultivars require ideal environmental conditions; hence, their performance will vary depending on the production environment (Reynolds et al., 2001). This relative change in performance of cultivars across environments is termed genotype by environment interaction (G×E). G×E reduces the association between phenotypic and genotypic values, which may cause selections from one environment to perform poorly in another. An understanding of G×E in multi-environment trials (MET) is important in identifying locations that are efficient in distinguishing tested genotypes and in providing information on the most representative and descriptive environment.

As G×E complicates the identification of ideal environments, and superior and stable genotypes, multiple data reduction tools such as cluster analysis, joint plots and biplots have been used by breeders (Yan and Tinker, 2006). Biplots allow the identification of associations between genotypes, environments, and their interaction. Biplots include the Additive Main effect and

Multiplicative Interaction (AMMI) and Genotype plus genotype by environment (GGE) analyses. GGE biplots are powerful tools that combine principal component analysis (PCA) and graphical explanation of G×E to identify patterns associated with genotypes and G×E in the evaluation of cultivars in MET (Hoyos-Villegas et al., 2016). GGE allows a better understanding of G×E, and in turn facilitates the identification of representative and discriminative environments (Yan and Tinker, 2006). In contrast, AMMI analysis explain less of the GGE variation compared to GGE biplots, and does not allow an accurate visualization of genotypes in a given environment since the inner-product property is not considered (Hoyos-Villegas et al., 2016; Yan et al., 2007). The objective of this study was to assess the performance, yield stability, G×E in soybean cultivars, and the informativeness of the tested environments in multiple locations in Southern Africa.

## MATERIALS AND METHODS

### Genotypes and evaluation environments

Fifty-five (F7) elite soybean lines developed by the International Institute of Tropical Agriculture (IITA) along with seven commercial checks (released varieties) from seed companies were used to generate the data used in this study. The genotypes being evaluated have an early to medium maturity group with determinate and indeterminate growth habit (Table 1). The multi-environment trials (MET) were carried out at six different locations in Zambia and Malawi. In 2018 season, the locations were Chipata, Seedco-Zambia, International Institute of Tropical Agriculture Southern African Region Administration (IITA-SARAH), and Chitedze (Malawi). In 2019 season, the locations were Bvumbwe (Malawi), Kabwe (Zambia), IITA-SARAH, and Chitedze (Malawi). According to their characteristics, locations were grouped into four environments (Table 2).

### Experiment design and data collection

The soybean genotypes were planted in a 6 × 5 and 5 × 5 alpha lattice design with three replications. Each genotype occupied a plot comprising four rows of 4 m long each, 0.5 m between rows, and 0.05 m between plants. Basal dressing fertilizer (25 kg N/ha, 30 kg K<sub>2</sub>O/ha, 60 kg P<sub>2</sub>O<sub>5</sub>/ha) was applied at planting, and Metolachlor and Imazethapyr were applied as pre-emergence herbicides for control of weeds. Quizalofop-p-ethyl and Fomesafen were applied as post-emergence control of weeds and hand weeding control was also done. Data collected involved grain yield (GY, kg ha<sup>-1</sup>), 100 seed weight (SWT100), plant height (PLHT), days to maturity (DM), days to flowering (DFFL), lodging, and shattering. Lodging was recorded for each plot based on a descriptive scale (Woods and Swearingin, 1977) where 1 = almost all plants erect, 2 = either all plants leaning slightly or a few plants down, 3 = either all plants leaning moderately (<45° angle) or 25-50% of plants down, 4 = either all plants leaning considerably (>45° angle) or 50-80% of plants down, and 5 = all plants down. For shattering the scale proposed by Krisnawati and Muchlish Adie (2017) was used, where 1 = No pod-shattering (Very Resistant); 2 ≤ 25% pod-shattering (Resistant); 3 = 25-50% pod-shattering (Moderately Resistant); 4 = 51-75% pod-shattering (Highly Susceptible); 5 ≥ 75% pod-shattering (Very Highly Susceptible).

**Table 1.** List of the evaluated soybean genotypes.

No.	Genotype	Source	Growth habit <sup>†</sup>	Maturity group <sup>§</sup>	No.	Genotypes <sup>†</sup>	Source	Growth habit <sup>†</sup>	Maturity group <sup>§</sup>
1	Dina	MRI	I	L	32	TGx2001_9DM	IITA	I	M
2	Kafue	IITA	D	E	33	TGx2002_14DM	IITA	D	M
3	Lukanga	ZAMSEED	D	M	34	TGx2002_17DM	IITA	D	M
4	SC_Safari	SEEDCO	I	E	35	TGx2002_23DM	IITA	I	M
5	SC_SAMBA	SEEDCO	D	M	36	TGx2002_35FM	IITA	D	E-M
6	SC_SPIKE	SEEDCO	I	M-L	37	TGx2002_3DM	IITA	D	E
7	SC_SQUIRE	SEEDCO	I	M	38	TGx2002_3FM	IITA	D	E
8	TGx1987_62F	IITA	D	M	39	TGx2002_4DM	IITA	D	E
9	TGx2001_10DM	IITA	I	M	40	TGx2002_5FM	IITA	I	M
10	TGx2001_11DM	IITA	I	M	41	TGx2002_6DM	IITA	I	M
11	TGx2001_13DM	IITA	I	M	42	TGx2002_6FM	IITA	I	M
12	TGx2001_13FM	IITA	I	M	43	TGx2002_7FM	IITA	I	M
13	TGx2001_14DM	IITA	I	M	44	TGx2002_8FM	IITA	I	M
14	TGx2001_14FM	IITA	I	M	45	TGx2002_9FM	IITA	I	M
15	TGx2001_15DM	IITA	I	M	46	TGx2014_15FM	IITA	I	M
16	TGx2001_16DM	IITA	I	M	47	TGx2014_16FM	IITA	D	E-M
17	TGx2001_18DM	IITA	I	M	48	TGx2014_17FM	IITA	I	M
18	TGx2001_18FM	IITA	I	M	49	TGx2014_19FM	IITA	I	M
19	TGx2001_19DM	IITA	I	M	50	TGx2014_21FM	IITA	I	M
20	TGx2001_1DM	IITA	I	M	51	TGx2014_23FM	IITA	I	M
21	TGx2001_20FM	IITA	I	M	52	TGx2014_24FM	IITA	D	E-M
22	TGx2001_22DM	IITA	I	M	53	TGx2014_27FM	IITA	I	M
23	TGx2001_24DM	IITA	I	M	54	TGx2014_31FM	IITA	I	M
24	TGx2001_24FM	IITA	I	M	55	TGx2014_33FM	IITA	I	M
25	TGx2001_26FM	IITA	I	M	56	TGx2014_34FM	IITA	I	M
26	TGx2001_2DM	IITA	I	M	57	TGx2014_38FM	IITA	I	M
27	TGx2001_5DM	IITA	I	M	58	TGx2014_43FM	IITA	I	M
28	TGx2001_5FM	IITA	I	M	59	TGx2014_44FM	IITA	I	M
29	TGx2001_6FM	IITA	I	M	60	TGx2014_4FM	IITA	I	M
30	TGx2001_8DM	IITA	I	M	61	TGx2014_5GM	IITA	I	E-M
31	TGx2001_8FM	IITA	I	M	62	TGx2014_9FM	IITA	I	M

\*I = Indeterminate, D = determinate; <sup>†</sup> TGx (Tropical glycine max); <sup>§</sup>M= Medium maturity (100-111 days), E= Early maturity (85-99 days), L= Late maturity (112-120 days).

## Statistical analysis

### Genotype plus Genotype × Environment biplots and stability analysis

Datasets from Chipata, IITA-SARAH, Chitedze, and Seedco-Zambia in 2018 were combined with datasets from Chitedze, IITA-SARAH, Bvumbwe, and Kabwe in 2019, respectively, to establish four environments (Table 1). Normality was also checked in data using residuals.

Multi-environment data were analyzed to visually identify GE interactions. The package *metan* in R (Olivoto and Lúcio, 2020) was used to calculate ANOVA-based stability statistics, variance components, and broad-sense heritability ( $H^2$ ) in mixed-effect models, and best linear unbiased prediction (BLUP) values for GY, SWT100, PLHT, DM, DFFL, lodging, and shattering, using the model:

$$y_{ijk} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \gamma_k + \varepsilon_{ijk}$$

where  $\mu$  is the grand mean,  $\alpha_i$  is the random effect of the  $i$ th line (genotype),  $\tau_j$  is the random effect of the  $j$ th environment,  $(\alpha\tau)_{ij}$  is the interaction of genotype  $i$  and environment  $j$ ,  $\gamma_k$  is the random effect of the  $k$ th block within the  $j$  environment, and  $\varepsilon_{ijk}$  is the random error. BLUP-adjusted means were obtained by resolving the random effects for each line (Olivoto and Lúcio, 2020).

Different biplots were created using the environment-centered model ( $Y_{ij} - \mu - \beta_j = \alpha_i + \phi_{ij}$ ) where the E main effect ( $\beta_j$ ) is removed and the biplot only considers G ( $\alpha_i$ ) and GE ( $\phi_{ij}$ ) as the relevant sources of variation. The biplot analyses were carried out using the package GGE Biplot GUI in R (Frutos et al., 2014). In the biplots, no scaling was used, except for the genotype by trait biplot where data were scaled with the standard deviation (SD). When evaluating test environments or traits, data were represented using column-preserving, singular-value; whereas, for

**Table 2.** Trial locations and their geographical information.

Environment	Location	Country	Longitude	Latitude	Elevation (mamsl)*	Rainfall (mm)		Temperature			Soil type	
						2018	2019	2018 (min)	2018 (max)	2019 (min)		2019 (min)
E1	Chipata	Zambia	E32°39'	S13°40'	1098	1249	234.46	13	28	12.6	29.089	Loamy sand
	Chitedze	Malawi	E33°38'	S13°59'	1100	929	114.533	12	28.6	13	30	Sand clay
E2	IITA-SARAH	Zambia	E28°30'	S15°30'	1193	703	127.98	12.6	27.98	11	30.88	Red clay loam
E3	Bvumbwe	Malawi	E35°067'	S15°.917'	1146	930	180.28	13	24.67	12.33	27.89	Sand clay
E4	Seedco. Zambia	Zambia	E28°33'	S15°67'	1301	826	138.04	11	29	13	28.46	Red clay loam
	Kabwe	Zambia	E28°4514'	S14°4285'	1182	804	104.517	12.38333	29.11667	13.1	28.95	Sandy loam

the comparison of G and GE, biplots were drawn using row-preserving, singular value partitioning (Yan and Tinker, 2006).

## RESULTS AND DISCUSSION

### Average agronomic performance, analysis of variance and heritability

Grain yield in E1 ranged from 1.89 t/ha (TGx2001-8FM) to 2.86 t/ha (TGx2014-9FM) with an average of 2.33 t/ha. In E2, grain yield ranged from 2.20 t/ha (TGx2001-19DM) to 2.99 t/ha (SC SQUIRE) with an average of 2.56 t/ha. E3 had a grain yield ranging from 1.71 t/ha (TGx2001\_9DM) to 2.69 t/ha (TGx2014\_24FM) with an average of 2.24 t/ha. In E4 grain yield ranged from 2.92 t/ha (TGx2001\_8DM) to 3.58 t/ha (TGx2001\_2DM) with an average performance of 3.21 t/ha. The agronomic performance of traits across environments showed that grain yield ranged from 2.27 t/ha (TGx2001-8FM) to 2.88 t/ha (TGx2014\_24FM) with an average of 2.58 t/ha. Similar findings were reported, in which soybean yield ranged from 1.77

(SC Semeki) to 2.43 t/ha (Lukanga) with an average of 1.95 t/ha (Manda and Maata, 2020). SWT100 ranged from 13.10 g (Dina) to 17.96 g (TGx2001\_14DM) with an average of 14.88 g, PLHT ranged from 67.49 cm (TGx2002\_14DM) to 85.19 cm (TGx2014\_23FM) with an average of 75.35 cm, DM oscillate between 103 days (TGx2014\_31FM) and 118 days (Dina) with an average of 111 days; these results indicate that most of our genotypes have a medium maturity period, following the scale developed by Kawuki et al. (2003), in which early maturing varieties takes 85 to 99 days, medium maturing varieties take 100 to 111 days to maturity and late maturing varieties take 112 to 121 days to maturity, DFFL varied from 45 days (TGx2014\_31FM) to 56 days (TGx1987\_62F) with an average of 51 days, lodging score ranged from 1.82 (TGx2014\_24FM) to 2.41 (TGx2002\_4DM) with an average of 2.09, and shattering score was between 1.17 (Dina) and 2.38 (TGx2002\_6FM) with an average of 1.68.

In the analysis of variance, all the sources of variation had a significant effect ( $p < 0.001$ ) on GY, DM, DFFL, and shattering. The environment did

not have a significant effect on PLHT or lodging. Lodging showed the highest variation (CV = 58.7%) followed by shattering (CV = 38.4%). In contrast, DM and DFFL had the lowest CV with 4.8 and 8.5%, respectively (Table 3).

Table 4 shows the proportion of variance explained by the Environment, Genotype, and G×E interaction, and the broad-sense heritability ( $H^2$ ) for the traits evaluated. Overall, the environment explained a higher proportion of the variance, accounting for 4.56 to 50.03%, followed by G×E (5.03-18.06%), and the genotype (0.54-15.46%). For yield, the environment explained 14.36% of the variability, genotype 1.02%, and the interaction 5.06%. The magnitude of the contribution of the E and G×E are attributed to contrasting conditions between testing environments (Ahakpaz et al., 2021). In the present study, environments in the two countries vary in terms of soil type, precipitation, temperature, and humidity (Table 2), differences that explain the observed effects. Previous studies have also reported a higher contribution of the environment than G×E on soybean yield (Gurmu et al., 2017; Rakshit et al., 2012; Temesgen et al.,

**Table 3.** ANOVA for soybean traits.

SOV	GY		SWT100		PLHT		DM		DFFL		Lodging		Shattering	
	df	Mean Sq.	df	Mean Sq	df	Mean Sq	df	Mean Sq	df	Mean Sq	df	Mean Sq	df	Mean Sq
Environment	3	80532.75***	3	388.3***	3	3980 <sup>ns</sup>	3	2163.9*	3	2991.1***	3	244.41 <sup>ns</sup>	3	34.068***
Rep (ENV)	68	12679.94***	61	22.5 <sup>ns</sup>	61	3806***	59	660.4***	65	79.9***	46	1.63***	61	2.158***
Genotype	61	1284.55***	61	51.4***	61	465***	61	350.3***	61	232.3***	61	1.52***	61	2.096***
G × E	183	860.99***	183	31.9***	183	299***	183	80.6***	183	62.8***	183	1.32***	183	0.833***
Error	1358	497.75	1152	22.8	1021	110	1158	26.5	1286	18.8	748	0.82	1155	0.412
CV (%)		29.5		31.7		14.7		4.8		8.5		40.8		38.4
BLUP-adjusted mean		2.58 t/ha		14.88 g		75.35 cm		111 days		51 days		2.09		1.68

\*, \*\*, and \*\*\* indicate significance at the 0.05, 0.01 and 0.001 probability levels, respectively. ns, not significant at  $p \leq 0.05$ . Mean Sq.: Mean square.

2015; Vaezi et al., 2017) and other traits (Li et al., 2020). Nevertheless, higher G×E contributions (Atnaf et al., 2013; Bhartiya et al., 2018; Mwiinga et al., 2020) or a higher genotype effect (Li et al., 2020) has also been reported.

The  $H^2$  can be divided into high ( $H^2$  greater than 50%), medium ( $H^2$  greater than 20% and less than 50%), and low ( $H^2$  less than 20%) (Sulistyo et al., 2018). According to this, GY, SWT100, and lodging showed a medium heritability (37, 44, and 21% respectively), and PLHT, DM, DFFL, and shattering showed a high  $H^2$  (50-76%; Table 4). For yield-related variables, the  $H^2$  was lower compared to those reported by Li et al. (2020) for soybean MET in China. However, our  $H^2$  values for yield and other variables are consistent with previous reports that showed PLHT, DM, and DFFL having a higher heritability compared to yield in soybean (Bianchi et al., 2020; Li et al., 2020).

### GGE analysis

Although the effect of the environment explained a large percentage of the total variance (Table 4),

significant contributions of G×E to total variation for all the variables evaluated were also identified (Table 3). When selecting stable genotypes and evaluating tester environments, the effect of the genotype and G×E are the main factors to be examined simultaneously (Yan and Kang, 2003; Yan et al., 2007). Here, the GGE analysis allowed to graphically evaluate the environments and genotypes.

### Environment evaluation

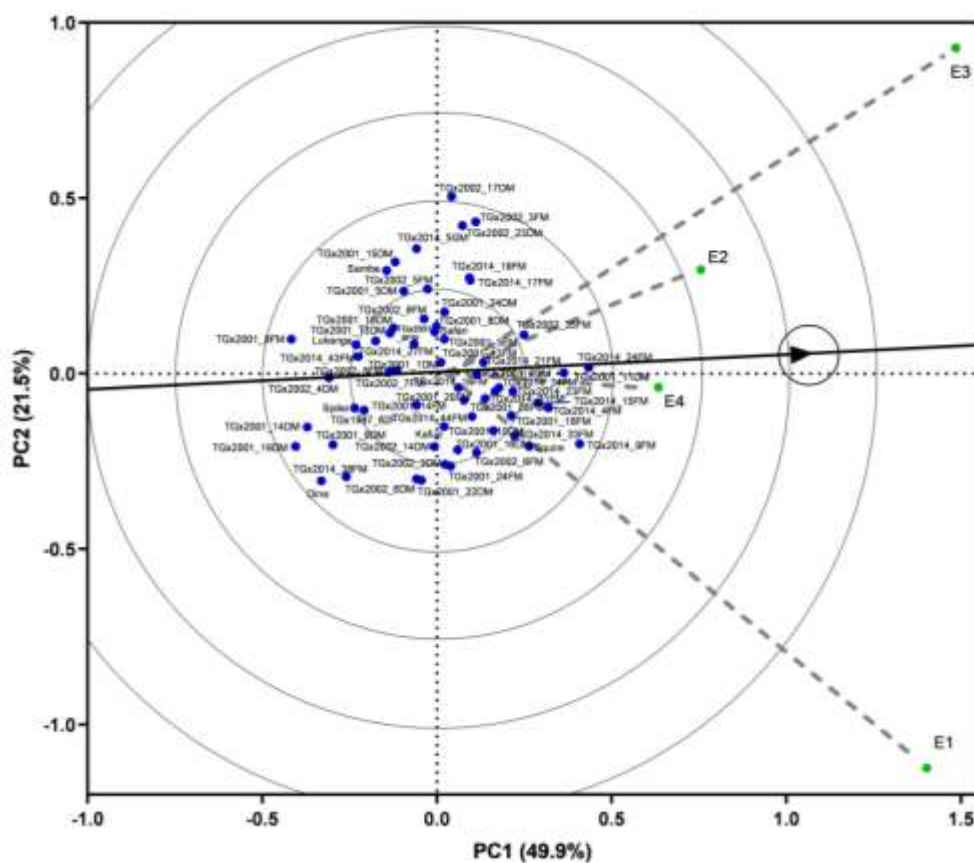
The discriminativeness vs representativeness is a tester-centered (G + GE) and column-preserving biplot that assesses environments. In this biplot, it is possible to identify environments that are similar and could be grouped into mega-environments. Similarly, the biplot provides information about which of the environments is more representative within the trials and which one has the best discriminating ability. Relationships among the four environments can be visualized by analyzing their vectors and the angles formed among them. Acute angles indicate positive correlations and obtuse angles negative

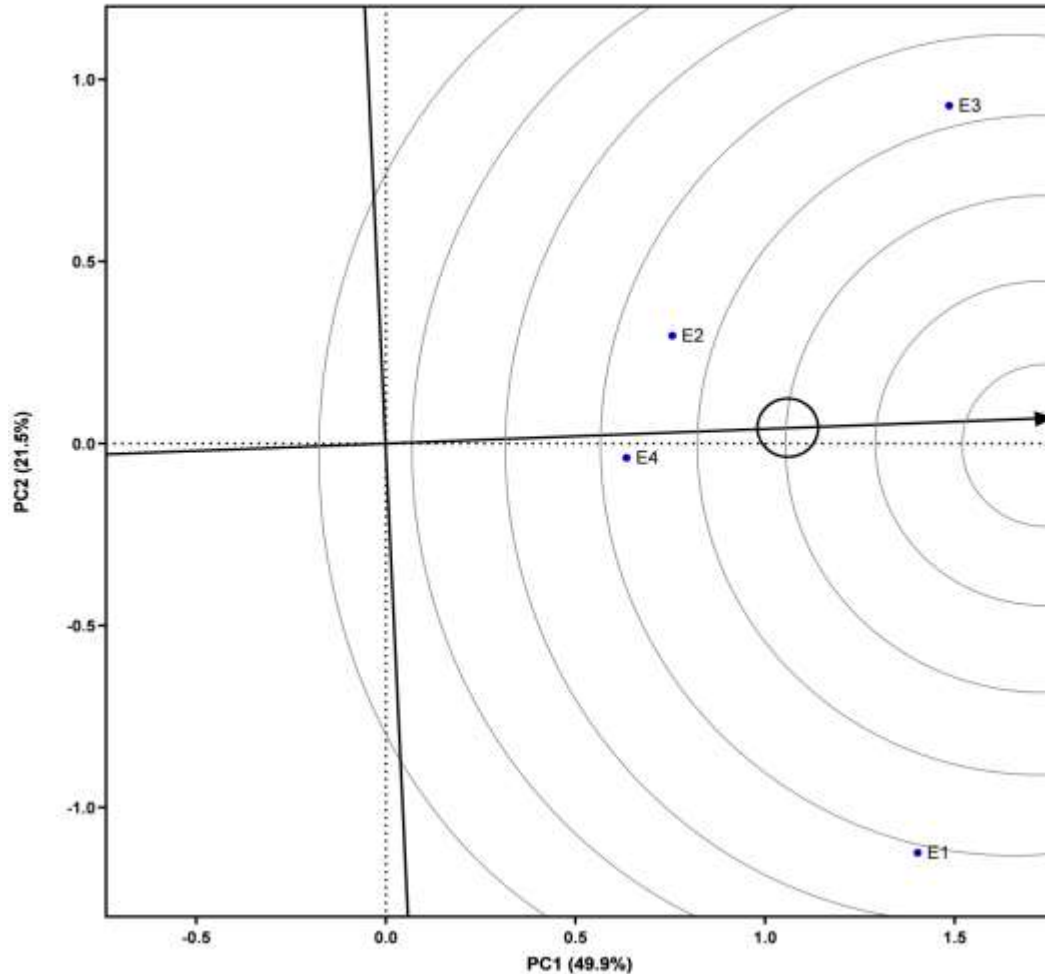
correlations. All the environments showed a positive correlation, where the angles formed among the environment vectors allowed to identify two major groups, one with E2, E3, and E4 and another with E1 (Figure 1). However, E4 is close to being in the middle between E2-E3 and E1. Information on mega-environments allows a better experimental design for the evaluation of multiple genotypes and to determine when locations can be predictable for a given environment (Gauch et al., 2008). As reported here, GGE and other biplot analysis such as AMMI have shown to be efficient to identify mega environments (Ahakpaz et al., 2021; Hoyos-Villegas et al., 2016; Li et al., 2020). The mega-environments identified here allow the use of repeatable G×E in soybean breeding in Southern Africa, and contribute to the crop production by releasing adapted cultivars according to mega-environments (Yan, 2019). These results also demonstrate the utility of GGE biplots to analyze multiple locations and their relevance in the routine MET conducted in breeding programs aimed to release cultivars with wide adaptation.

The concentric circles on the biplot are proportional to the standard deviation and help to

**Table 4.** Proportion of variance explained by the Environment, Genotype, and G×E interaction, and the broad-sense heritability (H<sup>2</sup>) for different traits evaluated in 2018 and 2019 season in Zambia and Malawi.

Trait	Proportion of variance explained (%)			H <sup>2</sup>
	Environment	Genotype	G×E	
Grain yield	14.36	1.02	5.03	0.37
SWT100	4.56	2.34	7.18	0.44
Plant height	15.67	2.20	7.31	0.50
Days to maturity	8.08	10.12	11.44	0.76
Days to flowering	13.24	15.46	18.06	0.75
Lodging	50.03	0.54	5.61	0.21
Shattering	8.53	8.27	10.09	0.71





**Figure 2.** Ranking of environments relative to the “ideal” environment. PC, Principal component; E1, Chipata and Chitedze; E2, IITA-SARAH; E3, Bvumbwe; E4, Seedco.Zambia and Kabwe.

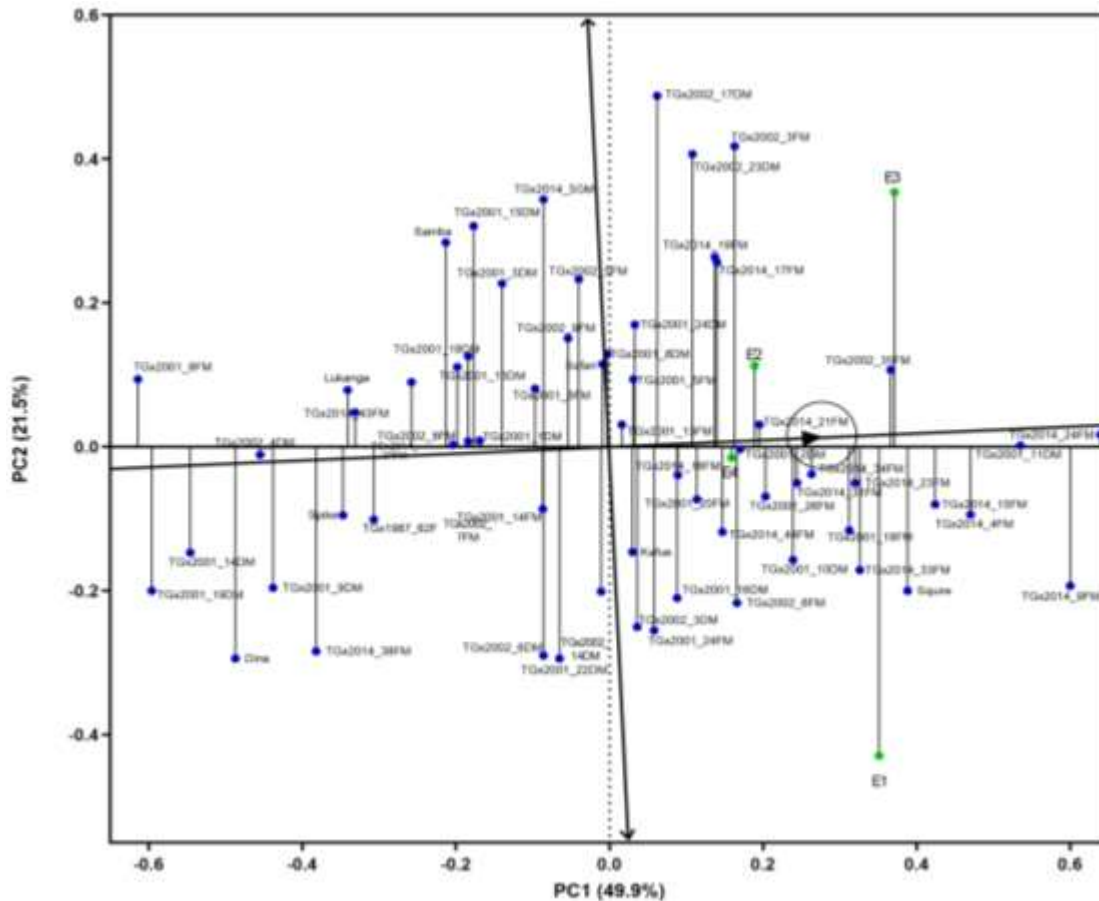
genotypes that have general adaptation. In contrast, discriminating but non-representative environments are valuable to select genotypes with specific adaptation if the environment can be divided into mega-environments or eliminate unstable genotypes if the target environment is a single mega-environment (Yan and Tinker, 2006). Here, the testing environments composed of single environments as described in materials and methods, hence E1 and E3, which were the most discriminating, but were not representative, would be useful environments to identify locally adapted genotypes and potential parents.

The ideal environment biplot ranks the tester environments based on discriminating ability and representativeness considering an “ideal environment”. An ideal environment is that with the highest discriminativeness and representativeness. Figure 2 shows the “ideal environment” with the arrowed line in the center of the concentric circles and it is located on the AEA with a distance from the biplot origin equal to the

longest vector. The E3 was the closest environment to an ideal, followed by E1, E2, and E4. The results are consistent with previous reports in multiple crop species where GGE biplots have been successfully applied to characterize environments that allow a better selection of genotypes and the optimization of resources in breeding programs (Baraki et al., 2020; Luo et al., 2015; Mwiinga et al., 2020).

### Genotype evaluation

The performance of genotypes in each environment can be analyzed by comparing the angles formed among the vectors of genotypes and environments. Acute angles indicate that a genotype performs better than average in a specific environment, obtuse angles suggest that the performance is lower than average, and if a right angle is formed between a genotype and an environment, the performance of the genotype is close to the average in



**Figure 3.** Mean vs stability biplot. PC, principal component; E1, Chipata and Chitedze; E2, IITA-SARAH; E3, Bvumbwe; E4, Seedco Zambia and Kabwe.

that environment (Hoyos-Villegas et al., 2016). TGx2002\_35FM, TGx2014\_24FM, and TGx2001\_11DM yielded better than average in E2, E3, and E4, while TGx2014\_9FM, Squire, TGx2014\_33FM, and TGx2001\_16DM are better adapted to E1 (Figure 1).

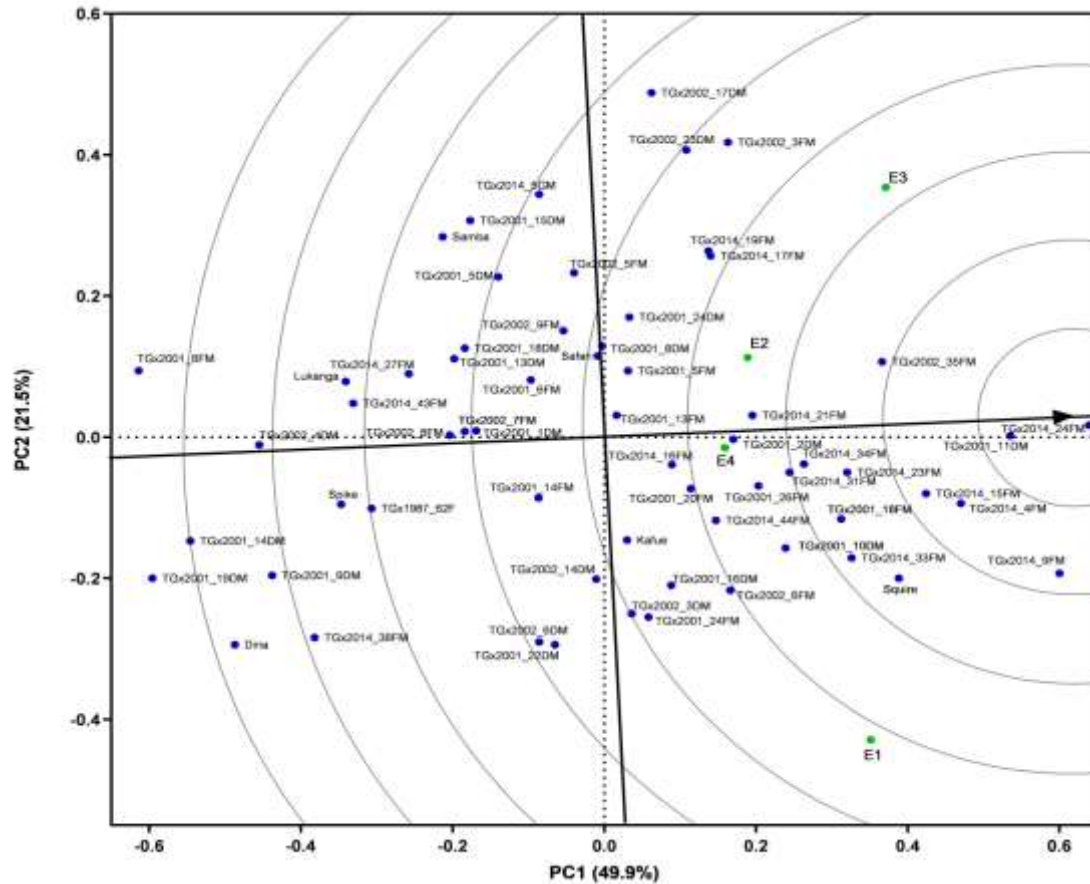
The origin of the biplot is referred to as a “virtual” genotype which has a performance close to the average in all environments (Hoyos-Villegas et al., 2016). Among the evaluated genotypes, TGx2001\_13FM and TGx2014\_16FM were near the biplot origin, which means that these genotypes have a lower contribution to genotype or GE interactions (Figure 1). Genotypes with vectors distal from the biplot origin would have higher contributions to genotype or GE interactions with a better or worse response across all environments depending on the direction of the vectors. TGx2014\_24FM, TGx2002\_17DM, and TGx2014\_9FM had the longest vectors in the same direction as the environments, indicating that these genotypes had the best performance across all environments. In contrast, TGx2001\_19DM, TGx2001\_8FM, Dina, and TGx2001\_14DM had the longest vectors in the opposite direction of the environments and thus, the lowest performance across

E1 – E4. Our results are consistent with those reported by Mwiinga et al. (2020) who evaluated some of the lines used in the present study, and found that TGx2014\_24FM and TGx2002\_17DM showed a positive performance in different environments.

By examining the angle formed between the AEA and vector of the genotype, it is possible to observe if the response of a genotype is mainly due to GE effects. Genotypes such as Safari, TGx2002\_5FM, TGx2014\_5GM, and TGx2002\_14DM showed right angles with the AEA, indicating that their response can be mainly attributed to GE interactions (Figure 1). Regarding the elite genotypes, Dina and Lukanga had lower than average performance in all environments, whereas Kafue had a better performance in E1 and Safari in E2 – E4.

The mean performance and stability biplot is appropriate for genotype evaluation. In this biplot, it is possible to identify the most stable genotype and the one with the highest mean performance. Figure 3 shows the Average-Environment Coordination (AEC) view of the genotype-metric preserving biplot. On the AEC, genotypes that are closer or above the AEC abscissa





**Figure 4.** Ranking of genotypes relative to the “ideal” genotype. PC, principal component; E1, Chipata and Chitedze; E2, IITA-SARAH; E3, Bvumbwe; E4, Seedco.Zambia and Kafue.

(single-headed line) in the same direction of environments, are expected to have the highest mean performance. In contrast, genotypes located farther from the AEC abscissa arrowhead would have the lowest performance. The genotypes TGx2014\_24FM, TGx2014\_9FM, TGx2001\_11DM, and TGx2014\_4FM had the highest mean performance, and TGx2001\_8FM, TGx2001\_19DM, and TGx2001\_14DM the lowest (Figure 3).

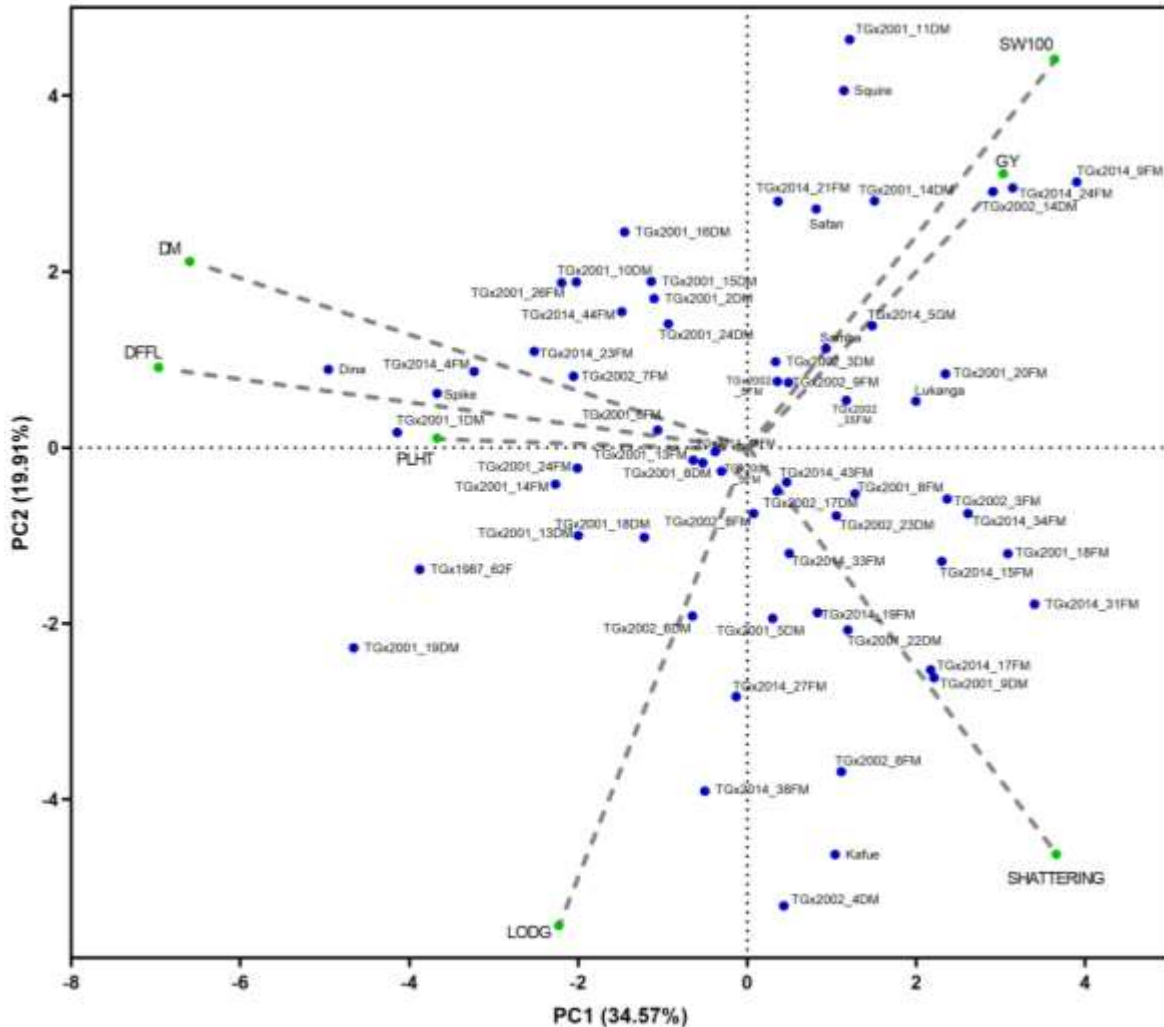
The AEC ordinate on axis y (double-headed line) points to poorer stability in either direction and genotypes near the AEC ordinate have a mean yield similar to the grand mean. Longer vectors correspond to genotypes with poorer stability and genotypes with short vectors are expected to be more stable. TGx2002\_17DM, TGx2002\_3FM, and TGx2002\_23DM were found to be the genotypes with the greatest variability, compared to TGx2002\_4DM, TGx2002\_8FM, TGx2001\_11DM, and TGx2014\_24FM that were highly stable. Similar results have been reported by Mwiinga et al. (2020) who found, for instance, that TGx2002\_17DM, had the lowest yields in some environments and was the winner in some of the environments evaluated. In this way, our results further

support previous findings and highlighting how variable genotypes can be successfully identified by the analyses performed here.

Six genotypes, including TGx2002\_5FM, Safari, TGx2001\_13FM, Kafue, TGx2002\_14DM, and TGx2001\_24FM performed at the grand mean (Figure 3). It is worth noting that the two first principal components explained 71.4% of the variation, which may not reflect the actual stability for some of the genotypes (Hoyos-Villegas et al., 2016). However, the variability captured in our analysis is slightly higher than previously reported in a similar experiments using AMMI biplots (Mwiinga et al., 2020).

Based on the performance and the stability, genotypes can be ranked relative to an “ideal” genotype, which is that with the highest performance and the lowest variability. Among the genotypes, TGx2014\_24FM was the closest to an ideal genotype, followed by TGx2001\_11DM (Figure 4). Therefore, TGx2014\_24FM and TGx2001\_11DM could be considered for release in multiple environments. Interestingly, the elite lines (Dina, Kafue, Lukanga, and Safari) showed yields lower or close to the grand mean and were considerably unstable, being





**Figure 6.** Genotype by trait biplot representing 62 genotypes evaluated for seven traits. DFFL, days to flowering; DM, days to maturity; PLHT, plant height; GY, grain yield; SW100, 100 seed weight; LODG, lodging; PC, principal component.

complexity, low heritability, and the effect of G×E interactions (Bianchi et al., 2020). Besides dissecting yield by decomposing the variance attributable to E, G and G×E, and the analysis of the G×E through GGE analysis, the identification of correlated variables can contribute to improving single yield components, and then the overall yield performance (Shahriari et al., 2018). Furthermore, strongly correlated variables with high heritability could help to indirectly select for yield and circumvent challenges of direct selection.

Multiple traits averaged across all environments can be correlated with genotypes using genotype × trait biplots. These biplots allow to visually recognize relationships between variables, genotypes, and their interactions; being possible to identify traits that are positively or negatively associated, traits that could be redundant or that can be used as indirect indicators of other

measurements (Yan and Tinker, 2006). Following the same approach in angle formation used for environment and genotype evaluation, PLHT, DM, and DFFL were positively correlated to each other and were negatively correlated to shattering (Figure 6). Regarding lodging, DFFL, PLHT, and shattering were positively correlated, but DM was independent. Yield variables (GY and SW100) were grouped in the biplot and showed a strong negative relation with lodging; and a slight negative correlation with shattering, PLHT, DM and DFFL. Despite GY and SW100 had medium  $H^2$ , correlated traits such as PLHT, DM, and DFFL had high  $H^2$ , reinforcing that they could be used for indirect selection for yield. Our results complement previous studies on soybean breeding programs in Southern Africa (Mwiinga et al., 2020), where the correlation between yield and other traits was not evaluated. Interestingly, the negative

correlation found between yield and PLHT differs from previous results in soybean where selection for plants with higher plant height improved overall yield (Li et al., 2020), which could be attributed to the adaptation of breeding lines to the environmental conditions in Southern Africa. In the cultivar development, further MET analyses aimed to address these correlations are necessary to incorporate them as indirect selection strategies. Considering associated variables can simplify breeding programs and yield improvement as redundant measurements, variables that may be easier to assess, and indirect sources of information can be identified, which makes genotype  $\times$  trait analysis a valuable decision-making tool for soybean breeding (Hoyos-Villegas et al., 2016; Li et al., 2020).

Regarding genotypes, TGx2014\_24FM, TGx2002\_14DM, and TGx2014\_9FM had the highest GY and SW100, and the lowest lodging, these could be used as potential parents for improving yield and reduce lodging score. On the other hand, TGx2001\_11DM and Squire, which had considerable yield performance, had a lower shattering score, which did not have an apparent effect on TGx2014\_24FM and TGx2002\_14DM. Among the checking lines, Safari and Lukanga showed a higher GY and SW100. These findings are consistent with previous studies where, as a checking line, Lukanga had shown to have a higher yield response to different environmental factors such as population density. Lukanga and Kafue had a higher shattering score and Dina had a higher PLHT, DM, and DFFL, accompanied by a lower GY, SW100, and shattering.

## Conclusions

In the present study, GGE analysis revealed that the yields of the 62 soybean genotypes tested were significantly affected by genotype, environment, and genotype by environment. E1 (Chipata and Chitedze) and E3 (Bvumbwe) were identified as the most discriminating environments, and E4 (Seedco.Zambia and Kabwe) as a highly representative environment. TGx2014\_24FM was identified as the most ideal genotype, which also had the highest yield across environments (2.88 t/ha), followed by TGx2001\_11DM (2.86 t/ha). Through genotype  $\times$  trait analysis, it was identified that yield was positively associated with plant height, days to flowering, and days to maturity, negatively associated with lodging, and independent of shattering. Our results support the used GGE analysis as an analysis tool to dissect complex traits largely affected by G $\times$ E such as yield by decomposing its variance and by identifying related variables.

## CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

## ACKNOWLEDGEMENTS

The authors would like to thank the International Institute of Tropical Agriculture IITA-Zambia through the soybean breeding program for the financial support and providing experimental lines. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the IITA.

## ABBREVIATIONS

AMMI, Additive Main effect and Multiplicative Interaction; GGE, Genotype plus genotype by environment; G  $\times$  E, genotype by environment interaction

## REFERENCES

- Abate T, Alene AD, Bergvinson D, Shiferaw B, Silim S, Orr A, Asfaw S (2012). Tropical grain legumes in Africa and South Asia: knowledge and opportunities. Nairobi, Kenya: International Crops Research Institute for the Semi-Arid Tropics. 112 pp.
- Ahakupaz F, Abdi H, Neyestani E, Hesami A, Mohammadi B, Mahmoudi KN, Abedi-Asl G, Noshabadi MRJ, Ahakupaz F, Alipour H (2021). Genotype-by-environment interaction analysis for grain yield of barley genotypes under dryland conditions and the role of monthly rainfall. *Agricultural Water Management* 245(12):2020.
- Atnaf M, Kidane S, Abadi S, Fisha Z (2013). GGE biplots to analyze soybean multi-environment yield trial data in north Western Ethiopia. *Journal of Plant Breeding and Crop Science* 5(12):245-254.
- Baraki F, Gebregergis Z, Belay Y, Berhe M, Teame G, Hassen M, Gebremedhin Z, Abadi A, Negash W, Atsbeha A, Araya G (2020). Multivariate analysis for yield and yield-related traits of sesame (*Sesamum indicum* L.) genotypes. *Heliyon* 6(10):e05295. <https://doi.org/10.1016/j.heliyon.2020.e05295>
- Bernardo R (2020). Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity* 125:6. <https://doi.org/10.1038/s41437-020-0312-1>
- Bhartiya A, Aditya JP, Kumari V, Kishore N, Purwar JP, Agrawal A, Kant L (2018). Stability analysis of soybean [*Glycine max* (L.) Merrill] genotypes under multi-environments rainfed condition of North Western Himalayan hills. *Indian Society of Genetics and Plant Breeding* 78(3):342–347.
- Bianchi MC, Bruzi AT, Soares IO, Ribeiro FDO, Gesteira GDS (2020). Heritability and the genotype  $\times$  environment interaction in soybean. *Agrosystems, Geosciences and Environment* 3(1):1-10. <https://doi.org/10.1002/agg2.20020>
- FAOSTAT (2019). Agriculture Organization of the United Nations Statistics Division (2018). Production Available in: <http://faostat3.fao.org/browse/Q/QC/S> [Review date: June 2018].
- Frutos E, Galindo MP, Leiva V (2014). An interactive biplot implementation in R for modeling genotype-by-environment interaction. *Stochastic Environmental Research and Risk Assessment* 28(7):1629-1641. <https://doi.org/10.1007/s00477-013-0821-z>
- Gauch HG, Piepho HP, Annicchiarico P (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science* 48(3):866-889. <https://doi.org/10.2135/cropsci2007.09.0513>
- Gurmu F, Hussein S, Laing M (2017). Genotype-by-environment interaction and stability of sweetpotato genotypes for root dry matter,  $\beta$ -carotene and fresh root yield. *Open Agriculture* 2(1):473-485. <https://doi.org/10.1515/opag-2017-0052>
- Hoyos-Villegas V, O'Connor JR, Heslop AD, Hilditch A, Jahufer MZZ, Barrett BA (2019). Rate of genetic gain for persistence to grazing and dry matter yield in white clover across 90 years of cultivar development. *Crop Science* 59(2):537-552.

- <https://doi.org/10.2135/cropsci2018.07.0471>
- Hoyos-Villegas V, Wright EM, Kelly JD (2016). GGE biplot analysis of yield associations with root traits in a mesoamerican bean diversity panel. *Crop Science* 56(3):1081-1094.
- Kawuki RS, Adipala E, Tukamuhabwa P (2003). Yield Loss Associated with Soya Bean Rust (*Phakopsora pachyrhizi* Syd.) in Uganda. *Journal of Phytopathology* 151(1):7-12.
- Krisnawati A, Muchlish Adie M (2017). Characterization and performance of agronomic characters of soybean genotypes resistant to pod shattering. *Biodiversitas* 18(3):1158-1164.
- Kumudini S (2010). Soybean growth and development. *The Soybean: Botany, Production and Uses* pp. 48-73.
- Li M, Liu Y, Wang C, Yang X, Li D, Zhang X, Xu C, Zhang Y, Li W, Zhao L (2020). Identification of Traits Contributing to High and Stable Yields in Different Soybean Varieties Across Three Chinese Latitudes. *Frontiers in Plant Science* 10(1):1-14.
- Luo J, Pan YB, Que Y, Zhang H, Grisham MP, Xu L (2015). Biplot evaluation of test environments and identification of mega-environment for sugarcane cultivars in China. *Scientific Reports* 5(9):1-11. <https://doi.org/10.1038/srep15505>
- Manda N, Maata M (2020). Responses of soybeans (*Glycine max* (L.) Merrill) associated with variable plant density stress applied at different phenological stages: Plasticity or elasticity? *African Journal of Biotechnology* 19(6):307-319.
- Merrick LF, Lyon SR, Balow KA, Murphy KM, Jones SS, Carter AH (2020). Utilization of evolutionary plant breeding increases stability and adaptation of winter wheat across diverse precipitation zones. *Sustainability (Switzerland)* 12(22):1-23.
- Mwiinga B, Sibiya J, Kondwakwenda A, Musvosvi C, Chigeza G (2020). Genotype x environment interaction analysis of soybean (*Glycine max* (L.) Merrill) grain yield across production environments in Southern Africa. *Field Crops Research* 256(5):107922. <https://doi.org/10.1016/j.fcr.2020.107922>
- Olivoto T, Lúcio ADC (2020). metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution* 11(6):783-789. <https://doi.org/10.1111/2041-210X.13384>
- Patil G, Vuong TD, Kale S, Valliyodan B, Deshmukh R, Zhu C, Wu X, Bai Y, Yungbluth D, Lu Y, Kumpatla S, Shannon JG, Varshney RK, Nguyen HT (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal* 16(11):1939-1953.
- Rakshit S, Ganapathy KN, Gomashe SS, Rathore A, Ghorade RB, Kumar MVN, Ganesmurthy K, Jain SK, Kamtar MY, Sachan JS, Ambekar SS, Ranwa BR, Kanawade DG, Balusamy M, Kadam D, Sarkar A, Tonapi VA, Patil JV (2012). GGE biplot analysis to evaluate genotype, environment and their interactions in sorghum multi-location data. *Euphytica* 185(3):465-479.
- Reynolds MP, Calderini DF, Condon AG, Rajaram S (2001). Physiological basis of yield gains in wheat associated with the Lr19 translocation from *Agropyron elongatum*. *Euphytica* 119(1):137-141.
- Shahriari Z, Heidari B, Dadkhodaie A (2018). Dissection of genotype x environment interactions for mucilage and seed yield in *Plantago* species: Application of AMMI and GGE biplot analyses. In *PLoS ONE* 13(5):e0196095. <https://doi.org/10.1371/journal.pone.0196095>
- Sulistyo A, Purwantoro, Sari KP (2018). Correlation, path analysis and heritability estimation for agronomic traits contribute to yield on soybean. *IOP Conference Series: Earth and Environmental Science* 102(1). <https://doi.org/10.1088/1755-1315/102/1/012034>
- Temesgen T, Kenei G, Sefera T, Jars M (2015). Yield stability and relationships among stability parameters in faba bean (*Vicia faba* L.) genotypes. *Crop Journal* 3(3):258-268.
- Vaezi B, Pour-Aboughadareh A, Mohammadi R, Armion M, Mehraban A, Hossein-Pour T, Dorii M (2017). GGE Biplot and AMMI analysis of barley yield performance in Iran. *Cereal Research Communications* 45(3):500-511.
- Woods SJ, Swearingin ML (1977). Influence of Simulated Early Lodging upon Soybean Seed Yield and its Components. *Agronomy Journal* 69(2):239-242.
- Yan W (2019). LG biplot: a graphical method for mega-environment investigation using existing crop variety trial data. *Scientific Reports* 9(1):1-8. <https://doi.org/10.1038/s41598-019-43683-9>
- Yan W, Kang MS (2003). *GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists*. CRC Press.
- Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop Science* 47(2):643-655. <https://doi.org/10.2135/cropsci2006.06.0374>
- Yan W, Tinker NA (2006). *Biplot analysis of multi-environment trial data: Principles and applications*. *Canadian Journal of Plant Science* 86(3):623-645.