Genotype plus genotype-by-environment interaction analysis of soybean (Glycine max (L. Merrill) across production environments in Southern Africa

CHRISTABELL NACHILIMA

A Dissertation Submitted to the School of Agricultural Sciences in Partial Fulfilment of the Requirements for the Award of a Degree of Master of Science in Plant breeding and seed system.

SCHOOL OF AGRICULTURAL SCIENCES,
DEPARTMENT OF PLANT SCIENCES

The University of Zambia

LUSAKA

April 22, 2021

“*I dedicate this thesis to my loving parents Petronella Sichilima and Greenwell Sichilima, and my son Kenny Kisu Saviye, with lots of love.*”

Contents

Acknowledgments

First and foremost, I want to thank the Almighty God for his wisdom, provision and for giving me the resilience during the course of my research. I also want to thank both the Ministry of Higher Education for the scholarship and the International Institute for Tropical Agriculture (IITA) for the financial support towards my research. Indeed, this was a great opportunity and I will forever be grateful. Special thanks goes to my research supervisor Dr. Kelvin Kamfwa for providing invaluable guidance and encouragement throughout this research. And also Dr. Godfree Chigeza my mentor and supervisor for the thorough, helpful suggestions, and comments made during the course of the study, deeply grateful to Dr. Chikoye for his support during my research, my special thanks goes to Selita Zulu, Sunfold Simachulu and Maxwell Siachunza for their assistant during data collection and trial management. I am extending my heartfelt thanks to the following people Emmanuel Ngonga, Henry Cordoba, Dr. Valerio Hoyos-Villegas, Prisca Chileshe, Noah Manda and Turnbull Chama for their positive contributions.

Last but not the least, great gratitude goes to my parents and my brothers and sisters for the care, love, prayers and their emotional support. They sacrificed a lot for me to succeed in my education, I will always remain indebted to them. I also want to thank my husband Kenny Saviye for showing keen interest and for being a constant source of inspiration and encouragement.

Declaration

I, Christabell Nachilima, declare that this dissertation titled, ‘Genotype and genotype-by-environment interaction analysis of soybean (Glycine max (L. Merrill) across production environments in Southern Africa’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a Master of Science degree in Plant breeding and seed system.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my work.
- I have acknowledged all primary sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed.

Signed: ______________________________

Date: ______________________________

Approval

This dissertation of Christabell Nachilima, has been approved as partial fulfilment of the requirements for the award of Master of Science in Agronomy by the University of Zambia.

Examiner's name	**Signature**	**Date**
....	....	....
....	....	....
....	....	....

List of Tables

List of Figures

Abstract

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 were to: (i) Determine yield stability and Genotype by Environment Interaction of soybean lines across testing environments in Zambia and Malawi using GGE biplots, (ii) Estimate broad-sense heritability for phenology, lodging, shattering, hundred seed weight (HSW) and seed yield, and (ii) Understand multiple trait associations among soybean lines grown in six environments in Malawi and Zambia. Sixty two soybean genotypes were evaluated in six locations in 2018 and 2019 growing seasons, using a 6 x 5 and 5 x 5 alpha lattice design, the collected data was analysed using a GGE biplot.. The additive main effect and multiplicative interaction model (AMMI) indicated that environments, genotypes and GEI significantly ($P<0.001$) affected grain yield and contributed 14.4 %, 1.0% and 5.1 %, respectively, to the total variation. Whereas E4 (Kabwe and SEEDCO.Zambia) was the most representative environment, E1 (Chipata and Chitedze) and E3 (Bvumbwe) were the most discriminating. Thus, E1 and E3 could be used to test for local adaption and selection of Parents. Bvumbwe in Malawi (E3) was the most informative test environment hence it is ideal for selecting broadly adapted genotypes. Out of the 62 genotypes that were tested `TGx2002_4DM`, `TGx2002_8FM`, `TGx2001_11DM`, and `TGx2014_24FM` were highly stable across the tested environments in Malawi and Zambia. Genotype TGX 2014-24FM was identified as the ideal genotype with high yield mean performance and high stability. Therefore, it could be recommended for cultivar release in multiple environments. Plant height, days to maturity, and days to flowering which were positively correlated to yield, had higher heritability of 50%, 75%, and 76%, respectively, than yield suggesting that they could be used to indirectly select for yield

Chapter 1

Introduction

Soybean (Glycine max L.) is an important protein and oilseed crop cultivated throughout the world. Its high protein content (40 - 42%) makes it suitable and desirable in human diet (Antalina et al., 1999). Soybean seeds have an oil content of approximately 18–22% (Patil, 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, soybean plays an important role in soil fertility through symbiotic nitrogen fixation. When integrated in a rotational system with cereals it enhances the level of nitrogen in the soil (Kumudini, 2010). Soybean is grown in most parts of the world, with Brazil being the leading producer, followed by USA and then Argentina. Africa produces about 1% of the global production with South Africa being the leading soybean producer followed by Nigeria and then Zambia. (Figure 1) (FAOSTAT, 2019).

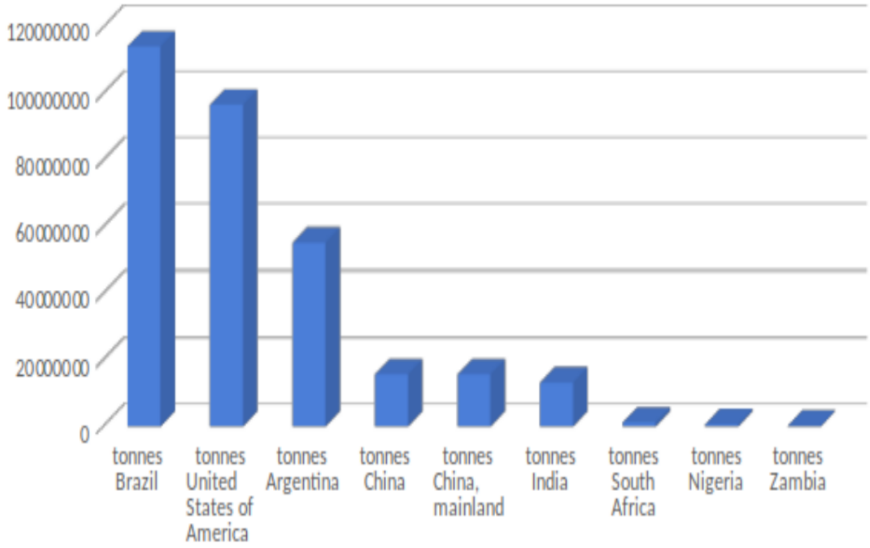

Figure 1.1: Global soybean production for 2019. Source: Data from FAOSTAT, (2019).

Soybean production has gradually been increasing in Zambia making it the leading exporter to southern Africa 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 limited availability of varieties that are adapted to local and regional climatic conditions.

Soybean cultivar productivity depends on its genetics, environment in which it is being grown and interaction of these two factors. The environment includes attributes such as temperature, rainfall, soil type, soil nutrients, and cultural practices, which have an impact on soybean cultivar productivity. For cultivars to fully express their genetic potential they require ideal environmental conditions, hence cultivar performance will vary depending on production environment (Reynolds et al., 2001). This relative change in performance of cultivars across environments is termed as genotype by environment interaction (GEI). GEI reduces the association between phenotypic and genotypic values and may cause selections from one environment to perform poorly in another. Measuring GEI is important to determining an optimum breeding strategy for releasing genotypes with adequate adaptation to target environments (Mwiinga, 2020). For breeders, the agro-ecological diversity of environments where most field crops are grown represents a 'double-edged sword'. On one hand diversity complicates breeding and testing of improved genotypes with adequate adaptation, on the other hand it permits identification of extreme environmental conditions that guarantee selection pressure from important stresses. An understanding of genotype by environment interaction (GEI) in multi-locational yield trials (MLYT) is important in identifying locations that are efficient in distinguishing tested genotypes and in providing information on the most representative and descriptive environment. The other concept of interest in plant breeding is varietal stability. Stability analysis aims at examining the performance of genotype relative to other genotypes in different enviroments (Bernado, 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 an organization such as the International Institute of Tropical Agriculture (IITA) that has a regional or global breeding mandate of developing soybean varieties adapted to multiple environments in multiple countries. Therefore, it is important to determine the stability of a given genotype in multiple production environments to inform variety release decisions.

Breeding for genetically complex traits such as seed yield, which generally has low heritability requires multiple testing sites. The more testing sites a breeding program has the higher the chances of making significant gains in breeding for yield. However, maintaining large number of testing sites is challenging because of resource constraints. It is important, therefore, for an organization such as IITA to identify fewer number of testing sites that are descriptive and representative (Ideal) of many environments.

Direct selection for yield is challenging because of its low heritability (citation). Various agronomic traits in soybean such as days to flowering, maturity, lodging, shattering and plant height affects yield. These agronomic traits have high heritability and could potentially be used to indirectly select for yield. It is important, therefore, to measure heritability of various agronomic traits and to identify those stongly correlated with yield.

Therefore, the objectives of this study were to:

- Determine yield stability and Genotype by Environment Interaction of soybean lines across testing environments in Malawi and Zambia using GGE biplots,
- Estimate broad-sense heritabilities for phenology, lodging, shattering, hundred seed weight (HSW) and seed yield
- Understand multiple trait associations among soybean lines across test enviroments in Malawi and Zambia.

Hypotheses:

1. Genotypes with stable seed yield accross testing environments in Malawi and Zambia exists.
2. Phenology, shattering, HSW, lodging have higher heritability than seed yield.
3. Phenology, shattering, lodging, and HSW are strongly associated with seed yield.

Chapter 2

Literature Review

2.1 Importance of Soybean

Soybean today is one of the most important and extensively grown crops in the world (Hungria et al). It is one of the most valued oilseed crops in the world. As per latest data available, soybean accounts for 36.65million of oil (FAO, 2009), It accounts for 29.7% of the world's processed vegetable oil and is a rich source of dietary protein both for the human diet and for the chicken and pork industries (Graham and Vance, 2003), putting it far ahead of all other field crops raised for oil extraction. Although its cultivation dates back to more than 5000 years ago in China, it came to prominence only during the last 200 years. It has been cultivated for varying purposes during different periods of history in different parts of the world. Its earlier uses have varied from a green manure crop to a forage crop and a nitrogen-fixing crop due to its ability to fix substantial quantities of atmospheric nitrogen in association with nodule-forming bacteria (Bradyrhizobium). The presence of Isoflavones in soybean may reduce the risks of cancer and lower serum cholesterol (Molteni et al., 1995; Kennedy, 1995). Soybean is also used as a milk substitute in weaning foods and baking and in ink and biodiesel fuels (Anonymous, 2000; 2001; Graham and Vance, 2003). A number of characteristics of soybean have made it an attractive crop in the globalized context. Soybean has low water content, high nutritive value, and the capacity to yield a variety of products (e.g., human food, animal food, oil, and industrial derivates (Burton, 2013). These characteristics reduce vulnerability to market fluctuations, reduce storage and transportation costs, and have contributed to its rapid expansion. Furthermore, the increasing global demand for soybean products has resulted in large investments in research and development and the widespread use of transgenic cultivars that have increased yields, reduced costs by reducing herbicide use, and increased the range of appropriate planting environments (Abate, 2012).

2.2 Botany

Soybean is an erect, annual plant that has green leaves covered with fine hairs. The first leaves are simple and grow opposite each other on the stem while the leaves that form subsequently are trifoliate. They have small flowers that consist of five separate; unequal petals that can vary in colour but are commonly violet or white. However, the morphology is diverse depending on the cultivar (Johnson and Bernard, 1963). The flowers and lateral branches form at the auxiliary buds at the point of contact between the leaf petiole and the main stem. The height of the plants can range from about 0.3 to 3.0 m.

The seed is made up of two parts, the seed coat which covers and protects the embryo and two cotyledons which form part of the embryo region. The bean is attached to the pod at the hilum (Kumudini, 2010). Soybean seeds occur in various sizes, and in many seed coat colours, including black, brown, blue, yellow, green and mottled. Varieties differ in hilum colour and can be yellow, imperfect yellow, grey, buff, brown, black or imperfect black. Yellow hilum/ clear hilum soybeans with large seed size and thin but strong seed coat that is free from cracking and discolouration are preferred (Gandhi, 2009). However, the yellow and green seeds are more common. Soybean pods are straight and sometimes slightly curved, reaching 20-70 mm long depending on the cultivar and environment, and they also form in clusters of 1 - 9. Young pods are green in colour, covered in fine transparent hairs and when they mature, they are also hairy and range in colours such as brown or tan, black and yellow. This colour change happens as the plant's leaves turn yellow and fall off. The pods may contain 1-4 seeds depending on the cultivar (Krisnawati and Adie, 2015).

The root system of soybean consists of a taproot, which can grow up to 1.2 m into the soil. Furthermore, there is a proliferation of secondary roots that are arranged in four rows along the taproot. Most of the effective roots are found in the top 600 mm of soil, therefore the soybean plant is a shallow feeder (Kumudini, 2010).

2.3 Soybean Production Constraints

Soybean is grown in most parts of the world, Brazil being the leading soybean producer, followed by USA, Argentina, China, and India. South Africa is the leading soybean producer in Africa followed by Nigeria and then Zambia (FAOSTAT, 2019). The challenges growers face in crop production, include unpredictable weather, diseases, pests, weeds and variable soil quality were previously reviewed (Lal 2009; Strange and Scott 2005). Soybean is affected by all of these variables. Researchers and growers work together on each of these challenges to improve and ensure the quantity and quality of soybean production. Some strategies for increasing yields include the use of fertilizers and pesticides, while others involve developing new plant varieties that best suit the needs of the farmers. In soybean, localized variety development is important so that growers use varieties that are well adapted to local conditions such as weather, preferred agronomic practices and photoperiod (Panthee 2010). However, increases in crop production due to varietal improvements are often offset by constraints caused by broadly categorized abiotic and/or biotic factors. Abiotic constraints Abiotic constraints affecting soybean production are those caused by the physical environment. This includes weather related phenomena, soil nutrient availability, salinity, and response to photoperiod. Farming practices may control some of these abiotic constraints, but many, such as drought, flooding, and frost, have few if any remedies. In more arid climates, drought can reduce both vegetative growth and time to maturation, causing fewer pods to form with fewer and/or smaller seeds per pod (Gibson, 1996).

2.4 Genotype by environment interaction and stability

Genotype-by-environment interaction is said to be present when different cultivars or genotypes respond differently to diverse environments, and for GE to be detected via statistical procedures, at least two genotypes (cultivars) must be evaluated in at least two environments. A basic model that includes GE interaction is P = G + E + GE. It follows from this model that for a given genotype, there can be many phenotypes depending upon the environment. The interactions of genotypes with environments (G x E) makes it difficult for breeders to identify the best genotypes, be it during selection or for cultivar recommendation Simmonds (1981). The presence of interactions indicates that the relative genotype performance in the tests depends essentially on the given environmental conditions. The phenotypic response of any genotype in relation to others could therefore be inconsistent, which is demonstrated by changes of the relative position of the genotypes from one environment to another. The GxE interaction can be partitioned in studies on the adaptability and phenotypic stability. Adaptability is the capacity a genotype has to make use of the environmental effects to warrant a high yield level; on the other hand stability is related with the yield maintenance or yield predictability in the diverse environments (Borém 1998).

2.5 Methods for analysing GEI and stability

There are various methodologies of analysis of G x E and stability designed to evaluate a genotype group tested in a series of environments. Among these the most widely used are the ones based on linear regression (Finlay and Wilkinson 1963, Eberhart and Russel 1966, Verma et al. 1978, Cruz et al. 1989), and a more recent application method called AMMI analysis (Additive Main effects and Multiplicative Interaction analysis) that combines a univariate method for the additive effects of genotypes and environments, with a multivariate method for the multiplicative effect of GxE interaction (Zobel et al. 1988). The AMMI method is being used in studies on the GxE interaction of soybean (Oliveira et al. 2003). This study aimed to evaluate the adaptability and stability of soybean cultivars in the state of Paraná and to test the efficiency of the AMMI multivariate method (Zobel et al. 1988) in comparison with the methods of Eberhart and Russel (1966) and Cruz et al. (1989) of evaluating cultivar stability.

Analysis of variance of multi-location trials is important for estimating variance components related to different sources of variation. These include genotypes and genotype by environment interaction (Crossa, 1990). The variance component analysis is crucial as it measures the errors that result from genotype by environment interaction in measuring traits such as yield. Hence, the knowledge of the magnitude of the interaction helps in estimating the genotypic effects and determining the optimum resource allocations in terms of number of sites and plots to be included in the next trial (Crossa, 1990). However, Romagosa et al. (1993) reported that when dealing with a large number of genotypes, estimating GEI using ANOVA is demanding and it fails to show the pattern of the GEI variance components.

Additive main effects and multiplicative interactions (AMMI) and genotype plus genotype by environment interaction (GGE) models effectively capture the additive (linear) and multiplicative (bilinear) components of GEI interaction and provides useful interpretation of multi-environment data in breeding programmes (Saini and Chetan, 2007). AMMI combines the additive components in a single model for the main effects of genotype, environment and multiplicative components for the interaction effect (Mitrovia et al., 2012). Genotypic performances and phenotypic stability of the cultivars are best expressed by their graphic analyses (Miranda et al., 2009) and it is useful in summarizing and emulating the response patterns which originally existed in the raw data. The GGE biplot analysis is another method which incorporates the genotype and GEI effects in the evaluation of cultivars and it uses graphic axes to identify best performing cultivars in the mega-environments (Akcura et al., 2011). Secondly, this model provides genotype estimates in different locations. The other strength of this model is that it also combines ANOVA and PCA by separating sum of squares of genotypes and GEI together using the PCA method (Abay and Bjornstad, 2009). AMMI and GGE statistical tools have huge importance and relevance to agricultural scientists because they deal with data that come from various types of experiments (Rad et al., 2013).

2.6 Multi-trait relationships

GGE biplots have been used for analysing multi environment trials and the concept can also be used to analyse data for multiple traits across locations. When performing multi-trait analyses, the genotypes are used as entries instead of using environments and traits are used as testers to construct genotype by trait (GT) biplots. This is an effective tool that graphically summarizes the genotype by trait data, visualizes relationships among the measured traits and visualizes the performance of genotypes based on the traits which influence selection of potential parents (Yan and Tinker, 2005). In addition, it helps identify less important traits that do not contribute directly to the trait of interest. Genotype by trait has been used in soybean yield analysis by Yan and Kang

(2002) who reported that one genotype performed the best across all locations.

2.7 Heritability

heritability and genetic advance are useful biometrical tools for determination of genetic variability. The grain yield is a complex character, quantitative in nature and an integrated function of a number of component traits (Melquiades and Reyes,1990; Karasu, 2009). Soybean yield has been reported to have low heritability of 26.7-28.6% (Melquiades and Reyes,1990; Karasu, 2009). Therefore, selection for yield per se may not be much rewarding unless other yield attributing traits are taken into consideration. Correlation study provides a measure of association between characters and helps to identify important characters to be considered while making elucidates selection. Knowledge of correlation between grain yield and other characters is helpful in selection of suitable plant type (Aditya et al, 2011). Indirect selection for seed yield was a function of selection towards the enlargement of the components of seed yield (Watson, 1952). In previous studies yield components that were found to be influential in soybean yield (Akheter & Sneller 1996), included number of plants/unit area, number of main stem nodes/plant, number of pods/node, number of seeds/pod, number of branches/plant, number of pods/branch and mean seed weight. Agronomic traits influencing soybean yield include days to flowering, maturity, Plant height, grain filling period, and shattering resistance (Johnson et al, 1955). According to their studies using two F3 populations conducted at two locations showed that selection based on the combination of agronomic traits influencing soybean yield can be as effective as selection for yield itself. Agronomic traits influencing soybean yield have higher heritability making it suitable for use in indirect selection for yield. Estimate of plant height heritability was 84% (Brim, 1973) and 91 per cent for days to 50 % flowering, number of primary branches per plant and Days to maturirty. Similar results were obtained by (Chauhan and Singh, 1984) and (Pushpendra and Ram, 1987).

Chapter 3

Materials and Methods

3.1 Study sites

The multi-locational trials (MLT) were carried out in two seasons 2017/ 2018 and 2018/2019 at six different locations, International Institute of Tropical Agriculture Southern African Region Administration (Hub-IITA-SARAH), Lusaka-Zambia, Chipata-Zambia, Seedco (Lusaka West, Zambia), Chitedze (Malawi) in 2018, and Chitedze (Malawi), IITA-SARAH (Zambia), BVUMBWE (Malawi), and Kabwe (Zambia) in 2019 to assess and confirm the effects of genotype, environment, and genotype by environment interaction. The locations have different agro-climatic conditions. Table 3.1 shows details about the locations.

Table 3.1: Trial locations and their geographical information

Location	**Country**	**Longitude**	**Latitude**	**Elevation (m)**	**Rainfall (mm)**		**Temperature (^{o}C)**			
					2018	2019	2018 Min.	2018 Max.	2019 Min.	2019 Max.
Chipata	Zambia	E32^{o}39’	S13^{o}40’	1098	1249	234.46	13	28	12.6	29.089
Chitedze	Malawi	E33^{o}38’	S13^{o}59’	1100	929	114.533	12	28.6	13	30
IITA-SARAH	Zambia	E28^{o}30’	15^{o}30’	1193	703	127.98	12.6	27.98	11	30.88
Bvumbwe	Malawi	E35^{o}06’	15^{o}917’	1146	930	180.28	13	24.67	12.33	27.89
SeedCo. Zambia	Zambia	E28^{o}33’	15^{o}67’	1301	826	138.04	11	29	13	28.46
Kabwe	Zambia	E28^{o}45’	14^{o}4287’	1182	804	104.517	12.38	29.12	13.1	28.95

3.2 Experiment design and measurements

Sixty-two genotypes of Soybean (Glycine max L.) were evaluated during the 2017/18 and 2018/19 seasons in six diverse locations; two in Malawi and four in Zambia and (Table 3.1). The sixty-two genotypes included seven Zambian released varieties (Table 3.3). The soybean genotypes were planted in a 6 x 5 and 5 x 5 alpha lattice design with three replications. Each genotype occupied a plot comprising of 4 rows, 4 m in length, 0.5 m between rows, and 0.05 m intra-row spacing. Basal dressing fertilizer (25 kg N/ha, 30 kg K2O/ha, 60 kg P2O5/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. Hand weeding was done to control weeds that could not be killed by herbicides. Data collected involved grain yield (GY), 100 seed weight (SWT100), plant height (PLHT), days to maturity (DM), days to flowering (DFFL), lodging, and shattering.

Table 3.2: The scoring rate for pod-shattering was as follows (Krisnawati & Adie, 2017;IITA, 1986).

Score	Description	Category pod shattering
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

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).

The scoring scale for lodging was based on Woods and Swearingin, Agron. J., (1977) as follows;
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.
5 = all plants down.

Table 3.3: List of the evaluated soybean genotypes

No.	Genotypes	Sources	Growth Habit	Maturity Period	No.	Genotypes	Sources	Growth Habit	Maturity Period
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

3.3 Statistical analysis

3.3.1 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 named E1 (Chipata and Chitedze) , E2 (IITA-SARAH) , E3 (BVUMBWE), and E4 (Kabwe and Seedco-Zambia), see table 1. Normality was also checked in data using residuals. Multi environment data were analyzed to visually identify GE interactions. The statistics software package metan in R (Olivoto and Lúcio, 2020) was used to calculate ANOVA-based stability statistics, Variance components and heritability in mixed-effect models, and Best linear unbiased prediction (BLUP) values for grain yield (GY), 100 seed weight (SWT100), plant height (PLHT), days to maturity (DM), days to flowering (DFFL), lodging, and shattering. In the BLUPs estimation for each replicate, it was assumed the effect of the genotype, and the interaction effect of the genotype and environment to be random effects within the following mixed-effect model:

$$y = x\beta + Z\mu + \varepsilon \tag{3.1}$$

Where y is a vector of response variable; β is a vector of known fixed effect; u is a vector of random effects; x is a design matrix relating to β; Z is a design matrix relating Y to u; ε is a vector of random errors (Olivoto and Lucio, 2020).

Different biplots were created using the environment-centered model ($Y_{ij} - \mu - \beta_j = \alpha_i + \phi_{ij}$) where the E main effect (β_i) is removed and the biplot only considers G(α_i) and GE (ϕ_{ij}) as the relevant sources of variation.he biplot analyses were carried out using the package GGE Biplot GUI in R (Frutos, Galindo and Leiva, 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 the comparison of G and GE, biplots were drawn using row-preserving, singular value partitioning (Yan and Tinker, 2006).

The discriminativeness vs representativeness is a tester-centred (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.

The ideal environment biplot ranks the tester environments based on discriminating ability and representativeness considering an “ideal environment” which is defined as the most discriminating and representative. On the other hand, in the ideal genotype biplot, genotypes are ranked based on mean performance and stability and are compared to the “ideal genotype”, a hypothetical genotype that has the highest performance in all the environments and, as result, is completely stable.

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. The length of the vector of each genotype is proportional to its variability. Longer vectors correspond to genotypes with poorer stability and genotypes with short vectors are expected to be more stable. Moreover, genotypes near the average-environment coordination (AEC) ordinates (double-arrowed line) have a mean yield similar to the grand mean. In the same way, genotypes with vectors near the AEC abscissa (single-arrowed line) have a mean like the higher mean yield across the environments (Kaya et al., 2006)

The which-won-where pattern shown by a GGE biplot represents concepts such as crossover GE, mega-environment differentiation, and specific adaptation (Yan and Tinker, 2006). The which-won-where biplot is constructed by drawing a polygon on genotypes that are furthest from the origin. The polygon will contain all the genotypes, and perpendicular lines to the sides allow to define sectors where the performance of the genotypes within environments can be analyze.

Finally, yield can be associated with other traits using a genotype-by-trait table. Grain yield was associated with SWT100, PLHT, DM, DFFL, lodging, and shattering and a genotype by trait biplot along with the genotypes was constructed. This biplot allows identifying traits that are positively or negatively correlated as well as identify the genotypes with a better performance in specific traits.

Chapter 4

Results

4.1 Analysis of variance and heritability

In the analysis of variance, all the sources of variation had a significant effect 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 4.2).

Table 4.1 shows the proportion of variance explained by the Environment, Genotype, and GxE interaction, and the broad-sense heritability (H2) for different traits. Overall, the environment explained a higher proportion of the variance, followed by the interaction (GxE), and the genotype. In the yield, the environment explained the 14.36% of the variability, the genotype the 1.02%, and the interaction the 5.06% with an H2 of 0.37. The highest H2 was observed for DM with 0.76 and the lowest for lodging with 0.21.

Table 4.1: Proportion of variance explained by the Environment, Genotype, and GxE in interaction, and the broad-sense heritability (H2) for different traits evaluated in 2018 and 2019 season in Zambia and Malawi

	Proportion of variance explained (%)			
Trait	**Environment**	**Genotype**	**GxE**	**H^2**
Grain yield	14.36	1.02	5 .03	0.37
HSW	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

GxE: Genotype by environment interactions, H^2 broad sense heritability, HSW Hundred seed weight.

Table 4.2: Anova for Soybean traits.

SOV		GY	SWT100	PLHT	DM	DFFL	Lodging	Shattering
	Df	**Mean Sq**	**Mean Sq**	**Mean Sq**	**Mean Sq**	**Mean Sq**	**Mean Sq**	**Mean Sq**
Environment	3	80532.75 ***	388.3 ***	3980 n.s	2163.9 *	2991.1 ***	244.41 n.s	34.068 ***
Rep(ENV)	68	12679.94 ***	22.5 n.s	3806 ***	660.4 ***	79.9 ***	1.63 ***	2.158 ***
Genotype	61	1284.55 ***	51.4 ***	465 ***	350.3 ***	232.3 ***	1.52 ***	2.096 ***
G × E	183	860.99 ***	31.9 ***	299 ***	80.6 ***	62.8 ***	1.32 ***	0.833 ***
Error	1358	497.75	22.8	110	26.5	18.8	0.82	0.412
CV (%)	29.5	29.5	31.7	14.7	4.8	8.5	40.8	38.4

*, **, and *** indicate significance at the 0.05, 0.01 and 0.001 probability levels, respectively. n.s., not significant at $P \leq 0.05$ GY Grain yield, SWT100 100 Seed weight, PLHT Plant height, DM Days to maturity, DFFL Days to 50% flowering, G x E Genotype by environment, SOV Source of variation, Rep; Replication.

4.2 Environment evaluation

Relationships among the four environments can be visualized by analysing 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 4.1). However, E4 is close to being in the middle between E2-E3 and E1.

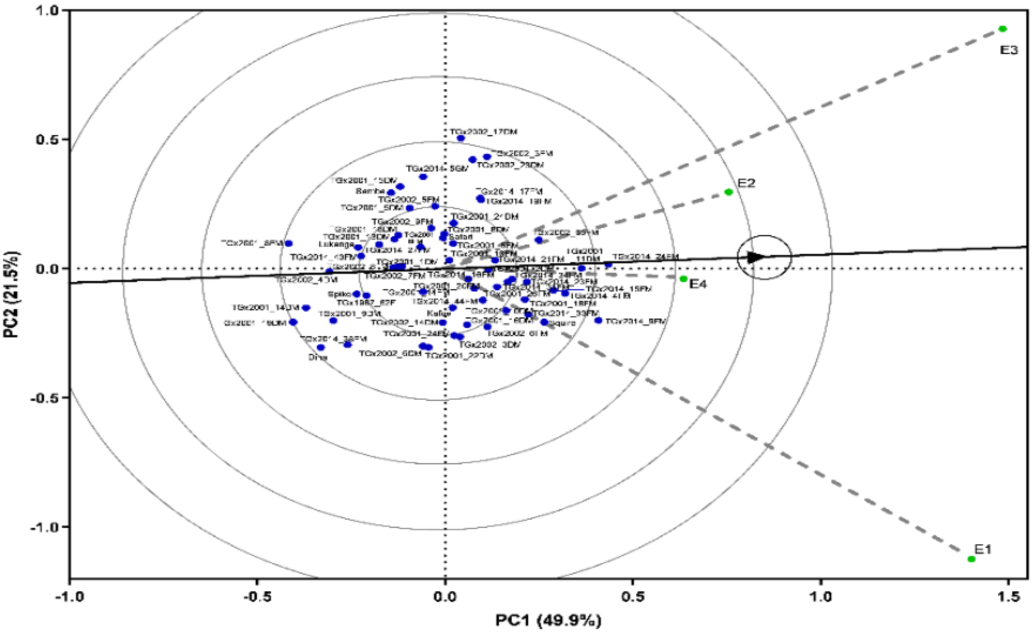

Figure 4.1: General view of relationship among environments and performance of genotypes. PC, principal component.

The concentric circles on the biplot are proportional to the standard deviation and help to visualize the length of the vectors. Environments with longer vectors have a higher discriminating ability. Thus, E1 and E3 were the most discriminating environments followed by E2 and E4 which had shorter vectors and may provide less information on the evaluated genotypes (Figure 4.1). In Figure 4.1 the average environment, indicated with an arrowhead and a circle, and an Average-Environment Axis (AEA) have also been included. The AEA is a line that passes through the average environment and the biplot origin and helps to categorize the environments based on their representativeness. A test environment with the smallest angle with the AEA is more representative. Therefore, E4 is the most representative environment, followed by E2, E3, and E1 (Figure 4.1). Test environments that are both representative and discriminating are useful to select 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 are 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.

An ideal environment is that with the highest discriminativeness and representativeness. Figure 4.2 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

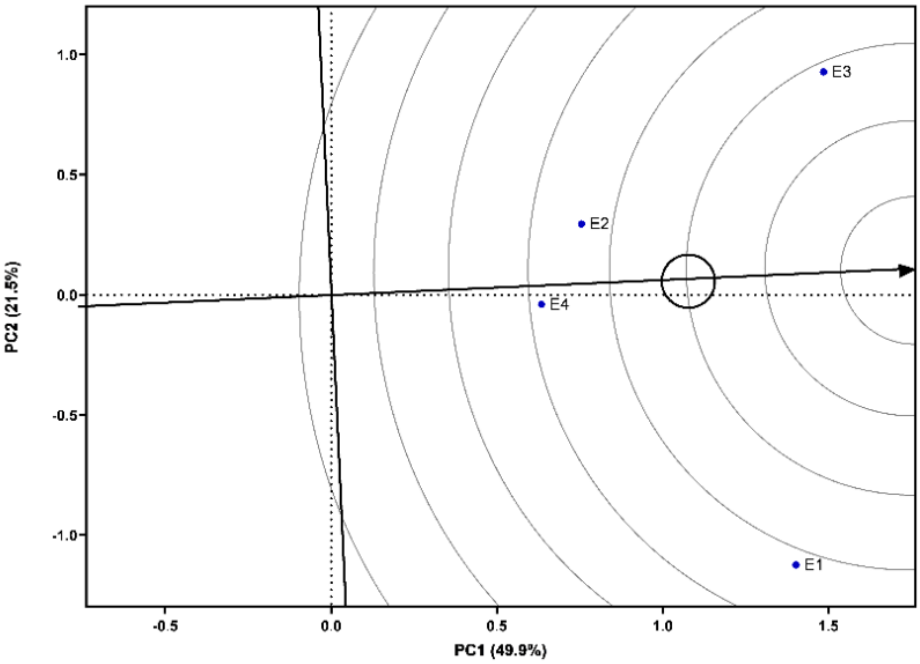

Figure 4.2: Ranking of environments relative to the “ideal” environment. PC, principal component.

4.3 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 that environment (Hoyos-Villegas, Wright and Kelly, 2016). `TGx2002_35FM` q, `TGx2014_24FM`, `TGx2001_11DM` yield better than average in E2, E3, and E4, while `TGx2014_9FM`, Squire, `TGx2014_33FM`, and `TGx2001_16DM` are better adapted to E1 (Figure 4.1)

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, Wright and Kelly, 2016). Among the evaluated genotypes, `TGx2001z_13FM`, and `TGx2014_16FM` were near the biplot origin, which means that these genotypes have a lower contribution to genotype or GE interactions (Figure 4.1). 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. By examining the angle formed between the AEA and the 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 4.1). 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.

Figure 4.3 shows the Average Environment Coordination (AEC) view of the genotype-metric preserving biplot, and it is appropriate to evaluate the mean performance and stability of genotypes. On the AEC, genotypes that are closer or above the AEC abscissa (single-arrowed line) in the same direction of environments, are expected to have the highest mean performance. In contrast, genotypes located farther from the AEA abscissa 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. Six genotypes, including `TGx2002_5FM`, Safari, `TGx2001_13FM`, Kafue, `TGx2002_14DM`, and `TGx2001_24FM` performed at the grand mean because they were located close to the AEC ordinate (Figure 4.3).

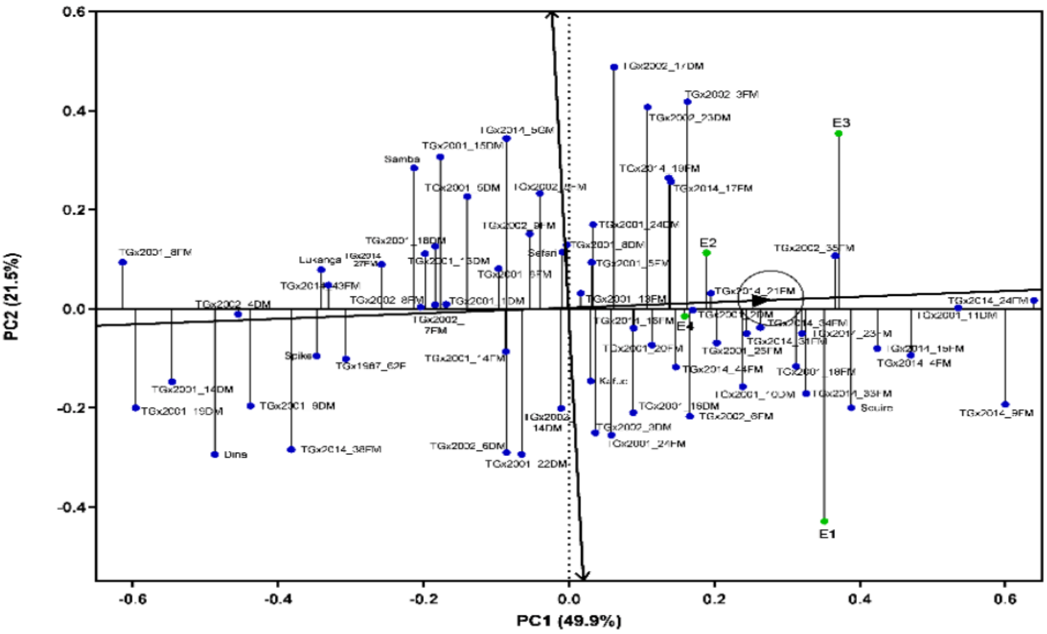

Figure 4.3: Mean vs stability biplot. PC, principal component.

The AEC ordinate on axis y (double-arrowed line) points to poorer stability in either direction. Consequently, `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 (Figure 4.3). 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, Wright and Kelly, 2016). 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 4.4). Interestingly the elite lines (Dina, Kafue, Lukanga, and Safari) showed yields lower or close to the grand mean and were considerably unstable, being far from the ideal environment (Figure 4.3 and 4.4).

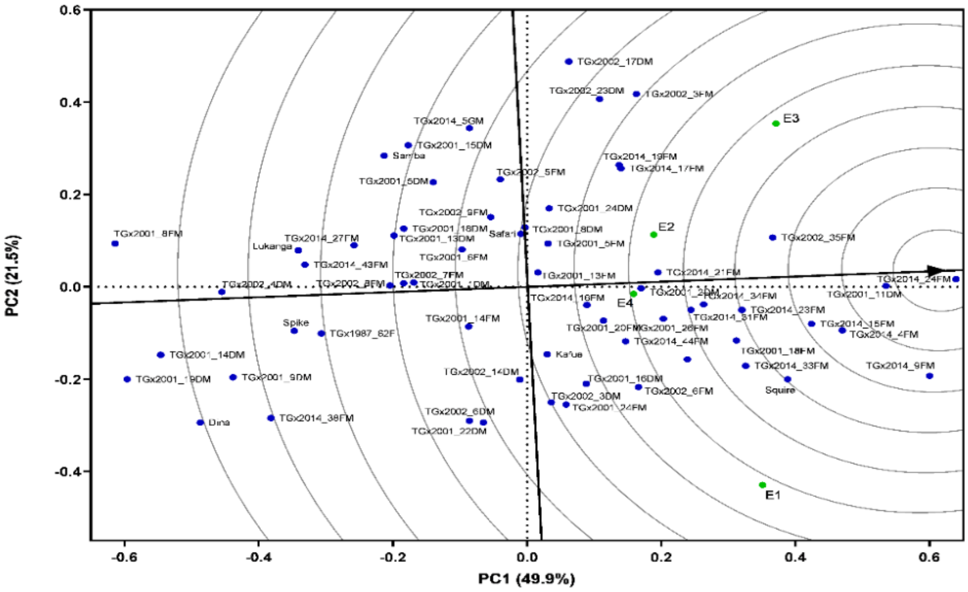

Figure 4.4: Ranking of genotypes relative to the "ideal" genotype. PC, principal component.

The which-won-where feature is displayed in figure 4.5. Genotypes within the same sector of an environment are expected to have a better response in that environment. A total of six sectors were identified, and the environments fell into two sectors. E2 – E4 were located in the same sector, where `TGx2014_24FM` had the best performance (genotype located at one of the vertices of the polygon), followed by `TGx2001_11DM`. In comparison, in the E1 the genotype with the best response was `TGx2014_9FM`, followed by `TGx2001_24FM`.

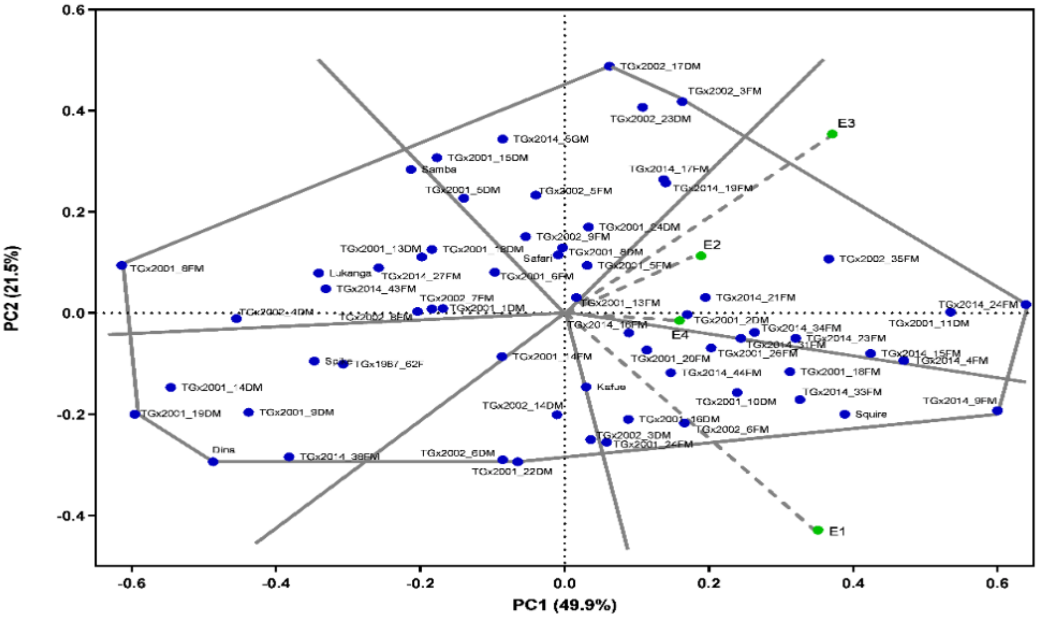

Figure 4.5: The which-won-where view of the GGE biplot. PC, principal component.

4.4 Genotype-by-trait analysis

Multiple traits averaged across all environments can be correlated with the genotypes using genotype x trait biplots. These biplots allow to visually identify 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 4.6). Yield variables (GY and SW100) were grouped in the biplot and showed a negative relation with lodging and independence of shattering. Regarding genotypes, `TGx2014_24FM`, `TGx2002_14DM`, and TGx2014_9FM had the highest GY and SW100, and the lowest lodging. 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 elite lines, Safari, and Lukanga showed a higher GY and SW100; 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.

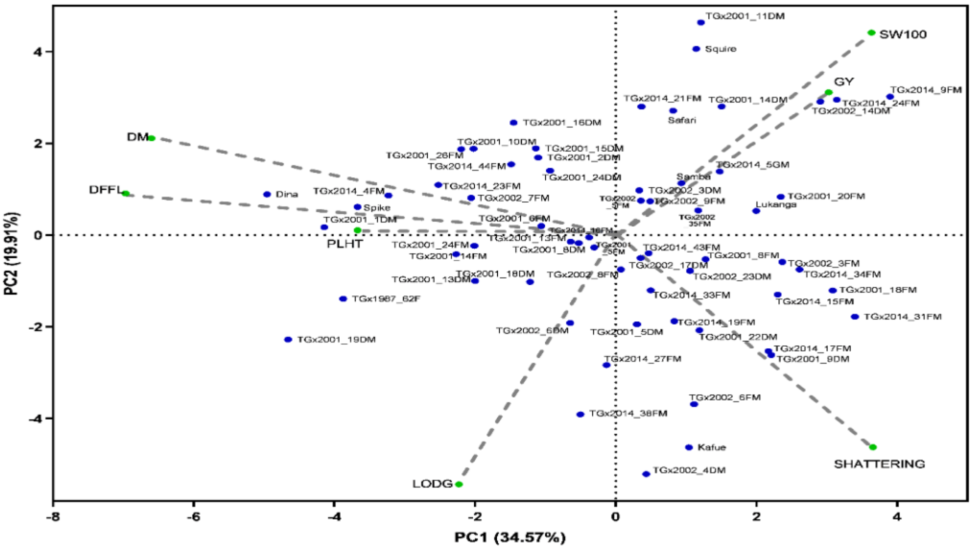

Figure 4.6: Genotype by trait biplot representing 62 genotypes evaluated for seven traits. DFFL, days to flowering; DM, days to maturity; PLHT, plant height; GY, grain yield; SW100, 100 seed weight; LODG, lodging. PC, principal component.

Chapter 5

Discussion

Understanding genotype by environment interaction is an important aspect in any breeding program. Breeding programs with a global or regional mandate such as IITA requires that genotypes that are being developed are extensively tested for yield and yield related traits in multiple environments within and outside targeted countries. This is to ensure that the materials are broadly adapted in these targeted environments. In this study a total of 62 genotypes from IITA except four from private breeding companies were evaluated in field trials conducted in 2018 and 2019 growing seasons in Malawi and Zambia.

In this study, the additive main effect and multiplicative interaction (AMMI) analysis of variance for yield showed that the environmental variance was significantly higher than both the genotype and genotype by environment interaction (GEI) variance. This result also showed that the environmental (E) main effect was the most important source of variation due to its large contribution to the total variation for yield. Environment and genotype accounted for 14.36 % and 1.02 %, of the total variation, respectively. The significant G×E interaction explained 5.06 % of the variation which was almost five times that of the genotypic effects and less than that of the environmental effect. This was expected because the environments under testing in the two countries vary in terms of soil type, precipitation, temperature, and humidity. Previous studies have also reported a higher contribution of environment than GEI on yield (Gurmu et al. 2009; Rakshit et al, 2012, Temesgen et al, 2015; Gurmu, 2017 and Vaezi et al, 2017). However, other studies have reported a higher GEI contribution to yield than environment (Mwiinga, 2020; Atnaf et al, 2013; Bhartiya et al. 2017).

Environmental evaluation is done through the depiction of vector length. E1 and E3 were the most discriminative environments since their GEI variation was larger than E2 and E4 as depicted by the length of their vectors. Therefore, E1 and E3 would be useful in the assessment of variations among genotypes. A test environment with the smallest angle with the AEA (Average-Environment Axis) is more representative. Therefore, E4 is the most representative environment, followed by E2, E3, and E1. Test environments that are both representative and discriminating are useful to selecting genotypes that have broader adaptation. In contrast, discriminating but non-representative environments are valuable for selecting genotypes with specific adaptation (Yan and Tinker, 2006), 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. An ideal environment is that with the highest discriminativeness and representativeness. E3 was the closest environment to an ideal, followed by E1, E2, and E4. The "which won where" pattern helps in visualizing the possible existence of mega environments in multi environmental trials and shows the best performing genotype in each environment (Kaya et al., 2006). The genotypes on the vertex of the polygon formed were either the best or poorest in the sectors and designated environments they fell in (Yan et al., 2007). A total of six sectors were identified, and the environments fell into two sectors. E2, E3, E4 were located in the same sector, where `TGx2014_24FM` had the best performance (genotype located at one of the vertices of the polygon), followed by `TGx2001_11DM`. In comparison, in the E1 the genotype with the best response was `TGx2014_9FM`, followed by `TGx2001_24FM`. Therefore, these genotypes are recommended to be released in E1.

In the current study, yield performance and stability analysis were defined using the first and second interaction principal component axis scores of all test locations symbolised by a small circle. 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, Wright and Kelly, 2016). Two lines pass through the origin of the biplot, the first one is the average environment axis, and this has an arrow pointing to greater GEI effect and reduced stability. The second one, the ordinate of the AEC runs perpendicular to the AEC (Kaya et al., 2006). Yield performance of genotypes `TGx2014_24FM`, `TGx2014_9FM`, `TGx2001_11DM`, and `TGx2014_4FM` was higher than average, and these genotypes can be recommended for all test locations provided that there is optimal climatic conditions and improved management practices. The AEC ordinate separates genotypes with below average yields to those with above average yields. `TGx2001_8FM`, `TGx2001_19DM`, and `TGx2001_14DM` had the lowest mean. Six genotypes, including `TGx2002_5FM`, Safari, `TGx2001_13FM`, Kafue, `TGx2002_14DM`, and `TGx2001_24FM` exhibited average performance because they were located close to the AEC ordinate. Based on performance and stability, genotypes can be ranked relative to an "ideal" genotype, which is that with the highest performance and the lowest variability. Among the tested genotypes, `TGx2014_24FM` was the closest to an ideal genotype, followed by `TGx2001_11DM`. Therefore, `TGx2014_24FM` and `TGx2001_11DM` could be considered for release in multiple environments.

Yield is a major breeding objective, however breeding for yield tends to be challenging because it is a genetically complex trait with low heritability. Therefore, identification of traits that are strongly correlated with yield and have high heritability could help to indirectly select for yield and circumvent challenges of direct selection. In this study the heritability for yield was estimated at 37% which is relatively low, but it is consistent with previous studies that reported heritability of 26-28.6% (Melquiades and Reyes,1990; Karasu, 2009). However, it was encouraging to note that the other traits which includes plant height, days to maturity, and days to flowering which were positively correlated to yield, had higher heritability of 50%, 75%, and 76% respectively, than yield suggesting that they could be used to indirectly select for yield.. Regarding genotypic by trait performance, `TGx2014_24FM`, `TGx2002_14DM`, and `TGx2014_9FM` had the highest grain yield and HSW, and the lowest lodging. On the other hand, `TGx2001_11DM` and Squire, which had high yield performance, had a lower shattering score, which did not have an apparent effect on `TGx2014_24FM`, and `TGx2002_14DM`. Among the checks, Safari, and Lukanga showed higher grain yield and HSW. These findings are consistent with previous studies on the density effects on yield components (Manda and Maata, 2020). Lukanga and Kafue had a higher shattering score; and Dina had a higher plant height, days to maturity, and days to flowering, accompanied by a lower grain yield, HSW, and shattering.

Chapter 6

Conclusion

In the present study AMMI analysis revealed that yield of the 62 genotypes tested were significantly affected by genotype, environment and genotype by environment. GGE analysis using the “which won where” pattern identified genotypes with specific adaptation and possibility of mega environments. `TGx2014_9FM`, followed by `TGx2001_24FM` were more suitably adapted to environment E1. The study also revealed that TGx2002_4DM, TGx2002_8FM, TGx2001_11DM, and TGx2014_24FM were highly stable across the tested environments in Malawi and Zambia. Hence, they could be recommended for all test locations in the current study provided that there are optimal climatic conditions and improved management practices. `TGx2014_24FM` was identified as the most ideal genotype followed by `TGx2001_11DM`. Environment E3 can be used as a location for breeding purposes if the breeding program decided to reduce the number of breeding locations or in efforts to optimize resources utilization. Plant height, days to maturity, and days to flowering could be used to indirectly select yield due positive correlation to yield, and higher heritability.

Bibliography

[1] AA. Karasu1, M. Oz1, A. T. Göksoy2* and Z. M. Turan. *Genotype by environment interactions, stability, and heritability of seed yield and certain agronomical traits in soybean [Glycine max (L.) Merr.]*. Mustafakemalpa_a Vocational School, Uluda_ University, 16059 Bursa, Turkey. Department of Field Crops, Faculty of Agriculture, Uluda_ University, 16059 Bursa, Turkey, (2009).

[2] Abate, T., A.D. Alene, D. Bergvinson, B. Shiferaw, S. Silim, A. Orr, et al. *Tropical grain legumes in Africa and south Asia: knowledge and opportunities International Crops Research Institute for the Semi-Arid Tropics*. 2012.

[3] Akcura, M., S. Taner and Y. Kaya. *Evaluation of bread wheat genotypes under irrigated multi-environment conditions using GGE biplot analyses*. Agriculture 98: 35-40, 2011.

[4] Alberts, M.J. *A comparison of statistical methods to describe genotype x environment interaction and yield stability in multi-location maize trials*. (Doctoral dissertation, University of the Free State), 2004.

[5] Alghamdi, S.S. *Yield stability of some soybean genotypes across diverse environments*. Pakistan Journal of Biological Sciences 7: 2109-2114. 2004.

[6] Allard, R.W. and A.D. Bradshaw. *Implications of genotype-environmental interactions in applied plant breeding*. Crop Science. 4: 503-508. 1964.

[7] Antalina. *Recent Research and Industrial achievement for soybean in Japan*. Proceeding of RIELT-JIRCAS. Workshop on soy research, Sep. 28, 2000.

[8] Arshad, M., A. Bakhsh and A. Ghafoor. *Path coefficient analysis in chickpea (Cicer arietinum L.)*. Pakistan Journal of Botany 36: 75-82. 2004.

[9] Atnaf, M., S. Kidane, S. Abadi and Z. Fisha. *GGE biplots to analyze soybean multi-environment yield trial data in North-western Ethiopia*. Journal of Plant Breeding and Crop Science 5: 245-254. 2013.

[10] Baker, J., L. Allen, K. Boote, P. Jones and J. Jones. *Response of soybean to air temperature and carbon dioxide concentration*. Crop Science. 29: 98-105. 1989.

[11] Beaver, J. and R. Johnson. *Yield stability of determinate and indeterminate soybeans adapted to the northern United States*. Crop Science. 21: 449-454. 1981.

[12] Bekheit, M. *Evaluation of some soybean genotypes in Upper Egypt.* Thesis submitted to Faculty of Agriculture in partial fulfilment of requirement for the degree of M.Sc (Agriculture) in Assuit University, Egypt. 2000.

[13] Bernardo, R. *Molecular markers and selection for complex traits in plants: learning from the last 20 years.* Crop Science. 48: 1649-1664. 2008.

[14] Bhartiya, A., J. Aditya, V. Kumari, N. Kishore, J. Purwar, A. Agrawal, et al. *GGE biplot & AMMI analysis of yield stability in multi-environment trial of soybean [Glycine max (L.) Merrill] genotypes under rain-fed condition of North-western Himalayan hills.* Journal of Animal and Plant Sciences 27: 227-238. 2017.

[15] Borrás, L., G.A. Slafer and M.a.E. Otegui. *Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal.* Field Crops Research 86: 131-146. 2004.

[16] Breene, W., S. Lin, L. Hardman and J. Orf. *Protein and oil content of soybeans from different geographic locations.* Journal of the American Oil Chemists' Society 65: 1927-1931. 1988.

[17] Burton, J., R. Wilson, G. Rebetzke and V. Pantalone. *Registration of N98–4445A mid-oleic soybean germplasm line.* Crop Science. 46: 1010-1012. 2006.

[18] Burton, J.W. and L. Miranda. *Soybean improvement: Achievements and challenges.* Ratarstvo i Povrtarstvo 50: 44-51. 2013.

[19] Carpentieri-Pípolo, V., F.A.M. da Silva and A.L. Seifert. *Popcorn parental selection based on genetic divergence.* Crop Breeding and Applied Biotechnology 3:90-106. 2003.

[20] Chauhan V.S., Singh B.B. *Genetic variability and heritability in soybean.* Indian J. Agric. Sci. (1984)53 (4): 273-276.

[21] Chianu, J., B. Vanlauwe, J. Mukalama, A. Adesina and N. Sanginga. *Farmer evaluation of improved soybean varieties being screened in five locations in Kenya: Implications for research and development.* African Journal of Agricultural Research 1: 143-150. 2006.

[22] Cianzio, S., P. Arelli, B. Diers, H. Knapp, P. Lundeen, N. Rivera-Velez, et al. *Soybean germplasm lines AR4SCN, AR5SCN, AR6SCN, AR7SCN, and AR8SCN. ISURF#* Iowa State University, Ames, IA, USA. 2007.

[23] Cooper, M., F. van Eeuwijk, S.C. Chapman, D.W. Podlich and C. Löffler. *Genotype-by-environment interactions under water limited conditions.* African Crop Science Journal 13: 41-47. 2006.

[24] Crossa, J. *Statistical analyses of multilocation trials.* Advances in Agronomy 44: 55-85. 1990.

[25] Deulvot, C., H. Charrel, A. Marty, F. Jacquin, C. Donnadieu, I. Lejeune-Hénaut, et al. *Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea.* BMC Genomics 11: 468. doi:10.1186/1471-2164-11-468. 2010.

[26] Dong, Y., L. Zhao, B. Liu, Z. Wang, Z. Jin and H. Sun. *The genetic diversity of cultivated soybean grown in China.* Theoretical and Applied Genetics 108: 931-936. 2004.

[27] Dong, Y.S., B.C. Zhuang, L.M. Zhao, H. Sun and M.Y. He. *The genetic diversity of annual wild soybeans grown in China.* Theoretical and Applied Genetics 103: 98-103. doi:10.1007/s001220000522.

[28] Dornbos, D. and R. Mullen. *Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature.* Journal of the American Oil Chemists' Society 69: 228-231. 1992.

[29] Eberhart, S.t. and W. Russell. *Stability parameters for comparing varieties.* Crop Science. 6: 36-40. 1966.

[30] Fabiyi, E. *Soyabean processing, utilization and health benefits.* Pakistan Journal of Nutrition 5: 453-457. 2006.

[31] FAO. *Food and agriculture data.* FAOSTAT, Rome. 2018.

[32] FAOSTAT. *Agriculture Organization of the United Nations Statistics Division (2014).* Production Available in: http://faostat3. fao. org/browse/Q/QC/S [Review date: April 2015]. 2016.

[33] Finlay, K. and G. Wilkinson. *The analysis of adaptation in a plant-breeding programme.* Australian Journal of Agricultural Research 14: 742-754. 1963.

[34] Gandhi, A. *Quality of soybean and its food products.* International Food Research Journal 16: 11-19. 2009.

[35] Gebeyehu, S. and H. Assefa. *Genotype X environment interaction and stability analysis of seed yield in navy bean genotypes.* African Crop Science Journal 11: 1-7. 2003.

[36] Gibson, L. and R. Mullen. *Soybean seed composition under high day and night growth temperatures.* Journal of the American Oil Chemists' Society 73: 733-737. 1996.

[37] Gurmu, F. *Stability analysis of fresh root yield of sweetpotato in southern Ethiopia using GGE bi-plot.* International Journal of Pure Agricultural Advances 1: 1-9. doi:10.20448/813.11.1.9. 2017.

[38] Gurmu, F., H. Mohammed and G. Alemaw. *Genotype x environment interactions and stability of soybean for grain yield and nutrition quality.* African Crop Science Journal 17:87-98. 2009.

[39] Haghi, Y., P. Boroomandan, M. Moradin, M. Hassankhali, P. Farhadi, F. Farsaei, et al. *Correlation and path analysis for yield, oil and protein content of Soybean (Glycine max L.) genotypes under different levels of nitrogen starter and plant density.* Biharean Biologist 6: 32-37. 2012.

[40] J.P. ADITYA, Pushpendra BHARTIYA, Anuradha BHARTIYA. *Genetic variability, heritability and character association for yield and component characters in soybean (G. max (L.) Merrill),* Department of Genetics and Plant breeding, College of Agriculture, GBPUA & T, Pantnagar, jayprakashaditya@gmail.com. 2011.

[41] Kehlenbeck, H., C. Krone, E.-C. Oerke and F. Schönbeck. *The effectiveness of induced resistance on yield of mildewed barley.* Journal of Plant Diseases and Protection101: 11-21. 1994.

[42] Kim, M., D.L. Hyten, A.F. Bent and B.W. Diers. *Fine mapping of the SCN resistance locus from PI 88788.* The Plant Genome 3: 81-89. 2010.

[43] Krisnawati, A. and M.M. Adie. *Selection of Soybean Genotypes by Seed Size and its Prospects for Industrial Raw Material in Indonesia.* Procedia Food Science 3: 355-363. doi:http://dx.doi.org/10.1016/j.profoo.2015.01.039. 2015.

[44] Kumawat, G., G. Singh, C. Gireesh, M. Shivakumar, M. Arya, D.K. Agarwal, et al. *Molecular characterization and genetic diversity analysis of soybean (Glycine max (L.) Merr.) germplasm accessions in India.* Physiology and Molecular Biology of Plants 21: 101-107. doi:10.1007/s12298-014-0266-y. 2015.

[45] Kumudini, S. *Soybean growth and development. The Soybean: Botany, Production and Uses.* British Library, London, UK: pp48-73. 2010.

[46] Meksem, K., P. Pantazopoulos, V. Njiti, L. Hyten, P. Arelli and D. Lightfoot. *'Forrest'resistance to the soybean cyst nematode is bigenic: saturation mapping of the Rhg1and Rhg4 loci.* Theoretical and Applied Genetics 103: 710-717. 2001.

[47] Miladinović, J., M. Vidić, V. Đorđević and S. Balešević-Tubić. *New trends in plant breeding-example of soybean.* Genetika 47: 131-142.

[48] Minmin Lit, Ying Liut, Chunsheng Wang†, Xue Yang, Dongmei Li, Xiaoming Zhang, Chongjing Xu, Yan Zhang, * and Lin Zhao* *Identification of Traits Contributing to High and Stable Yields in Different Soybean Varieties Across Three Chinese Latitudes, Key Laboratory of Soybean Biology of Ministry of Education, China (Key Laboratory of Biology and Genetics & Breeding for Soybean in Northeast China).* Northeast Agricultural University, Harbin, China

[49] Miranda, G.V., L.V.d. Souza, L.J.M. Guimarães, H. Namorato, L.R. Oliveira and M.O. Soares. *Multivariate analyses of genotype x environment interaction of popcorn.* Pesquisa Agropecuária Brasileira 44: 45-50. 2009.

[50] Misiko, M., P. Tittonell, J. Ramisch, P. Richards and K. Giller. *Integrating new soybean varieties for soil fertility management in smallholder systems through participatory research: Lessons from western Kenya.* Agricultural Systems 97: 1-12. 2008.

[51] Mitroviã, b., s. Treski, m. Stojakoviã, m. Ivanoviã and G. Bekavac. *Evaluation of Experımental Maize Hybrids Tested in Multi-Location Trials Using Ammi and GGE Biplot Analyses.* urkish Journal of Field Crops 17: 35-40. 2012.

[52] Mushoriwa, H. *Breeding gains diversity analysis and inheritance studies on soybean (Glycine max (L.) Merrill) germplasm in Zimbabwe.* PhD thesis, University of KwaZulu-Natal, South Africa. 2013.

[53] Nkongolo, K. and L. Nsapato. *Genetic diversity in Sorghum bicolor (L.) Moench accessions from different ecogeographical regions in Malawi assessed with RAPDs.* Genetic Resources and Crop Evolution 50: 149-156. 2003.

[54] Pathan, M. and D.A. Sleper. *Advances in soybean breeding. Genetics and genomics of soybean.* Springer p. 113-133. 2008.

[55] Patil G, Vuong TD, Kale S, Valliyodan B, Deshmukh R, Zhu C, Wu X, Bai Y, Yungbluth D, Lu F, Kumpatla S, Shannon JG, Varshney RK, Nguyen HT *Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping.* Plant Biotechnol J. 2018;16(11):1939–53. 2018.

[56] Purdy, R. and M. Langemeier. *International Benchmarks for Corn Production.* Center for Commercial Agriculture, Purdue University. 2018.

[57] Pushpendra, Ram H.H. *Genetic components of variation for certain yield contributing traits in soybean.* Indian J. Agric. Sci. (1987)57 (4).

[58] Rad, M.N., M.A. Kadir, M. Rafii, H.Z. Jaafar, M. Naghavi and F. Ahmadi. *Genotype environment interaction by AMMI and GGE biplot analysis in three consecutive generations of wheat (Triticum aestivum) under normal and drought stress conditions.* Australian Journal of Crop Science 7: 956. 2013.

[59] Rakshit, S., K. Ganapathy, S. Gomashe, A. Rathore, R. Ghorade, M.N. Kumar, et al. *GGE biplot analysis to evaluate genotype, environment and their interactions in sorghum multi-location data.* Euphytica 185: 465-479. 2012.

[60] Reinprecht, Y., S.-Y. Luk-Labey, K. Yu, V.W. Poysa, I. Rajcan, G.R. Ablett, et al. *Molecular basis of seed lipoxygenase null traits in soybean line OX948.* Theoretical and Applied Genetics 122: 1247-1264. 2011.

[61] Romagosa, I., P. Fox, L.G. Del Moral, J. Ramos, B.G. del Moral, F.R. De Togores, et al. *Integration of statistical and physiological analyses of adaptation of near-isogenic barley lines.* Theoretical and Applied Genetics 86: 822-826. 1993.

[62] Saini, P. and S. Chetan, *A Review on Genotype Environment Interaction and its Stability Measures.* Plant Breeding Reviews 15:155-204. 2007.

[63] Shi, A., P. Chen, D. Li, C. Zheng, B. Zhang and A. Hou. *Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers.* Molecular Breeding 23: 113-124. 2009.

[64] Shi, Z., S. Liu, J. Noe, P. Arelli, K. Meksem and Z. Li. *SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance.* BMCGenomics 16: 314-321. 2015.

[65] Sionit, N., B. Strain and E. Flint. *Interaction of temperature and CO2 enrichment on soybean: growth and dry matter partitioning.* Canadian Journal of Plant Science 67: 59-67. 1987.

[66] Sudarić, A., M. Vratarić, S. Mladenović-Drinić and M. Matosa. *Biotechnology in soybean breeding.* Genetika 42: 91-102. 2010.

[67] Tefera, H. *Breeding for promiscuous soybeans at IITA. Soybean-Molecular Aspects of Breeding.* InTech. 2011.

[68] Tefera, H., A. Kamara, B. Asafo-Adjei and K. Dashiell. *mprovement in grain and fodder yields of early-maturing promiscuous soybean varieties in the Guinea Savanna of Nigeria.* Crop Science 49: 2037-2042. 2009.

[69] Temesgen, T., G. Keneni, T. Sefera and M. Jarso. *Yield stability and relationships among stability parameters in Faba bean (Vicia faba L.) genotypes.* The Crop Journal 3: 258-268. doi:https://doi.org/10.1016/j.cj.2015.03.004. 2015.

[70] Vaezi, B., A. Pour-Aboughadareh, R. Mohammadi, M. Armion, A. Mehraban, T. Hossein-Pour, et al. *GGE biplot and AMMI analysis of barley yield performance in Iran.* Cereal Research Communications 45: 500-511. 2017.

[71] Vignal, A., D. Milan, M. SanCristobal and A. Eggen. *A review on SNP and other types of molecular markers and their use in animal genetics.* Genetics Selection Evolution 34: 275-281. 2002.

[72] Yan, J., T. Shah, M.L. Warburton, E.S. Buckler, M.D. McMullen and J. Crouch. *Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers.* PloS one 4: e8451. 2009.

[73] Yan, W. and L. Hunt. *Genotype by environment interaction and crop yield.* Plant Breeding Reviews 16: 135-178. 1998.

[74] Yan, W. and M.S. Kang. *GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists.* CRC press. 2002.

[75] Yan, W. and N.A. Tinker. *An integrated biplot analysis system for displaying, interpreting, and exploring genotype× environment interaction.* Crop Science 45: 1004-1016. 2005.

[76] Zhu, Y., Q. Song, D. Hyten, C. Van Tassell, L. Matukumalli, D. Grimm, et al. *Single-nucleotide polymorphisms in soybean.* Genetics 163: 112. 2003.