GENOTYPE BY ENVIRONMENT INTERACTION, GENETIC VARIABILITY AND PATH ANALYSIS FOR GRAIN YIELD IN ELITE SOYBEAN [*Glycine max* (L.) Merrill] LINES

By

Bubala Mwiinga

BSc. Agricultural Sciences (University of Zambia)

A dissertation submitted in partial fulfilment of the academic requirements for the award of Master of Science degree in Plant Breeding

College of Agriculture, Engineering and Science

School of Agricultural, Earth and Environmental Sciences

University of KwaZulu-Natal

Pietermaritzburg

South Africa

December, 2018

DISSERTATION ABSTRACT

Soybean [*Glycine max* (L.) Merrill] is the world's leading source of protein and vegetable oil. However, its productivity is still low in the region due to limited availability of stable and high yielding cultivars. Therefore, the objectives of this study were: (1) to determine the magnitude of genotype by environment interaction and stability of elite soybean lines for seed yield, (2) to establish trait profiles of 25 soybean genotypes and to study the associations among characters, their direct and indirect effects on grain yield and (3) to estimate genetic parameters of traits related to seed yield and to analyse genetic diversity among elite soybean lines. To achieve these objectives, 25 genotypes (20 elite soybean lines and five commercial checks) were evaluated in multi-location trials conducted in the 2017/18 rainy season using six sites in four countries *viz.* Zambia, Malawi, Zimbabwe and Mozambique.

Both AMMI and GGE biplot analyses indicated Lusaka West as the highest yielding and most informative environment and could be useful for selecting specifically adapted genotypes. Rattray Arnold Research Station was the most ideal environment as it was both informative and highly representative. The soybean lines TGx2002-17DM, TGx2001-10DM, TGx2001-18DM, TGx2014-24FM, TGx2001-6FM and TGx2002-3DM exhibited specific adaptation. Both GGE and AMMI models showed that TGx2014-5GM was more stable than the checks and was second to the highest yielding check.

The genotype by trait (GT) and correlation coefficient analyses revealed that pod number per plant and hundred seed weight were the most positively correlated traits with grain yield, while days to 50% flowering had a negative association with grain yield. Sequential path analysis, showed that the number of pods per plant and hundred seed weight had the highest positive and significant direct effects on seed yield, implying that these two traits could be used as selection criteria for seed yield in soybean. The soybean lines TGx2014-5GM and TGx2002-23DM had good combinations of high yields with large seed size and high pod number.

The analysis of genetic variability showed small differences between PCV and GCV values for all the traits except for pod clearance. This implied that there were minimal effects of the environment and high contribution of the genes in the phenotypic expression of the traits, except for pod clearance, which was more affected by the environment. Moderate GCV values of 13.45% and 13.49%, high heritability values of 70% and 69% and GAM values of 23.24% and 23.04% were recorded for grain yield and number of pods per plant, respectively. Only two principal components, PC1 and PC2 accounted for the variation, with a cumulative contribution of 68.25%. All the seven traits were useful in discriminating the genotypes as they had high eigenvalues in either PC1 or PC2. The 25 soybean genotypes were grouped into two

main clusters, which were further sub-divided into eight sub-clusters based on the seven morphological characters. The genotypes TGx2014-5GM, checks SC Safari and SC Squire in sub-cluster 6 had the highest means of the most desirable traits (large seed size, high pod number per plant and seed yield). The three genotypes could be used in hybridisation programmes for improvement of grain yield, seed size and number of pods of the genotypes.

Overall, the study identified soybean lines that could potentially be released as cultivars in the four southern African countries or used as parents in future soybean improvement programmes. It also revealed traits that could be used for indirect selection of seed yield and high genetic diversity among the genotypes for possible exploitation in soybean breeding programmes to increase seed yield.

DECLARATION

I, Bubala Mwiinga, declare that:

i) The research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

ii) This dissertation has not been submitted in full or in part for any degree or examination to any other university;

iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers.

   a) Where other written sources have been quoted, then their words have been re-written but the general information attributed to them has been referenced;
   b) Where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

v) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signature……..… ……Date: 07/12/2018

Bubala Mwiinga (candidate)

Signature……. Date:……20/02/2019……...

Dr Julia Sibiya (Supervisor)

Signature……. Date:……20/02/2019……...

Dr Godfree Chigeza (Co-supervisor)

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the Almighty God for keeping me in good health throughout this journey. It has not been easy but He gave me strength and guided me through to the end of this work.

I wish to express my heartfelt thanks to my supervisor, mentor and coordinator of the Improved Masters in Cultivar Development for Africa (IMCDA) programme at the University of KwaZulu-Natal, Dr Julia Sibiya for being there for me whenever I needed her help. She was very supportive and ensured that everything was correctly done in my research. I wish to sincerely thank Dr Cousin Musvosvi for supporting me throughout my research. He was very instrumental in the analysis of the data and correcting this dissertation.

My sincere gratitude goes to my co-supervisor and the International Institute of Tropical Agriculture (IITA) soybean breeder, Dr Godfree Chigeza and the Research associate, Christabell Nachilima for according me an opportunity to conduct my research at IITA. He and the soybean breeding team were very supportive to me during my research and provided all the materials I needed for this research, including the soybean lines.

I wish to thank the Alliance for a Green Revolution in Africa (AGRA) for funding the IMCDA programme and for the support rendered to me throughout my research. It would not have been possible for me to study at this university without the support from AGRA. To Andile Mshengu, the administrative officer for the IMCDA programme, thank you so much for ensuring that all the logistics were put in place in order to make this study a success. To my family, thank you so much for supporting and encouraging me during the two years I was away for my studies.

DEDICATION

I would like to dedicate this dissertation to my mother, Noria Habeenzu Mwiinga.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AEC:	Average environment coordination
AMMI:	Additive main effects and multiplicative interaction
ANOVA:	Analysis of variance
ASV:	AMMI stability value
CH:	Commercial check
cm:	Centimetre
CV:	Coefficient of variation
D:	Determinate growth habit
DF:	Degrees of freedom
DFFL:	Days to 50 percent flowering
DM:	Days to maturity
E:	Environment
FAO:	Food and agriculture organisation
G:	Genotype
GA:	Genetic advance
GAM:	Genetic advance as a percent of mean.
GE:	Genotype by environment
GCV:	Genotypic coefficients of variation
GEI:	Genotype by environment interaction
GGE:	Genotype plus genotype by environment
GT:	Genotype by trait
GYD:	Grain yield
ha:	Hectare

H^2: Broad sense heritability

I: Indeterminate growth habit

IITA: International Institute of Tropical Agriculture

IPCA: Interaction principal component axis

k: Selection intensity

kg: Kilogram

L: Late maturity

M: Medium maturity

m: Meter

masl: Metres above sea level

MLT: Multi-location trial

mm: Millimetres.

ns: Not significant

PC: Principal component

PCA: Principal component analyses

PCV: Phenotypic coefficients of variation

PLHT: Plant height

POD_CL: Pod clearance

POD_PL: Number of pods per plant

r: Pearson coefficient correlation

RARS: Rattray Arnold Research Station

RR: Round ready

SARAH: Southern African Region Administration Hub

SSA: Sub-Saharan Africa

SWT: Hundred seed weight

V_e: Error variance

$V_{g:}$ Genotypic variance

$V_{g.l}$: Variance for the interaction between the genotypes and environments

VIF: Variance inflation factor

V_p: Phenotypic variance

X: Grand mean

CHAPTER 1
DISSERTATION INTRODUCTION

1.1 Background

Soybean [*Glycine max* (L.) Merrill, 2n = 2x = 40], is the global most important source of protein and vegetable oil. It is a self-pollinating annual legume that belongs to the family Fabaceae, subfamily Faboideae (Tefera, 2011; Pawar, 2013; Jain et al., 2017). The crop was first domesticated in north- eastern China about 1700-1100 B.C (Hymowitz, 1990) from where it spread to other countries in Asia and other continents. It was only introduced in the United States in the 1700s but it was until the 1920s and 1930s that it became an important crop. Before then, it was mainly grown as a forage crop (Sleper and Poehlman, 2006). In sub- Saharan Africa (SSA), soybean is reported to have been introduced in the 19th century by Chinese traders (Mpepereki et al., 2000). The first country to record soybean cultivation in SSA was South Africa in 1903, followed by Zimbabwe, Tanzania and Malawi in 1906, 1907 and 1909, respectively. Its cultivation expanded to other SSA countries like Zambia in 1910 and Mozambique in 1915 (Shurtleff and Aoyagi, 2009).

1.2 Production of soybean

In comparison with other major crops in the world, soybean has been leading in terms of annual percentage increase in production area since 1970 (Hartman et al., 2011). Soybean area expanded to 121.5 million ha in 2016 from 26.5 and 61.1 million ha recorded in 1966 and 1996, respectively (FAO, 2017). The world soybean production was 334.9 million tonnes in 2016 (FAO, 2017) and 348.12 million tonnes in 2017 (USDA, 2018). The global average productivity stood at 2.75 tons/ha in 2017 (USDA, 2018). The USA, Brazil and Argentina are the dominant world producers with annual production of 117.2, 96.3 and 58.8 million tonnes, respectively, in 2016 (FAO, 2017). The three countries combined produced 81.3% of the world's total soybean production in 2016. China follows the three major producers with production of 11.97 million tonnes in 2016 (Figure 1.1; FAO, 2017).

There has been an exponential increase in soybean planting area in Africa. Production area rose to 1 979 024 ha in 2016 as compared to 222 630 ha recorded in the 1970s. The corresponding production increased from 97 211 tonnes in the early 1970s to 708 322 tonnes in 1996 and 2 119 814 tonnes in 2016. However, Africa's contribution is still less than 1% of global soybean output (FAO, 2017). South Africa and Nigeria are the leading producers in Africa, while Zambia and Uganda are ranked third and fourth, respectively (Figure 1.2). Soybean area and production have rapidly expanded recently in the Southern African region,

which contributes over 50% of Africa's total soybean production (Figure 1.2). Zambia's soybean production rose from 267 490 tonnes in 2016 to 302 720 tonnes in 2018 (MOA and CSO, 2018). Similar trends have also been recorded in Malawi, Zimbabwe and Mozambique, with production estimates of 190 000, 70 000 and 32 000 ton/ha, respectively (Meyer et al., 2018). The expansion in production is a response to the growing population in the region and world, which has resulted in increased demand for vegetable oil, biofuels and animal feed (Meyer et al., 2018).

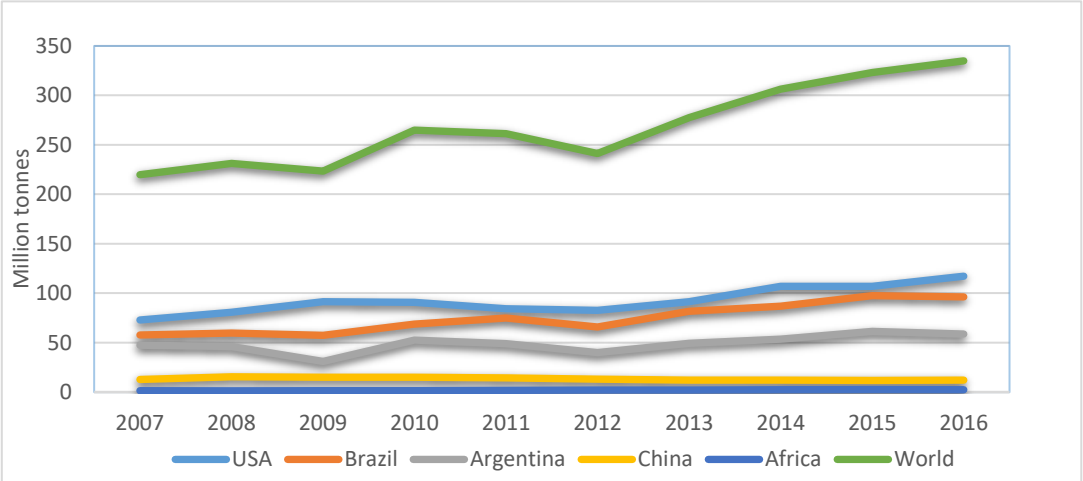

Figure 1.1 Top world soybean producers and Africa's contribution
Source: FAO (2017)

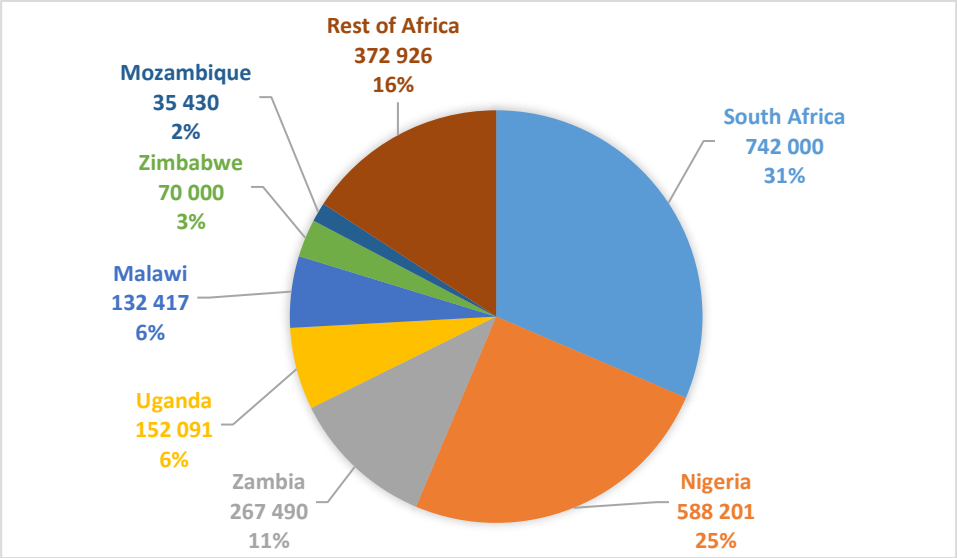

Figure 1.2 Soybean production in Africa (2016)
Source: Meyer et al. (2018)

1.3 Importance of soybean

Globally, soybean has become one of the most traded crops because of its many end uses (Hartman et al., 2011). These include animal and fish feed; human food; medicinal uses; and in the manufacturing of various industrial products, biofuels and pesticides (Liu, 1997). The seed has high oil (20%) and protein (40%) content (Sinclair et al., 2014). Oils in soybean contain high amounts of polyunsaturated fatty acids (8% α-linolenic acid and 55% linoleic acid), saturated fatty acids (stearic and palmitic acids), unsaturated fatty acid (oleic acid) and tocopherols (antioxidants) (Kanchana et al., 2015). About 95% of the oil extracted is used as cooking oil and margarines, while the remaining 5% is used in the manufacturing of such industrial products as cosmetics, biodiesel, paints, varnishes, printing inks, linoleum, disinfectant, soaps and plastics (Kanchana et al., 2015). Soy protein (about 90%) consists of two storage globulins, glycinin and conglycinin, which contain such amino acids as methionine, valine, tryptophan, tyrosine, cystine, threonine, phenylalanine, lysine, leucine and isoleucine (Kanchana et al., 2015). These amino acids are extremely important in human and animal nutrition. The seed also contains 35% carbohydrates (mostly the non-starch polysaccharides like pectin, cellulose and hemicelluloses), 1-3% phospholipids, sterols, saponins, ferritins and isoflavins (Liu, 1997). Despite being a good source of B-vitamins, soybean is deficient in vitamin C and B_{12}. Soybean also contains important minerals (about 5%) including Na, Fe, Zn, P, Mg, K and Ca (Kanchana et al., 2015).

According to Hartman et al. (2011), most of the soybean meal (about 98%) is used as aquaculture and livestock feeds and only a small portion is processed into human foods. Examples of soy protein rich foods include soymilk, tofu, soy chunks, soy yoghurt, tempeh and natto (TechnoServe, 2011; Mohamedkheir et al., 2018). Popularity of these foods has increased in recent years and they are now being consumed in many parts of Africa (Mohamedkheir et al., 2018). The immature seeds can be eaten in various ways, including being used as a side dish, after being blanched, boiled or steamed.

Soybean has for a long time been used as a preventive medicine. It has been proven to be important as a purifying food and improved organ functioning. Examples of its health benefits include dietary supplement for people suffering from diabetes (Azadbakht et al., 2003); weight management (Maskarinec et al., 2008); prevention of osteoporosis (Chen et al., 2003); and menopause symptom relief in women (Mateos-Aparicio et al., 2008). Others include lowering the risk of breast, uterus and prostate cancers (Hamilton-Reeves et al., 2007); prevention of arteriosclerosis and consequent reduction of risk of heart attack and stroke (Mateos-Aparicio et al., 2008). These health benefits have enhanced consumer

interest in soybean containing foods and thus providing the food industry with the impetus to develop such foods (Hartman et al., 2011).

Being a cash and highly nutritious crop, soybean helps smallholder farmers to be nutrition secure and gives them a better option for income diversification (Meyer et al., 2018). The crop also has potential to enhance the chemical, physical and biological soil properties through nitrogen fixation, decaying of root biomass, leaf shedding during growth of the crop and residues retained after harvesting (Athoni and Basavaraja, 2012). Hence, it is a good crop for rotation with cereals that do not have the capacity to fix nitrogen. Soybean has potential to fix between 44-300 kg N (Mohamedkheir et al., 2018), resulting in improvement of maize yields by 10%-20% (TechnoServe, 2011). This is a major benefit in African farming systems, where soils have been impoverished and fertilizers are either not available or are not affordable by smallholder farmers (Sinclair et al., 2014; Uwaoma, 2015). Studies have highlighted its importance in the control of *Striga hermonthica*, a parasitic maize weed (Kumar et al., 2015).

1.4 Soybean production constraints

The low soybean yields in SSA are mainly due to socio-economic, abiotic and biotic constraints. Biotic constraints include weeds, pests and pathogens. Weeds have been found to be the main detrimental factor, especially that some weed species have become resistant to glyphosate (Powles, 2010). Weeds can lead to serious yield losses of 80% or more, if not controlled (Billore et al., 2007). Dominant weed species in soybean production include *Cyperus rotundus, Caesulia axillaries, Digitaria sanguinalis, Echinochloa colona, Acalypha indica, Anotic monthuloni, Cynodon dactylon* and *Commelina benghalensis* (Hartman and Hill, 2010). Though soybean is attacked by more than 300 species of pathogens (Hartman et al., 2011), the major diseases resulting from pathogen attack in SSA include rust (*Phakopsora pachyrhizi*), frogeye leaf spot (*Cercospora sojina*), red-leaf blotch (*Phoma glycinicola*), bacterial blight (*Pseudomonas syringae* pv. *glycinea*), soybean mosaic virus and soybean cyst nematodes (*Heterodera glycines*) (Hartman and Hill, 2010). Apart from reducing the quality of the seed, pathogens also reduce seed yield of soybean. For example, yield losses of 55% or higher attributed to soybean rust have been reported in many soybean producing countries (Mueller et al., 2009). A number of pests can cause considerable economic damage to the crop, as they attack different parts of the plants, thereby reducing yield and seed quality. The major soybean pests include beetles, mites, loopers, aphids and stinkbugs (O'Neal and Johnson, 2010; Hartman et al., 2011).

Abiotic constraints affecting soybean production are drought, heat, floods, frosts, low soil P, K and pH, salinity, and photoperiod response. In many parts of Africa, drought and heat are

the primary abiotic constraints as they greatly contribute to lower yields. Recently, there has been an increase in droughts, further exacerbated by climate change (Ghosh et al., 2014). Soybeans are also sensitive to low soil phosphorous, potassium and pH (less than 5) and high soil salinity. These conditions lead to poor root development, reduced growth, leaf chlorosis and ultimate yield reduction (Katerji et al., 2003).

The socio-economic, technical and institutional factors facing soybean production in Africa include limited availability of stress resistant, high yielding and stable varieties (Lubungu et al., 2013); poor farming practices (late planting, poor disease management); and limited research and development (Opperman and Varia, 2011; Mohamedkheir et al., 2018). Others include low or no use of agricultural inputs (fertilizers, agricultural lime and inoculant); low technology transfer from research to farmers; and poor capital to increase the area under cultivation, infrastructure and crop marketing (Opperman and Varia, 2011). These constraints reduce cultivated area and profitability of the crop, and mainly affect smallholder farmers who are significant soybean producers in most African countries (Opperman and Varia, 2011; Meyer et al., 2018).

The above constraints can be addressed by better agronomic practices and crop improvement. Genetic improvement is instrumental in addressing these challenges through the development of cultivars that are able to grow in low pH, P and K soils, with level of tolerance to drought, heat, salinity, pests and diseases. Related *Glycine* species are a possible source of genes that can confer tolerance or resistance to biotic and abiotic constraints. It is absolutely necessary to localise soybean cultivar development so as to provide farmers with cultivars that are adaptable to prevailing conditions (Hartman et al., 2011).

1.5 Justification of the study

Despite the increase in soybean demand as a result of rapid population growth and increased livestock feed demand, yields in SSA are still low (1.1 ton/ha). Grain yield is a polygenic trait that is highly influenced by genotype by environment interaction (GEI) effects (Popovic et al., 2013). The potential genetic expression of a variety is often masked by the growing environment, resulting in low progress from phenotypic selection of polygenic characters like seed yield (Horn et al., 2018). Most soybean breeding programmes in Africa often overlook the effect of GEI and the concept of stability by selecting elite lines with good mean performance across testing environments and years (Gurmu et al., 2009). This approach can only be applied if there is no GEI and in this case a single environment can suffice for genotypes evaluation. However, such a situation rarely exists and GEI effect is, in most cases, the common phenomenon in yield trials and this limits the identification and selection

of superior genotypes for recommendation to target environments (Atnaf et al., 2013). The knowledge of the pattern and magnitude of GEI and stability of elite breeding lines during the final stages of testing is indispensable for identification and recommendation of superior cultivars (Horn et al., 2018). Analysis of GEI and stability has not received much attention in Southern Africa as evidenced by low availability of widely adapted soybean cultivars in the region (Mohamedkheir et al., 2018). Therefore, this study was undertaken in order to identify and recommend elite soybean lines for wide and specific production as well as to identify informative (discriminating) and representative test environments in four Southern African countries.

Developing cultivars that combine high yielding ability with many other desirable attributes is usually the ultimate goal of many soybean breeding programmes. However, grain yield is a complex trait and results from direct or indirect interaction of its component characters (Aditya et al., 2011). Generally, its response to selection is low due to low heritability (Gholizadeh and Dehghani, 2017). For these reasons, direct selection for seed yield has proven to be an ineffective approach (Aditya et al., 2011; Mulridharan, 2017). Therefore, it is imperative to carry out correlation and path coefficient analysis in order to know the yield contributing traits, the associations among them and with seed yield. This will help to develop an efficient breeding strategy for high grain yield in soybean (Pratap et al., 2012; Nwofia et al., 2016).

Correlation coefficient (r), measures the extent to which two characters are associated and helps to develop selection indices and permits prediction of correlated response (Islam and Rai, 2013). Path analysis on the other hand measures the contribution of various traits by partitioning their correlation coefficients into direct and indirect effects on grain yield (Rao, 2016). According to literature, path analysis has extensively been used to identify characters that contribute to seed yield in soybean but none of the researchers has identified a trait that is most related with yield (Machikowa and Laosuwan, 2011). This study, therefore sought to add to the already existing knowledge by identifying the most important traits for seed yield increase in soybean.

The success of a breeding programme in the genetic improvement of yield relies on the genetic diversity and the magnitude and nature of variability available in the germplasm. It is therefore, imperative to estimate genetic parameters, which are the measures of genetic variability. Genetic parameters indicate the amount of genetic gain that can be realised from selection and the best yield selection strategy to be employed (Mulridharan, 2017). The analysis of genetic diversity provides the breeder with key information for the identification and selection of suitable parents with desired alleles, which can be combined so as to increase seed yield in soybean (Rani et al., 2016).

1.6 Objectives

The overall objective of the study was to develop a selection criteria for seed yield through the studies of genetic variation and associations among important traits as well as to evaluate stability, adaptability and genetic diversity among elite soybean lines. The specific objectives were:

i) To determine the magnitude of genotype by environment interaction and stability of elite soybean lines for seed yield.
ii) To determine the nature of association among grain yield related traits, their direct and indirect effects on seed yield.
iii) To estimate genetic variation through genetic parameters of traits related to seed yield, and to analyse genetic diversity among elite soybean lines.

1.7 Hypotheses

The following hypotheses were tested:

i) Elite soybean lines respond differently in varied production environments in Southern Africa.
ii) There are no close associations between seed yield and its component characters, which can be used for indirect selection of seed yield.
iii) There is no genetic variation and diversity among elite soybean lines.

1.8 Dissertation outline

This dissertation consists of six independent chapters written in a journal paper format. For this reason, some information and references appear in more than one chapter. The Crop Science journal referencing style is used in this dissertation. The chapters are as follows:

Chapter 1: Dissertation introduction

Chapter 2: Literature review

Chapter 3: Genotype by environment interaction and stability of elite soybean lines across varied production environments in Southern Africa

Chapter 4: Genotype by trait associations, correlation and path coefficient analyses for seed yield and its component characters among elite soybean lines

Chapter 5: Genetic variability and diversity among elite soybean lines

Chapter 6: General overview of the study

1.9 References

Aditya, J., P. Bhartiya and A. Bhartiya. 2011. Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). Journal of Central European Agriculture 12: 27-34.

Athoni, B.K. and G.T. Basavaraja. 2012. Association analysis, genetic variability and genetic diversity studies in soybean [*Glycine max* (L.) Merrill]. Asian Journal of Bio Science 7: 114-118.

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

Azadbakht, L., R. Shakerhosseini, S. Atabak, M. Jamshidian, Y. Mehrabi and A. Esmaill-Zadeh. 2003. Beneficiary effect of dietary soy protein on lowering plasma levels of lipid and improving kidney function in type II diabetes with nephropathy. European Journal of Clinical Nutrition 57: 1292.

Billore, S., A. Vyas, O. Joshi, N. Pandya and I. Khan. 2007. Bio-efficacy of flumioxazin in soybean (*Glycine max*). The Indian Journal of Agricultural Sciences 77. 642-644.

Chen, Y.-m., S.C. Ho, S.S. Lam, S.S. Ho and J.L. Woo. 2003. Soy isoflavones have a favorable effect on bone loss in Chinese post-menopausal women with lower bone mass: a double-blind, randomized, controlled trial. The Journal of Clinical Endocrinology & Metabolism 88: 4740-4747.

FAO. 2017. Food and Agriculture Data. Food and Agriculture Organisation, Rome, Italy.

Gholizadeh, A. and H. Dehghani. 2017. Correlation and sequential path analysis between yield and related characters of wheat (*Triticum aestivum* L.) genotypes in non-stressed and salinity-stressed conditions. Romanian Agricultural Research 34: 37-49.

Ghosh, P., N. Kumar and A. Yadav. 2014. Role of resource conservation technologies in mitigating climate change. Resource Conservation Technology in Pulses: 312-325.

Gurmu, F., H. Mohammed and G. Alemaw. 2009. Genotype x environment interactions and stability of soybean for grain yield and nutrition quality. African Crop Science Journal 17: 87-99.

Hamilton-Reeves, J.M., S.A. Rebello, W. Thomas, J.W. Slaton and M.S. Kurzer. 2007. Soy protein isolate increases urinary estrogens and the ratio of 2: 16 α-hydroxyestrone in men at high risk of prostate cancer. The Journal of Nutrition 137: 2258-2263.

Hartman, G.L. and C.B. Hill. 2010. 13 Diseases of soybean and their management. The Soybean: Botany, Production and Uses: 276-299.

Hartman, G.L., E.D. West and T.K. Herman. 2011. Crops that feed the world 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. Food Security 3: 5-17.

Horn, L., H. Shimelis, F. Sarsu, L. Mwadzingeni and M.D. Laing. 2018. Genotype by environment interaction for grain yield among novel cowpea (*Vigna unguiculata* L.) selections derived by gamma irradiation. The Crop Journal 6: 306-313.

Hymowitz, T. 1990. Soybeans: the success story. Advances in new crops: 159-163.

Islam, A. and P. Rai. 2013. Character association of soybean (*Glycine max*) lines for yield contributing traits. Bangladesh Journal of Plant Breeding and Genetics 26: 33-38.

Jain, R.K., A. Joshi, H.R. Chaudhary, A. Dashora and C.L. Khatik. 2017. Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Legume Research- An International Journal 41: 532-536.

Kanchana, P., M.L. Santha and K.D. Raja. 2015. A review on *Glycine max* (L.) Merr. (soybean). World Journal of Pharmacy and Pharmaceutical Sciences 5: 356-371.

Katerji, N., J. Van Hoorn, A. Hamdy and M. Mastrorilli. 2003. Salinity effect on crop development and yield, analysis of salt tolerance according to several classification methods. Agricultural Water Management 62: 37-66.

Kumar, A., A. Pandey, C. Aochen and A. Pattanayak. 2015. Evaluation of genetic diversity and interrelationships of agro-morphological characters in soybean (*Glycine max* L.) genotypes. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences 85: 397-405.

Liu, K. 1997. Chemistry and nutritional value of soybean components. In: Soybeans. Springer, Boston, MA. p. 25-113.

Lubungu, M., W.J. Burke and N.J. Sitko. 2013. Challenges of smallholder soybean production and commercialization in Eastern Province of Zambia. Michigan State University, Department of Agricultural, Food and Resource Economics.

Machikowa, T. and P. Laosuwan. 2011. Path coefficient analysis for yield of early maturing soybean. Songklanakarin Journal of Science & Technology 33: 365-368.

Maskarinec, G., A.G. Aylward, E. Erber, Y. Takata and L.N. Kolonel. 2008. Soy intake is related to a lower body mass index in adult women. European Journal of Nutrition 47: 138-144.

Mateos-Aparicio, I., A. Redondo Cuenca, M. Villanueva-Suárez and M. Zapata-Revilla. 2008. Soybean, a promising health source. Nutricion Hospitalaria 23: 305-312.

Meyer, F., L.N. Traub, T. Davids, B. Chisanga, R. Kachule, E. Tostão. 2018. Modelling soybean markets in eastern and southern Africa, regional network of agricultural policy research institutes (RENAPRI). EUR 28978 EN, Publications Office of the European Union, Luxembourg, doi:10.2760/2059.

MOA and CSO. 2018. Crop Forecast Survey. Ministry of Agriculture and Central Statistics Office, Lusaka, Zambia.

Mohamedkheir K, Seifeldin I, Sapey E and H. Tianfu. 2018. History, current status, and prospects of soybean production and research in sub-Saharan Africa. The Crop Journal 3: 226-235.

Mpepereki, S., F. Javaheri, P. Davis and K. Giller. 2000. Soybeans and sustainable agriculture: promiscuous soybeans in southern Africa. Field Crops Research 65: 137-149.

Mueller, T., M. Miles, W. Morel, J. Marois, D. Wright, R. Kemerait. 2009. Effect of fungicide and timing of application on soybean rust severity and yield. Plant Disease 93: 243-248.

Mulridharan, P. 2017. Genetic analysis of traits determining seed yield over environments in soybean. PhD Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

Nwofia, G., R. Edugbo and E. Mbah. 2016. Interaction of genotype x sowing date on yield and associated traits of soybean [*Glycine max* (L.) Merrill] over two cropping seasons in a humid agro-ecological zone of South-Eastern Nigeria. Journal of Agricultural Sciences 11: 164-177.

O'Neal, M.E. and K.D. Johnson. 2010. 14 Insect pests of soybean and their management. The Soybean: Botany, Production and Uses: 300-324.

Opperman, C. and S. Varia. 2011. Technical report: soybean value chain. AECOM International Development. URL: www.satradehub.org (date accessed: July 19, 2018).

Pawar, N. 2013. Characterization and categorization of major growth and yield traits in soybean [*Glycine max* (L.) Merrill. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

Popovic, V., J. Miladinovic, M. Tatic, V. Djekic, G. Dozet, V. Đukić. 2013. Stability of soybean yield and quality components. African Journal of Agricultural Research 8: 5651-5658.

Powles, S.B. 2010. Gene amplification delivers glyphosate-resistant weed evolution. Proceedings of the National Academy of Sciences 107: 955-956.

Pratap, N., P. Singh, R. Shekhar, S. Soni and A. Mall. 2012. Genetic variability, character association and diversity analyses for economic traits in rice (*Oryza sativa* L.). South Asian Association for Regional Cooperation Journal of Agriculture 10: 83-94.

Rani, A., V. Kumar, B. Gill, V. Morya and R. Rawal. 2016. Assessment of genetic diversity of soybean genotypes differing in resistance against Yellow Mosaic Virus using Simple Sequence Repeat Markers. Legume Research 39: 674-681.

Rao, V.T. 2016. Genetic variability, correlation and path coefficient analysis under drought in groundnut (*Arachis hypogaea* L.). Legume Research 39: 319-322.

Shurtleff, W. and A. Aoyagi. 2009. History of soybeans and soyfoods in Africa (1857-2009): Extensively Annotated Bibliography and Sourcebook. Soyinfo Center.

Sinclair, T.R., H. Marrou, A. Soltani, V. Vadez and K.C. Chandolu. 2014. Soybean production potential in Africa. Global Food Security 3: 31-40.

Sleper, D.A. and J.M. Poehlman. 2006. Breeding field crops. Blackwell publishing, USA.

TechnoServe. 2011. Southern Africa Regional Soybean Roadmap: Final Report. URL: www.technoserve.org/resources (date accessed: August 20, 2018).

Tefera, H. 2011. Breeding for promiscuous soybean at international institute of tropical agriculture. Soybean-Molecular Aspects of Breeding. Institute for New Technologies 7: 147-162.

USDA. 2018. World agricultural supply and demand estimates. United States Department of Agriculture. ISSN: 1554-9089. Internet site: http://usda. mannlib. cornell. edu/reports/waobr/wasde-bb (Accessed July 19, 2018).

Uwaoma, I.G. 2015. Economics of small scale soybean processing firms in Anambra State, Nigeria. PhD Thesis, Department of Agricultural Economics, University of Nigeria, Nigeria.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

This chapter reviews the literature, which is related to the objectives of this study. The taxonomy, morphology, origin, domestication of soybean and soybean breeding in sub-Sahara Africa (SSA) are discussed. It also reviews the genotype by environment interaction (GEI) and its importance and implications in plant breeding. Adaptability, stability and methods of analysing GEI and stability (AMMI and GGE biplot) are also highlighted in this chapter. Correlation, path and genotype by trait analyses, genetic variability and diversity and their importance in soybean improvement are also discussed.

2.2 Taxonomy of soybean

Soybean is an annual herbaceous legume plant that belongs to the family Leguminosae, sub-family Papilionoideae, tribe Phaseoleae, genus *Glycine* Willd, subgenus *soja* (Moench) F.J. Herm., and species *Glycine max* (L.) Merr., (Nwokolo, 1996; Singh et al., 2007). Linnaeus was the first to propose the genus name *Glycine* in his first edition of Genera *Plantarum.* The cultivated species first appeared in the edition, 'Species *Plantarum*', under the name *Phaseolus max* (L.). In 1917 Merrill proposed the combination, *Glycine max* (L.) Merr., and this name has since become the valid name for cultivated soybean (Hymowitz and Newell, 1981).

The genus *Glycine* comprises two sub-genera, *Soja* and *Glycine.* The sub-genus *Soja* (Moench) F.J. Herm., consists of two species, including *Glycine max* (L.) Merrill., the domesticated and cultivated soybean, and its wild ancestor, *Glycine soja* Sieb. & Zucc (Singh, 2017). The two species are annuals and have 40 chromosomes (2n=2x=40). They are self-fertilising species with less than 1% outcrossing and grow in varied climates, including tropical, sub-tropical and temperate regions (Sleper and Poehlman, 2006). *Glycine soja,* which is native to China, Japan, Taiwan, Korea and Russia, has a viny prostate growth and tends to shatter a lot. Both species are cross-compatible but *Glycine soja* is rarely used in a crossing programmes due to its undesirable characteristics unless only for transfer of desirable genes to *Glycine max* (Sleper and Poehlman, 2006). The sub-genus *Glycine* is composed of 26 wild species which are perennial in nature and are found in Australia, New Zealand and other Pacific Islands (Singh, 2017). The species in this sub-genus have varied numbers of chromosomes (2*n* = 38, 40, 78, 80). Examples of species in this sub-genus include: *G. tabacina, G. tomentella, G. hirticaulis, G. canescens* and *G. dolichocarpam.*

Glycine max and species in the sub-genus *Glycine* are cross-incompatible, although limited success was obtained between *G. tomentella and G. max* (Sleper and Poehlman, 2006).

2.3 Morphology of soybean

Soybean is a bushy, erect herbaceous annual plant that grows 30-150 cm in height, depending on the type. Three growth types are common i.e. indeterminate, determinate and semi-determinate habits (Singh, 2017). The determinate growth type has a longer vegetative period than the other types. When the terminal bud of this type ceases its vegetative activity, it ends in axillary and terminal racemes with 5-12 pods attached. Indeterminate types continue to increase in number of nodes and plant height throughout the flowering period. Semi-determinate types are intermediate between the two types and have indeterminate stems that end vegetative growth in a raceme similar to the determinate ones (Bernard, 1972). It initially exhibits taproot growth, which is followed by secondary root development at a later stage. Apart from anchoring the plant, the roots play an important function in supplying the plant with nitrogen by forming nodules and establishing a symbiotic relationship with the bacterium that fixes nitrogen (*Bradyrhizobium japonicum*). Four different types of leaves have been reported in soybean. These include the seed leaves, simple primary leaves, pinnately trifoliolate leaves and the prophylls (Lersten and Carlson, 2004).

The reproductive stages follow vegetative growth and are characterised by development of clusters of flowers and pods. Flower colour may be purple or white (Nwokolo, 1996). Soybean is highly self-pollinated (less than 1% out-crossing) due to the stigma being surrounded by the stamens (Singh, 2017). Each inflorescence on the nodes of the plant develops into one to more than 20 pods. Hairy, green immature pods turn to yellowish-brown upon reaching maturity. There are one to three and rarely four seeds per pod. The seeds take about 50-80 days to mature after fertilization, depending on environmental factors and variety. The most common colours of the seed include yellow, brown and tan, while green and black are rare (Lersten and Carlson, 2004). Each seed has a hilum, which is the visible scar on the seed coat. Helium colour ranges from yellow, grey, buff, brown and black. The hilum has a micropyle, which is a small opening into the seed coat. The micropyle is important in seed germination as it is the entry point for water (Kanchana et al., 2015). Soybean requires 25-32ºC temperature and 400-800 mm rainfall for its optimal performance. It can be grown in a variety of ecological zones (Nwokolo, 1996). It requires fertile, well- drained and mildly acidic soils, although it can tolerate various soil types. Being a day length sensitive plant, it flowers quicker when grown under short day conditions.

2.4 Origin and domestication of soybean

Soybean is believed to have been domesticated from its wild ancestor, *Glycine soja* Sieb. & Zucc (Hermann, 1962; Nwokolo, 1996). The ancestor has a climbing growth habit, with pods containing black seeds that have lower oil content (9-12%) but rich in protein (31-52%) (Hymowitz et al., 1972). The ancestor *G. soja* exhibits high level of shattering as soon as it reaches maturity. It is native to Japan, Taiwan, Korea, far eastern Russia and China (Singh and Hymowitz, 1999).

The domestication of soybean is suggested by geographical, linguistic and historical evidence to have taken place in North-eastern China between 1700-1100 B.C (Hymowitz, 1990). From Northern China, soybean spread to other parts of China and then to the Korean peninsula by the first century. By the Age of Discovery (15-16thcentury), soybean had spread to many other several countries in Asia, including Japan, north India, Burma, Malaysia, Vietnam, Thailand, Indonesia, the Philippines and Nepal (these are referred to as secondary gene centres) (Hymowitz, 1990). From Asia the crop spread to many other parts of the world, including sub-Saharan Africa (SSA).

The crop has a short history of commercialisation in SSA, where it was only introduced through the eastern coast by Chinese traders in the 19th century (Mpepereki et al., 2000). South Africa was the first country to record soybean cultivation in SSA in 1903, followed by Zimbabwe in 1906, Tanzania in 1907 and Malawi in 1909. It is reported to have first been cultivated in Zambia in 1910 and Mozambique in 1915 (Shurtleff and Aoyagi, 2009).

2.5 Soybean breeding in sub-Saharan Africa

According to Mohamedkheir et al. (2018) soybean breeding in SSA was first started in Nigeria in the 1960s. The national research programmes, the International Institute of Tropical Agriculture (IITA), private breeders and universities have been the main players conducting soybean breeding in SSA (Mohamedkheir et al., 2018). The IITA started soybean improvement in 1974 and ever since the institute has been leading in developing soybean varieties that are high-yielding and self-nodulating (Tefera et al., 2009). Other private companies contributing to variety development in SSA include: ZamSeed, MRI Syngenta and SeedCo in Zambia; SeedCo in Zimbabwe; and Monsanto, Pannar Seeds, DuPont Pioneer and Agricol in South Africa (Mohamedkheir et al., 2018).

Investments in soybean improvement and the impacts of research in SSA have for a long time been lower and continue to be so compared to other regions of the world (Mohamedkheir et al., 2018). Adoption of latest breeding technologies (e.g. marker assisted breeding) in soybean improvement still lags behind in SSA compared to other soybean

producing regions like USA and South America. Another major challenge in SSA is the low numbers of qualified soybean breeders in many national soybean programmes (Alene et al., 2015). This has resulted in negligible contribution of national soybean programmes to the number of released cultivars in SSA (Mohamedkheir et al., 2018).

Lack of proper processing and utilisation methods, losses due to shattering, poor self-nodulation, low seed viability and yield are the main challenges that needed research attention and have continued to be addressed by most soybean breeding programmes in SSA (Tefera et al., 2009; Mohamedkheir et al., 2018). Therefore, soybean breeding in SSA has been aiming at developing stable cultivars that are both high yielding and tolerant to many abiotic and biotic challenges (Tefera, 2011). This is due to low yields being recorded in SSA compared to other continents (Mohamedkheir et al., 2018). Other important traits that are being considered for improvement include natural nodulation, quality (oil, protein), maturity, seed size, seed colour and pod shattering (Tefera, 2011; Mohamedkheir et al., 2018).

The number of released cultivars in SSA was over 195 by 2015 (Alene et al., 2015), with most of them being released by IITA, private companies and breeders and very few by national programmes (Mohamedkheir et al., 2018). In SSA, only South Africa has approved the release and production of genetically modified cultivars with genes that confer tolerance to glyphosate (Roundup Ready). The number of registered RR cultivars stood at 89 in 2015 (Sadie, 2015) and over 90% of soybean acreage in South Africa was occupied by RR varieties in 2017 (Mohamedkheir et al., 2018).

Significant genetic gains have been attained in soybean breeding in SSA. It was reported by Shurtleff and Aoyagi (2009) that yields were as low as 534 kg/ha among the earliest released cultivars. For example, Tefera et al. (2009) reported 24.1 and 23.6 kg/ha as the average annual increase of grain yield from 1980 to 1996 among the early and medium maturing cultivars, respectively. These represented genetic gains of 2.2% and 1.99% for early and medium maturing cultivars, respectively. They further reported 22.8 kg/ha as the increase in fodder yield per year and 1.72% genetic gain for natural nodulation in the same period. Cultivars with potential to yield as high as 5 ton/ha are available for commercial production (Mushoriwa, 2013) but their stability across all production areas in SSA is not known. These represent relatively high significant breeding gains over the years compared to 0.53 tons/ha reported among the earliest cultivars (Shurtleff and Aoyagi, 2009). However, there is still need for more investment and concerted effort in soybean improvement in SSA in order to accelerate genetic gains and reduce the yield gap between SSA and the Americas (Mohamedkheir et al., 2018).

To avoid narrowing the genetic base through repeated use of few elite lines in crosses (Mushoriwa, 2013), the approach used by breeding programmes in SSA is to combine high yielding breeding materials from USA, Argentina and Brazil with the promiscuous landraces from Asia and locally adapted lines in order to develop high-yielding tropical soybean varieties (Brink and Belay, 2006; Mohamedkheir et al., 2018). Hybridization, followed by selection is the main methodology used in conventional variety development (Tefera, 2011). Pedigree, backcross and modified single-seed descent are the main breeding methods used in soybean improvement (Sleper and Poehlman, 2006; Tefera, 2011). Among the three, the modified single-seed descent selection is the more commonly used method because it makes it possible to rapidly advance generations at a low cost (Sleper and Poehlman, 2006).

2.6 Genotype by environment interaction and stability

The term phenotype is the physical appearance or perceptible trait of an individual (Xu, 2010). The environment (E), genotype (G) and genotype by environment interaction (GEI) components determine the phenotypic expression of an individual (Sharifi et al., 2017). The genotype of an individual is its genetic composition, i.e., the transmissible DNA nucleotide sequence. The total sum of conditions surrounding an individual or a group of individuals is known as the environment (Yan and Kang, 2003). These conditions include soil type and fertility, precipitation, temperature, photoperiod, cultural practices, biotic and abiotic stresses (Kerby et al., 2000) which differ across years, seasons and locations. The genetic make-up of an individual generally remains unchanged in different environments. Therefore, any variation in phenotypic expression of a specific genotype is attributable to the environment (Yan and Kang, 2003). The inconsistent performance of a genotype in varied environments is called the genotype by environment interaction (GEI) (Rao et al., 2002; Sharifi et al., 2017). The GEI is only valuable if it is significant and causes genotypes to rank differently (crossover GEI) or if the genotypes' magnitude of differences change in varied environments (non-crossover GEI) (Haldane, 1946). The crossover type is important for selection of specifically adapted genotypes (Kaya et al., 2006). The non-crossover interactions are mostly associated with heterogeneity of the variance and non-additivity of the genotype by environment data matrix (Fernandez, 1991). The non-crossover type is important when selecting widely adapted genotypes (Kaya et al., 2006).

Generally, the contribution of GEI to genotypic variation is usually small. About 80% or more of the variation is explained by environment (E) and genotype (G) effects (Yan and Kang, 2003). If significant GEI is found to be present, the interest of researchers would be to know the causes of the interaction to enable them predict the performance of genotypes under varied environments (Yan and Kang, 2003). Environments in the four Southern African

countries (Zambia, Malawi, Mozambique and Zimbabwe) where the genotypes were tested are highly varied in terms of latitudes, altitudes, rainfall, temperature, wind, day length, diseases and pests, soil moisture, soil type and fertility (Mushoriwa, 2013). Variability in these environmental conditions across years, locations and seasons causes significant GEI among soybean genotypes (Bull et al., 1992; Mushoriwa, 2013).

2.7 Implications and importance of GEI in breeding of crops

The GEI confounds comparisons among genotypes and the definition of breeding objectives also becomes complicated (Xu, 2010). The Interaction between genotypes and the environment negatively affects heritability as it is the part of the denominator of the heritability equation (whether narrow or broad sense) (Yan and Kang, 2003). Genetic advance from selection is determined by the heritability of the trait of interest. When the heritability of a trait is low, progress from its selection is reduced (Kaya et al., 2006; Nwangburuka et al., 2011). Grain yield being a polygenic trait, is highly affected by the environment. Hence, improving yield is not always easy due to low genetic gain from its selection (Horn et al., 2018). The GEI confounds precise partitioning of the contributions of improved cultivars or environment to yield (Silvey, 1981; Crossa, 1990). The cost of testing of breeding lines increases due to the requirement of additional sites and years for testing. The correlation between genotypic and phenotypic values is reduced when there is a significant crossover GEI. This makes it difficult to identify and recommend superior genotypes across environments (Nwangburuka et al., 2011; Kumar et al., 2014).

Multi-environment trials are required for understanding and determining the usefulness of GEI. It has been reported by researchers that the knowledge of GEI is important in the allocation of resources for testing of breeding lines across environments and development of cultivars for specific purposes (specific adaptation) and improved (wide) adaptation to the target environment (Gurmu et al., 2009; Xu, 2010; Mustapha and Bakari, 2014).

2.7.1 Adaptability concepts

Adaptability is the ability of a cultivar to take advantage of the environmental effects to guarantee high yield, while stability is related with the yield maintenance and predictability of a cultivar in variable environments (Borém, 1998; Cruz et al., 2004). Multi-locational testing is a useful tool for variety adaptability and stability (Miladinovic et al., 2006). Two types of adaptation exist, i.e. specific or narrow and wide or general adaptation. Annicchiarico (2002) classified cultivars that perform well in specific environments as having specific adaptation. The benefits of specific adaptation to breeders is the high genetic gains achieved through exploiting positive interaction effects of cultivars with individual environments (Mushoriwa,

2013). In a study done by Annicchiarico et al. (2005), it was found that specific adaptation gave higher yield gains than wide adaptation. Specific adaptation gives assurance of food security in specific environments, where different varieties are commercialized. It helps to reduce GEI and increases crop yields (Annicchiarico, 2002). It is usually exploited by national programmes and seed companies that operate in several countries with different environmental conditions. However, specific adaptation increases cost of testing as many sites and genotypes are required (Yan and Kang, 2003).

A widely adapted or stable genotype is described as the one with the capacity to utilize the available resources in good environments and performs above the grand mean in all environments (Gurmu, 2017). Bekheit (2000) established that low yielding genotypes are widely adaptable, while high yielding genotypes have a high likelihood of being unstable. According to Gebeyehu and Assefa (2003) selecting only high yielding genotypes could lead to loss of several stable, low yielding genotypes. The merit of wide adaptation is the increased precision of the genotypic means since data are pooled from several environments (Atlin et al., 2000). Wide adaptation conserves resources because of the use of few representative and informative environments as opposed to testing genotypes in all environments (Yan and Kang, 2003).

2.7.2 Stability concepts

Generally, two stability concepts exist, i.e. the static and dynamic stability. In the static or biological stability concept, a stable genotype is described as one that performs constantly in spite of any changes in environmental conditions. Such a genotype does not respond to improvements in environmental conditions like increased input levels (Becker and Leon, 1988). Static stability is important for disease resistance and quality traits. This concept, however has received minimal attention from agronomists and breeders, who have preference for genotypes that are capable of responding to favourable environmental conditions such as improved input levels (Becker, 1981; Kaya and Ozer, 2014). The dynamic or agronomic concept describes a stable genotype as one with predictable performance across environments (Becker and Leon, 1988). Such genotypes respond to improvements in environmental conditions and preferred in commercial plant breeding (Alwala et al., 2010).

2.7.3 Methods of analysing GEI and stability

The analysis of GEI aids breeding programmes to identify and recommend specifically and broadly adapted cultivars to farmers (Sabaghnia et al., 2013). Numerous methods have been used to analyse the extent and nature of GEI under variable growing conditions and to determine the adaptation of genotypes. These methods fall into any of the following

categories: univariate parametric, non-parametric and multivariate methods (Lin et al., 1986; Tadege et al., 2014). Examples of univariate non-parametric methods include variance ranks ($S^{(2)}_i$) and mean absolute rank differences ($S^{(1)}_i$). The deviation from regression (S^2_{di}), cultivar superiority measure (P_i), coefficient of determination (r_i^2), Eberhart and Russel regression coefficient (b_i), Wricke's ecovalence (W_i), Shukla's stability variance (σ^2_i), coefficient of variation (cv), and environmental variance (S^2_{xi}) are examples of univariate parametric methods (Tadege et al., 2014).

The weakness of parametric methods like regression and analysis of variance (ANOVA) is that they cannot reliably predict the overall response and stability of genotypes (Alwala et al., 2010). This is because they consider the response of a genotype to environments to be a univariate problem, which in the actual sense is a multivariate situation (Alwala et al., 2010). The ANOVA being an additive model, only reveals if the GEI is significant or not. It is incapable of providing insights into the environments and genotypes that lead to the interaction. It is for this reason that the two multivariate methods called the genotype plus genotype by environment interaction (GGE) and Additive Main effects and Multiplicative Interaction (AMMI) models have recently been used to analyse GEI and stability of genotypes in variable environments. These two methods combine univariate methods for the additive genotype and environment effects with a multivariate method for the multiplicative effect of the GEI, thus providing a better interpretation of multi-environmental data set (Bhartiya et al., 2017). Both methods require that the genotypes should be tested in multi-locations and the trial may or may not be replicated (Gauch, 2006).The two methods were used in this study.

2.7.3.1 Additive Main effects and Multiplicative Interaction (AMMI)

The AMMI model proposed by Zobel et al. (1988), is an inclusive approach for genotype by environment and stability analysis (Gauch and Zobel, 1988; Vaezi et al., 2017). The results are presented graphically in a biplot, which shows the E, G and GEI effects on the same scatterplot (Crossa et al., 1991; Purchase, 1997). The interactive principal component axis (IPCA) scores are plotted against each other in a biplot, making it easier to visualise and interpret the GEI components (Mustapha and Bakari, 2014). The genotypes with IPCA scores closer to zero are stable across environments (Alwala et al., 2010). However, a poor yielding genotype can also have an IPCA score closer to zero. Therefore, these values should be used alongside the mean yields of genotypes across environments. The genotypes with high IPCA values and mean yields are specifically adapted to the environments with the same size IPCA 1 values (Xu, 2010).

Accuracy in the AMMI model is also achieved by separation of structural variation from noise (Nassir and Ariyo, 2011). It is, therefore a valuable method for easier understanding of

complex GEI and better estimation of yields of genotypes across environments (Alberts, 2004).

Several stability parameters like the AMMI stability value (ASV), AMMI stability measures (EVi and EVF) and sums of the absolute value of the IPCA scores (SIPCi and SIPCF), have been proposed for use in identifying stable genotypes across environments. The ASV for the genotypes is calculated from the first and the second IPCA scores and used to rank them in relation to their stability. Genotypes that have the lowest ASV (close to zero) are considered to be the most stable (Temesgen et al., 2015). The stable genotypes also have short vectors from the origin of the biplot. The ASV is similar to other stability methods such as those used by Shukla (1972) and Eberhart and Russell (1966). It has been reported to be a good criteria for identifying and selecting high yielding and stable genotypes (Sabaghnia et al., 2008; Karimizadeh et al., 2012).

There has been a wide use of the AMMI method in soybean research by many scientists in the effort to understand GEI and identifying stable genotypes. It was used by Gurmu et al. (2009) and Bhartiya et al. (2017) who identified Awassa-95 and C11, respectively, as the most stable soybean genotypes. It has also been widely used in other crops, like rice (Sharifi et al., 2017), faba beans (Temesgen et al., 2015) and cowpea (Simion et al., 2018).

2.7.3.2 Genotype plus genotype by environment interaction (GGE) biplot

Yield data collected from multi-location trails (MLT) are, in most cases, large and pose a challenge in understanding the general pattern of the GEI if they are not graphically presented (Farshadfar et al., 2013). The GGE biplot method was proposed by Yan et al. (2000) to solve this problem as it allows the visual inspection of the general pattern of GEI (Yan and Tinker, 2006). There are two concepts that are emphasised by the GGE-biplot. Firstly, despite yield being a function of the G, E, and GEI effects; the E must be excluded in the evaluation of genotypes and only the G and GEI (especially the repeatable GEI) are considered to be relevant. The G and GEI together are known as the GGE. Secondly, the biplot method, which was originally suggested by Gabriel (1971), displays and estimates the GGE of a MLT and hence the term GGE biplot. The biplot is formed by the first two principal components, PC1 (primary effect) and PC2 (secondary effect) (Kaya et al., 2006) derived from the subjection of the MLT yield data to singular value decomposition (Yan and Tinker, 2006).

The basic characteristics of a GGE biplot include a small circle at the centre, indicating the average environment coordination (AEC) and a single arrowed line that passes through the origin of the biplot, called the AEC abscissa. The arrow of AEC abscissa points in the

direction where yield is increasing (Alwala et al., 2010). Another feature of the biplot is the double arrowed line, called the AEC ordinate, passing through the biplot origin and is perpendicular to the AEC abscissa. The AEC ordinate is a measure of instability of genotypes as it moves away from the biplot origin and also separates the above average genotypes from those yielding below the grand mean. Stability of genotypes is indicated by their vertical distance from the AEC abscissa (Alwala et al., 2010). The longer the vertical distance from the AEC abscissa to the genotype, the more unstable the genotype is. The genotypes with shorter vertical distances from the AEC abscissa are highly stable (Yan and Kang, 2003).

The GGE biplot is a very effective tool for evaluation of genotypes (their stability and mean yield) and test-environments (their representativeness and discriminating capacity). It is also useful for mega-environment analysis, such as the "which won where" pattern, which helps to recommend genotypes that are specifically adapted to specific mega-environments (Amira et al., 2013). The GGE biplot technique has, in the recent years, been widely used to study GEI and stability in soybean (Amira et al., 2013; Atnaf et al., 2013; Adie et al., 2014; Bhartiya et al., 2017). It has also received attention in other crops like cowpea (Horn et al., 2018), bread wheat (Kaya et al., 2006), chickpea (Farshadfar et al., 2013), millet (Zhang et al., 2016) and barley (Vaezi et al., 2017).

When comparing the GGE biplot to AMMI, Adie et al. (2014) and Ndhlela et al. (2014) pointed out that the GGE biplot is founded on environment-centred PCA, while AMMI analysis is based on double-centred PCA. Despite the fact that both techniques combine G with GE in the analysis of mega-environments and evaluation of genotypes, the GGE biplot explains the G and GE more explicitly and has the inner-product property of a biplot (Yan et al., 2007). The GGE biplot was suggested to be better than AMMI by Yan and Kang (2003) in terms of being biological and logical in explaining PC1 scores, which are a representation of genotype effects as opposed to additive main effects. The discriminating power and representativeness power of test-environments is viewed better in GGE biplot than in AMMI analysis. Yan et al. (2007) argued that the AMMI model does not accurately identify 'which-won-where' genotype as does the GGE biplot. However, Gauch (2006) suggested that the AMMI model more uniquely separates the E, G main effects and GEI, and structural variation from noise.

2.8 Correlations and path coefficient analysis

Seed yield is a polygenic character that is highly affected by the environment in which a genotype is grown. As a consequence of this, it generally has low heritability, making its improvement through direct selection to be difficult (Mahawar et al., 2013; Jain et al., 2017). The common practice by breeders is to indirectly select for yield through its component

characters, which usually have high heritability and are less influenced by the environment (Maleki et al., 2011; Pawar, 2013). Hence, the success of a breeder in improving seed yield depends on the extent to which it is correlated with its component characters (Kearsey and Pooni, 1998; Sharma and Sharma, 2012). It is, therefore imperative to know the nature of correlations that exist between seed yield and its related traits and among the yield components themselves in order to predict the correlated response to selection (Rao, 2016).

Selection for a target trait like seed yield would be much easier if it is positively associated with it related traits. Nevertheless, key traits are not usually positively correlated with each other (Lewis, 2006). For this reason, correlations alone may not sufficiently explain how important each trait is in determining seed yield as the number of these independent variables increases (Maleki et al., 2011). Therefore, it is vital to partition the correlation coefficients into direct and indirect effects through path coefficient analysis (Malik et al., 2007; Chet et al., 2010; Athoni and Basavaraja, 2012). In path coefficient analysis, the predictor traits are considered as first-order variables and their effects on the dependable variable, such as grain yield, are analysed (Ali et al., 2009). Path coefficient analysis is important to breeders because it helps to clearly identify the key characters that determine yield. This in turn helps breeders to develop a selection criteria for grain yield improvement (Aditya et al., 2011; Jain et al., 2015; Bhartiya and Aditya, 2016).

Correlation and path analyses have extensively been conducted by many soybean researchers, however, none of these researchers has clearly identified a trait that mostly determines seed and can be used in all breeding programmes (Machikowa and Laosuwan, 2011). Different results have been obtained by many researchers and this could be attributed to the influence of the environment in different locations where the experiments are conducted and the use of different genotypes (Mushoriwa, 2013). Malik et al. (2007), Aditya et al. (2011), Machikowa and Laosuwan (2011), Mahawar et al. (2013), Mushoriwa (2013), Jain et al. (2015), Kuldeep (2015), and Bhartiya and Aditya (2016) reported 100 seed weight, pods per plant, plant height, days to flowering and maturity and number of branches per plant to be positively correlated with seed yield. They further pointed out that dry matter weight, number of branches per plant, pod number per plant, plant height, harvest index and 100 seed weight were the traits with the most direct effects on seed yield. On the contrary, Mulridharan (2017) found no correlation between grain yield and the following traits: dry matter per plant, number of branches per plant, 100 seed weight, plant height, pod number per plant and days to 50% flowering.

2.9 Genotype by trait analysis

The overall objective of many crop improvement programmes is to identify genotypes that are favourable for most desirable characters. In other words, the most desirable cultivars should combine high yield with other important traits to enhance their acceptability (Sharifi and Ebadi, 2018). A number of methods are used to understand inter-relationships among traits and genotypes in various crops. A more recent method called genotype by trait biplot, has been gaining popularity in exploring multi-character data with a view to studying the relationships among traits and genotypes (Yan and Kang, 2003). This method uses the principles of the GGE biplot analysis except that different units for the characters are removed by standardising the data before the biplot is constructed. The other difference with the GGE biplot is that traits are put in place of environments and are used as testers (Yan and Kang, 2003). It graphically shows how well traits are correlated with one another across genotypes and helps to visualise trait profiles (merits and weaknesses) of genotypes in order to identify superior genotypes in certain desirable traits (Yan and Rajcan, 2002). Such genotypes can be recommended for release or possibly be used as parents in improvement programmes (Oliveira et al., 2018). It also assists in identifying traits that can be used to indirectly select for the trait of interest (e.g. yield) and those that are not important in its improvement (Sofia, 2016; Atnaf et al., 2017).

The genotype by trait biplot has been employed in the studies of trait associations and evaluation of genotype in a number of crops. These include white lupins (Atnaf et al., 2017), forest trees (Ukalski and Klisz, 2016), coconut (Odewale et al., 2014), rice (Sharifi and Ebadi, 2018), green beans (Oliveira et al., 2018), maize (Kaplan et al., 2017) and soybean (Yan and Rajcan, 2002).

2.10 Genetic variability

The effectiveness of selection and improvement of any trait, mainly relies on size and nature of the genetic variability existing in the population (Pawar, 2013; Rao, 2016). If greater genetic variability is available in the population, there will be higher chances of developing a cultivar with broad genetic base and high resistance to biotic and abiotic challenges (Smith et al., 1991). The variability present in a population is determined by genetic parameters, which include phenotypic and genotypic coefficients of variation (PCV, GCV), heritability and genetic advance (Aditya et al., 2011). It is essential to estimate these parameters as they determine how best the available germplasm can be utilised in a crop breeding programme (Kuldeep, 2015).

Heritability is a measure of the proportion of the phenotypic variance that is heritable (Kearsey and Pooni, 1998). Since heritability is determined by all the variance components, its value is affected by change in any one of them (Bhandarkar, 1996). Genetic advance is a measure of progress that can be realised from selection of the desirable trait (Mulridharan, 2017). Atnaf et al. (2017) reiterated that heritability should be accurately estimated in order to successfully improve the trait of interest. Although heritability alone would be meaningless unless it is estimated alongside genetic advance in order to be useful in choosing the selection strategy to be employed (Johnson et al., 1955). High heritability is not always a suggestion of high genetic gain. However, high heritability together with high genetic advance, imply that the trait of interest can easily be improved by simple selection based on its phenotypic expression (Mulridharan, 2017).

Several researchers have indicated the importance of genetic parameters in determining genetic variability and yield improvement in soybean. Athoni and Basavaraja (2012) in their research found plant height to have the highest PCV, GCV, heritability and genetic advance. Aditya et al. (2011) reported dry matter weight per plant, grain yield per plant and pod number per plant have the highest GCV and PCV values. They also found pod number per plant to have the highest heritability and genetic advance. Jain et al. (2015) also found high heritability for 100 seed weight, days to 50% flowering and plant height. Bhartiya and Aditya (2016) also found 100 seed weight to have the highest heritability and genetic advance. Similarly plant height, harvest index and pod number per plant were reported to have the highest heritability and genetic advance by Jain et al. (2017). All these researchers suggested high potential of these characters to improve grain yield of soybean.

2.11 Genetic diversity based on morphological traits

Crops, such as soybean, which is highly self-pollinated, have narrow natural diversity, resulting in impediment of effective selection for seed yield. Genetic diversity is defined as the amount by which the heritable traits of crops differ within a population (Pervin et al., 2007). Genetic diversity studies, which should be carried out prior to implementation of soybean yield improvement, aid in identifying appropriate parents for hybridisation (combination of new alleles for the trait of interest) to develop superior cultivars (Rani et al., 2016). Hybridisation of parents that are genetically different would lead to increased genetic variation in the population and ultimate high genetic gains in breeding programmes (Brown-Guedira et al., 2000).

There are many multi-variate techniques and approaches that are used to analyse morphological diversity in soybean and other crops. These include dissimilarity measures, principal components analysis, cluster analysis and canonical variables (Pawar, 2013).

Cluster and principal component analyses are the most commonly used techniques and are particularly useful for analysing yield contributing components. Kumar et al. (2015) used the two techniques based on morphological characters and grouped the 40 genotypes used in their study into two major clusters. They also found four principal components that were significant and contributed 76% to the total variation. The number of filled pods per plant, pod number per plant and days to maturity were the traits that positively contributed to PC1. Hundred seed weight, fertility percentage and yield per hectare positively contributed to PC2, while branches per plant and length of pods contributed more to PC3 and PC4, respectively.

2.12 Summary

Soybean is the global most important source of protein and vegetable oil. Seed yield is a polygenic character that generally has low heritability as it is greatly influenced by the environment. Since improving soybean yield still remains the main objective of soybean breeding in SSA, knowledge of heritability, correlations between seed yield and its related traits, and the most yield contributing traits is fundamental in the development of an effective selection strategy for seed yield. This is because most breeding programmes in SSA still rely on phenotypic selection to improve yield. Knowledge of GEI is indispensable to soybean breeders in selecting desirable cultivars. Testing newly developed elite lines in multiple years and locations is key in identifying and selecting high-yielding soybean varieties that are widely or specifically adapted in SSA.

The review of the literature identified the following gaps:

- Yield of soybean is still low in SSA. One of the major reasons for the low productivity is limited availability of high-yielding and stable cultivars that can be grown across varied environments in SSA. Mostly stability does not receive as much attention as specific adaptation in SSA. Mostly the available cultivars are developed for specific regions or countries.
- There is no clear selection criteria for yield that has been developed and can be employed by all breeding programmes.
- There is narrow genetic base in soybean. Limited genetic improvement effort in tropical soybean could be one of the reasons for lower genetic gains in SSA than in the Americas. This is the reason why diversity studies and introductions are key in increasing genetic variability and yield of tropical soybean.

Therefore, this study sought to address some of the identified gaps and contribute to breeding of high-yielding soybean cultivars in SSA.

2.13 References

Adie, M.M., A. Krisnawati and A. Gatut-Wahyu. 2014. Assessment of genotype x environment interactions for black soybean yield using AMMI and GGE biplot. International Journal of Agriculture Innovations and Research 2: 673-678.

Aditya, J., P. Bhartiya and A. Bhartiya. 2011. Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). Journal of Central European Agriculture 12: 27-34.

Alberts, M.J. 2004. A comparison of statistical methods to describe genotype x environment interaction and yield stability in multi-location maize trials. MSc Thesis, Department of Plant Sciences (Plant Breeding), Faculty of Agriculture, University of the Free State, Bloemfontein, South Africa.

Alene, A., T. Abdoulaye, J. Rusike, V. Manyong and T. Walker. 2015. The effectiveness of crop improvement programmes from the perspectives of varietal output and adoption: cassava, cowpea, soybean and yam in Sub-Saharan Africa and maize in West and Central Africa. Crop Improvement, Adoption and Impact of Improved Varieties in Food Crops in Sub-Saharan Africa: 74-122.

Ali, M.A., N.N. Nawab, A. Abbas, M. Zulkiffal and M. Sajjad. 2009. Evaluation of selection criteria in *Cicer arietinum* L. using correlation coefficients and path analysis. Australian Journal of Crop Science 3: 65-70.

Alwala, S., T. Kwolek, M. McPherson, J. Pellow and D. Meyer. 2010. A comprehensive comparison between Eberhart and Russell joint regression and GGE biplot analyses to identify stable and high yielding maize hybrids. Field Crops Research 119: 225-230.

Amira, J., D. Ojo, O. Ariyo, O. Oduwaye and M. Ayo-Vaughan. 2013. Relative discriminating powers of GGE and AMMI models in the selection of tropical soybean genotypes. African Crop Science Journal 21: 67-73.

Annicchiarico, P. 2002. Defining adaptation strategies and yield stability targets in breeding programmes. Quantitative Genetics, Genomics, and Plant Breeding: 365-383.

Annicchiarico, P., F. Bellah and T. Chiari. 2005. Defining subregions and estimating benefits for a specific-adaptation strategy by breeding programmes. Crop Science 45: 1741-1749.

Athoni, B.K. and G.T. Basavaraja. 2012. Association analysis, genetic variability and genetic diversity studies in soybean [*Glycine ma*x (L.) Merrill]. Asian Journal of Bio-Science 7: 114-118.

Atlin, G., R. Baker, K.B. McRae and X. Lu. 2000. Selection response in sub-divided target regions. Crop Science 40: 7-13.

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

Atnaf, M., K. Tesfaye, K. Dagne and D. Wegary. 2017. Genotype by trait biplot analysis to study associations and profiles of Ethiopian white lupin (*Lupinus albus*) landraces. Australian Journal of Crop Science 11: 55-62.

Becker, H. 1981. Correlations among some statistical measures of phenotypic stability. Euphytica 30: 835-840.

Becker, H. and J. Leon. 1988. Stability analysis in plant breeding. Plant breeding 101: 1-23.

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

Bernard, R.L. 1972. Two genes affecting stem termination in soybeans1. Crop Science 12: 235-239.

Bhandarkar, S. 1996. Genetic variability analysis of soybean varieties under different seasons. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur (M.P), India.

Bhartiya, A. and J. Aditya. 2016. Genetic variability, character association and path analysis for yield and component traits in black seeded soybean lines under rain fed condition of Uttarakhand hills Of India. Legume Research 39: 31-34.

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

Borém, A. 1998. Improvement of plants, Editora UFV, Federal University of Vicosa, Brazil.

Brink, M. and G. Belay. 2006. Plant resources of tropical Africa 1. Cereals and pulses. Prota Foundation; Backhuys Publishers: Leiden CTA: Wageningen, Netherlands: 76-82.

Brown-Guedira, G., J. Thompson, R. Nelson and M. Warburton. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Science 40: 815-823.

Bull, J., M. Cooper, I. Delacy, K. Basford and D. Woodruff. 1992. Utility of repeated checks for hierarchical classification of data from plant breeding trials. Field Crops Research 30: 79-95.

Chet, R., K. Sanjay, Y.C. Yadav and K. Adesh. 2010. Correlation coefficient and path analysis in garden pea (*Pisum sativum* var Hortens). Progressive Agriculture 10: 119-123.

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

Crossa, J., P. Fox, W. Pfeiffer, S. Rajaram and H. Gauch. 1991. AMMI adjustment for statistical analysis of an international wheat yield trial. Theoretical and Applied Genetics 81: 27-37.

Cruz, C., A. Regazzi and P. Carneiro. 2004. Biometric models applied to genetic improvement. Imprensa Universitária, Viçosa, Brazil.

Eberhart, S.t. and W. Russell. 1966. Stability parameters for comparing varieties 1. Crop Science 6: 36-40.

Farshadfar, E., M. Rashidi, M.M. Jowkar and H. Zali. 2013. GGE biplot analysis of genotype x environment interaction in chickpea genotypes. European Journal of Experimental Biology 3: 417-423.

Fernandez, G.C. 1991. Analysis of genotype x environment interaction by stability estimates. Horticultural Science 26: 947-950.

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

Gauch, H.G. 2006. Statistical analysis of yield trials by AMMI and GGE. Crop Science 46: 1488-1500.

Gauch, H.G. and R.W. Zobel. 1988. Predictive and post-dictive success of statistical analyses of yield trials. Theoretical and Applied Genetics 76: 1-10.

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

Gurmu, F. 2017. Stability analysis of fresh root yield of sweet potato in Southern Ethiopia using GGE bi-plot. International Journal of Pure Agricultural Advances 1: 1-9.

Gurmu, F., H. Mohammed and G. Alemaw. 2009. Genotype x environment interactions and stability of soybean for grain yield and nutrition quality. African Crop Science Journal 17: 87-99.

Haldane, J.B. 1946. The interaction of nature and nurture. Annals of Eugenics 13: 197-205.

Hermann, F.J. 1962. A revision of the genus *Glycine* and its immediate allies. United State Department of Agriculture.

Horn, L., H. Shimelis, F. Sarsu, L. Mwadzingeni and M.D. Laing. 2018. Genotype-by-environment interaction for grain yield among novel cowpea (*Vigna unguiculata* l.) selections derived by gamma irradiation. The Crop Journal 6: 306-313.

Hymowitz, T. 1990. Soybeans: The success story. Advances in New Crops: 159-163.

Hymowitz, T. and C. Newell. 1981. Taxonomy of the genus *Glycine*, domestication and uses of soybeans. Economic Botany 35: 272-288.

Hymowitz, T., R. Palmer and H. Hadley. 1972. Seed weight, protein, oil and fatty acid relationships within the genus *Glycine*. Agris: International Information System for the Agricultural Science and Technology: 245-250.

Jain, R.K., A. Joshi, H.R. Chaudhary, A. Dashora and C.L. Khatik. 2017. Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Legume Research- An International Journal 41: 532-536.

Jain, S., S.C. Srivastava, S.K. Singh, Y.M. Indapurkar and B.K. Singh. 2015. Studies on genetic variability, character association and path analysis for yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Research: An International Journal 38: 182-184.

Johnson, H.W., H. Robinson and R.E. Comstock. 1955. Genotypic and phenotypic correlations in soybeans and their implications in selection 1. Agronomy Journal 47: 477-483.

Kanchana, P., M.L. Santha and K.D. Raja. 2015. A review on *Glycine max* (L.) Merr. (soybean). World Journal of Pharmacy and Pharmaceutical Sciences 5: 356-371.

Kaplan, M., K. Kokten and M. Akcura. 2017. Assessment of genotype x trait x environment interactions of silage maize genotypes through GGE biplot. Chilean Journal of Agricultural Research 77: 212-217.

Karimizadeh, R., M. Mohammadi, M.K. Shefazadeh, A.A. Mahmoodi, B. Rostami and F. Karimpour. 2012. Relationship among and repeatability of ten stability indices for grain yield of food lentil genotypes in Iran. Turkish Journal of Field Crops 17: 51-61.

Kaya, Y., M. Akçura and S. Taner. 2006. GGE-biplot analysis of multi-environment yield trials in bread wheat. Turkish Journal of Agriculture and Forestry 30: 325-337.

Kaya, Y. and E. Ozer. 2014. Parametric stability analyses of multi-environment yield trials in triticale (*Xtriticosecale wittmack*). Genetika 46: 705-718.

Kearsey, M.J. and H.S. Pooni. 1998. The genetical analysis of quantitative traits. Stanley Thornes (Publishers) Ltd.

Kerby, T., J. Burgess, M. Bates, D. Albers and K. Lege. 2000. Partitioning variety and environment contribution to variation in yield, plant growth, and fiber quality. National Cotton Council, Memphis. p. 528-532.

Kuldeep S. C, K.S.Baig., D. H. Sarang, PH. Kiihne Dumai, P. U. Dhone and AniL Kumar. 2015. Association analysis for yield contributing and quality parameters in soybean. An International Quarterly Journal of Environmental Sciences 7: 113-118.

Kumar, A., S. Kumar, C. Kapoor, R. Bhagawati, A. Pandey and A. Pattanayak. 2014. GGE biplot analysis of genotype x environment interaction in soybean grown in Neh regions of India. Environment and Ecology 32: 1047-1050.

Kumar, A., A. Pandey, C. Aochen and A. Pattanayak. 2015. Evaluation of genetic diversity and interrelationships of agro-morphological characters in soybean (*Glycine max* L.) genotypes. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences 85: 397-405.

Lersten, N.R. and J.B. Carlson. 2004. Vegetative morphology 1. Soybeans: improvement, production, and uses. American Society of Agronomy 3: 15-57.

Lewis, R.S. 2006. Identification of germplasm of possible value for confronting an unfavorable inverse genetic correlation in tobacco. Crop Science 46: 1764-1771.

Lin, C.-S., M.R. Binns and L.P. Lefkovitch. 1986. Stability analysis: Where do we stand? Crop Science 26: 894-900.

Machikowa, T. and P. Laosuwan. 2011. Path coefficient analysis for yield of early maturing soybean. Songklanakarin Journal of Science & Technology 33: 365-368.

Mahawar, R., L. Pawar, N. Koli, H. Chaudhary, D. Meena and M. Ali. 2013. Assessment of genetic variability, correlation and path analysis of quantitative traits in soybean [*Glycine max* (L.) Merrill]. Soybean Research 11: 70-78.

Maleki, H.H., G. Karimzadeh, R. Darvishzadeh and A. Sarrafi. 2011. Correlation and sequential path analysis of some agronomic traits in tobacco (*Nicotiana tabacum* L.) to improve dry leaf yield. Australian Journal of Crop Science 5: 1644-1648.

Malik, M.F.A., M. Ashraf, A.S. Qureshi and A. Ghafoor. 2007. Assessment of genetic variability, correlation and path analyses for yield and its components in soybean. Pakistan Journal of Botany 39: 405-413.

Miladinovic, J., H. Kurosaki, J.W. Burton, M. Hrustic and D. Miladinovic. 2006. The adaptability of short-season soybean genotypes to varying longitudinal regions. European Journal of Agronomy 25: 243-249.

Mohamedkheir K, Seifeldin I, Sapey E and H. Tianfu. 2018. History, current status, and prospects of soybean production and research in sub-Saharan Africa. The Crop Journal 3: 226-235.

Mpepereki, S., F. Javaheri, P. Davis and K. Giller. 2000. Soyabeans and sustainable agriculture: promiscuous soyabeans in southern Africa. Field Crops Research 65: 137-149.

Mulridharan, P. 2017. Genetic analysis of traits determining seed yield over environments in soybean. PhD Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

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

Mustapha, M. and H. Bakari. 2014. Statistical evaluation of genotype by environment interactions for grain yield in millet (*Pennisetum glaucum* (L.). International Journal of Engineering and Science 3: 7-16.

Nassir, A.L. and O.J. Ariyo. 2011. Genotype x environment interaction and yield-stability analyses of rice grown in tropical inland swamp. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 39: 220-225.

Ndhlela, T., L. Herselman, C. Magorokosho, P. Setimela, C. Mutimaamba and M. Labuschagne. 2014. Genotype x environment interaction of maize grain yield using AMMI biplots. Crop Science 54: 1992-1999.

Nwangburuka, C., O.B Kehinde and O. Denton. 2011. Genotype x environment interaction and seed yield stability in cultivated okra using the additive main effect and multiplicative interaction (AMMI) and genotype and genotype x environment interaction (GGE). Archives of Applied Science Research 3: 193-205.

Nwokolo, E. 1996. Soybean (*Glycine max* (L.) Merr.). Food and feed from legumes and oilseeds. Springer. p. 90-102.

Odewale, J., C. Agho, C. Ataga, E. Okolo, C. Ikuenobe, G. Odiowaya. 2014. Genotype by trait relations between yield and some morphological traits of coconut (*Cocos nucifera* L.) hybrid varieties based on GT biplot. International Journal of Plant and Soil Science 3: 270-280.

Oliveira, T.R.A.d., G.d.A. Gravina, G.H.F.d. Oliveira, K.C. Araújo, L.C.d. Araújo, R.F. Daher. 2018. The GT biplot analysis of green bean traits. Ciência Rural 48: 1-6.

Pawar, N. 2013. Characterization and categorization of major growth and yield traits in soybean [*Glycine max* (L.) Merrill. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

Pervin, M., M. Polash, S. Rahman and A. Deb. 2007. Study of genetic variability and GXE interaction of some quantitative traits in blackgram [*Vigna mungo* (L.) Hepper]. Journal of Biological Sciences 7: 169-175.

Purchase, J.L. 1997. Parametric analysis to describe genotype x environment interaction and yield stability in winter wheat. PhD Thesis, Department of Agronomy, Faculty of Agriculture of the University of the Free State, Bloemfontein, South Africa.

Rani, A., V. Kumar, B. Gill, V. Morya and R. Rawal. 2016. Assessment of genetic diversity of soybean genotypes differing in resistance against yellow mosaic virus using simple sequence repeat markers. Legume Research 39: 674-681.

Rao, M., B. Mullinix, M. Rangappa, E. Cebert, A. Bhagsari, V. Sapra. 2002. Genotype x environment interactions and yield stability of food-grade soybean genotypes. Agronomy Journal 94: 72-80.

Rao, V.T. 2016. Genetic variability, correlation and path coefficient analysis under drought in groundnut (*Arachis hypogaea* L.). Legume Research 39: 319-322.

Sabaghnia, N., M. Mohammadi and R. Karimizadeh. 2013. Yield stability of performance in multi-environment trials of barley (*Hordeum vulgare* L.) genotypes. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis 61: 787-793.

Sabaghnia, N., S. Sabaghpour and H. Dehghani. 2008. The use of an AMMI model and its parameters to analyse yield stability in multi-environment trials. The Journal of Agricultural Science 146: 571-581.

Sadie, J. 2015. South African variety list as maintained by the registrar of plant improvement. Department of Agriculture, Fisheries and Forestry, Republic of South Africa.

Sharifi, P., H. Aminpanah, R. Erfani, A. Mohaddesi and A. Abbasian. 2017. Evaluation of genotype x environment interaction in rice based on AMMI model in Iran. Rice Science 24: 173-180.

Sharifi, P. and A.A. Ebadi. 2018. Relationships of rice yield and quality based on genotype by trait (GT) biplot. Anais da Academia Brasileira de Ciencias 90: 343-356.

Sharma, B.B. and V.K. Sharma. 2012. Character association and path analysis studies for yield and horticultural traits in garden pea. Environment and Ecology 30: 1591-1598.

Shukla, G. 1972. Some statistical aspects of partitioning genotype environmental components of variability. Heredity 29: 237-245.

Shurtleff, W. and A. Aoyagi. 2009. History of soybeans and soyfoods in Africa (1857-2009): Extensively Annotated Bibliography and Sourcebook. Soyinfo Center.

Silvey, V. 1981. The contribution of new wheat, barley and oat varieties to increasing yield in England and Wales 1947-78. National Institute of Agricultural Botany 15: 399-412.

Simion, T., W. Mohammed and B. Amsalu. 2018. Genotype by environment interaction and stability analysis of cowpea [*Vigna unguiculata* (L.) Walp] genotypes for yield in Ethiopia. Journal of Plant Breeding and Crop Science 10: 249-257.

Singh, R. 2017. Botany and cytogenetics of soybean. The soybean genome. Springer. p. 11-40.

Singh, R.J. and T. Hymowitz. 1999. Soybean genetic resources and crop improvement. Genome 42: 605-616.

Singh, R.J., R.L. Nelson and G. Chung. 2007. Soybean (*Glycine max* (L.) Merr.). Genetic resources, chromosome engineering, and crop improvement. Oilseed crops 4: 13-50.

Sleper, D.A. and J.M. Poehlman. 2006. Breeding field crops. Blackwell publishing, USA.

Smith, S., A. Al-Doss and M. Warburton. 1991. Morphological and agronomic variation in North African and Arabian alfalfas. Crop Science 31: 1159-1163.

Sofia, S. 2016. Studies on genotype by trait biplot analysis and genetic diversity for yield and yield related traits in mungbean (*Vigna radiata* (L.) Wilczek). MSc Thesis, Department of Genetics and Plant Breeding, Acharya N.G Ranga Agricultural University, Guntur, India.

Tadege, M.B., H.Z. Utta and A.A. Aga. 2014. Association of statistical methods used to explore genotype x environment interaction (GEI) and cultivar stability. African Journal of Agricultural Research 9: 2231-2237.

Tefera, H. 2011. Breeding for promiscuous soybean at international institute of tropical agriculture. Soybean-molecular aspects of breeding. Institute for New Technologies 7: 147-162.

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

Temesgen, T., G. Keneni, T. Sefera and M. Jarso. 2015. 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.

Ukalski, K. and M. Klisz. 2016. Application of GGE biplot graphs in multi-environment trials on selection of forest trees. Folia Forestalia Polonica 58: 228-239.

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

Xu, Y. 2010. Molecular plant breeding. CABI, United Kingdom: P 381-414

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

Yan, W. and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Yan, W., M.S. Kang, B. Ma, S. Woods and P.L. Cornelius. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Science 47: 643-653.

Yan, W. and I. Rajcan. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Science 42: 11-20.

Yan, W. and N.A. Tinker. 2006. Biplot analysis of multi-environment trial data: Principles and applications. Canadian Journal of Plant Science 86: 623-645.

Zhang, P.-P., H. Song, X.-W. Ke, X. Jin, L. Yin, Y. Liu. 2016. GGE biplot analysis of yield stability and test location representativeness in proso millet (*Panicum miliaceum* L.) genotypes. Journal of Integrative Agriculture 15: 1218-1227.

Zobel, R.W., M.J. Wright and H.G. Gauch. 1988. Statistical analysis of a yield trial. Agronomy Journal 80: 388-393.

CHAPTER 3

GENOTYPE BY ENVIRONMENT INTERACTION AND STABILITY OF ELITE SOYBEAN [*Glycine max* (L.) Merrill.] LINES ACROSS VARIED PRODUCTION ENVIRONMENTS IN SOUTHERN AFRICA

Abstract

Soybean [*Glycine max* (L.) Merrill] is the global most important source of protein and oil. Developing stable cultivars with high potential for seed yield is the main objective of many soybean breeding programmes. The aim of this study was to evaluate the stability of 25 elite soybean lines for seed yield and the magnitude of their interaction with the environment in southern Africa using the GGE biplot and AMMI models. Twenty-five genotypes (five checks and 20 experimental lines) were evaluated in six locations during the 2017/18 rainy season in a 5x5 alpha lattice design, replicated three times at each location. The locations were: IITA-SARA, Lusaka West and Chipata in Zambia; Chitedze in Malawi; Nampula in Mozambique; and Rattray Arnold Research Station in Zimbabwe. Seed yield was recorded on per plot basis for each genotype and the data analysed using SAS and Genstat. The environment, genotype and genotype by environment interaction (GEI) effects were highly significant ($p<0.001$), with contribution of 21%, 32% and 47%, respectively to the total variation. The first two interaction principal component axes (IPCA1 and IPCA2) explained 44% and 22%, respectively of the variation due to GEI. The first two principal components, PC1 and PC2 in the GGE biplot model accounted for 52% and 18% of the variation, respectively. Twelve genotypes (48%) yielded above the grand mean and the remaining 52% had yields below the average yield. AMMI analysis showed that the lines TGx2002-17DM, TGx2001-10DM, TGx2001-18DM, TGx2014-24FM, TGx2001-6FM and TGx2002-3DM were specifically adapted to Chitedze, Nampula, IITA-SARAH, Lusaka West and Chipata, respectively. Both AMMI and GGE biplot analyses indicated that Lusaka West was the highest yielding and most informative environment. However, Rattray Arnold Research Station was identified as the most ideal environment for selecting widely adapted lines as it was both informative and highly representative of the environments. Both analyses showed that the line TGx2014-5GM was more stable than the checks and second to the highest yielding check, with seed yield of 4143 kg/ha. This line showed great potential and could be recommended for release as a cultivar in the four southern African countries.

Key words: AMMI, GGE biplot, genotype by environment interaction, Soybean, stability

3.1 Introduction

Soybean [*Glycine max* (L.) Merrill], is the global number one source of high quality, inexpensive protein and vegetable oil. The crop has extensively been used as human food, animal feed and raw material for manufacturing of various industrial products (Sinclair et al., 2014). It is a good crop for rotation with cereals as it improves soil fertility and breaks the build-up of pests and diseases (Athoni and Basavaraja, 2012). Soybean production has increased in the recent years in southern Africa, owing to the increased demand for vegetable oil, protein and soybean cake, although productivity is still low (1.1 ton/ha) (Mohamedkheir et al., 2018). The crop follows maize and wheat in terms of production area in the southern African region, which contributes over 50% of Africa's total soybean production (FAO, 2017). South Africa (850,000 tonnes) is the leading producer in the region, followed by Zambia (302,720 tonnes), Malawi (190,000 tonnes), Zimbabwe (70,000 tonnes) and Mozambique (32,000 tonnes) (Meyer et al., 2018; MOA and CSO, 2018).

The phenotype of a plant is determined by the genotype (G), the production environment (E) and the interaction between the environment and genotype (Yan and Tinker, 2006). The genetic make-up of an individual generally does not change regardless of the environment it is exposed to. Therefore, any variation in phenotypic expression of a specific genotype is caused by the environment (Yan and Kang, 2003). This variation in the response of a genotype in different environments is known as genotype by environment interaction (GEI) (Sharifi et al., 2017). The interaction between the environment and genotype is only important if it is significant and causes genotypes to rank differently (crossover GEI) or if the genotypes magnitude of differences change in varied environments (non-crossover GEI) (Fernandez, 1991). The crossover type is important for selection of specifically adapted genotypes, while the non-crossover type is important when selecting widely adapted genotypes (Kaya et al., 2006).

Environments in the four southern African countries (Zambia, Malawi, Mozambique and Zimbabwe) where the genotypes were tested are highly varied in terms of latitude, altitude and rainfall. Variability in these environmental conditions, across years and locations, could cause differential response of new elite soybean lines in the region (Bull et al., 1992; Mushoriwa, 2013). Definition of breeding objectives also becomes complicated in the presence of significant GEI. GEI lowers heritability, thereby leading to reduced progress from selection of the desired traits and genotypes (Yan and Kang, 2003). It also makes it difficult to compare genotypes among themselves and their relationship with the test environment (Xu, 2010). Therefore, multi-environment trials are required to understand the intricate GEI so that specifically adapted and stable or widely adapted genotypes can be identified and

recommended to the target environments (Gurmu et al., 2009). A stable genotype is one that has the capacity to make use of the available resources and performs above the grand mean in all environments (Eberhart and Russell, 1966; Gurmu et al., 2009).

Although methods such as PCA, regression and ANOVA measure GEI, they cannot detect the significant GEI components (Zobel et al., 1988). It is for this reason that multivariate methods, namely additive main effects and multiplicative interaction (AMMI) and genotype plus genotype by environment interaction (GGE) models are gaining popularity in the analysis of GEI and stability of genotypes in variable environments (Amira et al., 2013). The two methods combine univariate methods for the environment and genotype additive effects with a multivariate method for the multiplicative effect of GEI (Zobel et al., 1988; Bhartiya et al., 2017), thereby providing a better interpretation of multi-environmental data set (Bhartiya et al., 2017). The AMMI model is effective in assessing the adaptability and stability of genotypes (Pacheco et al., 2005). Accuracy in the AMMI model is achieved by separation of structural variation from noise (Nassir and Ariyo, 2011). The AMMI's stability value (ASV), is helpful in identifying stable genotypes across environments (Purchase, 1997). Lower AMMI stability values are an indication of great stability of genotypes (Anley et al., 2013). GGE biplot makes it possible to clearly visualise the complex GEI in a graph (Yan et al., 2000). The GGE biplot is effective in the evaluation of stability and mean yield of genotypes, test-environments assessment (their representative and discriminating capacity) and mega-environment analysis (the "which won where" pattern) (Yan and Kang, 2003).

Multi-location yield trials are usually conducted in the final stages of cultivar development. The 20 elite lines used in the current research were advanced from preliminary yield trials and testing them in multi-locations was required to identify candidates for release as cultivars. Therefore, the objective of this study was to determine the magnitude of GEI and stability of 20 elite soybean lines and five commercial cultivars (checks) for seed yield in southern Africa using the AMMI and GGE biplot models.

3.2 Materials and methods

3.2.1 Lines and cultivars used in the study

Twenty elite lines developed by the International Institute of Tropical Agriculture, and five commercial checks were used in the study. The five checks included Kafue (early maturing), SC Safari (medium maturing), SC Squire (medium maturing), MRI Dina (late maturing) and Lukanga (medium maturing). The information about the lines and checks used is indicated in Table 3.1.

Table 3.1 Experimental lines and checks tested in the four countries

Genotype code	Genotype name	Source	Maturity	Growth habit
G1	TGx2001-11DM	IITA	M	I
G2	TGx2014-21FM	IITA	M	I
G3	TGx2002-7FM	IITA	M	I
G4	TGx2014-5GM	IITA	M	I
G5	TGx2002-14DM	IITA	E	I
G6	TGx2001-24DM	IITA	M	I
G7	TGx2001-6FM	IITA	M	I
G8	TGx2001-13DM	IITA	M	I
G9	TGx2002-23DM	IITA	M	I
G10	TGx2014-19FM	IITA	M	I
G11	TGx2001-8DM	IITA	M	I
G12	TGx2001-1DM	IITA	M	I
G13	TGx2001-10DM	IITA	L	I
G14	TGx2002-5FM	IITA	M	I
G15	TGx2014-16FM	IITA	M	I
G16	TGx2014-24FM	IITA	M	I
G17	TGx2001-18DM	IITA	M	I
G18	TGx1987-62F	IITA	M	I
G19	TGx2002-3DM	IITA	M	I
G20	TGx2002-17DM	IITA	M	I
Checks				
CH1	Kafue	IITA	E	D
CH2	Lukanga	Zamseed	M	D
CH3	MRI Dina	MRI Syngenta	L	I
CH4	SC SAFARI	Seed Co	M	I
CH5	SC SQUIRE	Seed Co	M	I

Key: L= late maturity, M= medium maturity, D= determinate growth habit, I= indeterminate growth habit, IITA= International Institute of Tropical Agriculture.

3.2.2 Description of the six sites

The study was conducted in Zambia, Malawi, Mozambique and Zimbabwe in the 2017/18 rainy season. Three sites were used in Zambia and only one site in each of the other

countries, namely: Malawi, Mozambique and Zimbabwe. The sites are described in detail in Table 3.2.

Table 3.2 Information about the six sites in the four countries

Code	Environment name	Country	longitude	Latitude	Elevation (masl)	Rainfall (mm)
E1	IITA-SARAH	Zambia	E28º30′	S15º30′	1193	703
E2	Lusaka West	Zambia	E28º33′	S15º67′	1301	826
E3	Chipata	Zambia	E32º39′	S13º40′	1098	1249
E4	RARS	Zimbabwe	E31º14′	S17º40′	1341	880
E5	Chitedze	Malawi	E33º38′	S13º59′	1100	929
E6	Nampula	Mozambique	E39º19′	S15º16′	366	-

RARS = Rattray Arnold Research Station, IITA=International Institute of Tropical Agriculture, SARAH=Southern African Region Administration Hub, masl=metres above sea level, mm=millimetres.

3.2.3 Trial design and management

The 25 genotypes were evaluated in a 5 x 5 alpha lattice design, replicated three times per environment. A plot consisted of four rows that were 0.5 m apart and 5 m long. The intra-row spacing was 0.05 m, which gave a plot size of 5 m^2 and a target population of about 350,000 plants per hectare. Basal dressing fertilizer (25 kg N/ha, 30 kg K_2O/ha, 60 kg P_2O_5/ha) was applied at planting and pre-emergence herbicides (Metolachlor and Imazethapyr) were applied soon after planting to prevent weeds from germinating. Weeds that germinated later in the season were both mechanically (hand weeding) and chemically (Quizalofop-p-ethyl and Fomesafen) controlled. When the crop had fully matured, the net plots (two middle rows) in each replication were harvested and weight of the seed in kg was recorded for each plot. The weight was converted to yield in kg per hectare after being corrected to 11% moisture content (Mushoriwa, 2013).

3.2.4 Statistical analysis

3.2.4.1 Analysis of variance

The grain yield data collected at each site were subjected to analysis of variance (ANOVA) followed by combined analysis of variance for all the six sites using PROC GLM in SAS 9.4 software (SAS, 2013). The means were separated using Tukey test. The replications within location were considered as random effects, whereas the genotypes were taken to be fixed

effects so that the genotype, environment and GEI effects could be determined to be significant or not. The combined ANOVA model used is given as follows:

$$Y_{ijkl} = \mu + G_i + E_j + R_{k(j)} + B_{l(jk)} + GE_{ij} + \varepsilon_{ijlk} \quad \text{Equation 3.1}$$

Where Y_{ijkl} is the response of the i^{th} genotype in j^{th} environment and k^{th} replication within environment and l^{th} block within replication; μ is the grand mean, G_i is the genotype effect i; E_j is the environment effect j; $R_{k(j)}$ is the replication within environment effect k; $B_{l(jk)}$ is the block within replication effect l; GE_{ij} is the genotype × environment interaction effect; and ε_{ijkl} is the random error.

3.2.4.2 Additive main effects and multiplicative interaction (AMMI) analysis and AMMI's stability value (ASV)

The AMMI analysis was carried out in Genstat version 18.2 (VSNi, 2016). The AMMI model combines both ANOVA and PCA in assessing the stability and adaptability of genotypes. The genotype and environment main effects are taken to be additive using ANOVA, while the GEI is taken to have a multiplicative effect by PCA. The model used to determine the nature of GEI was adopted from Zobel et al. (1988) and the biplot was constructed using IPCA1 and IPCA2 scores. The model is given as follows:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_n \lambda_n \delta_{in} \gamma_{jn} + P_{ij} + \varepsilon_{ij} \quad \text{Equation 3.2}$$

Where Y_{ij} is the mean yield of the i^{th} genotype effect in j^{th} environment in all replications; and the additive components are μ (the grand mean), α_i (the i^{th} genotype effect) and β_j (the j^{th} environment effect). The multiplicative component consists of λ_n, δ_{in}, γ_{jn} *and* P_{ij} terms, where λ_n is the interaction principal component, δ_{in} is the eigen vector for the genotypic principal component, γ_{in} is the environmental principal component, P_{ij} are the AMMI residuals and ε_{ij} is the random error.

The AMMI stability values calculated using the formula proposed by Purchase (1997) were used to rank the 25 genotypes according to their stability. The formula is given as follows:

$$\text{ASV} = \sqrt{\left[\frac{SSIPCA1}{SSIPCA2}(IPCA1\ Score)\right]^2 + (IPCA2\ score)^2} \quad \text{Equation 3.3}$$

3.2.4.3 Genotype and genotype by environment interaction (GGE) biplot model

GGE biplot consists of two concepts, i.e., the GGE and biplot concepts. The GEI was further partitioned and analysed using the GGE model (Yan and Kang, 2003) in Genstat version

18.2 (VSNi, 2016). The GGE biplot was constructed using PC1 and PC2. The model used is based on singular value decomposition of PC1 and PC2 and is written as follows:

$$Y_{ij} - \mu - \beta_j = \lambda_1\gamma_{i1}\delta_{j1} + \lambda_2\gamma_{i2}\delta_{j2} + \varepsilon_{ij} \quad \text{Equation 3.4}$$

Where Y_{ij} is the mean of i^{th} genotype in the j^{th} environment, μ is the grand mean, β_j is environment main effect in the j^{th} environment and $\mu+\beta_j$ is the mean of all genotypes in j^{th} environment. The terms λ_1 and λ_2 are the singular values for PC1 and PC2, respectively; γ_{i1} and γ_{i2} are eigenvectors of the i^{th} genotype for PC1 and PC2, respectively. The components δ_{j1} and δ_{j2} are eigenvectors of the j^{th} environment for PC1 and PC2, respectively. ε_{ij} is the residual associated with the i^{th} genotype in the j^{th} environment.

For a meaningful biplot to be constructed, the partitioning of the singular values into the environment and genotype eigenvectors is cardinal. Therefore, the above model can be rewritten as follows:

$$Y_{ij} - \mu - \beta_j = g_{i1}\, e_{1j} + g_{i2}\, e_{2j} + \varepsilon_{ij} \quad \text{Equation 3.5}$$

Where g_{i1} and e_{1j} are PC1 scores for the i^{th} genotype and j^{th} environment, respectively, g_{i2} and e_{i2} are the PC2 scores for the i^{th} genotype and j^{th} environment, respectively. The other terms are as described in equation 3.4.

The results from the GGE biplot analysis are useful in displaying the relationships among genotypes and environments so that stable genotypes can be identified. The winning genotypes in each environment can easily be visualised and the ideal genotypes can be identified based on both their mean and stability.

3.3 Results

3.3.1 Combined analysis of variance

The combined analysis of variance (Table 3.3) shows that the genotype, environment and genotype by environment interaction (GEI) effects were highly significant ($P<0.001$). The environment main effect contributed 19.9% to the total sum of squares and the contributions of genotype and genotype by environment interaction effects were 24.5% and 35.6%, respectively. The grand mean and coefficient of variation (CV) were 3146.3 kg/ha and 7.46%, respectively.

Table 3.3 Combined analyis of variance for grain yield of 25 genotypes across the six sites

Source	DF	SS	MS
ENV	5	73316846.5	14663369.3***
REP(ENV)	12	2747445.5	228953.8***
BLK(ENV*REP)	72	5165481.2	71742.8ns
GEN	24	90320032.3	3763334.7***
ENV*GEN	120	131267734.1	1093897.8***
Error	216	11890128.3	55046.9
Total	449	368215531.1	
Mean yield	3146.31 kg/ha		
CV%	7.46		

MS = Mean square, SS = Sum of squares, DF = Degrees of freedom, ***Significant at P<0.001, ns = not significant, CV = Coefficient of variation.

3.3.2 Additive main effect and multiplicative interaction (AMMI) analysis

3.3.2.1 AMMI ANOVA for grain yield

The AMMI ANOVA (Table 3.4) revealed that the G, E and GEI effects were highly significant (P<0.001). The GEI effect accounted for 47.36% of the total variation. The contributions of the G and E effects to the total variation were 31.59% and 21.04%, respectively. The GEI was further partitioned into five IPCAs, which were all highly significant (P<0.001). IPCA1 contributed 44% and 23% to the GEI sum of squares and degrees of freedom (df), respectively. IPCA1 and IPCA2 together contributed 66% and 45% to the GEI sum of squares and degrees of freedom, respectively. The first three IPCAs explained about 81% of GEI variation and accounted for 65% of the GEI degrees of freedom.

Table 3.4 AMMI ANOVA for grain yield across the six sites

Source of variation	DF	SS	MS	Total variation %	GE explained %	GE Cumulative %
Treatments	149	348412476	2338339***			
Block (Env)	12	2747446	228954***			
Genotypes, G	24	110071631	4586318***	31.59		
Environments, E	5	73316846	14663369***	21.04		
GE Interactions	120	165023998	1375200***	47.36		
IPCA 1	28	73248019	2616001***		44.39	44.39
IPCA 2	26	36080975	1387730***		21.86	66.25
IPCA 3	24	24835867	1034828***		15.05	81.30
IPCA 4	22	16929608	769528***		10.26	91.56
IPCA 5	20	13929530	696476***		8.44	100.00
Error	288	17055610	59221			
Total	449	368215531	820079			

GE = Genotype by Environment interaction, IPCA = Interaction principal component axis, *** Significant at P<0.001, MS = Mean squares, SS = Sum of squares, DF = Degrees of freedom

3.3.2.2 IPCA scores, AMMI stability values and mean yields for environments and genotypes

Grain yield ranged from 2444 to 4251 kg/ha (Table 3.5). The check CH4 and line G4 were the highest yielding genotypes, with yields of 4251 kg/ha and 4143 kg/ha, respectively. The lines G20 and G2 recorded the lowest yields of 2492 kg/ha and 2445 kg/ha, respectively. Twelve genotypes (CH1, G6, CH3, G16, G7, G14, CH2, G1, CH5, G9, G4 and CH4) yielded above the grand mean (3146 kg/ha) and the remaining 13 were below the average yield. The environment E2 (Lusaka West) recorded the highest mean yield of 3250 kg/ha. Nampula (E6) was the least and recorded a mean grain yield of 2476 kg/ha. The IPCA scores for environments and genotypes were both negative and positive. The genotypes G7, G17 and G20 had larger positive or negative IPCA scores. The lowest IPCA scores were observed for genotypes G4, G14 and G11. The AMMI stability values for genotypes ranged from 1.3 for G14 to 61.4 for G20.

Table 3.5 Mean yields of 25 soybean genotypes across environments, IPCA scores and ASVs for genotypes

GEN CODE	E1	E2	E3	E4	E5	E6	Mean GYD	IPCAg 1	IPCAg 2	IPCAg 3	IPCAg 4	IPCAg 5	ASV
G1	3874.70	4251.63	2265.94	3345.75	3815.84	4241.82	3632.61	5.97	1.33	-14.86	-20.91	2.20	12.19
G2	1845.61	2698.46	2193.69	2193.67	3522.05	2214.48	2444.66	14.84	11.80	1.61	6.03	4.00	32.36
G3	2722.47	4289.67	3082.26	2504.79	2612.79	2064.00	2879.33	0.05	-7.37	16.60	3.59	0.94	7.37
G4	4078.05	5107.53	3659.42	4348.43	3796.52	3867.77	4142.95	0.12	-2.97	7.87	-9.24	-10.05	2.98
G5	3506.27	2431.22	3725.63	3123.46	2758.27	3139.57	3114.07	-11.58	24.54	2.60	-4.51	6.15	33.99
G6	2500.25	4662.54	1627.64	2405.79	3898.42	2048.43	2857.18	20.56	-16.17	-2.87	10.11	-4.99	44.76
G7	4748.02	3204.50	3288.17	3411.78	2691.59	2722.36	3344.40	-22.05	6.50	-7.66	0.72	-7.04	45.23
G8	2796.52	3767.63	1429.45	1554.80	3178.84	2817.78	2590.84	11.98	-11.72	-12.78	-7.88	15.07	26.99
G9	5291.80	4708.86	2882.64	3288.39	3432.74	3315.60	3820.00	-14.22	-12.44	-18.20	-1.84	0.12	31.43
G10	1990.23	2440.08	2690.01	2498.38	4094.60	2551.81	2710.85	14.52	24.15	-1.01	10.00	7.03	38.10
G11	3295.52	3755.09	3609.63	2280.02	3546.01	2965.36	3241.94	-0.40	5.76	6.48	10.66	19.71	5.81
G12	2056.75	4765.10	2483.56	2246.07	2886.39	2451.73	2814.93	13.54	-13.11	20.56	-1.58	2.02	30.46
G13	2399.79	3681.44	2257.79	3020.59	2467.68	3058.26	2814.26	6.29	2.78	9.05	-19.90	-7.36	13.08
G14	3311.87	4299.34	3455.50	2732.60	3239.96	3174.65	3368.99	0.47	0.89	9.70	-4.78	13.21	1.31
G15	4564.86	2812.15	2418.34	1913.22	2718.42	1444.65	2645.27	-21.10	0.75	-19.95	18.93	2.73	42.85
G16	3532.02	5055.68	2338.45	2208.06	4079.70	2671.69	3314.27	11.66	-22.14	-5.00	8.01	7.39	32.42
G17	4518.94	3992.86	2763.67	2571.74	1927.51	2856.53	3105.21	-22.43	-8.33	-5.60	-12.23	4.74	46.29
G18	2756.98	3250.38	1642.62	2119.49	2937.09	2475.86	2530.40	6.70	0.09	-7.52	-7.12	1.62	13.60

Table 3.5 continued

GEN CODE	E1	E2	E3	E4	E5	E6	Mean GYD	IPCAg 1	IPCAg 2	IPCAg 3	IPCAg 4	IPCAg 5	ASV
G19	2274.64	3178.74	2582.94	3051.89	3097.78	2098.15	2714.02	4.45	11.16	6.90	3.68	-7.86	14.35
G20	1732.13	3113.41	859.24	2693.94	4171.82	2383.41	2492.33	30.13	5.42	-13.22	0.64	-14.73	61.41
CH1	3228.87	3123.03	3148.72	3369.45	3929.37	2571.82	3228.54	2.35	17.05	-3.10	6.41	-5.50	17.70
CH2	4134.34	4596.80	3079.62	4068.69	3257.36	2186.79	3553.93	-9.48	-4.18	2.21	8.57	-24.20	19.70
CH3	3759.26	3649.61	3160.34	3400.87	2693.97	3168.27	3305.39	-10.15	2.96	2.77	-9.44	-4.69	20.82
CH4	4877.84	5656.33	4313.77	3758.90	3897.56	3000.80	4250.87	-9.48	-15.37	8.98	11.83	-3.25	24.64
CH5	4506.12	4420.10	4549.19	3336.51	2628.26	3003.16	3740.56	-22.74	-1.36	16.43	0.24	2.75	46.18
Mean	3372.15	3876.49	2780.33	2857.89	3251.22	2739.79	**3146.31**						
IPCAe1	-45.55	15.01	-24.36	-1.97	43.04	13.83							
IPCAe2	-13.63	-47.68	20.93	17.71	11.17	11.50							
IPCAe3	-31.36	19.21	33.31	4.56	-18.51	-7.20							
IPCAe4	6.08	-1.37	14.62	-9.31	26.50	-36.52							
IPCAe5	2.31	-2.61	15.48	-38.79	4.07	19.53							
ASV	93.47	56.58	53.70	18.15	88.08	30.35							

ASV= AMMI stability value, IPCAe= Interaction principal component axis scores for environments, IPCAg= Interaction principal component axis scores for genotypes, Mean GYD= Mean grain yield. E1= IITA-SARAH, E2= Lusaka West, E3= Chipata, E4= Rattray Arnold Research Station, E5= Chitedze and E6= Nampula.

3.3.2.3 First four AMMI selections in the six environments

The best four yielders in each environment are shown in Table 3.6. The genotype CH4 was the best as it appeared in the top four in environments E2, E1, E3 and E4. It was followed by G4, which appeared in the top four in environments E4, E2 and E6. Genotypes G20, G1, CH5 and G9 appeared in the top four only in one environment each, namely E5, E6, E3 and E1, respectively. The ranking of genotypes was different across environments.

Table 3.6 Best four genotypes in each of the six environments

Environment	Mean GYD	Score	1	2	3	4
E1	3372	-45.55	G9	CH4	G7	G15
E2	3876	15.01	CH4	G4	G16	G12
E3	2780	-24.36	CH5	CH4	G11	G5
E4	2858	-1.97	G4	CH2	CH4	G7
E5	3251	43.04	G20	G10	G16	G6
E6	2740	13.83	G1	G4	G14	G9

Mean GYD= Mean grain yield. E1= IITA-SARAH, E2= Lusaka West, E3= Chipata, E4= Rattray Arnold Research Station (RARS), E5= Chitedze and E6= Nampula

3.3.2.4 AMMI biplots

The first two principal components, IPCA1 and IPCA2, explained 66.25% of the total GEI variation (Figure 3.1). The length of the vector of an environment from the biplot origin is an indication of the environment's discriminative capacity and representativeness. The environments with longer vectors have higher discriminative ability, while those with shorter vectors are more representative. Figure 3.1 shows that E2 (Lusaka West) was the most discriminating environment. The environment E1 (IITA-SARAH) and E5 (Chitedze) also showed high discriminative capacity. The environment E6 (Nampula) was the most representative environment followed by E4 (RARS). The environments E6 and E5 were closely related as evidenced by the smallest angle between them. If the angle between vectors of two genotypes is smaller and the two genotype vectors are similar in length, it means that those genotypes have similar yield performance. Figure 3.1 shows that the genotypes G1 and G18 were close together and had similar yield performance. The closer the genotype is to the environment marker the more specifically adapted it is to that environment. The genotype G20 was the best yielder at E5, while G1 and G13 were adapted to E6. The genotypes G17 and CH5 performed well at E1, with G6 and G16 showing better adaptation to

E2. The genotypes G7 and CH3 showed specific adaptation to the environment E3. The genotypes G19 and CH1 were specifically adapted to E4.

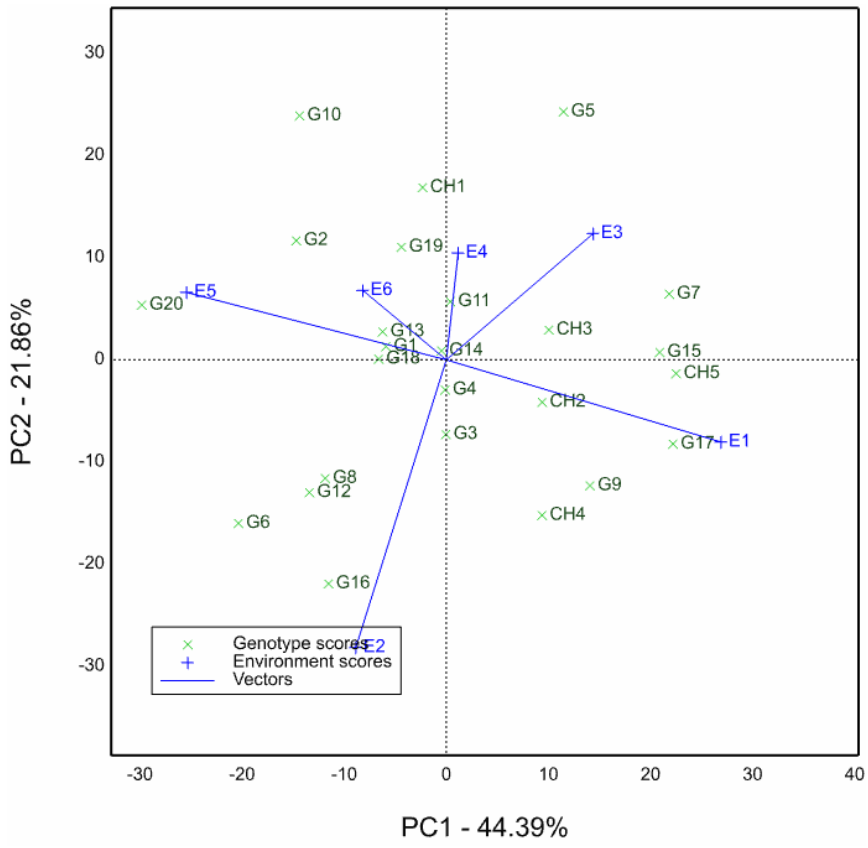

Figure 3.1 Biplot analysis of GEI based on AMMI2

Figure 3.2 shows means of genotypes and environments, and stability of genotypes. The environments and genotypes on the left side of the dotted line dividing the graph from bottom to top yielded below the grand mean, while those on the right side of this line yielded above the grand mean. Thus, environments E3, E4 and E6 yielded below average, while E1, E2 and E5 were above the grand mean. Out of the 25 genotypes, 12 (48%) genotypes yielded above the grand mean and 13 (52%) genotypes yielded below the grand mean. The highest yielding genotypes and environments were the ones on the further right (increasing mean yield direction) of the graph, while the poorest environments and genotypes lied on the furthest left side (decreasing mean yield direction) of the graph. Therefore, the best yielding genotypes in Figure 3.2 were CH14, G4 and G9. The genotypes G2, G20, G18 and G8 were the least across the six sites. The highest yielding environment was E2. E6 was the poorest yielding environment. The most stable genotypes were G4, G14, G3 and G11. The most unstable genotypes were G20, CH5, G7, G17 and G15.

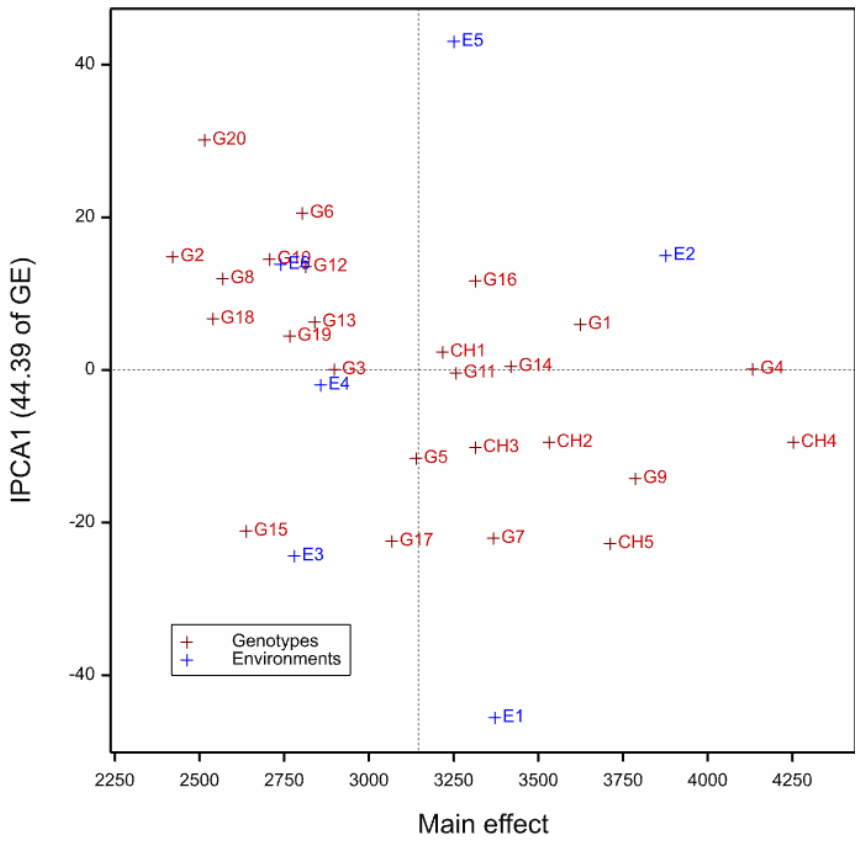

Figure 3.2 Biplot showing environment and genotype means plotted against IPCA1 scores

3.3.3 Genotype plus genotype by environment (GGE) analysis

The results showed that PC1 and PC2 together contributed 69.17% to the total variation.

3.3.3.1 Mega environment classification and the 'which won where' view

Figure 3.3 shows the winning genotypes in each mega-environment. It is important to visualise the 'which won where' pattern as it facilitates the identification of mega-environments that are possibly present in a region. The polygon view was created by connecting the furthest genotype markers from the origin of the biplot (most responsive genotypes). The rest of the genotypes fell inside the polygon (less responsive). The 10 rays (lines running perpendicularly to the sides of the polygon) divided the biplot into 10 sectors and the environments were in three of them. Three mega-environments were identified in the biplot: environments E6, E2, E1 and E4 formed the first mega-environment, while E3 and E5 constituted the second and third mega-environments, respectively. The vertex genotypes were either the poorest or best yielding in one or more of the test environments, while those inside the polygon (whose vectors were shorter) were less affected by the environments. The

genotype CH4 was the best performer in the first mega-environment. The best genotypes in the second and third mega-environments were CH5 and G6, respectively. The genotype G5, G7, G15, G16 and G20 performed the poorest in all environments. The genotypes G3 and G11 were closest to the biplot origin and were the least responsive.

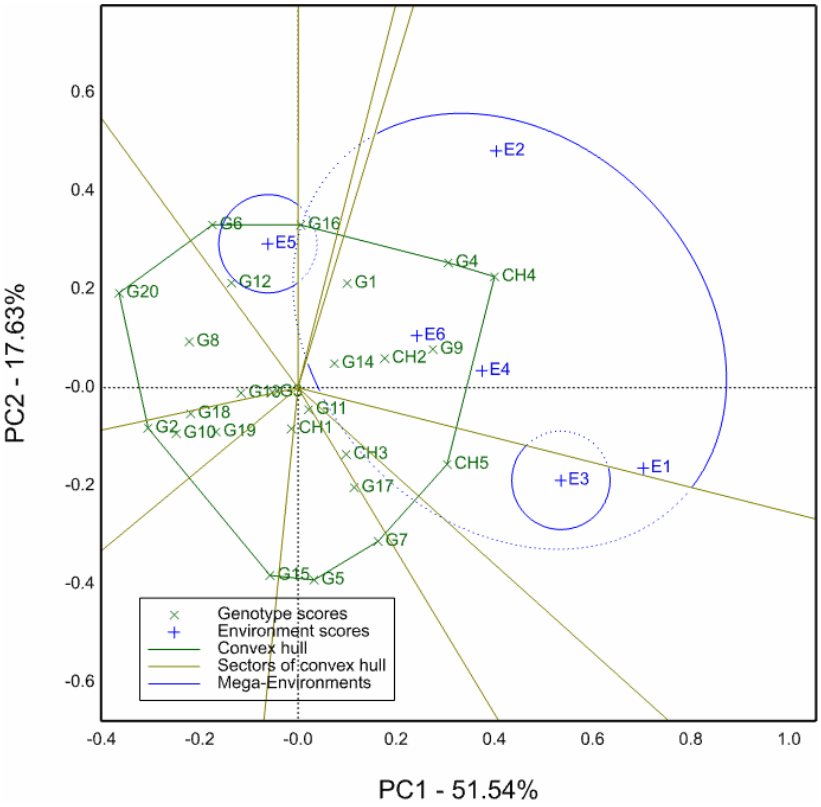

Figure 3.3 Biplot showing mega-environments and the best genotypes in each mega-environment

3.3.3.2 Genotypes ranking relative to the ideal genotype

The yield stability of genotypes was evaluated using the average environment coordination method suggested by Yan and Hunt (2001). In this method, a line called the average environment coordination (AEC) abscissa, passing through the biplot origin and the ideal genotype, was drawn and its single arrow points in the direction of increasing seed yield. The double arrowed line, passing through the biplot origin and is perpendicular to the AEC abscissa, was then drawn. This line is called the AEC ordinate; it is a measure of instability of genotypes as it moves away from the biplot origin and also separates the above average genotypes from those yielding below the grand mean. The yield stability of a genotype is estimated by a line connecting the genotype marker to the AEC abscissa. The longer this projection (vector) is from the AEC abscissa the more unstable the genotype is and vice versa.

Figure 3.4 shows that the genotypes on the left side of the AEC ordinate (such as G20, G2, G8, G10) were below the grand mean. The ones on the right side of the AEC ordinate, such as CH4, G4, G9 and G1, performed above the grand mean. It can also be visualised from Figure 3.3 that the most stable genotypes were G14, G4, G11, G3, G13, G19, CH2 and CH4. The genotype G4 and CH4 were the best genotypes as they combined both high yielding and stability. The genotypes G20, G5 and G6 were both unstable and low yielding. The stable but low yielding genotypes included G3, G13 and G19.

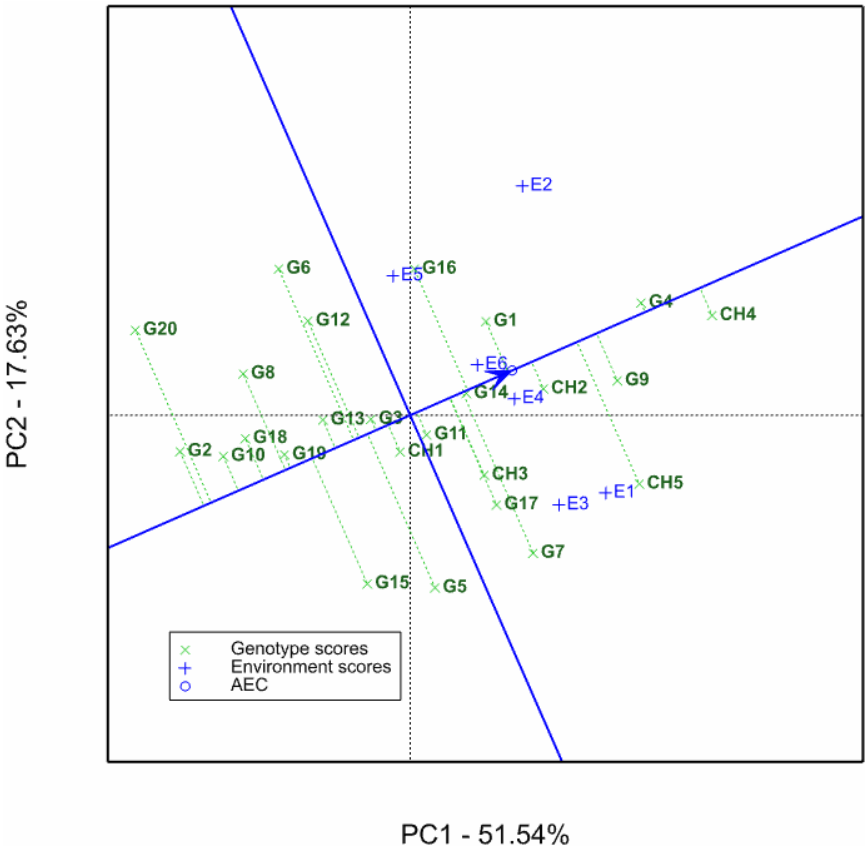

Figure 3.4 Genotypes ranking based on stability and mean perfomance

An ideal genotype should have both the longest vector of high performing genotypes and low GEI. The ideal genotype is represented by the arrow pointing it (Figure 3.5). In reality, such a genotype may not be found but it is important because of its use as the basis for genotype evaluation. The desirable genotypes are the ones proximal to the ideal genotype located at the centre. Concentric circles are used to visualise the distance between the ideal genotype and each of the tested genotypes. Figure 3.5 shows that the most desirable genotype was CH4 as it was located in the innermost concentric circle (closest to the ideal genotype). The genotypes G4, G9, CH2, G1 and CH5 were also high yielding and stable genotypes because they were proximal to the ideal genotype. The genotypes, like G20, G2, G10 and G18, on the

left side of the AEC ordinate performed below the grand mean and were furthest from the concentric circle with the ideal genotype.

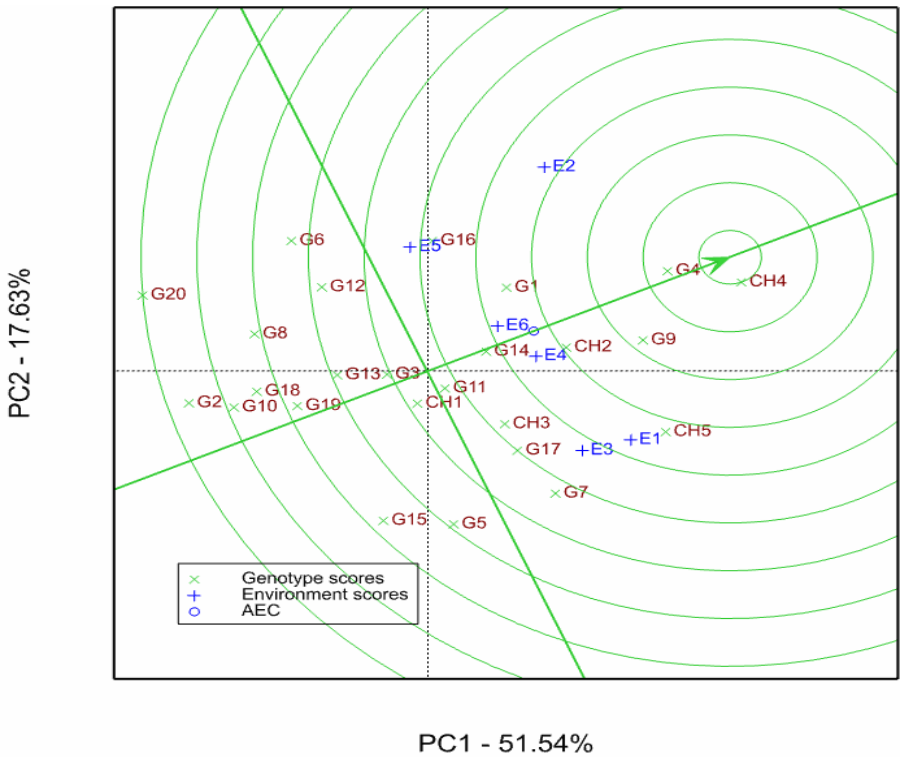

Figure 3.5 Biplot showing the genotypes ranking relative to the ideal

3.3.3.3 Discriminating ability and representativeness of test environments

The evaluation of test environments is vital as it assists in identifying ideal environments that can effectively give the best information about genotypes being tested. An ideal environment needs to have more discriminating power (indicated by high PC1 scores) as well as being capable of representing the other test environments (indicated by small PC2 scores) (Kaya et al., 2006). In Figure 3.6, the ideal environment is represented by the arrow pointing to it. The ideal environment is used to identify the most desirable environments and is a reference for the selection of genotypes in a multi-location yield trial. Hence, the more desirable environments are the ones situated nearest to the ideal environment positioned at the centre. Concentric circles are drawn to show how close or far the test environments are from the ideal environment (Yan et al., 2000). It is shown in Figure 3.6 that E4 was the most representative and discriminating environments (most useful in identifying the best genotypes). The other favourable environments included E1, E2 and E3 as they are in the concentric circles closer to the ideal environment. However, the most unfavourable environment was E6 as it fell in a

concentric circle furthest from the ideal environment and its position was on the left side of the AEC ordinate. All the favourable environments showed high mean yields for the genotypes.

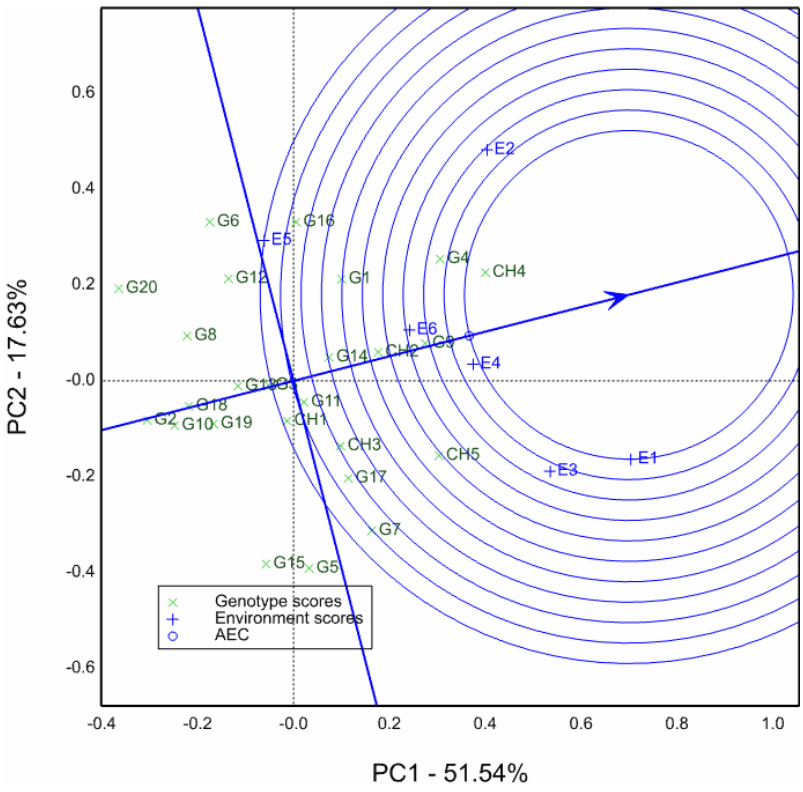

Figure 3.6 Biplot showing the ranking of the test environments based on ideal environment

3.3.3.4 Relationships among environments

In Figure 3.7 the lines connecting the environment marker and the origin of the biplot is called the environment vector. The angle between vectors of two environments indicates the association between them. An angle less than 90º means the two environments are similar and genotypes respond in a similar manner in both of them. If the angle is larger (>90º), then the two environments are negatively correlated and there is large GEI and crossover type of interaction. A right angle between environments shows that the two are unrelated (Bhartiya et al., 2017). The environment E1 and E3 were the most related followed by environments E4 and E6. The environments E5 and E4 were unrelated and E5 was negatively correlated with E1 and E3. Similarly, the association between E2 and E3 was negative. The six environments were separated into three groups on the basis of the size of angles among their vectors. The first group was composed of E1 and E3, while the second group consisted of E4 and E6. Group three consisted of E2 and E5. The most discriminating environments are the ones with the longest environment vectors, while those with very short vectors are non-informative (Yan

et al., 2007) and genotypes perform in a similar way in such environments (Yan and Kang, 2003). Figure 3.2 shows that the environments E1 and E2 discriminated the genotypes the most among the six environments.

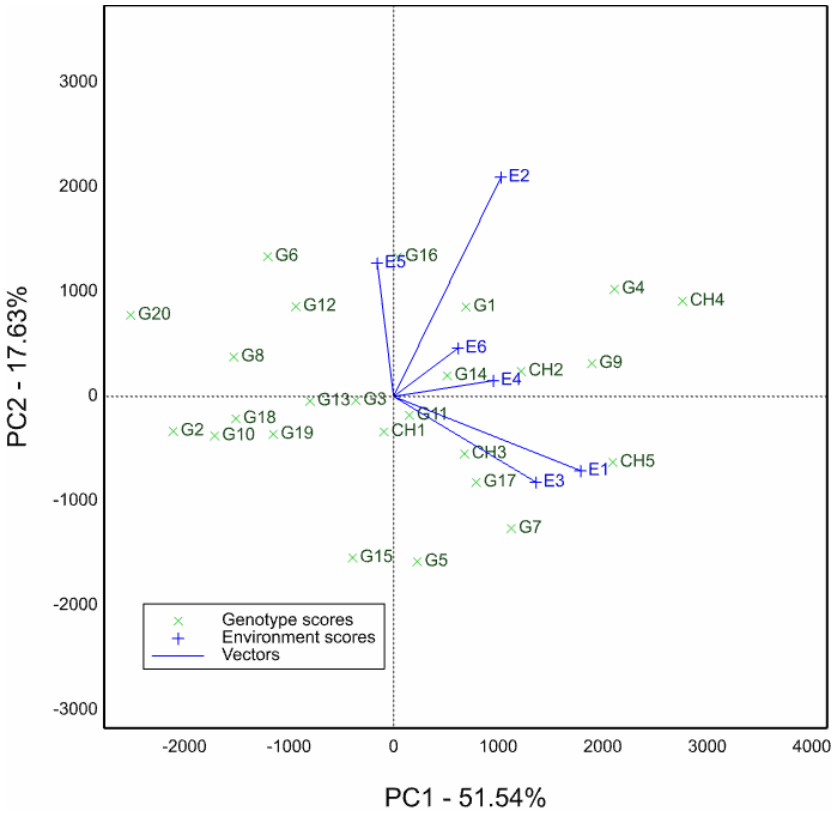

Figure 3.7 Biplot showing the relationship among the environments

3.4 Discussion

3.4.1 AMMI analysis

The AMMI ANOVA (Table 3.4) revealed that the G, E main effects and GEI were highly significant (P<0.001). The E, G and GEI effects contributed 21%, 32% and 47%, respectively, to the total variation. This means that the 25 genotypes were not the same in performance and the six environments also significantly differed from one another. Since GEI was the largest contributor to the total variation, it can be implied that the responses of the 25 genotypes were different across the six sites and there was crossover type of GEI. This was expected because the six sites in the four countries vary in terms of environmental conditions. Similar results were reported by Atnaf et al. (2013) and Bhartiya et al. (2017), who found that GEI had higher contribution to total variation than the E and G effects. On the contrary Gurmu

et al. (2009), Rakshit et al. (2012), Temesgen et al. (2015), Gurmu (2017) and Vaezi et al. (2017) all found the environment to be the highest contributor to the total variation. Five IPCAs (AMMI5) were found to be significant and IPCA1 accounted for 44% of the sum of squares of the GEI. IPCA1 and IPCA2 jointly contributed 66% to sum of squares of the GEI. The five IPCAs fully explained the GEI variation.

Table 3.5 and Figure 3.1 show that the genotypes with high IPCA1 values and whose projections to environment markers were short, were more specifically adapted to the environments. Thus, lines G20, G13, G17, G16, G7 and G19 were specifically adapted to Chitedze (E5), Nampula (E6), IITA-SARAH (E1), Lusaka West (E2) and Chipata (E3), respectively. These lines had high IPCA1 values and were closest to the above environments, hence they could be recommended for production to these environments. Different winners were identified for different environments (Figure 3.1), which confirmed that there was crossover interaction. Similar results are also shown in Table 3.6, where none of the genotypes won at least in two environments. Resulting from this, selection and recommendation of genotypes to environments would be hard. Tukamuhabwa et al. (2012) and Mushoriwa (2013) also found the presence of crossover GEI in their studies as they found different winning genotypes in the different test environments. The genotypes with the longest vectors in Figure 3.1 interacted the most with the environments and these included G20, G5, G10, G17 and G16.

Figure 3.2 shows the mean yield of environments and genotypes and how stable or unstable the genotypes were. The highest yielding environment was Lusaka West (E2): it had high mean yield (3876 kg/ha), which was above the average yield (3146 kg/ha) of all environments and genotypes. It was also found to be the most powerful in discriminating the genotypes due to having the longest vector in Figure 3.1. This environment was also reported to be good for conducting multi-location trials by Mushoriwa (2013). IITA-SARAH was second in terms of high yielding (3372 kg/ha) and discriminative ability. Figure 3.2 showed that the check (CH4) (4280 kg/ha), lines G4 (4143 kg/ha) and G9 (3820 kg/ha) were the highest yielding genotypes as their mean yields were above the grand mean yield (3146 kg/ha). The mean yields of lines G2 (2445 kg/ha), G20 (2492 kg/ha) and G18 (2530 kg/ha) were below average and the three genotypes were shown to be the least yielding across the six environments. The AMMI stability values (ASV) and IPCA scores in Table 3.5 were used to classify the genotypes according to stability. The lower ASVs and near zero IPCA scores are associated with great stability of genotypes. According to this criteria, the most stable lines were G14, G4, G11 and G3 as they had the lowest ASVs and near zero IPCA scores. These genotypes could potentially be used to breed for stability in breeding programmes, however farmers would not be attracted to them because of their low yields, except for G4. Despite being stable, G3 (2879 kg/ha) yielded below

the mean of all genotypes, while lines G14 (3369 kg/ha) and G11 (3242 kg/ha) yielded just slightly above the grand mean. The line that would appeal most to both farmers and the breeders is G4 (4143 kg/ha). It was the second highest yielding after the check SC Safari (CH4) (4280 kg/ha) and among the most stable genotypes. In view of this, this line can potentially be released as a cultivar in Zimbabwe, Zambia, Mozambique and Malawi or used in future breeding programmes as it was widely adapted (stable) and high yielding. However, these results need to be verified by testing the genotypes in many more environments and years (Yan and Kang, 2003).

3.4.2 GGE biplot analysis

Multi-location trials are conducted for the purpose of evaluating genotypes and test environments. The GGE biplot is a powerful technique for analysing data from multi-location yield trials and interpreting intricate GEI. Different principal components are used to simplify the complex GEI, whereby PC1 (which estimates the G mean performance) is plotted against PC2 (which estimates the GEI as an instability measure of each genotype) (Yan and Tinker, 2006). If PC1 and PC2 account for more than 60% of the G+GE variation and the combined effects of G and GE explains more than 10% of the total variation, the biplot is assumed to fully approximate the variability in the GE data (Yan et al., 2010). In this research, PC1 and PC2 together contributed 69% to the G+GE variation, while G and GE combined explained more than 10% of the total variation (Table 3.3). Therefore, the biplot was effective in approximating the variability in the data from the six sites. The contribution of the GE to the total variation was higher than that of the G, implying that there was possible existence of mega-environments in the testing sites. Similarly, Atnaf et al. (2013), also found that the G and GEI contributed for 15% and 60%, respectively, to the total variation and reported three mega-environments in their study.

The GGE biplot graphically presents the GEI in such a way that genotypes and environments that are close to the ideal genotype and environment, respectively, can easily be visualised (Yan and Tinker, 2006). Figure 3.4 shows that the genotype G14 was the most stable genotype as it had the shortest vector from the AEC abscissa, however it was not the highest yielding (3369 kg/ha) across environments and would not be desired by farmers. Stability alone is not meaningful if the mean yield of the genotypes is not considered. Therefore, the desirable genotype should combine stability with high mean yield (Kaya et al., 2006). In Figure 3.5, such genotypes are the ones closest to the ideal genotype positioned at the centre of the concentric circles. Thus, the genotype CH4 (check) (4251 kg/ha) had the highest combination of good yield potential and stability (widely adapted) followed by G4 (4143 kg/ha) and G9 (3820 kg/ha). The two genotypes were the most proximal to the ideal genotype (Figure 3.5).

If only the table of means (Table 3.5) is used, it would be difficult to identify the best genotypes in terms of high yielding potential and stability. This makes the GGE biplot a powerful tool in identifying superior genotypes. Since CH4 was a check, the genotypes G4 and G9 were the most desirable among the tested lines and they could be potential candidates for release in the four Southern African countries or used in future breeding programmes if the research is repeated and similar results are obtained.

The most distinctive attribute of GGE biplot is the “which won where” analysis, where the crossover GEI, mega-environment identification and specific adaptation of elite lines are visualised and easily addressed (Rakshit et al., 2014). In Figure 3.3, the polygon was divided into 10 sectors and three mega-environments were identified. The first mega-environment consisted of IITA-SARAH (E1), Lusaka West (E2), RARS (E4) and Nampula (E6), while the second and third were composed of Chipata (E3) and Chitedze (E5), respectively. The vertex genotypes in a mega-environment were the winning ones in that particular mega-environment. Thus, the check CH4 (SC Safari) and line G4 won in the first mega-environment, while CH5 and G6 won in the second and third mega-environments, respectively. The winning genotypes were different in the three mega-environments, implying that there was crossover type of GEI. Different winning genotypes in different mega-environments were also reported by Bhartiya et al. (2017) and Vaezi et al. (2017). In order to exploit the large GEI present, the winning genotype(s) (specifically adapted) could be recommended for cultivation to specific mega-environments. The genotypes, G3, G11 and G14, situated nearest to the biplot origin were the least responsive, hence they could be recommended for use in breeding for stability (wide adaptation). Figure 3.3 show that environments in the first mega-environment were similar, consequently the information obtained from the tested genotypes was similar. Therefore, a more representative environment with discriminating powers could be used to test genotypes in this particular mega-environment. This could reduce costs associated with testing genotypes and increase the efficiency of breeding. Nevertheless, this research needs to be repeated in more years and sites in order to verify this mega-environment pattern (Rakshit et al., 2012).

The biplot can also be used to understand the relationships among test environments, their discriminative capacity and representativeness by considering the length of their vectors and angles between two environment vectors. From Figure 3.7, it can be seen that the angle was smallest angle between IITA-SARAH (E1) and Chipata (E3) vectors, suggesting that the two environments were the most related. Thus, the response of genotypes to the two environments was similar and there was possible non-crosser type of GEI between the two environments. The obtuse angle between Chitedze (E5) and Chipata (E3) vectors indicated that the two environments were negatively associated and there was strong crossover type of GEI. As a

result genotypes ranked differently in the two environments e.g. genotype G16 was higher yielding in Chitedze (E5) but yielded lower in Chipata (E3). Presence of both of crossover and non-crossover types of GEI was also found by Kaya et al. (2006). The length of the environment vectors from the biplot origin in Figure 3.7 is indicative of the discriminative capacity of the environment. Lusaka West (E2) had longest vector followed by IITA-SARAH (E1), implying that the two had the most discriminative ability among the all the test environments. The highly informative but non-representative environments are important in selecting genotypes that are specifically adapted and culling unwanted genotypes, while representative environments are useful in selecting widely adapted genotypes (Rakshit et al., 2012). An ideal environment should have high discriminating capacity (large PC1 scores) as well as being representative (small PC2 scores) of other test environments. Therefore, the environment closest to the ideal environment was RARS (E4) (Figure 3.6) and this environment could be the most ideal for selecting widely adapted genotypes, though this study needs to be repeated to validate these findings.

3.5 Conclusion

AMMI and GGE biplot analyses were of great benefit in understanding the intricate GEI in this research. The study showed that GEI was the highest contributor to the total variation. The type of GEI present was crossover one and this resulted in genotypes ranking differently (different winners in each environment) and existence of mega-environments. In AMMI analyse, the lines TGx2002-17DM, TGx2001-10DM, TGx2001-18DM, TGx2014-24FM, TGx2001-6FM and TGx2002-3DM were revealed to be specifically adapted to Chitedze, Nampula, IITA-SARAH, Lusaka West and Chipata, respectively. These lines could potentially be recommended for cultivation to these environments. It was also found that 48% of the genotypes yielded above the grand mean and the rest were below the average mean. In AMMI analysis, ASV and IPCA scores were used to identify the most stable genotypes, which were TGx2002-5FM, TGx2014-5GM, TGx2001-8DM and TGx2002-7FM. These genotypes had the lowest ASVs and near zero IPCA scores.

Three mega-environments were identified by the GGE biplot, and the first mega-environment consisted of IITA-SARAH, Lusaka West, RARS and Nampula. The second and third mega-environments were composed of Chipata and Chitedze, respectively. The winning genotypes in the first, second and third mega-environments were SC Safari (check) and TGx2014-5GM; SC Squire (check); and TGx2001-24DM; respectively. GGE biplot and AMMI analyses identified Lusaka West followed by IITA-SARAH as the most powerful in discriminating genotypes and highest yielding environments. The two environments could be useful in selecting genotypes that are specifically adapted and culling unwanted genotypes. GGE biplot

analysis identified Rattray Arnold Research Station (RARS) as the most ideal environment for selecting widely adapted lines as it was both informative and highly representative of other environments. Both analyses showed that the most ideal genotype was TGx2014-5GM as it was both high yielding (4143 kg/ha) and stable across the six sites. Therefore, this medium maturing line could be a potential cultivar for release in the four southern African countries or possibly be used in future breeding programmes as a source of high yielding and stability genes.

3.6 References

Amira, J., D. Ojo, O. Ariyo, O. Oduwaye and M. Ayo-Vaughan. 2013. Relative discriminating powers of GGE and AMMI models in the selection of tropical soybean genotypes. African Crop Science Journal 21: 67-73.

Anley, W., H. Zeleke and Y. Dessalegn. 2013. Genotype x environment interaction of maize (*Zea mays* L.) across North-western Ethiopia. Journal of Plant Breeding and Crop Science 5: 171-181.

Athoni, B.K. and G.T. Basavaraja. 2012. Association analysis, genetic variability and genetic diversity studies in soybean [*Glycine max* (L.) Merrill]. Asian Journal of Bio Science 7: 114-118.

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

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

Bull, J., M. Cooper, I. Delacy, K. Basford and D. Woodruff. 1992. Utility of repeated checks for hierarchical classification of data from plant breeding trials. Field Crops Research 30: 79-95.

Eberhart, S.t. and W. Russell. 1966. Stability parameters for comparing varieties 1. Crop Science 6: 36-40.

FAO. 2017. Food and agriculture data. Food and Agriculture Organisation, Rome, Italy.

Fernandez, G.C. 1991. Analysis of genotype × environment interaction by stability estimates. Horticultural Science 26: 947-950.

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

Gurmu, F., H. Mohammed and G. Alemaw. 2009. Genotype x environment interactions and stability of soybean for grain yield and nutrition quality. African Crop Science Journal 17: 87-99.

Kaya, Y., M. Akçura and S. Taner. 2006. GGE-biplot analysis of multi-environment yield trials in bread wheat. Turkish Journal of Agriculture and Forestry 30: 325-337.

Meyer, F., L.N. Traub, T. Davids, B. Chisanga, R. Kachule, E. Tostão. 2018. Modelling soybean markets in eastern and southern Africa, regional network of agricultural policy

research institutes (RENAPRI). EUR 28978 EN, Publications Office of the European Union, Luxembourg, doi:10.2760/2059.

MOA and CSO. 2018. Crop forecast survey. Ministry of Agriculture and Central Statistics Office, Lusaka, Zambia.

Mohamedkheir K, Seifeldin I, Sapey E and H. Tianfu. 2018. History, current status, and prospects of soybean production and research in sub-Saharan Africa. The Crop Journal 3: 226-235.

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

Nassir, A.L. and O.J. Ariyo. 2011. Genotype x environment interaction and yield-stability analyses of rice grown in tropical inland swamp. Notulae Botanicae Horti Agrobotanici Cluj-Napoca 39: 220-225.

Pacheco, R., J. Duarte, R. Vencovsky, J. Pinheiro and A. Oliveira. 2005. Use of supplementary genotypes in AMMI analysis. Theoretical and Applied Genetics 110: 812-818.

Purchase, J.L. 1997. Parametric analysis to describe genotype x environment interaction and yield stability in winter wheat. PhD Thesis, Department of Agronomy, Faculty of Agriculture of the University of the Free State, Bloemfontein, South Africa.

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

Rakshit, S., K. Ganapathy, S. Gomashe, M. Swapna, A. More, S. Gadakh. 2014. GGE biplot analysis of genotype x environment interaction in Rabi grain sorghum [*Sorghum bicolor* (L.) Moench]. Indian Journal of Genetics and Plant Breeding 74: 558-563.

SAS. 2013. Base SAS 9.4 procedures guide. SAS Institute, Cary, NC, USA.

Sharifi, P., H. Aminpanah, R. Erfani, A. Mohaddesi and A. Abbasian. 2017. Evaluation of genotype × environment interaction in rice based on AMMI model in Iran. Rice Science 24: 173-180. doi:https://doi.org/10.1016/j.rsci.2017.02.001.

Sinclair, T.R., H. Marrou, A. Soltani, V. Vadez and K.C. Chandolu. 2014. Soybean production potential in Africa. Global Food Security 3: 31-40.

Temesgen, T., G. Keneni, T. Sefera and M. Jarso. 2015. 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.

Tukamuhabwa, P., M. Asiimwe, M. Nabasirye, P. Kabayi and M. Maphosa. 2012. Genotype by environment interaction of advanced generation soybean lines for grain yield in Uganda. African Crop Science Journal 20: 107-116.

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

VSNi. 2016. Genstat for windows: Eighteenth edition. Release 18.2. VSN international, ltd, Hemel Hempstead, United Kingdom.

Xu, Y. 2010. Molecular plant breeding. CABI, United Kingdom. P 381-414

Yan, W., J. Frégeau-Reid, D. Pageau, R. Martin, J. Mitchell-Fetch, M. Etienne. 2010. Identifying essential test locations for oat breeding in eastern Canada. Crop Science 50: 504-515.

Yan, W. and L. Hunt. 2001. Interpretation of genotype x environment interaction for winter wheat yield in Ontario. Crop Science 41: 19-25.

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

Yan, W. and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Yan, W., M.S. Kang, B. Ma, S. Woods and P.L. Cornelius. 2007. GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Science 47: 643-653.

Yan, W. and N.A. Tinker. 2006. Biplot analysis of multi-environment trial data: principles and applications. Canadian Journal of Plant Science 86: 623-645.

Zobel, R.W., M.J. Wright and H.G. Gauch. 1988. Statistical analysis of a yield trial. Agronomy Journal 80: 388-393.

CHAPTER 4
GENOTYPE BY TRAIT ASSOCIATIONS, CORRELATION AND PATH COEFFICIENT ANALYSES FOR SEED YIELD AND ITS COMPONENTS AMONG ELITE SOYBEAN LINES

Abstract

Developing stable cultivars with high potential for seed yield is usually the main goal for many soybean breeding programmes. The overall objective of this research study was to establish trait profiles of soybean genotypes and to study the associations among characters, their direct and indirect effects on grain yield. Twenty-five genotypes (five checks and 20 experimental lines) were evaluated in six sites in four southern African countries: Zambia (IITA-SARA, Lusaka West and Chipata), Malawi (Chitedze), Mozambique (Nampula) and Zimbabwe (Rattray Arnold Research Station) during the 2017/18 rainy season in a 5x5 alpha lattice design replicated three times in each environment. The data collected included the number of days to 50% flowering, pod clearance, plant height, days to maturity, pod number per plant, hundred seed weight and grain yield. Correlation and path coefficient analyses were conducted in SAS 9.4, while the genotype by trait (GT) associations were analysed using GEA-R software. Both GT biplot and correlation coefficient analysis revealed that pod number per plant and hundred seed weight had the highest significant correlations with grain yield, while days to 50% flowering was negatively associated with grain yield. In sequential path analysis, the number of pods per plant and hundred seed weight recorded the highest positive and significant direct effects on seed yield. Therefore, hundred seed weight and number of pods per plant can be used as grain yield selection criteria. The GT biplots revealed that lines TGx2014-5GM and TGx2002-23DM had a good combination of high yield with large seed size and high pod number. These two lines can be used as parents in future soybean improvement programmes focussing on increased seed yield or they can be released as cultivars since they were at an advanced stage of testing.

Key words: correlation, genotype by trait association, path coefficient analysis, soybean

4.1 Introduction

Globally, soybean is extensively used as food for humans, an important ingredient in livestock feed and raw material in the manufacturing of various industrial products (Sinclair et al., 2014). Developing stable cultivars with high seed yield potential is usually the main goal in many soybean breeding programmes. However, seed yield is a quantitative character greatly affected by the environment in which a genotype is grown. Consequently, seed yield has low heritability and direct selection for it may not be reliable and could be misleading (Mahawar et al., 2013). The common practice by breeders is to indirectly select for seed yield through its component characters, which are less affected by environmental conditions and have high heritability (Maleki et al., 2011). Therefore, it is imperative to know the associations that exist among yield components and how they are related to seed yield through correlation studies (Azam et al., 2018). Correlation studies enable breeders to identify traits with great influence on grain yield to be used as markers in selection programmes (Jain et al., 2017).

If all characters of interest in crop breeding were positively associated or their inheritance was independent of one another, then selection of target traits would be easier (Atnaf et al., 2017). However, negative associations among important characters exist in many cases and this complicates selection in crop breeding (Lewis, 2006). It is for this reason that correlations alone may not sufficiently explain how important each trait is in determining seed yield as the number of traits increases. Hence, it is vital to partition the correlation coefficients into direct and indirect effects through path coefficient analysis (Athoni and Basavaraja, 2012).

Path coefficient analysis explains the associations that occur within and between the causal factors that contribute to the response variable e.g. yield. In path coefficient analysis, the predictor characters are considered as first-order variables and their effects, either direct or indirect, on the dependable variable, such as yield, are analysed (Ali et al., 2009). However, this conventional approach of path analysis can possibly lead to multi-collinearity, which may make it difficult to interpret the real effect of each character due to mixed effects. This makes sequential path analysis to be more important to breeders because it helps to clearly identify the key characters that determine yield by lowering collinearity (Hair et al., 1995). Path coefficient analysis together with correlation studies help breeders to develop a selection strategy that builds up desirable alleles in the population in order to improve soybean grain yield (Aditya et al., 2011; Jain et al., 2015; Bhartiya and Aditya, 2016).

The identification of superior varieties is a key part of a soybean breeding programme. In order for a variety to be desirable, it must have all preferred traits by both farmers and consumers (Yan and Kang, 2003). A good soybean cultivar should be high yielding as well as being

superior for other important traits like disease and pest resistance, seed size, maturity, quality (oil and protein content), lodging and drought tolerance. Therefore, cultivars should be developed and selected based on key important traits to enhance their acceptability by end users. One method used to identify such genotypes with many desirable traits is the genotype by trait biplot, which is based on two concepts. Firstly, yield is the most valuable character and the other target characters are of great value only if they are combined together with high grain yield. Secondly, superior genotypes should be selected based on their ability to combine yield with other target characters and not only on their superiority in individual characters (Yan and Frégeau-Reid, 2018).

The genotype by trait method has been gaining popularity in exploiting multi-location data. It uses the same principles of the GGE biplot analysis except that different units for the characters are removed by standardising the data before the biplot is constructed. The other difference with the GGE biplot is that traits are put in place of environments before constructing the biplot (Yan and Kang, 2003). It graphically shows how well traits are correlated with one another across genotypes and helps to visualise trait profiles (merits and weaknesses) of genotypes in order to identify superior genotypes in certain traits or a group of traits (Yan and Rajcan, 2002). The superior genotypes can possibly be used in crossing programmes or released as commercial cultivars. The GT bipot also assists in identifying traits that can be used to indirectly select for the target trait e.g. seed yield and those that are not important in the improvement of the target trait. This method is also useful in culling inferior genotypes based on an individual character or group of characters. It also gives insights with regard to cultivars' usefulness for cultivation and assists in the recommendation of cultivars to specific regions (Odewale et al., 2014).

The genotype by trait biplot, correlation and path coefficient analyses have been used to study trait associations and genotype evaluation in soybean and many other crops (Yan and Rajcan, 2002; Bhartiya and Aditya, 2016; Kaplan et al., 2017; Oliveira et al., 2018). The objectives of the study were: (I) to determine the nature of correlations among characters, their direct and indirect influence on grain yield (II) to evaluate 25 elite soybean genotypes based on multiple characters and to establish winning lines in certain key characters.

4.2 Materials and methods

4.2.1 Lines and cultivars used in the study

Six sites in four southern African countries (Zambia, Malawi, Mozambique and Zimbabwe) were used to evaluate the twenty elite lines developed by the International Institute of Tropical Agriculture, and five commercial checks. The five checks included Kafue (early maturing), SC

Safari (medium maturing), SC Squire (medium maturing), MRI Dina (late maturing) and Lukanga (medium maturing). The detailed information of the lines and checks used is indicated in Table 3.1 of Chapter 3.

4.2.2 Description of the six sites used in the study

The study was conducted in Zambia, Malawi, Mozambique and Zimbabwe in the 2017/18 rainy season. Three sites were used in Zambia (IITA-SARAH, Lusaka west and Chipata) and only one site in each of the following countries: Malawi (Chitedze), Mozambique (Nampula) and Zimbabwe (RARS). The sites are described in detail in Table 3.2 of Chapter 3.

4.2.3 Trial design and management

The 25 genotypes were evaluated in a 5 x 5 alpha lattice design and each genotype was replicated three times in each environment. A plot consisted of four rows that were 0.5 m apart and 5 m long. The intra-row spacing was 0.05 m, which gave a plot size of 5 m^2 and a target population of about 350,000 plants per hectare. Basal dressing fertilizer (25 kg N/ha, 30 kg K_2O/ha, 60 kg P_2O_5/ha) was applied at planting and pre-emergence herbicides (Metolachlor and Imazethapyr) were applied soon after planting to prevent weeds from germinating. Weeds that germinated later in the season were both mechanically (hand weeding) and chemically (Quizalofop-p-ethyl and Fomesafen) controlled.

4.2.4 Data collection

The following data were collected from the trials:

i) Days to 50% flowering: number of days from planting until 50% of the plants in each plot had at least one flower.
ii) Days to maturity**:** number of days from planting to maturity. A genotype was considered mature when 95% of the pods in a plot had changed from yellow to brownish or grey.
iii) Height at harvest**:** length of the main stem (not including petioles and Leaves) at maturity. Five plants were randomly measured in each plot and the average height was recorded in cm.
iv) Pod clearance**:** average height of five plants from the ground to the first pod.
v) Hundred seed weight: mass in grams of hundred seeds from each plot.
vi) Pod number per plants: average number of pods from five random plants in each plot.

vii) Seed yield: weight in kg of air dried seed from each net plot. The weight was converted to yield in kg per hectare after being corrected to 11% moisture content (Mushoriwa, 2013).

4.2.5 Data analysis

4.2.5.1 Combined analysis of variance

The combined analysis of variance (ANOVA) was performed using PROC GLM in SAS 9.4 software (SAS, 2013). Tukey test was used to separate the means. The combined ANOVA model used is given follows:

$$Y_{ijkl} = \mu + G_i + E_j + R_{k(j)} + B_{l(jk)} + GE_{ij} + \varepsilon_{ijlk} \quad \text{Equation 4.1}$$

Where Y_{ijkl} is the response of the i^{th} genotype in j^{th} environment and k^{th} replication within environment and l^{th} block within replication; μ is the grand mean, G_i is the genotype effect i; E_j is the environment effect j; $R_{k(j)}$ is the replication within environment effect k; $B_{l(jk)}$ is the block within replication effect l; GE_{ij} is the genotype × environment interaction effect; and ε_{ijkl} is the random error.

4.2.5.2 Pearson correlation coefficients

The Pearson correlation coefficients were computed in SAS 9.4 (SAS, 2013). The general formula used for computing correlation coefficients is as given as follows:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad \text{Equation 4.2}$$

Where; r is the Pearson coefficient correlation, x and y are the two variables and n is the sample size.

4.2.5.3 Path coefficients

The following model, which was suggested by (Akintunde, 2012), was used to compute the path coefficients:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + U \quad \text{Equation 4.3}$$

Where: Y is the dependent variable (GYD), while $a+b_1X_1+b_2X_2+b_3X_3+U$ are the independent variables. Each variable is assumed to independently contribute to the dependent variable Y.

Two procedures were used to perform path analysis in SAS 9.4: conventional path analysis (PROC CALIS in SAS 9.4) and sequential path analysis using the sequential stepwise multiple regression (PROC REG in SAS 9.4). In the conventional path analysis, seed yield was considered as the response variable, while all the six yield components were taken to be first-order predictor variables. The tolerance values and variance inflation factors (VIF) were used to measure multi-collinearity, as proposed by (Hair et al., 1995). The VIF measures the extent to which the variance of a specific independent variable is affected by other independent variables affect. The tolerance value is a measure of variation of a specific independent variable, which is not contributed by the other independent variables. Hair et al. (2005) suggested that tolerance values below 0.1 and VIF values above 10 are an indication of high collinearity.

Sequential stepwise multiple regression was used to identify the independent variables with the highest contribution to total seed yield variation and having low collinearity. Such predictor variables were regarded as first-order paths, while those with low contribution to the total seed yield variation and having high collinearity as second-order paths for seed yield. This procedure was repeated by taking the variables in the first-order group as dependent variables so that their first-order variables (to be used as second-order variables for seed yield) could be identified. A path diagram was drawn using the first-and second-order variables for seed yield.

4.2.5.4 Genotype by trait biplots

Genotype by trait analysis, which is an application of GGE biplot analysis, was performed in GEA-R. The means of traits were standardised and then used to generate the biplots. The analysis was based on Model 2 (Transform=0, Scale=1 and Centering=2). This analysis revealed the winning genotypes and for what traits. The model equation suggested by Yan and Kang (2003), for genotype by trait biplot analysis is given below:

$$(Y_{ij}-\mu-\beta_j)/d_j=\lambda_1 g_{i1} e_{1j}+\lambda_2 g_{i2} e_{2j}+\varepsilon_{ij} \qquad \text{Equation 4.4}$$

Where: Y_{ij} is the genetic value resulting from the combination of the i^{th} genotype and j^{th} trait; μ is the mean value of all combinations of the j^{th} trait; B_j is the trait j main effect; λ_1 and λ_2 are the PC1 and PC2 singular values, respectively; g_{i1} and g_{i2} are the eigenvectors for PC1 and PC2, respectively, for i^{th} genotype; e_{1j} and e_{2j} are the eigenvectors for PC1 and PC2, respectively, for the j^{th} trait; d_j is the phenotypic standard deviation; and ε_{ij} is the residual associated with the combination of the i^{th} genotype and j^{th} trait.

4.3 Results

4.3.1 Combined analysis of variance

The combined analysis of variance (Table 4.1) shows significant effects of the environment on all the seven characters (P<0.001). The genotypes were significantly different from one another (P<0.001) in performance for all the seven characters. The GEI was also highly significant (P<0.001) for all the traits.

4.3.2 Comparison of genotype means for all traits across the six sites

The means of the 25 genotypes (Table 4.2) were separated using Tukey test at 5% level of significance. There is no significant difference among the means with the same letters. The range for days to 50% flowering was 48 days for genotype G16 to 60 days for genotype G18. The days to maturity varied from 105 days for the check CH1 to 123 days for the check CH3. Plant height ranged from 70 cm for G5 to 96 cm for the check CH3. Pod clearance ranged from 12 cm for the genotype G5 to 18 cm for the genotype G7. The genotype G20 had the least number of pods (42) and the highest number of pods was recorded for G4 (75). Seed size varied from 12 g/100 seeds for the check CH3 to 18 g/100 seeds for the check CH5. The genotype G2 was the least in grain yield (2445 kg/ha) and the check CH4 was the highest yielding genotype (4251 kg/ha).

Table 4.1 Combined analysis of variance for the seven traits in 25 genotypes across the six sites

Source	DF	DFFL	DM	PLHT	POD_CL	POD_PL	SWT	GYD
ENV	5	3088.89***	13712.61***	20447.74***	1479.61***	2386.85***	95.65***	14663369.30***
REP(ENV)	12	4.30ns	6.12*	14.52ns	5.31ns	25.61**	0.57ns	228953.80***
BLK(ENV*REP)	72	5.23***	3.55ns	12.98ns	4.33*	12.54ns	0.27ns	71742.80ns
GEN	24	192.98***	315.24***	583.69***	30.34***	1025.81***	38.99***	3763334.70***
ENV*GEN	120	35.80***	71.91***	221.78***	22.36***	325.33***	6.08***	1093897.80***
Error	216	2.87	2.93	14.02	3.10	9.60	0.42	55046.90
Total	449							

GYD=Grain yield, SWT=Hundred seed weight, POD_PL=Number of pods per plant, POD_CL=Pod clearance, PLHT=Plant height, DM=Days to maturity, DFFL=Days to 50% flowering, DF=degrees of freedom, ns = not significant (P>0.05), ***Significant at P<0.001, *significant at P<0.05, **significant at P<0.01

Table 4.2 Genotype means for all seven traits across environments

GENOTYPE	DFFL	DM	PLHT	POD_CL	POD_PL	SWT	GYD
G1	50.22i-l	119.72bc	77.84g	12.83fg	46.97hij	15.30e-h	3632.61bc
G2	56.03cd	116.92de	93.50ab	14.91b-f	43.81ijk	13.22mn	2444.66j
G3	52.36g-j	116.08e	79.22g	15.17b-f	45.94ijk	14.34ijk	2879.33fg
G4	50.91h-k	109.34jk	81.17fg	14.86b-f	75.34a	17.35bc	4142.95a
G5	49.02klm	106.98kl	69.91h	11.86g	45.46ijk	17.23bc	3114.07ef
G6	53.96d-g	117.18de	86.26def	15.99a-d	51.05fgh	14.31i-l	2857.18fg
G7	54.76def	113.17fgh	78.73g	18.01a	62.93bc	14.88ghi	3344.40cde
G8	57.58bc	118.87cd	88.83bcd	15.70a-d	43.54jk	15.96de	2590.84g-j
G9	52.76e-h	115.19ef	92.94ab	15.57b-e	63.42b	15.83def	3820.00b
G10	49.37klm	111.93hi	88.88bcd	16.97ab	45.23ijk	13.27mn	2710.85g-j
G11	54.90de	115.67e	85.46def	14.97b-f	44.37ijk	16.57cd	3241.94de
G12	55.96cd	116.83de	93.11ab	13.79d-g	52.45ef	13.87j-n	2814.93fgh
G13	58.93ab	121.46ab	87.74cde	16.53abc	44.78ijk	13.63k-n	2814.26f-i
G14	51.33h-k	112.84f-i	88.43bcd	13.87d-g	58.67cd	14.11i-m	3368.99cde
G15	48.50lm	112.77ghi	79.31g	15.46b-e	48.01ghi	14.70hij	2645.27g-j
G16	47.69m	105.82l	81.33fg	14.37c-f	53.21ef	15.74d-g	3314.27cde
G17	55.22d	110.76ij	92.12abc	16.84ab	51.53fg	14.25i-l	3105.21ef
G18	59.96a	116.62de	85.57def	16.15a-d	55.84de	13.02n	2530.40hij
G19	52.53f-i	109.33jk	93.19ab	15.39b-e	43.38jk	15.42e-h	2714.02g-j
G20	54.08d-g	115.04efg	82.81efg	13.21efg	41.86k	14.19i-l	2492.33ji
CH1	48.45lm	105.49l	85.49def	15.65a-d	46.03ijk	13.44lmn	3228.54de
CH2	49.50klm	110.98hij	77.95g	13.92d-g	45.37ijk	14.99f-i	3553.93bcd
CH3	59.09ab	122.56a	95.80a	16.96ab	61.58bc	11.99o	3305.39de
CH4	50.10jkl	111.06hij	85.01def	15.19b-f	60.76bc	17.70ab	4250.87a
CH5	50.01j-m	111.05hij	85.85def	14.22c-g	50.56fgh	18.43a	3740.56b
Mean	52.93	113.75	85.46	15.14	51.28	14.95	3146.31
SEm	0.34	0.34	0.75	0.35	0.62	0.13	46.92
Maximum	59.96	122.56	95.80	18.01	75.34	18.43	4250.90
Minimum	47.69	105.49	69.91	11.86	41.86	11.99	2444.70
P value	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
CV%	3.20	1.51	4.38	11.62	6.04	4.32	7.46

CV=coefficient of variation (%), SEm=standard error of means, GYD=Grain yield (kg/Ha), SWT=Hundred seed weight (g), POD_PL=Number of pods per plant, POD_CL=Pod clearance (cm), PLHT=Plant height (cm), DM=Days to maturity, DFFL=Days to 50% flowering

4.3.3 Associations among yield components and their correlations with grain yield

Table 4.3 shows the correlations between grain yield and its component characters and the association among yield components. Grain yield (GYD) showed significant and positive association with hundred seed weight (SWT) (0.229***) and pod number per plant (POD_PL) (0.666***). This trait (GYD) was also positively but non-significantly correlated with plant height (PLHT), days to maturity (DM) and pod clearance (POD_CL). The association between days to 50% flowering (DFFL) and grain yield was non-significant and negative (-0.055). Significant associations were also observed among the yield components. The trait DFFL was significantly and positively correlated with DM, PLHT and POD_CL. There was significant and positive association between DM and POD_CL. The traits PLHT and POD_CL were also significantly and positively correlated.

Table 4.3 Pearson's correlation coefficients between seed yield and its component characters

	DFFL	DM	PLHT	POD_CL	POD_PL	SWT	GYD
DFFL							
DM	0.655***						
PLHT	0.673***	0.635***					
POD_CL	0.188***	0.490***	0.334***				
POD_PL	0.031ns	-0.060ns	0.078ns	0.076ns			
SWT	0.002ns	0.129***	0.059ns	-0.084ns	0.012ns		
GYD	-0.055ns	0.017ns	0.034ns	0.086ns	0.666***	0.229***	

ns = not significant (P>0.05), ***Significant at P<0.001, GYD=Grain yield, SWT=Hundred seed weight, POD_PL=Number of pods per plant, POD_CL=Pod clearance, PLHT=Plant height DM=Days to maturity, DFFL=Days to 50% flowering

4.3.4 Path coefficient analysis

The conventional path analysis (Table 4.4) shows pod number per plant (0.676***) to have the highest positive and significant direct influence on grain yield and plant height to have the lowest direct effect, which was negative and non-significant (-0.018ns). Days to maturity (0.134*) and hundred seed weight (0.207***) also directly affected grain yield. The direct effect of pod clearance (0.022ns) on grain yield was positive but non-significant. The direct influence of DFFL on grain yield was negative but significant (-0.156**). All the indirect effects of the

traits were low and non-significant. All the traits showed low collinearity as they had less than 10 VIF values and higher than 0.1 tolerance values.

Table 4.4 The influence of all independent variables on grain yield and collinearity measures (all traits as first-order variables for grain yield)

	Direct effect	Indirect effect	Total	Tolerance	VIF
DFFL	-0.156**	0.101ns	-0.055ns	0.421	2.377
DM	0.134*	-0.117ns	0.017ns	0.371	2.699
PLHT	-0.018ns	0.052ns	0.034ns	0.465	2.149
POD_CL	0.022ns	0.064ns	0.086ns	0.673	1.486
POD_PL	0.676***	-0.010ns	0.666***	0.954	1.048
SWT	0.207**	0.022ns	0.229***	0.928	1.078

ns = not significant (P>0.05), **significant at P<0.01, ***Significant at P<0.001, *significant at P<0.05, GYD=Grain yield, SWT=Hundred seed weight, POD_PL=Number of pods per plant, POD_CL=Pod clearance, PLHT=Plant height, DM=Days to maturity, DFFL=Days to 50% flowering, VIF=variance inflation factors, DF=degrees of freedom

The first and second-order variables for seed yield were identified through sequential path analysis (Table 4.5 and Figure 4.1). The first-order traits were: days to DFFL (-0.17***), days DM (0.14**), POD_PL (0.68***) and SWT (0.20***). The direct effects of all the four first-order traits on seed yield were significant. The direct effect of DFFL was negative and the other three first-order traits (DM, SWT and POD_PL) had positive values. Among the first order variables, POD_PL was the most influential trait on grain yield followed by SWT. The four first-order variables explained about 65% of the total variation for grain yield. The second-order variables were PLHT and POD_CL, which indirectly affected grain yield through other traits. The variable PLHT showed significant and positive indirect effect on grain yield through DFFL and DM. The variable POD_CL also indirectly contributed to grain yield through DM. Both the first-and second-order variables showed low collinearity as they had less than 10 VIF values and higher than 0.1 tolerance values.

Table 4.5 Direct effects and collinearity measures for first- and second-order predictor variables.

Response variable	Predictor variable	Adjusted R^2	Direct effect	Tolerance	VIF
GYD	DFFL	0.65	-0.17***	0.56	1.79
	DM		0.14**	0.55	1.82
	POD_PL		0.68***	0.99	1.01
	SWT		0.20***	0.97	1.03
DFFL	PLHT	0.45	0.67***	0.89	1.13
DM	PLHT	0.49	0.53***	0.89	1.13
	POD_CL		0.31***	0.89	1.13
POD_PL	PLHT	0.05	0.06ns	0.89	1.13
	POD_CL		0.06ns	0.89	1.13
SWT	PLHT	0.11	0.10*	0.89	1.13
	POD_CL		-0.12**	0.89	1.13

ns = not significant (P>0.05), **significant at P<0.01, ***Significant at P<0.001, *significant at P<0.05, GYD=Grain yield, SWT=Hundred seed weight, POD_PL=Number of pods per plant, , POD_CL=Pod clearance, PLHT=Plant height, DM=Days to maturity, DFFL=Days to 50% flowering, VIF=variance inflation factors, DF=degrees of freedom

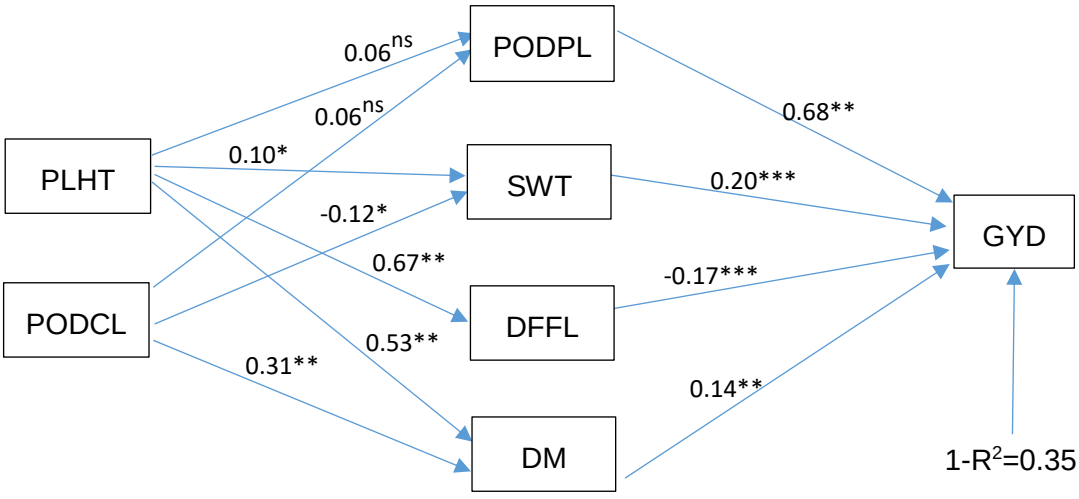

Figure 4.1 Sequential path diagram showing associations among traits contributng to seed yield.

4.3.5 Genotype by trait biplots

4.3.5.1 Relationships among traits and genotypes

In Figure 4.2 the angle between two trait vectors indicates the kind of association between two traits. An angle less than 90º between two trait vectors implies that the two traits are positively correlated. When the angle is greater than 90º, the two traits are negatively correlated, while a right angle means there is no relationship between the two characters (Yan and Tinker, 2006). Figure 4.2 shows that grain yield (GYD) was closely correlated with pod number per plant (POD_PL) and hundred seed weight (SWT). Plant height (PLHT), days to 50% flowering (DFFL), pod clearance (POD_CL) and days to maturity (DM) were negatively associated with grain yield. Hundred seed weight showed negative association with the rest of the traits except for GYD and POD_PL. Pod number per plant was positively correlated with all the traits except for DM. The biplot also shows strong associations among POD_CL, PLHT, DFFL and DM.

The genotypes with shorter projections to a trait marker were the best performers for that particular trait. Figure 4.2 shows that the genotypes G4, CH4 and G9 had the highest number of pods and grain yield. The genotypes G2, G5 and G20 were the lowest in terms of seed yield. The genotypes CH5, G4 and CH4 had the largest seed size. Among all the 25 genotypes, G13 and CH3 were the latest to reach maturity. The check CH3 was the tallest, latest to flower and had a high pod clearance value.

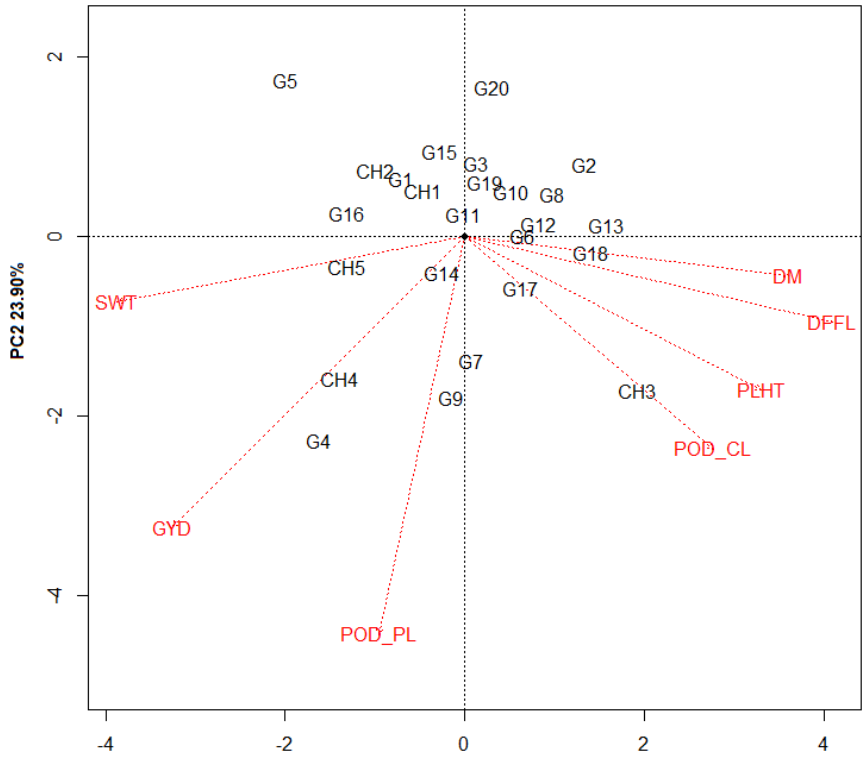

Figure 4.2 Inter-relationships among traits and genotypes

4.3.5.2 Genotype comparisons based on all the seven traits

In order to evaluate genotypes on the basis of all the seven characters, a polygon was created by joining the markers of the furthest genotypes from the biplot origin (most responsive genotypes for specific traits). The rest of the genotypes (less responsive for all traits) are contained inside the polygon. The biplot (Figure 4.3) was divided into six sectors by six rays (perpendicular lines to the polygon sides). The traits were in two sectors: the first sector contained GYD, POD_PL and SWT, while the second sector consisted of DM, DFFL, POD_CL and PLHT. The line G4, which is the vertex genotype in the first sector, was the best in combining high grain yield with high pod number per plant and large seed size. The other leading genotypes in the first sector were CH4 and G9. The vertex genotypes CH3 and G13 in the second sector flowered and matured late, were among the tallest and had high pod clearance values. Vertex genotypes in sectors without any trait were the poorest for all traits: such genotypes included G2, G5 and G20.

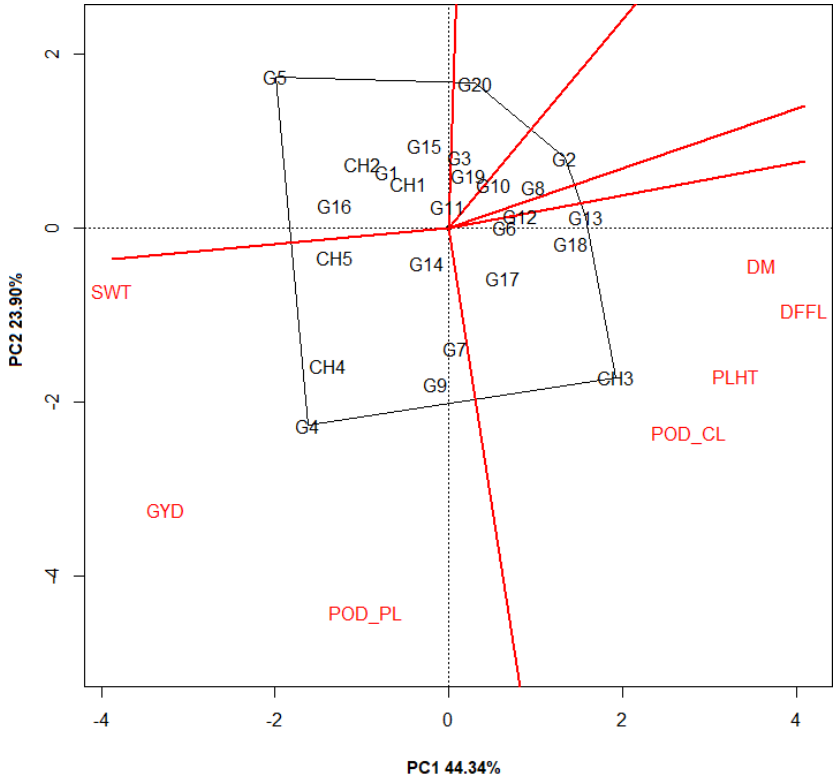

Figure 4.3 Comparison of genotypes based on multiple traits

4.3.5.3 Comparison of two genotypes

In order to compare two genotypes, a line connecting the two genotypes is drawn. Another line passing through the biplot origin and running perpendicularly to the line connecting two genotypes, is then drawn. This second line separates the characters and the two genotypes being compared. In Figure 4.4a, the highest yielding check CH4 was compared with the highest yielding genotype G4. The genotype G4 was similar to the check in terms of yield, number of pods, seed size, pod clearance and height. In Figure 4.4b, the same check CH4 was compared to the least yielding genotype G2. The genotype G2 was better than the check in terms of pod clearance, height, days to 50% flowering and days to maturity but poorer than the check in terms of grain yield, seed size and number of pods.

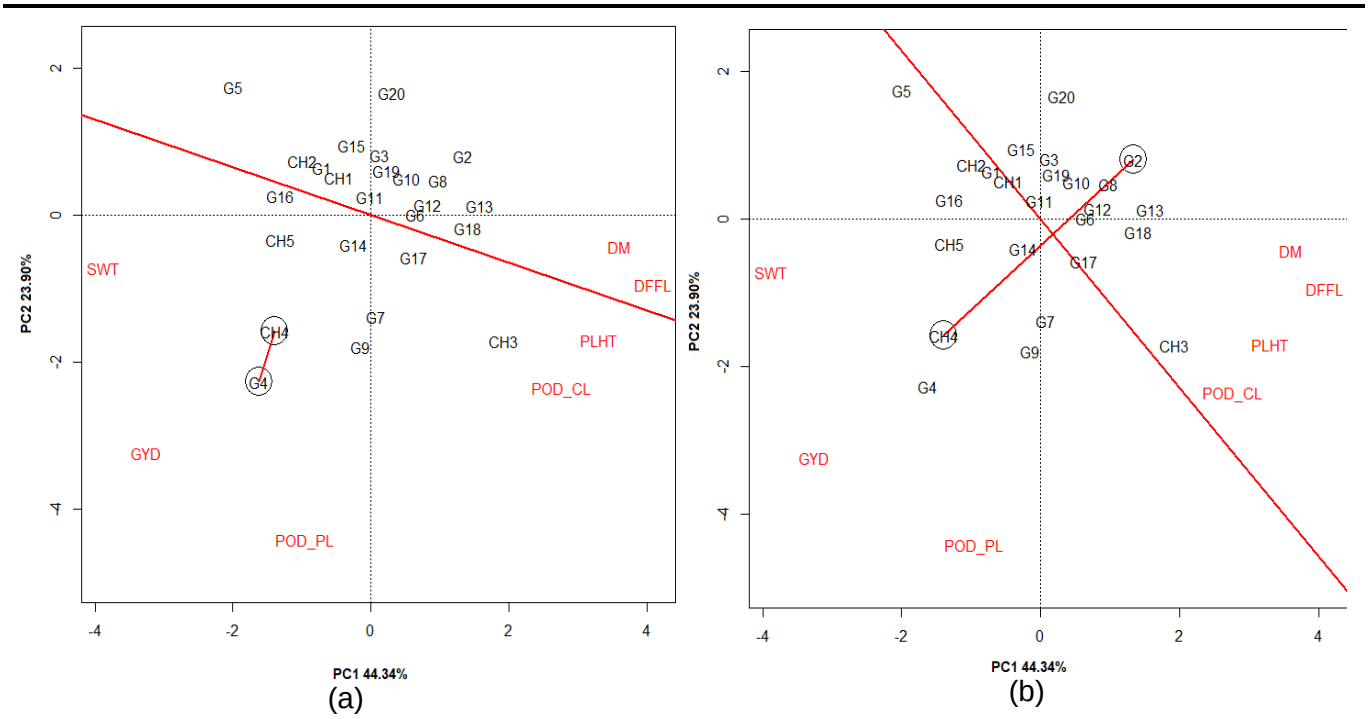

Figure 4.4 Comparison of the best check with (a) the highest and (b) poorest yielding lines

4.4 Discussion

4.4.1 Genotype mean performance for all the seven traits across the six sites

The combined ANOVA (Table 4.1) shows significant differences (P<0.005) in performance of the 25 genotypes for all the traits. In Table 4.2, it is shown that the check SC Safari (CH4), lines G4 and G9 recorded the highest mean grain yield over the six sites, with values of 4251, 4143 and 3820 kg/ha, respectively. These genotypes also had good seed size (above 15 g hundred seed weight) and high number of pods per plant (above 60). Hence, the three genotypes could be useful in crosses targeting to improve yield, seed size and pod load. The genotypes with the lowest pod clearance values, such as G5, G1 and G20, would not be ideal for combine harvesting as they started podding very close to the ground, which may result in losses during combine harvesting (less than 14 cm from the ground). Tall genotypes, such as the check MRI Dina (CH3), G2 and G19, would have a high chance of lodging, which can lead to losses during harvesting. The highest yielding genotypes (CH4, G4 and G9) were moderate in height and pod clearance. They were also medium maturing (Table 3.1 of chapter 3), hence they can potentially be ideal for medium rainfall areas in the region.

4.4.2 Correlations among traits

For a breeder to achieve yield improvement in soybean, it is important to study the associations that exist between grain yield and its related traits. Table 4.3 shows that seed yield was positively and significantly correlated with pod number per plant (0.666***) and hundred seed weight (0.229***). This means that these traits had the highest influence on seed yield among all the six components of seed yield. These findings are in conformity with those Athoni and Basavaraja (2012) and Jain et al. (2015), who reported strong association between seed yield and the two traits. Many other researchers reported similar findings (Aditya et al., 2011; El-Mohsen et al., 2013; Pawar, 2013). However, Mulridharan (2017) found the two traits to be negatively associated with seed yield. Machikowa and Laosuwan (2011) also found a negative correlation between hundred seed weight and grain yield.

Pod clearance, plant height and days to maturity had non-significant but positive correlations with seed yield. This implies that not much improvement in seed yield would be realised based on selection for these traits. Similarly, Jain et al. (2015) and Bhartiya and Aditya (2016) reported negative correlation between seed yield and plant height. On the contrary, Aditya et al. (2011) and Machikowa and Laosuwan (2011) found plant height to be an important trait for seed yield improvement in soybean. The association between days to 50% flowering and seed yield was negative and non-significant, meaning this trait may be considered to be unimportant for seed yield improvement. This is supported by Mushoriwa (2013) who also reported days 50% flowering to be negatively correlated with grain yield. The significant correlations observed among plant height, pod clearance, days to 50% flowering and days to maturity meant that selection for any of them would indirectly improve the other traits that are closely associated with it.

4.4.3 Path coefficient analysis

Correlation coefficient analysis on its own fails to provide information on the most important traits for seed yield improvement. For this reason, the correlation coefficients were used to generate path coefficients so that the traits that directly or indirectly contributed to grain yield could be identified. The conventional path analysis showed that pod number per plant (r=0.676***), hundred seed weight (r=0.207***), days to 50% flowering (r=-0.156**) and days to maturity (r=0.134*) directly influenced seed yield. Although the correlations between grain yield and days to 50% flowering and maturity were non-significant (Table 4.3), path analysis showed the two traits to have significant direct influence on grain yield. The contributions of pod clearance (r=0.022ns) and plant height (-0.018ns) to seed yield were negligible, hence

these two traits are not important for grain yield improvement. The highest direct positive contributors to seed yield were pod number per plant and hundred seed weight. Therefore, these characters could be useful in improving soybean yield and may be used as selection criteria for seed yield. Selecting for large seed size and high number of pods per plant can potentially increase soybean yield. These findings are similar to what was reported in earlier studies, where pod number per plant (Machikowa and Laosuwan, 2011; El-Mohsen et al., 2013; Jain et al., 2015) and hundred seed weight (Iqbal et al., 2003; Bhartiya and Aditya, 2016) were found to directly affect grain yield in soybean in the positive direction. Contrary to the findings of this study, direct effects in the negative direction for pod number per plant (Mushoriwa, 2013) and hundred seed weight (Machikowa and Laosuwan, 2011) on seed yield have been found.

Sequential path analysis was performed in order to lower multi-collinearity among variables so that the most grain yield contributing traits could be identified. The first-order traits were pod number per plant (r=0.68***), hundred seed weight (r=0.20***), days to maturity (r=0.14**) and days to 50% flowering (r=-0.17***). The four first-order variables explained about 65% of the total variation for seed yield and the collinearity effects were lower than in the conventional path analysis (less than 10 VIF values and higher than 0.1 tolerance values). The second-order variables, which indirectly contributed to grain yield through other traits, were plant height and pod clearance. Between the two second-order traits, plant height had higher significant and positive indirect effect on grain yield through days to 50% flowering and days to maturity. Pod clearance indirectly contributed to grain yield through days to maturity. Therefore, seed yield of soybean could indirectly be improved through plant height and directly via hundred seed weight and number of pods per plant.

4.4.4 Genotype by trait analysis

The genotype by trait biplot explained 68.24% (PC1= 44.34%, PC2=23.9%) of the total variation of traits across the 25 genotypes. Since the value is over 60%, the biplot was effective in approximating the variability in the traits data of the 25 genotypes (Yan et al., 2010). The biplot showed that number of pods per plant and hundred seed weight were the most influential traits on grain yield. This is evidenced by the smaller vector angles between the two traits and grain yield. This is supported by correlation and path analysis results, where the two traits showed the highest significant and positive association with grain yield and significant direct effects on grain yield. Therefore, the two characters could potentially be used as selection criteria for grain yield improvement. The other traits showed negative associations with seed yield and could be regarded to be less or not important in improving seed yield. Since all

variation was not accounted for by the biplot (only 68.24%), these results from the biplot may not have fully reflected what was observed in the correlation and path analysis. For example, days to maturity was found to be positively correlated with grain and was a first-order variable for grain yield in correlation and path analysis but the GT biplot showed the trait to be negatively correlated with yield. The biplot also showed that the lines TGx2014-5GM (G4) and TGx2002-23DM (G9) were comparable with the highest yielding check SC Safari (CH4) as they were high yielding and also had high pod numbers and large seed size. This can also be seen in the table of means for the traits (Table 4.2). Therefore, the two lines can potentially be released as cultivars or used in future breeding programmes for improvement of seed yield, pod load and seed size.

4.5 Conclusion

The genotype by trait biplots, correlation and path coefficient analyses were very useful in this study as they revealed the potentially useful traits in seed yield improvement of soybean. Both GT biplot and correlation coefficient analyses revealed that number of pods per plant and hundred seed weight were the most important traits, which could be used as selection criteria for seed yield. Sequential path analysis demonstrated that grain yield could be directly improved by selecting for large seed size and high pod number per plant and indirectly through plant height. The biplots revealed that lines TGx2014-5GM and TGx2002-23DM had large seed size, high pod number and high yield potential. These two lines could be potential parents for seed yield improvement in a soybean breeding programme or they could be released as cultivars since they were at an advanced stage of testing.

4.6 References

Aditya, J., P. Bhartiya and A. Bhartiya. 2011. Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). Journal of Central European Agriculture 12: 27-34.

Akintunde, A. 2012. Path analysis step by step using excel. Journal of Technical Science and Technologies 1: 09-15.

Ali, M.A., N.N. Nawab, A. Abbas, M. Zulkiffal and M. Sajjad. 2009. Evaluation of selection criteria in *Cicer arietinum* L. using correlation coefficients and path analysis. Australian Journal of Crop Science 3: 65-70.

Athoni, B.K. and G.T. Basavaraja. 2012. Association analysis, genetic variability and genetic diversity studies in soybean [*Glycine ma*x (L.) Merrill]. Asian Journal of Bio Science 7: 114-118.

Atnaf, M., K. Tesfaye, K. Dagne and D. Wegary. 2017. Genotype by trait biplot analysis to study associations and profiles of Ethiopian white lupin (*Lupinus albus*) landraces. Australian Journal of Crop Science 11: 55-62.

Azam, M., M. Hossain, M. Alam, K. Rahman and M. Hossain. 2018. Genetic variability, heritability and correlation path analysis in mungbean (*Vigna radiata* L. Wilczek). Bangladesh Journal of Agricultural Research 43: 407-416.

Bhartiya, A. and J. Aditya. 2016. Genetic variability, character association and path analysis for yield and component traits in black seeded soybean lines under rainfed condition of uttarakhand hills of India. Legume Research 39: 31-34.

El-Mohsen, A.A.A., G.O. Mahmoud and S.A. Safina. 2013. Agronomical evaluation of six soybean cultivars using correlation and regression analysis under different irrigation regime conditions. Journal of Plant Breeding and Crop Science 5: 91-102.

Hair, J., R. Anderson, R. Tatham and W. Black. 2005. Multivariate analysis of data. Porto Alegre: Bookman: 89-127.

Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black. 1995. Multivariate data analyses with readings. Englewood Cliffs, New Jersey, USA.

Iqbal, S., T. Mahmood, M. Tahira, M. Ali, M. Anwar and M. Sarwar. 2003. Path coefficient analysis in different genotypes of soybean (*Glycine max* (L.) Merrill). Pakistan Journal of Biological Science 6: 1085-1087.

Jain, R.K., A. Joshi, H.R. Chaudhary, A. Dashora and C.L. Khatik. 2017. Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Legume Research-An International Journal 41: 532-536.

Jain, S., S.C. Srivastava, S.K. Singh, Y.M. Indapurkar and B.K. Singh. 2015. Studies on genetic variability, character association and path analysis for yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Research: An International Journal 38: 182-184.

Kaplan, M., K. Kokten and M. Akcura. 2017. Assessment of genotype× trait× environment interactions of silage maize genotypes through GGE biplot. Chilean Journal of Agricultural Research 77: 212-217.

Lewis, R.S. 2006. Identification of germplasm of possible value for confronting an unfavorable inverse genetic correlation in tobacco. Crop Science 46: 1764-1771.

Machikowa, T. and P. Laosuwan. 2011. Path coefficient analysis for yield of early maturing soybean. Songklanakarin Journal of Science & Technology 33: 365-368.

Mahawar, R., L. Pawar, N. Koli, H. Chaudhary, D. Meena and M. Ali. 2013. Assessment of genetic variability, correlation and path analysis of quantitative traits in soybean [*Glycine max* (L.) Merrill]. Soybean Research 11: 70-78.

Maleki, H.H., G. Karimzadeh, R. Darvishzadeh and A. Sarrafi. 2011. Correlation and sequential path analysis of some agronomic traits in tobacco (*Nicotiana tabacum* L.) to improve dry leaf yield. Australian Journal of Crop Science 5: 1644-1648.

Mulridharan, P. 2017. Genetic analysis of traits determining seed yield over environments in soybean. PhD Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

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

Odewale, J., C. Agho, C. Ataga, E. Okolo, C. Ikuenobe, G. Odiowaya. 2014. Genotype by trait relations between yield and some morphological traits of coconut (*Cocos nucifera* L.) hybrid varieties based on GT biplot. International Journal of Plant and Soil Science 3: 270-280.

Oliveira, T.R.A.d., G.d.A. Gravina, G.H.F.d. Oliveira, K.C. Araújo, L.C.d. Araújo, R.F. Daher. 2018. The GT biplot analysis of green bean traits. Ciência Rural 48: 1-6.

Pawar, N. 2013. Characterization and categorization of major growth and yield traits in soybean [*Glycine max* (L.) Merrill. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

SAS. 2013. Base SAS 9.4 procedures guide. SAS Institute, Cary, NC, USA.

Sinclair, T.R., H. Marrou, A. Soltani, V. Vadez and K.C. Chandolu. 2014. Soybean production potential in Africa. Global Food Security 3: 31-40.

Yan, W. and J. Frégeau-Reid. 2018. Genotype by yield by trait (GYT) biplot: A novel approach for genotype selection based on multiple traits. Scientific reports 8: 1-10.

Yan, W., J. Frégeau-Reid, D. Pageau, R. Martin, J. Mitchell-Fetch, M. Etienne. 2010. Identifying essential test locations for oat breeding in eastern Canada. Crop Science 50: 504-515.

Yan, W. and M.S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists. CRC press.

Yan, W. and I. Rajcan. 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Science 42: 11-20.

Yan, W. and N.A. Tinker. 2006. Biplot analysis of multi-environment trial data: Principles and applications. Canadian Journal of Plant Science 86: 623-645.

CHAPTER 5
GENETIC VARIABILITY AND DIVERSITY AMONG ELITE SOYBEAN LINES

Abstract

Existence of genetic variability is important in cultivar development as it provides material for effective selection. The aim of this study was to estimate genetic variability for grain yield and its component traits, and to assess morphological diversity among elite soybean lines. Twenty-five genotypes (five checks and 20 experimental lines) were evaluated in six environments during the 2017/18 rainy season using a 5x5 alpha lattice design with three replications in each environment. Data collected included the number of days to 50% flowering, pod clearance, plant height, days to maturity, pod number per plant, hundred seed weight and grain yield. The combined analysis of variance, estimation of variance components, cluster and principal component analyses were performed. The combined analysis of variance revealed significant differences among genotypes ($P<0.001$). Pod number per plant and grain yield exhibited high values for both heritability and genetic advance as percent of mean, and moderate genotypic coefficient of variation values. Pod number per plant and 100 seed weight had high heritability values of 84% and 69%, respectively and were also highly correlated with seed yield, hence they can be considered as key traits for seed yield improvement through selection. In principal component analysis, only two principal components, PC1 and PC2 contributed to the total variation and both were significant and cumulatively accounted for 68.25% of the total variation. All the seven traits were useful in discriminating the genotypes as they had high eigenvalues in either PC1 or PC2. The 25 genotypes fell into two main clusters, which were further divided into eight sub-clusters based on the seven morphological characters. The most dissimilar genotypes were in sub-clusters 1 and 8. The sub-cluster 6 had the highest means for seed size (17.8 g per 100 seeds), pod number per plant (62) and seed yield (4045 kg/ha). Therefore, the three genotypes, TGx2014-5GM, SC Safari and SC Squire in sub-cluster 6 could be used as sources of desirable genes for improvement of seed yield, seed size and pod load in soybean breeding programmes.

Key words: cluster analysis, GAM, genetic variability and diversity, heritability, PCA, Soybean

5.1 Introduction

Increasing seed yield of soybean is the main objective of soybean breeding. However, grain yield being a polygenic trait, is greatly affected by the environment in which a genotype is grown. Generally, heritability for seed yield is low, which makes it difficult to directly select for it (Mahawar et al., 2013). The effectiveness of selection and improvement of yield or any other trait rely on the amount and nature of genetic variability in a breeding population (Pawar, 2013; Rao, 2016). The higher the genetic variability available in the population, the higher the possibility of developing a cultivar with broad genetic base and high resistance to biotic and abiotic challenges (Smith et al., 1991). Genetic variability is determined by genetic parameters, which include phenotypic and genotypic coefficients of variation (PCV, GCV), heritability and genetic advance (Aditya et al., 2011). It is essential to estimate these parameters as they determine how best a soybean improvement programme could utilise its germplasm (Kuldeep, 2015).

Phenotypic coefficient of variation indicates the effect of the environment on the phenotypic expression of a trait. The genotypic coefficient of variation gives insights to breeders on the available genetic variability of polygenic traits in breeding populations, but it is not reliable on its own with regard to measuring the amount of variation that is heritable. This can only be possible if GCV is estimated together with heritability. Heritability measures the size of the total phenotypic variance of a trait that can be transmitted from parents to their offspring (Kearsey and Pooni, 1998). Since heritability is determined by all the variance components, its value is affected by change in any one of the components (Bhandarkar, 1996). Atnaf et al. (2017) reiterated that heritability should be accurately estimated in order to successfully improve the trait of interest. However, heritability alone would be meaningless unless it is estimated alongside genetic advance when choosing the selection strategy to be employed (Johnson et al., 1955). Genetic advance provides estimates of the amount of progress that could be realised from one cycle of selection of the desired character (Mulridharan, 2017). High heritability is not at all times an indication of high genetic gain. When both heritability and genetic advance are high, it can be implied that the trait of interest can be improved through simple selection on the basis of its observed performance (Mulridharan, 2017).

In order for a breeding programme to be successful, there must be genetic diversity in the germplasm used to develop cultivars. Genetic diversity is defined as the amount by which the heritable traits of crops differ within a population (Pervin et al., 2007). Soybean has been reported to have a narrow genetic base due to being highly self-pollinated (Brown-Guedira et al., 2000). Crossing of elite lines followed by intensive selection in soybean breeding further

reduces genetic diversity. In most cases the released cultivars of soybean are derived by crossing elite parents, which may have similar genetic backgrounds. Hence, they may not be representative of the genetic diversity existing in the soybean genome (Chowdhury et al., 2002). This may lead to increased vulnerability of new cultivars to abiotic and biotic stresses. For this reason, it is vital to conduct genetic and morphological diversity studies before implementing a soybean improvement program. Such studies reveal genotypes with similar genetic composition and identify diverse parents that can be hybridised to develop highly diverse families from which superior cultivars can be selected (Rani et al., 2016). The use of such parents would lead to increased genetic gain from selection and ultimate rise in productivity. Incorporation of morphologically and genetically diverse germplasm from other breeding programmes would also lead to development of new diverse cultivars.

Numerous methods are used to study genetic diversity in crops. Examples of such methods include biochemical approaches, geographical origins, morphological characterisation and molecular markers (Dayaman et al., 2009; Adie and Krisnawati, 2017). Morphological characterisation is the most commonly used traditional method because of being easy, quick and cheap. Despite being influenced by the environment, morphological characters are still helpful in the estimation of genetic diversity because of their simplicity in use (Liu et al., 2011). Cluster and principal component analyses are the most commonly used multi-variate techniques for analysing morphological diversity. In cluster analysis, genotypes with similar characteristics are grouped together: minimal differences exist among genotypes within the same group, whereas genotypes in different groups are more different from each other (Hair et al., 2009). Principal component analysis (PCA) emphasises variation and reveals patterns in the dataset. It makes the visualisation and exploration of the data to be easy (similarities and differences among genotypes and traits are clearly visible). To achieve this, the data are transformed into fewer dimensions called principal components, with the aim of evaluating the usefulness of each character in relation to the total genotypic variation available. This method is capable of determining the most useful and redundant traits in the improvement of the desirable characters like seed yield (Cruz et al., 2004).

The objectives of this study were to estimate genetic parameters and to evaluate genetic diversity among elite soybean lines using morphological characters.

5.2 Materials and methods

5.2.1 Lines and cultivars used in the study

Twenty elite lines, which were developed by the International Institute of Tropical Agriculture, and five checks (Kafue, SC Safari, SC Squire, MRI Dina and Lukanga) were evaluated. The information about the lines and checks used is contained in Table 3.1 of chapter 3.

5.2.2 Description of locations

The study was conducted in Zambia, Malawi, Mozambique and Zimbabwe in the 2017/18 rainy season. Three sites were used in Zambia (IITA-SARAH, Lusaka west and Chipata) and only one site in each of the following countries: Malawi (Chitedze), Mozambique (Nampula) and Zimbabwe (RARS). The sites are described in detail in Table 3.2 of Chapter 3.

5.2.3 Trial design and management

The 25 genotypes were planted in a 5 x 5 alpha lattice design and each genotype was replicated three times per environment. Each plot consisted of four rows that were 5 m long. The rows were 0.5 m apart and the intra-row spacing was 0.05 m, which gave a plot size of 5 m^2 and a target population of about 350,000 plants per hectare. Basal dressing fertilizer (60 kg P_2O_5/ha, 25 kg N/ha, 30 kg K_2O/ha) was applied at planting and pre-emergence herbicides (Metolachlor and Imazethapyr) were applied soon after planting to prevent weeds from germinating. Weeds that germinated later in the season were both manually (hand weeding) and chemically (Quizalofop-p-ethyl and Fomesafen) controlled.

5.2.4 Data collection

The following data were recorded on each of the 25 genotypes:

i) Days to 50% flowering: number of days from sowing until 50% of the plants in each plot had at least one flower.
ii) Days to maturity**:** number of days from sowing to maturity. A genotype was considered mature when 95% of the pods in a plot had changed from yellow to brownish or grey.
iii) Height at harvest**:** average length of the main stem (not including petioles and Leaves) of five plants at maturity.
iv) Pod clearance**:** average height of five plants from the ground to the first pod.
v) Number of pods per plant: average pod number of five plants in each plot.
vi) Hundred seed weight: weight of 100 seeds in grams for each plot.

vii) Seed yield: weight in kg of air dried seed from each net plot. The weight was corrected to 11% moisture content and converted to yield in kg per hectare (Mushoriwa, 2013).

5.2.5 Data analysis

The combined analysis of variance (ANOVA) across the six sites was performed using PROC GLM of SAS 9.4 software (SAS, 2013). The means were separated using Tukey test. The model used is given as follows:

$$Y_{ijkl} = \mu + G_i + E_j + R_{k(j)} + B_{l(jk)} + GE_{ij} + \varepsilon_{ijlk} \quad \text{Equation 5.1}$$

Where Y_{ijkl} is the response of the i^{th} genotype in j^{th} environment and k^{th} replication within environment and l^{th} block within replication; μ is the grand mean, G_i is the genotype effect i; E_j is the environment effect j; $R_{k(j)}$ is the replication within environment effect k; $B_{l(jk)}$ is the block within replication effect l; GE_{ij} is the genotype × environment interaction effect; and ε_{ijkl} is the random error.

5.2.5.1 Estimation of variance components

The variance components (V_g, V_{gl} and V_e) were estimated using PROC VARCOMP in SAS 9.4 (SAS, 2013) and were used to calculate V_p, broad sense heritability, PCV, GCV, GA and GAM. The following formulae suggested by Singh and Chaudhary (1979) were used:

$$V_P = V_g + \frac{V_{g.l}}{l} + \frac{V_e}{r.l} \quad \text{Equation 5.2}$$

$$H^2 = \frac{V_g}{V_p} \quad \text{Equation 5.3}$$

$$PCV = \frac{\sqrt{V_P}}{\bar{X}} \text{ X } 100 \quad \text{Equation 5.4}$$

$$GCV = \frac{\sqrt{V_g}}{\bar{X}} \text{ X } 100 \quad \text{Equation 5.5}$$

$$GA = kH^2\sqrt{V_p} \quad \text{Equation 5.6}$$

$$GAM = \frac{GA}{\bar{X}} \text{ X } 100 \quad \text{Equation 5.7}$$

Where r and l are the numbers of replications and locations, respectively. V_e is the error variance, while $V_{g.l}$ is the variance for the interaction between the genotypes and

environments. V_g and V_p are the genotypic and phenotypic variances, respectively. X is the grand mean, k is the selection intensity at 5% (k=2.063). GCV is the genotypic coefficient of variation, PCV is the phenotypic coefficient of variation, H^2 is the broad sense heritability, GA is the genetic advance and GAM is the genetic advance as a percent of mean.

GCV and PCV were categorised using the scale suggested by Subramanian (1973), where values between 0-10%, 10-20% and higher than 20% were low, moderate and high, respectively. Heritability values were interpreted using the scale suggested by Robinson et al. (1949): values between 0-30% were low, while those between 30-60% and greater than 60% are moderate and high, respectively. The scale suggested by Johnson et al. (1955) was used to categorise the GAM values: the values between 0 -10%, 10-20% and higher than 20% were low, moderate and high, respectively.

5.2.5.2 Principal component and cluster analyses

IBM SPSS version 25 (IBM, 2017) was used for principal component analysis (PCA), based on the correlation matrix. The number of variables were reduced into few uncorrelated principal components and the contribution of each trait to the total variation was estimated. The PCA biplot was constructed in Genstat version 18.2 (VSNi, 2016) so that the relationships among genotypes based on the six traits could be visualised. Cluster analysis and construction of the dendrogram were performed using PROC CLUSTER of SAS 9.4 (SAS, 2013).

5.3 Results

5.3.1 Combined analysis of variance

The combined analysis of variance is presented in Table 4.1 of chapter 4.

5.3.2 Comparison of genotype means for all traits across the six sites

The means of the 25 genotypes are presented in Table 4.2 of chapter 4.

5.3.3 Estimates of genetic parameters

The genetic parameter estimates are shown in Table 5.3. The GCV values were lower than the PCV values for all the seven traits. The GCV ranged from 3.55% for days to maturity to 13.49% for pod number per plant. Only grain yield and pod number per plant had moderate GCV values of 13.45% and 13.49%, respectively. The rest of the traits had low GCV values (<10%). The range for PCV values was 4.01% for days to maturity to 16.29% for pod number per plant. Grain yield, pod number per plant and hundred seed weight had moderate to high

PCV values of 16.05%, 16.29% and 10.71%. All the remaining traits had low PCV values (<10%). All the traits had high broad sense heritability (>60%) except for pod clearance (<30%). Heritability varied from 27% for pod clearance to 84% for 100 seed weight. Pod number per plant and grain yield had high GAM values of 23.04% and 23.24%, respectively. Hundred seed weight and days to 50% flowering had moderate GAM values of 11.84% and 18.54%, respectively. The remaining three traits had low GAM values (less than 10%).

Table 5.1 Estimates of genetic parameters of seven characters across six locations

Tait	Mean	V_g	V_p	H^2	PCV	GCV	GA	GAM
DFFL	52.93	11.14	13.44	0.83	6.93	6.31	6.27	11.84
DM	113.75	16.26	20.84	0.78	4.01	3.55	7.35	6.46
PLHT	85.46	24.17	38.90	0.62	7.30	5.75	7.99	9.35
POD_CL	15.14	0.56	2.04	0.27	9.45	4.92	0.80	5.30
POD_PL	51.28	47.83	69.75	0.69	16.29	13.49	11.81	23.04
SWT	14.95	2.15	2.57	0.84	10.71	9.81	2.77	18.54
GYD	3146.31	178969.90	254978.19	0.70	16.05	13.45	731.19	23.24

GYD=Grain yield (kg/ha), SWT=Hundred seed weight (g), POD_PL=Number of pods per plant, DFFL=Days to 50% flowering, POD_CL=Pod clearance (cm), PLHT=Plant height (cm), DM=Days to maturity, V_p=phenotypic variance, V_g=genetic variance, H^2=broad sense heritability, PCV=phenotypic coefficient of variation (%), GCV=genotypic coefficient of variation (%), GA=genetic advance, GAM=genetic advance as a percent of mean (%).

5.3.4 Cluster analysis

The cluster analysis of 25 soybean genotypes based on the seven morphological characters revealed two major clusters, A and B (Figure 5.1), which were further divided into eight sub-clusters at 0.8 average distance between clusters. Table 5.4 shows the means of the seven traits within each sub-cluster. The sub-cluster 1 had only one genotype, CH3 (check), which flowered (mean = 59 days) and matured (123 days) late. It was also the tallest (95.8 cm), with high first pod height (17 cm), high number of pods (62), small seeds (12 g/100 seeds) and moderate seed yield (3305 kg/ha). Sub-cluster 2 had the highest number of genotypes (32%), which included G19, G17, CH1, G10, G11, G6, G15 and G3. The genotypes in this sub-cluster were early to flower (52 days), medium in maturity (112 days), medium in height (86 cm), low in number of pods (47), with small seeds (14.5 cm) and relatively moderate seed yield (2923 kg/ha). The genotypes G2, G12, G20, G8, G13 and G18 (24%) were in sub-cluster 3. These

genotypes were relatively late to flower (57 days), medium to late maturity (118 days), relatively tall (88.6 cm) and they had average first pod height (15 cm from the ground to the first pod). They were poor in terms of seed yield (2615 kg/ha), hundred seed weight (14 g) and pod number per plant (51). The sub-cluster 4 was made up of only one genotype, G7, which was moderate in number of days to 50% flowering (55 days) and maturity (113 days), short (78.7 cm), with relatively high number of pods (63), medium seeds (14.9 g) and above average seed yield (3344 kg/ha). The genotypes G14 and G9 were in sub-cluster 5 (8%). The two genotypes in this sub-cluster were relatively high yielding (3595 kg/ha), with medium seeds (15 g) and relatively high number of pods (61). They were moderate in terms of number of days to 50% flowering (52 days), maturity (114 days) and first pod height (14.7 cm) and were tall (90.7 cm). The sub-clusters 6 and 7 consisted of three genotypes each. The genotypes G4, CH4 and CH5 were in the sub-cluster 6 (12%), which was the highest yielding sub-cluster (4045 kg/ha), with high number of pods (62) and large seeds (17.8 g). The genotypes in this sub-cluster were moderate in days to 50 flowering (50 days), maturity (110 days), height (85 cm) and pod clearance (14.8 cm). The sub-cluster 7 was composed of the genotypes CH2, G16 and G5, which were moderate in seed yield (3327 kg/ha), with good seed size (16 g) and relatively low number of pods (48). The genotypes in this sub-cluster flowered early (48 days), were average in maturity (108 days), relatively short (76.4 cm) and low first pod height (13.4 cm). The sub-cluster 8 consisted of only one genotype, G1, which was relatively high yielding (3633 kg/ha), with medium seeds (15.3 g) and low number of pods (47). This genotype had a moderate number of days to 50% flowering (50 days), medium to late maturity (120 days) and relatively short (77.8 cm).

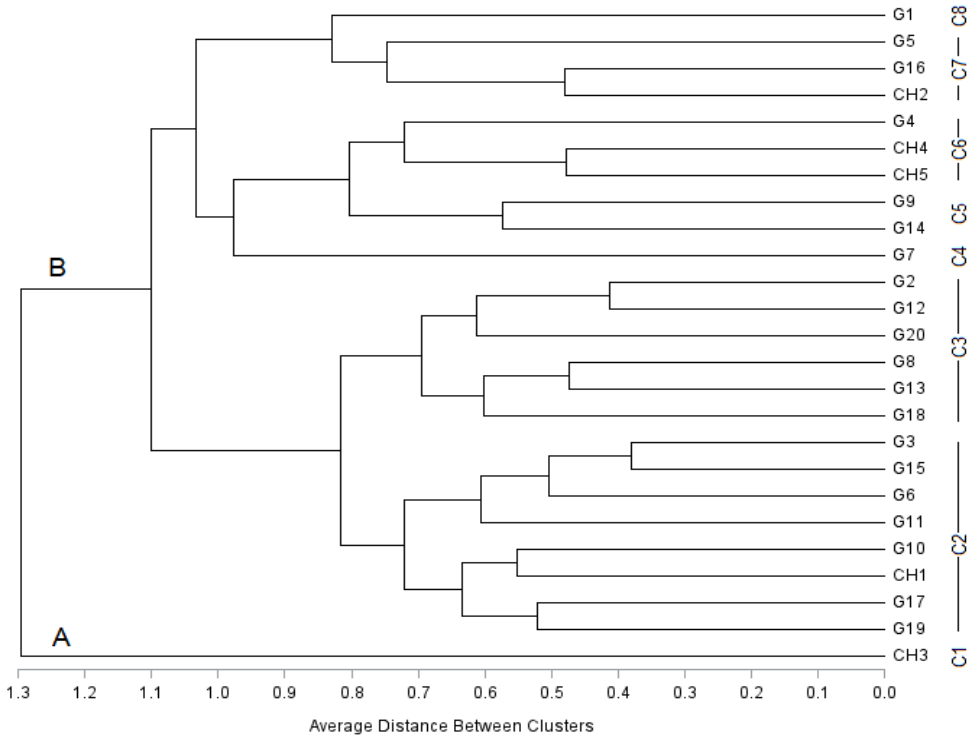

Figure 5.1 Dendrogram showing the grouping of five checks and 20 elite soybean lines based on seven morphological characters. A, B=main clusters, C1-C8=sub-cluster 1 to sub-cluster 8.

Table 5.2 Means of seven characters within each of the eight sub-clusters.

Sub-cluster	Mean						
	DFFL	DM	PLHT	POD_CL	POD_PL	SWT	GYD
Sub-cluster 1	59.09	122.56	95.80	16.96	61.58	11.99	3305.39
Sub-cluster 2	51.91	112.40	86.24	15.81	46.94	14.54	2922.79
Sub-cluster 3	57.09	117.62	88.59	15.05	47.04	13.98	2614.57
Sub-cluster 4	54.76	113.17	78.73	18.01	62.93	14.88	3344.40
Sub-cluster 5	52.05	114.01	90.68	14.72	61.04	14.97	3594.50
Sub-cluster 6	50.34	110.49	84.01	14.76	62.22	17.83	4044.79
Sub-cluster 7	48.74	107.93	76.40	13.38	48.01	15.99	3327.42
Sub-cluster 8	50.22	119.72	77.84	12.83	46.97	15.30	3632.61

GYD=Grain yield (kg/ha), SWT=Hundred seed weight (g), POD_PL=Number of pods per plant, DFFL=Days to 50% flowering, POD_CL=Pod clearance (cm), PLHT=Plant height (cm), DM=Days to maturity.

5.3.5 Principal component analysis

Table 5.5 shows the contribution of the first two significant PCs to the total variance and how the PCs were correlated with the traits. The first two PCs together contributed 68.25% (PC1=44.42%, PC2=23.83%) to the total variation. The traits that were highly correlated (high positive loadings) with PC1 mainly accounted for the PC1 variation. These included days to 50% flowering (0.852) pod clearance (0.719), days to maturity (0.703) and plant height (0.748). The most important traits in PC2 were grain yield (0.871), pod number per plant (0.903) and hundred seed weight (0.460).

Table 5.3 Principal component analysis showing eigenvectors of the seven traits, eigenvalues of the first two principal components and their contribution to the total variation.

Trait	Eigenvector	
	PC-1	PC-2
DFFL	0.852	-0.175
DM	0.703	-0.227
PLHT	0.748	0.035
POD_CL	0.719	0.208
POD_PL	0.198	0.903
SWT	-0.652	0.460
GYD	-0.328	0.871
Eigen value	3.109	1.668
Proportion of total variance (%)	44.415	23.831
Cumulative variance (%)	44.415	68.246

GYD=Grain yield (kg/ha), SWT=Hundred seed weight (g), POD_PL=Number of pods per plant, DFFL=Days to 50% flowering, POD_CL=Pod clearance (cm), PLHT=Plant height (cm), DM=Days to maturity.

The principal component biplot (Figure 5.2) shows the relationships among the seven traits and 25 genotypes. The angle between two trait vectors indicates the degree of association between two traits. The smaller the angle between two trait vectors, the higher the correlation between the two traits and vice versa. Figure 5.1 shows that grain yield (GYD) was highly correlated with pod number per plant (POD_PL) and hundred seed weight (SWT). A weak

correlation existed between grain yield and pod clearance (POD_CL), while plant height, days to 50% flowering and maturity were negatively correlated with grain yield. Strong correlations existed among plant height, pod clearance, days to 50% flowering and days to maturity. The best genotypes in a particular trait are the ones that are closer to the vector of that trait but further away from the biplot origin in the direction of that particular trait vector (often on the vertices of the convex hull). The genotypes G4, CH4, G9, CH5 and G14 were the best in terms of grain yield, hundred seed weight and pod number per plant. The genotypes CH3, G16, G18, G17 and G7 were more inclined to plant height, pod clearance, days to 50% flowering and maturity.

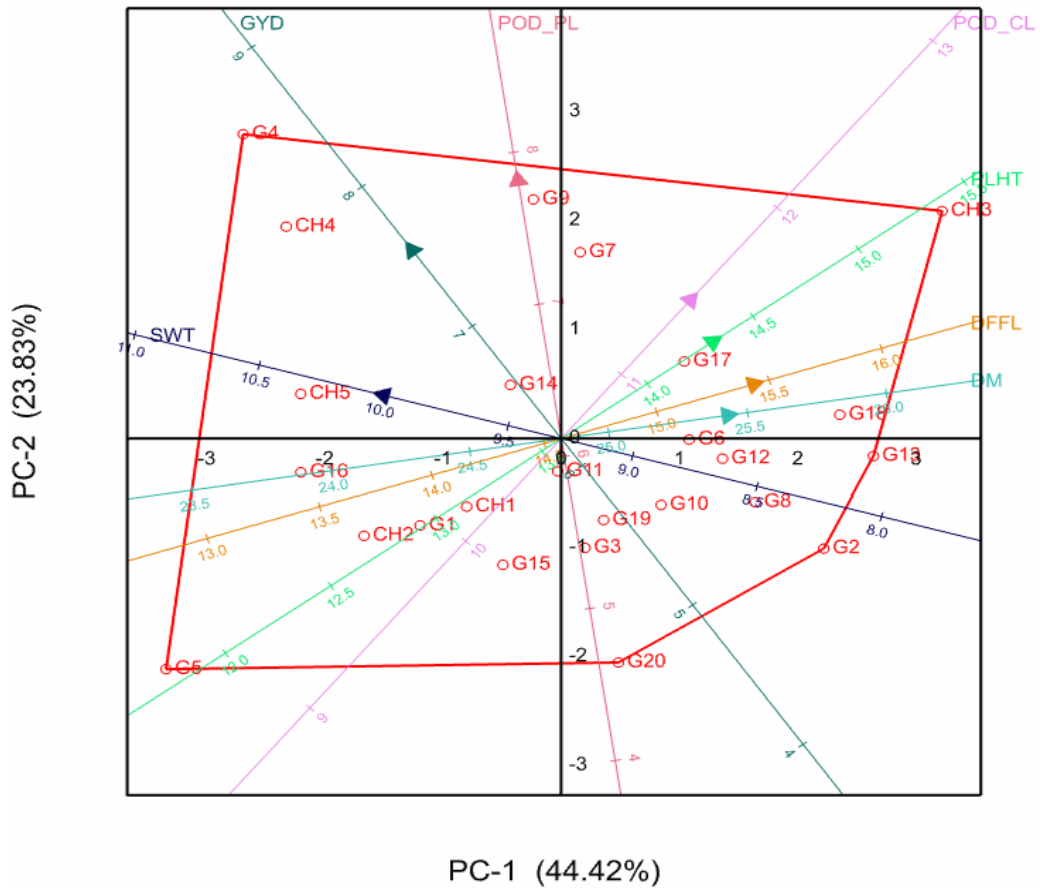

Figure 5.2 The principal component biplot shows the relationships among the seven traits and 25 genotypes. GYD=Grain yield, SWT=Hundred seed weight, POD_PL=Number of pods per plant, DFFL=Days to 50% flowering, POD_CL=Pod clearance, PLHT=Plant height, DM=Days to maturity.

5.4 Discussion

5.4.1 Genetic parameter estimates

A breeding programme can only be successful if the available amount of genetic variability in the population is known. This information indicates how heritable the trait of interest is (Majumder et al., 2008). Genotypic and phenotypic coefficients of variation are measures of genetic variability existing in a breeding population and the level of the environment's influence on the phenotypic expression of traits, respectively. In this study the PCV values were higher than the GCV values, which meant that all the traits were more influenced by the environment. However, small differences were observed between the two coefficients of variation for all the traits except for pod clearance. This was an indication that there were minimal effects of the environment and high contribution of the genes in the traits' phenotypic expression. These results concur with those obtained by Athoni and Basavaraja (2012), Jain et al. (2015) and Bhartiya and Aditya (2016), who all found that PCV values were slightly larger than GCV values for days to 50% flowering and maturity, plant height, hundred seed weight, seed yield and pod number per plant. These authors also suggested that these characters were minimally affected by the environment. Hence, these characters can genetically be improved through selection except for pod clearance. Pod clearance was the only trait that was highly affected by the environment, thus it would be difficult to improve it through selection, as it also gave low heritability (H^2=27%). The moderate GCV values recorded for grain yield and pod number per plant suggested that there was good variability, which could be exploited for the two traits. However, these results are contrary to those reported by Ghodrati (2013), who found large differences between GCV and PCV for grain yield, days to maturity, hundred seed weight and pod number per plant and suggested great influence of the environment on these traits.

In spite of the fact that GCV is a measure of the level of genetic variability in a population, it does not give the amount of genetic variation that is heritable and exploitable. This can only be possible if GCV and heritability are estimated together (Chandrawat et al., 2017). In this study all characters exhibited high broad sense heritability (>60%) except for pod clearance (<30%). This again confirms that these characters were more influenced by the genes and the environment had minimal influence on their expression except for pod clearance, which was highly affected by the environment as evidenced by its low heritability value. Moderate GCV and high heritability values were observed for pod number per plant and grain yield. This suggests that these characters could theoretically be improved through selection. Hundred seed weight also showed that it could be improved through selection, as it gave the highest heritability (84%) and its GCV value was closer to moderate (9.8%). Contrary to these findings,

Ghodrati (2013) found that grain yield, number of pods per plant and hundred seed weight had low values for both GCV and heritability, which suggested that they were highly influenced by the environment. This author instead reported plant height to have high values for heritability and GCV. Days to 50% flowering and days to maturity recorded low GCV values, despite having high heritability values, implying that they may not be improved through selection in early generations but only through hybridisation (Bello et al., 2012). Similarly, Bhartiya and Aditya (2016) reported low GCV and high heritability for days to 50% flowering and maturity.

In this study, the genetic advance for seed yield was 731 kg/ha. This means that the yield of the progeny would be increased by 731 kg if the best genotypes were selected as parents for the next cycle (selection intensity of 5%). The mean yield of the offspring would change from 3146 kg/ha to 3877 kg/ha. Similarly, number of pods per plant could be increased from 51 to 63 and hundred seed weight from 14.95 g to 17.72 g.

Broad sense heritability may not reliably be used as the only basis for selection as it is composed of both non-additive and additive effects. It is not useful on its own in indicating the amount of progress that would be realised from selection. In order for heritability to be useful, it should be estimated alongside genetic advance as a percent of the mean. In this study, pod number per plant and grain yield recorded high values for both heritability (>60%) and GAM (>20%). This was a reflection of high contribution of additive gene effects (which can be fixed from one generation to the next) to the expression of these traits. For this reason phenotypic selection based on these traits would be highly effective and high genetic gains would be achieved (Panse, 1957). Hundred seed weight also recorded moderate to high GAM and high heritability, suggesting that its expression was controlled by additive gene effects and its improvement through selection would also be possible. Jain et al. (2015), Chandrawat et al. (2017) and Jain et al. (2017) also obtained high heritability and GAM values for number of pods per plant, grain yield and hundred seed weight. The high heritability and low GAM values recorded for plant height and days to maturity were an indication of likely influence of non-additive gene effects on their expression, thus direct selection may not be effective (Abbas et al., 2018). Similarly, Bhartiya and Aditya (2016) also found high heritability but low GAM values for days to maturity and plant height. The low heritability and GAM values recorded for pod clearance suggested large influence of the environment on its expression and direct selection would be ineffective in improving it (Panse, 1957). Such a trait with low heritability and GAM values can only be improved through hybridisation (Bello et al., 2012).

Selection for seed yield in soybean can be hastened by using traits that have high positive correlation with grain yield and positive direct effects on grain yield, and high heritability.

According to the results in chapter 4, the number of pods per plant and hundred seed weight had positive correlations with seed yield and recorded the highest direct effects on grain yield. Since they had high heritability values in this chapter, the two traits could be important for seed yield improvement in soybean.

5.4.2 Genetic diversity based on morphological characters

The development of genetically dissimilar superior cultivars requires proper insight of genetic diversity in the germplasm. Diversity analysis is also useful in conserving and managing germplasm resources (Tahir and Karim, 2011; Khatab et al., 2016). In this study two multi-variate techniques i.e. cluster and principal component analyses were used to study genetic diversity among the 25 soybean genotypes. Cluster analysis is an important method for studying genetic diversity because it subdivides genotypes into clusters, such that genotypes in the same cluster are similar, while those in different clusters are dissimilar (Hair et al., 2005). In this study, two main clusters of the 25 genotypes were identified. Two clusters were further divided into eight sub-clusters based on the seven morphological characters. The morphological characters differed among clusters: each cluster had unique characteristics that differentiated it from the other clusters. The sub-cluster 2 had the highest number of genotypes (32%) followed by sub-cluster 3, which had six genotypes (24%). The sub-clusters 1, 4 and 8 each consisted of one genotype, i.e. CH3, G7 and G1, respectively. These genotypes showed high level of dissimilarity to other genotypes in terms of their characteristics, which is an indication of high variability. Figure 5.1 shows that the most dissimilar genotypes were in sub-clusters 1 and 8. The check CH3 (MRI Dina) in sub-cluster 1 was unique in the sense that it was the tallest (96 cm) among all the genotypes and was the latest to mature (123 days). It also had a fairly high number of pods but it was the least in terms of seed size (12 g/100 seeds). The line G1 in sub-cluster 8 was short (77.8 cm), medium to late in maturity (120 days), had low number of pods (47 pods/plant) and medium seeds (15.3 g/100 seeds). The two genotypes in sub-clusters 1 and 8 can possibly be hybridised followed by selection of genotypes exhibiting high variability in the successive generations because they were unrelated and fairly high yielding (CH3=3305 kg/ha, G1=3633 kg/ha).

The sub-cluster 6 had the highest means of the most desirable traits (large seed size, high pod number per plant and seed yield). The three genotypes in this sub-cluster i.e. line G4, the checks SC Safari (CH4) and SC Squire (CH5) had an average yield of 4045 kg/ha, seed size of 17.8 g per 100 seeds and 62 pods per plant. Within sub-cluster 6, the checks SC Safari (CH4) and SC Squire (CH5) were more similar to each other. The two checks are commercial cultivars that were released by SeedCo and this is a probable cause of their similarity. This is

supported by the results reported by Mushoriwa (2013), who also found the commercial SeedCo cultivars to be highly similar among themselves as they were grouped in the same cluster. The three genotypes in sub-cluster 6 would be potential sources of desirable genes for improvement of poor performing genotypes in sub-cluster 3, which had small seeds (14 g/100 seeds), low number of pods (51 pods per plant) and grain yield (2615 kg/ha). They can also be crossed with genotypes in sub-clusters 1 and 2 to improve their seed size and number of pods. This could ultimately increase grain yield variability of the progeny, from which highly diverse cultivars can be developed. However, the genotypes in sub-cluster 6 should not be crossed with those in sub-cluster 5 as this would result in narrow genetic variability among their progeny because they were similar despite being high yielding. Maximum diversity would be achieved if new lines from other breeding programmes are introduced.

Cluster analysis has been widely used by many researchers to estimate genetic diversity of soybean germplasm. For example Adie and Krisnawati (2015) used this method in their study and grouped 150 soybean genotypes into 10 sub-clusters based on 11 morphological characters. Similarly, Shadakshari et al. (2011) were able to group 50 soybean genotypes into 10 sub-clusters using 13 morphological traits. These authors reported considerable diversity among the soybean genotypes used in their studies.

Principal component analysis is helpful in determining interrelationships among characters and identifies combinations of traits that contribute the most to the total variation (Appiah-Kubi, 2012). In this study, only two principal components, PC1 and PC2 explained the variation and both of them were significant and together they accounted for 68.25% of the total variation. Since both principal components had eigenvalues greater than 1.0 (PC1=3.11, PC=1.67), they were considered to be relevant in explaining the variation among the 25 genotypes, as suggested by Kaiser (1958). All the seven traits were useful in distinguishing the 25 genotypes. The number of plant height, pod clearance, days to 50% flowering and days to maturity recorded high positive loading (eigenvalues) on PC1, which contributed 42.42% to the total variation. Hundred seed weight, number of pods per plant and grain yield had high positive loading (eigenvalues) on PC2, which contributed 23.83% to the total variation. Principal component analysis has been used in previous studies to estimate genetic diversity in soybean. In the study by Dayaman et al. (2009) it was found that the first three principal components contributed 60% to the total variation and hundred seed weight was among the useful traits in distinguishing the genotypes. In another study by Appiah-Kubi (2012), the first two principal components explained 89% of the total variation. This author further indicated

that plant height, number of seeds per pod and days to 50% flowering were the most important traits in explaining the total variation.

Figure 5.2 shows that the genotypes TGx2014-5GM (G4), SC Safari (CH4), TGx2002-23DM (G9), SC Squire (CH5) and TGx2002-5FM (G14) were the highest yielding genotypes, with the largest seeds and highest pod load. The hundred seed weight and number of pods per plant were highly correlated with grain yield due to smaller angles between their vectors and the one for grain yield. Therefore, these characters could be useful in selection programmes aiming at improving grain yield.

5.5 Conclusion

Genetic variability and diversity determine how successful a soybean breeding programme will be in improving the traits of interest, especially seed yield. Grain yield and pod number per plant exhibited moderate GCV, high values for both heritability and GAM. Number of pods per plant and hundred seed weight were highly correlated with seed yield in principal component analysis. Hence, these two traits have potential in improving seed yield because they also recorded high heritability values. The 25 genotypes were grouped into eight sub-clusters based on the seven morphological characters. The most dissimilar genotypes were in sub-clusters 1 and 8. Sub-cluster 6 had the highest means for large seed size (17.8 g per 100 seeds), pod number per plant (62) and seed yield (4045 kg/ha). The three genotypes, TGx2014-5GM, SC Safari and SC Squire in sub-cluster 6 were also confirmed to have high means for the three important traits (grain yield, number of pods per plant and hundred seed weight) in principal component analysis. Therefore, these genotypes could be used as sources of desirable genes for improvement of seed yield, seed size and pod load in soybean breeding programmes.

5.6 References

Abbas, G., M. Asghar, M. Rizwan, M. Akram, J. Hussain and F. Ahmad. 2018. Genetic analysis of yield and yield components for the improvement of mungbean germplasm. Pakistan Journal of Agricultural Research 31: 158-165.

Adie, M.M. and A. Krisnawati. 2015. Soybean yield stability in eight locations and its potential for seed oil source in Indonesia. Energy Procedia 65: 223-229.

Adie, M.M. and A. Krisnawati. 2017. Characterization and clustering of agronomic characters of several soybean genotypes. Nusantara Bioscience 9: 237-242.

Aditya, J., P. Bhartiya and A. Bhartiya. 2011. Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). Journal of Central European Agriculture 12: 27-34.

Appiah-Kubi, D. 2012. Diversity studies in soybean (*Glycine max* (L.) Merrill) and validation of shattering resistant markers for marker assisted selection. MSc Thesis, Kwame Nkrumah University of Science and Technology, Ghana.

Athoni, B.K. and G.T. Basavaraja. 2012. Association analysis, genetic variability and genetic diversity studies in soybean [*Glycine ma*x (L.) Merrill]. Asian Journal of Bio Science 7: 114-118.

Atnaf, M., K. Tesfaye, K. Dagne and D. Wegary. 2017. Genotype by trait biplot analysis to study associations and profiles of Ethiopian white lupin (*Lupinus albus*) landraces. Australian Journal of Crop Science 11: 55-62.

Bello, O., S. Ige, M. Azeez, M. Afolabi, S. Abdulmaliq and J. Mahamood. 2012. Heritability and genetic advance for grain yield and its component characters in maize (*Zea mays* L.). International Journal of Plant Research 2: 138-145.

Bhandarkar, S. 1996. Genetic variability analysis of soybean varieties under different seasons. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur (M.P), India.

Bhartiya, A. and J. Aditya. 2016. Genetic variability, character association and path analysis for yield and component traits in black seeded soybean lines under rainfed condition of uttarakhand hills of India. Legume Research 39: 31-34.

Brown-Guedira, G., J. Thompson, R. Nelson and M. Warburton. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Science 40: 815-823.

Chandrawat, K.S., K. Baig, S. Hashmi, D. Sarang, A. Kumar and P.K. Dumai. 2017. Study on genetic variability, heritability and genetic advance in soybean. International Journal of Pure and Applied Bioscience 5: 57-63.

Chowdhury, A.K., P. Srinives, P. Tongpamnak, P. Saksoong and P. Chatwachirawong. 2002. Genetic relationship among exotic soybean introductions in Thailand: consequence for varietal registration. Science Asia 28: 227-239.

Cruz, C., A. Regazzi and P. Carneiro. 2004. Biometric models applied to genetic improvement. Imprensa Universitária, Viçosa.

Dayaman, V., N. Senthil, M. Raveendran, N. Nadarajan, P. Shanmugasundaram and P. Balasubramanian. 2009. Diversity analysis in selected Indian soybean [*Glycine max* (L.) Merrill] using morphological and SSR markers. International Journal of Integrative Biology 5: 125-129.

Ghodrati, G. 2013. Study of genetic variation and broad sense heritability for some qualitative and quantitative traits in soybean (*Glycine max* L.) genotypes. Current Opinion in Agriculture 2: 31-35.

Hair, J., R. Anderson, R. Tatham and W. Black. 2005. Multivariate analysis of data. Porto Alegre: Bookman Publisher: 89-127.

Hair, J.F., W.C. Black, B.J. Babin, R.E. Anderson and R.L. Tatham. 2009. Multivariate data analysis. Bookman Publisher.

IBM. 2017. IBM SPSS statistics for windows, version 25.0. Armonk, NY: IBM Corporation.

Jain, R.K., A. Joshi, H.R. Chaudhary, A. Dashora and C.L. Khatik. 2017. Study on genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Legume Research- An International Journal 41: 532-536.

Jain, S., S.C. Srivastava, S.K. Singh, Y.M. Indapurkar and B.K. Singh. 2015. Studies on genetic variability, character association and path analysis for yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Research: An International Journal 38: 182-184.

Johnson, H.W., H. Robinson and R. Comstock. 1955. Estimates of genetic and environmental variability in soybeans 1. Agronomy Journal 47: 314-318.

Johnson, H.W., H. Robinson and R.E. Comstock. 1955. Genotypic and phenotypic correlations in soybeans and their implications in selection 1. Agronomy Journal 47: 477-483.

Kaiser, H.F. 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23: 187-200.

Kearsey, M.J. and H.S. Pooni. 1998. The genetical analysis of quantitative traits. Stanley Thornes (Publishers) Ltd, United Kingdom.

Khatab, I.A., A.R. Morsf, W.M. Fares and T. Kumamaru. 2016. Genetic diversity of soybean genotypes revealed by agro-morphological and SSR markers. Journal of the Faculty of Agriculture, Kyushu University 61: 79-87.

Kuldeep S. C, K.S.B., D. H. Sarang, PH. Kiihne Dumai, P. U. Dhone and AniL Kumar. 2015. Association analysis for yield contributing and quality parameters in soybean. An International Quarterly Journal of Environmental Sciences 7: 113-118.

Liu, M., M. Zhang, W. Jiang, G. Sun, H. Zhao and S. Hu. 2011. Genetic diversity of Shaanxi soybean landraces based on agronomic traits and SSR markers. African Journal of Biotechnology 10: 4823-4837.

Mahawar, R., L. Pawar, N. Koli, H. Chaudhary, D. Meena and M. Ali. 2013. Assessment of genetic variability, correlation and path analysis of quantitative traits in soybean [*Glycine max* (L.) Merrill]. Soybean Research 11: 70-78.

Majumder, D., A. Shamsuddin, M. Kabir and L. Hassan. 2008. Genetic variability, correlated response and path analysis of yield and yield contributing traits of spring wheat. Journal of the Bangladesh Agricultural University 6: 227-234.

Mulridharan, P. 2017. Genetic analysis of traits determining seed yield over environments in soybean. PhD Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

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

Panse, V. 1957. Genetics of quantitative characters in relation to plant breeding. Indian Journal of Genetics 17: 318-328.

Pawar, N. 2013. Characterization and categorization of major growth and yield traits in soybean [*Glycine max* (L.) Merrill]. MSc Thesis, Department of Plant Breeding and Genetics, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (MP), India.

Pervin, M., M. Polash, S. Rahman and A. Deb. 2007. Study of genetic variability and GxE interaction of some quantitative traits in blackgram [*Vigna mungo* (L.) Hepper]. Journal of Biological Sciences 7: 169-175.

Rani, A., V. Kumar, B. Gill, V. Morya and R. Rawal. 2016. Assessment of genetic diversity of soybean genotypes differing in resistance against yellow mosaic virus using simple sequence repeat markers. Legume Research 39: 674-681.

Rao, V.T. 2016. Genetic variability, correlation and path coefficient analysis under drought in groundnut (*Arachis hypogaea* L.). Legume Research 39: 319-322.

Robinson, H., R.E. Comstock and P. Harvey. 1949. Estimates of heritability and the degree of dominance in corn. Agris: International Information System for the Agricultural Science and Technology 41: 353-359.

SAS. 2013. Base SAS 9.4 procedures guide. SAS Institute, Cary, NC, USA.

Shadakshari, T., T. Kalaimagal, N. Senthil, M. Boranayaka, R. Kambe Gowda and G. Rajesha. 2011. Genetic diversity studies in soybean [*Glycine max* (L.) Merrill] based on morphological characters. Asian Journal of Bio Science 6: 7-11.

Singh, R.K. and B.D. Chaudhary. 1979. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi, India.

Smith, S., A. Al-Doss and M. Warburton. 1991. Morphological and agronomic variation in North African and Arabian alfalfas. Crop Science 31: 1159-1163.

Subramanian, S., Menon, M. 1973. Heterosis and inbreeding depression in rice. Madras Agricultural Journal 60: 1139-1145.

Tahir, N.A.R and H.F.H. Karim. 2011. Determination of genetic relationship among some varieties of chickpea (*Cicer arietinum* L) in Sulaimani by RAPD and SSR markers. Jordan Journal of Biological Sciences 4: 77-86.

VSNi. 2016. Genstat for windows: Eighteenth edition. Release 18.2. VSN international, Ltd, Hemel Hempstead, United Kingdom.

CHAPTER 6
GENERAL OVERVIEW OF THE STUDY

6.1 Introduction

Soybean [*Glycine max* (L.) Merrill] is the world's leading source of protein and vegetable oil. There has been an increase in production of the crop in recent years in Southern Africa as a response to the crop's ever increasing demand. Currently, soybean ranks third after maize and wheat in terms of production in the region. However, its productivity is still low in the region due to limited availability of stable and high yielding cultivars, which are tolerant to many abiotic and biotic stresses. It is important to test elite soybean lines in multi-locations in order to identify and recommend the high yielding ones for release as cultivars. Studies of genetic variability and associations among important traits, and diversity in breeding populations are useful for successful improvement of grain yield in soybean. Therefore, the objectives of this study were:

i) To determine the magnitude of genotype by environment interaction and stability of elite soybean lines for seed yield.
ii) To establish trait profiles of 25 soybean genotypes and to study the associations among characters, their direct and indirect effects on grain yield.
iii) To estimate genetic variation through genetic parameters of traits related to seed yield, and to analyse genetic diversity among elite soybean lines.

The aim of this chapter is to highlight the major findings of the study and their implications to soybean breeding in Southern Africa.

6.2 Summary of major findings

6.2.1 Genotype by environment interaction and stability of 25 soybean genotypes

The data collected from multi-location trials conducted in the 2017/18 rainy season using six sites in four countries *viz.* Zambia, Malawi, Zimbabwe and Mozambique were used to evaluate the stability of 20 elite soybean lines and five commercial checks for seed yield. The data were subjected to GGE biplot and AMMI analyses. The following were the findings from the two methods:

- The AMMI analysis revealed that the contribution of environment, genotypes and GEI effects to the total variation were 21%, 32% and 47%, respectively. Five IPCAs were

found to be highly significant (P<0.001). The GEI was the highest contributor to the total variation and the GEI present was of the crossover type.

- Both AMMI and GGE biplot analyses indicated Lusaka West as the highest yielding and most informative environment.
- According to the GGE biplot analysis, Rattray Arnold Research Station was the most ideal environment for testing genotypes.
- In the GGE biplot analysis, three mega-environments were identified. Mega-environment one was made up of IITA-SARAH, Lusaka West, RARS and Nampula, while the Mega-environments two and three were composed of Chipata and Chitedze, respectively.
- AMMI analysis showed that the lines TGx2002-17DM, TGx2001-10DM, TGx2001-18DM, TGx2014-24FM, TGx2001-6FM and TGx2002-3DM were specifically adapted to Chitedze, Nampula, IITA-SARAH, Lusaka West and Chipata, respectively.
- Both analyses revealed that the line TGx2014-5GM was the widely adapted and the second highest yielding (4143 kg/ha) genotype.

6.2.2 Genotype by trait associations, correlations and path analysis for seed yield

The data collected on the number of days to 50% flowering, pod clearance, plant height, days to maturity, pod number per plant, hundred seed weight and grain yield from the six locations were subjected to correlation and path coefficient analyses for seed yield and genotype by trait analysis. The findings were as follows:

- Both GT biplot and correlation coefficient analysis revealed that pod number per plant and hundred seed weight were the most positively correlated traits with grain yield, while days to 50% flowering had a negative association with grain yield.
- In sequential path analysis, the number of pods per plant and hundred seed weight recorded the highest positive and significant direct effects on seed yield, while plant height had a high indirect effect on seed yield.
- The GT biplots revealed that lines TGx2014-5GM and TGx2002-23DM had good combinations of high yields with large seed size and high pod number.

6.2.3 Genetic variability and diversity among elite lines of soybean

The combined data from the six locations on the seven morphological traits were subjected to cluster and principal component analyses in order to study genetic diversity among the 25 genotypes. The variance components estimated from the data (Vg, Vgl and Ve) were used to calculate Vp, broad sense heritability, PCV, GCV, GA and GAM. The findings were as follows:

- The GCV values were lower than the PCV values, which meant that all the traits were influenced by the environment. However, small differences were observed between the two coefficients of variation for all the traits except for pod clearance, which was highly affected by the environment.
- The genetic advance values for seed yield, 100 seed weight and pod number per plant were 731 kg/ha, 2.77 g and 12, respectively. This meant that the yield of the progeny would be increased by 731 kg if the best genotypes were selected as parents for the next cycle (selection intensity of 5%). The mean yield of the offspring would change from 3146 kg/ha to 3877 kg/ha. Similarly, hundred seed weight and number of pods per plant would be increased from 14.95 g to 17.72 g and 51 to 63, respectively.
- Moderate GCV values of 13.45% and 13.49%, high heritability values of 70% and 69% and GAM values of 23.24% and 23.04% were recorded for grain yield and number of pods per plant, respectively. This suggested that the two traits were highly influenced by additive gene effects, therefore phenotypic selection based on these traits would be highly effective and high genetic gains would be achieved.
- Hundred seed weight also showed that it could be improved through selection because it recorded the highest heritability (84%), moderate to high GAM (18.5%) and its GCV value was closer to moderate (9.8%).
- Although the traits plant height and days to maturity recorded high heritability values of 62% and 78%, they had low GCV values of 3.55% and 5.75% and GAM values of 6.46% and 5.75%, respectively. These were indications of likely influence of non-additive gene effects on these traits, thus direct selection may not be effective.
- The low GCV (4.92%), heritability (27%) and GAM (5.3%) values were recorded for pod clearance, suggesting that this trait was largely influenced by environmental effects and direct selection to improve it would be ineffective.
- The 25 genotypes were grouped into eight sub-clusters.
- The most dissimilar genotypes were the check MRI Dina and TGx2001-11DM in sub-clusters 1 and 8, respectively.
- The sub-cluster 6 had the highest means of the most desirable traits (large seed size, high pod number per plant and seed yield). The three genotypes in this sub-cluster i.e. line TGx2014-5GM, checks SC Safari and SC Squire had an average yield of 4045 Kg/ha, seed size of 17.8 g per 100 seeds and 62 pods per plant.
- Only two principal components, PC1 and PC2 were found to explained the variation and both of them were significant because they had eigenvalues greater than 1.0 (PC1=3.11, PC=1.67). Together they accounted for 68.25% of the total variation.

- The genotypes TGx2014-5GM (G4), SC Safari (CH4), TGx2002-23DM (G9), SC Squire (CH5) and TGx2002-5FM (G14) recorded the highest yields and number of pods and had the largest seeds.

6.3 Breeding implications and recommendations

The presence of crossover type of GEI could make it difficult to select and recommend genotypes for production due to differences in ranking of genotypes in the six environments. However, Lusaka West could be useful in selecting genotypes that are specifically adapted and culling unwanted genotypes, while Rattray Arnold Research Station could be useful for selecting widely adapted lines as it was both informative and highly representative of other environments. Testing genotypes in one environment that is both representative and informative could reduce costs and increase the efficiency of breeding. The lines TGx2002-17DM, TGx2001-10DM, TGx2001-18DM, TGx2014-24FM, TGx2001-6FM and TGx2002-3DM could potentially be recommended for cultivation to the environments where they showed specific adaptation. The line TGx2014-5GM could potentially be released as a cultivar in the four southern African countries on account that it had large seeds (17.4 g/100 seeds), high pod load (75 pods/plant) and showed high yielding potential (4143 kg/ha) and stability. This line could also be used as a parent in future soybean improvement programmes. However, these findings need to be validated by testing the genotypes in at least two years and many more sites.

Hundred seed weight and number of pods per plant have high potential for use as grain yield selection criteria on account that they were positively correlated with seed yield and had the highest direct effects on seed yield. They also recorded high heritability and GAM values. The three genotypes (TGx2014-5GM, checks SC Safari and SC Squire) in sub-cluster 6 are potential sources of desirable genes for improvement of grain yield, seed size and pod load. In addition, the principal component analysis showed high genetic diversity among the genotypes for possible exploitation in soybean breeding to increase seed yield. Although morphological traits were able to reveal genetic diversity and to group the genotypes into different clusters, molecular markers, are more reliable for diversity studied as they are not influenced by the environment. Therefore, molecular markers should be used to confirm if the results obtained in this study are accurate.