Breeding Gains, Diversity Analysis and Inheritance Studies on Soybean [*Glycine max* (L.) Merrill] Germplasm in Zimbabwe

By

Hapson Mushoriwa
BSc, MSc and MBA (University of Zimbabwe)

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD) in Plant Breeding

School of Agricultural Earth and Environmental Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

Republic of South Africa

29 November 2013

THESIS ABSTRACT

The soybean programme in Zimbabwe is over seventy years old. However, there is lack of information on breeding gains, genetic diversity, heritability, genetic advance, combining ability, gene action and relationships between grain yield and secondary traits available for breeding. Therefore, the aim of the present study was to characterise the genetic diversity of the available germplasm, determine gene action conditioning grain yield and estimate the breeding gains that have been realised since the inception of the breeding programme.

Evaluation of 42 soybean genotypes for genetic diversity conducted during 2010/11 and 2011/12 cropping seasons, using phenotypic and molecular characterisation approaches, revealed evidence of wide diversity among the genotypes. The phenotypic traits and SSR markers assigned the soybean genotypes to 8 and 15 clusters respectively. The SSR marker technique was more polymorphic, informative and highly discriminatory. The clustering pattern and relatedness from SSR data was in agreement with the pedigree data while the phenotypic clustering was divorced from pedigree data. Genotypes, G41 and G7; G41 and G1; G41 and G42 were the most divergent; therefore, they could be utilized as source germplasm in cultivar development and commercial cultivars.

Investigations on breeding gains involving 42 cultivars (representing a collection of all the varieties that were released in Zimbabwe from 1940 to 2013) showed that improvement in grain yield was slowing down. However, annual genetic gain was estimated to be 47 kg ha^{-1} $year^{-1}$ representing an annual gain of 1.67%. Furthermore, grain yield ranged from 2785 to 5020 kg ha^{-1}. Genotypes, G16, G15, G17, G1 and G42 exhibited superior performance in grain yield and other agronomic traits and are therefore, recommended for utilisation in the hybridisation programme. Seed protein concentration decreased by 0.02 $year^{-1}$ while oil increased by 0.02, 100 seed weight increased by 0.21 g $year^{-1}$ over time. In addition, number of days to 95% pod maturity and pod shattering increased by 0.35 and 0.38 days $year^{-1}$ respectively while lodging declined by 0.31%. Results indicated that emphasis should be refocused on grain yield to restore the original linear increase.

Assessment of the magnitude of GEI and stability of 42 released cultivars was done over 13 environments and two seasons using additive main effects and multiplicative interaction, cultivar superiority and rank analyses. Results showed that environment and GEI captured larger portion of the total sum of squares, which reveals the influence of the two factors on grain yield, hence, the need for evaluating soybean genotypes in multi-environment trials and over years. Further, the data revealed that GEI was of a crossover type because of differential yield ranking of genotypes. The three stability parameters selected two

genotypes, G1 and G15, as the most productive, consistent and stable, thus they could be produced in diverse environments while G2, G4, G5, G7, G16, G40, G17, G18 and G31 were identified as unstable and suitable for specific adaptation.

Correlation and path analyses showed that grain yield was positively and significantly correlated with number of branches per plant, number of nodes per plant, shelling percentage, and number of days from 95% pod maturity to first pod shattering, implying that breeding and selection for these traits probably improved grain yield. Number of nodes per plant, plant height and 100 seed weight exhibited highest direct effects on grain yield while, number of nodes per plant and plant height presented the highest indirect effects on grain yield. These results demonstrated that number of nodes per plant and plant height could be recommended as reliable selection traits for developing high yielding genotypes of soybean.

DECLARATION

I, Hapson Mushoriwa declare that

1. The research reported in this dissertation, except where otherwise indicated, and is my original work.

2. This dissertation has not been submitted for any degree or examination at any other university.

3. This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other people.

4. This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   a. 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.

5. Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

6. 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.

Signed:

_____

Hapson Mushoriwa (Candidate)

As the candidate's supervisor I agree to the submission of this thesis

_____

Professor J. Derera (Supervisor)

_____

Professor P. Tongoona (Co-supervisor)

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Professors P. Tongoona and J. Derera for their invaluable support, guidance, encouragement and constructive criticism from the outset or conception of this study, execution, through to final write up of this thesis.

I am deeply indebted to Seed Co and Pannar Seed management for funding this study. I am quite thankful to the Seed Co Group Head of Research for granting me the opportunity develop myself. Special thanks to Seed Co for providing and allowing me to use their germplasm. Many thanks to Mr. Jacob Tichagwa for the advice and unfailing support that made this study successful. My profound gratitude goes to the Seed Co soybean team for tirelessly manning my field trials and collecting data (Mr. Henry Ndlovu, Mr. Brighton Mbala, Mr.Shacky Mudzimande and Izet Nyamacheta). Special thanks to Dr. Antony Jarvie for accommodating me during the time of thesis writing. I would like to thank Mr. Caiphas Muyambo and Mr. Percival M. Sithole for assisting me to format the entire document.

I wish to extend my gratitude to Mr. Herbert Masole of Seed Co Zambia, Ms. Sophie Mahuwa of Seed Co Malawi, Mr. L. Mutemeri of ART farm, Zimbabwe and Mr. Zvarova of Crop Breeding Institute for taking care of my trials.

I thank my friends and college mates in particular Dr. Nyasha Kaseke, Gorden Mabuyaye and Lennin Musundire for their encouragement, great friendship and walking the journey together.

I owe special thanks to my family for their unwavering support, love and patience during the duration of study. To my dearest wife, Judith, many thanks for your understanding and standing in for me on a number of family issues, I am indebted forever! To my children, Nyasha and Hazel, this thesis should serve as an inspiration in your quests to fulfill your wishes. I am deeply indebted to my mother, Emilia Mushoriwa, father, Maraini Mushoriwa, my brothers, sisters and in-laws for rendering emotional, endless and moral support.

I am deeply grateful to the ALMIGHTY God for good health, strength and if anything enabling everything

DEDICATION

I would like to dedicate this work to my wife, Judith and children, Nyasha and Hazel Anesu.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AFLP : Amplified fragment length polymorphism

AMMI : Additive main effects and multiplicative interaction

ANOVA : Analysis of variance

cm : centimetres

CV : Coefficient of variation

CROIL : Percentage crude oil

CRPR : Percentage crude protein

D : Determinate

DFL : Days to 50% flowering

DMAT : Days to 95% pod maturity

DMS : Downy mildew scores

DNA : Deoxyribonucleic acid

DSH : Days from January 1 to first pod shattering

E : Environment

ECV : Environmental coefficient of variation

Hb : Heritability in the broad sense

g : grams

G : Genotype

GCA : General combining ability

GYLD : Grain yield

F : Significance of the variance ratio

GCV : Coefficient of genetic variation

GEI : Genotype by environment interaction

Ha : Hectare

He : Expected heterozygosity

Ho : Observed heterozygosity

I : Indeterminate

IITA : International Institute of Tropical Agriculture

IPCA : Interaction Principal Component Analysis

Kg : Kilograms

Kg/ha : Kilograms per hectare

LSD : Significance at alpha level of 5%

m.a.s.l : metres above sea level

MET : Multi-environmental trials

Mm : Millimetres

MSE : Mean square error

N : North

NAMC : National Agricultural Marketing Council of South Africa

NTSYS : Numerical taxonomy multivariate analysis system

PCA : Principal component analysis

PCV : Phenotypic coefficient of variation

PIC : Polymorphic information content

PDHT : Pod height in centimetres

PLHT : Plant height in centimetres

RAPD : Random amplified polymorphic DNA

RFLP : Restriction fragment length polymorphism

S : South

SAP : Seed appearance scores

SE : Standard error

SCA : Specific combining ability

SNP : Single nucleotide polymorphism

SSR : Simple sequence repeats

SVD : Singular value decomposition

US : United States

V_G : Genetic variance

Yr	: Year
%	: Percentage
%LODG	: Percentage lodged plants at maturity
%SS	: Percentage sum of squares
100SDMA	: 100 seed weight in grams

INTRODUCTION TO THESIS

1 Significance of soybean

Soybean [*Glycine max* (L.) Merrill] total production ranks first among the important oilseed crops worldwide (Jonas *et a*l., 2008). It is followed by cotton, sunflower, canola, palmoil and peanuts. Jonas *et al.* (2008) reported that the total annual global oilseed output was estimated to be approximately 310 million tonnes of which soybean accounted for 56%. Further, in terms of global crop production, it is ranked fourth after maize, wheat and rice (Kaga *et al.*, 2012). Production is dominated by the United States, Brazil and Argentina representing 35%, 30% and 27% of the global soybean respectively (NAMC, 2011). The continent of Africa produces a small quantity of soybean that is less than 1% of the total world production (Opperman, 2011). The wide range of uses has presented enormous and insatiable demand. Being a versatile crop, it has great potential to transform African economies as well. For example, Southern Africa shares similar agro-climatic conditions with Argentina and Brazil and possesses similar amount of land that could be put to soybean production (Opperman, 2011). Generally, Africa has great potential for soybean production given its favourable climate. Figure 1 below presents the areas where soybean has great potential for production in Africa.

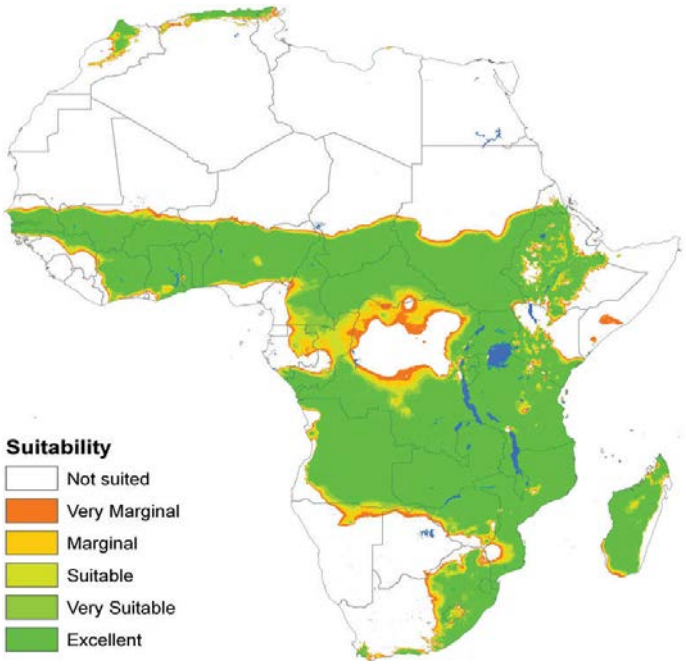

Figure 1: Soybean Suitability Map

Source: IITA (2008)

Given the total land area suited to soybean production in Africa, it means that the continent can improve its production and significantly increase its share of the annual global soybean

output. In order to make an impact in Africa, the development of superior cultivars remains a challenge and priority. This requires a detailed insight into genetic diversity of available germplasm and its extent of variability. Aditya *et al.* (2011) reported that information on genetic variability, heritability, genetic advance, correlation and path coefficient is necessary for soybean breeders to select the best parental stock and breeding methodology for genetic improvement. Bonato *et al*. (2006) stressed that availability of such information is a pre-requisite for organizing a working collection, identifying heterotic groups and selecting parents for crosses. By the same token, knowledge of the past and present breeding gains is crucial for establishing whether the newly developed and registered cultivars and elite lines are genetically more advanced than preceding cultivars. It presents opportunities to allocate resources efficiently and breeding effort to traits that have significant impact on grain yield. Further, it reveals gaps in respect of trait improvement.

2 Soybean production in Zimbabwe

In Zimbabwe, soybean is widely produced in agro-ecological regions [1]I, [2]II and [3]III where rainfall reliability and distribution are relatively better than other ecologies (Mugandani *et al.*, 2012). Given even distribution and adequate rainfall in all parts of the country, higher yields could be obtained under low elevations because of high heat units. The crop is used for processing edible vegetable oil, soybean cake and a variety of food products, with the first two dominating (Opperman, 2011). According to Kapuya *et al.* (2010), soybean accounts for about 30% of the total edible vegetable oil produced. There is a growing demand for soybean which is attributed to growth in the poultry and pig industries which are also driven by escalating demand for meat. Kapuya *et al.* (2010) estimated that the national annual soybean demand in Zimbabwe was pegged at 165 000 tonnes. Sadly, demand far outstrips supply given production volumes of 37 000 and 50 000 metric tonnes recorded in 2010-11 and 2011-12 seasons respectively (Esterhuizen, 2011). The statistics present an unhealthy situation and the huge supply demand gap has to be filled up with imports. Zimbabwe is therefore, a net importer of raw soybean, soybean meals, and vegetable oil. Furthermore, the crushing capacity is underutilised. The annual oilseed crushing capacity is estimated at 500 000 metric tonnes and only 40% will be utilised (Esterhuizen, 2011). Apparently, the shortfall is imported from Zambia, Malawi and India.

[1] Region I = Annual Rainfall > 1000mm, highly suitable soil texture, mean annual temperature of 15-18°C and length of growing season greater than 165 days
[2] Region II = Annual Rainfall 1000-700mm, highly suitable soil texture, mean annual temperature of 16 -19°C and length of growing season ranges from 150-165 days
[3] Region II = Annual Rainfall 700-550mm, highly suitable soil texture, mean annual temperature of 18-22°C and length of growing season ranges from 135-150 days

The said low output is partially attributed to low yields. Lower yields reduce the overall volumes directly and simultaneously reduce the production area because some farmers may abandon the crop. Furthermore, low productivity adversely affects economic viability and sustainability of the industry. Kapuya *et al.* (2010) estimated the productivity of soybean in Zimbabwe for 10 years and observed that yield ranged between 670 kg ha^{-1} and 2300 kg ha^{-1}. Although the yield levels are generally low, a few good commercial farmers are reported to be achieving over 4 000 kg ha^{-1} (TechnoServe, 2010). Moreover, seed yields of about 5 000 kg ha^{-1} have been recorded on research stations (Seed Co, 2008). Although yield gains are a function of cultivar genetic improvement and improved cultural practices, it may be argued that yield disparities shown here are largely attributed to differences in improved cultural practices. This is because potential yield gains can be clearly determined through comparing the farmer's yields versus those obtained when applying the best practices of management in high yielding environments (Egli, 2008a; Foulkes *et al.*, 2009). In the same breath, Louisiana Agricultural Extension Services (LAES) observed farmers' yields that ranged between 60 to 80% of the optimal levels (LAES., 2009). Arguably the yield gap may be due to environmental stresses such as weeds, nematodes, humidity, rainfall, temperature, soil factors, solar radiation, inadequate moisture, pests, diseases, nutrient deficiencies and low rates of fertilizers (Board *et al.*, 2010).

The factors constraining (productivity constraints) soybean production in Zimbabwe are basically biotic and abiotic. Biotic stresses, such as pests, weeds and diseases can reduce seed yield quite markedly. Mirsky *et al.* (2013) observed that weeds remain a perennial problem and impact negatively on soybean productivity. More importantly, soybean rust reduce seed yield significantly in regions where it has become endemic and where the conditions are favourable for its development and yield losses ranging from 40 to 60% have been reported (Mueller *et al.*, 2009). Only a few new cultivars have been registered with moderate seed yields in Zimbabwe. Semi-loopers damage the foliage, thus indirectly reducing seed. Apparently, heat and drought are recognized as the major damaging abiotic stresses. The Consultative Group on International Agricultural Research, (CGIAR, 2012) estimated the yield loss from heat and drought to range between 18 to 28% on soybean, which threatens food security. Drought has become more prevalent than before and is further aggravated by climate change as well as land degradation (Kashyapi *et al.*, 2012). Apart from heat and drought stresses, soil factors in particular low soil fertility, limit soybean yield. Further, both macro-nutrients and micro-nutrients are required in sufficient quantities for growth and development and simultaneously for maximum yield. The soybean crop removes a sizeable amount of phosphorous and potassium, thus inadequate application reduces the yield levels.

The above challenges can be addressed by plant breeding. From the perspective of crop improvement, each breeding cycle generates new lines that are perceived to be better than the previous cycle. If the incremental gains are accumulated over time, they constitute higher productivity. In this context, it becomes necessary to estimate both genetic gains and diversity and the results will serve as a basis for reviewing the breeding strategies. In the same vein, exploitation of combining ability analysis assists to select better parents that advance genetic gains.

3 Soybean breeding history in Zimbabwe

Before the introduction of soybean in 1940, the government of Zimbabwe used to import fish meal for feeding livestock. Thus, fish meal was the chief source of protein. However, the cost became unsustainable and at this point, the Zimbabwean government initiated soybean breeding with a view to using the soybean cake and fodder to feed livestock. Earlier efforts therefore, focused on breeding cultivars for fodder and this effort resulted in the release of Hernon strains (Shurtleff and Aoyagi, 2007). Apart from utility as a fodder crop, it was also used to improve soil fertility through the green maturing culture. Full support to the program began in the early 1960s, when a breeding team was engaged with a mandate to breed and develop both grain and fodder cultivars. More resources were channeled to the program with a view to achieve self- sufficiency. Over a period of 70 years, 42 cultivars were registered for commercial production in Zimbabwe. It is probably worth acknowledging Mr. Rex Tattersfield and Jacob Tichagwa who developed and registered all these cultivars. Over the last decade, the Zimbabwean soybean breeding has spread into Zambia and Malawi and has successfully intensified its testing. This was necessary because soybeans are photoperiodic sensitive and consequently, flowering and maturity are largely influenced by latitude and temperature. Moreover, variability in edaphic factors, other climatic conditions and other stresses called for direct evaluation under those environments. The evaluation effort has led to the release of a number varieties in these market domains.

Significant genetic gains have been attained in various agronomic traits (personal observation). Shurtleff and Aoyagi (2007) reported that some of the early releases were yielding 534 kg ha^{-1}. The present varieties have seed yield potential ranging between 4 000kg/ha to 5 000 kg ha^{-1} which represent highly significant breeding gains over seven decades. They further reported that earlier introductions were characterised by poor resistance to diseases, shattering and lodging, rendering them unfit for commercial production. In terms of phenological traits, the current releases flower and mature within the season lengths of the production domains. The first fodder cultivar was very late and would

require irrigation if rains tailed off earlier than normal whereas the first grain cultivar was early maturing, which was disasterous because it would mature in the midst of the rainy season. This could imply good progress in overcoming adaptation challenges. It could be argued that the achievements made represent sound breeding progress. It is also envisaged that the genetic gains that were realized over the given time-frame represent valuable germplasm improvement. Such germplasm can be exploited during hybridization or gene deployment. The modern cultivars can also be exploited to achieve the optimum productivity.

Further increases in soybean yield remains a primary issue in the interest of adequate food supplies (Egli, 2008b). Therefore, information on gene action and or combining ability, and heritability for soybean yield and other agronomic traits would be important in planning breeding strategies. Moreover, knowledge on the mechanisms controlling grain yield and its secondary traits is vital for developing high yielding lines and cultivars in general. Thus, these genetic parameters are exploited in advancing breeding gains through identifying and selecting superior genotypes for yield and its secondary traits (Machikowa *et al.*, 2011; Nassar, 2013). Selection on the basis of combining abilities helps to enrich the available germplasm.

However, advances in crop improvement of self-pollinating crops may result in narrowing of the genetic base through selection of the most preferred traits over a long period of time. Each year selection of the parental lines is limited to a chosen few and to make matters worse, breeders recycle a small subset of the elite commercial lines (Mikel *et al.*, 2010). The most common selection criteria used by soybean breeders is to identify elite lines with outstanding agronomic merit and use them to create new populations and rarely incorporate foreign material. Clearly, the approach has shown that the recombination of elite germplasm enhances the chances of improving the progeny. However, this practice repeated over time results in reduced genetic diversity. As a result, the commercial cultivars are thought to have originated from a limited number of key elite lines. The impact of such breeding strategy is felt by low genetic gains. The yield gains could either grow relatively slowly or remain stagnant ultimately implying low yields per unit area. This poses a challenge to continual genetic improvement. In other words, the major concern hinges on future breeding advances given diversity constraints (Gadde, 2006). Given the foregoing, the investigation was undertaken to estimate the genetic gains, levels of genetic diversity and generate information on combining ability, gene action and heritability meant to help accelerate the present breeding gains.

4 Problem statement

The major concern regarding the soybean breeding programme in Zimbabwe is that, the genetic gains have not been estimated ever since it was started. Precisely, this means that there is lack of information on genetic progress that has been accomplished. Lange and Federerizzi (2009) reiterated that genetic gain analysis serves to measure the success of a breeding programme. They also indicated that genetic gain analysis enables breeders to compare breeding progress realized from different breeding strategies applied at different times as well as different environments. Thus, a good understanding of the past events is equally useful and important. Lack of such valuable information, may adversely affect future breeding advances. Arguably, the availability of such information helps breeders to redefine, elaborate and adopt appropriate breeding strategies (Lange and Federerizzi, 2009).

In addition, there is also lack of information on genetic diversity available for breeding. The genetic diversity pattern for the germplasm that has been accumulated over seven decades is unknown. The quantification of genetic diversity is fundamental to crop improvement (Ojo *et al.*, 2012). It reveals information on diverse parents necessary to develop new populations. Divergent parents are known to produce segregants with wide genetic variability which is crucial for selection (Iqbal *et al.*, 2008). Thus, information on genetic diversity is useful in planning crosses. Although large germplasm collections may have been acquired over time, there could be a gap between the available germplasm and its use. This stems from lack of genetic diversity data which could imply minimal or poor utilisation of the available genetic resources. Genetic diversity studies could also generate information that enables breeders to classify the available germplasm and identification of subgroups of core collections with possible utility for specific breeding purposes (Mohammadi and Prasanna, 2003). Oliveira *et al.* (2010) noted that the estimation of genetic diversity through phenotypic traits helps to identify accessions of questionable value which essentially would be discarded and/or replaced.

The other challenge is that the Zimbabwean soybean programme seems to be experiencing stagnation in yield growth. This could be evidence of a yield plateau. The yields of the current varieties seem to be the same as varieties that were released ten years ago [Mutemeri, *personal communication*[4] and personal observations]. The appearance of yield plateaus may threaten food security. This may worsen the situation given that demand is apparently outstripping supply because the ability of agricultural production systems to expand the food supply will be hindered.

[4] = crop production specialist

Further, the soybean production environments are highly variable in the three countries where part of the study was conducted giving rise to complicated genotype by environment interactions (GEI). Contextually, there is no clear and documented information regarding the characterisation of the Zimbabwe germplasm in these multi-environments. Crossa (1990) asserted that multi-environment trials are instrumental in; *"assessing yield stability and pattern of response of genotypes in wide range of environments, precisely estimating and predicting yield based on limited experimental data and providing reliable guidance for selecting the best genotypes or agronomic treatments for planting in future years and at new sites."* Kang (2004) described GEI as a situation where the relative performance of genotypes changes from one environment to another. Thus, the germplasm should be characterised across multi-environments. The information generated would therefore, enable breeders to identify and select the genotypes to use for breeding purposes. Essentially, GEI serves to inform breeders of the breeding strategy to adopt; either specific or general adaptation (Fox *et al.*, 1997).

5 Rational for Research

Generally, the genetic diversity of soybean in most African breeding programmes has not been described, poorly understood, sometimes not well documented and underutilised. By the same token, the Zimbabwean germplasm has not been characterised. The improvement of soybean grain yield and quality is accomplished through hybridisation, biotechnological methods or mutagenesis followed by selection. Thus, an effective programme requires a broad genetic base upon which to select the parental lines for use in crosses. Hence, it would be beneficial to the soybean programme and the nation at large to have a good understanding of the relatedness of the available germplasm. Knowledge of the genetic diversity is also crucial for germplasm conservation. The material to be conserved as germplasm requires to be characterized for variability and validate its usefulness in crop improvement (Oliveira *et al.*, 2010). On another note, the documentation of breeding lines has been based on phenotypic characters, however, the scope of the present study is to characterise the germplasm using both molecular and phenotypic characters to determine the diversity as well the relatedness. More importantly, the Zimbabwean soybean breeding programme has expanded regionally and globally, which further necessitates the need to audit the existing germplasm.

Previous studies have shown that plant breeding exercised over a long period of time may result in narrowing of diversity because only a few parents contribute to the primary gene pool as well as repeated use of the same parents. Reyna and Sneller (2001) reported that

the breeding strategy involving the use of high yielding parental lines with good agronomic traits led to the reduction of the soybean genetic base in the United States germplasm. Furthermore, there has been an overuse of released cultivars because of their agricultural merit consequently decreasing genetic diversity and breeding progress (Hyten, 2005). Some of the observed consequences of reduced genetic diversity include; the exposure of the available germplasm to genetic vulnerability, yield plateaus, genotypes may be exposed to ever changing environmental stresses they may not cope with, climate change may demand new varieties and continued erosion of diversity. All these issues negatively affect breeding progress, thus jeopardizing future improvement.

Of particular interest, is the need to maintain continued genetic progress or gains in grain yield. Soybean seed yield gains have been quantified in major producing countries (Specht *et al.*, 1999; Duvick, 2005; Rowntree *et al.*, 2013), but information is lacking for tropical soybean particularly in Southern Africa where the climatic conditions, soil texture and depth seems to be quite favourable for soybean production. Sound knowledge of the genetic gains in grain yield is a prerequisite in order to have a good background of yield contributing traits. By the same token quantification of the genetic gains from all the agronomic traits cannot be overemphasized. The knowledge is useful in crafting successful strategies that ensure future yield increases. Egli (2008a) pointed out that advances in yield increases are partly a function of a good understanding of the past increases. There is great potential for yield improvement in soybean, given that highest grain yield of about 10 000 kg ha^{-1} has been recorded in United States on commercial farms (www.time.com/time/nation/article/0,8599,2084388,00.html, accessed on 26-06-2013).

7 Research Objectives

The specific objectives of the thesis were as follows;

1. To determine the level of genetic diversity among the soybean germplasm in the breeding programme in Zimbabwe.
2. To evaluate the level of genetic gains that has been made by the soybean breeding programme in Zimbabwe from 1940 to 2013.
3. To evaluate the adaptability and yield stability of soybean varieties in contrasting environments in Zimbabwe, Malawi and Zambia.
4. To identify the traits which have contributed significantly, and directly and indirectly to the high grain yield potential that has been realized over 70 years of soybean breeding in Zimbabwe.

8 Research Hypotheses

Below is a list of hypotheses that were tested;

1. The soybean germplasm available to the programme in Zimbabwe exhibits wide genetic variability hence further genetic gains can be realised.
2. The soybean programme in Zimbabwe has realized huge genetic gains over 70 years of breeding and selection (from 1940 to 2013).
3. Genetic improvements of the soybean germplasm over the past 73 years have enhanced adaptability and stability of the varieties in many production environments in Zimbabwe, Zambia and Malawi.
4. Soybean grain yield gains have been realized through direct contributions of the yield components such as 100 seed weight, number of pods per plant, number of pods per plant, plant height and number of branches and breeding emphasis on these traits will spur future gains.

9 Thesis Structure

The structure and outline of the thesis is as given below;

Chapter 1: Review of Literature

This chapter reviewed literature on the genetic diversity, genetic gains, relationships between soybean grain yield and secondary traits. Genotype by environment interaction was also reviewed and discussed. It also focused on the combining ability effects and their implications on breeding.

Chapter 2: Diversity of the soybean germplasm in the breeding programme in Zimbabwe

This chapter covered phenotypic and molecular characterisation aspects.

Chapter 3: Genotype by environment interaction and grain yield stability of the Zimbabwean soybean elite lines across 13 environments.

This chapter assessed the influence of GEI on grain yield and stability analysis

Chapter 4: Breeding gains, variability and heritability of soybean genotypes for the Zimbabwean soybean programme

This chapter focused on estimating the breeding progress, variance components, heritability and genetic advance of the available germplasm

Chapter 5: Correlation and path coefficient analyses of secondary traits on soybean.

The focus was mainly on relationships between grain yield and related components

Chapter 6: General overview and future directions.
The chapter summarised the findings of the study and their implications to breeding

Each chapter was presented separately. However, there is some overlap and repetition of both information and references between the chapters as they were written as independent journal papers covering the ideal information.

10 References

Aditya, J. P., 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.

Board, J. E., S. Kumudini, J. Omielan, E. Prior and C. S. Kahlon. 2010. Yield response of soybean to partial and total defoliation during the seed-filling period. Crop Science 50:703-712.

Board, J. E. and H. Modali. 2005. Dry matter accumulation predictors for optimal yield in soybean. Crop Science 45:1790-1799.

Bonato, A. L. V., Calvo, E.S., Geraldi, I. O., and C. A. A. Arias. 2006. Genetic similarity among soybean [(*Glycine max* (L) Merrill] cultivars released in Brazil using AFLP markers. Genetics and Molecular Biology 29: 692-704.

CGIAR, 2012. Leveraging legumes to combat poverty, hunger, malnutrition and environmental degradation. http://hdl.handle.net/10947/5316

Cho, Y. and R. A. Scott. 2000. Combining ability of seed vigour and seed yield in soybean. Euphytica 112:145-150.

Cregan, P. B. and R. W. Yaklich. 1986. Dry matter and nitrogen accumulation and partitioning in selected soybean genotypes of different derivation. Theoretical and Applied Genetics 72:782-786.

Crossa, J. 1990. Statistical analysis of multi-location trials. Advances in Agronomy 44:55-85.

Duvick, D. N. 2005. Genetic progress in yield of United States Maize (*Zea mays* L.). Maydica 50:193-202.

Egli, D. B. 2008a. Comparison of corn and soybean yields in the United States: Historical trends and future prospects. Agronomy Journal 100:79-88.

Egli, D. B. 2008b. Soybean yield trends from 1972 to 2003 in mid-western USA. Field Crops Research 106:53-59.

Esterhuizen, D. 2011. Assessments of commodity and trade issues.

Foulkes, M. J., M. P. Reynolds and B. R. Sylvester. 2009. Genetic improvement of grain crops: yield potential. In *Crop Physiology: Applications for Genetic Improvement and Agronomy* ELsevier, Burlington, MA, United Kingdom

Fox, P. N., J. Cross and I. Ramagosa. 1997. Multiple environment testing and genotype x environment interaction. In R. A. Kempton and P. N. Fox (eds.) Statistical Methods for Plant Variety Evaluation. Chapman and Hall, London.

Gadde. 2006. Genetic investigations in soybean (Glycine max (L.) Merrill), Dharwad Agricultural Sciences University. p. 1-75.

Hyten, D. L. 2005. Genetic diversity and linkage disequilibrium in wild soybean, landraces, ancestral and elite soybean populations. Department of Natural Resource Sciences and Landscape Architecture, University of Maryland. p.1-88.

IITA, 2008. 30 years of R4D in soybean: What's next?

Iqbal, Z., M. Arshad, M. Ashraf, T. Mahmood and A. Waheed. 2008. Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. Pakstan Journal of Botany 40:2323-2328.

Jonas, C. N., B. Vanlauwe, J. M. Mahasi, E. Katungi, C. Akech, F. S. Mairura, C. N. Justina, and and N. Sanginga. 2008. Soybean situation and outlook analysis: the case of kenya. .

Kaga, A., T. Shimizu, S. Watanabe, Y. Tsubokura, Y. Katayose, K. Harada, D. A. Vaughan and A. Tomooka. 2012. Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. Breeding Science Journal 61:566-592.

Kang, M. S. 2004. Breeding: Genotype-by-environment interaction. In: R.M. Goodman (ed.) Encyclopedia of Plant and Crop Science. Marcel-Dekker, New York, United States

Kapuya, T., D. Saruchera, A. Jongwe, T. Mucheri, T. Mujeyi, L. N. Traub and F. Meyer, 2010.The grain industry value chain in Zimbabwe.

Kashyapi, A., A. P. Hage and D. A. Kulkarni. 2012. Impact of climate change on world agriculture : A Review; Workshop Proceedings. Agricultural Meteorological Division, India. p.10-25.

LAES. 2009. Louisiana soybean and feed grain review. Research and Extension. Volume II. United States.p. 1-9.

Lange, C. E. and L. C. Federerizzi. 2009. Estimation of soybean genetic progress in the south of Brazil using multienvironmental yield trials. Science Agriculture 66:309-316.

Lark, K. G., K. Chase, F. Alder, L. M. Mansur and J. H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proceedings of the National Academic of Science 92:4656-4660.

Machikowa, T., C. Saetang and K. Funpeng. 2011. General and Specific Combining Ability for Quantitative Characters in Sunflower. Journal of Agricultural Science 3:91-95.

Mikel, M. A., B. W. Diers, R. L. Nelson and H. H. Smith. 2010. Genetic Diversity and Agronomic Improvement of North American Soybean Germplasm. Crop Science 50:1219-1229.

Mirsky, S. B., M. R. Ryan, J. R. Teasdale, W. S. Curran, C. S. Reberg-Horton, J. T. Spargo, M. S. Wells, C. L. Keene and J. W. Moyer. 2013. Overcoming Weed Management Challenges in Cover Crop-Based Organic Rotational No-Till Soybean Production in the Eastern United States. Weed Science Society of America 27:193-203.

Mohammadi, S. A. and B. M. Prasanna. 2003. Analysis of Genetic Diversity in Crop Plants-Salient Statistical Tools and Considerations. Crop Science 43:1435-1248.

Mueller, T. A., M. R. Miles, W. Morel, J. J. Marois, D. L. Wright, R. C. Kemerait, C. Levy and G. L. Hartman. 2009. Effect of Fungicide and Timing of Application on Soybean Rust Severity and Yield. American Phytopathological Society 93:243-248.

Mugandani, R., M. Wuta, A. Makarau and B. Chipindu. 2012. Reclassification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. African Crop Science Journal 20:361-369.

NAMC, 2011. The South African Soybean Value Chain by the Markets and Economic Research Centre of the NAMC: Strategic positioning of South African Agriculture in dynamic global markets. 1-94

Nassar, M. A. A. 2013. Heterosis and Combining Ability for Yield and its Components in Some Crosses of Soybean. Australian Journal of Basic and Applied Sciences 7:566-572.

Ojo, D. K., A. O. Ajayi and O. A. Oduwaye. 2012. Genetic Relationships among Soybean Accessions Based on Morphological and RAPDs Techniques. Pertanika Journal of Tropical Agricultural Science. 35:237-248.

Oliveira, M. F., R. L. Nelson, I. O. Geraldi, C. D. Cruz and J. F. F. Toledo. 2010. Establishing a soybean germplasm core collection. Field Crops Research 119:277-289.

Opperman, C., 2011.Technical Report: Soybean Value Chain. Southern Africa Trade Hub, USAID. p. 7-12.

Orf, J. H., K. Chase, F. R. Adler, L. M. Mansur and K. G. Lark. 1999. Genetics of Soybean Agronomic Traits: II. Interactions between Yield Quantitative Trait Loci in Soybean. Crop Science 39:1652-1657.

Pedersen, P. and J. G. Lauer. 2004. Response of soybean yield components to management system and planting date. Agronomy Journal 96:1372-1381.

Reyna, N. and C. H. Sneller. 2001. Evaluation of Marker Assisted Introgression of Yield QTL Alleles into Adapted Soybean. Crop Science 41:1317-1321.

Rowntree, S. C., J. J. Suhre, N. H. Weidenbenner, E. W. Wilson, V. M. Davis, S. L. Naeve, S. N. Casteel, B. W. Diers, P. D. Esker, J. E. Specht and S. P. Conley. 2013. Genetic Gain x Management Interactions in Soybean: I. Planting Date. Crop Science 53:1128-1138.

Seed Co, 2008. Annual report. p. 38-50.

Shibles, R. M. and C. R. Weber. 1966. Interception of solar radiation and dry matter production by various soybean planting patterns. Crop Science 6:55-59.

Shurtleff, W. and A. Aoyagi, 2007. History of Soybeans and Soyfoods in Africa, Soyinfo Centre, Lafeyette Publishers, USA. Part 1.

Specht, J. E., D. J. Hume and S. V. Kumudini. 1999. Soybean Yield Potential-A Genetic and Physiological Perspective. Crop Science 39:1560-1570.

TechnoServe, 2010. Southern Africa soya roadmap-Zimbabwe value chain analysis. www.technoserve.org/files/downloads/technoserve-bmgf-zimbababwe.pdf

Weber, C. R., R. M. Shibles and D. E. Byth. 1966. Effect of plant population and row spacing on soybean development and production. Agronomy Journal 58:99-102.

CHAPTER 1
LITERATURE REVIEW

1.0 Introduction

The objective of this chapter is to critically review literature in the context of the objectives of the research study. The review was also undertaken with a view to providing a reference point for the study. The chapter reviewed literature on the genetic diversity, genetic gains, and relationships between soybean grain yield and its secondary traits. Genotype x environment interaction was also reviewed and discussed. It also focused on the combining ability effects and their implications on breeding. The chapter starts by discussing the soybean botany, domestication and dissemination.

1.1 Soybean botany, domestication and dissemination

Soybean is a self-pollinating crop thought to have originated from north eastern China during or before the Shang dynasty 1700-1100 BC (Fukuda, 1993; Hymowitz, 2004). To this end, China is regarded as the centre of origin. It is classified in the Leguminosae family and falls under the genus, *Glycine* and species *max,* the cultivated soybean similar to its wild progenitor, *Glycine soja,* an annual plant native to north eastern Asia (Doyle *et al.*, 2004). Both *Glycine max* and *Glycine soja* are diploid and carry the chromosome number 2n=40. The subgenus *Glycine* consists of seven different species of the wild soybean, while the cultivated subgenus *Glycine soja* has only two species. The wild species are *G*. *clandestine*, *G. sericea; G. falcate, G.latifolia, G. latrobeana, G. canesscens, G. tabacina* and *G. tomentalla,* while the cultivated species are *G. soja and G. max.* In terms of distribution, the first five wild species are only found in Australia. The two cultivated species are cross fertile; however, *Glycine soja has* several undesirable traits making it unsuitable for cultivation by man. Hence, it is only useful as a donor parent for genes of interest to the locally adapted or recurrent parent (Hymowitz and Newell, 1981).

The domestication of soybean is reported to have begun in central, north east, south China and Peninsular Korea during the first century (Hymowitz and Newell, 1981). Later on, it spread across the provinces of China and Korea. The movement of soybean within the centre of origin was attributed to the development, consolidation of territories and degeneration of Chinese dynasties. For instance, when the Shanrong people were attacked by Lord Huan of the Qi state, he took soybean grain from them and introduced into the Central Plains and substantially changed the eating habits of the communities around the Yellow River. Later on soybean became the staple food along with millet and this took place

during the Warring States Period. Hymowitz (1990) postulated that soybean disseminated throughout Asia and landraces were developed in Japan, Thailand, north India, Burma, Vietnam, Indonesia, Philippines and Malaysia between the first century and 16th century. Further distribution to many countries is ascribed to the establishment of the sea and land trading, emigration of certain tribes from China and its acceptance as a food crop. It reached Europe and America in the late 1700s. It was reported to be cultivated in Illinois in 1851 and was subsequently distributed throughout the US Corn Belt. However, it only became popular as a grain crop in the 1920s (Hymowitz, 1990). Crop improvement was also initiated in the 1920s in US and several cultivars have been released for commercial production. Initially, it was grown as a forage crop in addition to green manuring meant to improve soil fertility. Soybean was introduced into Zimbabwe early in the 19th century primarily to improve soil fertility through green manure and as a forage crop.

Soybean is adapted to a wide range of climatic conditions and does well in areas of good rainfall (evenly distributed) or where irrigation is available. As a general rule, soybean grows well in the same areas where maize does well. Deep and well drained soils are recommended, varying in texture from sandy loam to clay loam (Tattersfield *et al.*, 1988). Heavy clays are also suitable provided that they are well drained and soil capping does not impede germination. Soybean is very sensitive to soil acidity and for maximum yields the soil pH ($CaCl_2$) should range between 5.3 to 5.5.

1.2 Soybean genetic diversity

Vaughan *et al.* (2007) defined genetic diversity as any variation in the nucleotides, genes, chromosomes, or whole genomes of organisms. The genetic diversity of any crop is essential to maximize on genetic improvement which is accomplished through hybridisation, mutagenesis or any biotechnological means (Mutengwa, 2004; Satyavathi *et al.*, 2006). The availability of genetic diversity enables breeders to select quality parents with respect to traits of interest for making combinations. Erasmus (2008) also reported that a breeding programme with a broad genetic base provides a valuable source of genes required for introgression purposes. This implies that breeders with a diverse germplasm can easily work out crossing plans where certain genes are introgressed into locally adapted varieties. Use of distant parents in making combinations further broadens the diversity and promotes increases in genetic gains.

China holds the largest soybean germplasm collections with an estimated total of 26 000 accessions for *Glycine max* and 6 200 accessions of *Glycine soja* (Oliveira *et al.*, 2010). It is

followed by the US which holds 16 999 accessions for *Glycine max*, 1116 accessions of *Glycine soja* and 919 accessions of the perennial *Glycine* species. Despite, huge germplasm accessions in various gene-banks, soybean has been described as one of the least diverse crops. The annual *Glycine* species lack diversity. Reports are available that give estimates of nucleotide diversity of soybean relative to the wild soybean, *Glycine soja* and other crop species. Tables 1.1 and 1.2 present the available data on nucleotide diversity of soybean versus other crops.

Table 1.1: Molecular genetic diversity values reported in studies comparing cultivated and wild soybean genotypes and the ratio of genetic diversity values in *Glycine max* versus *Glycine soja* as an estimate of diversity retained through the genetic bottleneck of domestication.

	Genetic diversity estimate		Proportion of diversity Retained after domestication
Data Source	*G. max*	*G. soja*	*G. max/G. soja*
Powell *et al.* (1996)	0.538	0.830	0.650
Li and Nelson (2002)	0.400	0.460	0.870
Xu and Gai (2003)	0.188	0.285	0.660
Hyten *et al.* (2006)	0.00115	0.00235	0.490
Kuroda *et al.* (2009)	0.496	0.87	0.570
Mean			0.650

Source:(Stacey, 2008)

Table 1.2: Reported estimates of nucleotide diversity of various species.

Species	$\Theta_w(x\ 10^{-3})$
Glycine max Zhu *et al.* (2003)	0.86
Glycine soja Scallon *et al.* (1987)	2.4
Zea mays spp Parviglumis Tenaillon *et al.* (2004)	13.4
Modern US. *Zea mays* inbred lines Tenaillon *et al.* (2004)	6.2
Maize landraces and inbreds Tenaillon *et al.* (2001)	9.6
Hordeum vulgare ssp. Spontaneum Kanazin *et al.* (2002)	9.8
Sorghum bicolor (Hamblin *et al.*, 2004)	2.3
Oryza Sativa (Feltus *et al.*, 2004)	1.81
Source: Source: Stacey (2008)	

The findings shown in the above tables clearly reveal that the level of DNA sequence variation in the cultivated soybean is lower compared to other species. Generally, humans have been found to have lower levels of nucleotide diversity compared to many plant species (Hyten *et al.*, 2006). Halushka *et al.* (1999) estimated nucleotide diversity in humans and found average nucleotide diversity of Θ = 0.00083 which is more or less similar to

soybean. Although the diversity of soybean is naturally low, further reduction is due to the processes of domestication, founder population effects and intensive selection during breeding.

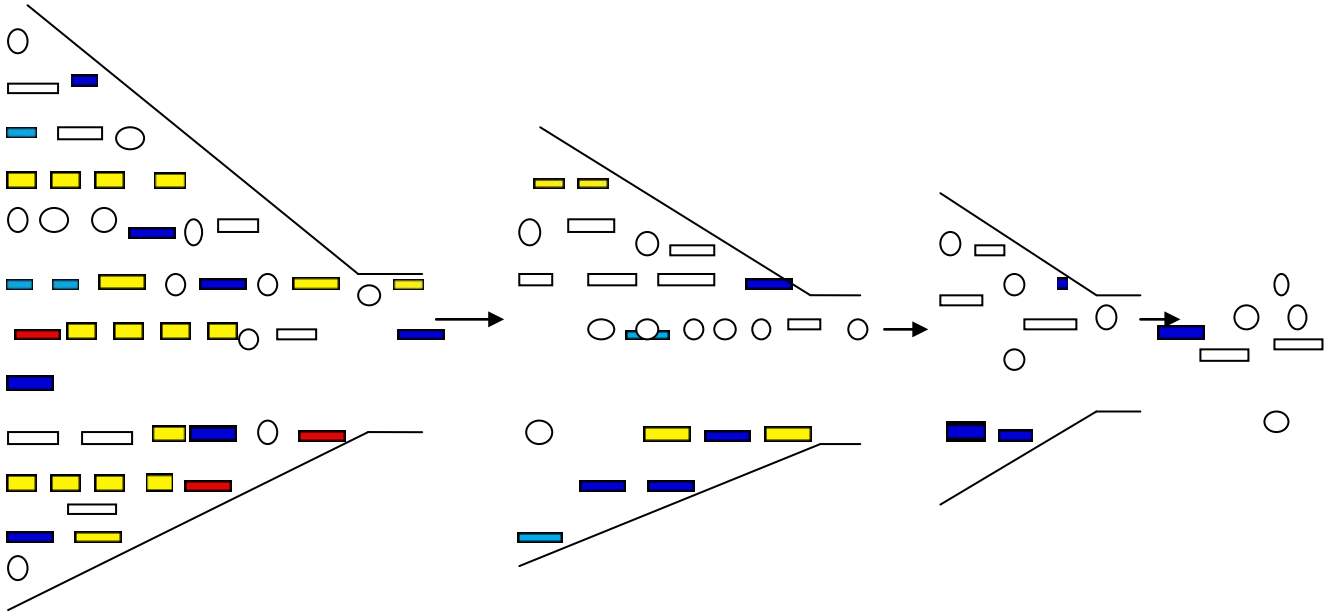

Figure 1.1: Genetic bottlenecks imposed on crop plants during domestication, by founder effects and through modern breeding practices.

N.B. Boxes represent allelic variations of genes originally found in the wild, but gradually lost through domestication, founder effects and breeding. Source: Tanksley and McCouch (1997)

The diagram above indicates that the wild *Glycine* species has the widest diversity, followed by early domesticates, which is succeeded by ancestors selected by breeders and lastly modern cultivars. Hence, movement from one level to another represents considerable loss in diversity. Although the domestication and founder population effects have contributed to increased productivity, they have simultaneously contributed to narrow diversity as reflected in the above diagram. Domestication was a long process that favored highly adapted lines with important agronomic traits. It entailed both positive and negative selection. For instance, lines that were shattering badly were selected against. Over time the approach reduced the number of rare alleles leading to genome-wide reduction of genetic diversity (Guo *et al.*, 2010). Founder effect refers to the loss of genetic variation arising from the new population established by few individuals originating from a larger population (Provine, 2004). This means that the few individuals will not be genetically representative of the parental

population and therefore, represent low genetic variation. According to Hyten *et al.* (2006), founding events occur when a few cultivars are introduced to a new production environment or used to create new populations. This has the same effect of narrowing variability. Both the domestication and founder effects contribute to genetic bottlenecks that alter the allele frequencies, eliminate rare alleles and diminish genetic diversity depending on the selection pressures and duration of these evolutionary events. On the other hand, both evolutionary events portray their value in crop evolution as well as the breeding prospects of the genetic variation embedded in the wild species (Ladizinsky, 1985).

1.3 Plant breeding and soybean genetic diversity

Plant breeding has made enormous impact in crop improvement and the global agricultural industry at large. Thus, breeders have continuously added value to the sector by introducing higher yielding cultivars (van de Wouw *et al.*, 2010a). Higher productivity translated to higher production levels, consequently making a significant contribution to fuel, feed and food security (Evenson and Gollin, 2004).

Despite its positive gains, plant breeding has been labeled as another evolutionary force that has contributed to reduction in genetic diversity. The bone of contention is centred on the fact that repeated use of parental stock of genetically related cultivars has decreased the genetic base of soybean. Criteria for parental stock selection is usually based on agronomic value and in the majority of cases, lines with such agronomic merit are recycled. This approach narrows the diversity and increases the frequency of fixed common alleles (Feng *et al.*, 2009). In order to understand the impact of plant breeding on soybean genetic diversity, it may be worth to revisit the history of soybean breeding. Amazingly, the pioneer breeders had an extensive plant pathology background (Hyten, 2005). As a result they focused more on breeding for disease resistance with the understanding that disease resistance would markedly enhance seed yield. This narrowed the diversity scope because resistance to given diseases usually traces to a single or few sources. Moreover, they also adopted the backcrossing breeding strategy where resistant varieties were crossed to a few agronomically superior cultivars or elite lines (Carter *et al.*, 2004). The few parents constituted a limited number of the founding ancestors. In addition, traditional soybean breeders have been reported to have made crosses between high-yielding parents and locally adapted varieties or elite experimental lines to generate high yielding varieties (Reyna and Sneller, 2001). In a similar regard, Magorokosho (2006) also asserted that annually, selection of parental lines is limited to a small group of individuals which is viewed to be possessing unique attributes. For example, in America a few outstanding cultivars were

overused. Classical examples include “Lincoln,” “Lee”, “Williams”, “Essex,” and “A3127.” Gizlice *et al.* (1994) observed that Lincoln accounted for the parentage of 18% of the cultivars that were bred between 1944 and 1988 because it was the highest yielding cultivar then. In the same vein, A3127 registered record yields when it entered the market in 1977 and was extensively used in many combinations (Sneller, 1994). Williams and Essex were again block-busters and they similarly contributed to the parentage of numerous cultivars in both North and South America. This practice resulted in only 80 cultivars accounting for 99% of the US parentage (Carter *et al.*, 2004).

The Zimbabwean soybean breeding programme has been in force for seven decades and given the American scenario, it deserves the investigation of diversity. The loss of diversity is a process and happens over a long period of plant breeding. Reif *et al.* (2005) postulated that extended breeding coupled with intensive selection leads to reduced diversity. The decrease in diversity results in effects such as genetic vulnerability, genetic erosion and yield plateaus and these are discussed below.

Soybean breeders have generated several elite breeding lines and cultivars. Because soybean is a self-pollinating crop, it implies that all the lines created, including the commercialized, will be genetically uniform. Given the practice where a few lines are used in making combinations and are thought to trace to common ancestors known to produce superior progenies, it then exposes the commercial products to attack by pests and diseases (van de Wouw *et al.*, 2010a). The topical issue here is that, varieties that trace to similar ancestors succumb to similar pests and diseases. The most well documented tragic case caused by genetic vulnerability is the Irish potato famine of the 1840s where more than one million Irish people starved to death due to yield loss caused by massive attack by late blight *(Phytophthora infestans)* (Xu, 2009). Other widely known examples are; the coffee rust *(Hemileia vastatrix)* epidemic in 1868 and Southern Corn Leaf Blight *(Helminthosporium maydis)* epidemic in United States in 1970. Ideally, the principal cause of such catastrophes stems from narrow genetic base.

Numerous studies have been carried out on genetic vulnerability. Duvick (1984) contended that genetic vulnerability is ascribed to few cultivars sharing the same genetic background available in the production systems. This is in agreement with the findings of Cooper and Hodgkin (2001) who reported that genetic vulnerability could be caused by growing single varieties to huge hectarage or varieties possessing the same resistant genes. Their results revealed that six cultivars of soybean planted 56% of soybean area, six cultivars of wheat accounted for 41% of the wheat area, six cultivars of cotton covered 68% of the total area

and six inbred lines of maize constituted more than 40% of the hybrids in United States. Cregan *et al.* (2006) concluded that resistance to a particular disease is in most cases conferred by a single source or few sources. The implication is that the "would be cultivars" share a common genetic background and present a high probability of being attacked by any mutant pathogenic strain.

The domination of the market by few varieties, which can be stressed back to common ancestors, threatens food security because of climate change. Fujisaka *et al.* (2011) pointed out that climate change could be viewed in terms of increases or decreases in temperature, increased frequency of droughts and floods and increases or decreases in precipitation (rainfall). The key issue is that, the said changes will have impact on adaptation of the crop cultivars. Kucharik and Serbin (2008) contended that climate change negatively affects productivity and phenological development of crop cultivars culminating in yield losses. The unpredictable and altered weather patterns have a high probability of exposing the crop to genetic vulnerability to pests, diseases and short duration seasons. For example, global warming can result in both new abiotic and biotic stresses in the business market domains. Lobell and Asner (2003) asserted that in a changing climate, pests and diseases mutate and become more active. This threatens breeding programs that have narrow diversity.

Another challenge that is associated with narrow genetic base is genetic erosion. Hammer and Laghetti. (2005) defined genetic erosion as loss of single genes, certain gene combinations or locally adapted landraces of plants or animals. It is possible that the introduction of foreign germplasm may cause displacement of the local varieties and populations resulting in reduced genetic variability. In addition, the adoption of modern varieties that are uniform and superior to old varieties may also result in reduction of genetic variability. The effect of plant breeding in the context of genetic erosion has been demonstrated (Ploetz, 2006; Singh *et al.*, 2006b; White *et al.*, 2008; van de Wouw *et al.*, 2010b). Reports of genetic erosion have been made for wheat in Italy to which an annual loss of 13.2% from 1920 to 1960 and 4% from 1980 to date was reported (Hammer and Laghetti., 2005). This observation showed huge loss of valuable diversity possibly defined in terms of quality traits and resistance or tolerance to both biotic and abiotic stresses. Putting together of combinations that involve genetically similar cultivars decreases genetic diversity, giving rise to genetic erosion. Elimination of rare alleles in the resulting population also contributes to narrowing of diversity. Climate change has a potential to cause genetic erosion of the unadaptable cultivars, landraces as well as wild species (Jarvis *et al.*, 2003).

Another negative effect of narrow genetic base is that breeders quickly get to a yield plateau. Fehr (1987)'s findings on pedigree analysis revealed that of the 136 varieties cultivated in United States, only five introductions were the cytoplasmic source for 121 cultivars and just ten accessions contributed 88% of their genome. These results concurred with investigations of (Gizlice *et al.*, 1994). In another assessment, Cober *et al.* (2005) demonstrated seed yield plateau for soybean cultivars released between 1934 to 1996 in United States (US). In the same vein, Gadde (2006) reported a constant rate of soybean yield increase in India. The observations reported herein negatively impact on subsequent yield improvement efforts.

Thus, the assessment of genetic diversity is fundamental to crop improvement. As such, there are several methods can be applied to estimate diversity and these include pedigree information, phenotypic, biochemical, and molecular characterisation. The method used determines the precision used. In this case, the methods chosen for this study are phenotypic and molecular. This is because they are the ones that commonly employed in diversity studies and the discussion will only focus on these.

1.4 Methods used to estimate genetic diversity

1.4.1 Genetic diversity based on phenotypic traits

This is the oldest and most exploited method. Its advantage is that it provides a simple, direct rapid and inexpensive way of characterising varieties (Mutengwa, 2004). Phenotypic characterisation is regarded as the classical approach and is viewed as the best determinant of the agronomic value and taxonomic classification of crop plants (Cholastova and Knotova, 2012). Apart from diversity studies, phenotypic characterisation is applied in the development, production and marketing of varieties. In markets where Plant Breeder's Rights exist, the registration of a new cultivar (s) occurs only if is distinct from other cultivars. Hence, different cultivars could be identified based on phenotypic descriptors (Govindarao, 2010).

Phenotypic characters that are commonly used to assess genetic variability in soybean include seedling, plant, phenological, seed and quality morphological characteristics. Numerous studies have explored the significance of phenotypic characterisation in estimating genetic diversity in soybean (Chen and Nelson, 2004; Dayaman *et al.*, 2009; Liu *et al.*, 2011; Malik *et al.*, 2011; Matsuo *et al.*, 2011; Salimi *et al.*, 2012). Hamzekhanlu *et al.* (2011) studied 34 mutant lines including one control cultivar and detected variability for number of leaves per plant, number of grains per plant, number of pods per plant, plant dry

weight (shoot dry weight), root dry weight, harvest index, number of nodules per plant, nodule dry weight, 100 seed weight and seed yield per plant. Further, the genotypes were clustered into four groups.

Manjaya and Bapat (2008) observed genetic variation during the characterisation of 55 soybean varieties using phenotypic traits viz; days to 50% flowering, days to maturity, plant height and number of branches per plant, number of pods per plant, number of seeds per plant, 100 seed weight and yield per plant. Using 52 morphological and agronomic characters, Antalikova *et al.* (2008) found variability for the traits measured on the 52 studied genotypes. In another study involving phenotypic characterisation of 139 soybean genotypes Iqbal *et al.* (2008), revealed significant differences among all the assessed traits.

Ravikumar (1999) characterised soybean genotypes and reported variability when he used seedling morphological traits. Tarasatyavathi *et al.* (2004) demonstrated that soybean can be characterised on the basis of leaf shape, leaf colour intensity, flower colour, pod pubescence, plant height and days to maturity. A study was also conducted to evaluate genetic diversity on 92 soybean genotypes originating from Asian Vegetable Research and Development Centre (AVRDC), United States and Pakistan and high CVs (coefficient of variation) coupled with wide ranges were obtained on pods per plant (29.5%), leaf area (44.8%), number of branches per plant (31.7%), 100 seed weight (39.0%) and grain yield per plant (46.6%) (Malik *et al.*, 2011). Wide ranges symbolized high level of diversity. Interestingly, the genotypes were classified into three distinct groups with the Pakistan germplasm forming its own cluster. In contrast to these findings, Ojo *et al.* (2012) performed a similar study and found seven clusters from 42 genotypes studied. They also revealed that the number of pods per plant, pod yield per plant, 100 seed weight and seed yield per plot accounted for the greatest phenotypic variation. These results implied broad diversity.

Although the phenotypic traits are influenced by the environment, the observations reported in various studies above indicate their usefulness and value in genetic diversity studies and crop improvement. Selection on the basis of phenotypic traits is still widely practiced and will continue to play a significant role in estimating diversity under the auspices of the application of ANOVA in crop species and their relatives. Results exhibiting high CVs and significant differences present high scope for selection. Furthermore, the clustering patterns obtained from phenotypic data, in respect of the number of clusters generated and genotypes contained in a cluster help to show diversity and the relatedness.

1.4.2 Genetic diversity based on molecular markers

The advancement of science has led to the discovery of molecular or DNA markers which among other applications are used to characterise genetic diversity in the germplasm pool for crop species. According to Xu and Gai (2003), genetic markers are biological features that are determined by allelic forms and can be used as experimental probes or tags to keep track of an individual, a tissue, cell, nucleus, chromosome or gene. Collard *et al.* (2005) defined genetic markers as representatives of genetic differences between individual organisms or species. Markers reveal sites of variation in DNA, in other words, they simply detect differences in genetic information carried by two or more individuals. Molecular diversity studies in soybean are primarily conducted for the purposes of; investigating phylogenic and evolutionary relationships of soybean and its relatives, analyzing diversity trends over time, baseline surveys to assess diversity in a given context, historical understanding of soybean in a particular area, analyzing the structure of diversity and formulation of germplasm maintenance and conservation strategies, varietal maintenance and relating diversity to agronomic performance (Moose and Mumm, 2008). A full understanding of the genetic relationships among the germplasm studied acts as a guide for parental selection in respect of population improvement (Moose and Mumm, 2008; Duran *et al.*, 2009; Kumar *et al.*, 2009).

There are numerous DNA based markers and the most common ones are; simple sequence repeats (SSR), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD) and single nucleotide polymorphism (SNP). According to Magorokosho (2006), the DNA markers can be categorized into three groups namely, (i) the hybridization based markers, i.e. the RFLPs, (ii) polymerase chain reaction based on DNA amplification i.e. RAPD, AFLP, SSR and (iii) single nucleotide polymorphisms (SNPs). Table 1.3 below shows the advantages and disadvantages of the five widely used DNA markers in plants.

Table 1.3: Comparison of the five widely used DNA markers in plants

Attribute	RFLP	RAPD	AFLP	SSR	SNP
Genomic coverage	Low copy coding region	Whole genome	Whole genome	Whole genome	Whole genome
Amount of DNA required	50-10ug	1-100ng	1-100ng	50-120ng	>120ng
Quality of DNA required	High	Low	High	Medium high	High
Type of polymorphism	Single base changes, indils	Single base changes, indils	Single base changes, indils	Changes in length of repeats	Single base changes, indils
Level of polymorphism	Medium	High	High	High	High
Effective multiplex ratio	Low	Medium	High	High	Medium to High
Inheritance	Co-dominant	Dominant	Dominant/ Co-dominant	Co-dominant	Co-dominant
Type of probes/primers	Low copy DNA or cDNA clones	Usually 10bp random nucleotides	Specific sequence	Specific sequence	Allele-specific PCR primers
Technically demanding	High	Low	Medium	Low	High
Radioactive detection	Usually yes	No	Usually yes	Usually no	No
Reproducibility	High	Low to Medium	High	High	High
Time Demanding	High	Low	Medium	Low	Low
Automation	Low	Medium	High	High	High
Development/startup cost	High	Low	Medium	High	High
Proprietary rights required	No	Yes and licensed	Yes and licensed	Yes and some licensed	Yes and some licensed
Suitable utility in diversity, genetics and breeding	Genetics	Diversity	Diversity and genetics	All purposes	All purposes

Source: Xu (2009).

Earlier genetic diversity studies exploited the RFLP markers. Interestingly, the first soybean genetic map was developed using the RFLP markers (Yang *et al.*, 2008). Of late, the PCR markers have taken over. However, the RFLP are still being used though on a limited scale. Keim *et al.* (1992) screened 38 soybean lines using 128 RFLP markers and observed that 69% of the 132 probes detected variation among the lines. They also found the average polymorphism index to be 0.30. However, cluster analysis and principal coordinate analysis showed lack of diversity among the cultivated lines. Skorupska *et al.* (1993) investigated genetic relationships between 108 soybean genotypes composed of breeding lines, elite cultivars as well as older cultivars using 83 RFLP probes and observed low levels of molecular diversity coupled with a gene diversity of $\geq$0.30. These results concurred with the findings of Lorenzen *et al.* (1995) who obtained a gene diversity of $\geq$0.30 among 64 soybean ancestors and milestone cultivars released in Southern United States using 217 RFLP markers.

Ude *et al.* (2003) demonstrated the usefulness of AFLP markers in soybean genetic studies when they assessed genetic diversity of 35 North America's soybean ancestors, 66 high yielding North America's soybean cultivars, 59 modern Chinese cultivars and 30 modern Japanese cultivars using five AFLP primer pairs. They obtained polymorphism information content (PIC) that ranged from 0 to 0.5. The average genetic distances between the Japanese cultivars, North America's soybean cultivars, North American soybean ancestors and Chinese cultivars were 6.3, 7.3, 7.5 and 8.5 respectively. Using 16 AFLP markers on 38 genotypes from US, Feng *et al.* (2009) concluded that the variability among the test genotypes was narrow. They found an average genetic distance of 0.124 among the 38 genotypes and recommended that the breeding programme should not use parents of the same genetic background. In a similar regard, Maughan *et al.* (1996) evaluated the diversity between *Glycine max* and *Glycine soja* using the AFLP markers and their results showed that the diversity values were greater in wild soybean compared to the cultivated soybean. The AFLP data was subjected to cluster and principal component analysis and the output revealed marked differences in the clustering patterns. They observed that the adapted soybean cultivars were tightly clustered together signifying relatively narrow genetic variation. Satyavathi *et al.* (2006) investigated the genetic diversity among 72 cultivars in India using 12 AFLP and concluded that there was a need to introduce diverse accessions from outside the country because the soybean cultivars were assigned to four clusters only, indicative of low levels of diversity.

Using 115 random amplified polymorphic DNA markers on 120 accessions from China, Japan and South Korea, Li and Nelson. (2001) observed that accessions from China formed their own cluster whereas accessions from Japan and South Korea were clustered together. Their results also showed that the genetic distances between genotypes ranged from 0.14 to 0.55 yielding a mean of 0.42. They also found the polymorphism information content (PIC) to range from 0.03 to 0.42 with an average of 0.32 and a standard deviation of 0.14. The study showed genetic similarity between the material from Japan and South Korea, probably because the material from Japan originated from Korea since historically soybean was first reported to be introduced to South Korea then Japan (Kihara, 1969). Lack of similarity between the material from China and those from both Japan and South Korea could imply that the germplasm used in the study did not originate from China. In another study involving 105 genotypes, Guedira *et al.* (2000) reported broad diversity among the sampled genotypes using 109 RAPD and SSR markers. The genotypes were allocated to 11 clusters and genetic distances ranged from 0.08 to 0.76 with an average of 0.52. The later study agreed with the findings of Thompson *et al.* (1998) who reported genetic diversity on 35 genotypes using 281 RAPD markers. The average genetic distance was 0.56. The two investigations support evidence of genetic diversity between the major ancestors of North American germplasm and plant introductions. Ojo *et al.* (2012) characterised 40 soybean accessions in Nigeria using ten RAPD primers and concluded that there was variability among the genotypes.

Single nucleotide polymorphisms (SNP) which have a capacity to genotype hundreds to thousands of SNPs in a single reaction can also be applied in genetic diversity studies. Using 496 SNPs, Hyten *et al*. (2006) estimated total nucleotide diversity in the context of expected heterozygosity per nucleotide site and number of polymorphic sites in the a given sample of soybean and reported higher total nucleotide diversity estimates in *Glycine soja* (2.2, 2.4), followed by landraces (1.4, 1.2) and North American ancestors and elite cultivars (1.1, 0.8). Assessment of the entire soybean genome revealed lower genetic diversity in the cultivated species relative to the *Glycine soja* (Akpertey, 2013). These results confirm changes in allele frequency through genetic bottlenecks.

The SSR markers have also been widely used in soybean genetic diversity studies. In Brazil, in particular at Embrapa Research Institute, Mulato *et al.* (2010) demonstrated the application of SSR markers in assessing genetic relationships between soybean cultivars. Using 20 SSR markers and ten EST SSR markers on 79 soybean accessions their results revealed high levels

of genetic diversity among the material examined. They obtained a total of 259 alleles, ranging from 2 to 21 alleles per locus. More importantly, the average was 8.63 alleles and the genotypes were assigned to five major clusters and numerous subgroups. Wang *et al.* (2006) assessed the genetic diversity among 129 accessions from the Chinese core collection using 60 SSR markers and concluded that the material was quite divergent. They observed a total of 732 alleles, PIC varied from 0.05 to 0.91 with a mean of 0.23 and the accessions were separated into five major clusters according to geographical origins (i.e. the Yellow River Valley ecotypes, two clusters from the Northern ecotypes, South ecotypes and a mixture of the Yellow River Valley ecotypes and Northern). Interestingly, they noted that the accessions from Yellow River Valley contained the most allelic richness and simultaneously were highly dispersed in their clustering pattern. These results were in tandem with the findings of Li *et al.* (2008) who registered 1 160 alleles among 1 863 landraces using 59 SSR markers and also found seven clusters with a higher genetic variation coming from Yellow River Valley. These findings support the argument or evidence that Yellow River Valley is thought to be the centre of origin of the cultivated soybean as well. In Japan, a comparative evaluation of 1 305 wild soybean collection and 53 cultivated soybean was carried out using 20 SSR markers and 28 alleles per locus were recorded in wild soybean as opposed to 5 alleles in the cultivated species (Kuroda, *et al.*, 2009). The results revealed less polymorphism in the cultivated soybean compared to its progenitor.

In this study, the SSR markers were selected to characterise genetic diversity and compare the relatedness of the germplasm under examination. They were preferred to other markers because; SSR markers are highly reproducible which is an important aspect in genetic analysis. They do not require restriction with enzymes as compared to RFLPs and they also do not require template DNA, a key requirement in AFLP analysis (Park *et al.*, 2009). They are co-dominant; hence, they are suitable for genetic analysis in segregating F2 populations or parentage analysis of hybrids (Yadav *et al.*, 2007; Missio *et al.*, 2010). Simple sequence repeats markers produce very high allelic variations and are highly polymorphic (Jeffreys *et al.*, 1994). Park *et al.* (2009) compared the utility of RFLP, RAPD, AFLP and SSR markers in germplasm diversity studies and concluded that the SSRs exhibited the highest heterozygosity and higher genetic variation was noted relative to RFLPs. They also observed alleles that ranged from 1 to 37 among 61 genotypes studied. On the other hand, Li *et al*. (2010) characterised the genetic diversity of 303 accessions of *Glycine max* and *Glycine soja* using 99 SSR and 554 SNP markers reported higher gene diversity of 0.77 from SSR markers compared to 0.35 for SNP markers

Further, SSR markers are abundant and well distributed in genomes compared to RAPD and AFLP which are often clustered in certain location of chromosomes or linkage (Robinson *et al.*, 2004). Simple sequence repeats markers require small quantities of DNA for screening including low starter costs, they can be genotyped easily and rapidly using numerous platforms for DNA fragment analysis and the analysis could be semi-automated (Cregan *et al.*, 1999; Robinson *et al.*, 2004). SSR markers are more cost effective when compared to RFLP and SNP which demand high costs related to large scale genotyping.

In addition, SSRs have shown high success rates in diversity studies (Simko *et al.*, 2012). In a comparative genetic diversity study involving the utility of DART, SNP and SSR on 54 sugar beet cultivars, success rate was highest for SSR markers (Simko *et al.*, 2012). In this context, it probably demonstrated their highly polymorphic nature. Peakall *et al.* (1998) demonstrated that soybean primers amplified SSRs with a success rate of 65% despite a lower rate of 3 to 13% outside the subgenus *Glycine.*

1.5 Performance of the soybean breeding programme

1.5.1 Genetic gain

The increase in soybean yield over the years is attributed to plant breeding, agronomic practices and environmental conditions. These elements interact and influence seed yield per unit area (Duvick, 2005; Rowntree *et al.*, 2013). Generally, crop improvement is an incremental process that entails breeding cycles composed of new breeding lines that contribute favourable alleles with small effects into cultivars. The small gains accumulated over time make an enormous impact on soybean productivity. High gains realized by some breeding programs have transformed the cropping systems to a higher level of productivity, whereas low levels of genetic gains impacts negatively on food security and perpetuates hunger and poverty. Hence, evaluation of the breeding gains attained over time remains critical.

Many researchers have estimated the genetic gains made in various breeding programs. The quantification of the breeding progress helps to measure the success of a particular programme over a given time period. Lange and Federerizzi (2009) estimated the genetic gains among three maturity groups of soybean and observed that yield gains ranged from 0.0 to 71.5 kg ha^{-1} $year^{-1}$ depending on maturity and the realized gains equated to 3.49% per year. In a similar regard, Justin (2010) also measured yield progress for soybean varieties released from 1928 to

2008 and reported an annual yield gain of 22.4 kg ha^{-1} for maturity group IV-V (from 1950s to 2000s), 15.7kg/ha for maturity group V (from 1940s to 2000s) as well as 12.4 kg/ha for maturity group VI (from 1920s to 2000s). In contrast, Jin *et al.* (2010) observed an annual yield gain of 0.58% over six decades (from 1950 to 2006 in China) in a yield gain evaluation exercise involving 45 cultivars ranging from maturity groups III to V. Their study showed that seed number per plant was the greatest contributor to seed yield, seed weight and size exhibited slight variations with cultivar year of release, lodging was reduced over the years, yield stability over the years was improved pointing to stable pod production across the environments, the photosynthetic rate was also enhanced coupled with improvements in resistance to abiotic stresses. Similarly, Morrison *et al.* (2000) reported a yearly yield gain of 0.5 over seven decades (released 1934-1920 in Canada).

However, Egli (2008a) reported stagnation of soybean grain yield growth in mid-west United States over thirty one years of breeding (from 1972-2003). This could be attributed to previous selection that narrowed the genetic base of the elite lines. On another note, on-farm yield gains were assessed for both the dry land and irrigated production and 24.9 kg ha^{-1} and 35.1 kg ha^{-1} gains were reported correspondingly (Specht *et al.*, 1999). The grain yield difference was 40% when expressed as percentage and this could be ascribed to insufficient moisture under the dry land production which ultimately impacts negatively on crop yield. These findings were in agreement with the observation of Egli (2008b) who obtained higher yields (3403 kg ha^{-1}) in irrigated production compared to rain fed production (1482 kg ha^{-1}). The assessment of genetic gains has been reported to be useful in identifying varieties that are productive for a long period of time. Such cultivars could be exploited in making new combinations thereby avoiding loss of rare and simultaneously desirable alleles. Justin (2010) cited few cultivars that demonstrated consistent performance and remained competitive for over 15 years. Nonetheless, such cultivars are tantamount to overuse which may compromise diversity.

Although genetic gains have been reported, it seems that most soybean breeding programs have annual gains that are under 1% (Gates and Gates, 2013). Farmers demand more yield and greater yield stability. More importantly, the estimated future demand for food seems to outweigh prevailing genetic gains. However, greater scope for increasing soybean genetic gains still exist and requires attention on increasing efficiency in utilizing soil nutrients, increasing "crowding effect" or plant populations, improving harvest index, changing plant architecture, and improving tolerance to biotic and abiotic stresses (Duvick, 2005). Therefore, time series

estimation of genetic gains for the Zimbabwean soybean germplasm would be a good starting point to establish the status quo. This would define the direction to be undertaken in view of crop improvement.

1.6 Grain yield in soybean

1.6.1 The genetic variability and potential for grain yield in soybean

The existence and magnitude of genetic variability in soybean breeding programs is a prerequisite for crop improvement (Warkard *et al.*, 2008; Aditya *et al.*, 2011). Generally, grain yield is a complex trait and a constituent of several components that are quantitative in nature, therefore, their expression is determined by the genetic, environmental conditions and their interactions (Sudaric and Vrataric, 2002). Contextually, it implies that genotypic coefficients of variation and phenotypic coefficient of variations should be employed in variability studies. Several scholars demonstrated the usefulness of these genetic parameters on assessing genetic variability in soybean (Singh *et al.*, 2000; Aravind, 2006; Malik *et al.*, 2007; Bhat *et al.*, 2012). Genetic variability for grain yield was observed with high significant differences among the cultivars studied (Karnwal and Singh, 2009). Thus, phenotypic selection on higher number of seeds per plant, seed weight and higher number of pods per plant should receive a lot of emphasis when deciding to develop high yielding cultivars. These traits help to accelerate genetic advance in grain yield.

Since the level of grain yield is a function of the combinations of its components, therefore, a good understanding of how yield components interact both phenotypically and genotypically in influencing yield is critical. Knowledge of this is important to identify yield components that can be used as selection criteria in advancing grain yield. Cultivars showing superiority for some of the yield components may be focused on and used as parental stock (Warkard *et al.*, 2008). Exploitation of such cultivars aids to raise the genetic gains. Unfortunately, no cultivar is superior in all the yield components. Earlier reports concluded that soybean grain yield components revolve around number of plants per area, number of pods per plant, number of seeds per pod and 100 seed weight (Johnson *et al.*, 1969). Board *et al.* (1997) contended that soybean grain yield is determined by seed size (100 seed weight) and seed number. He pointed out that seed number is a function of the number of seed per pod and pod number. On the other hand, Ramteke *et al.* (2010) asserted that soybean grain yield is an integrated function of plants per area, branches per plant, pods per branch, seeds per pod and 100 seed weight.

1.6.2 Relationship between seed yield and its secondary components

The observations above suggest that selection for grain yield should factor in yield determining traits. In this regard, the application of statistical tools such as correlation and path coefficient analyses help to reveal the interrelationships between yield and its secondary components. The bottom-line is that yield increase would be accomplished on the premise of the performance of its secondary traits and selection for closely related traits (Malik *et al.*, 2007). Correlation coefficient displays the relationship between the dependent and independent variables, the strength and direction of the relationship. Its pitfall is that it does not adequately predict the success of the selection. The strength of path coefficient analysis is that it measures the direct and indirect effect for one attribute on another and allows the partitioning of the correlation coefficient into direct and indirect influences that one variable has on another (Yagdi, 2009). In other words, it helps to identify the direct, indirect and total causal effect of the correlation (Hefny, 2011).

The relationships between morphological and phenological traits have been assessed using correlation and path coefficient analyses with the aim of determining the effects of important yield components. Iqbal *et al.* (2003) pointed out that pods per plant, followed by 100 seed yield and finally seeds per pod had maximum direct effect on seed yield, whereas plant height had negative direct effect on yield. Contrary to these findings, El-Badawy and Mehasen (2012) demonstrated that number of pods per plant and 100 seed weight had highest indirect effect to seed yield. Ariyo (1995) did not find any direct and indirect effects on all phenotypically related characters that they measured. However, Ariyo (1995) reported positive and significant correlation between; seed yield and plant height; seed yield and number of pods per plant; seed yield and 100 seed weight; and seed yield and seeds per pod. Arshad *et al.* (2006) investigated the association between yield and its secondary traits and observed positive and significant relationship between grain yield and the following; days to maturity, pod length, number of branches, number of unfilled pods, filled pods and total pods and 100 seed weight. No relationship was found between grain yield and days to 50% flowering and seed yield per five plants. Differences exhibited by the above-mentioned observations are possibly attributed to environmental influences emanating from different environments used. Therefore, it was found prudent to evaluate the association of seed yield and its components under the Zimbabwean, Malawian and Zambian conditions. Nonetheless, the tools assist to identify components that may be focused on and serve as selection criteria for higher yield.

1.7 Genotype x environment interactions in soybean

The soybean production environments in Zimbabwe, Zambia and Malawi are characterized by differences in latitudes, altitudes, climatic conditions, soil moisture, soil type and or fertility levels from location to location coupled with seasonal variations. This raises concern over the performance of cultivars under different environmental conditions. In the same vein, when genotypes are compared over several environments, the rankings change or differ and this presents challenges during selection as well as making cultivar recommendations (Cucolotto *et al.*, 2007). For this reason, genotype x environment interaction (GEI) is considered to be a hindrance to crop improvement. Variability in environmental conditions results in significant genotype x environment interactions in addition to the genotype main effects and the environment main effects during the testing of soybean genotypes. Rao *et al.* (2002) defined genotype x environment interaction as the failure of genotypes to achieve the same relative performance in different environments. Fox *et al.* (1997) defined GEI as differential genotypic expression across environments. Crossa (1990) and Fox *et al.* (1997) postulated that there are three types of GEI effects viz, cultivar x location interaction, cultivar x year interaction and cultivar x location x year interaction effects. The significance of these interactions is that they cause differences in the ranking order of genotypes under evaluation in the given multiple environment trials (METs). Therefore, it becomes prudent to test genotypes over several environments and seasons. This is especially important with quantitative traits such as yield because significant GEI is known to curtail the correlation between genotypic and phenotypic values which adversely affects response to selection (Comstock and Moll, 1963). There are generally two types of interactions that breeders encounter in GEI studies namely quantitative and qualitative (Gail and Simon, 1985). Quantitative interactions arise when there is variation in the response of genotypes to environments without rank changes while qualitative or cross over interactions occur when there are changes in rank order across the environments. In this case, qualitative interactions complicate selection and cultivar recommendations. Soybean breeders are concerned about the consistent expression of yield and all agronomic traits across a wide range of environments. Consistent performance is key in crop improvement and acceleration of genetic gains. It is also critical to farmers because they are assured of salvaging something irrespective of environmental and seasonal changes.

1.7.1 Adaptation Strategies

Studies involving genotype x environment interaction indicated that adaptability and stability assessments are crucial for identifying and recommending superior genotypes in specific environments and wide range of environments (Nascimento *et al.*, 2010). Miladinovic *et al.* (2006) showed that the multi-locational trials are a reliable tool for variety adaptability. Generally, there are two types of adaptation strategies viz; specific and general or wide adaptation strategies.

1.7.1.1 Specific adaptation strategies and evidence of GEI in soybean

Annicchiarico (2002) classified genotypes with good performance over a limited number of environments as possessing narrow or specific adaptation. Suffice to say that specific adaptation exists when GEI is significant (Reddy *et al.*, 2011). Its merit in plant breeding is centred on raising genetic gains through the exploitation of positive interaction effects of genotypes with individual locations. Specific adaptation is extensively exploited by national programs and large seed companies which have research operations in several countries and as such having varied environmental conditions. In this case, it becomes logical to target each country as a sub-region and tap on genotype x location (GL) interaction effects through adaptive traits coupled with high heritability of yield derived from reduced GL interaction, thereby increasing crop yields (Annicchiarico *et al.*, 2005). In a comparative study of wide versus specific adaptation strategies in terms of observed and predicted yield gains for 24 cultivars of wheat over 3 years and 47 environments, specific adaptation gave 2 to 7% yield gains above wide adaptation (Annicchiarico *et al.*, 2005).

A lot of investigations around specific adaptation in soybean have been conducted. Bekheit (2000) reported that high yielding cultivars had low stability symbolizing the existence of high GEI. Jandong *et al.* (2011) examined the adaptation and stability of seven cultivars under six different soil pH regimes and observed specific adaptation implying that each genotype had specific soil requirements. Xiong *et al.* (2011) examined the environmental adaptability and stability of 60 accessions over six sites and two seasons and obtained a strong GEI in several genotypes. Ceccarelli and Grando (2007) noted that selection for specific adaptation was also crucial for making inroads and accomplishing sound genetic gains in unfavourable environments. Contextually, specific adaptation presents opportunities to enhance food security. Where different varieties are commercialized for each sub-region or specific environment, it

becomes a valuable approach in the management of genetic vulnerability. In this case, it broadens the diversity of the cultivars in the market domains. However, specific adaptation is associated with high costs, probably because of increased field testing. Many sites coupled with a large number of cultivars are required for testing.

1.7.1.2 Wide adaptation strategies and GEI studies in soybean

The development of varieties that are high yielding with stable yields across multiple environments and seasons is topical to commercial soybean production. This has the advantage of increasing both the production area and production volumes. Conducting field testing of genotypes under several heterogeneous environments, affords researchers a chance to identify genotypes with high mean yield and low GEI (Sreedhar *et al.*, 2011). A genotype is said to have wide adaptation when its average performance is greater than the mean over multi-locations (Annicchiarico, 2002). Allard and Bradshow (1964) reiterated that the best genotype is the one that exhibits consistent performance across a multitude of production environments. Such cultivars that cope with broad range of environments are useful in breeding and are exploited in cropping systems. Gebeyehu and Assefa (2003) lamented that selection focused on high yielding genotypes appeared less stable than the average of all lines and selection for yield only results in throwing out stable genotypes. In view of the diversity of cultivar reactions to the characteristics of environments, it therefore becomes critical to have multi-environmental trials (MET) in order to obtain an accurate idea of their performance (Lecomte *et al.*, 2010). Another key element is yield stability. A stable genotype is defined as a genotype's ability to perform consistently and produce mean performance that is above average in all the locations (Gurmu *et al.*, 2009). In short, a high yielding stable genotype is characterised by reliable seed yield across environments. Therefore, phenotypic characterisation of genotypes under wide range of environments becomes critical in order to appreciate the pattern and magnitude of GEI and be able to observe genotypic responses and consequently identifying superior genotypes. According to Altin *et al.* (2000) the merit of wide adaptation is that the data from several environments is pooled, which increases the precision of the genotypic means.

Many researchers reported on wide adaptation studies that focused on soybean. Al-Assily *et al.* (2002) assessed the performance of five soybean genotypes and observed that three cultivars had mean yields that were above the trial mean with remarkable stability. In a similar regard, Cucolotto *et al.* (2007) found four cultivars out of 30 that combined good adaptation and stability. The variation in wide range of ecologies and seasons has been found to significantly

influence number of seeds per unit area (Egli, 1998). In order to advance genetic gains focus should be placed on physiological causes of GEI.

1.7.2 Genotype x environment evaluation tools

There are several statistical tools that can be employed to analyse genotype x environment data. These models include analysis of variance (ANOVA), multivariate techniques, simple linear regression, additive main effects and multiplicative interaction (AMMI), nonparametric tests such as rank, and Genotype main effects and genotype x environment (GGE) biplot. In this investigation, focus was placed on AMMI.

Apparently, AMMI is a recent but now frequently used in analysing GEI. It is a data visualization model whose strength is premised on its ability to clarify GEI, improves accuracy of yield estimates and is capable of diagnosing other methods as sub-cases when they are better for certain data sets (Gauch, 1988). Further, it integrates the strengths of both the ANOVA and principal component analysis (PCA) into one method (Zobel *et al.*, 1988; Crossa, 1990). Additive main effects and multiplicative interaction uniquely separates genotype main effects (G), environments main effects (E) and partitions the genotype x environment (G x E) interaction which is important for research purposes as opposed to GGE Biplot which gives all the components without environment main effects. Secondly, the AMMI model is capable of separating structural variation from noise with a view to achieve accuracy (Nassiri and Ariyo, 2011). However, there seems to be continued debate between the two models. The AMMI has been applied by several soybean researchers in GEI studies. Gurmu *et al.* (2009) employed the AMMI model and identified three high yielding and stable soybean cultivars. Cucolotto *et al.* (2007) produced similar observations. These results contrasted the findings of (Asfaw *et al.*, 2009) who found no superior cultivars across four sites and three seasons.

Considering that soybean production ecologies vary from one ecology to another, analysis of GEI would help to show patterns of adaptation and stability of the Zimbabwean soybean germplasm. The results could be useful in identifying genotypes that combine both wide adaptation and superior yield and such genotypes could present opportunities for use in the hybridisation and selection programme.

1.9 Summary

The review of literature has shown that plant breeding has the potential to decrease the genetic diversity of soybean. Other evolutionary forces, such as domestication and founder population effects also contribute to decrease in diversity. Furthermore, the consequences of reduced diversity were highlighted which include genetic vulnerability, genetic erosion and yield plateau. The utility of on phenotypic and molecular approaches in diversity studies were presented. Most of the previous diversity studies involving the application of the two methods revealed variability. From the molecular point of view, it was shown that genetic diversity of soybean is low relative to other crop species. The application of various molecular markers in diversity studies and their pros and cons were discussed.

The importance of improvement of genetic gains in soybean breeding was discussed. Significant genetic gains were reported in certain studies. However, it was noted that other breeding programs have annual gains that are below 1%. Therefore, a need exists to quantify the breeding progress that has been made by the Zimbabwean programme since inception.

The majority of traits of economic importance in soybean are polygenic in nature. From the different studies reported in literature, both additive and non-additive effects were found to be important in the inheritance of grain yield in soybean. Correlation and path coefficient analyses were found to be useful tools in that they reveal the associations between grain yield and its secondary traits. Environmental influence was noted to cause some variations in the results involving the application of these tools. This therefore, justifies the need to conduct a similar study under the Zimbabwean conditions.

Given the diverse nature of the multi-environments upon which soybean is grown, multi-locational testing becomes arguably crucial and imperative. METs help to establish whether GEI is significant or not. Significant GEI implies, breeding for specific adaptation. Breeding for general adaptation coupled with cultivar yield stability is critical because the soybean production environments are highly variable. The merits and demerits of the two adaptation strategies were dealt with.

In conclusion, the review of literature revealed the following gaps;

- the diversity of soybean germplasm in Zimbabwe has not been described, documented and is probably underutilized.

- the average yield levels are lower than major soybean producing countries in Africa suggesting that the breeding gains are probably lower. Moreover, there is no information on the breeding gains that have been accomplished by Zimbabwean soybean programme.

- there are no conclusive results on the relationship between grain yield and its secondary traits in the soybean cultivars.

- no studies have been conducted to quantify G x E and assess stability analyses in the heterogeneous production environments of Malawi, Zambia and Zimbabwe.

1.10 References

Aditya, J. P., 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.

Al-Assily, K. H. A., S. R. Saleeb, S. H. Mansour and M. S. A. Mohamed. 2002. Stability parameters of soybean genotypes as criteria for response to environmental conditions. Minufiya Journal of Agricultural Research 27:169-180.

Allard, R. W. and A. D. Bradshow. 1964. Implications of genotype-environment interactions in applied plant breeding. Crop Science 4:503-507.

Alliprandini, L. F. and N. A. Vello. 2004. Heritability and correlations in four-way soybean crosses. Euphytica 136:81-91.

Altin, G. N., R. J. Baker, K. B. McRae and X. Lu. 2000. Selection response in subdivided regions. Crop Science 40:7-13.

Annicchiarico, P. 2002. Defining adaptation strategies and yield-stability targets in breeding, p 365-383. *In* M.S. Kang (ed.). Quantitative genetics, genomics and plant breeding. CABI Publishing, Wallingford, United Kingdom.

Annicchiarico, P., F. Bellah and T. Chiari. 2005. Defining subreions and estimating benefits for a Specific-Adapation Strategy by Breeding Programmes. Crop Science 45:1741-1749.

Antalikova, G., M. Zakova and D. Benedikova. 2008. Characterisation of soybean traits variability by cluster analysis. Agriculture 54:45-53.

Aravind, G. 2006. Genetic Variability and Diversity Studies in Soybean [*Glycine max* (L.) Merrill] Genetics and Plant Breeding, Dharwad University of Agricultural Sciences. p. 1-103.

Ariyo, O. J. 1995. Correlation and path coefficient analysis of components of seed yield in soybeans. African Crop Science Journal 3:29-33.

Arshad, M., N. Ali and A. Ghafoor. 2006. Character correlation and path coefficient in soybean [*Glycine max* (L.) Merrill.]. Pakistan Journal of Botany 38:121-130.

Asfaw, A., F. Alemayehu, F. Gurumu and M. Atnaf. 2009. AMMI and SREG GGE biplot analysis for matching varieties onto soybean production environments in Ethiopia. Academic Journals 4:1322-1330.

Bekheit, M. A. 2000. Evaluation of some soybean genotypes in Upper Egypt Crop Science, Assuit University, Egypt.

Bhandarkar, S. 1999. Studies on genetic variability and correlation analysis in soybean. Mysore Journal of Agriculural Science 33:130-132.

Bhat, S., G. T. Basavaraja and P. M. Salimath. 2012. Analysis in segregating generation of soybean [*Glycine max* (L.) Merrill]. Karnataka Journal of Agricultural Science 25:176-178.

Board, J. E., M. S. Kang and B. G. Harville. 1997. Path analyses, identifying Indirect Selection criteria for yield of late-planted soybean. Crop Science 37:979-884.

Board, J. E., S. Kumudini, J. Omielan, E. Prior and C. S. Kahlon. 2010. Yield response of soybean to partial and total defoliation during the seed-filling period. Crop Science 50:703-712.

Board, J. E. and H. Modali. 2005. Dry matter accumulation predictors for optimal yield in soybean. Crop Science 45:1790-1799.

Carter, T. E., R. Nelson, C. H. Sneller and Z. Cui. 2004. Genetic Diversity in Soybean: improvement, production, and uses. *In* H.R. Boerma, et al. (eds.), US, Madison, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. p. 303-416

Ceccarelli, S. and S. Grando. 2007. Decentralised-participatory plant breeding: an example of demand driven research. Euphytica 155:349-360.

Ceyhan, E. and M. A. Avci. 2005. Combining ability and heterosis for grain yield and some yield components in pea (*Pisum sativum* L.). Pakistan Journal of Biological Sciences 8:1447-1452.

CGIAR, 2012. Leveraging legumes to combat poverty, hunger, malnutrition and environmental degradation. http://hdl.handle.net/10947/5316.

Chen, Y. and R. L. Nelson. 2004. Genetic variation and relationships among cultivated, wild and semiwild soybean. Crop Science 44:316-325.

Cho, Y. and R. A. Scott. 2000. Combining ability of seed vigour and seed yield in soybean. Euphytica 112:145-150.

Cholastova, T. and D. Knotova. 2012. Using Morhological and Microsatellite (SSR) Markers to Assess the Genetic Diversity in Alfalfa (*Medicago sativa* L.). World Academy of Science, Engineering and Technology 69:1-7.

Cober, E., M. J. Morrison, B. Ma and G. Butler. 2005. Genetic improvement rates of short-season soybean increase with plant population. Crop Science 45:1029-1034.

Collard, B. C. Y., M. Z. Z. Jahufer, J. B. Brouwer and E. C. K. Pang. 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142 169-196.

Comstock, R. F. and R. M. Moll. 1963. Genotype-environment interactions. *In* W.D. Hanson and H.F. Robinson (eds.). Statistical genetics and plant breeding, NAS-NCR, Washington, United States. p. 164-196.

Cooper, H. D. and T. Hodgkin. 2001. Broadening the Genetic Base of Crop production CABI, Wallingford, United Kingdom.

Cregan, P. B., D. L. Hyten, Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht. and R. C. Shoemaker. 2006. Domestication, founding effects, artificial-genetic

bottlenecks and soybean genetic variability. Plant and Animal Genomes XIV Conference, Town and Country Convention Centre, San Diego, CA, United States.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. l. Kahler, N. Kaya, T. T. Van Toai, D. G. Lohnes, J. Chung, and and J. E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Science 39:1464-1490.

Cregan, P. B. and R. W. Yaklich. 1986. Dry matter and nitrogen accumulation and partitioning in selected soybean genotypes of different derivation. Theoretical and Applied Genetics 72:782-786.

Crossa, J. 1990. Statistical analysis of multi-location trials. Advances in Agronomy 44:55-85.

Cucolotto, M., V. C. Pipolo, D. D. Garbuglio, N. S. F. Junior, D. Destro and M. K. Kamikoga. 2007. Genotype x environment interaction in soybean: evaluation through three methodologies. Crop Breeding and Applied Biotechnology 7:270-277.

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 Intergrative Biology 5:125-129.

Doyle, J. J., J. L. Doyle, J. T. Rauscher and A. H. D. Brown. 2004. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). New Pathologist 161:121-132.

Durai, A. A. and B. Subbalakshmi. 2007. Heterosis and combining ability in soybean for the traits of vegetable importance. Vegetable Science 51:51-54.

Duran, C., N. Appleby, D. Edwards and J. Batley. 2009. Molecular Genetic markers: Discovery, Applications, data storage and visualisation. Current Bioinformatics 4:16-27.

Duvick, D. N. 2005. Genetic progress in yield of United States Maize (*Zea mays* L.). Maydica 50:193-202.

Duvick, N. D. 1984. Genetic diversity in major farm crops on the farm and in reserve. Economic Botany 38:161-178.

Egli, D. B. 1998. Seed biology and the yield of grain crops CAB International, Oxford, United Kingdom. p: 178.

Egli, D. B. 2008a. Comparison of corn and soybean yields in the United States: Historical trends and future prospects. Agronomy Journal 100:79-88.

Egli, D. B. 2008b. Soybean yield trends from 1972 to 2003 in mid-western USA. Field Crops Research 106:53-59.

El-Badawy, M. M. and S. A. S. Mehasen. 2012. Correlation and path coefficient analysis for yield and yield components of soybean genotypes under different planting density. Asian Journal of Crop Science 4:150-158.

Esterhuizen, D., 2011. Assessments of commodity and trade issues. Global Agricultural Information Network Report, USDA Foreign Agricultural Service.

Evenson, R. and D. Gollin, 2004.The state of food and agriculture. FAO report

Fehr, W. R. 1987 Principles of cultivar development: Macmillan Publishing Company, New York.

Feltus, F. A., J. Wan, S. Schulze, J. C. Estill, N. Jiang and A. H. Paterson. 2004. An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. Genome Research 14:1812-1819.

Feng, C., A. Hou, P. Chen, B. Cornelius, A. Shi and B. Zhang. 2009. Genetic diversity among popular historial Southern U.S. soybean cultivars using AFLP markers. Journal of Crop Improvemet 22:31-46.

Foulkes, M. J., M. P. Reynolds and B. R. Sylvester. 2009. Genetic improvement of grain crops: yield potential. p.355-385 *In* Crop Physiology: Applications for Genetic Improvement and Agronomy ELsevier, Burlington, MA, United States.

Fox, P. N., J. Cross and I. Ramagosa. 1997. Multiple environment testing and genotype x environment interaction. *In* R. A. Kempton and P. N. Fox (eds.) Statistical Methods for Plant Variety Evaluation. Chapman and Hall, London. p. 116-138

Fujisaka, S., D. Williams and M. Halewood, 2011.The impact of climate change on countries' interdependence on genetic resources for food and agriculture. FAO conference paper. p. 19-26.

Fukuda, Y. 1993. Cytogenetical studies on the wild and cultivated Manchurican soybeans [*Glycine max* (L.) Merrill]. Japanese Journal of Botany 6:489-506.

Gadde. 2006. Genetic investigations in soybean (*Glycine max* (L.) Merrill), Dharwad Agricultural Sciences University. Student thesis. p.1-75.

Gail, M. and R. Simon. 1985. Testing for quantitative interactions between treatment effects and patient subsets. Biometrics 41:361-372.

Gates, B. and M. Gates. 2013. Delivering increased rates of genetic gain to smallholders: bringing the breeding technology revolution to Africa and South Asia-United States [Online] http://www.cgiarfund.org/sites/cgiarfund.org/ (verified on 30 July 2013). p. 1-5.

Gauch, H. G. 1988. Model selection and validation for yield trials with interaction. Biometrics 44:705-715.

Gavioli, E. A., D. Perecin and A. O. D. Mauro. 2008. Analysis of combining ability in soybean cultivars. Crop Breeding and Applied Biotechnology 8:1-7.

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

Gizlice, Z., T. E. Carter Jr and J. W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947-1988. Crop Science 34:1143-1551.

Govindarao, C. N. 2010. Characterisation of [*Glycine max* (L.) Merrill.] varieties through morphological, chemical, molecular markers and image analyser, Dharwad University of Agricultural Sciences. p.1-115.

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

Guo, J., Y. Wang, C. Song, J. Zhou, L. Qiu, H. Huang, and and Y. Wang. 2010. a single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. Annals of Botany 106:505-514.

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

Halushka, M. K., J. B. Fan, K. Bentley, L. Hsie, N. Shen, A. Weder, R. Cooper, R. Lisshutz and A. Chakravarti. 1999. Parterns of single nucleotide polymorphisms in candidate genes for blood-pressure homoestasis. Nature Genetics 22:239-247.

Hamblin, M. T., S. E. Mitchell, G. M. White, J. Gallego and R. Kukatla. 2004. Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics 167:471-483.

Hammer, K. and G. Laghetti. 2005. Genetic erosion – examples from Italy. Genetic Resources and Crop Evolution 52:629-634.

Hamzekhanlu, M. Y., A. I. Darbandi, P. Beiranvand, M. T. Hallajian and A. Majdabadi. 2011. Phenotypic and molecular analysis of M7 generation of soybean mutant lines through random amplified polymorphic DNA (RAPD) marker and some morphological traits. African Journal of Agricultural Research 6:1779-1785.

Hefny, M. 2011. Genetic parameters and path analysis of yield and its components in corn inbred lines (*Zea mays* L.) at different sowing dates. Asian Journal of Crop Science 3:106-117.

Hymowitz, T. 1990. Soybeans: The success story and advances in new crops *In* J. Janick, et al. (eds.), First national symposium of new crops, Indiana, October 23-26, 1988. p. 159-163.

Hymowitz, T. 2004. Speciation and cytogenetics. *In* H. R. Boerma and J. E. Specht (eds.) Soybeans: improvement, production and uses. American Society of Agronomy, Crop Science Society of America and Soil Science Society of America. p. 97-136.

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

Hyten, D. L. 2005. Genetic diversity and linkage disequilibrium in wild soybean, landraces, ancestral and elite soybean populations. Department of Natural Resource Sciences and Landscape Architecture, University of Maryland. p. 1-88.

Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht., R. C. Shoemaker. and P. B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. National Academy of Sciences of US 103:1666-1671.

IITA, 2008. 30 years Research and Development in soybean: What's next?

Iqbal, M. Z. and S. A. Khan. 2003. Line x Tester analysis in true seed of potato (*Solanum tuberosum* SPP tuberosum). Journal of Biological Sciences 3:674-680.

Iqbal, M. Z., F. A. Nehvi, S. A. Wani, R. Qadar and Z. A. Dar. 2007. Combining ability for yield and yield related traits in maize. International Journal of Plant Breeding and Genetics 1:101-105.

Iqbal, S., T. Mahmood, M. Ali, M. Anwar and M. Sarwar. 2003. Path Coefficient Analysis in Different Genotypes of Soybean [*Glycine max* (L.) Merrill]. Pakistan Journal of Biological Sciences 6:1085-1087.

Iqbal, Z., M. Arshad, M. Ashraf, T. Mahmood and A. Waheed. 2008. Evaluation of soybean [*Glycine max* (L.) Merrill] germplasm for some important morphological traits using multivariate analysis. Pakistan Journal of Botany 40:2323-2328.

Jandong, E. A., M. I. Uguru and B. C. Oyiga. 2011. Determination of yield stability of seven soybean (*Glycine max*) genotypes across diverse soil pH levels using GGE biplot analysis. Journal of Applied Biosciences 43:2924-2941.

Jarvis, A., D. Ferguson, L. Williams, P. Guarino, H. Jones, J. Stalker, R. Valls, C. Pittman, Simpson and P. Bramel. 2003. Biogeography of wild Arachis: assessing conservation status and setting future priorities. Crop Science 43:1100-1108.

Jeffreys, A. J., K. Tamaki, A. MacLeod, D. G. Monckton, D. L. Neil and J. A. L. Armour. 1994. Complex gene conversion events in germline mutations at human minisatellites. Nature Genetics 6:136-145.

Jin, J., X. Liu, G. Wang, L. Mi, Z. Shen, X. Chen and S. J. Herbert. 2010. Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in North East China. Field Crops Research 115:116-123.

Johnson, E. T., J. W. Pendleton, D. B. Peters and D. R. Hicks. 1969. Influence of supplemental light on apparent photosynthesis, yield and yield components of soybean (*Glycine max* L. Merrill). Crop Science 9:577-581.

Jonas, C. N., B. Vanlauwe, J. M. Mahasi, E. Katungi, C. Akech, F. S. Mairura, C. N. Justina, and and N. Sanginga. 2008. Soybean situation and outlook analysis: the case of Kenya.

Justin, R. 2010. Evaluation of current and historical soybean genotypes for yield and forage use. Crop Science 61:321-380.

Kaga, A., T. Shimizu, S. Watanabe, Y. Tsubokura, Y. Katayose, K. Harada, D. A. Vaughan and A. Tomooka. 2012. Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. Breeding Science Journal 61:566-592.

Kanazin, V., H. Talbert, D. See, P. DeCamp, P. Nevo and T. Blake. 2002. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Molecular Biology 48:529-537.

Kang, M. S. 2004. Breeding: Genotype-by-environment interaction. In: R.M. Goodman (eds.) Encyclopedia of Plant and Crop Science. Marcel-Dekker, New York, United States.

Kapuya, T., D. Saruchera, A. Jongwe, T. Mucheri, T. Mujeyi, L. N. Traub and F. Meyer, 2010.The grain industry value chain in Zimbabwe.

Karnwal, M. K. and K. Singh. 2009. Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Research 32:70-73.

Kashyapi, A., A. P. Hage and D. A. Kulkarni. 2012. Impact of climate change on world agriculture : A Review; Workshop Proceedings. Agricultural Meteorological Division, India. p. 10-25.

Keim, P., W. Beavis, J. Schupp and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germplasm. Theoretical and Applied Genetics 85:205-212.

Kenga, R., S. O. Alabi and S. C. Gupta. 2004 Combining ability studies in tropical sorghum [*Sorghum bicolor* (L.) Moench] Field Crops Research 88:251-260.

Khan, A., M. Hatam and A.Khan. 2008. Heritability and interrelationship among yield determining components of soybean varieties. Pakistan Journal of Agricultural Research 16:1-4.

Khattak, G. S. S., M. Ashraf and R. Zamir. 2004. . Gene action for synchrony in pod maturity and indeterminate growth habit in mungbean [*Vigna radiata* (L.)] Wilczek. Pakistan Journal of Botany 36:589-594.

Kihara, H. 1969. History of biology and other sciences in Japan in retrospect. International Congress. Genetics 3:49–70.

Kucharik, C. J. and S. P. Serbin. 2008. Impacts of recent climate change on Wisconsin corn and soybean yield trends. Environmental Research Letters 3:1-10.

Kumar, P., V. K. Gupta, A. K. Misra, D. R. Modi, and and B. K. Pandey. 2009. Potential of Molecular markers in plant biotechnology. Plant Omics Journal 2:141-162.

Kuroda, Y., N. Tomooka, A. Kaga, S. M. Wanigadeva and D. A. Vaughan. 2009. Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*Glycine max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. Genetic Resources and Crop Evolution 56:1045-1055.

Ladizinsky, G. 1985. Founder effect in crop-plant evolution. Economic Botany 39:191-199.

LAES. 2009. Louisiana soybean and feed grain review. Research and Extension. Volume II. United States.p. 1-9.

Lange, C. E. and L. C. Federerizzi. 2009. Estimation of soybean genetic progress in the south of Brazil using multienvironmental yield trials. Science Agriculture 66:309-316.

Lark, K. G., K. Chase, F. Alder, L. M. Mansur and J. H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proceedings of the National Academy of Science 92:4656-4660.

Lecomte, C., L. Prost, M. Cerf, and and J. M. Meynard. 2010. Basis for designing a tool to evaluate new cultivars. Agronomy Sustainable Development 30:667-677.

Li, Y., R. Guan, Z. Liu, Y. Ma, L. Wang, L. Li, F. Lin, W. Luan, P. Chen, Z. Yan, Y. Guan, L. Zhu, X. Ning, M. J. M. Smulders, W. Li, R. Piao, Y. Cui, Z. Yu, M. Guan, R. Chang, A. Hou, A. Shi, B. Zhang, S. Zhu and L. Qiu. 2008. Genetic structure and diversity of cultivated soybean [*Glycine max* (L.) Merrill] landraces in China. Theoretical and Applied Genetics 117:857-871.

Li, Z. and R. L. Nelson. 2002. RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. Crop Science 42:1737-1744.

Li, Z. and R. L. Nelson. 2001. Genetic diversity among soybean accessions from three countries measured by RAPDs. Crop Science 41:1337-1347.

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.

Lobell, D. B. and G. P. Asner. 2003. Climate and management of contributions to recent trends in United States agriculutural yields. Crop Science 299:1032.

Lorenzen, L. L., S. Boutin, N. Young, J. E. Specht. and R. C. Shoemaker. 1995. Soybean pedigree analysis using map based molecular markers. Tracking RFLP markers in cultivars. Crop Science 35:1326-1336.

Machikowa, T., C. Saetang and K. Funpeng. 2011. General and specific combining ability for quantitative characters in sunflower. Journal of Agricultural Science 3:91-95.

Magorokosho, C. 2006. Genetic diversity and performance of maize varieties from Zimbabwe, Zambia and Malawi. Plant Breeding Department, Texas A and M University. Student thesis.

Malik, M. F. A., M. Ashraf, A. Qureshi, S. and M. R. Khan. 2011. Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pakistan Journal of Biological Sciences 43:1249-1255.

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.

Manjaya, J. G. and V. A. Bapat. 2008. Studies on genetic divergenece in soybean, [*Glycine max* (L.) Merrill]. Journal of Oilseeds Research 25:178-180.

Matsuo, E., T. Sediyama, C. D. Cruz, R. D. de Lima Oliveira, R. de Cássia Teixeira Oliveira and A. P. O. Nogueira. 2011. Genetic diversity in soybean genotypes with resistance to *Heterodera glycines*. Crop Breeding and Applied Biotechnology 11:304-312.

Maughan, P. J., S. Maroof, M. A. Buss and G. M. Huestis. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance and near isogenic line analysis. Theoretical and Applied Genetics 93:392-401.

Mikel, M. A., B. W. Diers, R. L. Nelson and H. H. Smith. 2010. Genetic diversity and agronomic improvement of North American soybean germplasm. Crop Science 50:1219-1229.

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

Mirsky, S. B., M. R. Ryan, J. R. Teasdale, W. S. Curran, C. S. Reberg-Horton, J. T. Spargo, M. S. Wells, C. L. Keene and J. W. Moyer. 2013. Overcoming Weed Management Challenges in Cover Crop-Based Organic Rotational No-Till Soybean Production in the Eastern United States. Weed Science Society of America 27:193-203.

Missio, R. F., E. T. Caixeta, E. M. Zambolim, L. Zambolim, C. D. Cruz and N. S. Sakiyama. 2010. Polymorphic information content of SSR markers for *Coffea* spp. Crop Breeding and Applied Biotechnology 10:89-94.

Mohammadi, S. A. and B. M. Prasanna. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. Crop Science 43:1435-1248.

Moose, S. P. and R. H. Mumm. 2008. Molecular Plant Breeding as the Foundation for 21st Century Crop Improvement. Plant Physiology 147:969-977.

Morrison, M. J., H. D. Voldeng and E. R. Cober. 2000. Agronomic Changes from 58 Years of Genetic Improvement of Short-Season Soybean Cultivars in Canada. Agronomy Journal 92:780-784.

Mueller, T. A., M. R. Miles, W. Morel, J. J. Marois, D. L. Wright, R. C. Kemerait, C. Levy and G. L. Hartman. 2009. Effect of Fungicide and Timing of Application on Soybean Rust Severity and Yield. American Phytopathological Society 93:243-248.

Mugandani, R., M. Wuta, A. Makarau and B. Chipindu. 2012. Reclassification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. African Crop Science Journal 20:361-369.

Mulato, B. M., M. Möller, M. I. Zucchi, V. Quecini and J. B. Pinheiro. 2010. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. Pesquisa Agropecuaria Brasileira 45:276-283.

Mutengwa, C. 2004. The Search for Molecular Markers, Parental Characterisation and Inheritance Studies of Witchweed [*Striga asiatica* (L.) Kuntze] Resistance in Sorghum [*Sorghum bicolor* (L.) Moench]Crop Science, University of Zimbabwe, Harare.

NAMC, 2011.The South African Soybean Value Chain by the Markets and Economic Research Centre of the NAMC: Strategic positioning of South African Agriculture in dynamic global markets. p. 1-94

Nascimento, M., E. L. Finoto, T. Sediyama and D. C. Cruz. 2010. Adaptability and stability of soybean in terms of oil and protein content. Crop Breeding and Applied Biotechnology 10:48-54.

Nassiri, 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.

Ojo, D. K., A. O. Ajayi and O. A. Oduwaye. 2012. Genetic Relationships among soybean accessions based on morphological and RAPDs techniques. Pertanika Journal of Tropical Agricultural Science. 35:237-248.

Oliveira, M. F., R. L. Nelson, I. O. Geraldi, C. D. Cruz and J. F. F. d. Toledo. 2010. Establishing a soybean germplasm core collection. Field Crops Research 119:277-289.

Opperman, C., 2011.Technical Report: Soybean Value Chain. Southern Africa Trade Hub, USAID. p. 7-12.

Orf, J. H., K. Chase, F. R. Adler, L. M. Mansur and K. G. Lark. 1999. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Science 39:1652-1657.

Panhwar, S. A., M. J. Baloch, W. A. Jatoi, N. F. Veesar and M. S. Majeedano. 2008. Combining ability estimates from line x tester mating design in upland cotton. Pakistan Academy of Science 45:69-74.

Park, Y. J., J. K. Lee and N. S. Kim. 2009. Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crops. Molecules 14:4546-4569.

Peakall, R., S. Gilmore, W. Keys, M. Morgante and A. Rafalski. 1998. Cross-species amplification of Soyabean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for transferability of SSRs in plants. Molecular Biology and Evolution 15 1275-1287.

Pedersen, P. and J. G. Lauer. 2004. Response of soybean yield components to management system and planting date. Agronomy Journal 96:1372-1381.

Ploetz, R. C. 2006. Fusarium wilt of banana is caused by several pathogens referred to as *Fusarium oxysporum* f. sp. *cubense*. Journal of Phytopathology 96:653-656.

Powell, W. M., J. J. Doyle, J. W. McNicol, S. V. Tingey and A. J. Rafalski. 1996. Genepool variation in genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics Society of America 144:793-803.

Provine, W. B. 2004. Genetics and speciation. Genetics 167:1041-1046.

Ramteke, R., V. Kumar, P. Murlidharan and D. K. Agarwal. 2010. Study on genetic variability and traits interrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. Electronic Journal of Plant Breeding 1:1483-1487.

Rao, M. S. S., B. G. Mullix, M. Rangappa, E. Cebertd, A. S. Bhagsaria, V. T. Saprad, J. M. Joshie and R. B. Dadsone. 2002. Genotype × environment interactions and yield stability of food-grade soybean genotypes. Agronomy Journal 94:72-80.

Ravikumar, M. 1999. Identification of soybean varieties based on seed, seedling and plant characters, Univ. Agric. Sci.., Bangalore, Karnataka, India.

Reddy, D. K. R., O. Venkateswarlu, M. C. Obaiah and G. L. S. Jyothi. 2011. Genotype x Environment interaction for grain yield and its components in redgram [*Cajanus cajan* (L.) Millsp]. Legume Research 34:288-291.

Reif, J. C., P. Zhang, S. Dreisigacker, M. L. Warburton, M. van Ginkel, D. Hoisington, M. Bohn and A. E. Melchinger. 2005. Wheat genetic diversity trends during domestication and breeding. TAG Theoretical and Applied Genetics 110:859-864.

Reyna, N. and C. H. Sneller. 2001. Evaluation of marker assisted introgression of yield QTL alleles into adapted soybean. Crop Science 41:1317-1321.

Robinson, A. J., C. G. Love, J. Batley, G. Barker and D. Edwards. 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20:1475-1476.

Rowntree, S. C., J. J. Suhre, N. H. Weidenbenner, E. W. Wilson, V. M. Davis, S. L. Naeve, S. N. Casteel, B. W. Diers, P. D. Esker, J. E. Specht and S. P. Conley. 2013. Genetic gain x management interactions in soybean: Planting date. Crop Science 53:1128-1138.

Salimi, S., H. S. Lahiji, G. M. Abadi, S. Salimi and S. Moradi. 2012. Genetic diversity in soybean genotypes under drought stress condition using factor analysis and cluster analysis. World Applied Sciences Journal 16:474-478.

Satyavathi, C. T., K. V. Bhat, C. Bharadwaj, S. P. Tiwari and V. K. Chaudhury. 2006. AFLP analysis of genetic diversity in Indian soybean [*Glycine max* (L.) Merr.] varieties. Genetic Resources and Crop Evolution 53:1069-1079.

Scallon, B. J., C. D. Dickinson and N. C. Nielsen. 1987. Characterisation of a null allele for the *Gy4* glycinin gene from soybean. Molecular Genetics and Genomics 208:107-113.

Sharma, S. C. and S. R. Maloo. 2009. Heterosis and combining ability studies for oil and protein content in soybean [*Glycine max* (L.) Merrill]. Indian Journal of Plant Genetic Resources 22:253-259.

Shibles, R. M. and C. R. Weber. 1966. Interception of solar radiation and dry matter production by various soybean planting patterns. Crop Science 6:55-59.

Shurtleff, W. and A. Aoyagi, 2007. History of Soybeans and Soyfoods in Africa, Soyinfo Centre, Lafeyette Publishers, USA. Part 1.

Siahsar, B. and A. Rezai. 1999. Correlation and path coefficient analysis of morphological and phenotypic traits associated with yield in soybean ([*Glycine max* (L.) Merrill]. Indian Journal of Agricultural Science 39:685-695.

Simko, I., E. I. and T. J. v. Hintum. 2012. Empirical evaluation of DART, SNP and SSR marker-systems for genotyping, clustering and assigning sugar beet hybrid varieties into populations. Plant Science 184:54-62.

Singh, J., R. P. Parmar, and H. S. Yadav. 2000. Asssesment of genetic variability and selection parameters in early generation of soybean. Advances in Plant Sciences 13:227-232.

Singh, R. P., D. P. Hodson, Y. Jin, J. Huerta-Espino, M. G. Kinyua, R. Wanyera, P. Njau and a. R. W. Ward. 2006b. Current status, likely migration and strategies to mitigate the threat to wheat production from race Ug99 (TTKS) of stem rust pathogen. Agricultural Journal of Veterinary Science, Nutritional and Natural Resources 1:1-13.

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe and R. C. Bridges. 1993. Restriction fragment length polymorphism in soybean germplasm of the southern USA. Crop Science 33:1169-1176.

Sneller, C. H. 1994. Pedigree analysis of elite soybean lines. Crop Science 34:1515-1522.

Specht, J. E., D. J. Hume and S. V. Kumudini. 1999. Soybean yield potential-a genetic and physiological perspective. Crop Science 39:1560-1570.

Sreedhar, S., T. D. Reddy and M. S. Ramesha. 2011. Genotype x environment interaction and stability for yield and its componentsin hybrid rice cultivars (*Oryza sativa L*). International Journal of Plant Breeding and Genetics. 5:194-208.

Stacey, G. 2008. Genetics and Genomics of Soybean. Crops and Models, Volume II Springer Publishers, Manhattan, New York, United States. p.337-355.

Sudaric, A. and M. Vrataric. 2002. Variability and interrelationships of grain quantity and quality characteristics in soybean. Euphytica 53:137-142.

Sulodhani, D. E., N. B. Singh, A. B. Devi, N. G. Singh and G. M. Laishram. 2005 Gene action for fruit yield and its components in tomato (*Lycopersicon esculentum* Mill.). Indian Journal of Genetics 65:221-222.

Sultana, S., L. Rahman and S. H. Habib. 2005. Genetic variability, correlation and path analysis in soybean. International Journal of Sustainable Agriculural technology 1:16-20.

Tanksley, S. D. and S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066.

Tarasatyavathi, C., C. H. Bharadwaj, S. M. Husain, P. G. Karmaka and Y. Mohan. 2004. Identification key for soybean (*Glycine max*) varieties released or notified in India. Indian Journal of Agricultural Science 74:215-218.

Tattersfied, R. J., G. Hutchinson and J. S. Tichagwa. 1988. Oilseed Handbook. Commercial Oil Producers Association.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley and B. S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.) Proceedings of the National Academy of the Sciences, United States 1998. p. 9161-9166.

Tenaillon, M. I., J. U'Ren, O. Tenaillon, and and B. S. Gaut. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize: a multilocus investigation of the domestication process in maize. Molecular Biology and Evolution 21:1214-1225.

Erasmus T. E. 2008. Genetic Diversity of Proprietary Inbred Lines of Sunflower, Determined by Mapped SSR Markers and Total protein analysis, UKZN, Pietermaritzburg.

Thompson, J. A., L. N. Randall and L. O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Science 38:1348-1355.

Tichagwa, J. S. 2009. World Soyabean Research Conference VIII. Abstracts for oral presentations and posters. Developing a Global Soya Blueprint for a Safe Secure and Sustainable Supply, WSRC, 10-15 August, 2009, Beijing, China.

Ude, G. H., K. W.J., J. M. Costa, P. B. Cregan and J. Alvernaz. 2003. Genetic diversity of soybean cultivars from China, Japan, North America, and North American ancestral lines determined by amplified fragment length polymorphism. Crop Science 43:858-1867.

van de Wouw, M., T. van Hintum, C. Kik, R. van Treuren and B. Visser. 2010a. Genetic Diversity trends in twentieth century crop cultivars: a meta analysis. Theoretical and Applied Genetics 120:1241-1252.

van de Wouw, M., T. van Hintum, C. Kik, T. van Treuren and B. Visser. 2010b. Genetic diversity trends in twentieth century crop cultivars: a meta analysis. Theoretical and Applied Genetics. 120:1241-1252.

Vaughan, D. A., E. Balazs and J. S. H. Harrison. 2007. From crop domestication to super-domestication. Annals of Botany 100:893-901.

Wang, L., R. Guan, L. Zhangxiong, R. Chang and L. Qiu. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Science 46:1032-1038.

Warkard, Y. N., N. R. Potdukhe, A. M. Dethe, P. A. Kahate and R. R. Kotgire. 2008. Genetic variability, heritability and genetic advance for quantitative traits in sorghum germplasm. Agricultural Science Digest 28 28:165-169.

Weber, C. R., R. M. Shibles and D. E. Byth. 1966. Effect of plant population and row spacing on soybean development and production. Agronomy Journal 58:99-102.

White, J., J. R. Law, I. MacKay, K. J. Chalmers, J. S. C. Smith, A. Kilian and W. Powell. 2008. The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DART markers and considered by genome. Theoretical and Applied Genetics 116:439-453.

Xiong, L. Z., Y. C. Yan, L. W. Guo, Q. Yong, Z. L. Feng, C. R. Zhen and Q. L. Juan. 2011. Analysis of adaptability of soybean mini core collection in Huang-Huai rivers region in China. Science Direct 37:443-451.

Xu, D. H. and J. Y. Gai. 2003. Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. Plant Breeding 122:503-506.

Xu, Y. 2009. Molecular plant breeding. International Maize and Wheat Improvement Centre (CIMMYT), Mexico.

Yadav, O. P., S. E. Mitchell, A. Zamora, T. M. Fulton and S. Kresovich. 2007. Development of new simple sequence repeat markers for pearl millet. Open Access Journal 3:1-4.

Yadav, R. K. 2006. Genetic variability estimates in soybean. International Journal of Plant Science 2:9-11.

Yagdi, K. 2009. Path coefficient analysis of some yield components in durum wheat (*Triticum durum* Desf.). Pakistan Journal of Botany 41:745-751.

Yang, K., K. Moon, N. Jeong, B. Back, H. M. Kim and S. C. Jeong. 2008. Genome structure in soybean revealed by a genomewide genetic map constructed from a single population. 92:52-59.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli, D. R. Grimm, S. M. Hyatt, E. W. Fickus, Y. N.D. and P. B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. Genetics Society of America 163:1123-1134.

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

CHAPTER 2

GENETIC DIVERSITY OF ZIMBABWEAN SOYBEAN GENOTYPES BASED ON PHENOTYPIC AND SSR MARKERS

2.1 Abstract

Knowledge of the genetic diversity of the available germplasm is an important foundation for crop conservation, management and improvement. Most importantly, successful breeding strategies rely on a full understanding of the genetic diversity of the crop plant. The objectives of this study were to (i) evaluate and compare the genetic diversity estimates of 42 soybean genotypes as determined by phenotypic traits and SSR markers and (ii) assess parental potential of these genotypes in cultivar breeding and development. Field trials were conducted at 13 sites using a 6 x 7 rectangular lattice design with three replications in Zimbabwe, Malawi and Zambia during 2010/2011 to 2011/2012 seasons. The 42 genotypes constituted a collection of all the soybean genotypes that were registered for commercialization and are part of the germplasm for the soybean breeding programme in Zimbabwe. The soybean programme is over 73 years old (1940-2013); therefore, it was found prudent to investigate the levels of diversity that still exists within the germplasm pool. The genetic diversity and relatedness was estimated using ten phenotypic traits and 30 SSR markers. Wide ranges of values among all the traits were observed indicating great variability. Furthermore, the frequency distribution of the traits displayed broad variability. The phenotypic traits and SSR markers assigned the soybean genotypes into 8 and 15 clusters respectively. Most importantly, the clustering patterns from SSR derived dendrograms corresponded very well with the pedigree records. The phenotypic dendrogram showed that clusters **I, II, IV** and **V** had few number of genotypes suggesting that these genotypes exhibit maximum variability. The SSR analysis detected a total of 135 alleles with a mean of 4.56. The average gene diversity was 0.50 and the observed heterozygosity was 0.11. The polymorphic information content ranged from 0.0879 to 0.7669 with a mean of 0.45. The SSR primer, Satt012 was the most informative. Clearly, the results demonstrated that the SSR marker data exhibited the existence of wider genetic diversity compared to the phenotypic data. Thus, the SSR marker technique was more polymorphic, informative and highly discriminatory. Genotypes, G41 and G7; G41 and G1; G41 and G42 were the most divergent; therefore, they could be utilized as source germplasm in cultivar development.

Key words: SSR markers, soybean**,** phenotypic traits, PIC, genetic diversity

2.2 Introduction

Soybean [*Glycine max* (L.) Merrill] has food security, nutritional, medicinal and economic value with high industrial potential, which can be exploited by developing economies in Sub-Saharan Africa. The long history of soybean breeding and cultivation in Zimbabwe has contributed to the evolution of the crop. As such, thousands of breeding lines and numerous cultivars have been developed for exploitation in the breeding programme and cropping systems. However, the genetic base or diversity of soybean germplasm could be under threat due to the frequent use of the same genetic resources resulting in narrowing of the genetic base (Reyna and Sneller, 2001). Dayaman *et al.* (2009) also asserted that narrow spectrum of diversity existed among soybean cultivars. In the same vein, Gizlice *et al.* (1994) and Vello *et al.* (1988) reported about the narrowness of the North American and Brazilian germplasm. Through pedigree analysis Gizlice *et al.* (1994) observed that 35 ancestors contributed more than 95% of the alleles and only five lines accounted for more than 55% of the genetic background of public cultivars in North America. Furthermore, Gai *et al.* (2001) noted that out a total of 308 ancestral varieties only 38 accounted for 54.18 and 56.84% of nuclear and cytoplasmic genetic material of the 651 soybean cultivars released from 1923 to in China. Undoubtedly, intensive breeding and selection may contribute to reduced diversity. Arguably, remarkable success of any breeding programme depends on the availability and extent of diversity in the germplasm pool and choice of parental stock as well as the selection method being used. Clearly, broad diversity coupled with a collection of favourable alleles for traits of economic importance, result in significant breeding gains ultimately leading to enhanced food production (Khodadadi *et al.*, 2011).

It is critical for plant breeders to have a full understanding of the genetic diversity of the crop they are dealing with (Priolli *et al.*, 2010). Generally, the investigation of phenotypic and genetic diversity is important to reveal similar genetic backgrounds, better understand the evolutionary relationships among accessions, identification of diverse parental combinations to develop segregating progenies with maximum genetic variability for further selection, revealing duplication in germplasm, introgression of desirable genes or chromosome segments from diverse sources into elite germplasm, management of core collections and determination of uniqueness and distinctness of the phenotypic and genetic constitution of genotypes with the purpose of protecting the breeder's intellectual property rights (Franco *et al.*, 2001; Li *et al.*, 2008; Ojo *et al.*, 2012). The main issue is that parental selection is the initial step in crop breeding and genetic variability defines potential for improvement and breeding efficiency. Khodadadi *et al.* (2011) reported that genetic divergence is critical for the accomplishment of

transgressive segregation. Contextually, this suggests that allelic diversity and combinations made from such parental stock raises the mean performance of the base population. Knowledge of genetic diversity may demand the breeder to expand the genetic base thereby introducing new and unique genes into the existing gene pool (Narvel *et al.*, 2000).

There are several approaches that can be applied to characterise cultivars, accessions or elite lines in genetic diversity studies viz; phenotypic characterization, agronomic traits, geographic origins, biochemical methods, coefficient of parentage and molecular markers (Li *et al.*, 2001; Dayaman *et al.*, 2009; Priolli *et al.*, 2010; Liu *et al.*, 2011; Tantasawat *et al.*, 2011). The coefficient of parentage is generally limited by incomplete data as well possible errors associated with the pedigree information and origins of accessions (Tantasawat *et al.*, 2011). Phenotypic traits are traditionally the most widely used and their popularity is premised on their simple, speed and inexpensive nature (Bretting and Widrlechner, 1995). Liu *et al.* (2011) reported that morphological and agronomic traits are useful in estimating genetic diversity because of their visualization and ease of attaining. Irrespective of their limitation due to environmental influence, their application is still valid for farmers, breeders and curators coupled with registration and release and variety protection (Hershey and Ocampo, 1989; Elias *et al.*, 2001.) Earlier studies demonstrated the utility and significance of phenotypic characterisation in assigning soybean genotypes to well differentiated clusters (Liu *et al.*, 2011; Malik *et al.*, 2011; Shadakshari *et al.*, 2011). The observed phenotypic variations demonstrated that selection could be undertaken on the basis of the assessed traits. Using nine phenotypic characters to evaluate the genetic diversity of 19 soybean genotypes under drought stress, Salimi *et al.* (2012) obtained seven clusters. They also found significant differences in 100-seed weight, days to 50% flowering, days to maturity and grain yield. Thus, these traits could be useful in selecting for drought stress and the identified germplasm thereof could be useful sources of drought stress tolerance. These results demonstrate that phenotypic characterisation helps plant breeders to identify traits that can be used for breeding purposes (Singh, 1989).

The advent of molecular markers permits easy assessment of a considerable number of loci distributed throughout the genome of plants (Chakravanthi and Naravaneni, 2006). The commonly used DNA markers in diversity studies are; Restriction Fragment Length polymorphism (RFLP), Amplified Fragment Length polymorphism (AFLP), Random Amplified Polymorphism (RAPD), Simple Sequence Repeats (SSR) and SNPs. However, in this study, the SSR markers were considered as the markers of choice because of their abundance, the high

level of polymorphism, they are multiallelic and hyper variable, appear to be randomly and uniformly distributed throughout the eukaryote genomes, are simple and easy to use, are accessible to other research laboratories via published primer sequences, are dominant, require PCR based detection, are adaptable to automation and relatively inexpensive (Yu *et al.*, 2000; He *et al.*, 2003; Semagn *et al.*, 2006). Other studies have reported high levels of polymorphism using SSR markers (Narvel *et al.*, 2000; Hudcovicova and Karaic, 2003). Interestingly, Powell *et al.* (1996) investigated the discriminating capacity and effectiveness of various markers in soybean germplasm and observed that the SSR markers had the highest expected heterozygosity (information content). Although previous studies have also shown that there is a high correlation between RFLPs, AFLPs and SSRs, the SSRs were found to generate hyper variable polymorphisms (Rongwen *et al.*, 1995). Clearly, this demonstrates that the SSR markers are a powerful and useful tool in genetic diversity studies. Previous genetic diversity studies on soybean using SSR markers have shown broad genetic variation (Wang *et al.*, 2006; Kuroda *et al.*, 2009; Mulato *et al.*, 2010; Zhang *et al.*, 2013).

However, intensive selection may reduce the genetic base which has serious consequences for future breeding progress. Breeding strategies that emphasise the mating of elite strains can result in recombining genes contributed by a limited number of ancestral introductions leading to narrowing of the genetic base. This could be aggravated by low levels of diversity that naturally exist in the soybean crop (Dayaman *et al.*, 2009). In the same vein, Sandoval *et al.* (1997) observed that genetic variation for grain yield coupled with other agronomic characters in soybean populations derived from crosses among elite lines can be limited by lack of genetic diversity. Consequently, the germplasm becomes vulnerable to environmental challenges such as biotic and abiotic stresses. As a result of global climate change, there are bound to be more challenges that will impact on soybean breeding progress unless a diversified germplasm is maintained.

The soybean programme in Zimbabwe is over 70 years old; therefore, it was found prudent to investigate the available germplasm that is within the germplasm pool. Further, soybean was declared a strategic crop because of its diverse utility but sadly there is a perennial shortage of the crop, suggesting that higher yielding cultivars should be made available. Grain yield could be enhanced by exploiting the available and well adapted genetic resources through effective and efficient breeding and selection strategies. Therefore, characterisation at both phenotypic and molecular levels would provide more information about the degree of diversity, genetic

constitution of the available germplasm and would also reveal important traits to breeders resulting in optimum utilisation of the germplasm. Genetic variation among traits is crucial for breeding and selection (Malik *et al.*, 2011). In the same breath, Terron *et al.* (1997) indicated that classification of elite breeding lines into well-defined and distinct genetic groupings minimizes chances of generating and testing undesirable crosses. On the other hand, there is lack of information on the genetic variation of the available germplasm in Zimbabwe. Studies on soybean genetic diversity studies compared diversity for germplasm obtained elsewhere. Essentially, the characterisation of the germplasm generated over such a long period of time would be critical to identify and correctly interpret the existing genetic relationships (Pritchard and Rosenberg, 1999; Buckler and Thornsberry, 2002).

Given the forgoing, the objective of this study was (i) to evaluate and compare the genetic diversity estimates of 42 soybean genotypes as determined by phenotypic traits and SSR markers and (ii) assess parental potential of these genotypes in cultivar breeding and development.

2.3 Materials and methods

2.3.1 Phenotypic characterisation

2.3.1.1 Germplasm

A total of 42 soybean genotypes were evaluated for their genetic diversity and relatedness. These are as shown in Table 2.1. This sample represented a collection of all the cultivars that were introduced, developed and released in Zimbabwe from 1940 to 2013. Some these cultivars were also registered in Malawi and Zambia. The germplasm included all the cultivars from Crop Breeding Institute of Zimbabwe, Pannar Seed in Harare, Zimbabwe and Seed Company of Zimbabwe. These cultivars were all registered under the Second Schedule of the Certification of Crops in Zimbabwe.

Table 2.1: List of genotypes used in the study

Number	Code Name	Origin†	Year of Release	Growth Habit
1	G41	South Africa	1966	I
2	G40	United States	1972	D
3	G39	Zimbabwe	1973	D
4	G38	Zimbabwe	1974	D
5	G37	Zimbabwe	1977	I
6	G36	Zimbabwe	1977	D
7	G35	Zimbabwe	1980	I
8	G34	Zimbabwe	1982	I
9	G33	Zimbabwe	1985	D
10	G32	Zimbabwe	1988	I
11	G28	Zimbabwe	1989	I
12	G31	Zimbabwe	1992	D
13	G30	Zimbabwe	1992	I
14	G29	Zimbabwe	1995	D
15	G27	Zimbabwe	1994	D
16	G26	Zimbabwe	1997	D
17	G6	Zimbabwe	1997	I
18	G25	Zimbabwe	1998	I
19	G24	Zimbabwe	1999	D
20	G23	Zimbabwe	1999	D
21	G22	Zimbabwe	1999	D
22	G3	Zimbabwe	1999	I
23	G21	Zimbabwe	2000	I
24	G20	Zimbabwe	2000	D
25	G19	Zimbabwe	2001	I
26	G18	Zimbabwe	2003	I
27	G17	Zimbabwe	2005	D
28	G16	Zimbabwe	2005	D
29	G4	Zimbabwe	2005	I
30	G15	Zimbabwe	2006	I
31	G14	Zimbabwe	2007	I
32	G13	Zimbabwe	2007	I
33	G12	Zimbabwe	2007	D
34	G2	Zimbabwe	2008	D
35	G1	Zimbabwe	2008	I
36	G11	Zimbabwe	2008	I
37	G10	Zimbabwe	2008	I
38	G9	Zimbabwe	2010	D
39	G8	Zimbabwe	2012	1
40	G7	Zimbabwe	2012	D
41	G42	Zimbabwe	2012	D
42	G5	Zimbabwe	2013	D

† = Geographic origin; D = Determinate, I = Indeterminate, G = genotype

2.3.1.2 Experimental design, sites and management

The experimental study was conducted at 13 test sites during 2010/2011 and 2011/12 seasons. Site details are as given in, Table 2.2 below. A 6 x 7 row-column rectangular lattice design with three replications was used across the two cropping seasons. The gross plot was six rows, 45 cm apart and five metres long while the net plot was four rows, 45 cm apart and 4.4 m long giving a nett area of 7.92 m^2. A total of 79 viable seeds were planted per row resulting in a plant population of approximately 350 000 plants ha^{-1}. The planting dates are as shown in Table 2.2 below. Planting was done by hand. A basal fertilizer (Cotton Fert) was applied at a rate of 400 kg ha^{-1} supplying 28 kg ha^{-1} of Nitrogen, 68 kg ha^{-1} of P_20_5 and 40 kg ha^{-1} of K_20. The seed was inoculated with *Bradrhizobium japonicum* inoculant Grasslands strain 1491. Herbicides, (Lasso and Gramoxone) were applied at 4 l ha^{-1} and 1 l ha^{-1} respectively as pre-emergency sprays. The two are compatible so they were mixed and applied simultaneously. Where irrigation facilities were available, supplementary irrigation was applied to the crop in times of need. The trials were protected against soybean using a fungicide known as Shavit 25 EC (Tridimefon) at a rate of 500ml ha^{-1}. Shavit was used because it controls rust only, providing us the opportunity to see the reaction of the genotypes to other diseases. Three fungicide applications were done with the first application at 50 days after planting or at flowering and second and third applications were 20 days apart, at 70 and 90 days after planting to provide protection at pod fill (Levy, personal communication). All the trials were hand harvested.

Phenotypic data were recorded from the net plot for pod height, plant heights, % lodged plants at maturity, 100 seed weight, percentage crude protein, seed appearance scores, bacterial blight scores, red leaf blotch scores, downy mildew scores, percentage crude oil, days to 50% flowering, days to 95% pod maturity, days from 95% pod maturity to first pod shattering and grain yield. Grain yield was later adjusted to kg ha^{-1} at 11% moisture following standard practice used at Seed Co (Seed Co Research and Technology, personal communication) using the following formulae;

*Grain Yield (kg ha^{-1}) = [Grain Weight (Plot yield in kg ha^{-1})/(100 - %MC) *10/Plot Area*111/100]*
Where; %MC = Grain Moisture in percentage.

N.B. The disease rating scale of 1-9 was adapted from international rating scale used for patent and cultivar registrations (http://www.google.com/patentsUS8378178 accessed on 10 October 2010).

Table 2.2: Environments used for evaluations of the test entries during 2010/11 and 2011/12 cropping seasons

Environment	Country	Season	Code	Latitude	Longitude	Altitude (masl)	Planting Date	Rainfall[1] (mm)
RARS	Zimbabwe	2010/11	E1	17°40'S	31°14'E	1341	10-12-10	686
GVTC	Zimbabwe	2010/11	E2	17°68'S	30°86'E	1449	12-12-10	712
Lusaka	Zambia	2010/11	E3	15°67'S	28°33'E	1300	08-12-10	860
Mpongwe	Zambia	2010/11	E4	13°59'S	28°00'	1219	05-12-10	1000
Bvumbwe	Malawi	2010/11	E5	15°55'S	35°04'E	1228	15-12-10	950
RARS	Zimbabwe	2011/12	E6	17°40'S	31°14'E	1341	06-12-11	749
GVTC	Zimbabwe	2011/12	E7	17°68'S	30°86'E	1449	10-12-11	712
Lusaka	Zambia	2011/12	E8	15°67'S	28°33E	1300	13-12-11	700
Mpongwe	Zambia	2011/12	E9	13°59'S	28°00'	1199	10-12-11	800
Bvumbwe	Malawi	2011/12	E10	15°55'S	35°04'E	1250	12-12-11	768
Lilayi	Zambia	2011/12	E11	15°33'S	28°30'E	1090	01-12-11	688
ART	Zimabwe	2011/12	E12	17°43'S	31°05E	1527	27-11-10	780
Chitedze	Malawi	2011/12	E13	13°85'S	33°85'	1146	14-12-11	643

[1]rainfall refers to the amount received during the growing period including irrigation, RARS = Rattray Arnold Research Station, ART = Agricultural Research Trust, GVTC = Gwebi Variety Testing Centre; masl = metres above sea level; mm = millimetres

2.3.2 Molecular characterisation

2.3.2.1 Plant material and DNA extraction

The study used the same genotypes as during phenotypic characterisation and these were grown in pots in the green house in winter of 2011. These were not replicated. Each pot had four plants and DNA was extracted from these plants. Fresh leaf tissue was harvested from the young leaves (four weeks after planting) from each plant, hence, four leaf discs were sampled and bulked together to represent each genotype. The four sampled leaf discs were put into a single hole of the 96-well block or plate. After sampling all the 42 genotypes, the entire block or plate was sealed with air pore tape and then the block was placed into a plastic bag together with 50 g of silica gel meant to dry the leaf discs for 48 hours. The indicator silica gel turned blue, symbolizing that the sampled leaves had been dehydrated. The extracted DNA was then sent for profiling in the laboratory in Canada. DNA Landmarks was preferred because it was cheaper than other organisations. It is important to report that the protocol that was used for DNA extraction was supplied by DNA Landmarks.

2.3.2.2 Simple sequence repeats primer selection

A total of 30 SSR markers were used to genotype the lines (Table 2.3). These were chosen for their distribution across the soybean genome and amplification quality (Cregan *et al.*, 1999). The number of markers to use was largely a function of cost. However, Guichoux *et al.* (2011) asserted that 10 to 30 highly polymorphic markers would suffice to provide quality or precise data.

Table 2.3: 30 soybean SSR primer sequences used for genotyping

Primer Name	Forward 5' ⟶3	Reverse 3' ⟶5
Satt012	GCAATTAGTTTTAAAATGTTTC	AGAATAGAGCCTACATATAATCATA
Satt148	AATCCGGGACGCAAAATTATTATTAA	TGCAAATTCCCTAATTAACACCCTTTATAC
Satt156	CGCACCCCTCATCCTATGTA	CCAACTAATCCCAGGGACTTACTT
Satt172	AGCCTCCGGTATCACAG	CCTCCTTTCTCCCATTTT
Satt180	TCGCGTTTGTCAGC	TTGATTGAAACCCAACTA
Satt182	GGTCCACATGAAATGAAGGT	TCTCAGCCTGCAAAGAAAA
Satt184	GCGCTATGTAGATTATCCAAATTACGC	GCCACTTACTGTTACTCAT
Satt215	GCGCCTTCTTCTGCTAAATCA	CCCATTCAATTGAGATCCAAAATTAC
Satt242	GCGTTGATCAGGTCGATTTTTATTTGT	GCGAGTGCCAACTAACTACTTTTATGA
Satt294	GCGGGTCAAATGCAAATTATTTTT	GCGCTCAGTGTGAAAGTTGTTTCTAT
Satt372	CAGAAAAGGAATAATAACAACATCAC	GCGAAAACATAATTCACACAAAAGACAG
Satt387	GCGTTACGTTTCACTATTTATTTAACAT	GCGGCAGGCTAGCTACATCAAGAG
Satt394	GCGTTTTTTCAATTTAAAGAGAATTGAC	GCGTAACTTGCATGTGGTATATCGAGATG
Satt397	TCTCGGGATCCTTGTTAGAT	GCGAAGAAGAAGAGAACATGTGAA
Satt414	GCGTATTCCTAGTCACATGCTATTTCA	GCGTCATAATAATGCCTAGAACATAAA
Satt429	GCGACCATCATCTAATCACAATCTACTA	TCCCCATCATTTATCGAAAATAATAATT
Satt434	GCGTTCCGATATACTATATAATCCTAAT	GCGGGGTTAGTCTTTTTATTTAACTTAA
Satt441	AAACCCACCCTCAAAAATAAAAA	AAATGCACCCATCAATCACA
Satt459	TCGTGTTAGATTTTTACTGTCACATT	AACTGCATACCCTTTGTTTGAA
Satt477	GTTGGGAAAAGGTTACTACCATATC	GGTCCGTATGCAATTCTTGACTAATA
Satt490	GCGGCACGAGTCAACTTTCTGTTTCCT	GCGGAAGAAGATTTTCGTTTTTAT
Satt509	GCGCTACCGTGTGGTGGTGTGCTACCT	GCGCAAGTGGCCAGCTCATCTATT
Satt511	GCGACTTTACTGAAAACCTGGAAA	GCTTCAAACCAACAAACAACTTA
Satt522	GCGAACTGCCTAGGTTAAAA	TTAGGCGAAATCAACAAT
Satt530	CATGCATATTGACTTCATTATT	CCAAGCGGGTGAAGAGGTTTTT
Satt577	CAAGCTTAAGTCTTGGTCTTCTCT	GGCCTGACCCAAAACTAAGGGAAGTG
Satt590	GCGCGCATTTTTTAAGTTAATGTTCT	GCGCGAGTTAGCGAATTATTTGTC
Satt598	CGATTTGAATATACTTACCGTCTATA	CACAATACCTGTGGCTGTTATACTAT
Sct_034	AATTCTCACTCTCACAACTTC	CCATGGGAATAGTTGGGT
Sct_067	CTCCCCATCTCTCTCTAAC	GGATTTTGTTATTTATTTATTTGA

2.3.2.3 PCR amplification and detection

The DNA concentrations were measured using Hoechst dye and the quality of the DNA samples were checked on a 0.8% agarose gel. After passing the quality control, the DNA samples were then used for polymerase chain reaction (PCR) amplification with 30 SSR markers. Polymerase chain reaction (PCR) amplification reaction were in a total volume of 1µl containing 0.40µl of 2.5mM $MgCl_{2,}$ 1.50µl of 10X assay buffer, 1.00µl of 2mM dNTPs, 0,10µ of 5 units/µl Taq polymerase, 0.06µl of 20µM forward and 0.06µl of 20µM reverse of primers. Amplification was performed in PTC Thermal Cycle 100 (MJ Research Inc., 1987) programmed for an initial denaturation of 94°C for three minutes, followed by 35 cycles of 1 minute denaturation at 94°C, one minute, annealing at 49°C and extension of minute at 72°C. Final extension was done to a period of 30 minutes at 72°C and the product was then stored at 4°C. Given that the SSR markers are co-dominant, the data was then scored using a scoring scale of -1 to 1 where, -1 denoted missing alleles, 0 represented allele absence and one allele presence.

2.3.3 Data analysis

2.3.3.1 Phenotypic data analysis

The phenotypic data was analysed in GenSat 16th edition (Goedhart and Thissen, 2013). The principal component analysis was performed using all the phenotypic traits that were measured using the same package. Furthermore, the phenotypic dendrogram was also constructed using the GenStat 16th Edition. Ideally, the phenotypic data was analysed to determine the means, range, coefficient of variation and standard deviation.

2.3.3.2 SSR analysis

The binary data matrix was used to calculate the genetic similarity matrix using Dice Similarity Coefficient (Dice, 1945) with the help of the Numerical Taxonomy Multivariate Analysis System for personal computer (NTSYS-pc) version 2.1 (Rolf, 1998). The resultant similarity distance matrix data was used to construct a dendrogram using the agglomerative hierarchical un-weighted pair-group method with an arithmetic average (UPGMA) sub-programme of NTSYS-pc (Sokal and Michener, 1958; Rohlf, 1998). Powermarker V3.25 (Liu and Muse, 2005) was used to determine major allele frequency, gene diversity, observed heterozygosity and polymorphic information content (PIC) values for each SSR marker used in the study. The expected heterozygosity (He) and observed heterozygosity (Ho) were used to

evaluate the genetic diversity within the set of genotypes. Expected heterozygosity, i.e. the probability that two alleles from the same locus would be different when selected at random was estimated for each SSR locus according to (Nei, 1973);

$He = 1 - \sum(pi)^2$

Where $\sum$ stands for summation over all alleles; pi is the frequency of the i^{th} allele at a locus for individual p

Observed heterozygosity was estimated by dividing the number of heterozygous individuals by the number of individuals scored. Polymorphic information content for the SSR markers in the sample DNA was calculated as follows;

$PIC = 1 - \sum(p_i)^2$ where p_i is the frequency of the i^{th} allele in a locus for individual p

2.4 Results

2.4.1 Genetic analysis at phenotypic level

2.4.1.1 Phenotypic variation among the traits assessed

Table 2.4 presents means of the traits that were studied. The germplasm revealed highest variability on percentage lodged plants at maturity, followed by the number of days from 95% pod maturity to first pod shattering, grain yield and plant height. High CVs were observed on lodging (319%) days from 95% pod maturity to first pod shattering (16%) and plant height (8.8%). However, among the investigated traits, the quality traits exhibited the lowest variation. The percentage protein and oil in the seed on a dry matter basis was 2% and 3% respectively. The phenological traits (i.e. maturity and flowering) registered CV values of 2%. The extent of variability for grain yield ranged from 2723 kg ha^{-1} to 4823 kg ha^{-1}. Narrow range of variability was observed for protein and oil with values ranging from 38.65 to 40.9% and 16.35 to 17.7% respectively. The study showed that the extent of variability was maximum for percentage lodged plants at maturity (0% to 28%).

In addition, the distribution shown by the histograms (Figure 2.1) revealed variability among the traits. The results also exhibited that oil content; pod height and number days from 95% pod maturity to first pod shattering were normally distributed. Number of days from planting to 50% flowering, 100 seeds weight and percentage lodged plants at maturity were negatively skewed

while plant height, grain yield, percentage protein in the seed on a dry matter basis, and number of days from planting to maturity 95% pod maturity were positively skewed.

Table 2.4: Summary of statistical parameters for 10 quantitative traits of 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe over three consecutive cropping seasons starting in 2010/2011

Trait	Mean	±SE Mean	CV (%)	Min	Max	Variance	Probability
Pod height (cm)	14.0	1.27	8.8	11	18	3.41	**
Plant height (cm)	87.0	5.30	6.1	73	98	6.76	**
Percentage lodged plants	1.8	5.80	319.1	0	28	1.59	ns
100 seed weight (g)	22.0	1.75	8.1	19	26	2.94	**
Percentage oil in the seed	16.93	0.49	3.0	16.4	17.7	1.0	ns
Percentage protein in the seed	39.8	0.82	2.0	38.7	41.25	0.84	ns
Days to 50% flowering	53.0	0.86	2.0	48	64	28.70	**
Days to 95%pod maturity	125.0	1.83	2.0	110	135	13.73	**
Days to first pod shattering	30.0	4.95	16.0	20	40	2.52	**
Grain Yield (kg ha^{-1})	4200	266.2	6.0	2723	4823	6.75	**

**, NS, Significant at P≤0.01 and NS = not significant respectively; SE Mean = Standard error of the mean, Min = Minimum, Max = Maximum, CV (%) Coefficient variation; cm = centimetres; g = grams, kg ha^{-1} kilograms per hectare

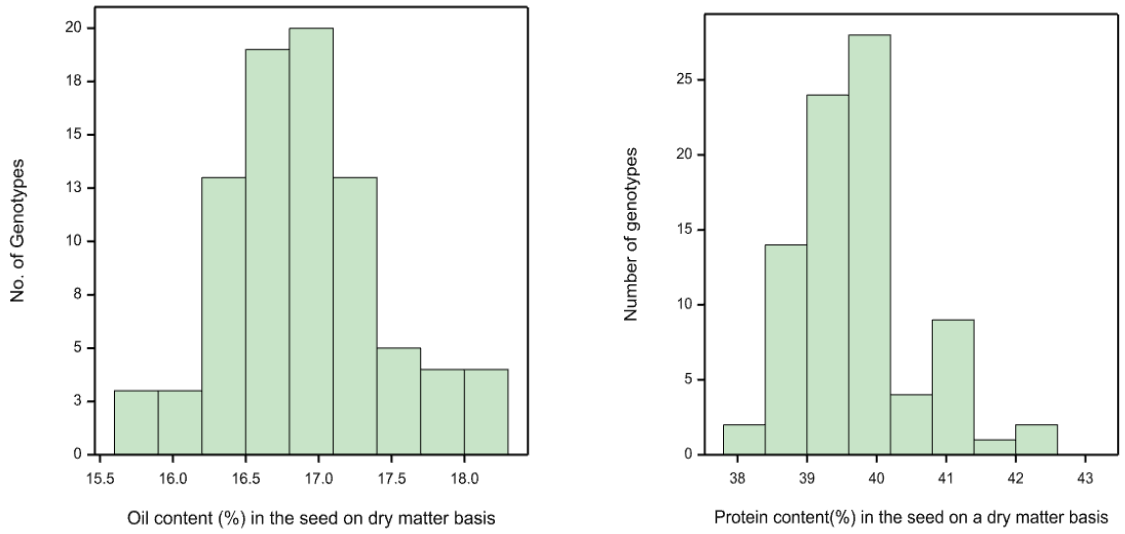

Figure 2.1a: The frequency distribution of percentage oil (left) and protein (right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2010/2011 and 2011/2012 cropping seasons

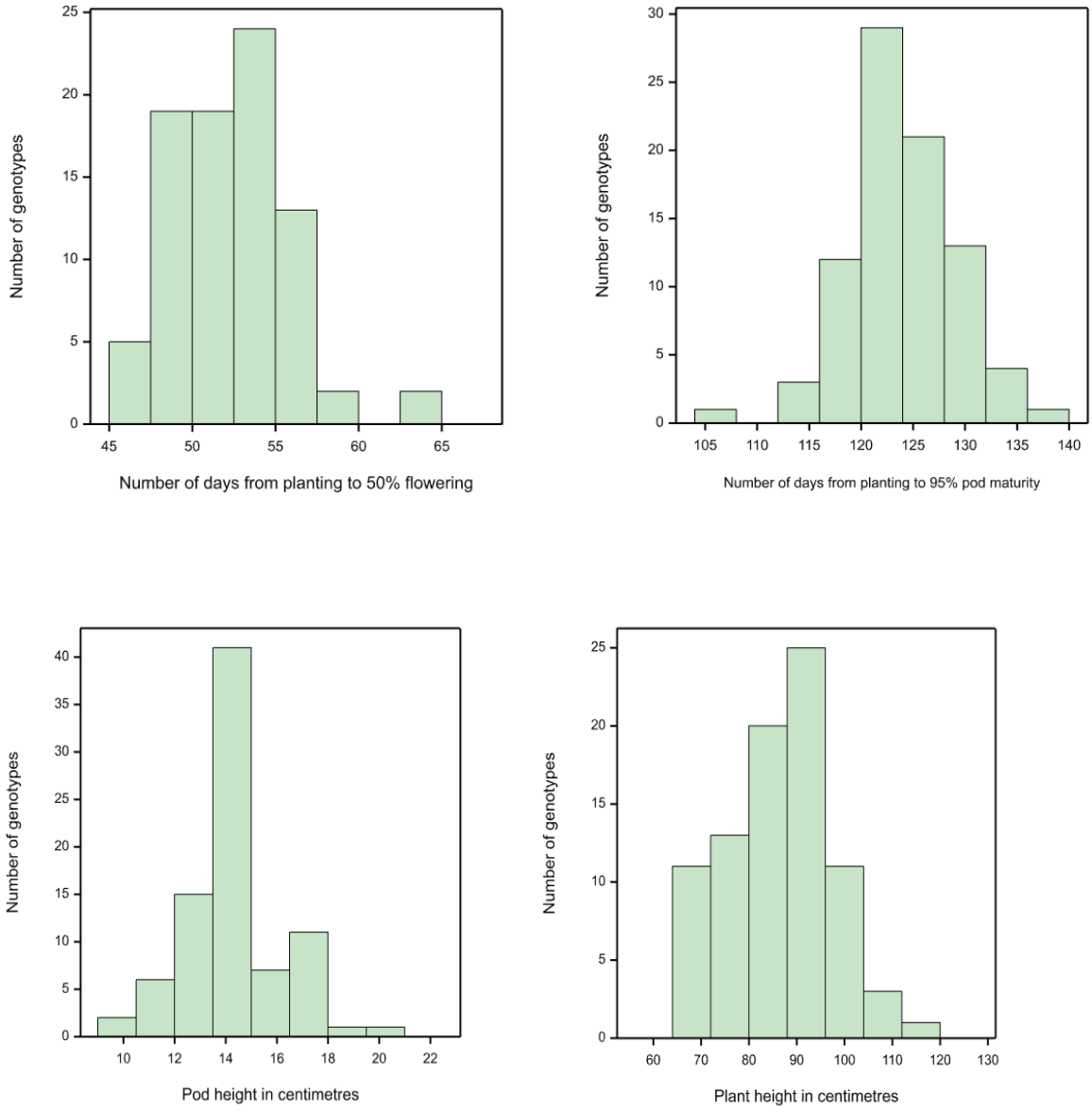

Figure 2.1b: The frequency distribution of number of days from planting to 50% flowering (top left), number of days from planting to 95% pod maturity (top right), pod height (bottom left) and plant height (bottom right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2010/2011 and 2011/2012 cropping seasons

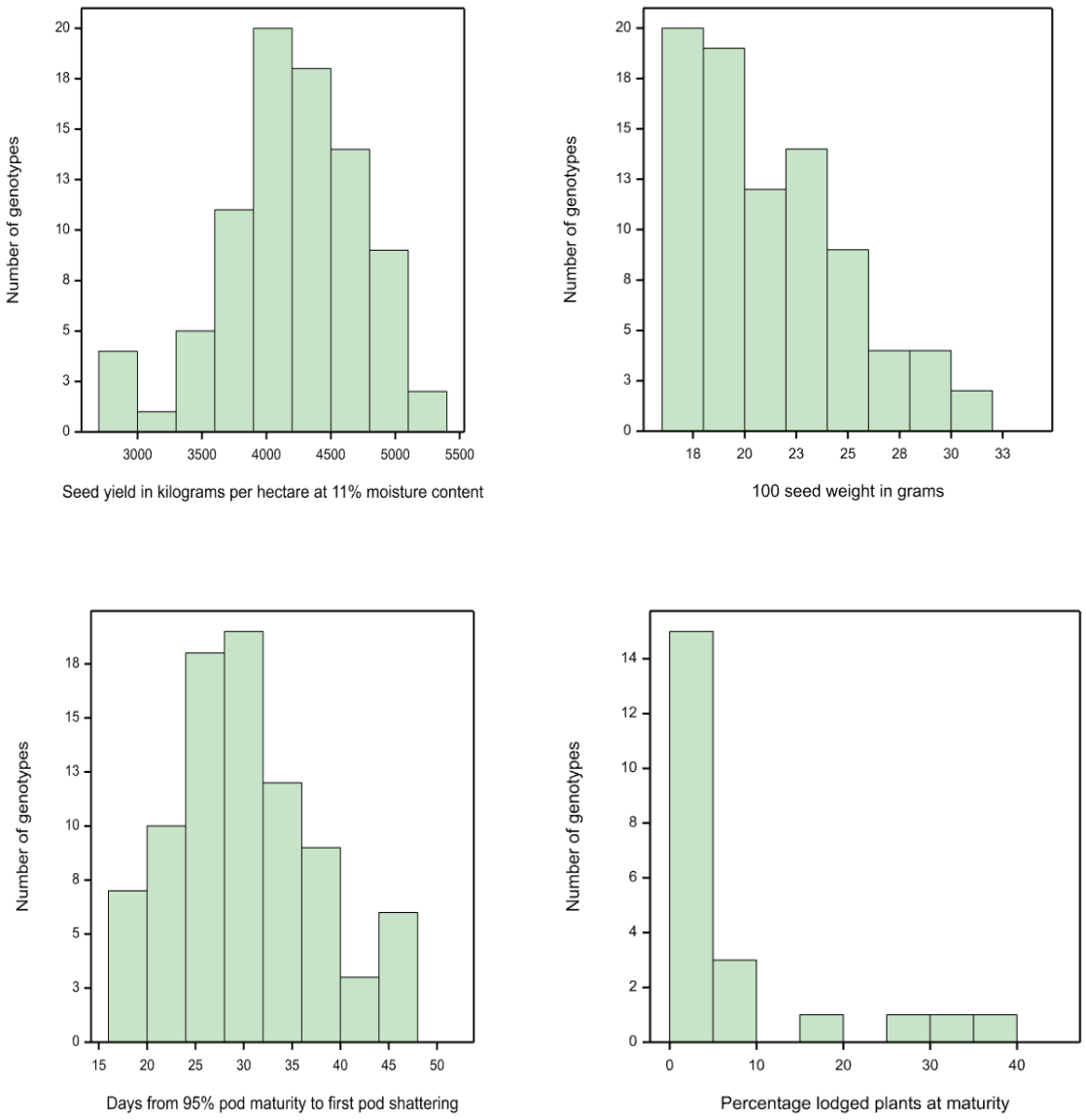

Figure 2.1c: The frequency distribution of seed yield in kg ha^{-1} at 11% moisture (top left), 100 seed weight in grams (top right), days from 95% pod maturity to first pod shattering (bottom left) and percentage lodged plants at maturity (bottom right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2010/2011 and 2011/2012 cropping seasons

2.4.1.2 Phenotypic cluster analysis

The clustering of soybean genotypes based on the variations across phenotypic traits defined two major clusters, A and B (Figure 2.1). Further division of these two clusters resulted in eight clusters at 0.87 genetic similarity. Sub-cluster **I** comprised of one genotype, G41 at a genetic similarity coefficient of 0.73. It was introduced from the Republic of South Africa. Sub-cluster **II** also contained one genotype (G31). Sub-cluster **III** consisted of seven genotypes (G38, G33, G3, G9 and G29), accounting for 16.7% of the total number of genotypes. Sub-cluster **IV** consisted of six genotypes which accounted for 14.3% of the total number of genotypes. Sub-cluster **V** contained five genotypes standing for 11.9% of the total population. Sub-cluster **VI** was composed of two genotypes which represented 4.8% of total genotypes. Sub-cluster **VII** comprised of eight genotypes which accounted for 19% while sub-cluster **VIII** was composed of 12 genotypes representing 28.6% of the total population. Generally, results revealed the pattern of registration and commercialization of the genotypes. Sub-cluster **VIII** consisted of genotypes (G1, G13, G14, G19, G21 and G6) which were all registered in Zambia between 2002 to 2007. Furthermore, Sub-clusters **1** to **V** consisted of genotypes that were released before 2000 whereas Sub-clusters **VI** to **VIII** contained all the genotypes that were registered and commercialized after 2000.

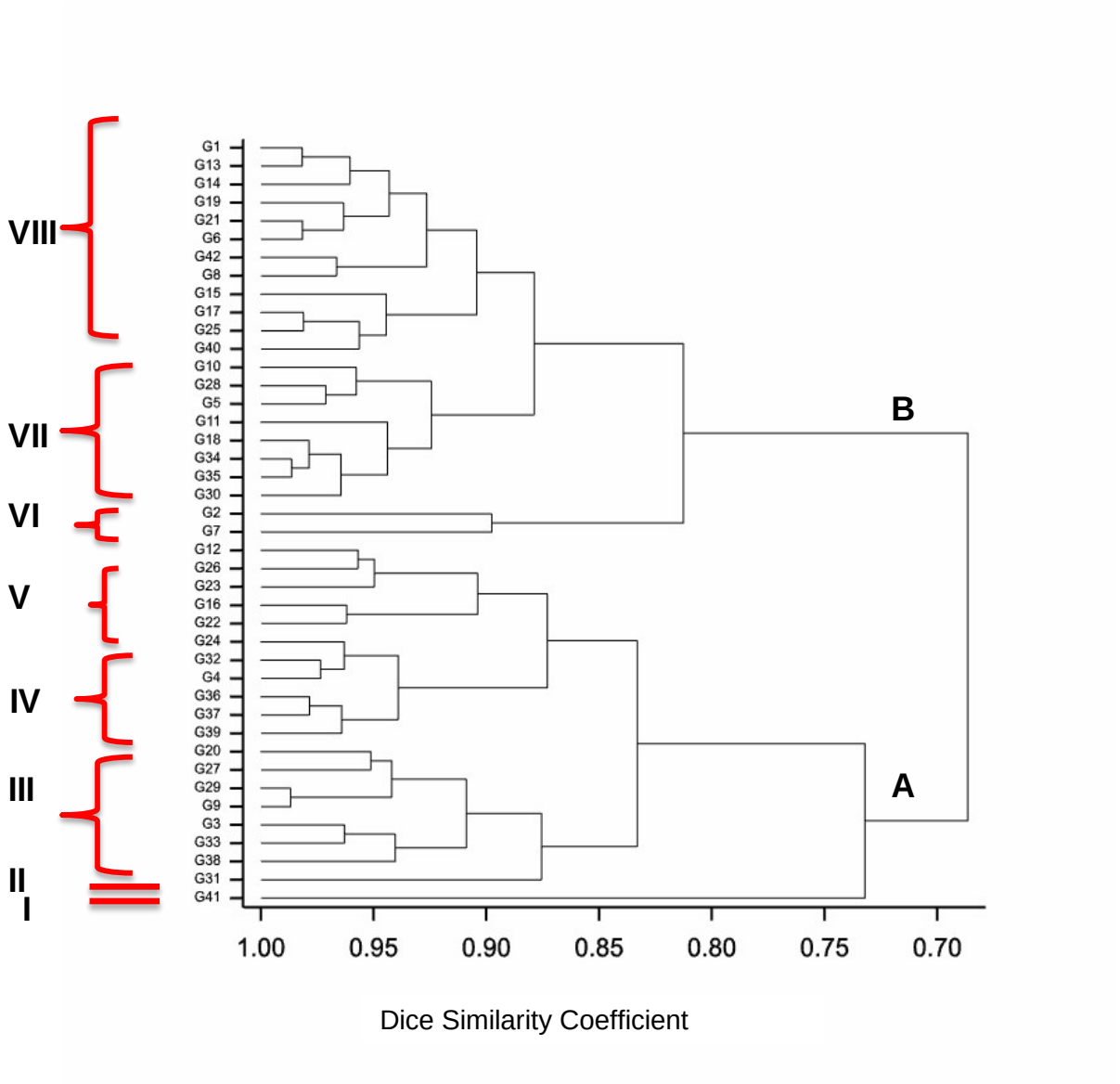

Figure 2.2: Dendrogram for phenotypic characterization of 42 analysed soybean germplasm using GenStat 16th, Edition. The eight clusters among the genotypes are denoted from I to VIII. Cluster A and B represented the two major clusters

2.4.1.3 Principal component analysis

Principal component analysis was performed using all the phenotypic traits and the score plot based on the first two principal components is presented in Figure 2.3. The first two principal components accounted for 52.0% of the total variation. The first principal component (PC1) accounted for 34.9% and had high contributing factor loadings (weights) from bacterial blight scores, red leaf blotch scores and downy mildew scores. The second principal component

contributed 17.2% to the total variance with high contributing factor loadings from percentage protein, 100 seed weight, grain yield and number of days to first pod shattering

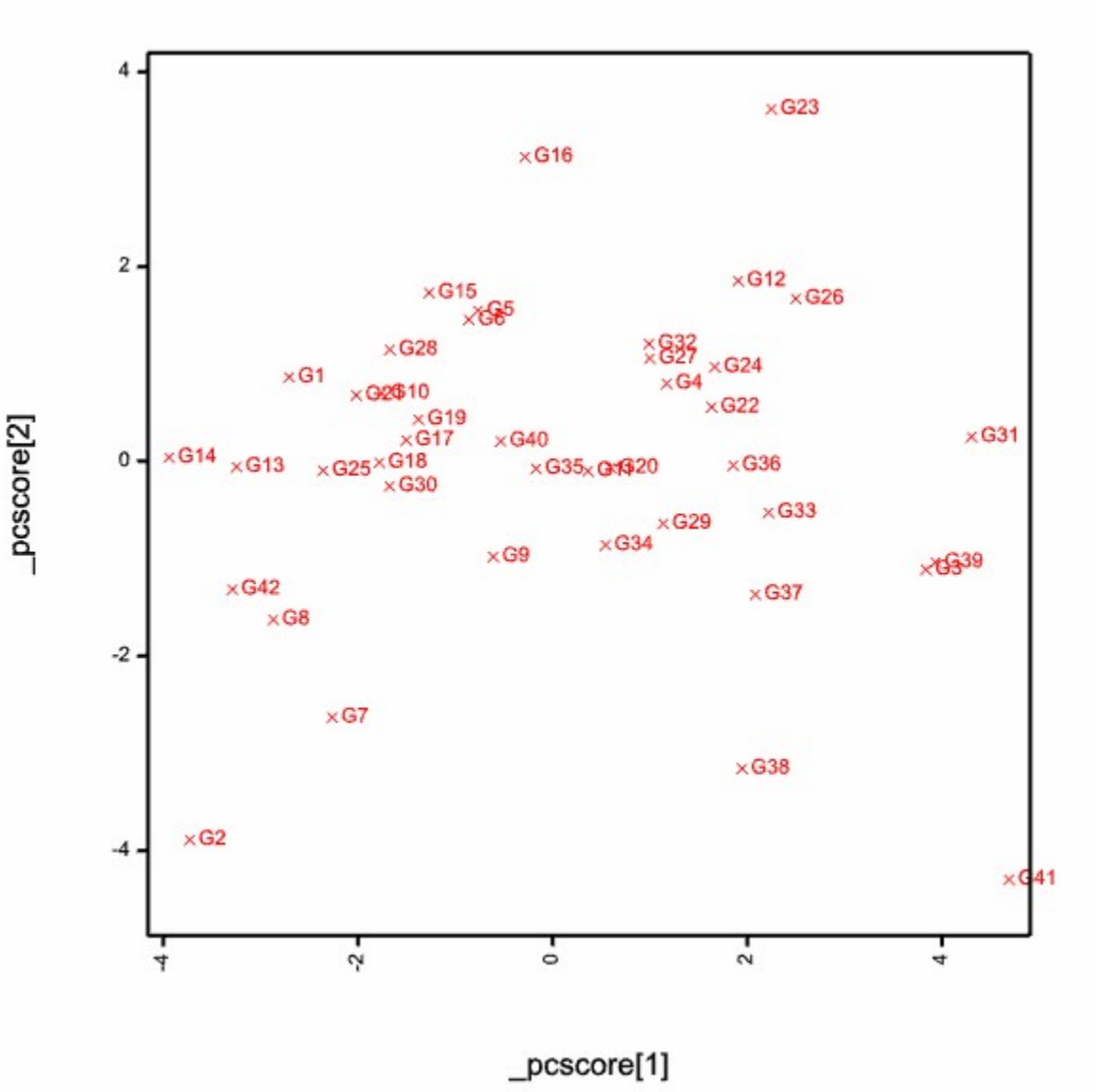

Figure 2.3: Distribution of 42 soybean genotypes as revealed by first two PCA analysis based on morphological data. Genotypes are coded from xG1 to xG42 as shown in plots where xG1 corresponds to G41 (the pre-fix “x” is slotted in by the software.

2.4.2 Genetic diversity at molecular level

2.4.2.1 Number of calls and call rate

A high call rate was generally observed for the markers used (Table 2.5).

Table 2.5: Summary of the number of calls and call rate for each marker

Name	Number of calls	Call rate
Satt012	49	100%
Satt148	48	98%
Satt156	49	100%
Satt172	49	100%
Satt180	49	100%
Satt182	49	100%
Satt184	49	100%
Satt215	49	100%
Satt242	49	100%
Satt294	49	100%
Satt372	49	100%
Satt387	49	100%
Satt394	49	100%
Satt397	49	100%
Satt414	49	100%
Satt429	49	100%
Satt434	49	100%
Satt441	49	100%
Satt459	49	100%
Satt477	49	100%
Satt490	49	100%
Satt509	49	100%
Satt511	49	100%
Satt522	49	100%
Satt530	49	100%
Satt577	49	100%
Satt590	49	100%
Satt598	49	100%
Sct_034	49	100%
Sct_067	49	100%

2.4.2.2 Polymorphism results

A total of 135 alleles were recorded and the number of alleles scored per locus ranged from two to nine, with a mean of 4.57 alleles per locus (Table 2.6). In addition, the gene diversity ranged from 0.0906 to 0.7937 with an average of 0.50 for the genotypes studied. The PIC estimated for all loci ranged between 0.1563 and 0.7669 with an average of 0.45. Results also showed that the means for the major allele frequency and heterozygosity (Ho) were 0.6252 and 0.1133 respectively. Among the primers that were used, Satt012, Satt414 and Satt372 were highly informative with PIC values of 0.76, 0.71% and 0.68% respectively.

Table 2.6: Details of polymorphisms and genetic analysis of 42 soybean genotypes using 30 SSR markers

Marker	Major Allele Frequency	Allele No.	Gene Diversity	Expected Heterozygosity (Ho)	PIC
Satt394	0.8491	5.0000	0.2727	0.0000	0.2634
Satt577	0.6792	3.0000	0.4656	0.1509	0.3979
Satt180	0.7358	5.0000	0.4313	0.0755	0.4016
Satt459	0.7642	3.0000	0.3847	0.0566	0.3477
Satt522	0.4434	4.0000	0.6517	0.1887	0.5859
Satt598	0.6698	2.0000	0.4423	0.1321	0.3445
Satt242	0.5566	5.0000	0.5675	0.1698	0.4897
Satt429	0.3962	6.0000	0.6897	0.1321	0.6366
Satt182	0.9057	3.0000	0.1746	0.0755	0.1661
Satt397	0.5000	2.0000	0.5000	0.0566	0.3750
Satt148	0.6154	4.0000	0.5533	0.1923	0.5001
Satt477	0.8396	2.0000	0.2693	0.0943	0.2330
Satt530	0.7453	7.0000	0.4304	0.0943	0.4145
Satt590	0.5755	8.0000	0.6358	0.1509	0.6131
Satt012	0.3396	8.0000	0.7937	0.1132	0.7669
Satt172	0.6132	3.0000	0.5331	0.1698	0.4630
Satt414	0.4057	7.0000	0.7419	0.0943	0.7053
Satt215	0.7830	4.0000	0.3708	0.0566	0.3509
Satt387	0.5000	3.0000	0.5502	0.1321	0.4490
Satt441	0.4906	9.0000	0.6917	0.1887	0.6573
Satt156	0.6509	5.0000	0.5158	0.1132	0.4641
Satt184	0.6698	4.0000	0.5093	0.1887	0.4706
Satt294	0.4340	6.0000	0.6823	0.1321	0.6270
Satt434	0.7453	6.0000	0.4274	0.0943	0.4081
Satt511	0.4906	4.0000	0.6728	0.1698	0.6262
Satt490	0.5000	3.0000	0.5822	0.0943	0.4950
Satt509	0.6509	5.0000	0.5392	0.0566	0.5068
Sct_034	0.9528	3.0000	0.0906	0.0189	0.0879
Satt372	0.3491	6.0000	0.7300	0.2075	0.6820
Sct_067	0.9057	2.0000	0.1709	0.0000	0.1563
Total	-	135	-	-	-
Mean	0.6252	4.5667	0.5024	0.1133	0.4562

2.4.2.3 Estimates of genetic distances

The estimates of genetic similarity are presented in Table, 2.7. Results exhibited both highest and lowest similarities within and between clusters. The pair wise genetic similarity coefficients among the 42 genotypes varied from 10% to 98%. The highest similarity coefficient was observed between genotypes EL25 and EL26 (coded as G20 and G19), while the lowest similarity coefficient was between genotypes, EL2 and EL37 and EL2 and EL39 (also coded as G41 and G7; and G41 and G1 respectively) implying that genetic distance (dissimilarity coefficient) was 90%.

Table 2.7: Similarity matrix of the 42 soybean genotypes generated by Numerical Taxonomy Multivariate Analysis System for personal computer (NTSYS-pc)

	EL1	EL2	EL3	EL4	EL5	EL6	EL7	EL8	EL9	EL10	EL11	EL12	EL13	EL14	EL15	EL16	EL17	EL18	EL19	EL20	EL21	EL22	EL23	EL24
EL1	1.00																							
EL2	0.19	1.00																						
EL3	0.29	0.30	1.00																					
EL4	0.22	0.27	0.69	1.00																				
EL5	0.65	0.36	0.26	0.19	1.00																			
EL6	0.30	0.60	0.46	0.54	0.27	1.00																		
EL7	0.28	0.33	0.56	0.49	0.29	0.42	1.00																	
EL8	0.28	0.40	0.60	0.74	0.26	0.64	0.57	1.00																
EL9	0.28	0.29	0.53	0.70	0.31	0.53	0.43	0.68	1.00															
EL10	0.31	0.38	0.66	0.67	0.34	0.53	0.66	0.64	0.64	1.00														
EL11	0.16	0.37	0.57	0.77	0.26	0.48	0.49	0.71	0.52	0.55	1.00													
EL12	0.28	0.30	0.58	0.61	0.31	0.52	0.51	0.62	0.59	0.56	0.52	1.00												
EL13	0.27	0.38	0.49	0.49	0.33	0.52	0.59	0.60	0.58	0.56	0.46	0.56	1.00											
EL14	0.22	0.37	0.39	0.43	0.35	0.37	0.36	0.53	0.58	0.46	0.40	0.61	0.57	1.00										
EL15	0.31	0.20	0.56	0.69	0.25	0.47	0.48	0.57	0.63	0.54	0.52	0.60	0.62	0.33	1.00									
EL16	0.24	0.44	0.58	0.74	0.33	0.58	0.73	0.72	0.68	0.68	0.71	0.66	0.59	0.52	0.61	1.00								
EL17	0.29	0.24	0.60	0.70	0.23	0.45	0.56	0.55	0.58	0.67	0.50	0.58	0.54	0.37	0.75	0.62	1.00							
EL18	0.32	0.39	0.55	0.69	0.33	0.59	0.54	0.58	0.64	0.60	0.59	0.73	0.57	0.48	0.75	0.76	0.62	1.00						
EL19	0.25	0.36	0.52	0.74	0.34	0.49	0.54	0.67	0.68	0.73	0.74	0.59	0.50	0.58	0.54	0.75	0.61	0.67	1.00					
EL20	0.36	0.25	0.59	0.66	0.24	0.56	0.55	0.60	0.55	0.52	0.53	0.58	0.62	0.38	0.83	0.61	0.72	0.71	0.52	1.00				
EL21	0.09	0.40	0.53	0.69	0.25	0.44	0.52	0.60	0.54	0.66	0.72	0.44	0.48	0.36	0.52	0.70	0.72	0.51	0.70	0.52	1.00			
EL22	0.23	0.48	0.57	0.69	0.32	0.54	0.53	0.67	0.55	0.61	0.72	0.52	0.52	0.39	0.56	0.72	0.63	0.52	0.67	0.54	0.79	1.00		
EL23	0.22	0.44	0.54	0.60	0.32	0.48	0.52	0.58	0.49	0.55	0.63	0.48	0.46	0.33	0.52	0.68	0.60	0.48	0.58	0.47	0.69	0.90	1.00	
EL24	0.22	0.23	0.65	0.79	0.25	0.53	0.48	0.65	0.69	0.63	0.62	0.73	0.54	0.49	0.81	0.73	0.72	0.78	0.63	0.77	0.58	0.59	0.52	1.00

EL = Experimental line (Genotype code); where EL1 = G1

	EL1	EL2	EL3	EL4	EL5	EL6	EL7	EL8	EL9	EL10	EL11	EL12	EL13	EL14	EL15	EL16	EL17	EL18	EL19	EL20	EL21	EL22	EL23	EL24
EL25	0.22	0.31	0.60	0.70	0.26	0.56	0.52	0.63	0.67	0.58	0.53	0.65	0.60	0.47	0.79	0.74	0.63	0.79	0.58	0.69	0.56	0.60	0.53	0.82
EL26	0.22	0.27	0.59	0.62	0.22	0.50	0.58	0.60	0.60	0.51	0.49	0.57	0.59	0.39	0.77	0.73	0.62	0.71	0.51	0.68	0.52	0.59	0.56	0.74
EL27	0.22	0.30	0.70	0.65	0.22	0.52	0.57	0.59	0.56	0.59	0.58	0.66	0.53	0.45	0.67	0.72	0.58	0.70	0.56	0.64	0.48	0.61	0.65	0.76
EL28	0.34	0.23	0.59	0.59	0.29	0.44	0.61	0.60	0.60	0.63	0.49	0.63	0.62	0.43	0.74	0.64	0.66	0.68	0.57	0.71	0.52	0.50	0.46	0.71
EL29	0.36	0.26	0.51	0.67	0.31	0.57	0.59	0.56	0.61	0.58	0.51	0.62	0.60	0.35	0.81	0.65	0.73	0.72	0.58	0.78	0.56	0.60	0.57	0.72
EL30	0.35	0.24	0.45	0.50	0.32	0.45	0.49	0.47	0.49	0.46	0.47	0.52	0.51	0.53	0.59	0.55	0.60	0.55	0.48	0.69	0.43	0.45	0.43	0.62
EL31	0.29	0.31	0.45	0.53	0.26	0.56	0.56	0.47	0.49	0.55	0.37	0.52	0.51	0.33	0.62	0.58	0.63	0.72	0.45	0.69	0.52	0.45	0.40	0.62
EL32	0.38	0.27	0.57	0.70	0.26	0.51	0.49	0.58	0.58	0.64	0.57	0.58	0.46	0.37	0.66	0.58	0.73	0.55	0.61	0.63	0.56	0.60	0.60	0.66
EL33	0.28	0.43	0.56	0.56	0.29	0.50	0.52	0.57	0.40	0.40	0.69	0.54	0.48	0.39	0.61	0.61	0.59	0.61	0.54	0.74	0.55	0.59	0.56	0.61
EL34	0.32	0.37	0.58	0.69	0.32	0.53	0.52	0.59	0.64	0.64	0.53	0.65	0.51	0.42	0.60	0.65	0.64	0.54	0.59	0.58	0.58	0.58	0.50	0.66
EL35	0.28	0.33	0.53	0.66	0.29	0.50	0.55	0.65	0.60	0.66	0.56	0.63	0.54	0.49	0.48	0.67	0.49	0.58	0.63	0.52	0.52	0.56	0.52	0.58
EL36	0.35	0.20	0.42	0.53	0.23	0.39	0.39	0.47	0.55	0.43	0.43	0.55	0.51	0.30	0.69	0.49	0.53	0.55	0.45	0.53	0.36	0.45	0.47	0.59
EL37	0.26	0.10	0.42	0.51	0.30	0.37	0.40	0.48	0.59	0.44	0.47	0.56	0.67	0.41	0.77	0.50	0.58	0.60	0.49	0.60	0.40	0.42	0.44	0.63
EL38	0.23	0.14	0.45	0.51	0.26	0.40	0.37	0.48	0.59	0.41	0.41	0.49	0.72	0.41	0.70	0.44	0.54	0.53	0.43	0.57	0.43	0.42	0.41	0.57
EL39	0.32	0.10	0.52	0.61	0.23	0.40	0.37	0.61	0.68	0.56	0.44	0.52	0.52	0.44	0.70	0.50	0.64	0.49	0.59	0.57	0.43	0.45	0.44	0.60
EL40	0.35	0.24	0.45	0.50	0.32	0.31	0.39	0.45	0.41	0.52	0.40	0.32	0.46	0.30	0.49	0.37	0.57	0.38	0.45	0.44	0.49	0.45	0.47	0.39
EL41	0.34	0.20	0.55	0.61	0.25	0.47	0.60	0.51	0.51	0.59	0.45	0.56	0.61	0.32	0.79	0.57	0.81	0.67	0.47	0.85	0.54	0.49	0.48	0.73
EL42	0.25	0.35	0.57	0.57	0.31	0.57	0.42	0.51	0.53	0.46	0.49	0.61	0.50	0.37	0.65	0.56	0.57	0.63	0.47	0.73	0.51	0.52	0.46	0.68
EL43	0.22	0.30	0.59	0.79	0.25	0.58	0.52	0.68	0.63	0.63	0.59	0.67	0.51	0.49	0.61	0.70	0.59	0.68	0.63	0.62	0.58	0.62	0.56	0.74
EL44	0.24	0.32	0.45	0.53	0.36	0.43	0.31	0.48	0.52	0.44	0.44	0.48	0.51	0.66	0.46	0.52	0.50	0.48	0.48	0.53	0.46	0.45	0.38	0.62
EL45	0.28	0.36	0.55	0.68	0.28	0.52	0.57	0.67	0.62	0.68	0.58	0.66	0.56	0.52	0.51	0.69	0.52	0.60	0.66	0.55	0.54	0.58	0.52	0.60
EL46	0.30	0.26	0.52	0.46	0.22	0.42	0.56	0.51	0.43	0.53	0.31	0.56	0.53	0.40	0.48	0.48	0.49	0.44	0.39	0.51	0.39	0.39	0.34	0.51
EL47	0.34	0.30	0.38	0.36	0.29	0.39	0.39	0.39	0.40	0.31	0.23	0.44	0.42	0.33	0.45	0.39	0.39	0.52	0.35	0.40	0.23	0.26	0.26	0.35
EL48	0.36	0.26	0.37	0.32	0.31	0.35	0.41	0.35	0.36	0.31	0.22	0.43	0.41	0.32	0.44	0.35	0.38	0.50	0.31	0.39	0.22	0.23	0.22	0.34
EL49	0.35	0.31	0.39	0.37	0.29	0.39	0.39	0.37	0.38	0.32	0.23	0.45	0.40	0.33	0.43	0.40	0.40	0.52	0.35	0.41	0.23	0.27	0.27	0.36

EL = Experimental line (Genotype code); where EL1 = G1

	EL25	EL26	EL27	EL28	EL29	EL30	EL31	EL32	EL33	EL34	EL35	EL36	EL37	EL38	EL39	EL40	EL41	EL42	EL43	EL44	EL45	EL46	EL47	EL48
EL25	1.00																							
EL26	0.92	1.00																						
EL27	0.74	0.83	1.00																					
EL28	0.62	0.65	0.67	1.00																				
EL29	0.70	0.69	0.62	0.66	1.00																			
EL30	0.50	0.49	0.55	0.66	0.67	1.00																		
EL31	0.70	0.62	0.52	0.56	0.67	0.57	1.00																	
EL32	0.53	0.52	0.58	0.69	0.70	0.67	0.53	1.00																
EL33	0.56	0.58	0.57	0.52	0.63	0.62	0.49	0.59	1.00															
EL34	0.61	0.55	0.51	0.55	0.64	0.53	0.56	0.72	0.49	1.00														
EL35	0.52	0.45	0.51	0.65	0.53	0.49	0.46	0.66	0.45	0.60	1.00													
EL36	0.53	0.52	0.52	0.62	0.54	0.43	0.37	0.63	0.43	0.53	0.46	1.00												
EL37	0.58	0.57	0.56	0.67	0.61	0.51	0.41	0.58	0.47	0.45	0.47	0.78	1.00											
EL38	0.61	0.60	0.52	0.57	0.58	0.41	0.44	0.44	0.37	0.45	0.40	0.68	0.86	1.00										
EL39	0.54	0.53	0.49	0.60	0.58	0.41	0.37	0.64	0.43	0.56	0.57	0.61	0.69	0.62	1.00									
EL40	0.47	0.46	0.39	0.43	0.44	0.30	0.40	0.50	0.36	0.44	0.36	0.47	0.47	0.58	0.47	1.00								
EL41	0.65	0.63	0.59	0.70	0.80	0.68	0.71	0.65	0.63	0.62	0.48	0.55	0.59	0.56	0.52	0.52	1.00							
EL42	0.63	0.56	0.56	0.56	0.63	0.57	0.66	0.49	0.56	0.59	0.51	0.43	0.49	0.52	0.43	0.37	0.69	1.00						
EL43	0.72	0.65	0.67	0.61	0.66	0.56	0.56	0.59	0.48	0.66	0.61	0.46	0.47	0.50	0.50	0.46	0.60	0.65	1.00					
EL44	0.56	0.49	0.48	0.43	0.51	0.69	0.47	0.44	0.49	0.53	0.40	0.31	0.38	0.41	0.35	0.41	0.52	0.54	0.55	1.00				
EL45	0.55	0.48	0.53	0.67	0.55	0.52	0.48	0.68	0.48	0.62	0.98	0.45	0.46	0.39	0.56	0.35	0.47	0.50	0.60	0.42	1.00			
EL46	0.46	0.45	0.47	0.59	0.52	0.40	0.46	0.46	0.34	0.51	0.42	0.40	0.35	0.38	0.38	0.34	0.53	0.45	0.54	0.38	0.44	1.00		
EL47	0.43	0.42	0.38	0.35	0.41	0.30	0.43	0.26	0.29	0.33	0.26	0.36	0.37	0.43	0.40	0.26	0.35	0.34	0.32	0.28	0.29	0.54	1.00	
EL48	0.41	0.41	0.37	0.38	0.39	0.29	0.41	0.22	0.28	0.29	0.22	0.35	0.35	0.45	0.35	0.32	0.37	0.33	0.31	0.27	0.25	0.55	0.94	1.00
EL49	0.40	0.39	0.39	0.33	0.41	0.30	0.43	0.27	0.30	0.33	0.26	0.33	0.34	0.41	0.37	0.23	0.35	0.34	0.33	0.28	0.29	0.54	0.98	0.92

EL = Experimental line (Genotype code); where EL1 = G1

2.4.2.4 Cluster analysis

Figure 2.4 shows the molecular dendrogram for the 42 genotypes constructed using the UPGMA clustering algorithm based on SSR markers. Clustering analysis helped to substantiate the results of pairwise or head to head analysis. The dendrogram clearly separated the cultivars into 15 clusters at 60% dice similarity. The closest distance was between genotypes G25 and G26 while the greatest genetic distance (dissimilarity) was between genotypes G41 and G7; and G41 and G1. Cluster 12 consisted of the majority of genotypes (17) accounting for 40% of the total population. Clusters 15, 14, 11 and 3 were each composed of two genotypes, while clusters 13, 8, 6, 5, 4 and two comprised of only one genotype each. Cluster 1 consisted of three cultivars. While clusters 7, 9 and 10 contained three, six and five genotypes respectively.

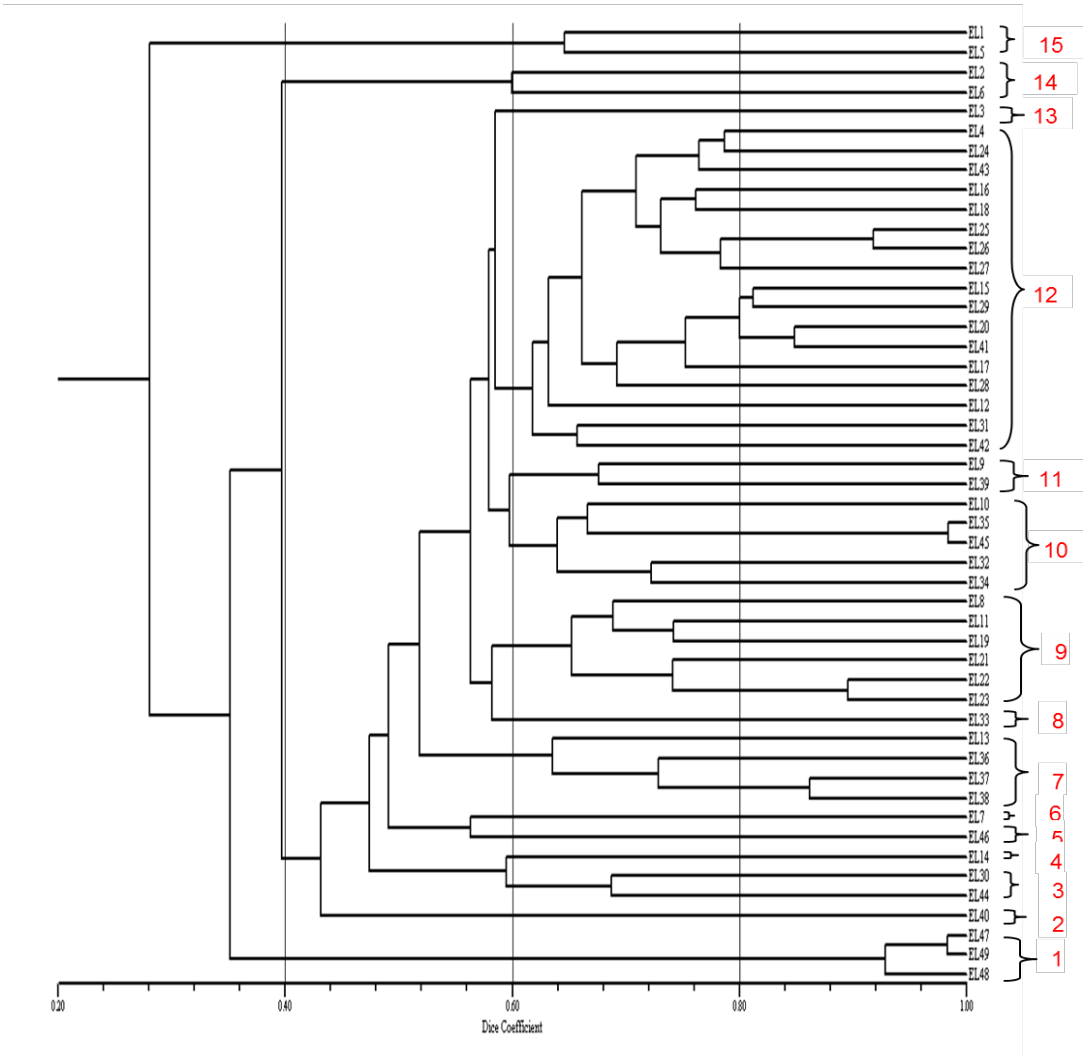

Figure 2.4: Dendrogram exhibiting genetic relationships among 42 soybean genotypes evaluated using 30 SSR markers

2.5 Discussion

2.5.1 Phenotypic diversity

2.5.1.1 Phenotypic variation among the genotypes

The results showed highly significant differences among most of the traits demonstrating existence of wide variation among the test genotypes. Fairly high CVs were obtained on percentage lodged plants at maturity, pod height and number of days from 95% pod maturity to first pod shattering. In addition, broad ranges revealed high levels of diversity among the genotypes for these traits suggesting that selection on the basis of these traits could beneficial. These results were in accordance with the findings of Sihag *et al.* (2004) and Chettri *et al.* (2005) who observed wide variability among the test soybean genotypes. However, in a similar study Malik *et al.* (2011) found higher CVs (46.5%) on grain yield compared to 6% observed in the current study. The lower CVs for most traits indicated that the extent of variability was low suggesting limited scope for selection. The quality traits (protein and oil) showed low CV values along with narrow ranges implying reduced genetic differentiation and it's further substantiated by the spreading pattern shown by the histograms (Figure 2.1). The limited diversity in these traits could be explained by selection practices where breeders sacrificed these traits for higher yield since there is generally, an inverse relationship (negative correlation) between grain yield and quality traits. Cho *et al.* (2008) examined the genetic interrelationships among 260 soybean genotypes and reported low CVs for protein and oil. Where breeders tend to chase higher productivity, quality traits are become compromised. However, Holbrook *et al.* (1989) as cited by Piovesan (2000), Wilcox & Cavins (1995) and Scott & Kephart (1997) showed the possibility of the selection of families bearing high protein content, without losses to the grain yield, or vice-versa. In a similar regard, Filho *et al.* (2004) reported a positive correlation between grain yield and protein in some of the populations under study. Therefore, it can be argued that it is possible to obtain lines with high grain yield, keeping the quality traits at the same level.

The observed variability displayed by the test genotypes in terms of ranges suggested that these genotypes could be exploited in the development of new cultivars and the introgression per se could assist to broaden the Zimbabwean germplasm diversity. Grain yield ranged exhibited from 2723 to 4823 ha^{-1} indicative of wide ranges. Basavaraja *et al.* (2005) presented similar results. The highest yielding genotypes with high agronomic value could be selected for making crosses. Contextually, the choice of parents should take into cognisance a combination

of genetic divergence and genotypic performance regarding important traits (Destro, 1991; Rangel *et al.*, 1991). Days to 50% flowering, days to 95% pod maturity and lodging showed wide ranges, giving room for selection. It was observed that narrow ranges represented the performance of the early releases whereas broader ranges were for the recent releases. Therefore, in the interest of genetic advance, recent releases would be elite lines of choice for cultivar development. However, depending on the breeding objective, for example, the first release G41 is early maturing and would be useful as a donor parent for such genes. Clearly, the results demonstrated the importance of evaluating the breeding lines before using them so that traits presenting higher levels of variability could be identified and simultaneously superior genotypes could be identified. Although genetic variation existed for some traits among the 42 elite lines, it is suggested to introduce exotic germplasm in order to increase diversity to the existing gene pool.

2.5.1.2 Cluster analysis and interrelationships among soybean genotypes

The 42 genotypes were assigned to eight distinct clusters, indicative of wide genetic diversity. Ideally, these results substantiated the wide ranges of the phenotypic variation shown in Table 2.4 as well as the frequency distribution of the various traits. These results were comparable to the findings of Ojo *et al.* (2012) and Dayaman *et al.* (2009) who reported seven and six genetic groups respectively. Contrary to these results, Malik *et al.* (2011) studied the genetic diversity among 92 genotypes and found three clusters demonstrating limited diversity. Among the various sub-clusters observed, sub-clusters **I and II** consisted of one genotype each, G41 and G31 respectively and sub-cluster **VI** contained 2 genotypes (G2 and G7) symbolising that these genotypes show maximum variability, hence they were quite dissimilar to other genotypes. Genotype, G41 was an introduction from South Africa with its pedigrees originating from United States. It was the first genotype to be registered and commercialized in Zimbabwe. Apart from the low grain yield from G41, it is an early maturing genotype and hence can be used for introgression purposes with a view to create segregates that are early, particularly for markets where earliness is an important attribute. Given the prevailing climate change, this genotype still has value due to perennial challenges of terminal drought being experienced in Sub-Saharan Africa. Cluster VIII contained the largest number of genotypes (12) implying high genetic resemblance among the members consequently reduced variability. The distribution of genotypes in other clusters showed the prevalence of broad diversity. Similar observations were reported by Dayaman *et al.* (2009).

Interestingly, the genotypes that were developed and selected for the Zambian market were spread in the same cluster viz, G1, G13, G14, G19, G21, G6, G42, G8, G15, G17 and G25. This collection could probably be possessing similar alleles for adaptation hence, being grouped together on grounds of eco-geographic location. The other exciting feature was that earlier cultivars were separated into the same clusters and vice versa. This could be attributed to genetic improvement implying that recently released cultivars are more superior and tend to perform similarly. Most importantly, the genetic clustering patterns shown by the results revealed that the breeding strategies that were applied over seven decades did not compromise diversity (instead they maintained). Furthermore, these grouping patterns present an excellent opportunity for breeders to make better choices when deciding to make combinations. The existence of many genetic clusters permits inter-cluster selection of parents for hybridisation. In the same breath, it has been observed that combinations made between two divergent clusters maximize variability for subsequent selection relative to within cluster mattings (Zhong-hu, 1991; Benesi, 2005). However, the down-side of the genetic clustering patterns from the phenotypic data was lack of correspondence with the pedigree information.

The principal component analysis (PCA) was also used to describe the genetic interrelationships among the test genotypes. Generally, PCA is regarded as a useful tool for observing the variability of the population under study in genetic studies (Malosetti and Abadie, 2001; Dayaman *et al.*, 2009). On another note, Crossa *et al.* (1990) postulated that PCA presents remarkably sound visualization of the clustering patterns which underpins the inherent or existing structure that has a physical meaning. In this study, the first two principal components explained 52.0% of the total variance. In a similar regard, Anna-Durai (2005) observed 66.1% of the total variance coming from three principal components. However, Mardia *et al.* (1979) reiterated that total variance compounded by the principal components close to 80% gives a satisfactory explanation about the diversity manifested by the population under study. The widespread distribution shown by scatter plot clearly demonstrated wide diversity among the genotypes. The classification of the genotypes was more or less similar to the phenotypic dendrogram (Figure 2.1). Classical examples are shown by genotypes, xG41 and xG31 which were assigned in their own clusters and hence, corresponding to the clustering pattern shown by the dendrogram derived from phenotypic data. Genotypes xG8 and xG42 had the smallest genetic distance in the phenotypic dendrogram and the same pattern is emerging on the PCA scatter plot. Genotypes xG16 and xG23 are still clustered together which again corresponds to the phenotypic dendrogram. However, genotypes XG38 and xG33 were far

apart. Probably the slight differences could be ascribed to the fact that the phenotypic means were used to construct the dendrograms as opposed to the non-averaged data (actual data) that was used to perform PCA.

2.5.2. SSR diversity

2.5.2.1 Genetic polymorphism

A high call rate was observed indicating that the markers that were used were highly reliable hence, these markers could be recommended for utilisation in genetic diversity studies and phylogenetic relationships in cultivated soybean as well as related *Glycine* species. In total, 135 alleles were detected among the 42 genotypes with an average of 4.57 alleles per locus. The number of alleles per locus ranged from two to nine which portrayed high levels of genetic diversity in the gene pool. The genetic diversity scored across the primers varied between 0.09 to 0.79. These results were comparable to the findings of Tantasawat *et al.* (2011) who found 4,82 alleles per locus in a genetic diversity study involving 25 soybean genotypes from Thailand. Similarly, Priolli *et al.* (2010) reported 5,06 alleles per locus among 168 Brazilian cultivars. The trend for the gene diversity was also in tandem with the findings of Narvel *et al.* (2000) who reported an average gene diversity of 0.5 within a range of 0.0 to 0.79 among 39 elite genotypes from United States. Therefore, the results revealed the existence of diversity among the Zimbabwean germplasm. This variation is primarily useful for genetic improvement. However, the mean number of alleles observed were few than 10.4 alleles per locus which were reported for 92 SSR markers in 260 accessions (Cho *et al.*, 2008). Furthermore, using 60 SSR markers on 122 soybean genotypes sampled from Chinese germplasm, Wang *et al.* (2006) reported 12.2 alleles per locus. This meant that a high level of diversity was present in those genotypes. Kuroda *et al.* (2009) assessed the genetic diversity of 1305 wild soybean accessions including 53 cultivated soyabean and found mean number of alleles per locus of 28 and 5 respectively. Arguably, the total number of alleles is thought to be proportional to the sample size (Xia-Su *et al.*, 2004). In that regard, higher diversity was partially attributed to number of genotypes used and high levels of diversity among the genotypes. The reduced diversity revealed by this study compared to these three findings could also be corresponding to the reports that intensive breeding and selection causes genetic bottlenecks (Hyten *et al.*, 2006). Essentially, that results in reduced diversity, changing of allele of frequencies and extinction of some of the alleles that were once existing in the ancestors.

The polymorphic information content (PIC) values ranged from 0.0879 and 0.7669 yielding an average PIC of 0.46. Dayaman *et al.* (2009) studied SSR polymorphism in 45 selected genotypes from India using 11 SSR primers and obtained values that ranged between 0.273 and 0.909. By comparison, Hudcovicova and Karaic (2003) reported PIC values ranging from 0.141 to 0.894 among 67 soybean genotypes with mean of 0.51. Hence, the PIC results were in agreement with earlier studies. Four out of thirty primers (Satt012, Satt414, Satt441 and Satt372) were highly informative (with PIC values greater than 0.65), along with high allele numbers (6-9) and indicating that they were highly effective in discriminating the soybean genotypes. These markers would be assets in molecular characterisation and could be recommended for utilisation in determining genetic diversity and relatedness among soybean genotypes. Interestingly, the larger proportion of the SSR markers (25/30) had PIC values $\geq$0.3, the value that has been commonly used to determine usefulness of RFLP, RAPD and AFLP markers in previous soybean germplasm diversity studies (Thompson *et al.*, 1998; Ude *et al.*, 2003; Tantasawat *et al.*, 2011). However, the highest percentage (76.7%) of polymorphism detected in this study was lower relative to the findings of (Kumar *et al.*, 2009; Khan *et al.*, 2010; Singh *et al.*, 2010). Narvel *et al.* (2000) and Dayaman *et al.* (2009) reported 97.0% and 97.9% polymorphism respectively. Probably the polymorphism in the present study was determined by the sequences of primers as well as the types used.

The study also showed that major allele frequency was characterised by a high proportion of alleles at very high frequency (Table 2.5) with an average of 0.62. The most frequent allele (frequency 0.95) was found at the locus, Satt034. Alleles with frequencies greater than 0.70 were found at 10 markers or loci. Compared to the results reported by Ristova *et al.* (2010), a higher frequency level was found in this study. Allele frequencies are used to illustrate the amount of genetic diversity in the population. Additionally, the expected heterozygosity values ranged between 0.0000 to 0.1923 with an average of 0.1133. The observed heterozygosity values were low suggesting little diversity which can be ascribed to species reproductive system inbreeding. These results were in agreement with Mulato *et al.*, 2010.

2.5.2.2 Genetic similarity and cluster analysis

The pair wise genetic similarity coefficients among the 42 cultivars ranged from 10,34% to 98,41%. These results indicated a high level of genetic diversity. Tantasawat *et al.* (2011) found 79% to 97% genetic similarity from the analysis of 15 certified soybean varieties in Thailand. Clearly, this demonstrated narrow genetic diversity. In a similar study, genetic similarity

coefficients varying from 0.26 to 0.93 for the elite x elite comparisons and from 0.18 to 0.94 for the plant introduction by plant introduction comparisons were reported (Narvel *et al.*, 2000). The genetic similarity about the introductions confirmed the past report that higher genetic diversity existed in exotic germplasm (Chowdhury *et al.*, 2002). For instance, EL25 and EL26 had a genetic similarity coefficient of 0.92 implying that the genetic distance (dissimilarity coefficient) between the two was 0.08. This simply meant that the two genotypes are very similar. According to the pedigree data, the two genotypes are sister lines. The results of similarity matrix were substantiated by the SSR dendrogram.

The dendrogram assigned the 42 cultivars into fifteen clusters (Figure 2.4). The genetic grouping patterns corresponded very well to the pedigree data. Cluster 15 contained two genotypes coded as EL1 and EL 5. Genotype, EL1 is the parent to EL5. Another example was demonstrated by cluster 14 where genotype EL2 is the parent to EL6. Further, cluster 13 contained one genotype which was introduced from United States and had its own cluster and constituted by pedigrees from US suggesting that it was one of the most divergent genotypes. Cluster 12 contained the highest numbers of genotypes. According to the pedigree information, there was an overuse of genotypes EL15 and EL20. EL20 was a progeny of EL15. Most of the combinations made traced back to EL4. It is important to point out that EL4 was the third genotype to be registered in Zimbabwe and it also traced back to the first grain variety to be registered G41 (coded EL2). EL4 revolutionalised soybean production in Zimbabwe. EL15 was an extremely good variety that had high grain yield potential combined with high agronomic value. As a result, most of the genotypes in this cluster had high genetic similarity coefficients signifying close relatedness. Members of clusters 3 were grouped together and they come from the same descendants. Individuals from cluster 9 were close in their pedigree justifying why they fell into the same cluster. Similarly, genotypes classified in cluster 10 shared a common ancestor confirming the relationship among the genotypes. Clusters 2, 4, 5, 6 and 8 consisted of one genotype in each which was rather distinct from other genotypes. Interestingly, cluster 1 was composed of genotypes from China and indeed they formed their own cluster. Thus, the clustering patterns exhibited evidence of wide genetic diversity. The results were comparable to previous studies (Mulato *et al.*, 2010; Ristova *et al.*, 2010; Ojo *et al.*, 2012). The dendrogram obtained from the SSR markers demonstrated that these markers were effective in distinguishing and separating genotypes.

2.5.3 Comparison of phenotypic versus SSR dendrograms

The dendrogram from the phenotypic data produced eight clusters while the SSR marker data separated the genotypes into 15 clusters which, apparently, was in agreement with the pedigree information. The dendrogram generated from the phenotypic data was generally divorced from the pedigree records. Between the two dendrograms a number of discrepancies were noted. On a comparative basis, it was observed that genotypes G41 was in its own cluster in the phenotypic dendrogram whereas in the SSR derived dendrogram it was assigned to the same cluster that contained its progeny G37 (EL6). Another classical example was for G19 and G20 (also coded as EL25 and EL26) which were clustered separately in the phenotypic dendrogram whereas in SSR dendrogram the two ended in the same cluster which was correct because the two are sister lines or full sibs. Apart from these two methods, principal component analysis was also employed to characterise the genetic diversity of the 42 genotypes using the phenotypic data on the grounds that it presents better visualization of the clustering patterns of studied material (Dayaman *et al.*, 2009). The results were from principal component analysis were comparable to the phenotypic approach. Clearly, the results demonstrated that the SSR markers were more discriminatory, highly polymorphic and informative. The ability of the SSR markers to detect polymorphisms at molecular level (DNA) justified their effectiveness and power in differentiating the genotypes (Tantasawat *et al.*, 2011). The genetic relationships that were displayed by the two methods could facilitate the selection of parental stock for hybridisation purposes in addition to the other breeding lines available in the programme. Selection of the most divergent parents would suggest that unique alleles for the desirable traits could be captured.

2.5.4 Implications of the study for breeding

Both the classical (phenotypic) and molecular characterisation support observation of good genetic diversity. The duo was able to reveal genetic variation among the soybean germplasm in Zimbabwe. Consequently, the available germplasm can be utilized for further genetic improvement. These results can aid breeders to select parents that are genetically distant from each other. Parents that have small genetic distances should be avoided because they fail to produce transgressive segregates and the recombinants will be less variable (Biswas et al., 2008). Ideally, genetically diverse parents will help to accelerate genetic gains as they are

assumed to possess complementary genes. The similarity matrix or head to head analysis revealed that EL2 and EL37; EL2 and EL39 (i.e. G41 and G7; G41 and G1) had the least genetic similarity coefficient of 0.10 possibly implying that they could be possessing unique alleles, presently a good opportunity for exploitation. Other potential combinations could come from G41 and G42 with 0.14 genetic similarity coefficients. However, it is suggested to acquire foreign germplasm from elsewhere in order to increase diversity.

2.6 Conclusions

Generally, the genetic diversity was observed at both phenotypic and molecular level among the 42 studied genotypes. The phenotypic dendrogram allocated the genotypes into eight clusters whereas SSR markers assigned them into 15 clusters with good compliance with the pedigree records. The SSR markers were found to more polymorphic, informative and discriminatory. The principal component analysis also revealed evidence of wide genetic diversity among the genotypes. Mostly importantly, potential genotypes with large genetic distance were identified viz, G41 and G7; G41 and 1; G41 and G42. Moreover, two molecular markers or loci had high PIC values that is Satt012 and Satt414 indicating their usefulness in genetic diversity studies. The genetic patterns obtained from the study could help the breeders to make better and reliable choices of distant parents when planning a crossing programme in order to obtain higher genetic variation among segregates. However, it would be prudent to introduce exotic germplasm in order to further enrich the available diversity. Germplasm enrichment and expansion are an opportunity!

2.7 References

Anna-durai, A. 2005. Genetic diversity, diallel analysis and induced mutagenesis in soybean [*Glycine max* (L.) Merrill]. MSc (Agric) Thesis. Tamilnadu Agricultural University, Coimbatore, India.

Basavaraja, G. T., G. K. Naidu and P. M. Salimath. 2005. Evaluation of vegetable soybean genotypes for yield and component traits. Karnataka Journal of Agricultural Science 17:27-30.

Benesi, I. 2005. Characterisation of Malawian cassava germplasm for diversity, starch extraction and its native and modified properties

Department of Plant Sciences (Plant Breeding) in the Faculty of Natural and Agricultural Sciences at the University of the Free State, South Africa.

Biswas, M. K., M. A. A. Mondal, M. Hossain and R. Islam. 2008. Utilisation of genetic diversity and its association with heterosis for progeny selection in potato breeding programs. Crop Science 6:882-887.

Bretting, P. K. and M. P. Widrlechner. 1995. Genetic markers and horticultural germplasm management. Horticultural science 30:1349-1356.

Buckler, E. S. and J. M. Thornsberry. 2002. Plant molecur diversity and applications to genomics. Plant Biology 5:107-111.

Chakravanthi, B. K. and R. Naravaneni. 2006. SSR marker based DNA finger printing and diversity study in rice (*Oriza sativa* L.). African Journal of Biotechnology 5:684-688.

Chettri, M., S. Mondal and R. Nath. 2005. Studies on genetic variability in soybean [*Glycine max* (L.) Merrill] in the mid hills of Darjeeling District. Journal of Interacademicia 9:175-178.

Cho, G. T., J. Lee, J. K. Moon, M. S. Yoon, H. J. Baek, J. H. Kang, T. S. Kim and N. C. Paek. 2008. Genetic diversity and population structure of Korean soybean landrace [*Glycine max* (L.) Merr.]. Journal of Crop Science Biotechnology 11:83-90.

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.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. van Toai, D. G. Lohnes, J. Chung and J. E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Science 39: 1464-1490.

Crossa, J., G. H. Gauch and R. W. Zobel. 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Science 6:36-40.

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 Intergrative Biology 5:125-129.

de Mello Holbrook, O. L., C. S. Sediyama, M. A. Moreira, M. S. Reis, G. A. Massoni and N. D. Piovesan. 1989. Grain yield and seed quality of soybean selected for high protein content. Pesquisa Agropercuaria Brasileira 39:445-450

Destro, D. 1991. Capacidade de combinação de genótipos de soja [*Glycine ma*x (L.) Merrill] apropriados para o consumo humano Piracicaba: ESALQ/USP, 158. Tese de Doutorado.

Dice, L. R. 1945. "Measures of the amount of Ecologic Association Between Species." Ecology 26:297-302.

Elias, M., D. McKey, O. Panaud, M. C. Anstett and T. Robert. 2001. Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): Perspective for on-farm conservation of crop genetic resources. Euphytica 120:143-157.

Franco, J., J. Crossa, J. M. Ribaut, J. Betran, M. L. Warburton and M. Khairallah. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. Theoretical and Applied Genetics 103:944-952.

Gai, J. Y., and and T. H. Zhao. 2001. The core ancesters of soybean cultivars in China. Journal of Nanjing Agricultural University 24:20-23.

Gizlice, Z., T. E. Carter Jr and J. W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947-1988. Crop Science 34:1143-1551.

Goedhart, P. W. and T. N. M. Thissen 2013. Biometrics GenStat Procedure Library Manual 16th Edition.

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Leger, O. Lepais, C. Lepoittevin, T. Malausa, E. Revardel, F. Salin and R. J. Petit. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11:591-611.

He, G., R. Meng, M. Newman, G. Gao, R. N. Pittman and C. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). Plant Biology 3:1-6.

Hershey, C. H. and C. Ocampo. 1989. New marker genes found in cassava, Cassava Newsletter 13: 1-5.

Hudcovicova, M. and J. Karaic. 2003. Utilisation of SSRs for characterisation of the soybean [*Glycine max* (L) Merril] genetic resources. Czech Journal of Genetic and Plant Breeding 39:120-126.

Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht., R. C. Shoemaker. and P. B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. National Academy of Sciences of US 103:1666-1671.

Khan, R., H. Khan and H. K. Farhatullah. 2010. Evaluation of microsatellite markers to discriminate induced mutation lines, hybrid lines and cultigens in chickpea (*Cicer arietinum* L). Austrialian Journal of Crop science 4:301-308.

Khodadadi, M., M. H. Fotokian and M. Miransari. 2011. Genetic diversity of wheat (*Triticum aestivum* L.) geneotypes based on cluster and principal component analyses for breeding strategies. Australian Journal of Crop Science 5:17-24.

Kumar, P., V. K. Gupta, A. K. Misra, D. R. Modi, and and B. K. Pandey. 2009. Potential of molecular markers in plant biotechnology. Plant Omics Journal 2:141-162.

Kuroda, Y., N. Tomooka, A. Kaga, S. M. Wanigadeva and D. A. Vaughan. 2009. Genetic diversity of wild soybean (*Glycine soja* Sieb. et Zucc.) and Japanese cultivated soybeans [*Glycine max* (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. Genetic Resources and Crop Evolution 56:1045-1055.

Li, Y., R. Guan, Z. Liu, Y. Ma, L. Wang, L. Li, F. Lin, W. Luan, P. Chen, Z. Yan, Y. Guan, L. Zhu, X. Ning, M. J. M. Smulders, W. Li, R. Piao, Y. Cui, Z. Yu, M. Guan, R. Chang, A. Hou, A. Shi, B. Zhang, S. Zhu and L. Qiu. 2008. Genetic structure and diversity of cultivated soybean [*Glycine max* (L.) Merrill] landraces in China. Theoretical and Applied Genetics 117:857-871.

Li, Z., and and R. L. Nelson. 2001. Genetic diversity among soybean accessions from three countries measured by RAPDs. Crop Science 41:1337-1347.

Liu, K. and S. Muse. 2005. PowerMarker: an intergrated analysis environment for genetic marker analysis. Genetics and Population Analysis 21:2128-2129.

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.

Malik, M. F. A., M. Ashraf, A. S. Quresh and M. R. Khan. 2011. Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pakistan Journal of Botany 43:1249-1255.

Malosetti, M. and T. Abadie. 2001. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits Genetic Resources and Crop Evolution 48:381-390.

Mardia, K. V., J. T. Kent and J. M. Bibby. 1979. Multivariate analysis Academic press, London, United Kingdom.

MJ Research Inc. 1987. Suppliers of biotechnological equipment MJ Research Inc., San Francisco, United States.

Mulato, B. M., M. Möller, M. I. Zucchi, V. Quecini and J. B. Pinheiro. 2010. Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. Pesquisa Agropecuaria. Brasileira 45:276-283.

Narvel, J. M., W. R. Fehr, W. C. Chu, D. Grant and R. C. Shoemaker. 2000. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Science 40:1452-1458.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, China. p. 3321-3323.

Ojo, D. K., A. O. Ajayi and O. A. Oduwaye. 2012. Genetic Relationships among Soybean Accessions Based on Morphological and RAPDs Techniques. Pertanika Journal of Tropical Agricultural Science. 35:237-248.

Piovesan, N. D. 2000. Aplicacao de cruzamentos dialelicos no melhoramento genetico do teor proteico em soja. Unversidade Federal de Vicosa, Vicosa. Thesis. p. 1-91.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238.

Priolli, R. H. G., J. B. Pinheiro, M. I. Zucchi, M. M. Bajay and N. A. Vello. 2010. Genetic diversity among brazilian soybean cultivars based on SSR loci and pedigree data. Brazilian Archives of Biology and Technology 53:519-531.

Pritchard, J. K. and N. A. Rosenberg. 1999. Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics 65:220-228.

Rangel, P. H. N., C. D. Cruz, R. Veconviski and R. P. Ferreira. 1991. Selection of lowland cultivars based in multivariance genetic divergence Revista Brazileira de Genética 14:437-453.

Reyna, N. and C. H. Sneller. 2001. Evaluation of marker assisted introgression of yield QTL alleles into adapted soybean. Crop Science 41:1317-1321.

Ristova, D., H. Sarcevic, S. Simon, L. Mihajlov and I. Pejic. 2010. Genetic diversity in southeast European soybean germplasm revealed by SSR markers. Agriculturae Conspectus Scientificus 75:21-26.

Rohlf, F. J. 1998. On Application of Geometric Morphometrics to Studies of Ontogeny and Phylogeny. Systematic Biology 47:147-158.

Rolf, F. J. 1998. NTSYSpc Numerical Taxonomy and Multivariate Analysis System Version 2.0 User Guide, Applied Biostatistics Inc., Setauket, New York. p. 1-44.

Rongwen, J., M. S. Akkaya, U. Lavi and P. B. Cregan. 1995 The use of microsatellite DNA markers for soybean genotype identification Theoretical Applied Genetics 91:1001–1007.

Salimi, S., S. H. Lahiji, G. M. Abadi, S. Salimi, and and S. Moradi. 2012. Genetic diversity in soybean genotypes under drought stress condition using factor analysis and cluster analysis. World Applied Sciences Journal 16:474-478.

Sandoval, M. P., T. E. Carter Jr, D. M. Webb and J. W. Burton. 1997. RFLP genetic similarity and coefficient of parentage as genetic variance predictors for soybean yield. Crop Science 37:698-703.

Scott R. A. and K. D. Kephart. 1997. Selection for yield, protein and oil in soybean crosses between adapted and introduced parents. Field Crops Research 49: 177-185.

Semagn, K., A. Bjørnstad and M. N. Ndjiondjop. 2006. An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540-2568.

Shadakshari, T. V., T. Kalaimagal, N. Senthil, M. B. Boranayaka and R. K. Gowda. 2011. Genetic diversity studies in soybean [*Glycine max* (L.) Merril] based on morphological characters. Asian Journal of Biological Science 6:7-11.

Sihag, R., J. S. Hooda, R. D. Vashishtha and B. P. S. Malik. 2004. Genetic divergence in soybean [*Glycine max* (L.) Merrill]. Annals of Applied Biology 20:17-21.

Singh, R. K., S. K. Mishra, S. P. Singh, N. Mishra and M. L. Sharma. 2010. Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. Australian Journal of Crop Science 4:116-125.

Singh, S. P. 1989. Pattern of variation in cultivated dry bean. Annual Report. Bean improvement cooperative. 180-182

Sokal, R. R. and C. D. Michener. 1958. A statistical methods for evaluating systematic relationships. Science Bulletin, Volume XXXVIII. p. 1409-1437

Tantasawat, P., J. Trongchuen, T. Prajongjai, S. Jenweerawat and W. Chaowiset. 2011. SSR analysis of soybean (*Glycine max* (L.) Merr.) genetic relationship and variety identification in Thailand. Australian Journal of Crop Science 5:283-290.

Terron, A., E. Preciado, H. Cordova, H. Mickelson and R. Lopez. 1997. Determinacion del patron heterotico de 30 lineas de maiz derivadas del la poblacion 43 SR del CIMMYT. Agronomy 8:26-34.

Thompson, J. A., R. L. Nelson and L. O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers Crop Science 38 1348-1355.

Ude, G. H., K. W.J., J. M. Costa, P. B. Cregan and J. Alvernaz. 2003. Genetic diversity of soybean cultivars from China, Japan, North America, and North American ancestral lines determined by amplified fragment length polymorphism. Crop Science 43:858-1867.

Vello, N. A., D. M. Hiromoto and A. J. B. V. A. Filho. 1988. Coefficient of parentage and breeding of Brazilian soybean germplasm. Revista. Brasileira Genetics 11:679-697.

Wang, L., R. Guan, L. Zhangxiong, R. Chang and L. Qiu. 2006. Genetic Diversity of Chinese Cultivated Soybean Revealed by SSR Markers. Crop Science 46:1032-1038.

Wilcox, J. R. and F. J. Cavins. 1995. Backcrossing higher seed protein to a soybean cultivar. Crop science 35: 927-931

Xia-Su, X., J. Liu and G. Liu. 2004. The use of SSRs for predicting the hybrid yield and yield heterosis in 115 key inbred lines of Chinese maize. Hereditas 141:207-215.

Yu, K., S. J. Park, V. Poysa and P. Gepts. 2000. Integression of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus Vulgaris* L.) The American Genetic Association 91:429-434.

Zhang, G. W., S. C. Xu, W. H. Mao, Q. Z. Hu and G. Y.M. 2013. Determination of the genetic diversity of vegetable soybean [*Glycine max* (L.) Merr.] using EST-SSR markers. Journal of Zhejiang University of Science 4:279-288.

Zhong-hu, H. 1991. An investigation of the relationship between the F1 potential and measures of genetic distance among wheat lines Euphytica 58:165-170.

CHAPTER 3

BREEDING PROGRESS, VARIABILITY AND HERITABILITY OF GENOTYPES IN THE SOYBEAN BREEDING PROGRAMME IN ZIMBABWE

3.1 Abstract

Estimation of genetic variability, heritability and performance of the available germplasm assist breeders in selection decisions that culminate in advancing genetic gains in the traits of interest. The objectives of the present study were to: (i) evaluate the genetic variability and agronomic performance of historical and current cultivars, (ii) measure the genetic gain of soybean grain yield and other agronomic traits for the cultivars developed in Zimbabwe over seven decades and (iii) to identify genotypes with good agronomic value that can be used as parents in current and future breeding programme. Forty two genotypes representing all the cultivars that have been released in Zimbabwe over 73 years of breeding were grown at 13 environments. The trials were arranged in a 6 x 7 rectangular lattice design with three replications. Highly significant differences ($P \leq 0.001$) were observed among genotypes for grain yield, plant height, pod height, 95% pod maturity, days to 50% flowering, days to first pod shattering, bacterial blight scores except seed oil and protein percentages, demonstrating broad spectrum of genetic diversity. Grain yield varied between 2785 and 5020 kg ha^{-1}. Genotypes, G15, G16, G17, G1, G42, G28 and G25 exhibited high productivity coupled with good desirable attributes. These could further be exploited in a cultivar improvement programme. Generally, PCV were higher than GCV and ECV, but, small differences were shown on grain yield, downy mildew scores, plant height and days to 95% pod maturity implying less influence of the environment on these traits. Broad sense heritability estimates were moderate to high for grain yield, pod height, plant height downy mildew and 100 seed weight. Interestingly, 100 seed weight combined high heritability with moderate genetic advance indicating additive genetic control; hence, selection could be useful. The realized genetic gain for grain yield was 47 kg ha^{-1} $year^{-1}$ representing annual rate of 1.67%. However, the regression trend line showed that improvement in grain yield was slowing down. One hundred seed weight increased by 0.21 g yr^{-1} over time and responses to year of release and yield differences between cultivars showed that breeders have been selecting for genotypes with greater 100 seed weight. Overall, results indicated that emphasis should be refocused on grain yield to restore the original linear increase.

Keywords: soybean, grain yield, variability, genetic gain

3.2 Introduction

High yields have contributed significantly to the total production of soybean across the globe. In order to sustain food security for the growing population, higher growth rates in seed yields should be maintained. The increase in soybean yield over the years is attributed to plant breeding and agronomic practices. Ideally, an interaction exists between the cultivar and cultural practices coupled with prevailing growing or environmental conditions. The incremental breeding gains compounded over time have had a significant impact in the soybean industry. If anything, the realized breeding gains have contributed to the transformation of the soybean industry in Sub-Saharan Africa. Thus, genetic gains serve as shear universal expression of the expected breeding progress (Moose *et al.*, 2008). Essentially, quantification of the breeding gains helps to measure the success of a given breeding programme. Additionally, it assists in reviewing the breeding strategies in use. Miladinovic *et al.* (2010) concluded that estimation of genetic gain provides an indication of the efficiency of selection for seed yield and other agronomic traits. Furthermore, periodic assessment of breeding gains helps to reveal the significance of crop improvement to the public as well as identifying traits that demand attention (Cox *et al.*, 1988).

Soybean breeding efforts in Zimbabwe have been in force since 1940. The improvement efforts resulted in the release of 42 cultivars including introductions from 1940 to 2013. These cultivars were developed through sequential crosses with a view to pyramid yield genes as well as other agronomic traits from diverse sources. More importantly, these cultivars have contributed to the Zimbabwean soybean value chain and the region at large. Given such a long period of breeding, quantifying the breeding progress would be vital to measure the success of the breeding efforts as well as understanding the past events (Lange and Federerizzi, 2009). On another perspective, Bhatia *et al.* (2008) reiterated that assessment of genetic gains realized over a given time period could present opportunities for reviewing the prevailing breeding methods with a view to enhancing crop productivity. Numerous studies have been carried out regarding the time series evaluation of soybean breeding gains (De Bruin and Pedersen, 2009; Lange and Federerizzi, 2009; Rowntree *et al.,* 2013). Jin *et al.* (2010) evaluated the genetic gains over 56 years of soybean breeding and selection in China and reported a positive correlation between seed yield and cultivar year of release with yearly gain of 0.58%, lodging score was reduced from 3.2 to 1.0%, oil concentration rose from 16.7 to 22%, protein went up from 37.0% to 45.5%, resistance to diseases and pests was equally improved. Investigating 45

cultivars released from 1928 to 2008, Justin (2010) observed a linear increase in seed yield with an annual gain of 22 kg ha^{-1} yr^{-1}.

Combining high grain yield with other agronomic traits may sometimes become problematic. Kopisch-Obuch et al., (2005) reported a linkage drag between grain yield and soybean cyst nematode. The negative correlation was ascribed to disruption of linkats (a collection of favourable alleles that are linked) caused by introgression of the genes (Demarly, 1979). Secondly, the genes for resistance to soybean cyst nematode are unfavourable to seed yield so that selection of new linkats at other genomic locations are needed to ameliorate those negative effects (Yuan *et al.*, 2002). In the same breath, the quantification of grain yield improvement over cultivar year of release will help to reveal the association between grain and soybean leaf rust. This is because the breeders initiated breeding for soybean rust resistance dating back to 13 years. Interestingly, Khanh, *et al.*, 2013 reported excellent soybean rust resistance and simultaneously observed grain yield retention coupled with grain quality traits in soybean rust resistance studies.

However, continued breeding and development of high yielding cultivars with good agronomic attributes remains a challenge. This requires a diverse pool of genotypes upon which to select the best types, hence genetic variability is a key component of the process. Contextually, the variability in the available germplasm is important for crop improvement. Arguably, genetic variation among traits is critical for selecting the desirable types. This view is supported by Ramteke *et al.* (2010) who asserted that the development of superior cultivars is premised on the existence and extent of genetic variability for the desirable traits. The characterisation of the germplasm results in the identification of genotypes that are divergent. Introgression that entails diverse germplasm further increases genetic variability and leads to greater gains from selection (Khan *et al.*, 2011). It is important to note that genotypes vary at genetic level, consequently exhibiting different phenotypic performances suggesting that breeders should have knowledge of the variability of their material. Dilnesaw *et al.* (2013) pointed out that variability existing in a given set of germplasm is the sum total of heredity effects of the concerned genes and environmental influence. This warrants the need to partition the observed variability into heritable and non-heritable components. These genetic parameters are measured as genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability and genetic advance as a percentage (Aditya *et al.*, 2011; Dilnesaw *et al.*, 2013). Crop improvement requires sound manipulation of these genetic parameters. By the same token, Chand *et al.*

(2008) reported that genetic improvement of crops is dependent on accurate estimates of genetic diversity, heritability and genetic advance. Heritability of a particular trait is defined as the proportion of observable differences in a trait between individuals within a population that is due to genetic differences (Wills, 2007). The significance of heritability is that it generates information regarding the proportion of genotypic variance from the total phenotypic variance. In other words, the heritability values inform the breeder about how much of a trait has been passed down to the subsequent generation.

Heritability can be viewed as narrow or broad sense (Falconer and Mackay, 1996; Gebre, 2005). The broad sense heritability (H) can be defined as the proportion of the total genetic variation including dominance and epistasis effects to total phenotypic variation computed as $H = \delta^2G/\delta^2p$, while narrow sense heritability (h^2) is the ratio of additive variance to phenotypic variance also computed as ($h^2 = \delta^2A/\delta^2p$ (Wray and Visscher, 2008). Given a higher heritability for a given trait, it means selection becomes easier and simultaneously response to selection will be greater. Previous studies have been done to estimate variance components, heritability and genetic advance in various soybean characters. For instance, Malik et al. (2007) and Aditya et al. (2011) presented high heritability for days to 50% flowering, branches per plant and 100 seed weight along with high genetic advance on number of pods per plant and dry weight per plant. This gives the impression that variation is largely depended on additive effect implying that selection would be useful. The study also revealed highest estimate of PCV (47.74 and GCV (41.83) for seed yield and number of pods per plant (PCV = 33.48; GCV = 30.16). These results suggest that remarkable improvement could be made through phenotypic selection. Dilnesaw *et al.* (2013) defined genetic advance as the measurement of the expected genetic progress that is attributed to selecting the best performing genotypes for a particular character. Genetic gain for a given trait is the product of its heritability, phenotypic standard deviation and section differential (Burton and Devane, 1953; Nyquist, 1991). Thus, the assessment of these genetic parameters is important because they assist breeders to establish the amount of breeding progress or genetic gain that has been made via particular breeding and selection strategies (methods).

However, the genetic gain for the Zimbabwean soybean breeding programme has not been estimated. The assessment of genetic gain would entail the comparison of morphological characteristics of old and modern cultivars, which essentially provides an indication of the

breeding value of the germplasm (Feil, 1991). Cultivars with superior performance demonstrate higher breeding value, hence can be utilized to advance the breeding gains. Further, knowledge of the attribute(s) that impact most on grain yield gain would receive more attention (Luque et al., 2006). Although many studies on genetic gain have been conducted in major soybean producing countries grain yield response and other various secondary traits entirely depend upon genotype x environment interaction genotype by environment interaction (GEI). Thus, studies conducted in other environments may produce different results under the Sub-Saharan Africa (SSA). By the same token the genetic parameters discussed above have not been quantified. The information and magnitude of these components (PCV, GCV, heritability and genetic advance) are key to crop improvement.

Given the importance of soybean in Zimbabwe and the region at large coupled with lack of information on genetic improvement, the estimation of the breeding gains would be vital. Therefore, the objectives of the study were to: (i) evaluate the genetic variability and agronomic performance of historical and current cultivars, (ii) measure the genetic gain of soybean grain yield for the cultivars developed in Zimbabwe over seven decades and (iii) to identify genotypes with good agronomic value that can be used as parents in current and future breeding programme.

3.3 Materials and Methods

3.3.1 Germplasm sources

The planting material included all the cultivars that were developed by the Zimbabwean soybean breeding programme from 1940 to 2013. It is important to note that three promising lines were included in this study but were later registered in 2012 and 2013 (G42 and G8 were registered in 2012 while G5 was released in 2013). A total of 42 cultivars were used which represented a collection of the released cultivars. The details of test genotypes are as given in Table 2.1 under Chapter 2.

3.3.2 Experimental site and design

A total of 13 test environments were used during 2010/11 and 2011/12 seasons in Zimbabwe, Zambia and Malawi. Site information is as presented in Table 2.2 under Chapter 2. The cultivars were evaluated in these three countries because the breeding team has similar research

operations in all three countries, consequently some of the cultivars are registered in these markets. Planting was done in December in all the seasons. The test entries were arranged in a 6 x 7 rectangular lattice design with three replications at all the sampled sites. The trials were grown in 6 row plots that were 5 metres long with row to row spacing of 45 cm and an in-row spacing of 6.3 cm, giving 79 plants per row. A perfect stand should, therefore, have 476 plants per plot, meaning that 350 000 viable seeds were planted per hectare.

3.3.3 Management

The trials were managed according to the best husbandry practices. Basal fertilizer (Compound L) was applied at a rate of 400 kg ha^{-1}, supplying 28 kg ha^{-1} of Nitrogen, 68 kg ha^{-1} of P_20_5 and 40 kg ha^{-1} of K_20. The seed was inoculated with *Bradrhizobium japonicum* inoculant Grasslands strain 1491. During the growing season, the trial plots were sprayed with the fungicide three times to protect them against rust should it occur. Integrated weed management was applied to control the weeds. Where irrigation facilities were available, supplementary irrigation was applied to the crop in times of need. In some sites the crop was guarded against animals and humans. All the trials were hand planted and harvested.

3.4 Data Collection

The following measurements were taken on each plot

- Percentage lodged plants at maturity (% LODG): Visual estimate to the nearest 10% of plants leaning more than 45°
- Seed appearance scores (SAP): From 1 = very good to 9 = very poor quality with much discolouration, moulding and cracking.
- Days from planting to 50% flowering (50% DFL): When 50% of the plants have at least one open flower.
- Days from planting to 95% pod maturity (DMAT): When 95% of the pods have dried.
- Days from 95% pod maturity to first pod shattering (DSH): Number of days from 95% pod maturity to first pod shattering
- Bacterial blight: *Pseudomonas savastanoi pv glycinea* was scored using a 1-9 scale where 1 is resistant and 9 very susceptible.
- Downy Mildew: *Peronospora manshurica* was scored using a 1-9 scale where 1 is resistant and 9 very susceptible.

- Red leaf blotch (RLB): *Dactuliochaeta glycines* was scored at the R6 stage of growth using a 1-9 scale where 1 is resistant and 9 very susceptible.
- 100 Seed Weight (SDMA): Mass of 100 seeds on a dry matter basis
- Seed yield (SYLD):Plot yield converted to kg/ha at 11% moisture.
- Percentage crude protein (%CRPR) content in the seed on a dry matter basis
- Percentage crude oil content (CROIL) in the seed on a dry matter basis.

It is important to note the trial plots were sprayed against soybean rust because most of lines are susceptible to it, hence it was not measured.

Statistically derived variables included;

1. Grain yield (SYLD) was adjusted to kg ha^{-1} at 11% moisture using the following fomular;

Grain Yield (kg ha^{-1}) = [Grain Weight (Plot yield in kg ha^{-1})/(100 - %MC) *10/Plot Area*111/100]

2. Crude Protein Content (CRPR) was adjusted to 11% moisture content using the following formular;

CRPR = Measured Protein/(100-%MC)*100

3. Crude Oil Content (CROIL) was adjusted to 11% moisture content using the following formular,

CROIL = Measured Crude Oil Content/ (100-%MC)*100

4. 100 Seed mass (SDMA) was adjusted to 11% moisture content using the following formular;

SDMA = Measured 100 seed weight/(100-%MC)*100

Where; %MC = Grain Moisture in percentage.

N.B. The disease rating scale of 1-9 was adapted from international rating scale used for patent and cultivar registrations (http://www.google.com/patentsUS8378178 accessed on 10 October 2010.

3.5 Data analysis

Analysis of variance was performed in GenStat RELM. Analysis across environments was carried out to analyse the effect of years, genotypes and interactions. The genotypes were considered to be fixed and both the replications within environments and environments were regarded as random. Data were combined across years using the following model;

$$Y_{ijk} = \mu + S_i + G_j + \beta_k (S_i) + (GxS)_{ji} + e_{ijk}$$

Where, Y_{ijk} = yield; μ = overall population mean; S_i = site; G_j = genotypes (entries); β_k (S_i) = blocks within sites; $(GxS)_{ji}$ = genotype x site interaction, which was considered as random; and e_{ijk} = random experimental error. The GxS interaction mean square was used as the error term to the F-test for the across site analysis.

The phenotypic and genotypic variances of each trait were estimated from ANOVA using GenSat. The expected mean squares under the assumption of random effects model was computed from linear combinations of the mean squares and the phenotypic and genotypic coefficient of variations were computed as per the methods suggested by (Burton and Devane, 1953) as follows;

Genotypic variance ($\delta^2 g$) = $\frac{Msg - Mse}{r}$

Phenotypic variance (δ^2_p) = $\delta^2 g + \delta^2_e$ and δ^2_e = MSe

Where, Msg and Mse are the mean sum of squares for the genotypes and error mean square (environmental variance) in the analysis of variance, respectively.

δ^2_p = phenotypic variance

$\delta^2 g$ = genotypic variance

δ^2_e = environmental variance

r = is the number of replications.

The genotypic coefficient of variation (GCV) and phenotypic coefficient of variation (PCV) as well as error coefficient of variance (ECV) were calculated according to the formular of Kumar et al. (1985 Bezaweletaw et al. (2006)):

Genotypic Coefficient of Variation (GCV) = (σg /grand mean) * 100

Phenotypic Coefficient of Variation (PCV) = (σ_p/grand mean) * 100
Error coefficient of Variance (ECV) = σ_e/grand mean

Where, σg and σ_p are genotypic and phenotypic standard deviations, respectively.

Heritability in broad sense (h^2) for all characters was computed as per the following formula adopted from (Allard, 1960);

$h^2 = \delta^2 g / {}^2\delta_p * 100$

Then, the genetic advance for selection intensity (k) at 5% (2.063) was estimated by the following formulae of (Johnson et al., 1955)

$GA = k * \delta_p * h^2$

Where k = selection differential at 5%; δ = phenotypic standard deviation of the mean yield of n genotypes and h^2 = broad sense heritability

Genetic gain was estimated as described (Nyquist, 1991);

$R = HS = K(\delta g^2 / \sqrt{\ } \delta p^2)$

where R= genetic gain, K= selection intensity for a normal population, δg^2 = genetic variance, and δp^2 = phenotypic variance, H= heritability and S= selection differential.

Genetic gain was also calculated as the difference between the original population mean and the population mean after a selection (Nyquist, 1991);

$\mu_2 - \mu_1 = R = \Delta G$

Where μ_2 = population mean after selection, μ_1 = original population mean, R= response to selection, and ΔG = genetic gain. In this study, genetic gain was estimated using the second formula.

Estimation of breeding progress- the following factors were considered;

- Year of first grain soybean release/registration
- Last cultivar release
- Age difference between the two releases
- Average performance for the first grain soybean release
- Average performance for the last grain soybean release
- Gain (difference between the 2 mean performances)

Genetic gain = gain/breeding period

% Genetic gain=(gain/mean performance of the first grain cultivar)/breeding period

In order to observe cultivar changes across time, the cultivar means of the agronomic traits were plotted against the year of cultivar release. A straight line was fitted through the points using simple linear regression. Head to head analysis and column graphs were plotted using raw data in GenStat to show the yield gains between the founder and modern varieties.

3.6 Results

3.6.1 Mean squares for the agronomic traits

The results of variance mean squares for the agronomic traits are as shown in table 2.1. Highly significant differences among the test genotypes were observed on all the traits. Similarly sites, seasons and all the interactions revealed significant effects for all the traits.

Table 3.1: Wald Statistic mean square of all traits studied

Source	D.F.	PDHT	PLHT	%LOD	SAP	%PS	BBS	RLB	100 SWGT	DFL	CROIL	CRPRO	DMAT	DSH	GYLD
Genotype	41	275.7***	2105.6***	301.5***	758.6***	190.04***	74.4	417.2***	831.8***	2329.1***	198.6***	198.3***	1668.5***	310.6***	433.2***
Site	12	2614.2***	4983.2***	614.4***	653.8***	120.81***	72.3***	1830.3***	1239.4***	14264.8***	125.93***	210.1***	21594.3***	4975.0***	1835.5***
Year	1	225.0***	89.3***	27.1***	0.9	148.89***	1.3	0.1	16.1***	790.6***	69.06***	5.1***	687.7***	297.5***	1288.1***
Genotype * site	287	534.9***	965.2***	560.8***	1756.5***	493.64***	368.7***	791.4***	923.7***	3142.4***	108.4	108.1	1543.9***	1180.5***	574.3***
Genotype * year	41	304.4***	1431.4***	331.9***	723.58***	270.8***	18.1	218.1***	919.2***	1033.5***	168.43***	144.7***	1632.3***	222.3***	414.5***
Site * year	4	656.0***	153.3***	27.1***	82.04***	445.5***	30.1***	399.5***	229.8***	219.3***	0.9	47.8***	3426.1***	143.8***	2895.5***
Genotype * site * year	164	212.6***	307.7***	251.3***	539.46***	277.9***	28.5	379.3***	128.8	247.4***	51.0	48.7	673.2***	227.2***	374.3***

D.F. = degrees of freedom; PDHT = pod height in centimetres; PLHT = plant height in centimetres; %LOD percentage lodged plants at maturity; SAP Seed appearance scores; %PS percentage purple stained seed; BBS = bacterial blight scores, Red leaf blotch scores; 100 Seed weight; DFL = days from planting to 50% flowering; CROIL = percentage oil in the seed; CRPRO = percentage protein in the seed, DMAT = days from planting to 95% pod maturity; DSH = days from 95% pod maturity to first pod shattering; GYLD = grain yield in kilograms per hectare at 11% moisture

3.6.2 Mean performance of the studied traits

Table 3.2 presents the mean performance of the examined genotypes. The pod height ranged from 12 to 19 cm with genotype, G41 showing the lowest pod height while genotype G2 had the highest pod height. Clearly, genotype 2 was the tallest (113 cm) and G31 was the shortest (71cm). The majority of the genotypes exhibited good standability, while G2 gave the highest mean value (28%) for lodging. In terms of 100 seed weight, G15 and G16 recorded the highest mean value whereas, G41 produced the lowest value. The highest oil content was observed on G16, while the minimum was given by G26. For protein content, G41 gave the highest mean value (41.25) and G22 gave the least mean value (38.45). However, the range for the two quality traits was very narrow. The phenological traits in view of days to 50% flowering and 95% pod maturity revealed that G2 flowered (48 days) and matured (111 days) earlier than the rest while G2 was relatively late maturing (137 days). G6 and G16 gave the highest mean value for days to pod shattering (42 days) compared to 20 days from G18. The range of shattering was quite wide (20-42 days). The grain yield ranged from 2785 to 5020 kg ha^{-1}. The top yielding genotype was G16 with a mean yield of 5020 kg ha^{-1} while G41 gave lowest yield 2785 kg ha^{-1}.

Table 3.2: Mean performance of the selected soybean traits studied

Genotypes	PDHT	PLHT	%LODG	SDMA	CROIL	CRPRO	DFL	DMAT	DSH	GYLD
G16	15	78	000	28	17.85	38.95	56	128	42	5020
G15	17	98	000	28	17.40	39.00	52	128	40	4870
G17	16	84	000	21	17.10	39.60	56	129	34	4728
G1	18	97	3.35	22	17.25	39.10	55	126	30	4710
G5	14	86	000	26	16.95	39.50	53	125	36	4688
G21	14	90	0.05	20	17.35	39.20	55	125	31	4686
G28	14	91	0.05	21	16.65	39.80	56	125	29	4680
G14	15	98	000	22	17.30	39.45	58	135	31	4680
G25	16	99	000	22	16.75	39.80	55	125	36	4672
G42	16	91	000	19	17.00	40.00	57	132	33	4592
G7	15	91	23.35	22	17.05	39.10	54	130	21	4530
G40	16	90	000	21	16.85	39.90	52	128	36	4508
G19	16	90	0.10	20	16.85	39.35	55	126	37	4482
G8	15	100	6.70	20	16.95	39.75	57	128	39	4464
G20	14	77	000	19	16.55	39.85	52	125	30	4456
G6	14	94	000	23	17.25	39.60	50	124	42	4440
G18	15	95	3.35	21	16.90	39.65	55	127	20	4434
G4	13	88	000	23	16.85	39.80	49	121	30	4376
G32	14	88	000	21	16.70	40.20	51	120	25	4332
G11	15	95	000	22	16.95	40.40	50	125	23	4330
G10	16	102	000	26	17.20	39.45	52	126	29	4324
G23	15	80	0.10	25	16.75	39.20	51	124	38	4288
G22	14	78	000	23	17.25	38.45	50	122	26	4284
G13	18	104	000	23	16.80	39.20	55	128	33	4228
G26	13	75	0.10	23	16.10	40.00	51	123	32	4204
G12	12	77	000	23	16.65	39.80	48	120	38	4152
G29	13	74	000	19	17.05	39.80	52	124	34	4134
G31	12	71	000	20	17.15	40.10	48	120	32	4091
G24	14	80	000	22	17.15	39.70	52	121	29	4043
G27	15	78	0.05	22	16.70	40.75	52	125	40	4015
G35	15	95	000	21	17.50	39.65	49	125	22	4012
G9	17	79	0.10	20	16.55	40.50	55	128	38	3974
G30	15	94	000	21	16.80	40.40	55	129	31	3935
G34	16	87	0.10	21	17.40	39.80	53	124	26	3887
G2	19	113	27.50	20	16.55	41.25	64	137	23	3744
G33	14	80	1.85	24	17.00	39.95	49	117	32	3633
G36	13	80	000	21	16.75	40.00	49	122	27	3615
G3	14	87	0.05	22	16.25	40.25	49	121	28	3549
G37	12	76	000	21	16.30	40.65	50	116	34	3307
G38	13	82	16.85	20	16.60	40.20	54	127	26	3264
G39	13	85	000	19	17.05	39.90	48	118	25	3254
G41	12	74	15.00	18	16.90	39.55	48	111	24	2785
Mean	14	87	2.35	21.68	16.93	39.77	52	122	31	4200
Range	12-19	71-113	0-28	18-28	16.1-17.8	38-41.2	48-64	111-137	20-42	2785-5020
S.E ±	0.9	3.75	3.04	1.24	0.35	0.58	0.61	1.292	3.5	188.2
LSD (5%)	2.57	10.7	8.68	3.54	0.99	1.67	1.76	3.69	9.99	537.6
C.V (%)	9	6	183	8	3	2	2	1	16	6
F. Sign.	***	***	***	***	ns	ns	***	***	***	***

***Significant at P≤0.001; ns = Not significant; SE = Standard error, LSD = Least significant difference at 5%; CV = Coefficient of variation; G = Genotypes; PDHT = Pod height; PLHT = Plant height; %Lodg = %Lodged plants; SDMA = 100 seed weight; CROIL = %Oil, CRPRO = % Protein, DFL = 50% flowering, DMAT = Days to maturity, DSH = Days to shattering; GYLD = Grain yield (kg ha^{-1})

3.6.3 Estimation of genetic parameters

The estimates of genotypic, phenotypic coefficients of variation, heritability and genetic advance are shown in Table 3.3. The estimates of genotypic and phenotypic coefficients of variation as percentage of the trait means exhibited broad variability. It was observed that the phenotypic coefficient of variation (PCV) was greater than the genotypic coefficient variation (GCV). However, exceptions were noted on downy mildew scores, pod maturity and grain yield, where the variance values were close. High GCV were noticed for downy mildew scores (20.88); grain yield (16.37); seed appearance scores (12.49) and red leaf blotch scores (10.93). On the other hand, high PCV were recorded for purple seed stain (92.82); bacterial blight scores (64.46); red leaf blotch scores (30.79); seed appearance scores (30.54) and pod height (23.78).

High heritability in the broad sense was recorded for 100 seed weight (68.40%), closely followed by seed appearance scores (57.58%), grain yield (49.88%); and pod height 49.42%, while the rest of the characters had heritability values ranging from 12.2 to 36.26%. Results also revealed that the highest genetic advance as percentage of the mean were observed on percentage purple stained seed (60.98); bacterial blight scores (35.23; pod height (24.21) and 100 seed weight (17.48). The lowest genetic advance was observed on oil content (0.70%).

Table 3.3: Genetic parameters of the measured traits of 42 soybean genotypes across eight sites

Genetic Parameter	PDHT	PLHT	SAP	%PS	SDMA	BBS	DMS	RLB	CROIL	CRPRO	DFL	DMAT	DSH	YIELD
Mean	14.00	87.00	2.104	1.361	21.68	2.157	1.74	2.668	16.93	39.77	52.52	122.2	31.00	4200
Vg	1.30	66.6	0.069	0.097	1.942	0.04	0.132	0.085	0.0443	0.1719	4.24	7.06	1.88	472930
Vp	11.08	127.1	0.413	1.596	7.21	1.933	0.136	0.675	0.246	0.864	15.05	8.368	58.38	473282
GCV (%)	8.11	9.38	12.49	2.29	6.42	9.27	20.88	10.93	1.24	1.04	3.92	2.17	4.42	16.37
PCV (%)	23.78	12.96	30.54	92.82	12.39	64.46	21.19	30.79	2.93	2.34	7.39	2.37	24.64	16.38
H^2_B (%)	49.42	33.13	57.58	31.84	68.4	26.67	36.26	21.52	12.20	15.78	17.9	20.52	21.15	49.88
GA	3.39	7.71	0.76	0.83	3.79	0.76	0.28	0.36	0.12	0.30	1.43	1.22	3.33	707.92
GAM (%)	24.21	8.86	36.12	60.98	17.48	35.23	16.09	13.49	0.70	0.76	2.72	1.00	10.74	16.85

PDHT = pod height in centimetres; PLHT = plant height in centimetres; %LOD percentage lodged plants at maturity; SAP Seed appearance scores; %PS percentage purple stained seed; BBS = bacterial blight scores, Red leaf blotch scores; 100 seed weight; DFL = days from planting to 50% flowering; CROIL = percentage oil in the seed; CRPRO = percentage protein in the seed, DMAT = days from planting to 95% pod maturity; DSH = days from 95% pod maturity to first pod shattering; YIELD = Grain yield in kilograms per hectare at 11% moisture; Vg = genetic variance; Vp = phenotypic variance; GCV = genotypic coefficient of variation; PCV = phenotypic coefficient of variation; H^2_B = heritability in the broad sense; GA = genetic advance; GAM = percentage genetic gain

3.6.4 Changes in soybean cultivar traits and yielding ability

3.6.4.1 Soybean yield trends in Zimbabwe from 1940 to 2013

The grain yields of the time series soybean cultivars showed a positive increase against year of registration from 1966 up to 2005 (Figure 3.1). From 2006 through 2013, the grain yields seem to be declining. The trend also showed a marked decline in grain yield in 1999 which occurred when G3 was registered. As shown on Figure 3.1, G5 represented the latest release and was registered in 2013. The modern genotypes achieved higher yields compared to the older ones. The lowest yield (2785 kg ha^{-1}) was given by the founder cultivar (G41), which was released in 1966 while the highest yield (5020 kg ha-1) was achieved in 2005 when G16 was registered. Relative to G16, subsequent releases showed decline in grain yields. Although a decline was witnessed,

yields remained above four tonnes per hectare and were also greater than the early generation cultivars. Overall, the model showed an annual rate of increase of 35.8 kg ha^{-1} with R^2 value of 0.52.

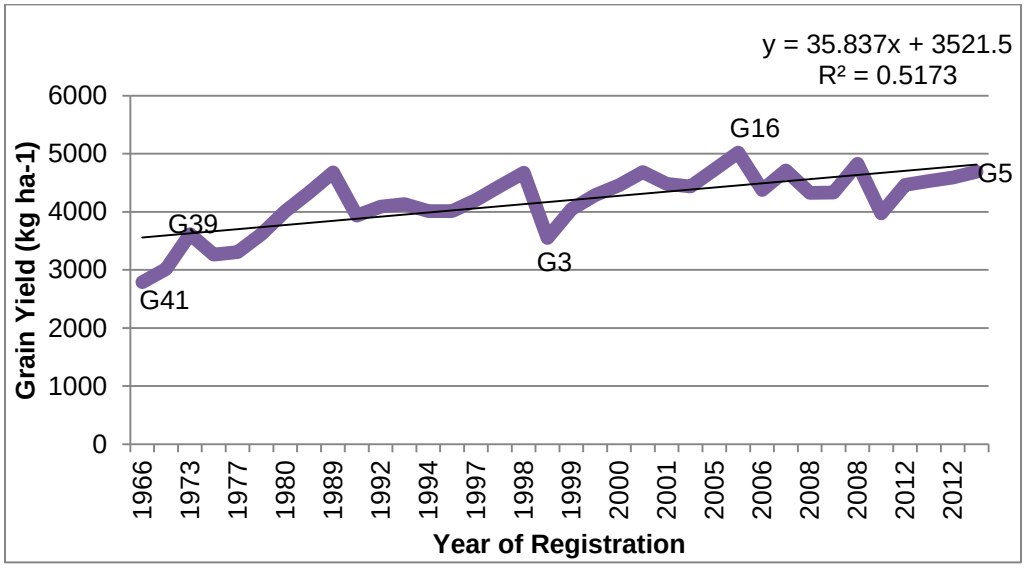

Figure 3.1: Regression of grain yield (kg ha^{-1}) over soybean cultivar year of registration across 13 environments during 2010/11 and 2011/12 cropping season

Figure 3.2 presents the comparison of mean grain yield of the genotypes that were registered in different decades. Although the trend line is positive, the annual rate of increase in yield is decreasing. The mean yields increased from the first decade through the fourth decade and thereafter, mean yields started to go down. According to the model, grain yield was increasing at a rate of 452.90 kg ha^{-1} $decade^{-1}$ with R^2 value of 0.85.

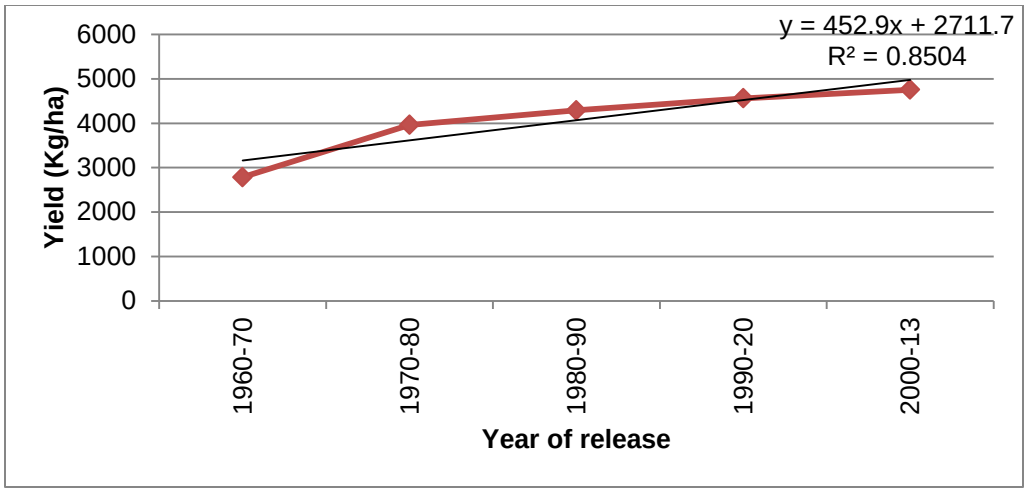

Figure 3.2: Mean grain yield trend across 5 decades

Table 3.4 shows the relative performance of the top 10 genotypes against the founder variety (G41). The highest yield advantage was given by G16 which was registered in 2005 closely followed by genotypes G15 which was released in 2006. Interestingly, one of the varieties from the rust resistant or tolerant background (G42) displayed highly competitive yield as it was part of top 10 yielding genotypes.

Table 3.4: Relative performance of the top ten genotypes against G41

Genotype	Year of Registration	Mean Yield (kg ha^{-1})	Yield of G41 (Founder Variety)	Differential (kg ha^{-1})	% yield advantage G41
G16	2005	5020	2785	2235	80.3
G15	2006	4870	2785	2085	74.9
G17	2005	4728	2785	1943	69.8
G1	2008	4710	2785	1925	69.1
G5	2013	4688	2785	1903	68.3
G21	2000	4686	2785	1901	68.3
G28	1989	4680	2785	1895	68.0
G14	2007	4680	2785	1895	68.0
G25	1998	4672	2785	1887	67.8
G42	2012	4592	2785	1807	64.9
Relative performance of all the rust tolerant genotypes against G41					
G42	2012	4592	2785	1925	64.9
G7	2012	4530	2785	1745	62.6
G8	2012	4464	2785	1679	60.3
G11	2008	4330	2785	1545	55.5
G10	2008	4324	2785	1539	55.3
G9	2010	3974	2785	1189	42.7

Table 3.5 and 3.6 present a direct comparison of the founder variety against the best yielding genotype and the latest rust tolerant release across all the environments that were used. Results showed that the founder variety was outperformed by both. The founder variety was significantly outperformed by the 2 genotypes in 2011/12 cropping season.

Table 3.5: Pairwise comparison of G16 versus G41 (kg ha^{-1})

	2010/11 season						2011/12 season							
	Environments						Environments							
Genotype	1	2	3	4	5		6	7	8	9	10	11	12	13
G16	2785	4566	5466	5838	5508		5084	4523	4130	4553	3680	5899	5260	3899
G41	1795	2042	2726	1222	3699		2241	1546	1947	2365	1620	3422	2684	1839
% Yield advantage	55.1	123.6	58	109	31.7		126.9	192.6	112.1	92.52	127.2	66.5	95.9	112
G41 = Founder Variety							G16 = Recent Variety							

Environment 1= Rattray Arnold Research Station 2010/11; 2 = Gwebi Variety Testing Centre 2010/11; 3 = Lusaka Farm 2010/11; 4 = Mpongwe Development Centre 2010/11; 5 = Bvumbwe Research Station 2010/11; 6 = Rattray Arnold Research Station 2011/12; 7 = Gwebi Variety Testing Centre 2011/12; 8 = Lusaka Farm 2011/12; 9 = Mpongwe Development Centre 2011/12; 10 = Bvumbwe Research Station 2011/12; 11 = Lilayi Farm 2011/12; 12 = ART Farm 2011/12; 13 = Chitedze Research Station

Table 3.6: Pairwise comparison of G42 versus G41 (kg ha^{-1})

	2010/11 season						2011/12 season							
	Environments						Environments							
Genotype	1	2	3	4	5		6	7	8	9	10	11	12	13
G42	3448	3526	4376	6505	5197		4823	3645	4695	4553	3444	4476	5567	3638
G41	1795	2042	2726	1222	3699		2241	1546	1947	2365	1620	3422	2684	1839
% Yield advantage	92.1	72.7	34.9	125	26.3		115.2	135.8	141.1	92.5	112.6	20.3	107.4	97.8
G41 = Founder Variety							G42 = Recent rust tolerant variety							

Environment 1= Rattray Arnold Research Station 2010/11; 2 = Gwebi Variety Testing Centre 2010/11; 3 = Lusaka Farm 2010/11; 4 = Mpongwe Development Centre 2010/11; 5 = Bvumbwe Research Station 2010/11; 6 = Rattray Arnold Research Station 2011/12; 7 = Gwebi Variety Testing Centre 2011/12; 8 = Lusaka Farm 2011/12; 9 = Mpongwe Development Centre 2011/12; 10 = Bvumbwe Research Station 2011/12; 11 = Lilayi Farm 2011/12; 12 = ART Farm 2011/12; 13 = Chitedze Research Station

3.6.4.2 Days to 95% pod maturity

Figure 3.3 showed that maturity increased from 1966 to 1989 in successively newer genotypes and subsequently became clustered; hovering between 121 and 125 days between 1995 and 2005. Further modest increases were observed among the genotypes representing the current generation (between 2005 and 2013). The rate of increase was 0.22 days year^{-1} accompanied by R^2 value of 0.4367. The genotypes were classified according to maturity where; genotypes that matured in less than 120 days were classified as early, those that matured between 120 and 125 days were classified as medium and above 125 days were categorised as late. Table 3.7 showed that grain yield differences were observed with the late maturing genotypes producing the highest mean yield (4562 kg ha^{-1}). The mean yields for the medium and early maturing groups were 4235 kg ha^{-1} and 3235 kg ha^{-1} respectively. The yield advantage between the late maturing and early maturing genotypes was 40.6% while the yield advantage of the medium over early maturing genotypes was 30.5%. The medium maturing cultivars were outperformed by the late maturing genotypes with a yield gain of 7.7%.

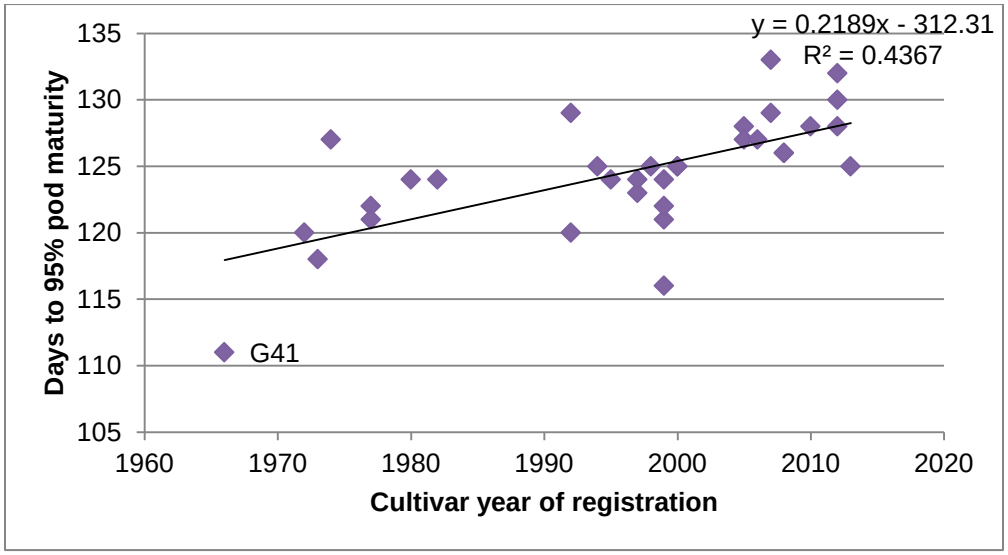

Figure 3.3: Regression of maturity over soybean cultivar year of registration across 13 environments during 2010/11 and 2011/12 cropping season

N.B. G41 = founder variety

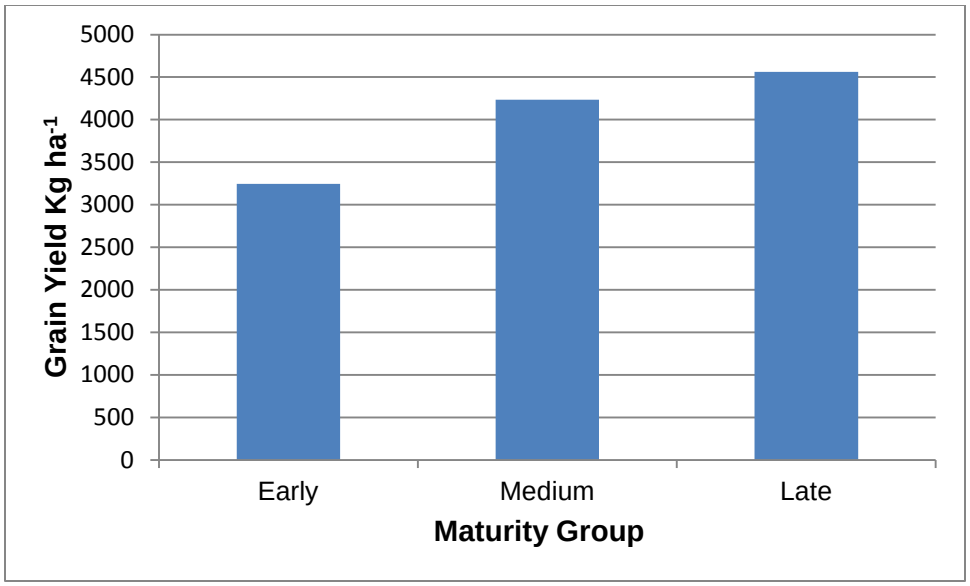

Figure 3.4: Comparison of the genotypes by maturity group

Table 3.7: Relative yield grain yield and yield advantage according to maturity group

Maturity group	Grain yield	Differential	Yield advantage over early (%)	Yield advantage over medium (%)
	Kg/ha			
early	3245	-	0	(23.4)
medium	4235	990	30.5	0
late	4562	327	40.6	7.7

3.6.4.3 100 seed weight

There was generally an increase in 100 seed weight over time in successive genotypes (Figure 3.5) as given by the regression model. The effect of cultivar year of release resulted in 100 seed weight increasing by 0.076 g year^{-1}

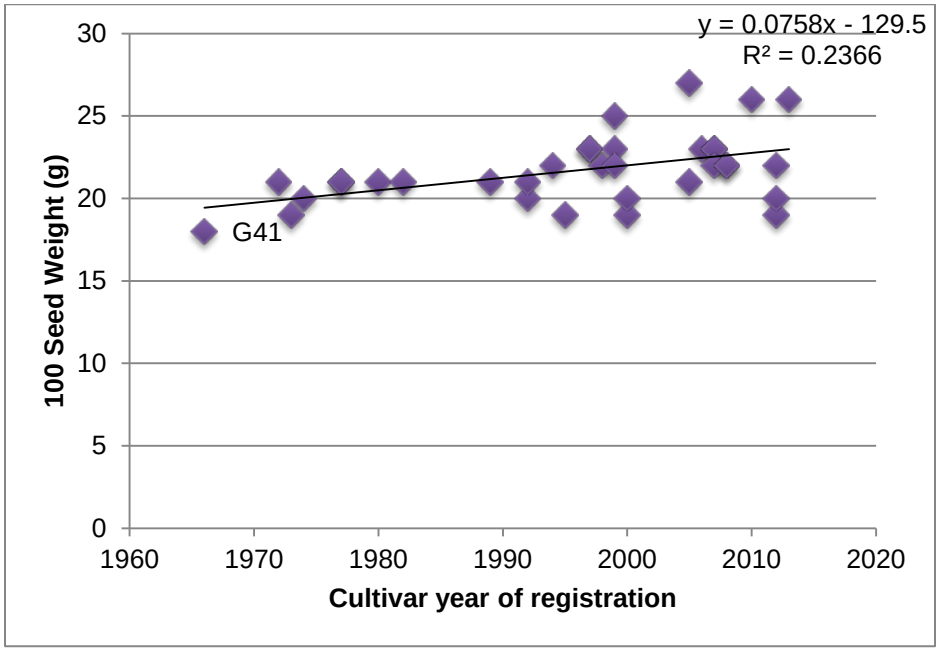

Figure 3.5: Regression of 100 seed weight over soybean cultivar year of registration across 13 environments during 2010/11 and 2011/12 cropping season.

N.B. G41 = founder variety

3.6.4.4 Resistance to stem lodging

The experiment showed a negative trend in percentage lodged plants at maturity over time. The percentage of lodged plants showed a tendency to level off at 0%. Percentage lodged plants at maturity decreased by 0.06% per year (Figure 3.6).

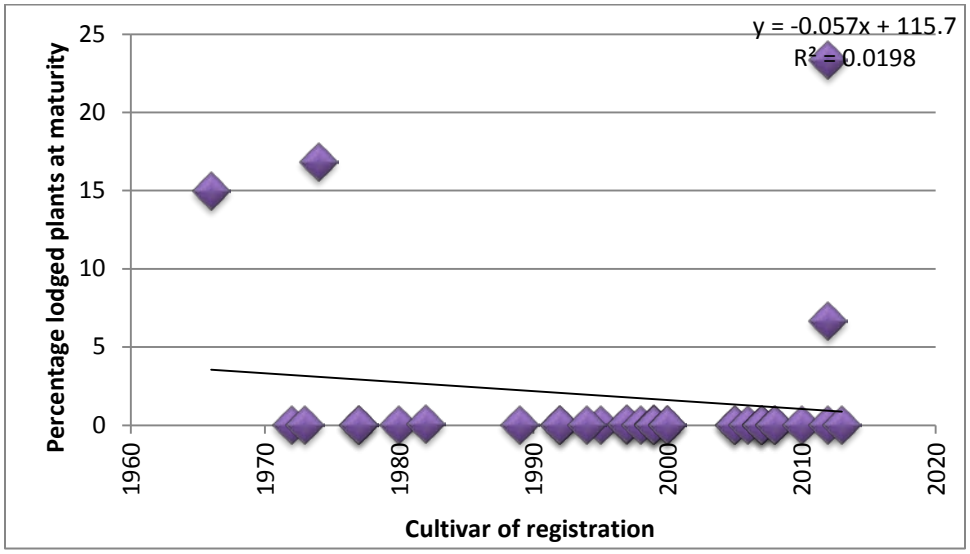

Figure 3.6: Regression of resistance to stem lodging over soybean cultivar year of registration

3.6.4.4 Quality traits

Grain oil percentage showed a positive change over time (Figure 3.7) and the rate of increase was 0.004% per year with an R^2 value of 0.02.

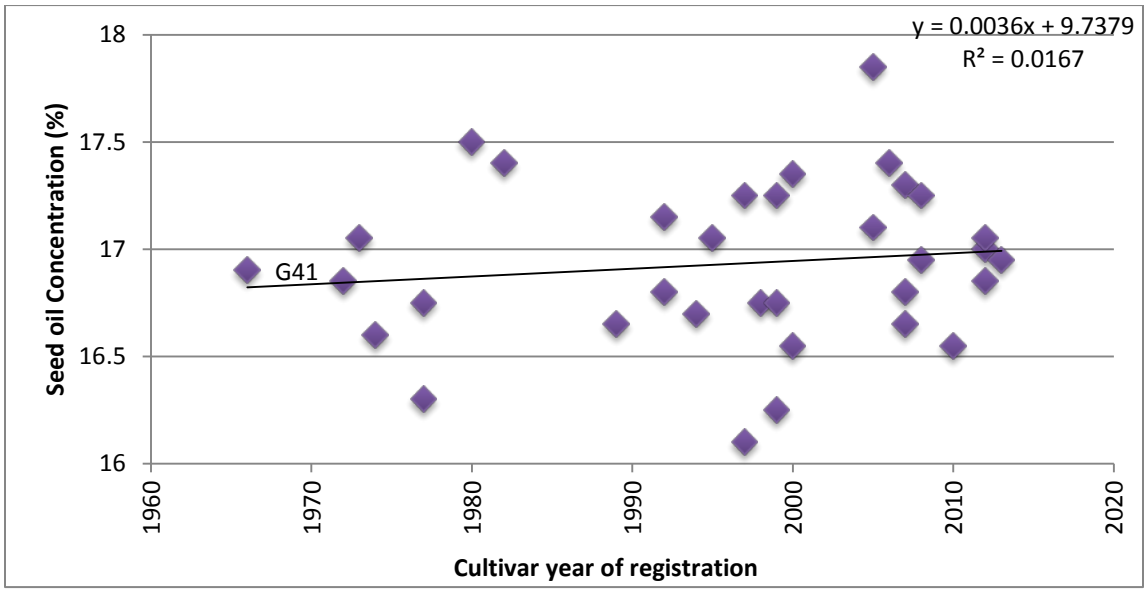

Figure 3.7: Regression of grain oil percentage over soybean cultivar year of release

The seed protein percentage decreased as revealed by the negative trend line (Figure 3.8). The seed protein declined at a rate of 0.01% per annum.

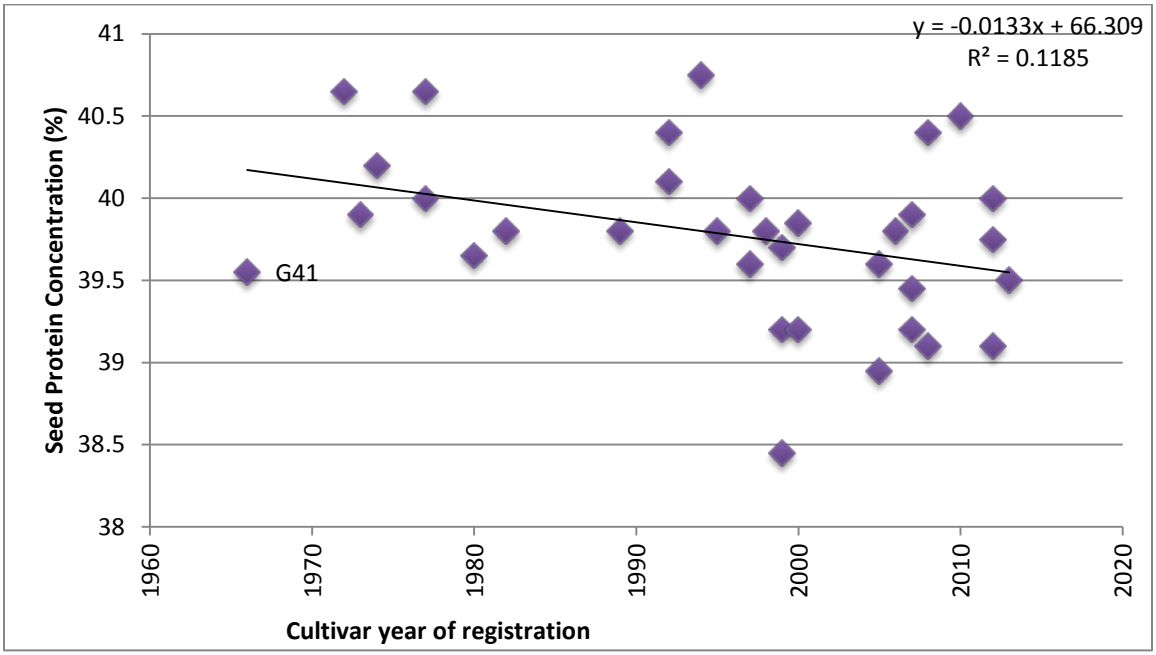

Figure 3.8: Regression of seed protein percentage over soybean cultivar year of release

N.B. G41 = founder variety

3.6.4.5 Resistance to pod shattering

The results showed a positive pattern and a slight increase in the number days from pod maturity to first pod shattering over time. Resistance to pod shattering increased by 0.17 days annually. However, the model is weak and only explained 18.82% of the factors contributing to increase in pod shattering resistance.

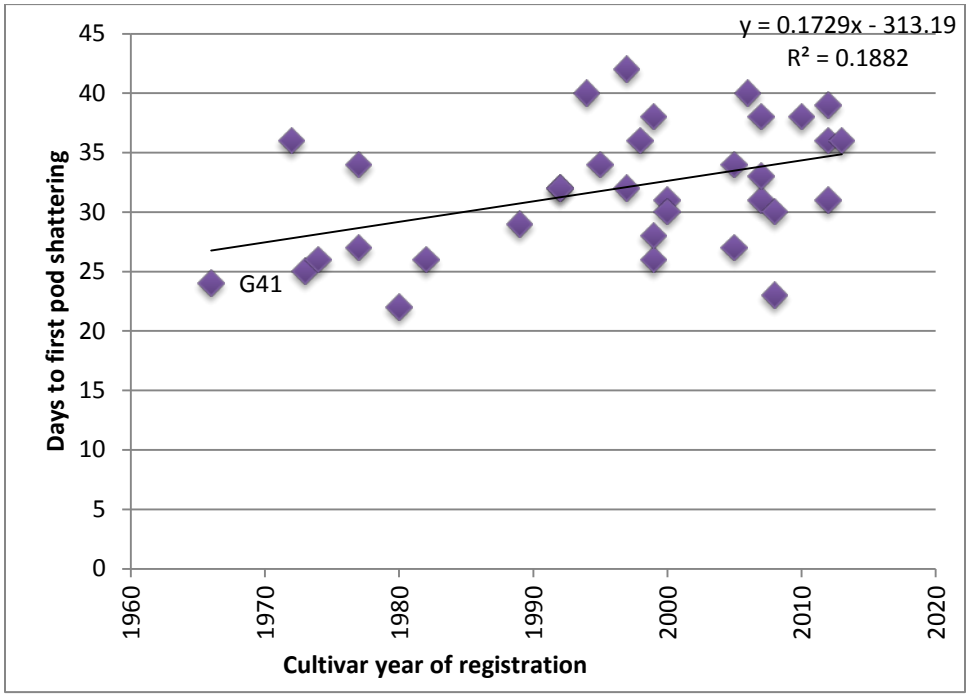

Figure 3.9: Regression pod shattering over soybean cultivar year of release

N.B. G41 = founder variety

3.6.4.6 Annual genetic gain estimates

Table 3.7 summarised estimates of the realised genetic gain for ten selected traits based on the mean performance of the best yielding genotype (G16) compared to the first release (G41). G16 was registered in 2005. The results revealed a genetic progress of 47 kg ha^{-1} $year^{-1}$ for grain yield over a breeding period of 48 years. This represented an annual gain of 1.67%. Resistance to pod shattering registered an annual genetic gain of 0.46 days, translating to an annual genetic gain of 1.56%, closely followed by 100 seed weight (0.21 g $year^{-1}$) and pod height (0.10 cm $year^{-1}$). The lowest annual genetic gain over time was recorded on seed protein percentage (-0.01%) whose annual percentage gain was -0.03% while percentage oil content recorded annual genetic gain of 0.11%.

Table 3.8: Annual genetic gains based on the best yielding genotype (G16) over 48 years of breeding

Trait	**A** Mean values for G41	**B** Mean for G16	**C** Differential (B-A)	Annual Genetic gain (Realised/Actual) **C/48 years**	Annual genetic gain (%) **[(C/A) * 100]/48 years**
PDHT (cm)	12	15	5	0.10 cm $year^{-1}$	0.87
PLHT (cm)	74	78	4	0.1 cm $year^{-1}$	0.11
%LODG	15	0	(15)	(0.31%) $year^{-1}$	(2.08)
SDMA (g)	18	28	10	0.21g $year^{-1}$	1.16
CROIL (%)	16.9	17.8	0.9	0.02% $year^{-1}$	0.11
CRPRO (%)	39.6	38.9	(0.6)	(0.01%) $year^{-1}$	(0.03)
DFL (days)	48	56	8	0.17 days $year^{-1}$	0.35
DMAT (days)	111	128	17	0.35 days $year^{-1}$	0.32
DSH (days)	24	42	18	0.38 days $year^{-1}$	1.56
YIELD (kg ha^{-1})	2785	5020	2235	47 kg ha^{-1} $year^{-1}$	1.67

Key: 48 years represent the breeding period from 1966 upon which the founder variety was registered up to 2013 where the last cultivar was registered.

N.B. Values in parenthesis = are negative values; PDHT = pod height, PLHT = plant height; %LODG = percentage lodged plants at maturity; SDMA = 100 seed weight; CROIL = percentage crude oil in the seed; CRPRO = percentage crude protein in the seed; DFL = days to 50% flowering; DMAT = days to 95% pod maturity; DSH = days from 95% pod maturity to first pod shattering; YIELD = grain yield in kilograms per hectare; g = grammes; kg ha^{-1}; cm = centimetres; % = percent

Table 3.8 also shows the genetic gain estimates of the same traits based on the mean performance of the latest released rust tolerant genotype (G42) relative to the founder genotype (G41). G42 was registered in 2012. Results showed an annual yield gain of 41.02 kg ha^{-1} $year^{-1}$ which corresponded to 1.47% $year^{-1}$. Maturity recorded an annual yield gain of 0.3 days yr^{-1} representing 0.27 %. The lowest gain was observed on seed protein percentage (-0.004%).

Table 3.9: Annual genetic gains based on the recent rust tolerant cultivar (G42) over 46 years of breeding

Trait	**A** Mean Values for G41	**B** Mean for G42	**C** Differential (B-A)	Annual Genetic Gain (Realised/Actual) **(C/46 years)**	Annual Genetic Gain (%) **[(C/A)*100]/46years**
PDHT (cm)	12	16	4	0.09 cm $year^{-1}$	0.07
PLHT (cm)	74	91	17	0.37 cm $year^{-1}$	0.50
%LODG	15	0	(15)	(0.33 %) $year^{-1}$	(0.02)
SDMA (g)	18	22	4	0.09 $year^{-1}$	0.48
CROIL (%)	16.93	16.75	(0.18)	(0.0039) $year^{-1}$	(0.02)
CRPRO (%)	39.55	39.80	0.25	0.005 $year^{-1}$	0.01
DFL (days)	48	55	7	0.15 days $year^{-1}$	0.32
DMAT (days)	111	125	14	0.30 days $year^{-1}$	0.27
DSH (days)	24	36	12	0.26 days $year^{-1}$	0.01
YIELD (kg/ha)	2785	4672	1887	41.02 kg ha^{-1} $year^{-1}$	1.47

Key: 46 years represent the breeding period from 1966 upon which the founder variety was registered up to 2012 where the recent rust tolerant cultivar was registered.

3.7 Discussion

3.7.1 Performance of genotypes

The obtained results revealed that genotype mean squares were highly significant ($P \leq 0.001$) for all the evaluated traits except bacterial blight scores indicating the presence of adequate diversity in respect of morphological, phenological and agronomical variability among the studied genotypes. This was also buttressed by the wide ranges that were shown by most characters assessed. Narrow ranges that were observed on quality traits were in conformity with the results of Aravind, 2006. Contrary to these results, Wiggins (2012) observed broader ranges for both seed protein and oil percentages. Arguably, there is still scope for increasing these traits possibly by making crosses involving exotic lines. The mean yield across 13 environments was 4200 kg ha^{-1} ranging from 2785 to 5020 kg ha^{-1}. The genotype, G16 superseded all the genotypes with highest yield of 5020 kg ha^{-1}. It was further observed that this genotype had greater 100 seed weight. Based on mean performance, the five genotypes G16, G15, G17, G1 and G5 were significantly superior in grain yield. These genotypes could be used as donor parents for desirable characteristics in the hybridisation programme. These results are supported with the previous work where broad variability for seed yield was found (Sirohi et al., 2007). It was observed that the early generation genotypes produced less grain than the current generation. However, there were two genotypes, G28 and G25 which displayed competitive

performance despite the fact that they were released a long time ago. G28 was released in 1989 whereas G25 was registered in 1997. Based on their agricultural merit, they are recommended for use as parental stock so as to avoid loss of favourable alleles.

Regarding phenological traits, the later genotypes flowered later and all the same they matured later than the early releases. This could suggest that breeders were developing and selecting genotypes with longer reproductive periods with a view to improve the grain yield potential (Duvick, 2005). On a comparative basis, the best yielding genotype, flowered and matured in 56 and 128 days relative to 48 and 111 days taken by the founder cultivar (G41). The plant height ranged from 71 to 113 cm. Plant height is an important trait because it is positively correlated with grain yield (Khan *et al.*, 2011). In the present study, the best yielding genotype was relatively short suggesting that grain yield was explained by other yield contributing characters. Another important trait is 100 seed weight, which varied between 18 to 28 g. The maximum weight was observed on G15 and G16 as opposed to the minimum weight which was recorded on the founder variety (G41). The results were at variance with the findings of Khan et al. (2011) and Malik *et al.* (2007) who reported lower values for 100 seed weight ranging from 11.8 to 20.2 g and 3.8 to 17.6 g respectively. The highest percentage of lodged plants at maturity (28%) was given by G2. Probably, it succumbed to lodging because of its height (113cm).

3.7.2 Observed variances

In general, the phenotypic coefficient of variation was higher than the genotypic coefficient of variation for all the traits. According to Shivasubramanian and Menon (1973), GCV and PCV values are classified as low, moderate and high in the ranges of 0 to 10%, 10 to 20% and >20% respectively. In this study, only downy mildew scores showed both high GCV and PCV while seed appearance scores, red leaf blotch scores and seed yield fell in the moderate group and the rest exhibited low values. Low GCV and PCV values for phenological traits as well as 100 seed weight obtained in this study concurred with the findings of several studies (Agarwal *et al.*, 2001; Basavaraja, 2002; Dilnesaw *et al.*, 2013). In addition, low values for seed protein and oil percentages obtained were in accordance with the results of (Aravind, 2006; Shrivastava and Shukla, 1998). This suggests that there is need to enrich variability for these traits. High GCV and PCV estimates for downy mildew scores indicated that there is greater scope for selection based on this character. Furthermore, it means that phenotypic expression would be a good sign of the genotypic potential (Singh *et al.*, 1994). Contrary to this, Dilnesaw *et al.* (2013) advocated for the comparison of genetic coefficient of variation with phenotypic coefficient of

variation because small differences between the two imply that environment has less influence on a given trait in question. Therefore, it can be judged that environmental influence was less on grain yield, plant height, downy mildew scores, seed oil and protein percentages and phenological traits because narrow differences were observed between the two components. Contextually, it means that there is further scope for improvement of these characters.

Results of the current work revealed low, moderate and high heritability estimates in the broad sense for the traits evaluated. Robinson *et al.* (1949) classified heritability values as low (0-30%), moderate (30-60%) and high (>60%). In this case, bacterial blight scores, red leaf blotch scores, seed oil and protein percentages, days to 50% flowering, days to 95% pod maturity and resistance to pod shattering showed low heritability estimates. Grain yield, pod height, plant height, seed appearance scores, percentage purple stained seed and downy mildew scores exhibited moderate values while 100 seed weight had high heritability estimates. Earlier research work, presented high heritability estimates for 100 seed weight, plant height, phenological traits and grain yield (Aditya *et al.*, 2011; Aravind, 2006; Okonkwo and Idahosa, 2013). High heritability estimates signifies the existence of a higher proportion of the fixable additive variance in the population. The bottom line is that selection can be exploited to make further improvements on the basis 100 seed weight.

The study showed low genetic advance as percentage of mean for plant height, seed oil and protein percentages, days to 50% flowering and days to 95% pod maturity. Days to first pod shattering, grain yield, red leaf blotch, downy mildew scores and 100 seed weight exhibited moderate genetic advance as percentage of the mean. High genetic advance as percentage of the mean was found on pod height, seed appearance scores, percentage purple stained seed and bacterial blight scores. Johnson *et al.* (1955) categorized genetic advance as a percentage of the mean into three groups *viz,* low (0-10%), moderate (10-20%) and high (>30%). Dilnesaw *et al.* (2013) and Aditya *et al.* (2011) reported high genetic advance for grain yield and plant height. The high genetic advance coupled with high heritability shown by seed appearance scores suggests that additive gene effects are conditioning the inheritance of this trait. Generally, a combination of high heritability and high genetic advance as percentage of the mean is more useful and meaningful in the context of genetic gain. The association of the two components results in high genetic gain from selection of such characters. Further, the association also influences the type of selection method to be applied (Bhat *et al*., 2012).

3.7.3 Breeding gains in soybean over 48 years of breeding (1966 through 2013)

3.7.3.1 Changes in grain yield

The current study revealed that over 48 years of plant breeding and selection in Zimbabwe, the annual rate of genetic yield gain averaged 1.67% which corresponded to 47 kg ha^{-1} $year^{-1}$. The results also showed that the yield advantage over the first grain cultivar was 80.3%. In a similar regard, Jin *et al.* (2010) obtained an annual rate of genetic gain of 0.58%. Further, the USDA-NASS (2011) and Rowntree *et al.* (2012) reported an annual yield gain of 23.4 kg ha^{-1} which is equivalent to 1.27% for a breeding period of over 80 years. In a separate examination of grain yield gains in United States (US), Egli (2008) reported 15.1 to 38.3 kg ha^{-1} $year^{-1}$ representing a yield advantage of between 30-45% and growing at an annual rate above 1%. However, Tefera *et al.* (2010) observed larger yearly genetic gain estimates of 3.33% over 20 years of breeding effort. The results revealed a remarkable improvement in grain yield compared to other breeding programmes. It also worth mentioning that the genetic gains are as a result of classical breeding tools compared to US where modern technology is largely applied such as marker assisted breeding. The increase in grain yield was probably attributed to increase in the number of nodes per plant and 100 seed weight and it can be concluded that selection for higher yield favoured genotypes with higher number of nodes and greater 100 seed weight. The genetic gain estimates were also done for cultivars that were bred from the rust tolerant background. The estimated annual genetic gain was 41.02 kg ha^{-1} $year^{-1}$ resulting in percentage gain of 1.47. This represented a yield advantage of 64.9% over the founder variety (G41). Although the trend revealed a decline in yield gain, it remained above numerous observations made previously (Justin, 2010; Morrison *et al.*, 2000). However, the regression trend showed that the breeding progress was slowing down. This could be attributed to linkage drag between grain yield and leaf rust resistance. As the breeders diverted their effort to leaf rust, this resulted in a decline in yield gains. Thus, a negative relationship between the two traits developed.

The model was fitted to show the response of grain yield thus, revealing the progress pattern over time. It was noted that grain yield leaped from 1966 to 1973 when the genotype, G39 was registered and accordingly resulted in a dramatic increase in production consequently the soybean industry advanced (Tichagwa, personal communication). It was also observed that cultivar registrations that were done post 2005 had lower grain yields. This was so because the breeding programme shifted its focus to breeding for soybean leaf rust resistance or tolerance. Ideally, soybean leaf rust breeding was started in 1999/2000 and the first breakthrough was in

2008 when two rust tolerant genotypes were registered, hence the programme is still at infancy. This meant that the effort to combine high grain yield with resistance to soybean rust resulted in a linkage drag causing diminished yield levels (Figure 3.1). The rust tolerant cultivars showed lower yields relative to the conventional or non-rust tolerant. Given, that the present germplasm comes from the leaf rust tolerant background, it therefore, means that the breeding effort should be focused on increasing grain yield. Improvement in tolerance to biotic stresses is one way of increasing yield.

In order to observe the yield improvements as a result of plant breeding, a head to head analysis was done between the founder cultivar and the modern cultivars (highest yielding cultivar) was performed (Table 3.5). A similar comparison was carried out between the founder cultivar and the recent leaf rust tolerant cultivar (Table 3.6). Clearly, the founder (G41) was markedly outperformed by G16 in all the 13 environments. This is a clear testimony that modern cultivars are by far better. It also reflected that the soybean breeders were using the best selection methods to improve grain yield demonstrating breeding efficiency. Similar results were observed by earlier researchers (Diers, 2010). A similar trend was shown between the founder cultivar and soybean leaf rust tolerant cultivar although the yield advantage was lower. Although the grain yield of the rust tolerant cultivars is less than the conventional cultivars, their mean yields are still significantly better than the founder cultivar. It was also observed that the magnitude of yield response was location and year dependent and is consistent with previous work (Rowntree et al., 2013). The location and seasonal differences also impacted on mean yield differences (De Bruin and Pedersen, 2008a).

3.7.3.2 Changes in maturity

There was a modest increase in crop duration over cultivar year of release. When maturity was regressed over the founder genotype (G41), an annual rate of maturity gain was estimated to be 0.35 days $year^{-1}$ using the best yielding genotype (G16). This annual rate of maturity improvement observed represented an average breeding gain of 0.32% $year^{-1}$. While the estimated yearly gains from the leaf rust tolerant genotype (G42) when regressed to G41 was 0.30 days which equated to 0.27% $year^{-1}$. Moreover, it was observed that genotypes falling into the late maturing category were all recent releases (from 2000 to 2013). Similarly, the early releases were classified under the early maturing group while the medium maturity group consisted of genotypes that were registered between 1977 and 1998. It could be argued that late maturing genotypes have a longer reproductive and seed filling periods compared to their

counterpart's hence higher productivity per unit area. The results were in conformity with published literature (Rowntree *et al.,* 2012). Thus, breeding and selection of genotypes that spend more time in reproductive growth stages of soybean (R1 to R7) resulted in yield improvement.

3.7.3.3 Changes in 100 seed weight

Results exhibited an annual breeding progress of 0.21 g $year^{-1}$ over cultivar year of release which was equivalent to an annual rate of 1.16%. In the same breath, annual breeding improvement of 0.09 grammes $year^{-1}$ representing an annual rate of 0.48% was observed on the rust resistant or tolerant breeding programme. This implied that 100 seed weight had a contribution to grain yield. These results were at variance with some past research work (Egli, 2008; Jin *et al.*, 2010; Kahlon, 2001) who found no consistent relationship between 100 seed weight and cultivar year of release, instead, they reported that yield improvement was achieved by increasing number of seeds per plant. Contrary to these reports, Gay *et al.* (1980) and Justin (2012) observed that higher yields were primarily due to increase in seed size which supports the current results. Similarly, Specht and Williams (1984) and Rowntree *et al.* (2013) found a yearly increase in 100 seed weight of 0.10 g $year^{-1}$ across all the maturity groups evaluated and 0.017 ($\pm$0.008) for maturity group III respectively. Over and above this, the study observed significant variations among the test genotypes. Therefore, selection for higher yield over time possibly favoured genotypes with greater seed weight.

3.7.3.4 Changes in lodging resistance

Generally, a decrease in percentage lodged plants at maturity was observed. The study showed that the past 48 years has resulted in an actual decline of percentage lodged plants at maturity of 0.31% per year when regressed to the founder cultivar which represented annual declining rate of 2.08%. The realized gain that was obtained from the rust tolerant cultivar was -0.33% $year^{-1}$ translating to annual improvement rate of -0.02%. In all cases, lodging resistance was enhanced. Similar results were presented by Voldeng *et al.* (1997). Justin (2012) postulated that soybean cultivars that succumb to lodging negatively impact on grain yield. Results revealed that 4 out of 42 cultivars showed lodged plants at maturity that varied between 5% and 28%. It can be argued that soybean breeders were practising negative selection where genotypes showing lodging were discarded, consequently increasing the gene frequency for lodging resistance which culminated in higher grain yield over time.

3.7.3.5 Changes in quality traits (seed protein and oil concentration)

Results revealed an annual decline in seed protein concentration over cultivar year of release. The estimated decline over the 48 year period was found to be 0.01% which equated to an annual decreasing rate of 0.03%. The annual decreasing rate in seed protein concentration was accompanied by an improvement in seed oil concentration. Seed oil concentration increased with year of release by 0.02%. This amounted to an annual improvement rate of 0.11% over 48 year period of breeding and selection. Probably, the decrease in seed protein concentration suggested that nitrogen was used to produce more nodes and was equally diverted to the seed where it was used to increase the size. These results were in agreement with previous research (Morrison *et al.,* 2000; Rowntree *et al.*, 2013). This trend also supported the well documented and long standing evidence that seed protein and oil percentages are negatively correlated (Panthee *et al.*, 2005; Yaklich *et al.*, 2002). The relationship between seed protein and oil content is that, a 1-unit increase in oil content is associated with a 2-unit decrease in protein percentage (Specht *et al.*, 1999).

3.7.3.6 Changes in resistance to pod shattering

The data showed that the number of days from 95% pod maturity to first pod shattering has been increased in relation to cultivar year of release, which essentially means that resistance to pod shattering was improved. The observed improvement rate for the number of days over 48 year period was 0.38 days year^{-1} and this represented an annual percentage gain of 1.56. It was also observed that the rust tolerant genotype showed an improvement of 0.26 days year^{-1} representing an annual breeding gain of 0.01%. Pod shattering has been reported to cause seed losses of 50-100% in susceptible cultivars (Tukamuhabwa *et al.*, 2002). They also quantified the seed loss per hectare and found yield losses that ranged from 0 to 186 kg ha^{-1} depending on genotype, location, season and harvesting date. These findings justify the importance of improving pod shattering resistance. The present study showed a tendency towards improving pod shattering resistance over time consequently increasing grain yield, implying that breeders were selecting for high levels of pod shattering resistance.

3.7.3.8 Changes in pod height, plant height and diseases

The study also showed that over 48 years of breeding and selection, improvements have been made in pod height and plant height. Increase in pod height meant that all the pods could be harvested by a combine, minimizing grain yield losses.

3.8 Conclusion

The objectives of this study were to: (i) evaluate the genetic variability, heritability and agronomic performance of historical and current cultivars, (ii) measure the genetic gain of soybean grain yield and other agronomic traits for all the cultivars that were introduced, bred and developed in Zimbabwe over 70 years and (iii) to identify genotypes with good agronomic value that can be used as parents in current and future breeding programme. The study revealed wide variability among the cultivars in respect of analysis of variance, coefficient variation and ranges for the studied traits except seed protein and oil percentages. This means that there is considerable and exploitable genetic variability by breeders. The new cultivars performed better than the old ones. Genotypes, G16, G15, G17, G1 and G42 exhibited superior performance in grain yield and other agronomic traits and are therefore, recommended for utilisation in hybridisation program. Interestingly, the genotypes G28 and G25 which were released in 1989 and 1998 respectively remained competitive and accordingly should be used for cross breeding with a view to avoid loss of genetic diversity. Generally, phenotypic coefficient of variation was higher than the corresponding genotypic and environmental coefficient of variation for all the characters indicating relatively large influence by the environment. However, small differences were shown on grain yield, downy mildew scores, plant height and days to 95% pod maturity implying less influence of the environment on these traits. The magnitude of heritability in the broad sense was from moderate to high for grain yield, pod height, plant height, seed appearance scores, percentage purple stain seed, downy mildew scores and 100 seed weight. Days to first pod shattering, grain yield, red leaf blotch, downy mildew scores and 100 seed weight exhibited moderate genetic advance as a percentage of the mean while pod height, seed appearance scores, percentage purple stained seed and bacterial blight scores showed high genetic advance as a percentage of the mean. High heritability accompanied by moderate genetic advance was observed on 100 seed weight indicating the presence of additive gene effects, therefore, selection would be useful and further exploitation could be done to improve grain yield.

The estimated average breeding gain over cultivar year of release was found to be 47 kg ha^{-1} $year^{-1}$ and 41 kg ha^{-1} $year^{-1}$ for the best conventional genotype and the best and most recent leaf rust tolerant genotype respectively. However, the introgression of rust resistance into the locally adapted material resulted in a decline in grain yield suggesting that a linkage drag was achieved. Therefore, it is suggested to investigate interactions between linkats and loci underlying seed yield and soybean rust resistance using marker assisted selection method.

Soybean yield gain over cultivar year of was associated with 100 seed weight. Improvement in lodging resistance, pod shattering resistance, disease tolerance (shown by low rating scores) and increase in pod height contributed to increase in grain yield. Maturity also showed a tendency of increasing seed yield. The decrease in seed protein percentage was associated with an increase in seed oil concentration. Based on the improvements that have been observed, it can be concluded that the soybean breeders were applying the best selection strategies. However, further research is required to quantify the yield gains in high stress, low-yielding environments and farmers' fields as this study was done on high yielding environments. It is further suggested to evaluate the improvement that has been realized on diseases. Overall, results indicated that emphasis should be refocused on grain yield to restore the original linear increase. It is proposed that the Zimbabwean soybean program should adopt new technologies in order to enhance yield growth.

3.9 References

Aditya, J. P., 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.

Agarwal, A. P., S. A. Patil and P. M. Salimath. 2001. Identification of potential soybean genotypes for pod shattering resistance and seed yield. Crop Improvement 27:236-239.

Allard, R. W. 1960. Principles of Plant Breeding, John Wiley and Sons. Inc. New York, United States. p. 160-172

Aravind, G. 2006. Genetic Variability and Diversity Studies in Soybean [*Glycine max* (L.) Merrill]. Genetics and Plant Breeding, MSc thesis. Dharwad University of Agricultural Sciences, Karnataka, India. p. 1-103.

Basavaraja, G. T. 2002. Studies on induced mutagenesis in soybean. Genetics and Plant Breeding, MSc thesis. Dharwad University of Agricultural Sciences, Karnataka, India. p. 1-80

Bezaweletaw, K., P. Sripichitt, W. Wongyai, and V. Hongtrakul. 2006. Genetic variation, heriatbility and path-analysis in Ethiopian Finger Millet [*Eleusine coracana* (L.) Gaertn] Landraces. Kasetsart Journal of Natural Science 40:322-334.

Bhat, S., G. T. Basavaraja, and P. M. Salimath. 2012. Analysis in segregating generation of soybean [*Glycine max* (L.) Merrill]. Karnataka Journal of Agricultural Science 25:176-178.

Bhatia, V. S., P. Singh, S. P. Wani, G. S. Chauhan, R. Kesava, A. V. R. Mishra and and Srinivas K. 2008. Analysis of potential yields and yield gaps of rainfed soybean in India using CROPGRO-Soybean model. Agricultural and Forest Meteorology 148:1252-1265.

Burton, C. W. and E. H. Devane. 1953. Estimating heritability in tall Festuca (*Restuca arundinaceae*) from donar material. Agronomy Journal 45:1476-1481.

Chand, N., S. R. Vishwakarma, O. P. Verma and M. Kumar. 2008. Worth of genetic parameters to sort out new elite barley lines over heterogenous environments. Barley Genetics Newslett 38:10-13.

Cox, T. S. J. P. Shroyer, L. B. Hui, R. G. Sears and T. J. Martin. 1988. Genetic improvement in agronomic traits of hard red winter wheat cultivars 1919 to 1987. Crop Science 28:756-760.

De Bruin, J. L. and P. Pedersen. 2008. Soybean seed yield response to planting date abd seeding rate in the Upper Midwest. Agronmoy Journal 100:696-703.

De Bruin, J. L. and P. Pedersen. 2009. New and old cultivar responses to plant density and intercepted light. Crop Science 49:2225-2232.

Diers, B.W. 2010. Progress from 80 years of breeding soybean in the United States, agronomy Day, Department of Crop Science, University of Illinois. p.1-2

Dilnesaw, Z., S. Abadi, and A. Getahun. 2013. Genetic variability and heritability of soybean [*Glycine max* (L.) Merrill] genotypes in Pawe district, Metekel zone, Benishangule Gumuz regional state, northwestern Ethiopia. Wudpecker Journal of Agricultural Research 2:240-245.

Demarly, Y. 1979. "The concept of linkat". *In* Proceedings of the conference on broadening genetic bases of crops (eds). PUDOC, Wageningen, Netherlands. p. 257-265.

Duvick, D. N. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). Maydica 50:193-202.

Egli, D. B. 2008. Soybean yield trends from 1972 to 2003 in mid-western USA. Field Crops Research 106:53-59.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics, Longman, Harlow, United Kingdom. p. 120.

Feil, B. 1991. Breeding progress in small grain cereals-A comparison of old and modern cultivars. Plant Breeding 108:1-11.

Gay, S., D. B. Egli, and D. A. Reicosky. 1980. Physiological aspects of yield improvement in soybeans. Agronomy Journal 72:387-391.

Gebre, B.G. 2005. Genetic variability and inheritance of drought and plant density adaptive traits in maize. MSc (Agric) thesis. Department of Plant Sciences, Plant Breeding, University of the Free State, Bloemfontein, South Africa.

Hyten, D. L., Q. Song, Y. Zhu, I. Y. Choi, R. L. Nelson, J. M. Costa, J. E. Specht, R. C. Shoemaker and P. B. Cregan P.B. 2006. Impacts of genetic bottlenecks on soybean genome diversity. National Academy of Sciences of US 103:1666-1671.

Jin, J., X. Liu, G. Wang, L. Mi, Z. Shen, X. Chen and S. J. Herbert. 2010. Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in North East China. Field Crops Research 115:116-123.

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

Justin, J. S. 2012. Genetic gain x management interactionin soybean: seeding rate effect. MSc thesis. Crop Science, University of Illinois, Urbana-Champaign, United States. p. 1-54.

Justin, R. 2010. Evaluation of current and historical soybean genotypes for yield and forage use. Crop Science 61:321-380.

Kahlon, C. S. 2001. Analysis of genetic improvement for soybean from 1950-2000. PhD thesis. Plant, Environment and Soil Sciences, Louisiana State University, United States. p.1-120.

Khan, S., A. Latif, S. Q. Ahmad, F. Ahmad and M.Fida. 2011. Genetic variability analysis in some advanced lines of soybean (*Glycine max* L.). Asian Journal of Agricultural Sciences 3:138-141.

Khanh, T. D., T. Q. Anh, B. C. Buu and T. D. Xuan. 2013. Applying molecular breeding to improve soybean rust resistance in Vietnamese elite soybean. American Journal of Plant Sciences:4:1-6.

Kopisch-Obuch, F. J., R. L. McBroom and B. W. Diers. 2005. Association between soybean cyst nematode resistance loci and yield in soybean. Crop Science 45:956-965.

Lange, C. E. and L. C. Federerizzi. 2009. Estimation of soybean genetic progress in the south of Brazil using multienvironmental yield trials. Science Agriculture 66:309-316.

Luque, S. F. A. G. Cirilo and M. E. Otegui. 2006. Genetic gains in grain yield and related physiological attributes in Argentine maize hybrids. Field Crops Research 95:383-397.

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., J. W. Burton, T. Balesevic, D. Miladinovic, V. Djordjevic and V. Djukic . 2010. Soybean breeding: comparison of the efficiency of different selection methods. Turkey Journal of Agriculture 35:469-480.

Moose, S. P. and R. H. Mumm. 2008. Molecular Plant Breeding as the Foundation for 21st century crop improvement. Plant Physiology 147:969-977.

Morrison, M. J., H. D. Voldeng and E. R. Cober E.R. 2000. Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. Agronomy Journal 92:780-784.

Nyquist, W. E. 1991. Estimation of heritability and prediction of selection response in plant populations. Critical Review in Plant Science 10:235-322.

Okonkwo, F. A. and D. O. Idahosa. 2013. Heritability and correlation studies of yield characters in some soybean (*Glycine max*) varieties in Ekpoma. American Journal of Research Communication:1-9.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton and C. E. Sams. 2005. Quantitative trait loci for seed protein and oil concentration and seed size in soybean. Crop Science 45:2015-2022.

Ramteke, R., V. Kumar, P. Murlidharan and D. K. Agarwal. 2010. Study on genetic variability and traits interrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. Electronic Journal of Plant Breeding 1:1483-1487.

Robinson, H. F., R. E. Comstcok and V. H. Harvey. 1949. Estimates of heritability and degree of dominance in corn. Agronomy Journal 41:353-359.

Rowntree, S., S. Conley and P. Esker. 2012. Breeders vs agronomists: What we learned from the soybean decades study, Proceedings of the 2012 Wisconsin Crop Management Conference, United States. p.1:4

Rowntree, S. C., J. J. Suhre, N. H. Weidenbenner, E. W. Wilson, V. M. Davis, S. L. Naeve, S. N. Casteel, B. W. Diers, P. D. Esker, J. E. Specht and S. P. Conley. 2013. Genetic gain x management interactions in soybean: I. Planting Date. Crop Science 53:1128-1138.

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

Shrivastava M. K. and R. S. Shukla. 1998. Genetic analysis for yield and its components in soybean under different environment. Crop Research 16:196-201.

Singh, G. P., K. R. Maurya, B. Prasad and A. K. Singh. 1994. Genetic variability in *Capsicum annum* (L.). Journal of Applied Biology 4:19-22.

Sirohi, S. P. S., M. Sanjai, S. P. Singh and Ramasharya Y. 2007. Genetic variability, correlation and path coefficient analysis for seed yield and its components in soybean [*Glycine max* (L.) Merrill]. Progressive Agriculture 7:119-123.

Specht, J. E., D. J. Hume and S. V. Kumudini. 1999. Soybean yield potential-A genetic and physiological perspective. Crop Science 39:1560-1570.

Specht, J. E. and J. H. Williams. 1984. Contribution of genetic technology to soybean productivity-Retrospect and prospect (eds.). *In*: W.R. Fehr, editor, Genetic contributions to yield grains of five major crop plants, CSSA Spec. Publ. 7. CSSA and ASA, Madison, Wisconsin, United States. p. 49-74.

Tefera, H., A. Y. Kamara, B. Asafo-Adjei and K. E. Dashiell. 2010. Breeding progress for grain yield and associated traits in medium and late maturing promiscuous soybean in Nigeria. Euphytica 175:251-260.

Tukamuhabwa, P., P. R. Rubaihayo and K. E. Dashiell. 2002. Genetic components of pod shattering in soybean. Euphytica 125:29-34.

USDA-NASS. 2011. National Agricultural Statistics Service, USDA-NASS, Washington, DC. www.nass.usda.gov/quick-stats/.

Voldeng, H. D., E. R. Cober, D. J. Hume, C. Gillard and M. J. Morrison. 1997. Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. Crop Science 37:428–431.

Wiggins, B. T. 2012. Heritability and Genetic Gain of Seed Protein, Oil, and Yield among RIL of soybean, Plant Science, University of Tennessee, Knoxville. p. 1-93

Wills, C. (2007) Principles of Population Genetics. Journal of Heredity 98:382.

Wray, N., and P. Visscher. 2008. Estimating trait heriability. Nature Education 1:1-29.

Yaklich, R. W., B. Vinyard, M. Camp and S. Douglas. 2002. Analysis of seed protein and oil from Northern and Southern Region Uniform Tests. Crop Science 42:1504-1515.

Yuan, J., V. N. Njiti, K. Meksem, M. J. Iqbal, K. Triwitayakorn, M. A. Kassem, G. T. Davis, M. E. and D. A. Lightfoot. 2002. Quantative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Science 42:271-277.

CHAPTER 4

GENOTYPE X ENVIRONMENT INTERACTION AND YIELD STABILITY ANALYSIS OF THE ZIMBABWEAN SOYBEAN GERMPLASM ACROSS DIFFERENT ENVIRONMENTS

4.1 Abstract

The occurrence of genotype x environment interaction (GEI) complicates selection of superior cultivars for breeding and commercialization in heterogeneous environments consequently slowing breeding progress from selection. In this regard, multi-environmental trials (METs) are conducted to identify superior and stable cultivars for production. The objectives of this study were to identify the best performing genotypes for wide and specific adaptation and evaluate the GEI and level of yield stability of the test genotypes. Forty two genotypes from the Zimbabwe soybean breeding programme were grown for two seasons at 13 environments. Additive main effects and multiplicative interaction (AMMI), cultivar superiority index, and rank analysis were used to evaluate the stability performance of the genotypes. Mean grain yield ranged from 3347 kg ha^{-1} for genotype G2 to 5208 kg ha^{-1} for G28. Additive main effects and multiplicative interaction analysis showed that grain yield variation due to genotypes, environments and (GEI) were highly significant ($P \leq 0.001$). Environments explained the greatest proportion (77%) of the total treatment sum of squares followed by GEI (17.4%) and genotypes (5.6%), justifying the need to conduct METs and testing over many seasons. The magnitude of GEI was three times larger than that for genotypes indicative of differences in genotypic responses to test sites. The data set revealed that GEI was of a crossover type because of differential yield ranking of genotypes. Results of AMMI identified genotypes G28, G1, G15, G25 and G14 as the most productive and stable across the test locations, while cultivar superiority analysis identified genotypes G25, G1, G28, G15, G21 as the most productive and stable and rank analysis results identified genotypes G34, G1, G15, G41, and G24 as the genotypes for cultivation across 13 environments and seasons because they combined stability and above average yield. Genotypes, G1 and G15 were the most productive, consistent and stable as revealed by the three models. These could be recommended for deployment across the test environments and used as breeding sources. Overall, the trend in grain yield coupled with adaptation, demonstrated significant improvement in these traits over time. The study observed that the application of AMMI model, cultivar superiority and rank analysis are important for facilitation of comparison and identification of superior genotypes either for specific or wide adaptation.

Key words: AMMI, genotype x environment interaction, yield stability, specific adaptation, wide adaptation, soybean

4.2 Introduction

Soybean [*Glycine max* (L) Merrill] occupies an important position among the leguminous crops due to its nutritional value. Today, soybean has a special position as a source of edible vegetable oil and protein. It is considered as one of the crops that have enormous potential to provide the world's food and forage security (Alghamdi, 2004). This has triggered interest in soybean production in Sub-Saharan Africa (SSA) and the phenomenal growth in production is mainly driven by increased utilisation of the crop as a supplement in livestock feed (Jandong *et al.*, 2011). Therefore, the use of stable genotypes that combine superior yields is critical for sustainable production in the region.

However, soybean growing environments in SSA are located in ecologies that vary considerably with respect to climatic conditions, soil texture, soil depth, latitude, season length, seasonal variation, varying pH status, management regimes, altitude etc. Given the diverse and wide ecologies in Sub-Saharan Africa, it means that the varieties are exposed to the influence of genotype by environment interaction (GEI). Genotype by environment interaction has the potential to affect the productivity and profitability of the soybean cropping systems. Furthermore, GEI complicates testing, selection and release decision of superior genotypes consequently reducing genetic progress (Mohammadi *et al.*, 2010). As a result, GEI alters the genotype rankings from one environment to the other. This view supports Fox *et al.* (1997) who argued that significant GEI suggested that selections from a particular environment may not do well in another environment. Essentially, the existence of GEI hampers progress from selection because breeders would need to identify different genotypes for each environment. Abay and Bjørnstad (2009) observed that a significant GEI can be exploited by selecting stable genotypes for specific environments. Ramburan *et al.* (2012) reiterated the need to conduct trials over a range of locations, years and management regimes in order to guarantee that superior and stable genotypes are identified. To this end, multi-environmental trials (METs) are often conducted by breeders in order to identify superior genotypes or high yielding genotypes with consistent performance. Sreedhar *et al.* (2011) supported this view and pointed out that multi-location testing is a critical exercise meant to assess the adaptability of genotypes and their yield stability over various environments. According to Yan *et al.* (2001), METs also serve to identify test sites that best represent target environments. He also contended that the genotype main effects coupled with genotype by environment interaction account for the sources of variation relevant in cultivar testing. Gauch Jr. and Zobel (1996) observed that the distribution ratio of the treatment elements in GEI studies was 70:20:10 and in respect of this, they stressed

the need to conduct studies on GEI with a view to validate whether this is met. In a nutshell, the analysis of GEI has become a topical issue that helps to evaluate varieties for adaptation and selecting parents for base populations (Aina *et al.*, 2007).

Variability due to GEI may be attributed to either crossover interaction (COI) or non COI. However, crop breeders are more concerned with non COI which is relevant and applicable when selecting genotypes for general adaptation, whereas, COI is essential for identifying genotypes for specific adaptation (Matus-Cadiz *et al.*, 2003).There are several statistical methods that can be employed to detect and quantify the GEI. Of late, most researchers are using the Additive Main effects and Multiplicative Interaction (AMMI) and Genotype and Genotype by Environment Interaction (GGE) models. The AMMI was developed by Zobel *et al.*, 1988b; Zobel *et al.*, 1988a; Gauch, 1992; Gauch *et al.*, 1997 while GGE was developed by Yan, 2002; Yan and Kang, 2003; Ma *et al.*, 2004; Yan and Tinker, 2005; Yan *et al.*, 2007. Before the advent of these two models, crop scientists used to use the ANOVA developed by Snedecor and Cochran (1980), linear regression developed by Finlay and Wilkinson, 1963; Eberhart and Russel, 1966 and principal component analysis developed by (Hill and Godchild, 1981). Apparently, the AMMI and GGE models are widely used in mega-environment analysis, genotype evaluation, and test-environment evaluation.

Both GGE and AMMI models have an advantage over other tools in that they analyse complex patterns of genotype x environment interaction as opposed to linear regression which is appropriate when a linear response of the genotypes to the environments exists. However, there are some differences between AMMI and GGE. Firstly, AMMI uniquely separates genotype main effects (G), environments main effects (E) and partitions the genotype x environment (G x E) interaction, which is important for research purposes while GGE biplot gives all the components without environment main effects. Secondly, the AMMI model is capable of separating structural variation from noise with a view to achieving accuracy (Nassiri and Ariyo, 2011). It is against this background that AMMI was selected as a model of choice in this study. Cucolotto *et al.* (2007) assessed the adaptability and stability of 30 soybean genotypes using AMMI and simple linear regression and concluded that AMMI offered better precision and was more effective in explaining the environments and stability of the cultivars. Nwangburuka *et al.* (2011) employed AMMI and GGE models in genotype x environment interaction studies and produced similar results.

Given that the Zimbabwean soybean programme has expanded into other Sub-Saharan countries and the challenges of GEI mentioned above, it would prudent to investigate the performance and adaptation of the existing germplasm. Before utilisation of the germplasm, it would be logical to estimate its adaptability and suitability to varied production environments as a prime step. Moreover, the germplasm comprised of the pre-commercial genotypes and hence, it was important to quantify their adaptation and stability in the target environments. This is critical and logical because new soybean cultivar releases are expected to show consistent performance across sites and seasons.

The objectives of the study were to; (i) compare the performance of 42 genotypes and identify the best performing for wide and specific adaptation, and (ii) evaluate the GEI and level of yield stability of the 42 genotypes using the AMMI model. The hypothesis was that performance of test entries was the same and all cultivars were stable and adapted to the subtropics.

4.3 Materials and methods

4.3.1 Germplasm

Forty two soybean cultivars were evaluated in 2010/2011 and 2011/2012 cropping seasons. These represented a collection of all the soybean varieties that were bred and released in Zimbabwe from 1940 to 2013. Three cultivars (G35, G36 and G38) were entered in the trials as promising lines and were subsequently registered as illustrated in Table 4.1. Details of the germplasm used are included in Table 4.1 below.

Table 4.1: Test entries used in genotype x environment interaction studies

Entry No.	Code Name	Year of Release	Growth Habit
1	G2	1966	I
2	G3	1972	D
3	G4	1973	D
4	G5	1974	D
5	G6	1977	I
6	G7	1977	D
7	G8	1980	I
8	G34	1982	I
9	G10	1985	D
10	G11	1988	I
11	G31	1992	D
12	G13	1992	I
13	G29	1995	D
14	G15	1989	I
15	G16	1994	D
16	G17	1997	D
17	G37	1997	I
18	G27	1998	D
19	G19	1999	I
20	G20	1999	D
21	G22	1999	D
22	G40	1999	I
23	G21	2000	D
24	G23	2000	I
25	G24	2001	D
26	G18	2003	I
27	G26	2005	I
28	G25	2005	D
29	G39	2005	I
30	G28	2006	D
31	G14	2007	I
32	G30	2007	I
33	G12	2007	I
34	G41	2008	D
35	G42	2008	I
36	G32	2008	I
37	G33	2008	D
38	G9	2010	I
39	G35	2012	D
40	G36	2012	D
41	G1	2012	I
42	G38	2013	I

Key: I = Indeterminate; D = Determinate; G = Genotype

4.3.2 Environments

The field evaluation of the 42 cultivars was conducted in Zimbabwe, Malawi and Zambia over two cropping seasons (2010/11 and 2011/12). Six test environments were used during 2010/11 cropping season, while nine environments were used in 2011/12 season. Thus, 15 test environments were used in total but only 13 environments were successful. More details about the test environments are described in table 4.2 below;

Table 4.2: Environments used for evaluations of the test entries between 2010/11 and 2011/12 cropping seasons

Environment	Country	Year	Code	Latitude	Longitude	Altitude (masl)	Rainfall[1] (mm)
RARS	Zimbabwe	2010/11	E1	17°40'S	31°14'E	1341	686
GVTC	Zimbabwe	2010/11	E2	17°68'S	30°86'E	1449	712
Lusaka	Zambia	2010/11	E3	15°67'S	28°33'E	1300	860
Mpongwe	Zambia	2010/11	E4	13°59'S	28°00'	1219	1000
Bvumbwe	Malawi	2010/11	E5	15°55'S	35°04'E	1228	950
RARS	Zimbabwe	2011/12	E6	17°40'S	31°14'E	1341	749
GVTC	Zimbabwe	2011/12	E7	17°68'S	30°86'E	1449	712
Lusaka	Zambia	2011/12	E8	15°67'S	28°33E	1300	700
Mpongwe	Zambia	2011/12	E9	13°59'S	28°00'	1199	800
Bvumbwe	Malawi	2011/12	E10	15°55'S	35°04'E	1250	768
Lilayi	Zambia	2011/12	E11	15°33'S	28°30'E	1090	688
ART	Zimabwe	2011/12	E12	17°43'S	31°05E	1527	780
Chitedze	Malawi	2011/12	E13	13°85'S	33°85'	1146	643

[1]rainfall refers to the amount received during the two seasons including irrigation, RARS = Rattray Arnold Research Station, ART = Agricultural Research Trust, GVTC = Gwebi Variety Testing Centre; masl = metres above sea level; mm = millimetres

4.3.3 Experimental design

The experiment was designed as a rectangular lattice (6 x 7) design with three replications for each environment. The experimental unit consisted of six rows, 5 m long spaced at 0.45 m equating to a gross plot size of 13.5 m^2. The in-row spacing was 6.3 cm implying that 79 viable seeds were planted per row resulting in a plant population of approximately 350 000 plants ha^{-1}. The nett plot consisted four rows, 4.4 m long giving nett plot size of 7.92 m^2.

At harvest grain yield was measured on the nett plot basis following standard practice used at Seed Co (Seed Co Research and Technology, personal communication) and adjusted to kg ha^{-1} at 11% moisture using the following formulae;

*Grain Yield (kg ha^{-1}) = [Grain Weight (Plot yield in kg ha^{-1})/(100 - %MC) *10/Plot Area*111/100]*
Where; %MC = Grain moisture in percentage.

4.3.4 Management

Standard cultural practices for soybean production that included land or seedbed preparation, hand planting, weeding, herbicide application and crop protection were applied at each test environment. Fertilizer (compound L) was applied at a rate of 400 kg ha^{-1} supplying 28kg ha^{-1} of Nitrogen, 68kg ha^{-1} of P_20_5 and 40kg ha^{-1} of K_20. The seed was inoculated with *Bradrhizobium japonicum* inoculant Grasslands strain 1491 (Tichagwa, personal communication). Where irrigation facilities were available, supplementary irrigation was applied to the crop in times of need. At some sites the crop was guarded against animals and humans. All the trials were hand harvested.

4.3.5 Data analysis

All the 13 test environments (year-location combinations) were used for analysis. The grain yield data in kg ha^{-1} was analysed using the Additive Main Effects and Multiplicative Interaction model (AMMI) macro in GenStat 14 (VSN, 2011). According to Talbot *et al.* (2008), the AMMI model assumes all data to be from the randomized block design and utilizes the adjusted the means. The following model (Ebdon and Gauch Jr., 2002a) was used for the ANOVA;

"$Y_{ger} = \mu + \alpha_g + \beta_e + \theta_{ge} + \varepsilon_{ger,}$" and the AMMI model is as follows;

$$Y_{ger} = \mu + \alpha_g + \beta_e + {}^{N}\sum_{n=1} \lambda_n \varsigma_{gn} \eta_{en} + \rho_{ge} + \varepsilon_{ger,}$$

Where Y_{ger} is the grain yield level for genotype g in environment e for replicate r, μ is the grand mean, α_g are genotype mean deviations (mean minus the grand mean), β_e are the environment mean deviations, N is the number of singular value decomposition (SVD) axes retained in the model, λ_n is the singular value for SVD axis η, ς_{gn} are the genotype singular vector values for SVD axis n, θ_{ge} are the interaction residuals, ρ_{ge} are the AMMI residuals, and ε_{ger}, is the error term. The term ${}^{N}\sum_{n=1} \lambda_n \varsigma_{gn} \eta_{en} + \rho_{ge}$ is equivalent to the interaction term in the ANOVA model.

Ebdon and Jr. (2002a) noted that the eigenvalue for a given SVD axis is the sum of squares (SS) retained by that axis and it is equal to the square of the singular value, λ^2. The sum of the eigenvalues $^{N}\sum\lambda^2$ for the N axes, plus the residual SS for a reduced model, is equal to the GE interaction SS. Therefore, the interaction SS is partitioned by SVD into interaction SS and associated degrees of freedom, which allow for the use of F-Tests to determine the significance of a given SVD axis (Gauch Jr, 1992a).

There are numerous approaches that can be employed to arrive at the maximum number of interaction principal component axes (IPCAs) to use in the AMMI model. In this case, a full model was adopted (Gauch Jr, 1992b). Ideally, a full model is one with all the significant IPCAs. The computation suggested by Gauch Jr (1992b) targets to estimate the level of noise using statistics from the AMMI ANOVA. He defined noise as the difference between yield estimate and its true mean.

Therefore, the following equation Gauch Jr. (1992a) was used to estimate the percent level of noise in the GE interaction component;

[100 x (Interaction DF x EMS)]/Interaction Sum of Squares (SS)

Where: interaction DF is equal to interaction degrees of freedom, Expected Mean Square (EMS) is equal to the expected error mean square for the AMMI ANOVA, Interaction SS is equal to interaction sum of squares.

The number of IPCAs in the final model selected was the one with a residual sum of squares value either equal or close to the corresponding sum of squares for the estimated level of noise (Table 4.3). However, in the quest to determine the optimum number of IPCAs to use, Sivapalan *et al.* (2000) asserted that AMMI2 should be the highest in explaining the biological patterns.

One AMMI Biplot was plotted. This was AMMI1 Biplot where IPCA1 scores were plotted against genotype and environment means. IPCA scores against both genotype and environment means were all done for grain yield using IPCA1 scores. No attempts were made to construct biplots above AMM2 because they are generally considered to be complex and uneasy to produce on a two dimensional graph. Genotypes with IPCA scores closer to zero were classified as stable

with wide adaptation as opposed to genotypes with significantly larger IPCA scores (>0) which were classified as unstable, exhibiting specific adaptation to the prevailing environments (Crossa *et al.*, 1990; Crossa *et al.*, 1991).

Stability coefficients displaying cultivar superiority were also computed on GenStat 14th Edition (Payne *et al*., 2011). Stability of the genotypes across the environments were estimated by the cultivar superiority index (P*i*) in accordance with Lin and Binns (1988) as follows;

$Pi = \sum_{j=1}^{n}(Xij - Mj)2/2n$

Where n = number of locations; Xij = yield of the i^{th} cultivars in the j^{th} environment;

M_j = maximum yield recorded in the j^{th} environment.

Cultivar rank analysis was performed in GenStat 14th Edition in accordance with Huehn, 1990; Nassar and Huhn, 1987 as follows;

$S^1_i = \sum_{j<i} |rij - rij'| / [n(n-1)/2]$

$S^2_i = \sum_{j=i}(r_{ij} - r_i)2/n-1$

Where;

- S^1_i = mean of the absolute differences among the classification of the i^{th} cultivar in the n^{th} environments,
- r_{ij} = classification of the i^{th} cultivar in the **j**th environment,
- n = number of environments
- S^2_i = variance of the classifications of the **i**th cultivar in the environments.

Interpretations for the S^1 and S^2 rank analyses (Huehn, 1990; Nassar and Huhn, 1987)

- Genotypes with lowest S^1 and S^2 values would be the most stable,
- For a cultivar with the maximum stability $S^1 = S^2 = 0$

Simple ranking (a non-parametric measure) was also performed. Accordingly, rank stability measures define the ability of a genotype to stabilize itself in different environments; hence, it has the same concept as genotype x environment interaction measures (Schoeman, 2003). Genotypes with similar rankings across environments are classified as stable (Farshadfar *et al.*, 2011).

4.4 Results

4.3.1 AMMI ANOVA results

The results of the AMMI analysis of variance are shown in Table 4.3. The treatments (genotypes + environments + interactions) accounted for 87.7% of the total grain yield sums of squares using approximately 33.3% of the total degrees of freedom (Table 4.3). The genotypes captured 4.9% of the total sums of squares and 5.6% of the treatments sums of squares. On the other hand, the environments explained 67.5% of the total sums of squares and 77.0% of the treatments sums of squares. The interactions explained 15.3% of the total sums of squares and 17.4% of the treatments sums of squares. Therefore, the environments accounted for more variation followed by the interactions (genotype x environment interactions) and the genotypes captured the least variation.

For the interactions, IPCA1 explained 46.1% of the variation (sum of squares) using about 10.6% of the total interaction degrees of freedom. Addition of the second IPCA or when IPCA2 was fitted, the two IPCAs explained 58.5% of the total interaction variation using approximately 20.8% of the total interaction degrees of freedom. When the third IPCA was added, the model explained 69.1% of the total interaction using about 30.6% of the total interaction degrees of freedom. The first four IPCAs explained 76.6% of the total interaction variation using approximately 39.9% of the total interaction degrees of freedom.

Table 4.3: ANOVA for full AMMI model for grain yield (kg ha^{-1}) of 42 cultivars evaluated across 13 test environments during 2010/11 and 2011/12 cropping seasons

Source	DF	SS	MS	%Total SS Explained	% Treatment Explained	% Interaction SS Explained
Treatments	545	3795437270	6964105***	87.7		
Genotypes	41	210926715	5144554***	4.9	5.6	
Environments	12	2922339401	243528283***	67.5	77	
Block	26	62836136	2416774***			
Interactions	**492**	**662171153**	**1345876*** **	**15.3**	**17.4**	**1.00**
IPCA1	52	305345002	5872019***			46.1
IPCA2	50	82263682	1645274***			12.4
IPCA3	48	70168338	1461840***			10.6
IPCA4	46	49375483	1073380***			7.5
IPCA5	44	41744444	948737***			6.3
IPCA6	42	29959161	713313**			4.5
IPCA7	40	21735111	543378			3.3
IPCA8	38	20014050	526686			3.0
IPCA9	36	16705469	464041			2.5
Residuals	252	113274204	449501	2.6		
Error	1066	469556239	440484			
Total	1637	4327829645	2643757			

, *, indicates that the term is significant at $P \leq 0.01$, $P \leq 0.001$; IPCA = Interaction principal component axis term1 to 9; DF = Degrees of freedom; SS = Sum of squares; MS = Mean square

4.3.2 Levels of noise and pattern in AMMI ANOVA

The level of noise reported herein of 32.73% (Table 4.4) was close to the one that was reported by Gauch Jr. (1992b). Results showed that the level of pattern to noise was about two fold higher than the noise. In this case, the best model was AMMI3 because the noise sum of squares of 216 728 618 fell between AMMI2 and AMMI3 but close to AMMI3 making AMMI3 the best model. Model selection was based on two methods that were; the proportional contribution of each IPCA to genotype x environment interaction and the ratio of noise sum of squares to residual sum of squares. The proportion of noise sum of squares for AMMI3 to its residual sum of squares was almost one (Table 4.5) making AMMI3 the most suitable model. Although AMMI3 was the best model, the biplot analysis was generated from IPCA1 because it explained 46.1% of the total interaction sum of squares. In a similar regard, it also accounted for the highest percentage of the level of pattern (0.461 x 67.27= 31%) relative to other IPCAs.

Table 4.4: Levels of noise and pattern in the grain yield interaction sums of squares

Attribute	%Level	Sum of Squares
Noise	32.73	216728618
Pattern	67.27	445442535

Table 4.5: Residual sums of squares for grain yield in each AMMI model

AMMI model fitted	Residual Sum Squares (RSS)	Noise Sum of Squares (NSS)	NSS/RSS	Pattern Sum of Squares (PSS)	PSS/R SS
AMMI-1	356826151	216728618	0.61	445442535	1.3
AMMI-2	274562469	216728618	0.79	445442535	1.6
AMMI-3	204394131	216728618	1.06	445442535	2.2
AMMI-4	155018648	216728618	1.40	445442535	2.9
AMMI-5	113274204	216728618	1.91	445442535	3.9
AMMI-6	83315043	216728618	2.60	445442535	5.3
AMMI-7	61579932	216728618	3.52	445442535	7.2
AMMI-8	41565882	216728618	5.21	445442535	10.7
AMMI-9	24860413	216728618	8.72	445442535	17.9

4.3.3 AMMI selections for the highest four yielding cultivars across 13 environments

Table 4.6 presents the best four selections from each test environment. The genotype which appeared in the top four environments in at least seven environments (env.) was G28, which was followed by; G27 (four env.), G16 (four env), G26 (four env.), G21 (three env.), G14 (three env.), and G15 (three env.). The other cultivars, G1, G36, G39, G29, G22, G42, G35G31, G5, G23, G8, G25, G17, G6, G40 and G8 appeared once or twice .

Table 4.6: Ranking of the first four AMMI selections per environment for grain yield (kg ha^{-1})

Environment	Environment Code	Season	Mean (kg ha^{-1})	IPCA Score	Rank 1	2	3	4
RARS	E1	2010/11	3112	16.8037	G29	G22	G15	G42
GVTC	E2	2010/11	3291	11.1488	G27	G26	G28	G42
LUSAKA	E3	2010/11	5717	42.8910	G27	G26	G28	G37
MPONGWE	E4	2010/11	7628	60.9470	G1	G36	G39	G27
BVUMBWE	E5	2010/11	6714	13.8021	G26	G28	G35	G27
RARS	E6	2011/12	4662	-19.2379	G14	G15	G6	G16
GVTC	E7	2011/12	3617	-10.0569	G16	G23	G28	G21
LUSAKA	E8	2011/12	3898	-42.3535	G40	G7	G8	G25
MPONGWE	E9	2011/12	3300	-18.2713	G22	G16	G28	G7
BVUMBWE	E10	2011/12	3941	-16.9078	G28	G25	G21	G8
LILAYI	E11	2011/12	4924	-25.7578	G14	G7	G21	G28
ART	E12	2011/12	4476	5.8355	G16	G31	G5	G26
CHITEDZE	E13	2011/12	3753	-18.8430	G15	G17	G14	G1

kg ha^{-1} = kilograms per hectare

4.3.4 Ranking of genotypes according to AMMI ANOVA

The mean yields of the top and bottom 21 yielding genotypes grown in 13 environments across two cropping seasons including the first IPCA scores, are shown in Table 4.7. The table also shows the mean yield and minimum and maximum grain yields for each environment. Generally, the mean grain yield of the test genotypes differed from one environment to the other ranging from 3 112 to 7 628 kg ha^{-1}. The highest yielding environment was Mpongwe, 2010/11 (E4) having a mean yield of 7 628 kg ha^{-1} while the lowest yielding environment was RARS, 2010/11 (E1) with a mean grain yield of 3 112 kg ha^{-1}. Environment 1 (RARS, 2010/11) recorded the lowest minimum grain yield (1771 kg ha^{-1}).

Table 4.7: AMMI IPCA1 scores and Grain Yield (kg ha^{-1}) of selected top and bottom yielding genotypes across 13 environments

				RARS (E1)	GVTC (E2)	Lusaka (E3)	MDC (E4)	Bvumbwe (E5)	RARS (E6)	GVTC (7)	Lusaka (8)	MDC (E9)	Bvumbwe (E10)	Lilayi (E11)	ART (E12)	Chitedze (E13)
RANK	**NAME**	**IPCA1**	**MEAN**	**10/11**	**10/11**	**10/11**	**10/11**	**10/11**	**11/12**	**11/12**	**11/12**	**11/12**	**11/12**	**11/12**	**11/12**	**11/12**
Top 21 Yielding Genotypes																
1	G28	0.5	5208	3594	3839	5497	5155	5827	4905	4126	4628	4395	**5436**	5618	4127	4056
2	G1	7.3	5098	**4229**	3610	4770	6**440**	5172	4863	4019	4624	3329	4765	5379	4816	4482
3	G25	-3.3	5084	3122	3507	4182	5649	5447	5175	3963	**4725**	3931	5107	5285	4798	4461
4	G15	4.4	5059	3692	3661	5028	5685	4865	5295	4071	4427	3031	4406	5334	4919	**4849**
5	G27	17.8	4937	3148	**4264**	**5854**	5915	5650	4718	3645	3543	3772	3264	5558	4805	3539
6	G21	-7.3	4924	3323	3439	4785	4590	4649	4930	4121	4547	4212	5022	5655	4158	4081
7	G14	-5.0	4868	2502	3280	4464	5147	5068	**5754**	4014	4181	3117	4704	**5880**	4160	4512
8	G42	2.0	4853	3630	3703	4877	5424	5385	4822	3761	4706	4023	4325	5419	2854	3659
9	G16	-21.2	4810	3040	3401	3165	2869	5110	5210	**4369**	4199	4398	4852	5349	**6292**	3772
10	G23	9.5	4801	3408	3688	4959	5133	5362	5132	4295	3371	2762	4025	5206	5033	3541
11	G35	10.9	4755	3187	3442	5181	5133	5825	4754	3543	3537	2622	4537	4589	4827	4132
12	G22	10.2	4748	3958	3692	4620	5401	4837	4232	3215	3730	**4402**	3959	4240	5082	3852
13	G26	16.9	4729	3384	3926	5742	4837	**6039**	4264	3704	3360	2641	3644	4753	5266	3414
14	G18	4.9	4722	3480	3494	4747	5685	4046	4336	3782	4410	3662	3842	5352	4179	3865
15	G17	4.2	4684	3206	3622	4253	5661	4258	4943	3493	4349	3017	2902	5263	4909	4516
16	G11	-5.7	4652	3085	3602	4217	4585	4477	4866	3390	4590	3956	3328	5395	4256	4223
17	G30	-1.7	4633	3547	3172	4739	3836	5542	5137	3933	3816	2475	4882	4572	4022	4050
18	G29	17.0	4632	4229	3566	4669	5424	5244	4318	3430	3329	2973	3852	3754	5182	3748
19	G24	12.1	4613	3557	3435	4459	5653	4764	4606	3771	3352	3207	3897	4639	4698	3434
20	G32	-0.7	4587	2989	3235	4131	4873	4282	4482	3991	4101	2968	3957	5285	5068	3764
21	G19	-7.0	4573	2519	3496	3949	4087	5204	4848	3723	4206	3396	3577	5542	4779	3618
Mean			4541	3112	3291	4680	5104	5098	4662	3617	3898	3300	3941	4924	4476	3753
Min		0.1	3347	1771	2294	2469	4184	5490	2357	1708	1762	2156	2099	2503	2305	1953
Max		25.6	5208	4229	4264	5854	6440	6039	5754	4369	6502	4402	5436	5880	6292	4849

RANK	NAME	IPCA1	MEAN	RARS (E1) 10/11	GVTC (E2) 10/11	Lusaka (E3) 10/11	MDC (E4) 10/11	Bvumbwe (E5) 10/11	RARS (E6) 11/12	GVTC (7) 11/12	Lusaka (8) 11/12	MDC (E9) 11/12	Bvumbwe (E10) 11/12	Lilayi (E11) 11/12	ART (E12) 11/12	Chitedze (E13) 11/12
Bottom 21 Yielding Genotypes																
22	G20	-1.9	4553	3572	3664	3861	4158	4931	4640	3662	4102	3532	3185	4789	5018	3575
23	G31	6.3	4548	3346	3180	4924	4613	3742	4225	4041	3179	3107	4239	4826	5593	3612
24	G8	-18.0	4538	3034	2996	4067	3049	5117	4867	3657	4795	3493	4947	5083	3312	4070
25	G10	-11.1	4480	3018	2746	3476	4127	4340	5089	4041	4228	2615	4708	4999	4248	4108
26	G36	16.3	4465	3228	3094	4039	6191	5165	4649	3281	3015	2854	4158	4092	4280	3503
27	G7	-27.9	4460	2432	2734	3089	3304	3899	4594	3691	5349	4227	4782	5730	3595	4050
28	G5	-18.3	4436	2131	2623	2811	4058	3529	4744	3780	4241	3661	4563	5420	5453	4157
29	G34	-11.2	4430	3187	3397	3962	3294	4499	4154	4086	4605	3356	3548	5419	4475	3108
30	G37	14.8	4368	3216	3207	5230	4887	4006	3766	3031	3358	2524	3861	4073	4456	3674
31	G13	-4.9	4362	2810	2699	3778	4092	4281	5012	3871	3232	2902	4711	4727	4435	3652
32	G6	-9.9	4341	3016	2966	3253	4540	3490	5229	3666	4036	3637	3541	5221	3536	3801
33	G39	21.6	4334	3006	3320	4817	6169	4090	3577	3025	2812	3582	3269	4390	4785	3001
34	G12	0.1	4293	2264	3504	2652	5190	5201	5118	3452	3825	2781	2099	5328	4740	3157
35	G3	12.6	4289	3393	3476	4510	4908	3957	4622	3537	2787	2858	2521	4572	4852	3269
36	G38	20.1	4258	3385	3382	4966	5908	4990	4051	2636	3409	2986	2995	4116	2732	3294
37	G40	-43.6	4206	2668	2586	2469	2249	4109	4931	3899	6502	2788	3637	5499	4023	4317
38	G33	22.2	4201	3610	3253	4897	4870	4721	4491	3888	1762	2156	3387	3974	4678	2430
39	G9	-2.5	4167	2793	2810	3705	3834	4128	4684	3294	3160	2824	3707	4333	4668	3730
40	G4	-28.2	4027	2107	2835	2918	1447	3954	4369	3008	4239	3988	3350	4905	4947	3781
41	G2	-28.0	3657	1771	2294	3245	1184	3664	5056	2285	3433	3959	3528	4487	2305	3832
42	G41	25.6	3347	2617	2401	3663	5116	4137	2357	1708	1910	2473	2521	2503	3646	1953

Underlined are the highest yield levels; IPCA = Interaction Principal Component Analysis 1; E1 = Rattray Arnold Research Station, 2010/11; E2 = Gwebi Variety Testing Centre, 2010/11; E3 = Lusaka West Farm, 2010/11; E4 = Mpongwe Development Centre, 2010/11; E5 = Bvumbwe Research Station, 2010/11; E6 = RARS, 2011/12; E7 = Gwebi Variety Testing Centre 2011/12; E8 = Lusaka Farm West, 2011/12; E9 = Mpongwe Development Centre, 2011/12; E10 = Bvumbwe Research Station, E11 = Lilayi Farm, 2011/12; E12 = Agricultural Research Trust, 2011/12; E13 = Chitedze Research Station, 2011/12. G1 to G42 represents the test genotypes.

4.3.5 AMMI Biplots

The biplot of the AMMI-1 result is presented in Figure 4.1. The abscissa shows main effects (genotype and environment means) while the ordinate represents the first IPCA scores. Genotype G27 had the largest positive (>10) interaction with the environments while G40 had the largest negative interaction with environments (~-43). Genotype G28 was the overall best performer; combining relative stability and high yield. Genotypes G28, G25, G15, G42 and G14 were above average in yield and relatively stable while G1 was above average, but relatively unstable. The most unstable and lowest yielding genotypes were G2 and G4. Genotype G12 was considered the most stable because its mean yield was equal to the grand mean (i.e. its mean yield was located at the origin of the graph), followed by genotypes G13, G20, G34, and 32, which were also very close to the grand mean.

The test environments showed variability in both main effects and interaction. Interestingly, all the environments from 2010/11 cropping season had IPCA scores that were greater than 0. The other environments had negative IPCA scores and were from 2011/12 season except E12. Environments E3, E4, E5, E6 and E11 were classified above the mean grain yield of all the environments whereas environments E1, E2, E7, E8, E9, E10 and E13 had below average seed yield (i.e. less than the grand mean yield level). The highest yielding environment was E4 while the lowest yielding environment was E1. However, environment E12 was classified as the average yielding environment and in this case, the environment average yield was 4541 kg ha^{-1}. Genotypes and environments with the same sign on the IPCA axis interacted positively whereas those with different signs interacted negatively. Environments E1 and E2 showed similarity in their interaction with the genotypes. Similarly, environments E7, E9, E10 and E13 also displayed similarity in their interaction with the genotypes. Genotype G40 had above average yield in environment E8. By the same token genotypes G16, G5 and G8 also had above average yields in environments E6 and E11. Similarly, genotypes G31, G17 and G18 had above average yields in environment E12. A considerable number of genotypes were clustered around the grand mean.

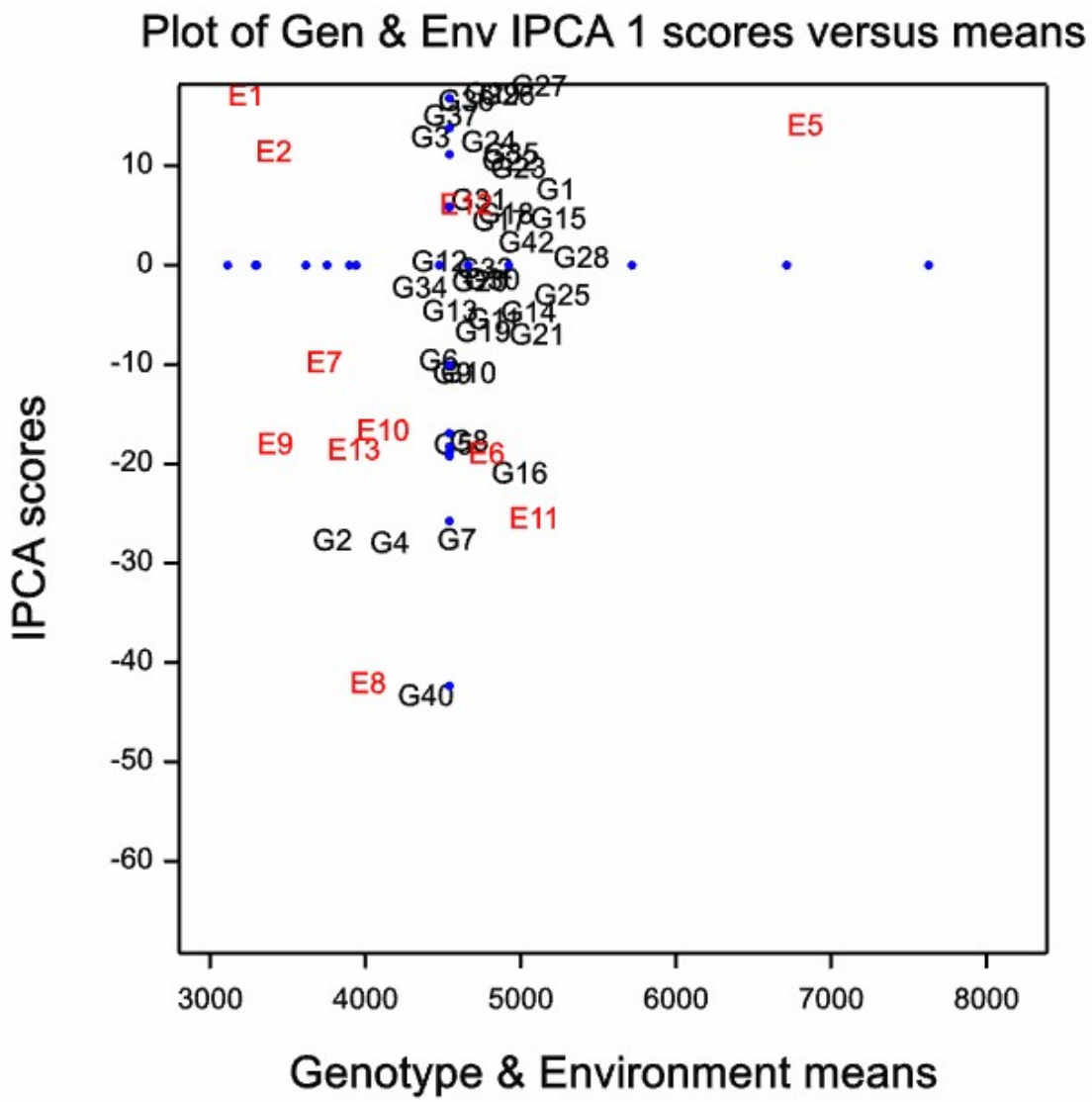

Figure 4.1: AMMI-1 biplot of IPCA 1 scores against grain yield for 13 environments.

The environments are; E1 = Rattray Arnold Research Station, 2010/11; E2 = Gwebi Variety Testing Centre, 2010/11; E3 = Lusaka West Farm, 2010/11; E4 = Mpongwe Development Centre, 2010/11; E5 = Bvumbwe Research Station, 2010/11; E6 = RARS, 2011/12; E7 = Gwebi Variety Testing Centre 2011/12; E8 = Lusaka Farm West, 2011/12; E9 = Mpongwe Development Centre, 2011/12; E10 = Bvumbwe Research Station, E11 = Lilayi Farm, 2011/12; E12 = Agricultural Research Trust, 2011/12; E13 = Chitedze Research Station, 2011/12. G1 to G42 represents the test genotypes

4.3.6 Cultivar Superiority Index values

Cultivar superiority index varied from 521 833 to 1 292 158 (Table 4.5). The smallest value represented cultivar G25 whereas the greatest value was for cultivar G16. The relative yield of the top 21 cultivars ranged between 100.7 to 114.7%. Genotype G28 had the highest relative yield followed by G1, while G19 was ranked number 21.

Table 4.8: Cultivar Superiority Index of top yielding genotypes across 13 environments

Code name	Yield kg ha^{-1}	%Mean	Cultivar Superiority Index values
G25	5084	111.9	521 833
G1	5098	112.2	524 178
G28	5208	114.6	565 077
G15	5059	111.4	579 880
G21	4924	108.4	738 208
G14	4868	1077.2	855 125
G27	4937	108.7	865 860
G42	4853	106.8	918 895
G18	4722	104.0	945 962
G23	4801	105.7	954777
G35	4755	104.7	982 732
G22	4748	104.5	989 227
G11	4652	102.4	1 037 688
G17	4684	103.1	1 060 160
G32	4587	101.0	1 094 081
G26	4729	104.1	1 108 735
G24	4613	101.5	1 120 157
G19	4573	100.7	1 196 375
G29	4632	102.0	1 211 673
G30	4633	102.0	1 235 065
G16	4810	105.9	1 292 158

G = denotes genotype; kg ha^{-1} = kilograms per hectare

4.3.7 Rank analysis

Table 4.9 illustrates the top 20 cultivars that were ranked according to ranking analysis. The rank indices ranged from 42.1 to 239.4. According to rank analysis, the lower the rank value the better the stability. In this case, the most stable variety was G34, followed by G1 with a high relative yield (112.27%). Although G28 had a slightly high rank value, it was the highest yielding (relative mean yield of 114.69%).

Table 4.9: Rank Analysis of top yielding genotypes across 13 environments

Name	Yield		Rank Analysis Values
	kg ha^{-1}	%Mean	
G9	4167	91.76	42.1
G1	5098	112.27	42.2
G15	5059	111.41	42.4
G41	3347	73.71	47.5
G24	4613	101.59	57.1
G32	4587	101.01	59.7
G13	4362	96.06	75.6
G19	4573	100.70	78.1
G21	4924	108.43	82.1
G3	4289	94.45	88.3
G11	4652	102.44	88.7
G25	5084	111.96	91.6
G18	4722	103.99	92.0
G23	4801	105.73	101.3
G17	4684	103.15	101.6
G28	5208	114.69	104.0
G20	4553	100.3	106.7
G10	4480	98.7	107.2
G9	4430	97.6	108.0
G42	4853	106.9	108.1

G = denotes genotype; kg ha^{-1} = kilograms per hectare

4.5 Discussion

4.4.1 AMMI ANOVA

The AMMI analysis of variance for the 42 genotypes evaluated across two seasons and 13 test environments revealed strong evidence that environment, genotype and genotype x environment effects were highly significant ($P \leq 0.001$) and accounted for 77%, 5.6% and 17.4% of the total treatment sum of squares respectively. This meant that the 42 genotypes and all the environments used were significantly different from each other. Furthermore, the results revealed that the environment component had larger influence on the performance of soybean genotypes, indicating the necessity for testing soybean genotypes at multi-location sites and over years. This is supported by the work of Gurmu *et al.*, 2009. In a similar regard, Asfaw *et al.* (2009) observed that environment main effects, explained 61.1%, genotype main effects 4.8% and GEI 34.1% of the total treatment variation. The magnitude of GEI effect was five times larger than that for genotypes indicating differences in genotypic responses to test environments.

The present data set revealed that the GEI were of the crossover type because of differential yield ranking of genotypes. However, the treatment elements failed to satisfy the expected distribution ratio of 70:20:10 for environment, genotypes x environment interaction and genotypes respectively that was observed by Gauch Jr and Zobel (1996). In this case, it is the environment which accounted for higher variation. One would have expected the GEI to be higher than what was obtained because the environments were sampled from three SADC countries thus, greater variability was expected. Generally, the more variable the environments, the greater the GEI. Gauch Jr and Zobel (1996) reported that when the environments are very different GEI can be expected to reach 60%. The observed genotype variation (5.6%) was below the expected. Genotypic grain yield (averaged across environments) ranged between 3 347 to 5 208 kg ha^{-1}. The GEI variation (17.4%) was close to the expected, possibly implying that the majority of the genotypes were widely adapted. This was supported the clustering pattern of a considerable number of genotypes around the grand mean yield with IPCA values close to zero (Figure 6.1). Moreover, it was also supported by the fact that the sum of squares for genotypes relative to IPCA 1 sum of squares was low. In addition, the results were in agreement with the findings of Gauch Jr (1992b) and Gauch Jr and Zobel (1996) who found sum squares for genotypes that were less than sum of squares for IPCA 1.

Given that the results of the study revealed a low GEI in respect of the expected proportion of the components of the treatment sum of squares (70:20:10), therefore, one would be able to recommend cultivars that combined both high grain yield and stability. Hence, G1 and 15 is recommended for cultivation in all the test environments. On the other hand, genotypes G2, G4, G5, G7, G16, G40, G17 G18 and G31 had high IPCA scores and therefore, are recommended for specific adaptation.

4.4.2 AMMI Biplots: Classification of genotypes and environments

There were several genotypes clustered around the grand mean with IPCA scores that were close to zero. Low IPCA values revealed low interaction with the test environments used, hence less responsive to environmental changes. The results of AMMI revealed that G28, G1, G25, G15, G14, G23, G18 and G42 were classified as the best genotypes combining both high stability and above average performance and these could be grouped under high yielding category. According to Fox *et al.* (1997), genotypes found in the top third of the yield table when are regarded as generally well adapted and in this case, the said cultivars were all in the top 14. Genotype G12 had the highest stability rating (IPCA score =0), thus, it was identified as having a combination of low GEI and average yield. This made it the most suitable variety for cultivation across sites and seasons. However, its low yield potential makes it unattractive to farmers. Genotypes G13, G6, G5, G7, G8, G31, G40 and G37 also showed average yields and could be classified under the medium yielding category. However, genotypes G16, G5, G7, G8 G13 and G37 had high IPCA values indicating that they were responsive to changes in environments. Genotypes G2 and G4 recorded low yields with high IPCA values and these were classified under the low yielding category. Genotypes, G5, G8 and G16 showed high interaction with environments E6 and E11 implying specific adaptation. Similarly, genotypes, G31, G17 and G18 also showed high interaction with environment E12. The genotypes that showed large interaction with particular environments were found to be unpredictable in performance (unstable) and hence could be recommended for specific adaptation.

The differences among the test environments and their clustering pattern could be explained by latitude, altitude, climatic conditions, season length and seasonal effects. Higher rainfall was received in environments; E3, E4 and E5 compared to E1 and E2 during 2010/2011 cropping seasons. Moreover, the latter environments were from Zimbabwe, whereas the other sites were sampled from Zambia. As a result, environments from Zambia were placed in a separate quadrant and classified as the highest yielding environments (Appendix 4.1). Environments from

2011/12 were distributed on the negative side of the horizontal dotted line (interaction score line of zero) as opposed to test environments from 2010/11 cropping season hence, had negative IPCA scores except E12. Environment, E12 (ART Farm was sampled from Zimbabwe and was expected to be clustered together with E1 and E2. Though it is difficult to explain why E12 was categorized in the same quadrant as E3, E4 and E5, it is however, thought that soybeans at ART farm have a longer growing period which is presumably similar to the Zambian sites. In the same vein, long growing season is reported to be correlated to high yielding potential (Miladinovic *et al.*, 2006). Environments E1, E2, E7, E8, E9, E10 and E10 had below average seed yield while E3, E4, E5 and E12 had above average seed yield. The unstable environments included E6, E8, E11 and E12. These exhibited a large interaction effect with certain genotypes. Most of the environments produced the least interaction effect suggesting that they were ideal for evaluation and selection in these environments would be effective because the performance of the genotypes will be fairly stable.

4.5.1 Relationship between IPCA scores and Cultivar Superiority

Both IPCA scores and cultivar superiority index identified genotype G1 and G25 to be highest yielding and most stable genotypes. The two approaches revealed positive correlation with high yield. This suggested that genotype; G1 and G25 were widely adapted and could be recommended for production in all the test environments. Therefore, the two stability parameters could be used for simultaneously selecting for high yield and stability. Rank analysis also identified G1 and G25 as the best candidates. However, few similarities were observed in genotype rankings among individual environments. Furthermore, high yielding and stable genotypes accounted for a significantly smaller proportion to the genotype x environment interaction than the medium and low yielding lines. These results are consistent with the findings of Mut *et al.*, 2009. There was agreement between the parametric and non-parametric methods in classifying the genotypes according to their stability. Hence, only one of these statistical tools would be adequate and useful to identify stable genotypes in a breeding programme for both growers and breeders attempting to select cultivars with predictable yield across environments.

4.5.2 Grain Yield Stability and adaptation

The study showed highly significant variation among the genotypes for grain yield which was attributed to genotype main effects, environmental main effects and genotype x environment

interactions. These results were consistent with numerous studies (Yothasiri and Somwang, 2000; Rao *et al.*, 2002; Gurmu *et al.*, 2009). Considering the AMMI results, genotypes G28, G1, G25, G14 and G15 were selected as the top five yielding and most stable genotypes. Interestingly, cultivar superiority analysis recorded G25, G1, G28, G15, and G21 as the top performers. Analysis of variance results revealed high yields from these genotypes G28 (5 208 kg ha^{-1}), G1 (5 098 kg ha^{-1}) , G25 (5059 kg ha^{-1}), G15 (5059 kg ha^{-1}), G21 (4924 kg ha^{-1}), G14 (4868 kg ha^{-1}) and above all, their relative yields ranged between 110 to 114.7% (Table 6.8 and 6.9) when compared to the grand mean. Based on their grain yield potential these genotypes can be classified into high yielding category. These genotypes were released between 2005 and 2012, except for G15 which was released in 1989. Clearly, the results showed that the newly released cultivars were superior in performance than the earliest generation cultivars. It can therefore, be argued that there has been sound breeding progress, implying that there was a significant built up of gene frequency for adaptation, productivity and stability. The performance pattern has increased significantly because the varieties from the earlier decade were poorly adapted and unstable as revealed by large IPCA scores and high cultivar superiority indices. Genotype G15 demonstrated wide adaptation because it was registered in Zimbabwe, Zambia and lately it has shown promise in Ethiopia. Thus, the results of cultivar superiority analysis complemented AMMI analysis results.

On the other hand, rank analysis (non- parametric method) results selected G9, G1, G15, G41 and G24 as the best genotypes. These genotypes were released between 2001 and 2013. The same picture is emerging where cultivars belonging to the current generation were identified. Therefore, the common genotypes that were selected by the three models are G1 and G25. These genotypes were also among the best four in 13 environments. Thus, the high mean performance coupled with high phenotypic stability shown by G1 and G25 suggested that these genotypes could be recommended for commercial production in wide range of environments. Although genotype, G28 was the best performing cultivar across the test environments, it failed to qualify as the most stable and productive genotype. Besides, it was the best in 7 of the 13 environments (Table 4.6) implying that it was the best performer. However, it was not selected in the top five high yielding and stable genotypes by the non- parametric stability measures. Ideally, a low rank index value indicates a combination of high yield and high stability (Mut *et al.*, 2009). In this case, G28 had a slightly high rank score. This may imply that it possessed dynamic stability suggesting that it was responsive to environmental changes. Mohammadi *et al.* (2009) described a genotype exhibiting dynamic stability as one that responds to improved

growing conditions and management practices with increased yield. Therefore, it would not be logical to recommend it for deployment across environments. It could however, be recommended for production in high yielding environments with long seasons because the mean number of days to maturity was above the grand mean.

Considering IPCA scores, Voltas *et al.* (2002) reported that distances from the origin (0,0) indicate the amount of interaction that genotypes exhibit over the environments or environments over the genotypes. Considering the current data set, genotypes G40, G16, G5, G8, G31, G17 and G18 expressed a highly interactive behaviour. These genotypes had large IPCA1 scores (positively or negatively). These genotypes were generally released between 1980 and 1990s. From the yield table, they fall under both low and medium yielding categories. These results demonstrated that these genotypes could be recommended for specific adaptation. However, genotypes G2 (1966 release) and G4 (1973 release) had the lowest mean yields coupled with large negative IPCA scores indicative of poor adaptation. These were classified into the low yielding category and this was confirmed with ANOVA, AMMI and cultivar superiority analyses.

As a generalization, AMMI analysis showed that 26 genotypes out of 42 (61.9%) had IPCA values between -10 and 10, possibly indicating average stability across environments. Clearly, it showed that stability was accumulated over time, bearing in mind that the breeding programme is over 70 years old indicating that adaptation and stability have been improved through breeding, extensive evaluation and selection. This was supported by the fact the founder variety, G2 (1966 release) showed poor stability (about 30 IPCA score) relative to G28 (2006 release) which had IPCA value close to zero. Studies have shown that yield stability is heritable and conditioned by additive gene action (Spehar, 1999). This implies that simple selection methods could be applied to advance yield stability and plasticity for cultivation over a wide range of environments. These results suggested that seed yield could be maximized through selecting genotypes showing consistently high yield performance across heterogeneous growing environments.

4.6 Conclusion

The objectives of the present study were to; (i) compare the performance of 42 genotypes and identify the best performing for wide and specific adaptation; and. (ii) evaluate the GEI and level of yield stability of 42 genotypes using AMMI model. The AMMI revealed the relative magnitude and significance of GEI effects and its interaction terms in relation to genotype and environmental effects. Results showed that GEI was a vital component of soybean yield variation and the biplots provided a good visualization of the response patterns of genotypes and environments. In addition, AMMI analysis was able to show the best genotypes across contrasting environments. The results identified genotypes G28, G1, G15, G14 and G25 as widely adapted. By the same token, cultivar superiority index selected G25, G1, G28, G15, and G21 as the superior genotypes that combined high yield, high stability and wide adaptability. Both selections from the AMMI and cultivar superiority analysis were classified under the high yielding category. On the contrary, rank analysis identified G9, G1, G15, G41 and G24 as the ideal genotypes. Genotypes G40, G16, G5, G8, G31, G17, G18 G2 and G4 were identified as unstable and suitable for specific adaptation. These selections fell into medium and low yielding categories.

The best model was AMMI3. Overall, the stability measures demonstrated that 62% of the genotypes had average stability across the 13 test environments while the rest were found to be unstable and suitable for specific adaptation. However, 5% exhibited below average stability viz, G2 and G4. Considering the three parameters (AMMI model, cultivar superiority index, rank analysis and mean yield), G1 and G25 were the best genotypes. Consequently these genotypes could be recommended for cultivation across the three countries and most importantly, can be used as breeding stock. The results also revealed an increase in grain yield over time coupled with improvement in adaptation as demonstrated by modern genotypes (G1 and G25), which combined relatively high productivity and stability.

The AMMI analysis identified E3 (Lusaka Farm), E4 (Mpongwe Development Centre) and E5 (Bvumbwe Research Station, Malawi) (Appendix 4.1) as the high yielding environments while E12 (ART Farm) was identified as the most stable environment with an IPCA score close to zero indicative of high yield and stability. Furthermore, the distribution and classification of the environments on the four quadrants symbolized diversity of the test environments. This implied that their influence on genotypic performance also differed.

4.7 References

Abay, F. and A. Bjørnstad. 2009. Specific adaptation of barley varieties in different locations in Ethiopia. Euphytica 167:181–195.

Aina, O. O., A. G. O. Dixon, and and E. A. Akinrinde. 2007. Additive main effects and multiplicative interaction (AMMI) analysis for yield of cassava in Nigeria. Journal of Biological Sciences 7:796-800.

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

Asfaw, A., F. Alemayehu, F. Gurumu and M. Atnaf. 2009. AMMI and SREG GGE biplot analysis for matching varieties onto soybean production environments in Ethiopia. Academic Journals 4:1322-1330.

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

Crossa, J., G. H. Gauch and R. W. Zobel. 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Science 30:493-500.

Cucolotto, M., V. C. Pipolo, D. D. Garbuglio, N. S. F. Junior, D. Destro and M. K. Kamikoga. 2007. Genotype x environment interaction in soybean: evaluation through three methodologies. Crop Breeding and Applied Biotechnology 7:270-277.

Ebdon, J. S. and H. Gauch Jr. 2002. Additive main effect and multiplicative interactionanalysis of national turfgrass performance trials 1. Interpretation of genotype x environment interaction. Crop Science 42:489-496.

Eberhart, S. A. and W. A. Russel. 1966. Stability parameters for comparing varieties. Crop Science 6:36-40.

Farshadfar, E., Mahmodi, N. and A. Yaghotipoor. 2011. AMMI stability value and simultaneous estimation of yield and yield stability in bread wheat (*Triticum aestivum* L.). . Australian Journal of Crop Science 5:1837-1844.

Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. Australian Journal of Biological Science 14:497-506.

Fox, P. N., J. Cross and I. Ramagosa. 1997. Multiple environment testing and genotype x environment interaction. In R. A. Kempton and P. N. Fox, (eds). Statistical Methods for Plant Variety Evaluation. Chapman and Hall, London, United Kingdom. p:742-754.

Gauch, G. H. 1992. Statistical Analysis of Regional Yield Trials-AMMI analysis of factorial designs. Elsevier Science, New York, United States. p. 1-278

Gauch, G. H., and and R. W. Zobel. 1997. Identifying mega environment and targetting genotypes. Crop Science 37:311-326.

Gauch Jr, H. G. 1992. Statistical analysis of regional yield trials. AMMI analysis of factorial designs. Elsevier, Amsterdam, The Netherlands. p. 53-110.

Gauch Jr, H. G. and R. W. Zobel. 1996. AMMI analysis of yield trials. *In* M.S. Kang and H.G. Gauch Jr. (eds). Genotype-by-environment interaction CRC Press, Boca Raton, Florida, United States. p. 85-122.

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

Hill, J. and N. A. Godchild. 1981. Analysing environments for plant breeding purposes as exemplified by multivariate analysis of long term wheat yields. Theoretical and Applied Genetics 59:317-325.

Jandong, E. A., M. I. Uguru and B. C. Oyiga. 2011. Determination of yield stability of seven soybean (*Glycine max*) genotypes across diverse soil pH levels using GGE biplot analysis. Journal of Applied Biosciences 43:2924-2941.

Lin, C. S. and M. R. Binns. 1988. A superiority measure of cultivar performance for cultivar location data. Canadian Journal of Plant Science 68:193-198.

Ma, B. L., W. Yan, L. M. Dwyer, J. Fregeau-Reid, H. D. Voldeng, Y. Dion and H. Nass. 2004. Graphic analysis of genotype, environment, nitrogen fertiliser and their interaction on spring wheat yield. Agronomy Journal 96:169-180.

Farshadfar, E., Mahmodi, N. and A. Yaghotipoor. 2011. AMMI stability value and simultaneous estimation of yield and stability in bread wheat (*Triticum aestivum* L.). Australian Journal of Crop Science 5:1837-1844

Huehn, M. 1990. Non-parametric measures of phenotypic stability. Part 1: theory. Euphatica 47: 189-194

Matus-Cadiz, M. A., P. Hucl, C. E. C.E. Peron, and R. T. Tyler. 2003 Genotype x environment interaction for grain colour in hard white spring wheat. Crop Science 43:219-226.

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.

Mohammadi, R., M. Aghaee, R. Haghparast, S. S. Pourdad, M. Rostaii, Y. Ansari, A. Abdolahi and A. Amari. 2009. Association among non-parametric measures of phenotypic ptability in four annual crops. Middle Eastern and Russian Journal of Plant Science and Biotechnology 3:20-24.

Mohammadi, R., M. Mohammadi, R. Karimizadeh and A. Amri. 2010. Analysis of genotype-by-environment interaction for grain yield of rainfed durum wheat genotypes in warm winter areas of Iran. Journal of Crop Science Biotechnology 13:267-274.

Mut, Z., N. Aydin, H. O. Bayramoglu and H. Ozcan. 2009. Interpreting genotype x environment interaction in bread wheat (*Triticum aestivum* L.) genotypes using non-parametric measures. Turkey Journal of Agriculture 33:127-137.

Nassa, R. and M. Huhn. 1987. Studies of estimation of phenotypic stability: tests of significance for non-parametric measures of phenotypic stability. Biometrics 43:45-53

Nassiri, 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 Agrobotanical Cluj-Napoca 39:220-225.

Nwangburuka, C. C., O. B. Kehinde, D. K. Ojo and O. A. 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). Archieves of Applied Science Research 3:193-205.

Ramburan, S., M. Zhou and M. T. Labuschagne. 2012. Investigating test site similarity, trait relations and causes of genotype x environment interactions of sugarcane in the Midlands region of South Africa. Field Crops Research 129:71-80.

Rao, M. S. S., B. G. Mullinix, M. Rangappa, E. Cebert, A. S. Bhagsari, V. T. Sapra, J. M. Joshi and R. B. Dadson. 2002. Genotype x environment interactions and yield stability of food-grade soybean genotypes. American Society of Agronomy 94:72-80.

Schoeman, L. J. 2003. Genotype x environment in sunflower (*Helianthus annus)* in South Africa. MSc thesis (Plant Breeding). Free State University, South Africa. p. 1-84.

Sivapalan, S., L. O. Brien, G. L. Ferrara, G. L. Hollamby, I. Barclay and P. J. Martin. 2000. An adaptation analysis of Australian and CIMMYT/ICARDA wheat germplasm in Australian production environments. Australian Journal of Agricultural Research 51:903-915.

Snedecor, G. W. and W. G. Cochran. 1980. Statistical methods, 7th Edition Iowa State University, Ames, Iowa State University Press, United States. p.1-593.

Spehar, C. R. 1999. Aiallel analysis for grain yield and mineral absorption rate of soybeans grown in acid brazilian savannah soil. Pesquisa Agropecuaria Brasileira 34:1003-1009.

Sreedhar, S., T. D. Reddy and M. S. Ramesha. 2011. Genotype x environment interaction and stability for yield and its componentsin hybrid rice cultivars (*Oryza sativa L*). International Journal of Plant Breeding and Genetics. 5:194-208.

Talbot, M., K. Brown and M. F. Smith. 2008. AMMI. Reference Manual Release 11.3. Procedure Library PL19 VSN International, Ltd, Hemel, Hempstead, United Kingdom.

Voltas, J., E. F. Van, E. Igartua, M. Garcia, L.F., J. L. Molina-Cano and I. Romagosa. 2002. Genotype x environment interaction and adaptation in barley breeding: Basic concepts and methods of analysis. *In* Safer G.A., Molina-Cano, J.L., Savin R., Araus J.L., and Romagosa, I (eds). Barley Science: Recent advances from molecular biology to agronomy of yield and quality. The Harworth Press Inc., New York, p. 205-241.

VSN. 2011. GenStat for Windows: fourteenth edition. Release 14.1. VSN International, Ltd, Hemel Hempstead, United Kingdom.

Yan, W. 2002. Singular-value partitioning in biplot analysis of multi-environment trial data. Agronomy Journal 94:990-996.

Yan, W., P. L. Cornelius, J. Crossa and L. A. Hunt. 2001. Two types of GGE biplots for analyzing multi-environment trial data. Crop Science 41:656–663.

Yan, W., B. M. Kang, S. Woods and P. L. Cornelius. 2007. GGE biplot versus AMMI analysis of genotype-by-genotype environment data. Crop Science 47:643-655.

Yan, W. and M. S. Kang. 2003. GGE biplot analysis: A graphical tool for breeders, geneticists and agronomists CRC Press, Boka Raton, Florida, United States. p. 1-271.

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

Yothasiri, A. and T. Somwang. 2000. Stability of soybean genotypes in Central Plain Thailand. National Science 34:315-322.

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

4.8 Appendices

Appendix 4.1

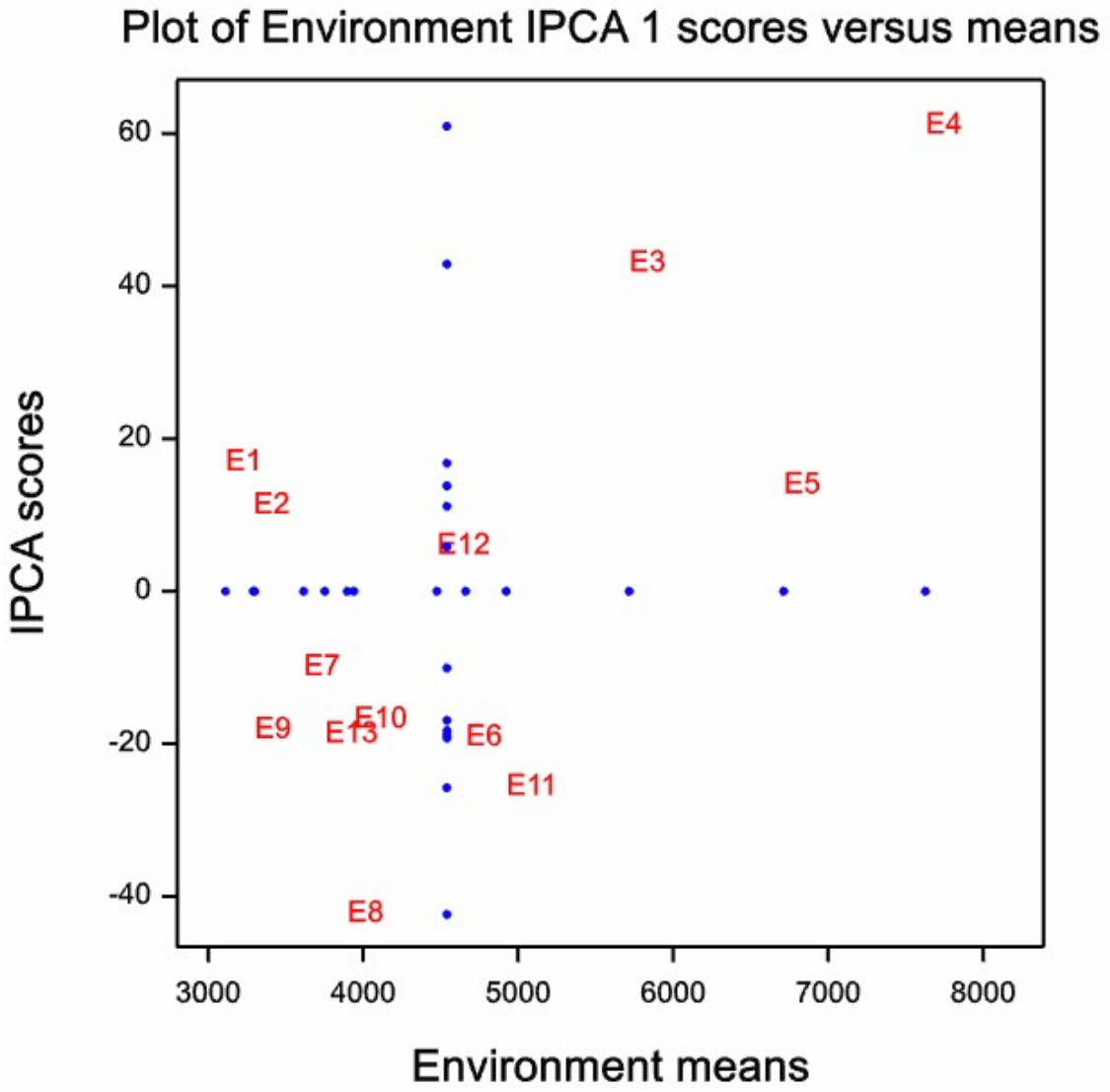

Figure 4.2: Plot of environment IPCA1 scores versus means

The environments are; E1 = Rattray Arnold Research Station, 2010/11; E2 = Gwebi Variety Testing Centre, 2010/11; E3 = Lusaka West Farm, 2010/11; E4 = Mpongwe Development Centre, 2010/11; E5 = Bvumbwe Research Station, 2010/11; E6 = RARS, 2011/12; E7 = Gwebi Variety Testing Centre 2011/12; E8 = Lusaka Farm West, 2011/12; E9 = Mpongwe Development Centre, 2011/12; E10 = Bvumbwe Research Station, E11 = Lilayi Farm, 2011/12; E12 = Agricultural Research Trust, 2011/12; E13 = Chitedze Research Station, 2011/12.

CHAPTER 5

CORRELATION AND PATH COEFFICIENT ANALYSIS FOR GRAIN YIELD AND ITS SECONDARY TRAITS OF ELITE ZIMBABWEAN SOYBEAN GERMPLASM

5.1 Abstract

High grain yield is partially a function of plant breeding. Knowledge of relationships between grain yield and the secondary traits in soybean is important particularly for selection. The objective of this study was to determine the correlations between grain yield and associated traits and direct and indirect effects of these traits on soybean yield. Forty two elite soybean breeding lines, replicated three times were planted at nine test environments during 2011/12 cropping season. A 6 x 7 rectangular lattice design was used. Significant differences were observed among the test genotypes for grain yield and most of the secondary traits studied. The mean grain yield ranged from 2441 to 5771 kg ha^{-1}. Five genotypes namely; G14, G25, G15, G40 and G21 were found significantly superior in yield and other major yield complementary traits. Grain yield showed positive and significant correlations with number of branches per plant (0.77***), number of nodes per plant (0.75***), shelling percentage (0.90***), red leaf blotch (0.31***) and number of days from 95% pod maturity to first pod shattering (0.54***). However, negative and significant correlation coefficients were found between grain yield with; plant height (-0.42***), percentage lodged plants at maturity (-0.46***), green stem scores (-0.31***), number of pods per plant (-0.31***), days from planting to 50% flowering (-0.60***), days from planting to 95% pod maturity (-0.43***), and percentage crude protein in the seed on a dry matter basis (-0.20*). Path coefficient analysis indicated that the number of nodes per plant gave the highest direct positive effect (0.48) on grain yield, followed by plant height (0.27) and 100 seed weight (0.20). Similarly, number of nodes per plant and plant height had the highest indirect effects on grain yield. Clearly, the results demonstrated that number of nodes and plant height could be recommended as reliable selection indices for developing high yielding genotypes of soybean. Overall, the results showed that genotype; G14 appeared to be the best candidate for future hybridizations due to high mean performance for both grain yield and associated traits.

Keywords: soybean, correlation, path coefficient analysis, grain yield

5.2 Introduction

Global annual soybean production is estimated to reach 371.3 million tonnes by 2030 (Haghii *et al.*, 2012). The increases in the said output would be a result of genetic and agronomic improvements. This gives the impression that high yielding cultivars should be made available for cultivation. However, the level of grain yield is a function of the interrelationships between yield and its secondary traits indicating that yield improvement entirely depends on the interaction of the various traits in influencing yield at both phenotypic and genotypic levels (Sudaric and Vrataric, 2002). Contextually, selection for grain yield alone may not bear fruits unless its associated characters are considered (Aditya *et al.*, 2011). Essentially, correlation and path coefficient analyses assist to reveal the important traits to be exploited, consequently qualifying the selection criteria. Therefore, accelerated grain yield increase can be accomplished through the performance of its secondary traits.

Grain yield is a quantitative trait with low heritability value, making the response to selection low. Thus, breeders should have good a understanding of the relationship between yield and its related characters in order to ensure effective and efficient manipulation of the given traits. The study of associations, direct and indirect effects of the secondary traits on grain yield is important for the success of the programme. Machikowa and Laosuwan (2011) reported that correlation and path coefficient analyses are useful tools employed in crop breeding to improve seed yield. In the same vein, Ariyo (1995) pointed out that correlation and path coefficient analyses help breeders to single out characters that advance genetic gains in seed yield. This view was also supported by Cyprien and Kumar (2011) who reiterated that two analyses assist breeders to define the appropriate selection criteria for seed yield during breeding. Thus, the adoption and application of these tools would help to exploit selection strategies that concentrate and build desired alleles in the base population.

Correlation analysis reveals the strength of the relationship between the given traits. Correlation analysis informs the breeder of the magnitude and direction of either the positive or negative value. Given r values that are closer to +1 or -1, would imply a stronger relationship between the measured variables irrespective of the direction. Malik et al. (2007) assessed the correlation of soybean yield and its components using 27 genotypes and found positive correlation coefficient for soybean yield with leaf area (r= 0.38), first pod height (r = 0.30), days to flowering (r = 0.44), days to maturity (r = 0.42), plant height (r = 0.38) and number of branches per plant (r = 0.44). In a similar study involving 40 soybean genotypes, Showkat and Tyagi (2010) observed positive and significant association for seed yield with total dry matter (r = 0.93), days to maturity (r = 0.45), number of branches per plant (r =

0.30), harvest index (r = 0.31) pod filling period (r = 0.36), pods per plant (r = 0.35), 100 seed weight (r = 0.32) and clusters per plant (r = 0.31). Sarutayophat (2012) reported positive and significant correlation between plant height and number of pods per plant (r = 0.82), plant height and pod yield (r = 0.52), number of pods per plant and pod yield (r = 0.82). These results demonstrated that selection for seed yield could be based on these characters.

Although correlation is a useful tool, its weakness is that it only shows whether simple linear relationship exists between measured traits or not; without revealing the cause and effect relationship among the traits. Makanda (2009) argued that correlation analysis does not show a good representation of the association if used in isolation. Sarutayophat (2012) also contended that the qualification of relationships among the measured traits would be incomplete if the characters responsible for causation are unknown. Furthermore, a correlative finding does not reveal the variable that has more influence over the other. In view of these pitfalls, breeders use correlation in conjunction with path coefficient analysis in order to derive maximum benefit. The beauty about path coefficient analysis is that it presents both the direct and indirect effects of the causal component on the effect components. The direct and indirect effects help to qualify the relative significance of the given characters while correlation defines the degree of relationship (Babe *et al.*, 2012).

Correlation and path coefficient have been widely applied to determine relationships among yield and its components in rice (Cyprien and Kumar, 2011; Babu *et al.*, 2012; Haider *et al.*, 2012), edible vegetable soybean (Sarutayophat, 2012), wheat (Kashif and Khaliq, 2004; Joshi, 2005; Khan and Dar, 2010), maize (Rafiq *et al.*, 2010; Sreckov *et al.*, 2010; Alaei, 2012; Zarei *et al.*, 2012), sorghum (Makanda, 2009; El-Din *et al.*, 2012; El-Naim *et al.*, 2012; Yazdani, 2012) and soybean (Iqbal *et al.*, 2003; Arshad *et al.*, 2006; Malik *et al.*, 2007; El-Badawy and Mehasen, 2012). Malik *et al.* (2007) studied the interrelationships between yield and its components in soybean and observed that number of pods per plant had the optimum positive direct effect on yield per plant followed by 100 seed weight and number of seeds per plant. These results were consistent with the findings of Arshad et al. (2006) who reported that days to maturity, number of branches, pod length, number of pods per plant and 100 seed weight had positive direct effects on seed yield. They also noted that pod length had high indirect effect on seed yield and high positive correlation was found between seed and unfilled pods, seed weight and filled pods as well as seed weight and total number pods per plant.

Though extensive studies have been conducted regarding the association of soybean seed yield and its components, sadly, these characters are largely influenced by the environment

Further, the genetic material used were different, hence the inferences were based on totally different germplasm and environmental conditions. This necessitated the need to carry out such studies because the results obtained elsewhere may not apply locally in the above context. More importantly, such investigations have not been done locally and as such, there are no documented reports available on this particular subject. Moreover, seed yield being a complex character, which is dependent on several variables required the investigation of its relationship with secondary traits for further development of high yielding cultivars and breeding efficiency. In view of this, the study was carried out with the objective of determining the relationships between grain yield and its associated components and direct and indirect effects of these traits on soybean grain yield using correlation and path coefficient analyses.

5.3 Materials and methods

5.3.1 Germplasm

The study was conducted using forty two soybean cultivars that represented all the cultivars that were introduced, developed and released in Zimbabwe from the birth of the breeding programme (i.e. 1940 to 2013). Table 2.1 in Chapter 2 shows detailed information about these cultivars.

5.3.2 Environments

The genotypes (42) were evaluated at eight sites viz; Rattray Arnold Research Station (RARS) (1341 masl; $17^0$40' S; $31^0$13' E); Agricultural Research Trust (ART) farm (1527 masl; $17^0$43' S; $31^0$05' E); Gwebi Variety Testing Centre (GVTC) (1449 masl., $17^0$68' S; $30^0$86' E); Mpongwe Development Centre ($13^0$02' N, $31^0$22'); Lusaka West farm (1300 masl., $15^0$25'S ; $28^0$17 E'); Lilayi farm $14^0$68'; $29^0$80' E), Chitedze Research Station in Malawi (1097 masl., $13^0$59' S $33^0$38' E;) and Bvumbwe Research Station in Malawi (1228 masl., $15^0$55' S; $35^0$04' E). The experiments were conducted at each site for one cropping season (2011-12).

5.3.3 Experimental Design

The trials were laid out in a 6 x 7 rectangular lattice design with three replications. The trials were grown in six row plots that were 5 m long with row to row spacing of 45 cm and an in-row spacing of 6.3 cm, giving 79 plants per row. A perfect stand should therefore have 476 plants per plot, meaning that 350 000 viable seeds were planted per hectare. Nett plot size was four rows, 5.4 m with an inter-row spacing of 0.45 m (9.72 m^2). The seed was inoculated

with *Bradrhizobium japonicum* inoculant Grasslands Strain 1491. Planting of the trials was done by hand and the trials were planting dates are as shown in Table 2.2 of Chapter 2.

5.3.4 Management

All the recommended management practices were applied at all research stations and farmer's management practices were applied at farmer's sites. A basal fertilizer (Cotton Fert) was at rate of 400 kg ha^{-1} supplying 28 kg ha^{-1} of nitrogen 68 kg ha^{-1}, phosphorus 40 kg ha^{-1} potassium. Integrated weed management was applied which included the hand weeding, hand pulling and application of herbicides. The amount of rainfall together with supplementary irrigation supplied varied from one place to another. The rainfall and irrigation totals were; 680 mm (RARS), 749 mm (ART farm), 423 mm (KRC), 712 mm (GVTC), 700 mm (Lusaka West farm), 800 mm (MDC), 643 mm (Chitedze Research Station) and 768 mm (Bvumbwe Research Station). The trials were sprayed with Shavit 25 EC (Tridimefon) at a rate of 500 ml ha^{-1} in order to control soybean rust. Three applications were done with the first application done at flowering and two subsequent sprays after an interval of 21 days. Shavit was preferred because it controls soybean rust only giving us the opportunity to assess the response of the genotypes to other diseases. At some sites the crop was guarded against animals and humans. All the trials were hand harvested.

5.3.5 Agronomic Data

The following data was recorded on each plot at each site;

- Pod clearance score (PDHT): The clearance between the soil and the bottom of the lowest pods in centimetres.
- Plant height (PLHT): Mean height of 3-5 modal plants to the top of the main stem of the upright plants, in centimetres.
- Percentage lodging at maturity (% LODG): Visual estimate to the nearest 10% of plants leaning more than 45^0
- Green stem scores at harvesting (GS): 0 = all stems dry, 1 = up to 50 % of the stems green, 2 = most stems green with leaves on the plants.
- Seed appearance scores (SAP): From 1 = very good to 9 = very poor quality with much discolouration, moulding and cracking.
- Number of purple stained seed in 100 seeds sample expressed as a percentage (%PS)
- Days from planting to 50% flowering (50% DFL): When 50% of the plants have at least one open flower.

- Days from planting to 95% pod maturity (DMAT): When 95% of the pods have dried.
- Days from 95% pod maturity to first pod shattering (DSH): When 3-5 plants on a plot had some pods shattering.
- Number of branches: Mean number of branches of 5 modal plants per plot
- Number of nodes: Mean number of nodes of 5 modal plants per plot
- Number of pods: Mean number of pods of 5 modal plants per plot
- Pod weight: Mean weight of pods of 5 modal plants per plot in grammes
- Number of seed per pod: Mean number of seeds of 5 modal plants per plot
- 100 Seed weight: Mean seed weight of 100 seed sample in grammes
- Shelling percentage: When seed weight was expressed as percentage of the total weight of the unshelled pods.
- Measured protein content in the seed on a dry matter basis
- Measured oil content in the seed on a dry matter basis.
- Grain yield (GYLD): Grams of air dry seed per nett plot

There were statistically derived variables and these were;

1) Grain yield (GYLD) was adjusted to kg ha^{-1} at 11% moisture using the following fomular;

   Grain Yield (kg ha-1) = [Grain Weight (Plot yield in kg ha-1)/(100 - %MC) *10/Plot Area*111/100]
2) Crude Protein Content (CRPR) was adjusted to 11% moisture content using the following formular;

   CRPR = Measured Protein/(100-%MC)*100
3) Crude Oil Content (CROIL) was adjusted to 11% moisture content using the following formular,

   CROIL = Measured Crude Oil Content/ (100-%MC)*100
4) 100 Seed mass (SDMA) was adjusted to 11% moisture content using the following formular;

   SDMA = Measured 100 seed weight/(100-%MC)*100

   Where; %MC = Grain Moisture in percentage.

N.B. The disease rating scale of 1-9 was adapted from international rating scale used for patent and cultivar registrations (http://www.google.com/patentsUS8378178 accessed on 10 October 2010).

5.3.6 Statistical Analysis

In the present study, the phenotypic correlations (rp) were assumed to be the same as the genetic correlations (rg). This is because the sample size was large (42 genotypes) evaluated across eight environments with a total of 24 replications. Effective sample sizes above 40 coupled with many test environments result in good correspondence between rp and rg because environmental effects are removed by multi-location effects indicating high levels of precision (Cheverud, 1988; Watt and Levin, 1988). Combined analyses were done across eight sites. All quantitative data was subjected to analysis of variance using GenStat 14th Edition (Payne *et al.*, 2011). Histograms exhibiting distribution of genotypes were plotted for grain yield and its components.

Correlation coefficients (r) between all the traits were computed in GenStat computer package (Payne *et al.*, 2007). Path-coefficients (P) were calculated by regression method based on the work of (Wright, 1921; 1960), (Dewey and Lu, 1959), and (Cramer *et al*., 1999). In this procedure, all the independent variables (1 to n) are regressed against the dependent variable (X12). The regression coefficient (b) of each of the independent traits (1 to n) is its direct effects to the dependent variable X (Cramer *et al.*, 1999). The indirect effects are then computed by multiplying the correlation coefficient between each of the independent variables (1 to n) and the variable in its path (1 to n) by the direct effect (b) of the independent variable in the path to the dependent variable (Cramer et al., 1999). The equations for the multiplications are given below:

$r_{112} = P_{112} + r_{12}P_{212} + r_{13}P_{312} + r_{14}P_{412} + r_{15}P_{512} + r_{16}P_{612} + r_{17}P_{712} + r_{18}P_{812} + r_{19}P_{912} + r_{110}P_{1012} + r_{111}P_{1112}$

$r_{212} = P_{212} + r_{12}P_{212} + r_{23}P_{312} + r_{24}P_{412} + r_{25}P_{512} + r_{26}P_{612} + r_{27}P_{712} + r_{28}P_{812} + r_{29}P_{912} + r_{210}P_{1012} + r_{211}P_{1112}$

$r_{312} = P_{312} + r_{13}P_{212} + r_{23}P_{312} + r_{34}P_{412} + r_{35}P_{512} + r_{36}P_{612} + r_{37}P_{712} + r_{38}P_{812} + r_{39}P_{912} + r_{310}P_{1012} + r_{311}P_{1112}$

$r_{412} = P_{412} + r_{14}P_{212} + r_{24}P_{312} + r_{34}P_{412} + r_{45}P_{512} + r_{46}P_{612} + r_{47}P_{712} + r_{48}P_{812} + r_{49}P_{912} + r_{410}\ P_{1012} + r_{411}P_{1112}$

$r_{512} = P_{512} + r_{15}P_{212} + r_{25}P_{312} + r_{35}P_{412} + r_{45}P_{512} + r_{56}P_{612} + r_{57}P_{712} + r_{58}P_{812} + r_{59}P_{912} + r_{510}P_{1012} + r_{511}P_{1112}$

$r_{612} = P_{612} + r_{16}P_{212} + r_{26}P_{312} + r_{36}P_{412} + r_{46}P_{512} + r_{56}P_{612} + r_{67}P_{712} + r_{68}P_{812} + r_{69}P_{912} + r_{610}P_{1012} + r_{611}P_{1112}$

$r_{712} = P_{712} + r_{17}P_{112} + r_{27}P_{212} + r_{37}P_{312} + r_{47}P_{412} + r_{57}P_{512} + r_{67}P_{612} + r_{78}P_{712} + r_{79}P_{712} + r_{710}P_{712} + r_{711}P_{712}$

$r_{812} = P_{812} + r_{18}P_{112} + r_{28}P_{212} + r_{38}P_{312} + r_{48}P_{412} + r_{58}P_{512} + r_{68}P_{612} + r_{78}P_{712} + r_{89}P_{812} + r_{910}P_{912} + r_{811}P_{1012}$

$r_{912} = P_{912} + r_{19}P_{112} + r_{29}P_{212} + r_{39}P_{312} + r_{49}P_{412} + r_{59}P_{512} + r_{69}P_{612} + r_{79}P_{712} + r_{89}P_{812} + r_{910}P_{912} + r_{1011}P_{1012}$

$r_{1012} = P_{1012} + r_{110}P_{112} + r_{210}P_{212} + r_{310}P_{312} + r_{410}P_{412} + r_{510}P_{512} + r_{610}P_{612} + r_{710}P_{712} + r_{810}P_{812} + r_{910}P_{912} + r_{1011}P_{1012}$

$r_{1112} = P_{1112} + r_{111}P_{112} + r_{211}P_{212} + r_{311}P_{312} + r_{411}P_{412} + r_{511}P_{512} + r_{611}P_{612} + r_{711}P_{712} + r_{811}P_{812} + r_{911}P_{912} + r_{1011}P_{1012}$

Where: 1 = Plant height; 2 = Number of pods per plant; 3 = Number of seeds per plant; 4 = Number of branches per plant; 5 = Number of nodes per plant; 6 = Shelling percentage; 7 = Days from planting to 50% flowering; 8 = Days from planting to 95% pod maturity; 9 = 100 seed weight; 10 = Crude protein content; 11 = Crude oil content; 12 = Grain yield, the dependent variable.

Taking equation (1) above for example,

r_{12} = the correlation coefficient between 1 (plant height) and 12 the dependent variable (grain yield used as the response);

P_{112} = the direct path coefficient of plant height on the dependent trait 12 (grain yield);

$r_{12}P_{212}$= the indirect path coefficient of plant height on trait 12 through trait used as 2 (number of pods per plant);

$r_{13}P_{312}$ = the indirect path coefficient of plant height on trait12 through trait 3 (number of seeds per plant);

$r_{14}P_{412}$= the indirect path coefficient of plant height on 12 through trait 4 (number of branches per plant);

$r_{15}P_{512}$ = the indirect path coefficient of plant height on trait 12 through trait 5 (number of nodes per plant);

$r_{16}P_{612}$ = the indirect path coefficient of plant height on 12 through trait 6 (shelling percentage);

$r_{17}P_{712}$ = the indirect path coefficient of plant height on 9 through trait 7 (days to 50% flowering);

$r_{18}P_{812}$ = the indirect path coefficient of plant height on trait 12 through trait 8 (days to maturity)

$r_{19}P_{912}$ = the indirect path coefficient of plant height on trait 12 through trait 9 (100 seed weight)

$r_{110}P_{1012}$ = the indirect path coefficient of plant height on trait 12 through trait 10 (protein percentage) and

$r_{111}P_{1112}$ = the indirect path coefficient of plant height on trait 12 through trait 11 (oil percentage)

As has been indicated above, grain yield was selected as the dependent variable and other traits as the independent variables. Path coefficients were estimated using computer software SAS version 9.2 (SAS Institute Inc, 2010). Figure 5.1 below presents the relationships between dependent variable, grain yield (X12) and the 11 independent variables, plant height (X1), number of pods per plant (X2), number of seeds per plant (X3), number of branches per plant (X4), number of nodes per plant (X5), shelling percentage (X7), days from planting to 50% flowering, (X7), days from planting to 95% pod maturity (X8), 100 seed weight (X9), crude protein content (X10) and crude oil content (X11). The path coefficients are represented by P112, P212, P312, P412, P512, P612, P712, P812, P812, P912, P1012 and P112 corresponding to direct effects on yield from plant height, number of pods per plant, number of seeds per plant, number of branches per plant, number of nodes per plant, shelling percentage, days from planting to 50% flowering, days from planting to 95% pod maturity, 100 seed weight, crude protein content and crude oil content respectively. The direct effect of each yield component on yield is the path coefficient from this component to yield. The indirect effect of one component through a second component is the product of the path coefficient from the second component and the correlation from the second component and the correlation between the two components (Dewey and Lu, 1959).

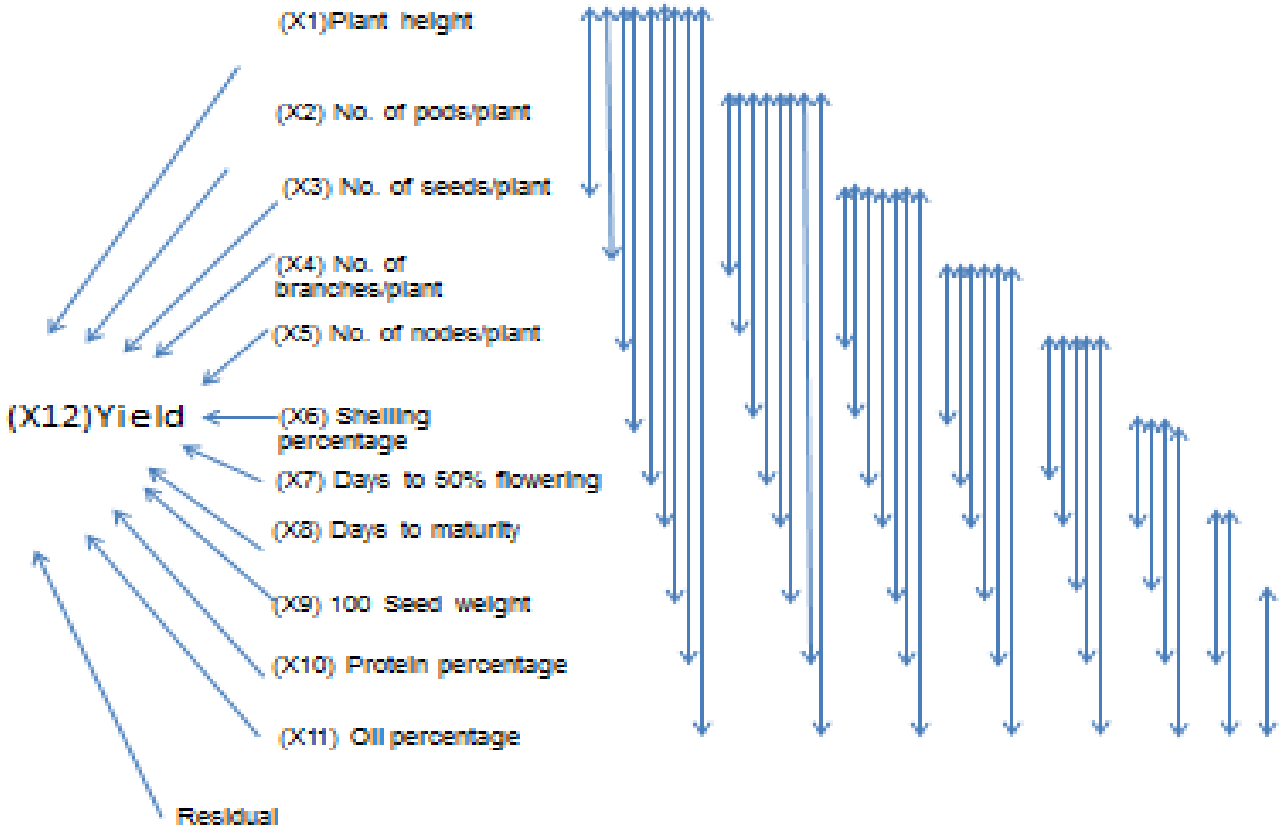

Figure 5.1: Path diagram showing the cause and effect relationships in the path analysis. Single arrows indicate the direct of one variable upon another, pij. Double arrows indicate the mutual association between variables, rij.

5.4 Results

5.4.1 Analysis of variance for plant traits of soybean lines

Means pertaining to the traits are given in Table 5.2. The coefficient of determination (R^2) value was highly significant (0.86). Highly significant variations were shown among the genotypes for grain yield, number of nodes per plant, days from planting to 95% pod maturity, number of branches per plant, shelling percentage and number of pods per plant. Plant height, 100 seed weight, percentage crude protein and oil and number of seeds per pod showed non-significant differences. The average grain yield per hectare ranged from 2441 to 5771 kg ha-1. Phenologically, G41, G4, G31 and G39 flowered earlier (49 days) than all other cultivars and crop duration varied from 108 to 129 days. Genotype, G41 was found to be the earliest maturing; whereas, G2 and G14 were late. The number of pods per plant ranged from 36 to 67. The highest pod producing genotypes were G2, G7, G12, G19 and G8 with 67, 66, 62, 58 and 57 pods per plant. Low pod counts were observed on genotypes; G32, G42, G31, G18 and G10 which recorded 36, 36, 36, 37 and 38 pods respectively. There was less variation on the number of seeds per pod as it was either 2 or 3 seeds per pod. The results showed that the number of branches per plant ranged from 3 to 8. Highest branches (eight) were produced by G2 while G41 had the minimum (three). The minimum and maximum number of nodes per plant were 17 and 28, registered by genotypes; G41 and G12 and G14 respectively. Shelling percentage was relatively high varying from 73 to 77%. The highest 100 seed weight was observed from G1 and G24 (29g) whereas G41 recorded the minimum (18g). Regarding quality attributes, the protein content ranged from 47 to 51% and the oil content fell between 16.6 and 18.7%. Based on mean performances for grain yield and associated traits, genotypes; G14, G25, G15, G40 and G21 were found to be the top 5 performers. Genotypes G38, G9, G37, G41 and G41 were the bottom 5 performing genotypes.

Table 5.1: Mean performances for various plant traits of soybean lines planted at eight environments during 2011/12 cropping season.

Genotype	Plant Height (cm)	Pods/ Plant	Seeds/ Pod	Branches/ Plant	Nodes/ Plant	Shelling Percent	Days to Flowering	Days to pod Maturity	100 Seed Weight (g)	% Protein	% Oil	Yield (kg ha-1)
G14	118	42	3	7	28	74	61	129	22	48.9	18.4	5771
G25	109	51	3	5	25	76	57	119	24	49.6	17.3	5368
G15	105	45	2	5	22	74	55	125	26	49.2	17.9	5266
G40	111	41	2	6	24	77	52	117	24	50.4	17.4	5162
G21	108	42	3	5	23	73	55	117	26	49	18.2	5126
G6	97	48	3	5	27	76	51	120	28	49.1	18.3	5125
G2	120	67	3	8	24	70	69	129	23	51.1	17.6	5092
G16	78	39	2	6	26	74	53	123	26	48	18.2	5084
G13	118	41	3	6	24	75	57	125	23	48.7	18.1	5079
G23	88	42	2	7	24	77	53	118	25	48.5	17.7	5072
G10	112	38	2	6	26	73	52	117	26	49.3	18.4	5036
G12	86	62	2	5	28	75	51	111	24	49.4	17.6	5026
G30	116	40	3	5	22	71	57	126	23	49.8	17.8	4977
G11	100	45	3	6	24	75	50	117	23	49.1	17.3	4925
G34	92	40	2	5	21	74	53	117	24	49.7	17.6	4889
G17	96	50	2	5	23	73	57	123	22	49.1	17.8	4859
G42	110	36	2	7	23	76	57	126	24	49.8	17.7	4823
G5	98	41	3	6	23	76	52	116	26	48.7	17.8	4815
G8	120	57	2	5	26	75	64	121	26	49.6	17.8	4801
G27	86	43	2	5	22	76	53	118	26	49.4	17.9	4799
G19	96	58	3	5	23	75	56	117	27	49.2	17.9	4796
G1	100	43	2	5	20	74	56	117	29	49.5	17.9	4776
G24	88	43	2	5	24	77	54	114	29	49.5	17.5	4696
G35	105	42	2	5	24	74	50	120	24	49	17.9	4682
G28	96	47	2	5	22	73	55	117	25	49.6	17.7	4681
G33	95	50	2	5	24	73	50	114	21	48.6	18.1	4610
G36	100	51	2	4	25	73	50	117	25	47.9	18.7	4550
G3	76	43	2	5	22	75	50	108	27	49.9	17.3	4531
G7	100	66	2	4	24	76	57	126	21	48.9	17.9	4519
G32	97	36	2	5	21	75	51	115	24	48.8	17.9	4491
G20	78	49	3	4	26	76	54	117	24	49.5	17	4427
G18	105	37	3	4	18	75	55	119	20	49.9	17.5	4366
G4	93	40	3	5	24	71	49	114	25	49.6	17.8	4362
G26	74	40	2	5	22	76	51	114	24	50.2	16.6	4322
G31	74	36	3	4	19	75	49	113	25	50.1	17	4318
G22	78	44	3	5	19	72	51	117	22	48.5	17.7	4300
G29	79	45	3	4	18	75	53	120	22	49.9	17.2	4296
G38	76	46	2	5	24	71	57	126	23	50.7	17.4	4243
G9	85	50	3	5	19	74	57	122	24	51.2	16.7	4154
G37	85	44	2	5	18	71	50	114	22	49.1	17.7	3763
G39	80	46	2	5	22	75	49	111	23	48.9	17.9	3398
G41	79	40	2	3	17	73	49	108	18	49.3	17.9	2441
Mean	94	45	2	5	23	75	54	119	24	49.4	17.7	4662
LSD (5%)	9.13	14.96	0.62	1.32	6.45	2.82	1.68	3.63	5.85	1.39	0.77	607.3
CV (%)	6	20	16	16	17	2	2	2	15	2	3	8
F. Sign.	NS	**	NS	***	*	***	NS	***	NS	NS	NS	***

LSD = Least significant difference at 5% alpha level; CV (%) = Coefficient of variation in percentage; F. sign. = significance of the variance ratio; NS = not significant; kg ha^{-1} = kilograms per hectare

5.4.2 Variation of genotypes for yield and secondary traits

The distribution of the test genotypes across eight environments is shown in Figure 5.2. The data showed positive skewness for grain yield, number of pods per plant, number of branches per plant, shelling percent, number of days from planting to 95% pod maturity and 100 seed weight. Number of nodes per plant, plant height, percentage crude oil and protein in the seed displayed normal distribution while number of seeds per plant and days from planting to 50% flowering exhibited negative skewness.

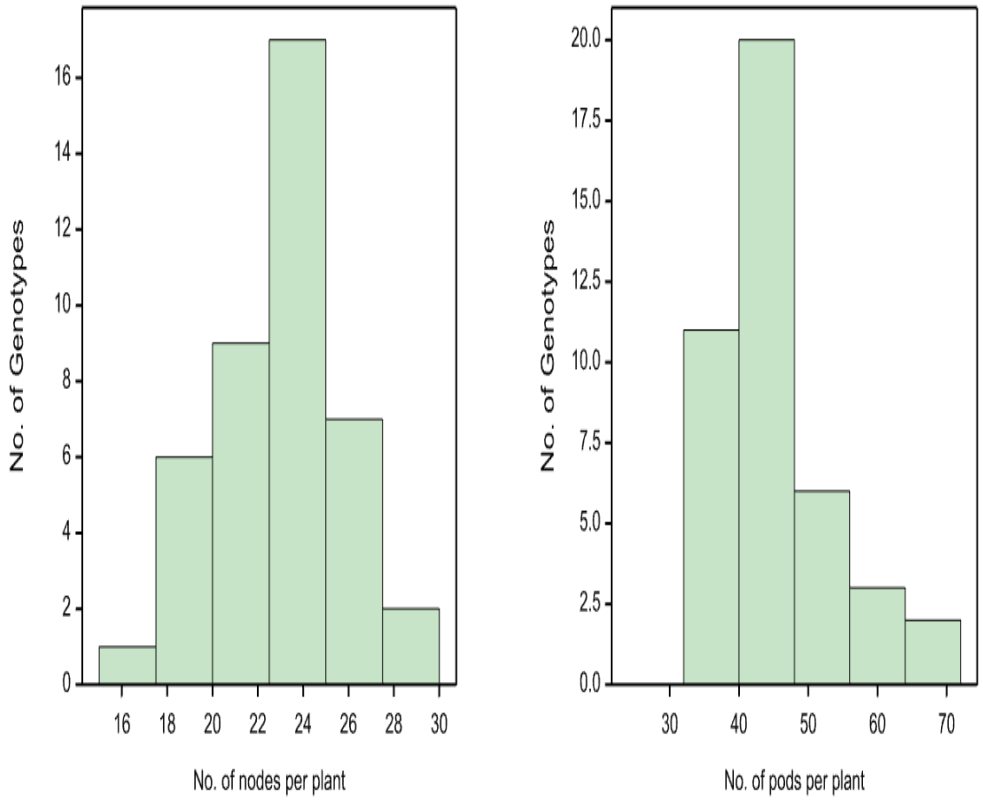

Figure 5.2a: The frequency distribution of number of nodes per plant (left) and pods per plant (right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2011/2012 cropping season

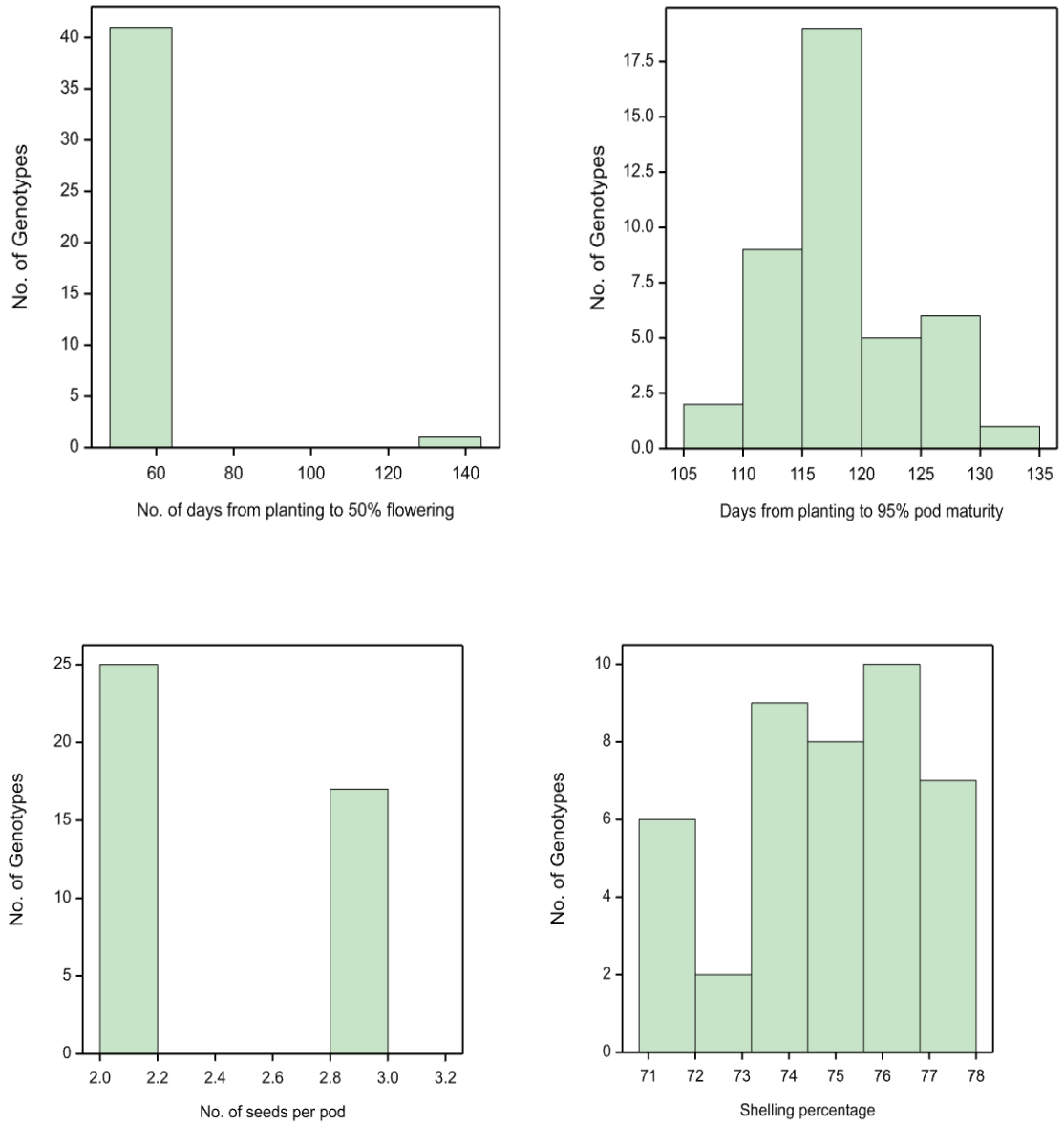

Figure 5.2b: The frequency distribution of number of days from planting to 50% flowering (top left), days from planting to 95% pod maturity (top right), number of seeds per pod (bottom left) and shelling percentage (bottom right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2011/2012 cropping season

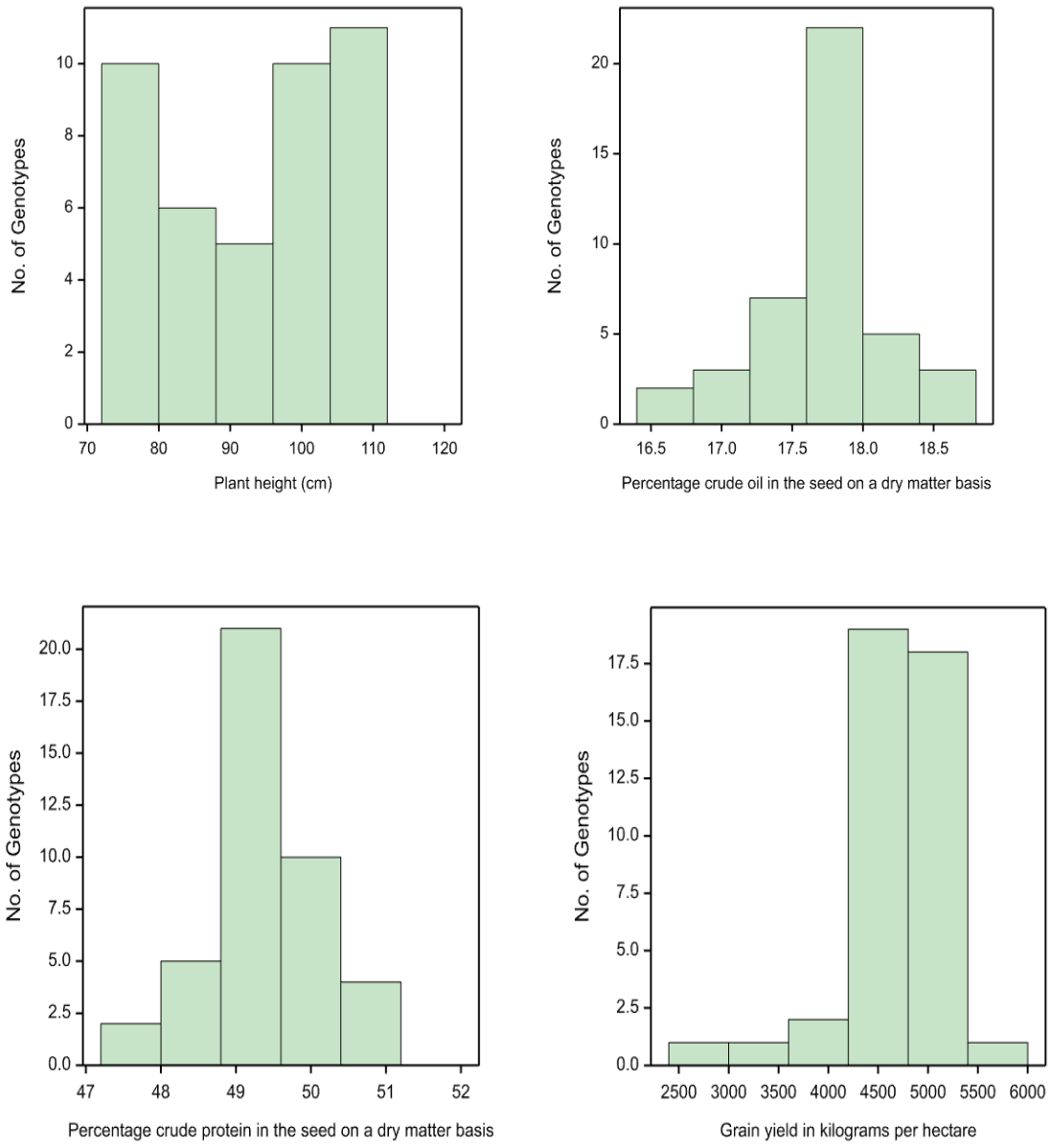

Figure 5.2c: The frequency distribution of plant height (top left), percentage crude oil in the seed on a dry matter basis (top right), percentage crude protein in the seed on a dry matter basis (bottom left) and grain yield (bottom right) among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2011/2012 cropping season.

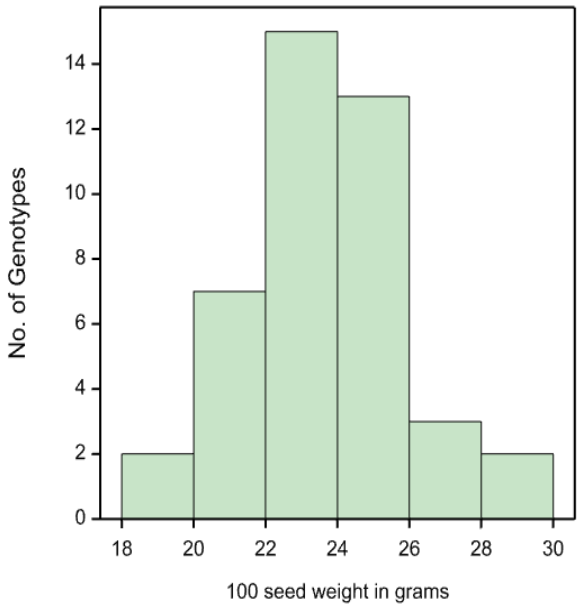

Figure 5.2d: The frequency distribution of number of 100 seed weight among 42 soybean genotypes evaluated at 13 testing locations in three countries (Malawi, Zambia and Zimbabwe) during the 2011/2012 cropping season.

5.4.3 Phenotypic correlations between grain yield and secondary traits

The results for correlation analysis are given in Table 5.2. The study showed that grain yield was positively and significantly correlated with number of branches per plant (0.77***), number of nodes per plant (0.75***), shelling percentage (0.90***), red leaf blotch (0.31***) and number of days from 95% pod maturity to first pod shattering (0.54***). However, negative and significant correlation coefficients were found between grain yield with; plant height (-0.42***), percentage lodged plants at maturity (-0.46***), green stem scores (-0.31***), number of pods per plant (-0.31***), days from planting to 50% flowering (-0.60***), days from planting to 95% pod maturity (-0.43***), and percentage crude protein in the seed on a dry matter basis (-0.20*). Inter-character correlations were also found among several traits for instance, number of branches per plant gave significant and positive correlation (0.60***) with number nodes per plant and shelling percentage (0.73***). Maturity exhibited significant and positive correlation with plant height (0.33) and days to 50% flowering (0.82). However, number of nodes displayed a significant and negative correlation with lodging (-0.38***) and days to 50% flowering (-0.47%).

Table 5.2: Phenotypic correlation coefficients between grain yield and its secondary traits

TRAIT	PDHT	PLHT	%LODG	GSS	#PODS	#SEED	#BRAN	#NODES	%SH	SAP	%PS	BBS	RLB	DFL	DMAT	SDM	DSH	CP	COIL	GYLD
PDHT	-																			
PLHT	0.08	-																		
%LODG	-0.02	0.29***	-																	
GSS	0.01	-0.07	-0.17*	-																
#PODS	0.15*	0.01	-0.39***	-0.05	-															
#SEEDS	0.08	-0.01	0.11	0.09	0.1	-														
#BRAN	0.01	-0.29***	-0.23**	0.24**	-0.2	-0.07														
#NODES	0.04	-0.29***	-0.38***	0.41***	-0.19	-0.08	0.60***	-												
%SH	0.049	-0.41***	-0.50***	0.3***	-0.3	-0.05	0.73***	0.69***	-											
SAP	-0.15	0.18*	0.14	0.02	0.04	-0.08	0.1	0	0.02	-										
%PS	-0.18*	0.09	0.12	0.11	0.02	-0.08	-0.03	0.13	0.1	0.40***	-									
BBS	-0.14	-0.07	-0.09	0.04	-0.16	0.06	-0.03	0.09	0.1	0.08	0.15	-								
RLB	-0.16*	-0.24**	-0.15*	-0.01	-0.25	0.08	0.18*	0.09	0.33***	0.08	0.19*	0.37***	-							
DFL	0.05	0.37***	0.42***	0.22*	0.34	0.03	-0.34***	-0.47***	-0.61***	0.16*	-0.21*	-0.31***	0.47***	-						
DMAT	0.14	0.33***	0.20*	0.15	0.18	0.03	-0.23**	-0.29***	-0.42	0.23**	-0.24**	-0.22*	0.45***	0.82***	-					
SDM	-0.22*	-0.13	-0.07	-0.15*	-0.07	-0.01	0.05	-0.12	0.02	0	-0.1	-0.05	0.13	-0.06	-0.08	-				
DSH	-0.12	-0.25**	-0.24**	0.23**	-0.31	-0.05	0.34***	0.38***	0.50***	0.01	0.16*	0.41***	0.31***	-0.42***	-0.25**	-0.05	-			
CP	0	0.08	0.06	0.06	0.04	0.02	-0.24**	-0.021	-0.18*	0.1	-0.03	0.03	-0.2	0.11	0.05	-0.06	-0.14	-		
COIL	0.041	0.05	-0.08	0.12	0.15	0.05	-0.05	0.06	0.039	-0.01	0.18*	-0.05	0.01	-0.12	-0.01	-0.07	0.1	-0.03	-	
GYLD	0.041	-0.42***	-0.46***	0.31***	-0.31	-0.07	0.77***	0.75***	0.90***	0.05	0.11	0.13	0.31***	-0.60***	-0.43***	0.05	0.54***	-0.20*	0.032	-

*, **, *** Significant at P<0.05, 0.01 and 0.001, respectively. PDHT = Pod height; PHLT = Plant height; %LODG = Percentage lodged plants; GSS = Green stem scores; #PODS = number of pods per plant; #SEEDS = Number of seeds per plant, #BRAN = Number of branches per plant; #NODES = Number of Nodes per plant; %SH = Shelling percent; SAP = Seed appearance scores; %PS = Percentage purple stained seed; BBS = Bacterial Blight scores; RLB = Red leaf blotch scores; DFL = Days to 50% flowering; DMAT = Days to 95% pod maturity; SDMA = 100 Seed mass; DSH = Days from 95% pod maturity to first pod shattering; CP = Perecentage protein; COIL = Percentage crude oil; GYLD = Grain yield

5.4.4 Path coefficient analysis

The estimations of direct and indirect effects of the yield components on grain yield are shown in Table 5.3. Number of nodes per plant had a positive and significant (P<0.01) correlation coefficient. In addition it had the highest direct effect (0.48**) towards soybean grain yield, followed by plant height (0.27*), 100 seed weight (0.20*), number of seeds per plant (0.17**) and days to 95% pod maturity (0.15). Number of pods per plant (-0.15), number of branches per plant (-0.01), days to 50% flowering (-0.06), protein percentage (-0.01) exhibited negative and non-significant direct path towards grain yield. On the other hand, number of nodes had the highest indirect effect through number of branches per plant, followed by plant height through days to 95% pod maturity.

Table 5.3: Direct and indirect effects of different secondary traits on grain yield

TRAIT component	PLHT	#PODS	#SEEDS	#BRANC	#NODES	%SH	DFL	DMAT	100SDWT	CRPRO	CROIL	TOTAL r to GYLD
PLHT	**0.27***	-0.03	0.02	0	0.19	0	0	0.08	0.01	0	0	0.59
#PODS	0.06	-0.15	0	0	0.18	-0.01	0	0.03	-0.01	0	0	0.13
#SEEDS	0.03	0	**0.17****	0	0	-0.01	0	0.04	-0.02	0	0	0.18
#BRANC	0.1	0	0.01	-0.01	0.22	0	0	0.08	0.03	0	0	0.59
#NODES	0.11	-0.05	-0.01	0	**0.48****	0.01	0	0.05	0.06	0	0	0.66
%SH	0.01	0.02	-0.02	0	0.06	0.1	0	-0.03	0.06	0	0	0.18
DFL	0.09	-0.07	0.04	0	0.07	-0.03	-0.06	0.07	-0.02	0.01	0	0.22
DMAT	0.14	-0.03	0.04	0	0.17	-0.02	0	0.15	-0.03	0	0	0.56
100SDWT	0.01	0.01	-0.02	0	0.17	0.03	0	-0.02	**0.20***	0	0	0.41
CRPRO	0.02	-0.01	0.03	0	-0.1	-0.02	0	0.01	0	-0.01	0	-0.09
CROIL	0.04	0	-0.03	0	0.19	-0.02	0	0.04	0.01	0.02	0-01	0.27

Bold and *, ** = Significant at P<0.05 and P<0.01 respectively, PHLT = Plant height; #PODS = number of pods per plant; #SEEDS = Number of seeds per plant, #BRAN = Number of branches per plant; #NODES = Number of Nodes per plant; %SH = Shelling percent; DFL = Days to 50% flowering; DMAT = Days to 95% pod maturity; SDMA = 100 Seed mass; CRPRO = Perecentage protein; CROIL = Percentage crude oil; GYLD = Grain yield.

5.5 Discussion

5.5.1 Variation of genotypes for grain yield and secondary traits

The results exhibited highly significant differences among the evaluated genotypes in the mean values of grain yield and some of the secondary traits suggesting high level of genetic diversity among the test genotypes. Genotype, G14 had the highest mean values for most of the analysed yield components (plant height, number of seeds per pod, number of branches per, number of nodes per plant, days to 50% flowering and days to 95% pod maturity) and it average yield was 5771 kg ha^{-1}. Three distinct groups of genotypes were identified with respect to grain yield. The first group comprised of genotypes with high seed yields i.e. the top 12 with mean yields above 5 000 kg ha^{-1} (from 5026 to 5771 kg ha^{-1}). The results showed these genotypes had relatively high mean values for most of the yield components. The second group had moderate mean values for the analysed secondary traits with average yields ranging from 4519 to 4977 kg ha^{-1}, while the bottom 14 genotypes were constituted by genotypes possessing low mean seed yields varying from 2441 to 4491kg ha^{-1}. The pattern shown by the results indicated that grain yield and associated traits were improved progressively. This meant that the modern genotypes have better performance compared to the earlier generations. Hence, the genotypes that were released between 1966 and 1980 exhibited the lowest mean performance, whereas genotypes that were registered between 1980 and 1995 showed moderate mean values for grain yield and associated traits. While the genotypes that were registered between 1995 to 2013 revealed high mean values. The relative variability for the analysed traits suggested that the yield components were more stable and reliable as selection criteria for higher yield in the soybean breeding programme.

The grain yield and the secondary traits showed continuous variation in respect of distribution. Clearly, this demonstrated that all the traits are polygenic indicating that they are quantitatively inherited. This observation was in accordance with the findings of (Khan *et al.*, 2011). Thus, each trait contributes relatively small effects to grain yield. These results were also consistent with the findings of earlier studies in soybean (Aslam *et al.*, 1992; Sudaric and Vrataric, 2002; Malik *et al.*, 2007; Karnwal and Singh, 2009). The observed positive skewness on grain yield probably demonstrated that the breeders exploited these secondary traits in their effort to improve grain yield. Number of branches, pods per plant, nodes per plant, days from planting to 95% pod maturity and shelling percentage showed high genotypic variability. However, the

mean number of branches and number of pods per plant were not in harmony with the previous studies in soybean (Malik *et al.*, 2006; Khan *et al.*, 2011) where ranges of 4-20 branches per plant and 47 to 167 pods per plant were reported compared to 4 to 8 branches and 36 to 67 pods per plant obtained in the present study. These traits are important yield components because the more the number of branches, the higher the number of nodes consequently, the greater the number of pods to be formed. Probably the low number of branches and pods could be attributed to the genetic backgrounds of the genotypes. The phenological data was at variance with the results of Malik *et al.* (2007) who obtained a range of 43 to 48 days for flowering and 85 to 109 days for maturity. However, the quality traits did not show any significant variability. There was no variability on plant height, number of seed per pod and 100 seed weight and this was not in agreement with the results of (Karnwal and Singh, 2009; Machikowa and Laosuwan, 2011) and this variation could be attributed to differences in the genotypes used.

5.5.2 Phenotypic correlations between grain yield and secondary traits

Among the 20 characters that were evaluated, grain yield showed positive and highly significant correlation with number of branches per plant, number of nodes per plant, shelling percentage, red leaf blotch and number of days from 95% pod maturity to first pod shattering. This meant that higher mean values for these traits could increase soybean grain yield. The results were in conformity with previous studies (Board *et al.*, 1997; Machikowa and Laosuwan, 2011; Ghodrati *et al.*, 2013). Therefore, these characters could be considered as selection criteria in the soybean improvement programme. Low positive and non-significant correlations were obtained between pod height, seed appearance scores, percentage purple stained seed, bacterial blight scores, 100 seed weight and oil percentage with grain yield. These results were in agreement with the report by Malik *et al.* (2007) who found weak positive correlation between grain yield with pod height and 100 seed weight. However, earlier and numerous studies indicated that there were positive correlations with high magnitude between number of pods per plant and number seeds per plant (Jadhav *et al.*, 1995; Shinde *et al.*, 1996; Ramgiry and Raha, 1997; Malik *et al.*, 2006). On the other hand, negative and significant correlations were observed between grain yield and (i) plant height (ii) percentage lodged plants at maturity (iii) number of pods per plant (iv) days to 50% flowering (vi) days from planting to 95% pod maturity and (vii) protein percentage. These observations were at variance with the results of (Bizeti *et al.*, 2004; Malik *et al.*, 2007; Ramteke *et al.*, 2010; Ghodrati *et al.*, 2013). This meant that increases in the mean values of these characters would decrease grain yield. For instance, plant height was

negative and significantly correlated with grain yield possibly because the photosynthetic from the leaves were used to build the height and thick stems. However, the fact that both days to 50% flowering and maturity were negatively associated with grain yield was at variance with the findings of Sharma *et al*., 1983 who argued that the two traits contributed most to grain yield. Nevertheless, Arshad *et al.* (2006) also reported negative and significant associations.

5.5.3 Phenotypic correlations among secondary traits

In general, inter-relationships between characters are important because they show the traits that can be improved simultaneously or concurrently. Thus, traits showing positive and significant associations could be improved concurrently. In the present study, plant height showed significant positive correlation with days to 50% flowering and maturity indicating that increase in plant height delayed flowering and maturity. Green stem scores showed positive and significant association with number of nodes per plant, shelling percentage and grain yield suggesting that genotypes which give dry stem at maturity produce more nodes coupled with higher shelling percentage and consequently higher grain yield. The association between number of branches with; number of nodes, shelling percent, number of days from maturity to first pod shattering and grain yield was positive and significant. More number of branches led to increase in number of nodes, shelling percentage, long pod shatter free period and grain yield. Malik *et al.* (2007) and Karnwal and Singh (2009) also observed that number of branches increased number of pods and grain yield.

In addition, the association of number of pods and days to 50% flowering was positive implying that late flowering increased the number of pods per plant and similar report was made by Bizeti *et al.*, (2004). There was a significant positive correlation between red leaf blotch scores with; days to flowering, maturity and shattering period. This revealed that increased tolerance or resistance to red leaf blotch resulted in delayed flowering, maturity and long pod shatter free period which essentially would increase grain yield. There was a positive and significant association between days to flowering and maturity which meant that an increase in the number of days to 50% flowering resulted in more number of days to maturity. Malik *et al.* (2007) and Ghodrati *et al.* (2013) observed similar results. Shelling percentage was positively and significantly correlated with red blotch scores, days to 50% flowering and days to first pod shattering. High shelling percentage was as a result of increased tolerance to red leaf blotch, delayed flowering and long pod shatter free period. Plant height was negatively and significantly correlated with number of branches per plant, number of nodes per plant, red leaf blotch scores,

days to 50% flowering, days to 95% pod maturity and days from 95% pod maturity to first pod shattering because the photo-assimilates were used to build thick stems and height. Ghodrati *et al.* (2013) also found a negative correlation between days to maturity and 50% flowering, whereas previous studies obtained positive and significant correlations (Lal and Hague, 1971; Malik *et al.*, 2007; Ghodrati *et al.*, 2013). On the other hand, this implies that breeding for high yield can be achieved indirectly through breeding for reduced plant height. Moreover, negative and significant correlations were observed between number of branches and days to 50% flowering, maturity and protein percentage. Number of nodes per plant was also negatively and significantly correlated to days to flowering and maturity suggesting that breeders should focus on early maturing cultivars in order to increase grain yield.

Thus, phenotypic correlations among the secondary traits have shown that number of pods per plant, number nodes per plant and number of days from 95% pod maturity to first pod shattering contributed high value for selection for grain yield in soybean. Consequently, these traits should be paid attention to in the soybean breeding programme.

5.5.4 Path coefficient analysis

Number of nodes, plant height and 100 seed weight showed the largest influence directly towards grain yield. The highest positive and significant direct effect was displayed by number of nodes per plant (0.48**), followed by plant height (0.27*) and 100 seed weight (0.20*). The findings of the current investigation are in harmony with previous studies for number of nodes per plant (Ghodrati *et al.*, 2013), plant height (Aslam *et al.*, 1992; Malik *et al.*, 2007), 100 seeds weight (Lal and Hague, 1971; Arshad *et al.*, 2006; Karnwal and Singh, 2009; El-Badawy and Mehasen, 2012), maturity (Arshad *et al.*, 2006), number of seed per pod (Oz *et al.*, 2009). Clearly, the results imply that number of nodes could be considered as the major selection criteria for grain yield because it is least affected by indirect factors. It is also important to note that 100 seeds weight had no phenotypic correlation with grain yield (Table 5.3) because high positive direct effect of 100 seed weight on seed yield were nullified by its negative indirect effects via number of pods, number of seeds and both days to 50% flowering and 95% pod maturity. Direct effects of other characters such as number of seeds per pod were low, signifying low level of influence to grain yield. Contrary to this, it was observed that number of pods per plant, number of branches per plant, days to 50% flowering, and protein percentage had negative indirect effects indicating that breeders should select against these traits as they

largely reduce grain yield. These results confirm the findings of Malik et al. (2007) in respect of number of branches per plant, days to 50% flowering and protein content.

Number of nodes per plant and plant height had the highest indirect contribution to grain yield. In this case, number of nodes per plant had positive indirect effects (0.22) via number of branches per plant, while for plant height it was through days from planting to 95% pod maturity. In a similar regard, Lal and Hague (1971) also reported positive high indirect effects from plant height. It is interesting to note that the character number of nodes per plant, showed a highly positive and significant correlation with grain yield and, in this case, it showed positive indirect effect on grain yield. Therefore, the two secondary traits, number of nodes per plant and plant height, appeared to be of positive value as an aid in selection for higher grain yield in soybean improvement programs. Although the phenotypic correlation between plant height and number of nodes per plant revealed a negative and significant relationship, as well as low but positive and non-significant direct effects, the mean performance of the two top genotypes (G14 and G25) (Table 4.2) demonstrated that these two traits can be combined in one cultivar. Hence, success of any breeding programme entirely depends on identifying genotypes possessing desirable traits. It could be argued that a negative correlation was probably brought about by some genotypes that combined high performance for one trait and poor performance for the other trait. A classic example of such genotypes includes G41, G9, G37, G39 and G38 (Table 5.2). It is also important to note that these genotypes were released between 1966 and 1975 compared to the top five genotypes which were registered between 1998 and 2013. The modern cultivars showed high mean values for the two secondary traits demonstrating that selection for higher grain on the basis of these two traits is possible.

Breeding for higher number of nodes per plant, plant height, 100 seed weight, number of seeds per pod, shelling percentage and maturity directly improve grain yield as shown by the positive and direct effects as well as positive correlation coefficient for some of the traits. A reasonably tall plant may produce more branches, more nodes and more leaves translating to a higher photosynthetic factory which is further enhanced by breeding for late maturing cultivars and when aggregated together leads to high productivity. The negative direct effects of number of pods per plant, number of branches per plant days to 50% flowering and percentage protein suggested that improvement in these traits directly increase grain yield. For instance, the more the number of grain filled pods, the higher the grain yield.

5.6 Conclusion

The objective of the study was to determine the relationship between grain yield and its secondary traits and direct and indirect effects using correlation and path coefficient analyses. Significant differences were observed among the test genotypes for grain yield and most of the secondary traits studied. The mean grain yield ranged from 2441 to 5771 kg ha^{-1}. Five genotypes namely; G14, G25, G15, G40 and G21 were found to be significantly superior in grain yield and other major yield complementary traits. In addition, the genotypes exhibited different patterns of variations for yield contributing traits studied. Results showed that all the traits are quantitatively inherited indicating that grain yield can be improved through different selection methods. Appropriate selection strategies help to raise the gene frequency for grain yield.

The findings revealed that there were positive and highly significant correlation between grain yield with; number of branches per plant, number of nodes per plant, shelling percentage, red leaf blotch and number of days from 95% pod maturity to first pod shattering. Therefore, during selection main emphasis should be given to these traits for the enhancement of genetic potential for seed yield in soybean. Path coefficient analysis indicated that number of nodes per plant, plant height and 100 seed weight had positive and significant direct effect on grain yield but 100 seed weight had high negative indirect effects through number of pods, number of seeds and both days to 50% flowering and 95% pod maturity via number of pods, number of seeds and both days to 50% flowering and 95% pod maturity. Furthermore, number of nodes per plant and plant height had highest indirect effects via plant height and maturity respectively suggesting that indirect selection of these traits will be effective to improve grain yield. Clearly, it meant that number of nodes per plant could be recommended as the most efficient selection criteria and reliable selection indices for soybean grain yield improvement. Among the test genotypes, genotype, G14 exhibited the highest number of nodes per plant and was the second tallest genotype with the highest yield making it a superior candidate for hybridisation with a view to advancing grain yield. Overall, the investigation showed that there is great potential to breed cultivars with high grain yield because the latest generation of cultivars revealed higher mean values for yield contributing traits compared to founding early generation genotypes.

5.7 References

Aditya, J. P., 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.

Alaei, Y. 2012. Correlation analysis of corn genotypes morphological traits. International Research Journal of Applied and Basic Sciences 3:2355-2357.

Ariyo, O. J. 1995. Correlation and path coefficient analysis of components of seed yield in soybeans. African Crop Science Journal 3:29-33.

Arshad, M., N. Ali and A. Ghafoor. 2006. Character correlation and path coefficient in soybean [*Glycine max* (L.) Merrill.]. Pakistan Journal of Botany 38:121-130.

Aslam, M., N. A. Khan, M. S. Mirsa and A. R. Khan. 1992. Correlation and path coefficient analysis for yield components in soybean. Pakistan Journal of Agricultural Research 13:20-25.

Babu, V. R., K. Shreya, K. S. Dangi, G. Usharani and A. S. Shankar. 2012. Correlation and path analysis studies in popular rice hybrids of India. International Journal of Scientific and Research Publications 2:1-5.

Bizeti, H. S., C. G. P. de Carvalho, J. R. P. de Souza and D. Destro. 2004. Path analysis under multicollinearity in soybean. Brazilian Archives of Biology and Technology 47:669-676.

Board, J. E., M. S. Kang and B. G. Harville. 1997. Path analyses identify indirect selection criteria for yield of late-planted soybean. Crop Science 37:979-884.

Cheverud, J. M. 1988. A comparison of genetic and phenotypic correlations. Evolution 42: 42:958-968.

Cramer, C. S., T. C. Wehner and S. B. Donaghy. 1999 PATHSAS: A SAS computer program for path coefficient analysis of quantitative data. Journal of Heredity 90:260-262.

Cyprien, M. and V. Kumar. 2011. Correlation and path coefficient analysis of rice cultivars data. Journal of Reliability and Statistical Studies 4:119-131.

Dewey, D. R. and K. H. Lu. 1959. A Correlation and path-coefficient analysis of components of crested wheat-grass seed production. Agronomy Journal 51:515-518.

El-Badawy, M. M. and S. A. S. Mehasen. 2012. Correlation and path coefficient analysis for yield and yield components of soybean genotypes under different planting density. Asian Journal of Crop Science 4:150-158.

El-Din, A. A. T., E. M. Hessein and E.A.Ali. 2012. Path Coefficient and Correlation Assessment of Yield and Yield Associated Traits in Sorghum (*Sorghum bicolar* L.) Genotypes. American Eurasian Journal of Agriculture and Environmental Science 12:815-819.

El-Naim, A. M., I. M. Ibrahim, M. E. A. Rahman and E. A. Ibrahim. 2012. Evaluation of some local sorghum (*Sorghum bicolar* L. Moench) genotypes in rain fed. International Journal of Plant Research 2:15-20.

Ghodrati, G. R., R. Sekhavat, S. H. Hahmoodinezhadedezfully and A. Gholami. 2013. Evaluation of correlations and path analysis of components seed yield in soybean. International Journal of Agriculture 3:795-800.

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

Haider, Z., A. S. Khan and S. Zia. 2012. Correlation and path coefficient analysis of yield components in rice (*Oryza sativa* L.) under simulated drought stress condition. American Eurasian Journal of Agriculture and Environmental Science 12:100-104.

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

Jadhav, A. S., P. J. Jadhav and S. M. Bachehhav. 1995. Correlation and path coefficient analysis in soybean. Journal of Maharashtra Agricultural University. 20:150-151.

Joshi, B. K. 2005. Correlation, regression and path coefficient analyses for some yield components in common and Tartary buckwheat in Nepal. Fagopyrum 22:77-82.

Karnwal, M. K. and K. Singh. 2009. Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. Legume Research 32:70-73.

Kashif, M. and I. Khaliq. 2004. Heritability, correlation and path coefficient analysis for some metric traits in wheat. International Journal of Agriculture and Biology 6:138-142.

Khan, M. H. and A. N. Dar. 2010. Correlation and path coefficient analysis of some quantitative traits in wheat. African Crop Science Journal 18:9-14.

Khan, S., A. Latif, S. Q. Ahmad, F. Ahmad and M.Fida. 2011. Genetic variability analysis in some advanced lines of soybean (*Glycine max* L.). Asian Journal of Agricultural Sciences 3:138-141.

Lal, V. S. and M. F. Hague. 1971. Path anaylsis of yield components in soybean. Indian Journal of Genetics and Plant Breeding 31:357-362.

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

Makanda, I. 2009. Combining ability and Heterosis for Stem Sugar Traits and Grain Yield Components in Dual Purpose Sorghum [*Sorghum bicolar* (L) Moench) Germplasm Science and Agriculture, University of KwaZulu Natal.

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.

Malik, M. F. A., A. S. Qureshi, M. Arshraf and A. Ghafoor. 2006. Genetic variability of the main yield related character in soybean. International Journal of Agriculture and Biology 8:815-819.

Oz, M., A. Karasu, T. Goksoy and Z. M. Turan. 2009. Interrelationships of Agronomical Characteristics in Soybean (*Glycine max)* Grown in Different Environments. International Journal of Agriculture and Biology 11:85-88.

Payne , R. W., D. A. Murray, S. A. Harding, D. B. Baird and D. M. Soutar. 2011. An Introduction to GenStat for Windows (14th Edition) VSN International, Hemel Hempstead, UK.

Rafiq, C. M., M. Rafique, A. Hussain and Altaf. 2010. Studies on heritability, correlation and path analysis in maize (*Zea mays* L.). Journal of Agricultural Research 48.

Ramgiry, S. R. and P. Raha. 1997. Correlation and path analysis for seed yield and quality attributes in soybean Crop Research Hissar 13:137-142.

Ramteke, R., V. Kumar, P. Murlidharan and D. K. Agarwal. 2010. Study on genetic variability and traits interrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. Electronic Journal of Plant Breeding 1:1483-1487.

Sarutayophat, T. 2012. Correlation and path coefficient analysis for yield and its components in vegetable soybean. Songklanakarin Journal of Science Technology 34:273-277.

SAS Institute Inc. 2010. SAS proprietary software. Release 9.1.9 SAA Institute, Cary, North Carolina, US.

Sharma, S. M., S. K. Rao and U. Goswami. 1983. Genetic variation, correlation and regression analysis and their implications in selection of exotic soybean. Mysore Journal of Agricultural Science 17:26-30.

Shinde, A. K., S. P. Birari, S. G. Bhave and R. M. Joshi. 1996. Path coefficient analysis of seed yield and quantitative traits in soybean (*Glycine max* (L.) Merrill. Annl. Agric. Res 17:28-32.

Showkat, M. and S. D. Tyagi. 2010. Correlation and Path Coefficient Analysis of Some Quantitative Traits in Soybean [*Glycine max* (L). Merrill.]. Research Journal of Agricultural Sciences 2010 1:102-106.

Sreckov, Z., J. Bocanski, A. Nastasic, I. Dalovic and M.Vukosavljev. 2010. Correlation and path coefficient analysis of morphological traits of maize (*Zea mays* L.). Research Journal of Agricultural Science 42:292-296.

Sudaric, A. and M. Vrataric. 2002. Variability and interrelationships of grain quantity and quality characteristics in soybean. Euphytica 53:137-142.

Taylor, R. 1990. Interpretation of the Correlation Coefficient: A Basic Review. Journal of Diagnostic Medical Sonography 6:35-39.

Watt, E. D. and D. A. Levin. 1988. Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. Heredity 80:310-319.

Wright, S. 1921. Correlation and causation. Journal of Agricultural Research 20:557-587.

Wright, S. 1960. Path coefficients and path regressions: alternative or complementary concepts? Biometrics 16:189-202.

Yazdani, M. 2012. Correlation and path coefficient analysis studies on main agronomic characters with its yield of *sorghum bicolor*. international Journal of Agronomy and Plant production 3:645-650.

Zarei, B., D. Kahrizi, A. P. Aboughadareh and F. Sadeghi. 2012. Correlation and path coefficient analysis for determining interrelationships among grain yield and related characters in corn hybrids. International Journal of Agriculture and Crop Sciences 4:1519-1522.

CHAPTER 6
GENERAL OVERVIEW OF RESEARCH AND WAY FORWARD

6.1 Introduction

Knowledge of the genetic variability of the existing germplasm coupled with a good understanding of the nature, relative magnitude of the combining ability of parents and their progeny performance is a pre-requisite for genetic improvement and selection of traits of interest. Furthermore, estimation of the genetic gains over time help to assess the success of breeding effort, at the same time providing an opportunity to review the existing breeding strategies with a view to accomplish overall breeding efficiency. In the same vein, correlation and path analysis investigations target to achieve breeding efficiency through the exploitation of multiple traits selection approach. The objective of this chapter is to highlight the major findings, challenges and implications of the study to breeding.

The objectives of the present study were to;

- determine the level of genetic diversity among the soybean germplasm in the breeding programme in Zimbabwe.
- evaluate the level of genetic gains that has been made by the soybean breeding programme in Zimbabwe from 1940 to 2013.
- investigate the level of grain yield stability of soybean varieties released over 70 years and identify the best performing genotypes for general and specific adaptation Zimbabwe, Malawi and Zambia.
- identify the traits which have contributed significantly, and directly and indirectly to the high grain yield potential that has been realized over 70 years of soybean breeding in Zimbabwe.

6.2 Summary of the major findings

6.2.1 Genetic diversity analysis

The findings of diversity analysis were;

- Both the phenotypic (classical) and molecular characterisation revealed high levels of genetic diversity

- SSR markers were highly polymorphic, informative and more discriminatory than the phenotypic data and revealed 15 and 8 clusters respectively.
- There was good correspondence between the SSR clustering pattern and pedigree data suggesting that molecular characterisation was more reliable than the phenotypic characterisation.
- Genotypes, G41 and G7; G41 and G1; G41 and G42 were the most divergent; therefore, they could be used as source germplasm in cultivar development and they could as well be exploited as commercial cultivars.

6.2.2 Estimation of breeding progress

The study revealed the following observations;

- Highly significant differences associated with high ranges for all the traits studied were obtained demonstrating broad spectrum of diversity among the genotypes.
- Phenotypic coefficient of variation (PCV) was larger than genotypic coefficient of variation (GCV) and environmental coefficient of variation. However, small differences Between PCV and GCV were observed on grain yield, downy mildew scores, plant height and days to 95% pod maturity implying less influence of the environment on these traits.
- One hundred seed weight combined high heritability with moderate genetic advance indicating additive genetic control; hence, selection could be useful.
- The study revealed that annual genetic gains for grain yield were slowing down
- Realized genetic gain for grain yield was 47 kg ha^{-1} $year^{-1}$ over 48 years representing annual rate of 1.67%. Seed protein content decreased by 0.03% $year^{-1}$ while oil content increased by 0.11% $year^{-1}$. Hundred seed weight increased by 0.21 g yr^{-1} over time. Number of days to 95% pod maturity and first pod shattering increased by 0.35 and 0.38 days yr^{-1} respectively while lodging declined by 2.08% annually.

6.2.3 Genotype by environment interaction

The following observations were made;

- The first two interaction principal component scores (IPCAs) in AMMI analysis accounted for 58.5% of the total interaction variation

- Common genotypes that were selected by AMMI, cultivar superiority and rank analyses as highly productive and stable were; G1 and G15.
- Unstable genotypes but high yielding in specific environments were G2, G4, G5, G7, G16, G40, G17, G18 and G31.

6.2.4 Relationship between grain yield and its secondary traits

The findings of this investigation were;

- Grain yield showed positive and significant correlations with number of branches per plant, number of nodes per plant, shelling percentage and number of days from 95% pod maturity to first pod shattering.
- There were negative and significant correlation coefficients between grain yield with; plant height, percentage lodged plants at maturity, green stem scores, number of pods per plant, days from planting to 50% flowering, days from planting to 95% pod maturity, and percentage crude protein content in the seed on a dry matter basis.
- Number of nodes per plant and plant height had the highest direct effects on grain yield. Furthermore, results showed that number of nodes per plant and plant height had highest indirect effects via plant height and maturity respectively.
- Overall, genotype G14 revealed the highest mean performance on grain yield and associated components. Therefore, it could be used as source germplasm in cultivar development programme.

6.3 Implications of the research findings for breeding soybean cultivars and recommendations

- Existence of a broad spectrum of genetic variability for a variety of traits suggests that it can be utilized for further genetic improvement. Wide genetic diversity represents opportunities and room for selection of superior elite breeding lines in a hybridisation programme.
- Although a breeding progress of 57 kg ha^{-1} $year^{-1}$ has been made, further breeding progress can be accelerated by;
- The positive relationship between grain yield and number of nodes per plant accompanied by the observed highest direct and indirect effects of number of nodes per plant demonstrated that breeders should use them as reliable selection traits for developing high yielding genotypes of soybean.

- Existence of genotype x environment interaction justifies the need to conduct multi-environment trials (METs) and over many years with a view to selecting superior genotypes.

6.4 Challenges encountered

- Although the pedigrees of the genotypes used in the study were found, further records about the descendants could not be found. This limited the scope of tracing and comparing the relationships. A proper data-base management computer software should be put in place, of course backed up by hardcopies, electronic servers and external hard-drives.
- Limited resources dictated the number of environments to be used for genotype by environment interaction study i.e. few environments were used. In the same vein, two sites were used for combining ability studies which is not the ideal because these studies and inheritance of polygenic traits vary with environment. This is because few environments increase the contribution of pooled error and additive by environment variances (Sofi et al., 2006).

6.5 Recommendations and way forward

- Although genetic diversity exists, it is proposed to continue enriching it by bringing foreign germplasm. Pirra et al. (2009) postulated that gene pool expansion is an opportunity for creating variability in any breeding programme. This has the advantage of introducing unique alleles that raise the gene frequency for traits of interest which ultimately is necessary for improving the breeding gains. If resources are available, it is suggested to carry out diversity studies first and use the results to select parental lines that have a high genetic distance.
- Furthermore, the gains that have been made on disease resistance should also be quantified.
- It is also proposed to investigate interactions between linkats and loci underlying seed yield and soybean resistance.
- It is recommended to also focus on short statured or dwarf cultivars that can accommodate significantly high plant populations with a view to raise grain yield levels. Moreover, attention should also be devoted to “slow wilting” types of soybean in the light of an increase in the frequency and severity of droughts in the region.

- Based on the observations that were made on GEI study, cultivars G1 and G15 could be recommended for good general adaptation, hence they can be commercialized in Zimbabwe, Zambia and Malawi while genotypes G2, G4, G5, G7, G16, G40, G17 and G18 could be recommended for specific adaptation (Zimbabwe).
- Number of nodes per plant and plant height could be recommended as reliable selection traits for developing high yielding genotypes of soybean. The magnitude and direction of correlation between grain yield and number of nodes per plant was high and positive and again, it showed the highest positive and significant direct and indirect effects.

6.6 Conclusion

The aim of the present investigation was to characterise the genetic diversity of the available germplasm, determine gene action controlling grain yield and estimate the breeding gains that have been realized since the inception of the breeding programme in Zimbabwe. The specific objectives of the study were successfully accomplished as shown below;

6.6.1 Determination of genetic diversity

The investigation revealed evidence of a broad spectrum of diversity among the genotypes presenting large scope for cultivar improvement and selection.

6.6.2 Estimation of breeding gains

Annual genetic gain for grain yield was estimated to be 57 kg ha^{-1} $year^{-1}$ which was representing an annual improvement 2.06% but slowing down

6.6.3 Genotype by environment interaction

Results showed that environment and GEI captured larger proportion of the total sum of squares which essentially demonstrated the influence of the duo on grain yield indicating the necessity of evaluating soybean genotypes in multi-environment sites.

6.6.4 Correlation and path analyses

The study demonstrated that number of nodes per plant and plant height could be used as selection traits for developing high yield cultivars.

6.7 References

Pirra M. T., P. Sartre, R. Nelson, S. Santoni, N. Texier and P. Roumet. 2009. Genetic Diversity in a Soybean Collection. Crop Science 49:895-902.

Priolli R. H. G., J. B. Pinheiro, M. I. Zucchi, M. M. Bajay and Vello. 2010. Genetic Diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. Brazilian Archives of Biology and Technology 53:519-531.

Sofi P., A. G. Rather and S. A. Wani. 2006. Combining ability and gene action studies over environments in field pea (*Pisum sativum* L.). Pakistan Journal of Biological Sciences 9:2689-2692.

Spehar C.R. 1999. Diallel analysis for grain yield and mineral absorption rate of soybeans grown in acid Brazillian Savannah soil. Pesquisa Agropecuria Brasileira 34:1003-1009.