USE OF MOLECULAR MARKERS, AGRONOMICAL AND RELATED ATTRIBUTES ON DIVERSITY ANALYSIS OF SOYBEAN (*Glycine max*) GENOTYPES

By

KEITUMETSE KUJANE

Dissertation submitted in fulfilment of the requirements for the degree

MASTER OF AGRICULTURE

in the

Department of Agriculture

Faculty of Health and Environmental Sciences

at the

CENTRAL UNIVERSITY OF TECHNOLOGY, FREE STATE

Supervisor: Prof MM Sedibe

Co-supervisor: Dr MA Mofokeng

January 2021

DECLARATION

I, **Keitumetse Kujane,** do hereby declare that this research project submitted to the Central University of Technology, Free State for the degree Master of Agriculture, is my own independent work; complies with the Code of Academic Integrity, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology, Free State; and has not been submitted before to any institution by myself of any other person in fulfilment (or partial fulfilment) of the requirements for the attainment of any qualification.

SIGNATURE **DATE**

ABSTRACT

Soybean (*Glycine max* (L.) Merrill) is one of the most important leguminous crops grown globally for food, oil and animal feed. It is undoubtedly of a great economic and social importance worldwide. Soybean provides about 64 percent of the world's oilseed meal supply and is the major source of oil, accounting for about 28 percent of total production. Studies have revealed that South Africa is the leading producer of soybean in Africa and very few small-holder farmers practice soybean production. The characterisation of diverse soybean genotypes using agronomic, molecular markers and nutritional quality traits have not been fully utilised in order to achieve breeding, conservation and management goals. The objectives of the study were to: (i) determine the presence of genetic diversity among the soybean genotypes using agro-morphological traits; (ii) assess the level of genetic diversity present among the soybean genotypes using nutritional quality traits; and (iii) to determine the presence of genetic diversity among the soybean genotype making use of simple sequence repeat (SSR) markers. Thirty soybean genotypes were randomly selected from the Agricultural Research Council-Grain Crop (ARC) gene bank and were grown in a growth chamber until they reached the 4^{th} leaf stage. The leaves were then collected and freeze-dried, then subjected to genotyping using 20 polymorphic SSR markers. The SSR analysis revealed extensive variation among the soybean genotypes. The genotypes Santa Rosa and PR 165-52 had the closest distance (similarity), whereas B 66 S 31, 69 S 7 and R-5-4-2 M showed the highest dissimilarity index. The number and size of alleles ranged from 4 to 22bp and 2 to 33bp, respectively. The polymorphic information content (PIC) varied between 0.46 and 0.85; while the heterozygosity data points ranged between 0.50 and 0.87. The second trial was carried out with 30 single lines of soybean of 3m each in row length were planted in a non-controlled environment and replicated 3 times using a complete randomized block design. The soybean genotypes were subjected to characterization using agro-morphological traits. Principal component analysis revealed that the three most important components contributed 21.3%, 14.9% and 9.1% to the total variation in the field trial where 30 lines of soybean were evaluated for agro-morphological traits and nutritional

quality. The traits that contributed most to the variation were pod weight before threshing, number of branches per plant, pod number per plant, and yield per plant. These were further analysed for nutritional quality using near-infrared spectroscopy (NIR). The genotypes that had a higher protein and oil contents were Columbia M8A (37.54%) and B 66 S 256 (17.83%). Overall, the study found considerable levels of genetic variability among the soybean germplasm found at the Agricultural Research Council-Grain Crop gene bank using agro-morphological, SSR markers and nutritional quality traits. The selected lines should be useful for future breeding programmes while the knowledge of the genetic diversity can be used to direct efforts to conserve the diversity of soybean germplasm present locally and globally.

ACKNOWLEDGEMENTS

- I would like to express my sincere gratitude to specific individuals and institutes for support and assistance in the completion of this study, especially:
    - My most sincere gratitude and appreciation goes to my main supervisor, Professor Moosa Sedibe, for taking me under his wing and never growing tired of guiding and supporting me in making sure that this study is a success. I do not know anyone who is as patient, kind-hearted, wise but most of all prides themselves in seeing others excel and succeed.
    - I would also like to extend my profound gratitude to my co-supervisor, Dr. Alina Mofokeng, for valuable lessons and her efforts and contribution towards this research study.
    - My deepest gratitude goes to the Agricultural Research Council for the use of its facilities, without it the study would not have been possible. I am also very thankful for the staff and their timely help and cooperation.
    - I am eternally grateful to the Central University of Technology, Free State and, National Research Foundation for funding this project. Without them, none of this would have been possible.
    - Special thanks to Dr Zaid Bello for the reading and linguistic care of this dissertation.
- To the most supportive family, loving partner and the world's best daughter: thank you for the prayers, encouragement and sacrifices. All of this has come to be because of your presence in my life.
- Above all, praises unto God for carrying me through this journey and showing me that all things are possible.

DEDICATION

To the Most High God

Praises be unto the Lord for always answering me with wide-open spaces when I am in tight circumstances. His steadfast love and mercy never come to an end and He has proven just that by giving me strength to complete this study.

To my parents

Keolebogile and Otlotleng Kujane, I genuinely have not known love and support until I met you. Thank you for always believing in my hopes and dreams because today one of them has come true.

To my grandmother

Lontshamang Maphatse, thank you for your love and constant prayers. I am what I am today because God heard you when you wished good over my life.

To the father of my child

Ngwako Ndindwa, your unwavering love and support has not gone unnoticed. Thank you for always pushing and encouraging me to do better than yesterday.

To my daughter

Kganya Kujane, this study is especially dedicated to you. I have fought and soldiered on because your presence has taught and reminded me that giving up is never an option. Please continue being my light.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER ONE

GENERAL INTRODUCTION

1.1. Background

Soybean [*Glycine max* (*L.*) *Merr*.] is one of the most cultivated leguminous crop worldwide. Soybean is a source of vegetable oil and protein meal used by humans and livestock (Hartman et al., 2011). Grain SA (2017), reported a global increase of about 345 million tons for 2017/2018 season, estimated to be about 10.1% increase. The key contributors to this increased production are the United States of America, which accounts for 34%, followed by Brazil with 30%, Argentina 18%, China 4%, India 3.95%, Paraguay 3%, and small proportion from other related countries (Sihlobo and Kapuya, 2016). South Africa is the leading producer of soybean in Africa followed by Nigeria, Zambia, Malawi, Benin and Ethiopia (Sihlobo and Kapuya, 2016). Soybean production occurs in almost all nine provinces of South Africa, but significant production takes place in Mpumalanga and Free State provinces. Despite the increased production, South Africa still imports oil, oil cake and other soybean products from other countries (Sihlobo and Kapuya, 2016).

Soybean is used for both animal and human consumption as well as for industrial purposes. Vegetable oil, soybean cake and food products are processed from soybean seeds, with vegetable oil and soybean cake being the most dominant product. On a small scale, biodiesel, candles, soy ink and soy rayons are also produced from soybean (Opperman, 2011). Soybean accounts for 30% of the total edible vegetable oil. Reports also show that there is a growing demand for soybean in the poultry and piggery industries, which are also driven by escalating demand for meat products (Kapuya et al., 2010).

According to Fageria et al. (2005), cover crops have the ability to provide sufficient soil cover, produce organic matter with low-residue carbon/nitrogen ratio and

absence of allelopathic effects on a crop while fixing atmospheric oxygen. Soybean thus acts as a good cover crop because it has a large leaf biomass, which improves soil fertility and will benefit the next crop. Weed growth can also be suppressed by the large leaf biomass of soybean leaves (Mpepereki et al., 2000).

Soybeans are high in quality protein and like meat they contain all essential building blocks, like amino acids, in amounts needed for good health. Soybean is an excellent source of the minerals calcium and iron, which are responsible for building and maintaining strong bones and teeth as well as carrying oxygen to the muscle and tissue cells (Dieticians of Canada, 2014). There are two areas where soybean has been found to be most beneficial and that is for the heart and breast cancer (Dieticians of Canada, 2014). Research has also shown that replacing high fat animal foods in a diet with soy can be beneficial for menopausal women as it will cause a decrease in their blood pressure, lead to fewer hot flushes, lower the risk of osteoporosis and help prevent breast cancer (Dieticians of Canada, 2014).

The variations in nucleotides, genes, chromosomes, or whole genomes of an organism make up genetic diversity (Vaughan et al., 2007). This makes genetic diversity essential in order to maximize genetic improvement, which is accomplished through hybridisation, mutagenesis or any biotechnological means (Mutengwa, 2004; Satyavathi et al., 2006). Breeders are able to select quality parents from genetic diversity with respect to traits of interest for making combinations. A breeding programme with a broad genetic base also provides a valuable source of genes required for introgression purposes, as indicated by, (Erasmus, 2008). Breeders can work out crossing plans from a diverse germplasm where certain genes are introgressed into locally adapted varieties. Diversity can be broadened and also promote the increase of genetic gains if crosses are made from distant parents.

1.2. Problem statement and justification

The major concern regarding the soybean breeding programme 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,

particularly in South Africa. Lack of such valuable information, may adversely affect future breeding advances. Evaluation of genetic resources is a first step towards efficiency in utilization of the genetic resources through introduction of new genes as well as for their maintenance. In addition, the evaluation and characterization of the collection on the basis of morphological and agronomic traits are the starting point of any breeding program (Fundora, 1998). The genetic diversity of soybean genotypes in most African breeding programs has not been fully exploited, understood and well documented. The Agricultural Research Council-Grain Crops is maintaining more than two thousand soybean accessions that have not been fully characterized using molecular, morphological and nutritional quality traits. There is little information reported on diversity in relation to nutritional quality traits of these accessions. There is also a need for genotyping of the soybean accessions using molecular markers such as simple sequence repeat. This is because the SSRs are easily maintained and can be shared among laboratories (Maughan et al., 1995). They are highly reproducible, co-dominant, with low cost, and are abundant in the plant genome (Mofokeng, 2015). They are applied widely in genome and genetic mapping analysis, quantitative trait locus and gene analysis (Li et al., 2000) and in marker assisted breeding.

For a successful breeding programme to function a complete understanding of the genetic diversity of the crop is required. Better knowledge of the genetic similarities and dissimilarities of breeding material could aid breeders and curators in maintaining genetic diversity, sustain long-term selection gain and conserve the germplasm. Monitoring the genetic diversity among genetic resources of elite breeding material could make crop improvement more efficient by the directed accumulation of favoured alleles thus reducing the amount of material to be screened. For farmers to benefit from newly developed soybean cultivars there is need to have a clear understanding of the genetic relatedness of the available germplasm maintained in the gene banks (Mushoriwa, 2013). Ojo et al. (2012) stated that quantification of genetic diversity is fundamental to have excellent crop improvement program.

1.3. Purpose of the study

1.3.1. Aim

The aim of this study was to assess the presence of genetic diversity among selected soybean genotypes maintained by the Agricultural Research-Council-Grain Crops using simple sequence repeat, agronomic and related attributes.

1.3.2. Objectives

The specific objectives of the study were:

i. To determine the presence of genetic diversity among the soybean genotype using SSR markers.

ii. To determine the presence of genetic diversity among the soybean genotype using agro-morphological traits, and

iii. To determine the presence of genetic diversity among the soybean genotype using nutritional quality traits analysed using near-infrared spectroscopy (NIR).

Research hypotheses

i. There is agro-morphological diversity among the soybean genotypes.
ii. There are differences among the soybean genotypes based on nutritional quality traits and related attributes.
iii. There is vast diversity among the soybean genotypes based on SSR markers.

1.4. References

Dieticians of Canada (2014): What are the health benefits of soy? https://www.dieticians.ca. (Accessed: 8 March 2018).

Fundora, M.Z. (1998): La convention et l'utilisation des ressources phylogénétiques: Résumé de la troisième réunion latino-américaine sur la biotechnologie végétale, REDBIO'98. Palais des conventions de la Havana, Cuba. pp. 199-200.

Grain SA. (2017): The rise of global soybean production. http://www.grainsa.co.za/the-rise-of-global-soybean-production (Accessed: 8 March 2018).

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

Li, Y.C., Fahima, T., Peng, J.H., Roder, M.S., Kirzhner, V.M., Beiles, A. and Nevo, E. (2000): Edaphic microsatellite DNA divergence in wild emmer wheat, *Triticum dicocoides*, at a microsite: Tagigha, Israel. Theoretical and Applied Genetics 101:1029-1038.

Maughan, P.J., Saghai-Maroof, M.A. and Buss, G.R. (1995): Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715-723.

Mofokeng, M.A. (2015): Diversity analysis of South African sorghum genotypes using agronomic traits, SSR Markers and protein content and amino acid Composition. Doctorate thesis, University of KwaZulu-Natal, Pietermaritzburg.

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

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] Department of Crop Science, University of Zimbabwe, Harare.

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

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

Sihlobo, W. and Kapuya, T. (2016): South Africa's soybean industry: A brief overview: http://www.grainsa.co.za/south-africa-s-soybean-industry:-a-brief-overview. (Accessed 2017/05/03)

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

CHAPTER TWO

LITERATURE REVIEW

2.1. Soybean origin, description and domestication

Soybean is a self-pollinating crop that is believed to have originated from north-eastern China during 1700-1100 BC (Fukuda, 1993; Hymowitz, 2004), of which is therefore regarded as its centre of origin. This crop is a member of the Leguminosae family (Fabaceae) and its genus is *Glycine* and species *max*. *Glycine soja* is an annual plant native to north eastern Asia and it is a cultivated soybean similar to its wild progenitor (Doyle et al., 2004). *Glycine max* and *Glycine soja* carry chromosome number 2n=40 and they are both diploid. Furthermore, the cultivated subgenus *Glycine soja* consists of two species which are *Glycine soja* and *Glycine max* while the subgenus *Glycine* consists of seven species of the wild soybean and they are; *Glycine falcate, Glycine sericea, Glycine clandestine, Glycine latifolia, Glycine tabacina, Glycine latrobeana, Glycine canesscens and Glycine tomentalla* respectively. It is believed that current cultivated soybean was domesticated from wild soybean (*Glycine soja*) around 6000 – 9000 years ago in East Asia (Carter et al., 2004; Kim et al., 2012). Studies also suggest soybean was first introduced from the north-eastern China. According to (Dong et al., 2004; (Li et al., 2010), soybean landraces with the highest genetic diversity are found in the Huanghe region around the Huanghe River which is also known as the Yellow River. A study by (Guo et al., 2010) suggested southern China to be the place of birth for soybean based on clustering and phylogenetic analyses using microsatellites and nucleotides diversity. Lee et al. (2011) opposed the study on southern China stating that there is no current archaeological evidence that supports southern China as the origin of soybean domestication.

The history of soybean began in Southeast Asia were it was domesticated by Chinese farmers during the 1100 BC and was later grown in Japan as well as many other countries. Research studies revealed that soybean seed first reached America in 1765 when it was planted as a garden crop in Georgia around 1770. Later on the

seeds were distributed to farmers in Illinois and gained popularity leading to farmers growing them as forage for their livestock. The plants performed very well under hot and humid temperatures of North Carolina causing the United States Department of Agriculture to perform tests on them and also encourage farmers to cultivate them mainly for animal feed. The valuable source of protein and oil from soybean was further discovered and also led to the realization of benefits it had on preserving the quality of soil. Soybean grew so much that the US started off with 20 proven varieties only to realize later on that there is more than 10000 varieties that encouraged agricultural scientists to study this plant and this meant new and improved varieties for better production for farmers all over the world. Soybean farming reached a peak in the United States around the 1940's while production in China came to a halt due to World War II. Studies show that there thirty one states in the U.S that are involved in soybean production and they include Indiana, Illinoi, Minnesota and Iowa being the top producers, supplying to Africa, Asia and other parts of the world. One-tenth of the volume of soybean produced in Iowa compares to that of North Carolina but as a net importer of soybean meals and soybean itself, North Carolina ranks as high as many other countries (NC Soybean Producers Association, 2018).

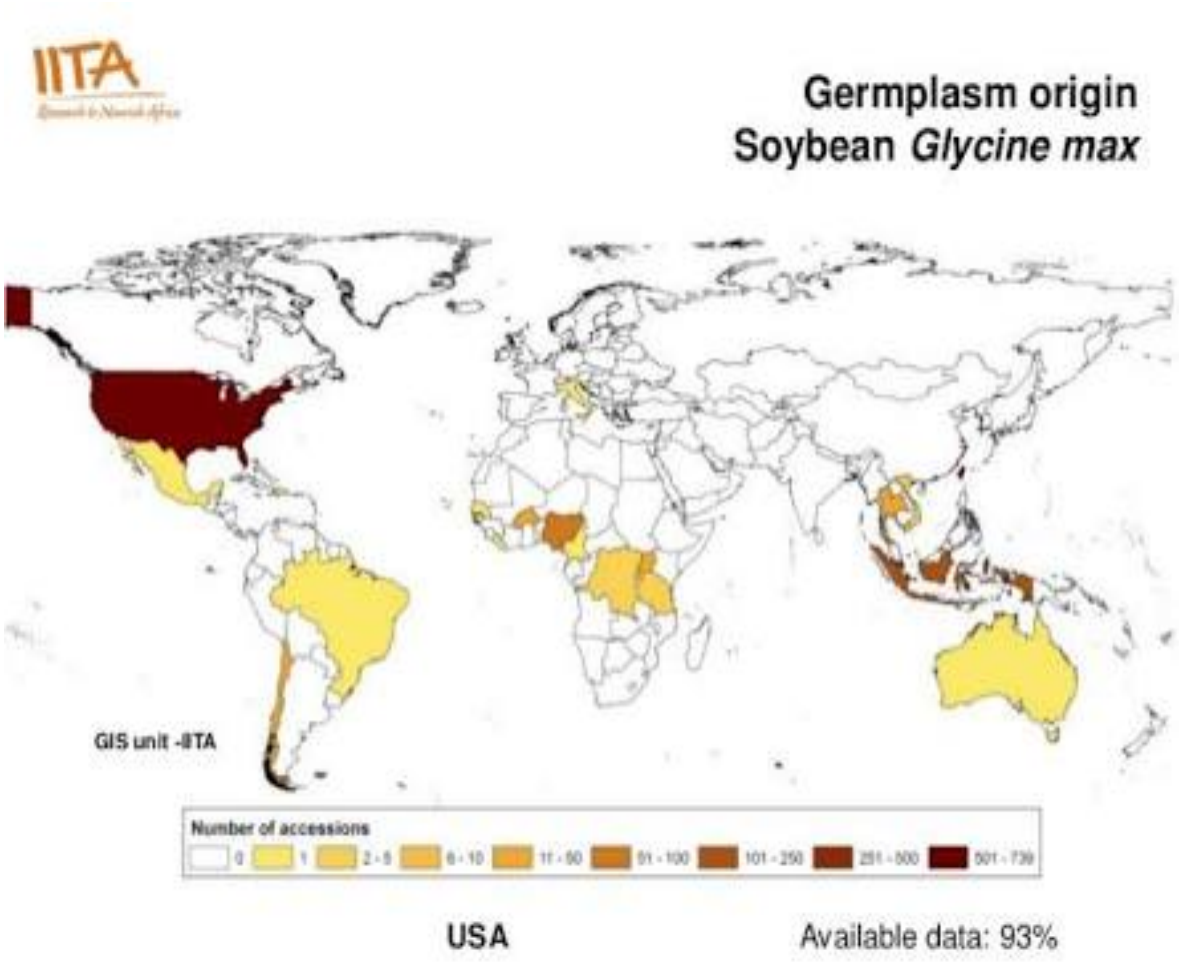


Figure 2.1: Germplasm collections, conservation and distribution of soybean.

Source: IITA (2008)

2.2. Soybean production in South Africa

Soybean is grown in a wide range of climatic conditions and it performs well under dryland and irrigation. It grows well in most high potential maize areas. Deep and well-drained soil with a sandy loam to clay loam texture is suitable for soybean production (Tattersfield et al., 1988). For maximum yield, soybean should be planted in soil that has a pH ($CaCl_2$) ranging from 5.3 to 5.5 (DAFF, 2010).

Soybean production in South Africa ranges from 450 000 to 500 000 tons per annum at an average yield of 2.5 to 3 ton/ha under dryland conditions (DAFF, 2010). Out of all nine provinces in South Africa, Mpumalanga produces the largest quantity of soybean by 42%, followed by the Free State with 22%, KwaZulu-Natal (15%),

Limpopo (8%), North West (5%) and Gauteng with 2%. According to a study by DAFF (2011), it has been proven that the Eastern and Western Cape provinces are the lowest producers of soybean with the Western Cape being the least producer since 2007/08 and 2009/10 production seasons.

Maphosa (2015) stated that very few small-holder farmers across South Africa practice soybean production. The author further elaborated that in 2006 a project run by Mapfura-Makhura Incubator (MMI) in partnership with Limpopo Department of Agriculture (LDA), was started with the intention to increase the soybean yield for biodiesel production in the Limpopo Province. The project assisted 76 farmers from the Sekhukhune district with production inputs such as seeds, herbicides and fertilizers as well as test and train their business and technical skills ability when coming to soybean and sunflower production (Maluleke, 2008). The project yielded great improvements for the farmers because production was increased.

2.3. Soybean genetic diversity and germplasm distribution

Knowledge on genetic diversity is of great significance for planning an efficient breeding programme for the improvement in any crop. Studies have revealed that thousands of breeding lines and varieties are developed each year through the soybean hybridization programs all over the world. Since China is rumoured to be the place of origin for soybean, Oliveira et al. (2010) reported that it holds the largest germplasm collection that is estimated to have 26 000 accessions of *Glycine max* and 6 200 of *Glycine* soja. The ranking is then followed by the United States of America which has about 16 999 *Glycine max* accessions, 1116 *Glycine soja*, and 919 perennial *Glycine* species. However, soybean is said to be the least diverse crop with the *Glycine* species lacking diversity. Such conditions are a result of domestication processes, intensive selection during breeding and also founder population effects (Mushoriwa, 2013). Provine (2004) explained founder effect as a loss of genetic variation caused by an establishment of a new population by few individuals coming from a large population. The few individuals will not represent the parental population genetically; they therefore represent low genetic variation. A research study done by Hyten et al. (2006) revealed that founding events occur

and new populations and production environments are created when few cultivars are introduced to them. There are alterations to the allele frequencies due to domestication and founder effects that contribute to genetic bottlenecks. This also eliminates rare alleles and lessens genetic diversity, which depends on the selection pressures.

2.4. Methods used to determine genetic diversity

Various methods are used to determine genetic diversity including agro-morphological and nutritional quality traits, and the use of DNA based markers.

Genetic diversity based on agro-morphological traits

Determination of genetic diversity using agro-morphological traits in soybean is one of the commonly used methods by researchers. Although it may be influenced by environment, is affordable and may be used for improving quantitative and qualitative traits preferred by the farmers and end user's. Phenotypic characters that are normally assessed for genetic variability include quantitative traits such as number of seeds per plant/pod, number of pods per plant, 100 seed weight, seed weight before and after threshing, days to flowering/maturity, plant height and other related traits (Chen and Nelson, 2004; Dayaman et al., 2009, Liu et al., 2011, Malik et al., 2011, Matsuo et al., 2011; Salimi et al., 2012). The qualitative traits include flower colour, hilum colour, pod hair colour, pubescence colour, leaflet shape, stem determination, pubescence type and density and nodes at flowering (Mushoriwa, 2013).

Manjaya and Bapat (2008) carried out a study on agro-morphological traits of soybean and they observed the genetic variation among the 55 soybean genotypes using the above-mentioned phenotypic traits. A similar study was conducted by Antalikova et al. (2008), using 52 agro-morphological traits to measure variability among the 52 genotypes that were studied. According to Ojo et al. (2012) a study was performed on 42 genotypes and seven clusters were found. The results showed great phenotypic variation among 100 seed weight, seed yield per plot,

number of pods per plant and pod yield per plant. Moe and Girdthai (2013) conducted an agro-morphological diversity study on 94 soybean accessions where traits such as flower colour, mature pod colour, pubescence colour, seed coat colour and hilum colour, leaf shape, days to 50% flowering, days to pod formation, days to maturity, plant height at maturity, number of seeds per pod, number of filled pods per plant, 100 seed weight, seed yield and harvest index were recorded. According to the study, the tested traits were positive and highly significantly correlated with the seed yield (days to 50% flowering, days to pod formation, days to maturity, plant height, number of filled pods per hill, 100-seed weight and harvest index). The study also indicated that the positive correlation between seed yield and seed per pod was non-significant.

Marconato et al. (2016) also reported the presence of genetic diversity among the soybean accessions maintained by EMBRAPA germplasm bank using qualitative and quantitative traits.

2.5. Genetic diversity based on molecular markers

Molecular markers are segments of DNA that breeders use to detect the presence or absence of specific alleles of interest and thus use them as selection tools (Beckmann and Soller., 1983; Darvasi et al., 1994). Molecular markers can be used in various ways including mapping of genes of interest and to assess genetic diversity in crop species. DNA markers make it possible to observe and exploit genetic variation in the entire genome. According to Moose and Mumm (2008), studies on molecular markers and diversity in soybean are conducted to analyse trends over time, investigate relationships within the soybean population, perform varietal maintenance, analyse diversity and formulation of germplasm conservation and maintenance strategies, and then lastly to analyse the historical understanding of soybean in a particular area.

Various DNA marker techniques are used for assessing genetic diversity in crop species. The techniques include amplified fragment length polymorphism (AFLP), SSR, single nucleotide polymorphism (SNP), Diversity Array Technology (DArT),

Microarray, genotype by sequencing (GBS), genome wide association studies (GWAS) and next generation sequencing (NGS) (Bonin et al., 2005).

Amplified fragment length polymorphism

The amplified fragment length polymorphism (AFLP) technique is the most important marker that is cost-effective and also an informative fingerprinting method. According to Bonin et al. (2005), AFLPs produce polymerase chain reaction-based multi-locus genotypes that are helpful in the areas of population genetics. DNA is cut by the restriction enzymes that cause the adaptors to attach to the end of the fragments; this then leads to fragments being amplified using polymerase chain reaction and their varying lengths can be visualised on gel or capillary-based platforms. AFLP is known to require large amounts of high-quality DNA but is also sensitive for detecting genetic polymorphism; therefore, AFLPs are not ideal for genotyping but can and has been used for population genetics of plants (Foster et al., 2011). According to a study by Mendelson and Shaw (2005), AFLP can be used to estimate relationships among closely related species; it also has a finer resolution that can indicate greater differentiation. This type of technique has also been used to assess the genetic diversity of 72 soybean varieties cultivated in India and 12 selected AFLP primer pairs produced 1319 products of which 1257 were polymorphic (95%) (Satyavathi et al., 2006).

Simple sequence repeat

Simple sequence repeats have been extensively used in soybean, having both advantages and limitations. They are distributed over the genome, having sequences of short tandem repeats. SSRs serve as a tool for genotype differentiation, genotype identification, pedigree analysis and genetic evaluation because of their hyper-variable number of repeat sequence (Satyavathi et al., 2006). These markers are known for their ability to perform successful identification of genetic diversity and relationships among soybean genotypes in different populations (Guan et al., 2010; Wang et al., 2010). A study carried out by Kumar and Lal (2015) on 91 soybean germplasm lines showed results of 28 alleles per loci, with an average of 2.8 alleles per locus and the polymorphic information content

(PIC) value of 10 polymorphic markers ranged from 0.40 to 0.67, with an average of 0.46. SSR markers have also been used to perform analysis in crops such as cowpea (Adetiloye et al., 2013), maize (Shiri et al., 2014), rice (Yao et al., 2015), wheat (Sutapa et al., 2014), barley (Wang et al., 2010), cassava (Njoku et al., 2013) as well as cotton (Zhao et al., 2015).

Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) have high frequency across the whole genome. They are easy to detect, co-dominant and also cost efficient. They represent a DNA sequence variant of a single base pair where the minor allele occurs in more than 1% of a population (Angaji, 2011). SNPs are reported to be the third-generation molecular markers where their application has sparked a need for effective instruments for SNP detection (Angaji, 2011). SNP detection can be used to scan DNA sequences for unknown polymorphisms and genotyping known polymorphisms. It has also been reported that there are many options for SNP genotyping although there are technologies capable of scanning DNA for new polymorphisms used in screening individuals for known polymorphism (Angaji et al., 2017). According to (Dudley and Karczewski, 2013), SNPs are the most widely tested markers in genetic diversity studies. A recent study by (Wang et al., 2018) has reported that 4 961 SNP markers out of 5 039 were mapped into 20 chromosomes and 31 scaffolds. Excluding the 78 unmapped markers, 4 930 were successfully onto the 20 soybean genome chromosomes. A further 459 unbiased SNP was deleted and it was defined as the frequency of two homogenous nucleotide at a given locus in a batch that showed a value of 0.85 or higher. The study described that 4 471 polymorphic SNP of the 0.85 threshold were enclosed for GWAS and population structure analysis.

Diversity Array Technology

Kilian et al. (2003) describes diversity arrays technology, also known as DArT, as a hybridization-based genotyping technology that is created to rapidly and simultaneously type and identify DNA polymorphism on the microarray platform. It was initially developed to overcome limitations that hinder the application of

microarray technology to non-model organisms. Dominant markers that result from single nucleotide polymorphism (SNP) at restriction sites can be detected by DArT (Wittenberg et al., 2005) at hundreds to thousands of arbitrary genomic loci (Wittenberg et al., 2005; Wenzl et al., 2004). Genetic diversity studies that were performed using DArT to infer diversity were for rice (*Oryza sativa* L.) (Jaccoud et al., 2001), grand eucalyptus (*Eucalyptus grandis* Hill ex Maiden; Lezar et al., 2004), wheat (*Triticum aestivum L.)* (Akbari et al., 2006), cassava (*Manihot esculenta* Crantz) (Xia et al., 2005), barley (*Hordeum vulgare* L.) (Wenzl et al., 2004), and in the model organism thale cress (*Arabidopsis thaliana* L. Heynh*)* (Wittenberg et al., 2005).

Microarray

A DNA microarray can be described as a collection of microscopic DNA spots that are attached to a solid surface and they are known as a chip or biochip. DNA microarrays are used by scientists to genotypes multiple regions of a genome or to measure expression levels of large numbers of genes. DNA microarrays are usually used for the detection of DNA or RNA that may or may not be translated into proteins. A study revealed that a microarray is the hybridization of two DNA strands having complementary nucleic acid sequences that form hydrogen bonds between complementary nucleotide base pairs (Elshire et al., 2011).

Genotype by sequencing

Genotype by sequencing (GBS) is a rapid approach for a reduced-representation sequencing of multiplexed samples that form a combination of genotyping and genome-wide molecular markers discovery (Elshire et al. (2011). GBS are flexible and relatively affordable, making them an excellent tool for research questions and application in breeding and plant genetics (Poland and Rife, 2012). GBS produces multiplex libraries of samples ready for next generation sequencing (NGS) using enzyme-based complexity reduction and DNA barcoded adapters. Elshire et al. (2011) and Poland et al. (2012a) describes the approach as robust across a range of species making it capable to produce tens of thousands to hundreds of molecular markers. Due to GBS being flexible, it is regarded as an ideal tool for plant genetic study in regards to research objectives, species and also populations. Studies on

whole-genome re-sequencing include rice (*Oryza sativa L*.), and maize (*Zea mays L*.) (Huang et al., 2009; Ashelford et al., 2011; Gan et al., 2011; Chia et al., 2012; Jiao et al., 2012; Xu et al., 2012) as well as *Arabidopsis thaliana* (L.) Heynh. Management also becomes difficult once genomes enlarge and become more complex, causing them to lack a solid reference genome (Morrell et al., 2011).

Next generation sequencing

Next generation sequencing is a technique developed in response to a need for capability to sequence larger numbers of samples at a relatively low cost. It also facilitates the development of methods that are used to genotype large numbers of single-nucleotide polymorphisms (SNPs) (Vlk and Řepková, 2017). Next generation sequencing technologies are tools used for crop reference genomes assembly, whole-genome molecular marker development, identifying markers in known-function genes, and also transcript-tome sequencing for the study of gene expression (Vlk and Řepková, 2017). A study by Ando et al. (2012) revealed that genes associated with phases of development using 454 sequencing were identified in the transcriptome analyses of cucumber (*Cucumis sativus L*.). Nigam et al. (2014) also conducted a study using both the combination of Roche technology and microarray with the aim of identifying genes and their products associated with cotton fibre quality.

2.6. Genetic diversity based on nutritional quality traits

Shenk et al. (2001) described the near-infrared spectroscopy (NIR) as a powerful, non-destructive tool that is able to estimate the level of chemical entities in a range of environments following careful calibration. NIR is used for monitoring quality and it is mostly used by plant breeders. It is responsible for providing quantitative and qualitative data on all the constituents obtained from a sample in a single spectral reading (Frenzel, 2003). A report by De Boever et al., (1994) states that NIR has been used to asses plant materials and their wide variety of characteristics. Cereals which include: oats, wheat, maize, barley and sorghum have had their total and phytate and phosphorus levels analysed using NIR. Nutritional traits such as protein, fat, fibre, starch and ash have been successfully measured in 22 different types of feed (Gerlach, 1990). According to Velasco et al. (1999), Orman et al. (1991) and Baye et al. (2004), NIR is a reliable technique that makes bulk grain sample screening possible even without prior preparation. However; bulk grain analysis does not allow the identification of an individual grain which is highly desired by plant breeders (Baye et al., 2006). Studies on soybean, sunflower meal, peas, fish meal, poultry and meat meal products have successfully calibrated for essential amino acids using NIR (Fontaine et al. 2001). A study by Fontaine et al. (2001) and Fontaine et al. (2002) highlights NIR as more superior to other methods/tools when estimating the levels of amino acids. It was even used to classify gene combinations and endosperm genes in a study of milled barley conducted by Jacobsen et al. (2005).

2.7. Grain yield of soybean

Genetic variability and grain yield in soybean

According to Karnwal and Singh (2009), the genetic variability for soybean is observed with high significant differences among the cultivars that are studied. Therefore, when deciding to develop a cultivar that is high yielding, traits such as seeds per plant, number of pods and seed weight should be the most emphasised because these traits help accelerate genetic advance in grain yield. Warkard et al. (2008) and Aditya et al. (2011) emphasised the importance of genetic variability in

soybean breeding programs for crop improvement. Since the level of grain yield is a function of the combinations of its components, knowledge of this is very important to identify yield components that can be used as selection criteria in advancing grain yield. An understanding of how yield components interact both genotypically and phenotypically in influencing yield is of great importance. A study by Ramteke et al. (2010) suggested that soybean grain yield is a link between seeds per pod, 100 seed weight, plants per area, branches per plant, as well as pods per branch. Cultivars showing superiority should be the focal point and be used as parental stock (Warkard et al., 2008).

2.8. Conclusion

Studies on soybean genetic diversity are intended to highlight genotypes with good agronomic performance and high nutritional quality that will be documented. They will also be used in the selection of parents' hybridization and soybean crop improvement programs. Genetic diversity in soybean can be decreased by plant breeding domestication and founder effects. Decreased diversity has a downfall, which causes things such as gene vulnerability, genetic erosion and plateau. More studies have revealed that genetic diversity of soybean genotypes in most African breeding programmes has not been fully exploited, understood and/or well documented. It is crucial for plant breeders to have a full understanding of the genetic diversity of crops, especially ones they are dealing with (Priolli et al., 2010).

2.9. References

Adetiloye, I.S., Ariyo, O.J., Alake, C.O., Oduwaye, O.O. and Osewa, S.O. (2013): Genetic diversity of some selected Nigeria cowpea using simple sequence repeats (SSR) marker. African Journal of Agricultural Research 8:586-590.

Aditya, J. P., Bhartiya, P. and Bhartiya, A. (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.

Akbari, M., Wenzl, P., Caig, V., Carling, J. and Xia, L. (2006): Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theoretical and Applied Genetics 113:1409–1420.

Ando, K., Carr, K.M. and Grumet, R. (2012): Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. BMC Genomics13:518.

Angaji, S.A. (2011): Genomics at a Glance: Molecular Markers & Genome Mapping. Lambert Academic Publishing.

Angaji, S.A., Ahmadzadeh, M., Yazdi Rouholamini, S.E. and Mohammadi, S. (2017): Single Nucleotide Polymorphisms (Applications & Techniques). Asar-e-Nafis.

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

Ashelford, K., Eriksson, M.E., Allen, C.M., D'Amore, R., Johansson, M., Gould, P., Kay, S., Millar, A.J., Hall, N., and Hall, A. (2011): Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. Genome Biology 12:R28. doi:10.1186/gb-2011-12-3-r28

Baye, T., and Becker, H.C. (2004): Analyzing seed weight, fatty acid composition, oil, and protein contents in Vernonia galamensis germplasm by near infrared reflectance spectroscopy. Journal of the American Oil Chemists Society 81: 641–645.

Baye, T.M., Pearson, T.C., and Settles, A.M. (2006): Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. Journal of Cereal Science 43:236-243.

Bonin, A., Pompanon, F. and Taberlet, P. (2005): Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. Methods Enzymology; 395: 145-61.

Carter, T Jr., Hymowitz, T. and Nelson, R. (2004): Biogeography, local adaptation, Vavilov, and genetic diversity in soybean. In: Werner D, ed. Biological resources and migration. Berlin, Germany: Springer, 47-59.

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

Chia, J.M., Song, C., Bradbury, P.J., Costich, D., de Leon, N., Doebley, J., Elshire, R.J., Gaut, B., Geller, L., Glaubitz, J.C., Gore, M., Guill, K.E., Holland, J., Hufford, M.B., Lai, J., Li, M., Liu, X., Lu, Y., McCombie, R., Nelson, R., Poland, J., Prasanna, B.M., Pyhäjärvi, T., Rong, T., Sekhon, R.S., Sun, Q., Tenaillon, M.I., Tian, F., Wang, J., Xu, X., Zhang, Z., Kaeppler, S.M., Ross-Ibarra, J., McMullen, M.D., Buckler, E.S., Zhang, G., Xu, Y. and Ware, D. (2012): Maize HapMap2 identifies extant variation from a genome in flux. Nature Genetics 44:803–807. doi:10.1038/ng.2313

Dayaman, V. N., Senthil, M., Raveendran, N., Nadarajan, P., Shanmugasundaram and Balasubramanian, P. (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 Boever, J.L., Eeckhout, W., and Boucque, C.V. (1994): The possibilities of near infrared reflection spectroscopy to redict total-phosphorus, phytate-phosphorus and phytase activity in vegetable feedstuffs. Netherlands Journal of Agricultural Science 42: 357–369.

Department of Agriculture, Forestry and Fisheries (2010): Soybean production guidelines. Government printers: Pretoria.

Department of Agriculture, Forestry and Fisheries (2011): Soybean market Value Chain Profile 2010–2011. http://www.daff.gov.za/docs/AMCP/SoybeanMVCP2010-2011.pdf. Accessed 22/01/2020.

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

Dudley, J.T. and Karczewski, K.J. (2013): Exploring personal genomics. Oxford University Press.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011): A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379. doi:10.1371/journal.pone.0019379.

Fontaine, J., Horr, J., and Schirmer, B. (2001): Near-infrared spectroscopy (NIRS) enables the fast and accurate prediction of the essential amino acid contents

in soy, rape seed meal, sunflower meal, peas ,fishmeal, meat meal products and poultry meal. Journal of Agricultural Food Chemistry 42:2726-2731.

Fontaine, J., Schirmer, B., J., Horr. (2002): Near-infrared spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran and sorghum. Journal of Agricultural and Food Chemistry 50:3902–3911.

Foster, J.T., Bull, R.L. and Keim, P.S. (2011): Ricin Forensics: Comparisons to Microbial Forensics. Microbial Forensics 12: 315-326.

Frenzel, T. (2003): Safety assessment of genetically modified food – New methodologies for the analytical characterization of rice. PhD Thesis, Technical University of Munich, Germany.

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

Gan, X., Stegle, O., Behr, J., Steffen, J.G., Drewe, P., Hildebrand, K.L., Lyngsoe, R., Schultheiss, S.J., Osborne, E.J., Sreedharan, V.T., Kahles, A., Bohnert, R., Jean, G., Derwent, P., Kersey, P., Belfield, E.J., Harberd, N.P., Kemen, E., Toomajian, C., Kover, P.X., Clark, R.M., Rätsch, G., and Mott, R. (2011): Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature 477:419-423. doi:10.1038/nature10414

Gerlach, M. (1990): NIR measuring technology for quality evaluation of feeds. Kraffuner 2: 67–74.

Guo, J., Wang, Y., Song, C., Zhou, J., Qiu, L., Huang, H. and Wang, Y. (2010) A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implicationsfrommicrosatellitesandnucleotidesequences. Annals of Botany 106:505–514.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T. and Han, B. (2009): High-throughput genotyping by whole-genome re-sequencing. Genome Resesearch. 19:1068–1076. doi:10.1101/gr.089516.108

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.

Hyten, D. L. (2006): 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.

Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. (2001): Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29:1–7.

Jacobsen, S., Sondergaard, I., Moller, B., Desler, T., and L. Munck. (2005): A chemometric evaluation of the underlying physical and chemical patterns that support near infrared spectroscopy of barley seeds as a tool for explorative classification of endosperm genes and gene combinations. Journal of Cereal Science 42:281–299.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W., Zhang, M., Xie, S. and J. Lai. (2012): Genome-wide genetic changes during modern breeding of maize. Nature Genetics 44:812-815. doi:10.1038/ng.2312

Kumar, M. and Lal, S.K. (2015): Molecular analysis of soybean varying in water use efficiency using SSRs markers. Journal of Environmental Biology 36:1011-1016.

Kapuya, T., Saruchera, D., Jongwe, A., Mucheri, T., Mujeyi, T., Traub, L.N. and Meyer, F. (2010): The Grain Industry Value Chain. Zimbabwe.

Karnwal, M. K. and Singh, K. (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.

Kim, M.Y., Van, K., Kang, Y.J., Kim, K.H. and Lee, S.H. (2012): Tracing soybean domestication history: from nucleotide to genome. Breeding Science 61:445–452.

Kilian A, Huttner E, Wenzl P, Jaccoud D and Carling, J. (2003): The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In: Tuberosa R, Phillips RL, Gale M, eds. In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution. Bologna, Italy: Avenue Media. pp 443–461.

Lee, G-A., Crawford, G.W., Liu, L., Sasaki, Y and Chen, X. (2011): Archaeological soybean (*Glycine max*) in East Asia: does size matter? PLoS ONE 6: e26720.

Liu, M., Zhang, M., Jiang, W., Sun, G., Zhao, H. and Hu, S. (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., Ashraf, M., Qureshi, A.S. and Khan, M.R. (2011): Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pakistan Journal of Biological Sciences 43:1249-1255.

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

Maphosa, T.K. (2015): Growth, nodulation and yield response of promiscuous and non-promiscuous soybean cultivars to inoculation in different soil types under glasshouse and field conditions. Master's dissertation. University of Limpopo.

Maluleke, N. (2008): Sekhukhune farmers receive seeds and fertilizers. Official newsletter of Mapfura-Makhura Incubator. Volume 1.

Marconato, M.B., Pereira, E.M., Silva, F.M., Bizari, E.H., Pinheiro, J.B., Mauro, A.O. and Unêda-Trevisoli, S.H. (2016): Genetic divergence in a soybean (*Glycine max*) diversity panel based on agro-morphological traits. Genetics and Molecular Research 15 (4). Doi: 10.4238/gmr15048980.

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

Moe, S. and Girdthai, T. (2013): Relationships of soybean [*Glycine max* (*L*.) Merrill] accessions based on physiological and agro-morphological traits. International Journal of Chemical, Environmental & Biological Sciences 1(2): 2320–4079.

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

Morrell, P.L., E.S. Buckler, and J. Ross-Ibarra (2011): Crop genomics: Advances and applications. Nature Reviews Genetics. 13:85–96.

Mpepereki, S., Javaheri, F., Davis, P. and Giller, K.E. (2000): Soybeans and sustainable agriculture: Promiscuous soybeans in Southern Africa. Field Crops Research, 65:137-149.

Mushoriwa, H. (2013): Breeding gains, diversity analysis and inheritance studies on soybean [*Glycine max (L.)* Merrill] Germplasm in Zimbabwe. Doctorate thesis, University of KwaZulu-Natal, Pietermaritzburg.

Nigam D., Kavita P., Tripathi R.K., Ranjan A., Goel R., Asif M., Shukla A., Singh G., Rana D., Sawant S.V. (2014): Transcriptome dynamics during fibre development in contrasting genotypes of Gossypium hirsutum L. Plant Biotechnology Journal 12:204–18.

Njoku, D.N., Egesi, C.N., Okogbenin, E., Gracen, E.V., Asante, I. and Offei, S. (2013): Assessment of genetic diversity in cassava based on SSR markers and morphological characters. Journal of Sustainable Agriculture and the Environment 14(1/2):1-16.

Oguntibeju, O.O., Meyer, S., Aboua, Y.G. & Goboza, M. (2016). *Hypoxis hemerocallidea* Significantly Reduced Hyperglycaemia and Hyperglycaemic-Induced Oxidative Stress in the Liver and Kidney Tissues of Streptozotocin-Induced Diabetic Male Wistar Rats. Evidence-Based Complementary and Alternative Medicine 1-10.

Ojewole, J. A. O. (2006). Antinociceptive, anti-inflammatory and antidiabetic properties of *Hypoxis hemerocallidea* Fisch. & C.A. Mey. (Hypoxidaceae) corm ["African Potato"] aqueous extract in mice and rats. Journal of Ethnopharmacology 103(1):126-134.

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

Okigbo, R.N., Eme, U.E. & Ogbogu, S. (2008). Biodiversity and conservation of medicinal and aromatic plants in Africa. Biotechnology and Molecular Biology Reviews 3 (6):127-134.

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

Orman, B.A, Schumann, R.A. (1991): Comparison of near-infrared spectroscopy calibration methods for the prediction of protein, oil, and starch in maize grain. Journal of Agricutural and Food Chemistry 39:883–886.

Opperman, C. (2011): Technical Report: Soybean Value Chain. Southern Africa Trade Hub. USAID. p7-12.

Otunola, G.A. & Afolayan, A.J. (2019). Proximate and elemental composition of leaf, corm, root and peel of *Hypoxis hemerocallidea*: A Southern African multipurpose medicinal plant. Pakistan Journal of Pharmaceutical Sciences 32 (2):535-539.

Owira, P. M. O., & Ojewole, J. A. O. (2009). "African potato" (*Hypoxis hemerocallidea* corm): a plant-medicine for modern and 21st century diseases of mankind? – A review. Phytotherapy Research 23(2):147-152.

Pereus, D., Otieno, J.N., Ghorbani, A., Kocyan, A., Hilonga, S. & De Boer, H.J. (2018). Diversity of *Hypoxis* species used in ethnomedicine in Tanzania. South African Journal of Botany 122:336-341.

Pii, Y., Cesco, S., & Mimmo, T. (2015): Shoot ionome to predict the synergism and antagonism between nutrients as affected by substrate and physiological status. Plant Physiology and Biochemistry 94:48-56.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sanchez-Villeda, H., Sorrells, M. and Jannink, J.L. (2012a): Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Gen. (in press). doi:10.3835/ plantgenome2012.06.0006

Potgieter, M.J., Erasmus, L.J.C., Semenya, S.S., Rankoana, S.A. And Potgieter, A. (2017): The Use of *Hypoxis Hemerocallidea* and *H. Obtusa* in Bapedi Phytomedicine in the Limpopo Province, South Africa. International Journal of of Ayurveda 2(6):1-9.

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

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

Raimondo D. (2011): The Red List of South African plants-A global first. *South* African Journal of Science 107(3/4):1-2.

Ramawat, K.G. (2007): Secondary plant products in nature. In: Biotechnology: Secondary Metabolites, Merillon, J.M. (Ed.), ProQuest Ebook Central, 21-57. [Accessed 23 July 2020].

Ramteke, R., Kumar, V., Murlidharan, P. and Agarwal, D.K. (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. R., Palada, M.C., & Becker, B.N. (2004): Medicinal and aromatic plants in agroforestry systems. Agroforestry Systems 61-62(1-3):1071-22.

Rao, B.R.R & Rajput, D.K. (2010): 'Global Scenario of Medicinal Plants.' National Conference on Conservation of Medicinal Plants - Herbal products and their Uses (CMP-HPU) (1), Arts and Science College for Women, Hyderabad, India, November, 2010. https://www.researchgate.net/publication/270512088_Global_Scenario_of_Medicinal_Plants [Accessed 17 August, 2020].

Rietra, R. P.J.J., Heinen, M., Dimkpa, C.O., & Bindraban, P.S. (2017): Effects of

Nutrient Antagonism and Synergism on Yield and Fertilizer Use Efficiency. Communications in Soil Science and Plant Analysis 48(16):1895-1920.

Roosta, R.A., Moghaddasi, R. & Hosseini, S.S. (2017): Export target markets of medicinal and aromatic plants. Journal of Applied Research on Medicinal and Aromatic Plants 7:84-88.

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

Rungqu, P., Oyedeji, O.O. & Oyedeji, A.O. (2019): Chemical Composition of *Hypoxis hemerocallidea* Fisch. & C.A. Mey from Eastern Cape, South Africa. In: Chemistry for a Clean and Healthy Planet. Ramasami, P., Gupta Bhowon, M., Jhaumeer Laulloo, S., & Li Kam Wah, H. (Eds.). Springer, 111-122.

Sathekge, N.R. (2010): Comparison of secondary metabolite content and antimicrobial activity of four *Hypoxis* species used in traditional medicine. Magister Scientiae Dissertation, University of Pretoria, South Africa**.**

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

Shakya, A.K. (2016): Medicinal plants: Future source of new drugs. International Journal of Herbal Medicine 4: 59-64.

Sharma, B.L., Singh, D., Sharma, S.K., Hashmi, A., Singh, A. & Bansal, A. (2013): Studies on Some Primary Metabolite's Extraction and Quantification in Different Plant Parts of Selected *Cassia* Species. Asian Journal of Pharmaceutical and Clinical Research 6:309-314.

Shenk, J.S., Workman, J.J., and Westerhaus, M.O. (2001): Application of NIR Spectroscopy to agricultural products. In: Burns, D.A., and E.W. Ciurczak (eds). Handbook of Near-infrared analysis: Second edition, revised and expanded. Practical Spectroscopy series vol 27. Marcel Dekker, New York.

Sher, H., Aldosari, A., Ali, A. & De Boer, H.J. (2014). Economic benefits of high value medicinal plants to Pakistani communities: an analysis of current practice and potential. Journal of Ethnobiology and Ethnomedicine 10: 1-16.

Shinwari, Z.K. (2010): Medicinal plants research in Pakistan. Journal of Medicinal Plants Research 4:161-176.

Shinwari, Z.K. & Qaiser, M. (2011). Efforts on Conservation and Sustainable use of Medicinal Plants of Pakistan. Pakistan Journal of Botany 43:5-10.

Shiri, M.R., Choukan, R. and Aliyev, R.T. (2014): Study of genetic diversity among maize hybrids using SSR markers and morphological traits under two different irrigation conditions. Crop Breeding Journal 4(1):65-72.

Singh, Y. (2009). Systematics of *Hypoxis* (Hypoxidaceae) in southern Africa. Philosophiae Doctor Dissertation, University of Pretoria, Pretoria, South Africa.

Smith-Hall, C., Larsen, H.O., & Pouliot, M. (2012: People, plants and health: a conceptual framework for assessing changes in medicinal plant consumption. Journal of Ethnobiology and Ethnomedicine 8:1-11.

Srivastava, A.K. (2018). Significance of Medicinal Plants in Human Life. In: Synthesis of Medicinal Agents from Plants. Elsevier, 1-24.

Street, R.A. & Prinsloo, G. (2013): Commercially Important Medicinal Plants of South Africa: A Review. Journal of Chemistry 1-16.

Subbarao, G. V., Ito, O., Berry, W. L. & Wheeler, R. M. (2003) Sodium-A Functional Plant Nutrient. Critical Reviews in Plant Sciences 22:391-416.

Sutapa, S., Singh, A.M., Mridul, C., Singh, S.K., Ahlawat, A.K. and Singh, G.P. (2014): Analysis of genetic diversity among the Indian bread wheat cultivars using microsatellite (SSR) markers. Indian Journal of Genetics and Plant Breeding 74:502-505.

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

Thakur, M., Bhattacharya, S., Khosla, P. K., & Puri, S. (2018). Improving production of plant secondary metabolites through biotic and abiotic elicitation. Journal of Applied Research on Medicinal and Aromatic Plants 12:1-12.

Trejo-Téllez, L. I. & Gómez-Merino, F. C. (2012): Nutrient Solutions for Hydroponic Systems, Hydroponics - A Standard Methodology for Plant Biological Researches.

Van Andel, T., Myren, B., & Van Onselen, S. (2012): Ghana's herbal market. Journal of Ethnopharmacology 140:368-378.

Van Wyk, B. E. (2008). A broad review of commercially important southern African medicinal plants. Journal of Ethnopharmacology 119:342-355.

Van Wyk, B. E. (2011). The potential of South African plants in the development of new medicinal products. South African Journal of Botany 77:812-829.

Van Wyk, B. E. (2015). A review of commercially important African medicinal plants. Journal of Ethnopharmacology 176:118-134.

Velasco, L., Mollers, C., and Becker, H.C. (1999): Estimation of seed weight, oil content and fatty acid composition in intact single seeds of rapeseed (*Brassica napus* L.) by near-infrared spectroscopy. Euphytica 106:79-85.

Verma, N., & Shukla, S. (2015). Impact of various factors responsible for fluctuation in plant secondary metabolites. Journal of Applied Research on Medicinal and Aromatic Plants 2:105-113.

Vlk, D. and Řepková, J. (2017): Application of next-generation sequencing in plant breeding. Czech J. Genet. Plant Breeding 53:89−96.

Wang, J.J., Yang, J., Zhu, Q., Jia, and Tao, Y. (2010): Assessment of genetic diversity by simple sequence repeat markers among forty elite varieties in the germplasm for malting barley breeding. Journal of Zhejiang University Science B 11:792-800.

Wang, M., Zheng, Q., Shen, Q., & Guo, S. (2013): The Critical Role of Potassium in Plant Stress Response. International Journal of Molecular Sciences 14: 7370-7390.

Wang, Y., Li, Y., Wu, H., Hu, B., Zheng, J., Zhai, H., Lv, S., Liu, X., Chen, X., Qiu, H., Yang, J., Zong, C., Han, D., Wen, Z., Wang, D. and Xia, Z. (2018): Genotyping of soybean cultivars with medium-density array reveals the population structure and QTNs underlying maturity and seed traits. Frontier Plant Science. 9:610. doi:10.3389/fpls.2018.00610

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

Wenzl, P, Carling, J, Kudrna, D, Jaccoud, D and Huttner, E. (2004): Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences of the United States of America 101:9915–9920.

Williams, V. L., Victor, J. E. & Crouch, N. R. (2013): Red Listed medicinal plants of South Africa: Status, trends, and assessment challenges. South African Journal of Botany 86: 23-35.

Williams, V.L., Raimondo, D., Crouch, N.R., Cunningham, A.B., Scott-Shaw, C.R., Lötter, M. & Ngwenya, A.M. (2016): *Hypoxis hemerocallidea* Fisch, C.A.Mey. & Avé-Lall. National Assessment: Red List of South African Plants version 2017.1. http://redlist.sanbi.org/species.php?species=1603-21 [Accessed May, 2020].

Wittenberg, A.H.J., van der Lee, T., Cayla, C, Kilian, A. and Visser, R.G.F. (2005): Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Molecular Genetics and Genomics 274:30–39.

Xego, S., Kambizi, L., & Nchu, F. (2016). Threatened Medicinal Plants of South Africa: Case of the Family Hyacinthaceae. African Journal of Traditional, Complementary and Alternative Medicines 13:169-180.

Xia, L., Peng, K.M., Yang, S.Y., Wenzl, P and de Vicente, M.C. (2005): DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. Theoretical and Applied Genetics 110:1092–1098.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, J., Wright, M., McCouch, S., Nielsen, R., Wang, J and Wang, W. (2012): Re-sequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature Biotechnology 30:105–111. doi:10.1038/nbt.2050

Yao, G., Duan, J., Zou, L., Li, C., Lu, L. and Huang, M. (2015): Research on genetic diversity of glutinous rice in Hubei by SSR marker. Journal of Henan Agricultural Sciences 44:19-22.

Zhao, Y., Wang, H., Chen, W., Li, Y., Gong, H., Sang, X., Huo, F. and Zeng, F. (2015): Genetic diversity and population structure of elite cotton (*Gossypium hirsutum L*.) germplasm revealed by SSR markers. Plant Systematics and Evolution 301:327-336.

Zimudzi, C. (2014). African Potato (*Hypoxis* Spp): Diversity and Comparison of the Phytochemical Profiles and Cytotoxicity Evaluation of four Zimbabwean Species. Journal of Applied Pharmaceutical Science 4:079-083.

.

CHAPTER THREE

GENETIC DIVERSITY ANALYSIS OF SOYBEAN (*Glycine max* (L.) Merr.) GENOTYPES MAKING USE OF SIMPLE SEQUENCE REPEAT (SSR) MARKERS

ABSTRACT

The aim of the study was to investigate the genetic diversity among 30 soybean genotypes maintained by the ARC using SSR markers. Soybean genotypes were characterized using 20 SSR primers. DNA was extracted using the standard cetyl trimethylammonium bromide method and amplified using polymerase chain reaction (PCR). Allele size was determined via comparison with a 100 base pair (bp) DNA ladder. Molecular data were analyzed, and a dendrogram and matrix were generated using graphical genotyping (GGT) 2.0 software. A total of 216 alleles with an average of 10.8 alleles per locus were detected. The allele sizes ranged between 2 and 33 bp with an average of 18.7 bp. The polymorphic information content among genotypes varied from 0.85 (Satt001) to 0.75 (Satt43) with an average of 0.716, and heterozygosity ranged from 0.87 to 0.78 with an average of 0.7485. The most diverse genotypes were B 66 S 31, 69S 7, and R5-4-2 M, which indicated the efficiency of the SSR markers for the detection of genetic diversity. The results of the current study depicted on the dendrogram revealed the diversity among the soybean genotypes tested, which might aid breeders in the future in the selection of parents for breeding.

Keywords: Allele; base pairs; dendrogram; DNA; PCR

Abbreviations: ARC_ Agricultural Research Council, bp_ base pairs, CTAB_ cetyl trimethylammonium bromide, UPGMA_ unweighted pair group method with arithmetic mean

3.1. Introduction

Soybean (*Glycine max* (L.) Merr.) is one of the oldest cultivated crops. It is relatively the cheapest source of protein and is in considerable demand for food and feed supply. Soybean seeds contain 40-42% protein, 18-20% oil, 11% soluble carbohydrates, and dry matter (Devi et al., 2012). Although the demand for soybean is high, there is a slight decline in the genetic improvement activities of its cultivars. It has a few limitations, such as susceptibility to diseases and pests, adverse environmental conditions, low yield, and poor handling/management. However, all these limitations can be ameliorated in many different ways, including implementing a variety improvement program, cultural practices, post-harvest technology, and field selection.

Genetic diversity plays an important role in the survival and adaptability of plant species. When the practice of farming started, farmers used selective breeding to pass on desirable traits of the crops while omitting the undesirable ones. Studies have recently shown that DNA markers have become an essential tool to implement a soybean improvement program because molecular markers are robust and not affected by the environment (Shoemaker and Specht, 1995). Molecular markers aid plant breeders to indirectly select individuals from different populations that carry a gene for a favorable trait if a tight linkage exists between a marker locus and the genetic locus controlling that trait. Moreover, several molecular maps have been generated with some morphological traits using amplified restriction fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), and restriction fragment length polymorphism (RFLP) markers in soybean (Apuya et al., 1988; Shoemaker and Specht, 1995; Keim et al., 1997; Cregan et al., 1999). Simple sequence repeat markers, also known as microsatellites, are useful, reliable, and easy-to-use markers for soybean characterization Shoemaker and Specht (1995). Simple sequence repeat markers have become important for genetic mapping and genetic diversity determination in soybean because they are co-dominant and spread throughout the genome, exhibit high levels of polymorphism, undergo polymerase chain reaction (PCR) easily, and can be analyzed using gel electrophoresis (Ott et al., 2011). Thousands of SSR markers have been developed

over the years from both expressed sequence tags (ESTs) and genomic DNA, and these add to the available SSRs. Soybean genome sequence have recently been developed from approximately 33,000 putative SSR markers, and some of them have been mapped (Ott et al., 2011). Marker distribution is often associated with gene distribution. To support this association, a study of microsatellites in Arabidopsis, rice, maize, wheat, and soybean clearly showed that microsatellite distribution was much higher in gene-rich, single-copy regions than in repetitive sections of the genome (Ott et al., 2011). A comparison between RFLP and SSR markers in soybean showed that RFLPs tended to be more closely associated with the gene-rich regions, whereas SSR loci were closely associated with the actual genetic sequences. Ott et al. (2011) showed that many major crops exhibited clustered genes in the gene-rich regions on chromosomes, which followed the patterns of clustered markers. The variability and large number of repeat sequences makes them an excellent tool for pedigree analysis, genotype differentiation, particular genotype identification, and genetic distance evaluation among organisms. Simple sequence markers characterize genetic diversity and compare the relatedness of germplasm. They are preferred over other markers because they are highly reproducible, an important aspect in genetic analysis. SSR markers require low starter costs and small quantities of DNA for screening; in addition, 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). Simko et al. (2012) showed that SSRs exhibited high success rates in diversity studies. A study on comparative genetic diversity using SNP, DART, and SSR on 54 sugar beet cultivars showed that the success rate was the highest for SSR markers (Simko et al., 2012) owing to their highly polymorphic nature. According to Guan et al. (2010) and Wang et al. (2010), SSRs have been used successfully in the identification of genetic diversity and relationships among soybean genotypes in different populations. Soybean primers were used to amplify SSRs with a success rate of 65 %, despite a lower rate of 3 to 13 % found outside the subgenus *Glycine* (Peakall et al., 1998). Wang et al. (2010) and Guan et al. (2010) showed that SSRs could be successfully used to identify genetic diversity and relationships among soybean genotypes within a population. According to Vinu et al. (2013), environmental fluctuations did not have a considerable effect on

molecular profiling because of their authenticity and reliability when it comes to breeding. Although there have been studies on soybean diversity using SSR markers, there are limited studies carried out in South Africa to determine the genetic diversity of soybean collected globally using molecular markers. Hence, in this study, we aimed to investigate genetic diversity among 30 soybean genotypes using SSR markers.

3.2. Material and methods

3.2.1 Plant materials and study site

The experiment was conducted at the Agricultural Research Council – Grain Crops, South Africa located at 26°44′43.16″S – 27°04′47.71″E with an altitude of 1340 metres above sea level. Thirty soybean genotypes that are sourced and maintained by the Agricultural Research Council were grown under controlled conditions in a growth chamber until the 4th leaf stage. Day and night temperatures of the growth chamber were kept constant at 29 °C so that growing temperature stayed the same in order to not hinder/delay the growth of the plants, and plants were irrigated to field capacity every fourth day depending on soil moisture depletion. Two seeds of each accession were planted in a 5 litre pot containing locally obtained loamy soil. This procedure was replicated twice for each genotype. The genotypes were coded as illustrated in Table 3.1.

Soybean nutritional quality can be analysed using the DA 7250 (Perten Instruments) laboratory instrument which represents a new generation of NIR analyser. The instrument performs multi-component analysis in 6 seconds and requires little or no sample preparation. In order to achieve the high performance, three tasks that are crucially important are accomplished: analyse representative samples, perform precise measurements and eliminate user errors. Representative samples will be analysed in rotating tray that allows nearly the entire sample surface to be presented to the large analysis area. Any error introduced by variability between sample dishes or issues relating to clean up are removed by the sample presentation principles (User Manual DA 7250, 2013). DA 7250 is a versatile instrument that can analyse

many different products like powders, meals, whole grain, pellets, finished goods, slurries, pasts and liquids. In the present study, whole seeds of soybean genotypes will be analysed according to the following nutritional quality traits: moisture, ash, fibre, linoleic acid, linolenic acid, palmitic acid, stearic acid, and oleic acid (User Manual DA 7250, 2013).

Table 3.1: A list of accessions obtained from the Potchefstroom Agricultural Research Council-Grain Crops gene bank used in the study.

No.	Code	Origin/place of collection	Accession name	No.	Codes	Origin/place of collection	Accession name
1	434-001	Unknown (ARC-GCI)	69 S 7	16	434-016	Zimbabwe	Oribi
2	434-002	Unknown (ARC-GCI)	B 66 S 31	17	434-017	Unknown (ARC-GCI)	ND 85
3	434-003	Unknown (ARC-GCI)	Lee Ex RHOD	18	434-018	USA	AGS 239
4	434-004	China	Columbia M 8 A	19	434-019	Unknown (ARC-GCI)	61 S 156
5	434-005	South Africa (ARC-GCI)	IBIS 2000	20	434-020	Brazil	Santa Rosa
6	434-006	USA	R-5-4-2 M	21	434-021	Unknown (ARC-GCI)	B 66 S 41
7	434-007	South Africa (ARC-GCI)	Egret	22	434-022	South Africa (ARC-GCI)	Egret
8	434-008	Unknown (ARC-GCI)	15/06/2012	23	434-023	Unknown (ARC-GCI)	Crawford
9	434-009	South Africa (ARC-GCI)	Dundee	24	434-024	Unknown (ARC-GCI)	B 66 S 37
10	434-010	Unknown (ARC-GCI)	Solar 12	25	434-025	Unknown (ARC-GCI)	B 66 S 387
11	434-011	USA	Hawkeye	26	434-026	Unknown (ARC-GCI)	B 66 S 24
12	434-012	USA	N69-2774	27	434-027	Unknown (ARC-GCI)	B 66 S 256
13	434-013	Asia	Maksura	28	434-028	USA	Kahala
14	434-014	USA	DB 1601	29	434-029	Unknown (ARC-GCI)	B 66 S 8
15	434-015	USA	Yeluanda	30	434-030	USA	PR 165-52

DNA extraction and PCR analysis

At the fourth leaf stage, leaves were harvested and kept in 1.5-mL polypropylene sample tubes prior to freeze-drying for three days. Freeze-dried leaf samples of each genotype were grounded into powder using a Qiagen tissue lyser grinder (Rogstad, 2003). DNA was extracted using the standard cetyl trimethylammonium bromide (CTAB) method (Saghai-Maroof et al., 1984). DNA quality was determined using 1 % agarose, and DNA concentration was quantified using a NanoDrop machine. DNA samples were diluted to a final concentration of 50 ng/µL using Tris-EDTA (TE) buffer. DNA was amplified by PCR using a Gene Amp PCR system 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA). Twenty primers were used for diversity analysis of the soybean genotypes. Amplified products were electrophoresed on 3.5 % agarose gel and detected using ethidium bromide staining. The primers and their sequences are shown in Table 3.2. The alleles were quantitatively scored as present or absent. Allele sizes were estimated in comparison with a 100 bp DNA ladder.

Table 3.2. Description of the 20 SSR markers used to analyze diversity of soybean genotypes (http://www.soybase.org).

Marker	Forward primer (5'→3)	Reverse Primer (3'→5)	Repeat Type
Satt001	AAAGTCTTTAAAAGTGTGTCTTA	TTAAAAGAAAAATGCAACAT	(ATT)2
Satt173	TGC GCC ATT TAT TCT TCA	AAGCGAAATCACCTCCTCT	(ATT)18
S45035	TTTGTGAACGATAGAAATTTAT	AGGGGAAAATTTTTAAAGA	
Satt373	TCCGCGAGATAAATTCGTAAAAT	GGCCAGATACCCAAGTTGTACTTGT	(ATT)21
SOYPRP1	CGTGCCAAATTACATCA	TGATGGGAACAAGTACATAA	(TAT)20
Satt534	CTCCTCCTGCGCAACAACAATA	GGGGGATCTAGGCCATGAC	(ATT)25
Satt005	TATCCTAGAGAAGAACTAAAAAA	GTCGATTAGGCTTGAAATA	(ATT)19
Satt242	GCGTTGATCAGGTCGATTTTTATTTGT	GCGAGTGCCAACTAACTACTTTTATGA	(ATT)26
Satt009	CCAACTTGAAATTACTAGAGAAA	CTTACTAGCGTATTAACCCTT	(ATT)14
Sat001	GCGGATACGACCAAAAATTGTT	GCGAACTGCGAAGATACTACCC	(AT)17
BarcSatt100	ACCTCATTTTGGCATAAA	TTGGAAAACAAGTAATAATAACA	(ATT)33
SATT9	ATTACTAGAGAAATTAGTTTA	CTTACTAGCGTATTAACCCTT	(AAT)12
Satt038	GGGAATCTTTTTTTCTTTCTATTAAGTT	GGGCATTGAAATGGTTTTAGTCA	(ATT)17
Satt002	TGTGGGTAAAATAGATAAAAAT	TCATTTTGAATCGTTGAA	(ATT)25
GMSE0634	TGGGTAGGTTTTTCAGCAATG	GCAAAGGGACCCAAAGGTAT	
Sat417	GCGAATATGGCGTTGAAAATAGTGAT	GCGACCCAGATTCTGTGCTAAGA	(AT)16
Satt156	GCGGTGTGGATCCAAAAACTCAAACTT	GCGTGCTAGTTCGATCAGCTTAGTTTC	(AT)17
Satt36	AAAGTCATAAGTGGCACTCCAAGTTT	GAACATAACAATAATAAATATAGCTC	(AT) I9
SAT1	CTGGTGGACTATTGATACGACC	AACTGCGAAGATACTACCCTCC	(AT)17
Satt43	AAATTCTGTTCATTGTCCGTC	CATTTTAATATCCCGAGTAGG	(AT)20

3.3. Data analysis

Data were captured using the GGT 2.0 software (Van Berloo, 2007). The resulting fragments were analyzed, and the alleles were scored using GeneMapper® version 4.1 software (Applied Biosystems). A dissimilarity matrix was generated using DARwin 5.0 software (Perrier and Jacquemoud-Collet, 2006). The data matrices of the genetic distances were used to create the dendrogram using the UPGMA algorithm. The assay efficiency index, referred to as the PIC value, was calculated using the formula PIC = 1 - ∑fi, where fi is the frequency of the allele (Smith et al., 1997).

3.4. Results and discussion

3.4.1. Polymorphism and allelic diversity of SSR markers

Table 3.3 summarizes the number of alleles, size ranges, polymorphic information content (PIC) values, and heterozygosity. SSR markers used in this study generated 216 alleles among the 30 soybean genotypes with an average of 10.8 alleles per locus. The markers also revealed marked genetic diversity among soybean genotypes. Four markers (Satt156, Satt001, Satt36, and SAT1) generated 17 alleles each. The allele size ranged from 2 to 33 base pairs (bp) with an average value of 18.7 bp. Hipparagi et al. (2017) analyzed 75 genotypes using 21 SSR markers and reported 60 alleles with an average of 2.85 alleles per locus. In another study, 38 soybean genotypes were analyzed using 16 SSR markers resulting in 51 alleles with an average of 2.22 alleles per locus (Bisen et al., 2015). In addition, Li et al. (2008) reported 19.7 alleles per locus upon characterizing 1863 genotypes using 59 SSR markers, and a genetic diversity analysis of 205 Chinese landraces yielded 16.2 alleles per locus (Guan et al., 2010). The allele number detected in this study is higher than those reported in the studies above. The great number of alleles generated by SSR markers suggested allelic richness, a useful indicator of genetic worthiness for subsequent selection and conservation strategies (Wang et al., 2006).

The PIC values ranged from 0.46 (GMSE0634) to 0.85 (Satt001) with an average of 0.716. High PIC values suggested a potential to detect differences among soybean

genotypes. Hipparagi et al. (2017) showed an average PIC value of 0.36 with the markers, Sat554, Sat180, Sat600, and Sat478 having 4 alleles per locus each. The PIC values ranged from 0.55 to 0.66. Various studies showed PIC values ranging from 0.199 to 0.87 (Wang et al., 2006; Hisano et al., 2007; Zhang et al., 2013; Kim et al., 2014; Bisen et al., 2015). The results of this study are in line with the findings of other studies. Genotype GMSE0634 (0.5) had the lowest heterozygosity value, whereas Satt001 (0.87) had the highest value. The average heterozygosity value obtained was 0.7485 and were much higher than results reported values on soybean (Wang et al., 2006; Li et al., 2008; Liu et al., 2010; Wang et al., 2010; Zhang et al., 2013; Hipparagi et al., 2017).

Table 3.3. Genetic information generated by 20 SSR markers on 30 soybean genotypes.

Marker	Allele number	Allele size (bp)	PIC	Heterozygosity
Satt001	10	2	0.85	0.87
Satt173	16	18	0.79	0.81
S45035	5	-	0.51	0.59
Satt373	16	21	0.80	0.82
SOYPRP1	4	20	0.65	0.70
Satt534	9	25	0.72	0.75
Satt005	11	19	0.81	0.83
Satt242	7	26	0.73	0.76
Satt009	6	14	0.65	0.70
Sat001	19	17	0.83	0.85
BarcSatt100	6	33	0.80	0.82
SATT9	5	12	0.51	0.54
Satt038	6	17	0.61	0.66
Satt002	7	25	0.71	0.75
GMSE0634	4	-	0.46	0.50
Sat417	13	16	0.72	0.76
Satt156	22	17	0.80	0.82
Satt36	17	19	0.82	0.84
SAT1	17	17	0.80	0.82
Satt43	16	20	0.75	0.78
Average	10.8	18.7	0.72	0.749

**PIC – Polymorphic Information Content*

** bp – base pairs*

3.4.2. Cluster analysis

Figure 3.1 shows a dendrogram for the 30 soybean genotypes constructed using the unweighted pair group method with arithmetic mean (UPGMA) clustering algorithm based on 20 SSR markers. Clustering analysis aids to substantiate the results of pairwise analysis. The dendrogram divided the cultivars into two major clusters, 2.6 (a) and 0.12 (b). The closest distance (similarity) was between genotypes Santa Rosa (020) and PR 165-52 (030), whereas the greatest genetic distance (dissimilarity) was between genotypes 69 S 7 (001) and Yeluanda (015), as well as R-5-4-2 M (006) and B 66 S 31 (002).

The two major clusters were further divided into two sub-clusters, 0.16 and 0.33. Sub-cluster 0.16 consisted of two sub-sub clusters, which were 2.3, having one genotype (R-5-4-2 M), and 0.10 consisting of 16 genotypes (Lee Ex RHOD (003), Oribi (016), Yeluanda (015), B 66 S 37 (024), B 66 S 41 (021), 61 S 156 (019), B 66 S 24 (026), 15/06/2012 (008), Dundee (009), Kahala (028), Egret (022), Crawford (023), Columbia M 8 A (004), B 66 S 387 (025), PR 165-52 (030), and Santa Rosa (020)). The dendrogram also showed sub-sub clusters 0.28 and 0.15, which were formed from sub-cluster 0.33.

The first sub-sub cluster comprised five genotypes (IBIS (005), B 66 S 256 (027), Solar 12 (010), Egret (007), and ND 85 (017)). In addition, sub-sub cluster 0.15 had seven genotypes (69 S 7 (001), AGS 239 (018), B 66 S 8 (029), DB 1601 (014), Hawkeye-USSR (011), Maksura (013), and N69-2774 (012). Recently, Bisen et al. (2015) reported two major clusters, which were further divided into two sub-groups in the analysis of 38 soybean genotypes using 16 SSR markers. In addition, Hipparagi et al. (2017) found three distinct clusters in 75 genotypes using 21 SSR markers, and Hirota et al. (2012) reported two distinct clusters.

Moreover, Tantasawat et al. (2011) found four major clusters in 25 soybean genotypes using SSR markers, whereas Wang et al. (2006) and Ghosh et al. (2014) observed two major clusters. A previous study on revolutionary relationship between *Glycine soja* and *Glycine max* revealed two clusters. The results of the present study showed similar findings as those of Wang et al. (2006), Wen et al. (2008), Hirota et al. (2012), Ghosh et al. (2014), and Bisen et al. (2015). The genotypes used in this study might

have showed genetic variation due to different sources and/or area from which they were collected; hence, they belonged to different clusters. Farmers usually concentrate on mass production rather than maintaining gene purity, resulting in the production of new lines.

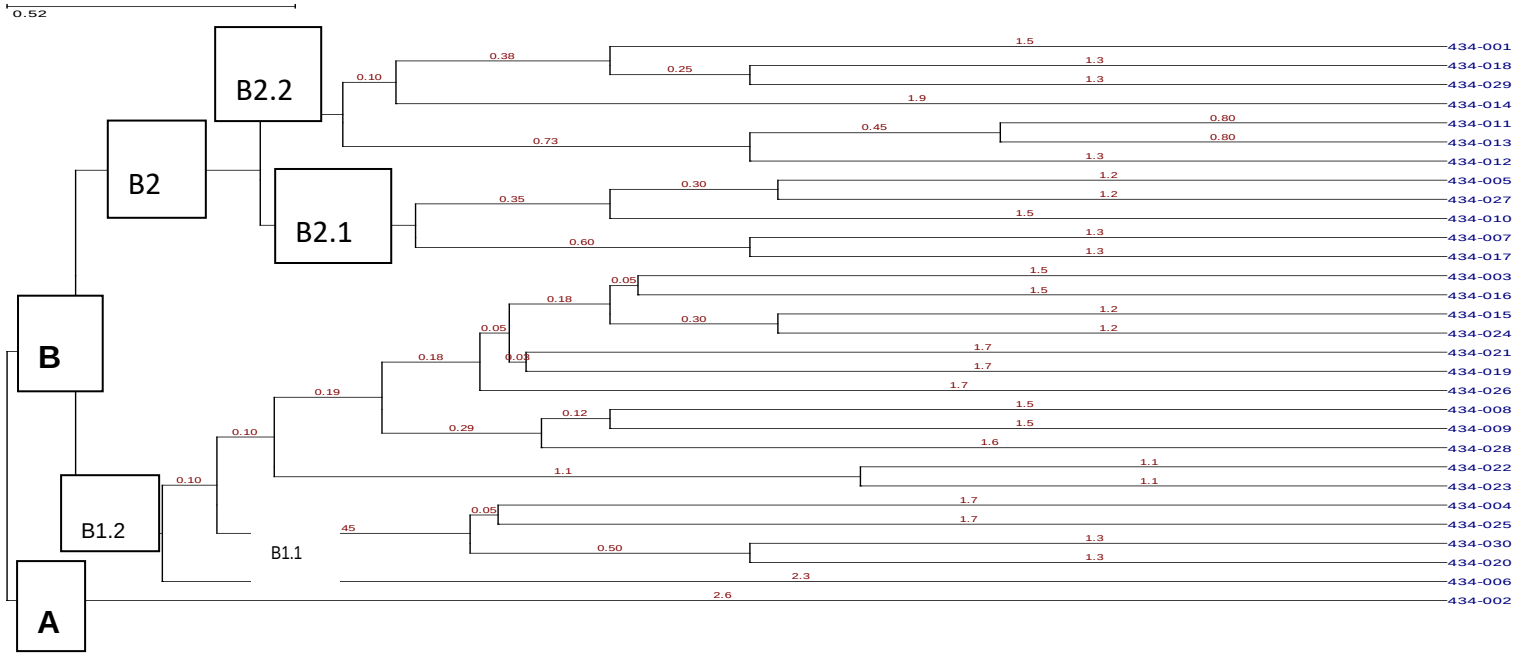


Figure 3.1: Dendrogram showing genetic relationships among 30 soybean genotypes evaluated using 20 SSR markers.

Various studies have assessed genetic diversity in soybean using SSR markers. At the Embrapa Research Institute in Brazil, Mulato et al. (2010) used SSR markers to assess genetic relationships between soybean cultivars. Twenty SSR markers and 10 EST SSR markers were used to analyze 79 soybean accessions, and results revealed high levels of genetic diversity among them. A total of 259 alleles were obtained, ranging from 2 to 21 alleles per locus, with an average of 8.63. The genotypes were assigned to five major clusters and numerous subgroups. In addition, Wang et al. (2006) assessed the genetic diversity of 129 accessions from the Chinese core collection by using 60 SSR markers, which suggested that the material was quite divergent. A total of 732 alleles were observed, and the PIC values varied from 0.05 to 0.91 with a mean of 0.23. These accessions were divided into five major clusters according to geographical origin (two clusters from the Northern ecotypes, one cluster from the Yellow River Valley ecotypes, one cluster from a mixture of the Yellow River Valley and Northern ecotypes, and one cluster from the Southern ecotypes). The accessions obtained from the Yellow River Valley exhibited the highest allelic richness and were simultaneously highly dispersed in their clustering pattern (Wang et al., 2006). Li et al. (2008) reported 1,160 alleles among 1,863 landraces using 59 SSR markers and found seven clusters collected from the Yellow River Valley with a high genetic variation. These findings provided evidence that the Yellow River Valley might be the origin of the cultivated soybean. In Japan, a study was carried out to compare between 1,305 wild soybean and 53 cultivated soybean using 20 SSR markers, and 28 and 5 alleles per locus were obtained for wild and cultivated soybean, respectively, indicating that cultivated soybean had less polymorphism than its progenitor (Li et al., 2008).

According to Jeffreys et al. (1994), SSR markers that produce very high allelic variations are also highly polymorphic. The use of SSRs and other molecular markers in germplasm diversity studies was investigated by Park et al. (2009), who concluded that SSRs exhibited the highest heterozygosity and genetic variation. Among 61 genotypes studied, 1 to 37 alleles were also observed. The genetic diversity of 303 accessions of *Glycine max* was analyzed using 99 SSR markers, which showed high gene diversity of 0.77 (Li et al., 2008). High levels of heterozygosity suggest a high proportion of genetic diversity, which increases selection response in breeding

programs. The high level of heterozygosity observed among the genotypes used in this study indicated that they were collected from various geographical areas with different levels of selection pressure. Farmers maintain a large number of landraces on a single plot to cope with the diverse environmental conditions, resulting in a continuous exchange of genes through pollen flow (Manzelli et al., 2007; Barnaud et al., 2008). In addition, farmers exchange seeds as gifts and *via* markets to renew old seed stocks or to acquire new varieties. Consequently, there may be a continuous exchange of genes among genotypes (Mofokeng et al., 2014).

3.5. Conclusion

Analysis of 30 soybean genotypes using 20 SSR markers revealed that there was genetic diversity among these genotypes. Two major clusters and two sub-clusters were detected. The first sub-cluster comprised 17 genotypes that were closely related, whereas the second sub-cluster included 12 genotypes that also exhibited close genetic relation. Assessment of genetic diversity is important for efficient management and protection of available genetic variability, as well as for crop improvement. The preferred method for breeding is molecular profiling because this method is authentic, reliable, and less affected by environmental changes.

Diversity studies play a major role in categorizing the population into diverse groups, which results in the development of gene pool. The SSRs are multi-allelic and generally more informative than most of the marker techniques, and are based on heterozygosity values (Powell et al., 1996). They serve as an ideal marker system for genetic analysis. Hence, this technique has been used widely in genetic diversity studies of various crops including soybean (Mofokeng et al., 2014). The SSR reveals a large number of polymorphisms. Hence, they can be used to study plant species in which previous methods have found little or no variation (Echt et al., 1998). The technique is inexpensive once the primers have been developed. It is repeatable, easily automated, requires small quantities of DNA, and gel runs can be multiplexed and not influenced by environmental conditions.

Phenotypic characterization and evaluation of soybean genotypes are dependent on records of qualitative or quantitative morphological traits (Upadhyaya et al., 2010). Characterization of the accessions gives an overview of the traits and helps to understand similarities and differences among the accessions under investigation (IBPGR and ICRISAT, 1993). The grouping of accessions was not based on the source of collection, hence were mixed. This shows the presence of genetic diversity among the accession within and among the South African provinces and/or globally. It may also be due to gene flow from the neighboring areas/provinces and sharing of seeds by farmers amongst themselves (Manzelli et al., 2005). Moreover, farmers share seeds and name the same accessions differently in various areas or regions (Chakauya et al., 2006). Farmer's practices may also influence the handling and conservation of the genetic material on their fields.

3.6. References

Apuya, R., Frazier, N.L., Keim, B., Jill, P., Roth, E. and Lark, K. (1988): Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (*L*.) Merr. Theor Appl Genet. 75:889-901.

Barnaud, A., Trigueros, G., McKey, D. and Joly, H.I. (2008): High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? Heredity. 101:445-452.

Bisen, A., Khare, D., Nair, P. and Tripathi, N. (2015): SSR analysis of 38 genotypes of soybean (*Glycine Max* (L.) Merr.) genetic diversity in India. Physiol Mol Biol Plants. 21:109-115.

Chakauya, E., Tongoona, P., Matibiri, E.A., and Grum, M. (2006): Genetic diversity assessment of sorghum landraces in Zimbabwe using microsatellites and indigenous local names. International Journal of Botany 2:29-35.

Cregan, P.B., Jarvik, T. and Bush, A.L. (1999): An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Devi, K.N., Singh, L.N.K., Singh, M.S., Singh, S.B. and Singh, K.K. (2012): Influence of sulphur and boron fertilization on yield, quality, nutrient uptake and economics of soybean (*Glycine max*) under upland conditions. J Agric Sci. 4:1-10.

Echt, C.S., DeVerno, L.L., Anzidei, M. and Vendramin, G.G. (1998): Chloroplast microsatellites reveal population genetic diversity in red pine, Pinus resinosa Ait. Molecular Ecology 7:1633-1643

Ghosh, J., Ghosh, P.D. and Choudhury, P.R. (2014): An assessment of genetic relatedness between soybean [*Glycine max* (*L*.) Merrill] cultivars using SSR markers. Am J Plant Sci. 5:3089-3096.

Guan, R., Chang, R., Li, Y., Wang, L., Liu, Z. and Qiu, L. (2010): Genetic diversity comparison between Chinese and Japanese soybeans (*Glycine max* (*L*.) Merr.) revealed by nuclear SSRs. Genet Resour Crop Evol. 57:229-242.

Hipparagi, Y., Singh, R., Choudhary, D.R. and Gupta, V. (2017): Genetic diversity and population structure analysis of Kala bhat (*Glycine max* (*L*.) Merrill) genotypes using SSR markers. Hereditas. 154:9.

Hirota, T., Sayama, T., Yamasaki, M., Sasama, H., Sugimoto, T., Ishimoto, M. and Yoshida, S. (2012): Diversity and population structure of black soybean landraces originating from Tanba and neighboring regions. Breed Sci. 61:593-601.

Hisano, H., Sato, S. and Isobe, S. (2007): Characterisation of soybean genome using EST-derived microsatellite markers. DNA Res. 14:271-281.

IBPGR and ICRISAT. (1993): Descriptors for Sorghum [Sorghum bicolor (L) Moench]. International Board for Plant Genetic Resources. Rome, Italy. International Crop Research Institute for the Semi- Arid Tropics, Patancheru, India.

Jeffreys, A.J., Tamaki, K. and MacLeod, A. (1994): Complex gene conversion events in germ line mutations at human mini satellites. Nat Genet. 6:136-145.

Keim, P., Schupp, J.M., Travis, S.E., Clayton, K., Zhu, T., Shi, L., Ferreira, A. and Webb, D.M. (1997): A high-density soybean genetic map based on AFLP markers. Crop Sci. 37:537-543.

Kim, K.S., Chirumamilla, A., Hill, C.B., Hartman, G.L. and Diers, B.W. (2014): Identification and molecular mapping of two soybean aphid resistance genes in soybean PI 587732. Theor Appl Genet. 127:1251-1259.

Li, Y., Guan, R., Liu, Z., Ma, Y. and Wang, L. (2008): Genetic structure and diversity of cultivated soybean [*Glycine max* (*L*.) Merrill] landraces in China. Theor. Appl. Genet. 117:857-871.

Liu, Y.L., Li, Y.H., Zhou, G.A., Uzokwe, N., Chang, R.Z., Chen, S.Y. and Qui, L.J. (2010): Development of soybean EST-SSR markers and their use to assess genetic diversity in the Subgenus soja. Agr Sci China. 9:1423-1429.

Manzelli, M., Benedettelli, S. and Vecchio, V. (2005): Agricultural biodiversity in Northwest Somalia–an assessment among selected Somali sorghum (Sorghum bicolor (L.) Moench) germplasm. Biodiversity and Conservation 14:3381-3392.

Manzelli, M., Pileri, L., Lacerenza, N., Bendettelli, S. and Vecchio, V. (2007): Genetic diversity assessment in Somalia sorghum (*Sorghum bicolor L. Moench*) accessions using microsatellite markers. Biodiversity Conserv. 16:1715-1730.

Mofokeng, M.A., Shimelis, H., Toongona, P. and Laing, M. (2014): A genetic diversity analysis of South African sorghum genotypes using SSR markers. South African Journal of Plant and Soil 31:145-152.

Mulato, B.M., Möller, M., Zucchi, M.I., Quecini, V. and Pinheiro, J.B. (2010): Genetic diversity in soybean germplasm identified by SSR and EST-SSR markers. Pesqui Agropecu Bras. 45:276-283.

North Carolina Soybean Producers Association (2018): History of soybeans (online). Available at http://ncsoy.org/media-resources/history-of-soybeans/ (Accessed 18 August 2018).

Ott, A., Trautschold, B. and Sandhu, D. (2011): Using microsatellites to understand the physical distribution of recombination on soybean chromosomes. PLoS One. 6:e22306.

Park, Y.J., Lee, J.K. and Kim, N.S. (2009): Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules. 14:4546-4569.

Peakall, R., Gilmore, S., Keys, W., Morgante, M. and Rafalski, A. (1998): Cross-species amplification of soybean (*Glycine max*) simple sequence repeats

(SSRs) within the genus and other legume genera: Implications for transferability of SSRs in plants. Mol. Biol. Evol. 15:1275-1287.

Perrier, X. and Jacquemoud-Collet, J.P. (2006): DARwin software. Available at http://www.darwin.cirad.fr/darwin.html [accessed 10 January 2018].

Powell, W., Machray, G. and Provan. J. (1996): Polymorphism revealed by simple sequence repeats. Trends in Plant Science 1:215-222.

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

Rogstad, S.H. (2003): Plant DNA extraction using silica. Plant Mol. Biol. Report. 21:463.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. and Allard, R.W. (1984): Ribosomal DNAsepacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U S A. 81:8014-8019.

Shoemaker, R.C. and Specht, J.E. (1995): Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35:436-446.

Simko, I., Eujayl, I. and van Hintum, T.J. (2012): Empirical evaluation of DArT, SNP and SSR marker- systems for genotyping, clustering and assigning sugar beet hybrid varieties into populations. Plant Sci. 184:54-62.

Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.H., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S. and Ziegle, J. (1997): An evaluation of the utility of SSR loci as molecular makers in maize (*Zea mays L*.): comparison with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173.

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

Upadhyaya, H.D., Sharma, S., Ramulu, B., Bhattacharjee, R., Gowda, C.L.L., Reddy, V.G. and Singh, S. (2010): Variation for qualitative and quantitative traits and identification of trait-specific sources in new sorghum germplasm. Crop and Pasture Science 61:609- 618.

User Manual DA 7250. (2013): Hägersten, Sweden.

Van Berloo, R. (2007): GGT graphical genotypes. Laboratory of plant breeding. Wageningen University. The Netherlands.

Vinu, V., Singh, N., Vasudev, S., Yadava, D.K., Kumar, S., Naresh, S., Bhat, S.R. and Prabhu, K.V. (2013): Assessment of genetic diversity in Brassica juncea (*Brassicaceae*) genotypes using phenotypic differences and SSR markers. Rev. Biol. Trop. 61:1919-1934.

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

Wang, M., Li, R.Z., Yang, W.M. and Du, W.J. (2010): Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. Afr. J Biotechnol. 9:4857-4866.

Wen, Z.X., Zhao, T.J., Zheng, Y.Z., Liu, S.H., Wang, C.E., Wang, F. and Gai, J.Y. (2008): Association analysis of agronomic and quality traits with SSR markers in *Glycine max* and *Glycine soja* in China: I. Population structure and associated markers. Acta Agron. Sin. 34:1169-1178.

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

CHAPTER FOUR

ASSESSMENT OF GENETIC DIVERSITY AMONG SOYBEAN (*Glycine max* (L.) Merr.) GENOTYPES MAKING USE OF AGRO-MORPHOLOGICAL AND NUTRITIONAL QUALITY TRAITS

ABSTRACT

The aim of this study was to assess 30 soybean genotypes and their genetic diversity using agronomic and quality parameters. The experiment was laid out in a randomized complete block design with three replicates in Potchefstroom. Agro-morphological traits were measured using soybean standard descriptor list. Quality traits were measured using near infrared spectroscopy, DA 7250. The results showed significant ($P<0.001$) differences on days to flowering, where Oribi, Crawford, BD 1601, Egret, B 66 S 41, B 66 S 387, Dundee, B 66 S 37 and Santa Rosa that took shorter days to flower, whereas B 66 S 256, Solar 12 and B 66 S 8 took longer to flower. Yeluanda, 15/06/2012, B 66 S 31 had the highest fibre and height. DB 1601 had the highest number of seeds per pod. Oribi had the highest number of branches with PR 165-52 and B 66 S 24 having the highest yields. Number of branches per plant and pod weight before threshing were highly significant and highly correlated with plant height, days to 50% flowering, nodes at flowering, pod length as well as oil, fiber, oleic acid and moisture acid, linolenic acid, palmitic acid, stearic acid and ash. Dundee and N69-2774 were associated with high oil content while Ex RHOD, R 5-4-2 M, N69-2774 and DB 1601 were associated with both ash and stearic acid. Crawford, Egret and B 66-387 accessions were associated with palmitic acid. B 66 S 8 had the highest oleic acid content. PR 165-52 and B 66 S 8 contained the largest oil content. Principal component (PC1) accounted for 21.28% while principal component (PC2) accounted for 14.99% of the total variation. The multivariate analysis helped us understand important agronomic and genetic traits when selecting soybean genotypes.

Keywords: Agro-morphological diversity; genetic diversity, multivariate analysis quality, variation.**Abbreviations:** PCA=principal component analysis

4.1. Introduction

Genetic diversity based on agronomic traits is one of the oldest and most exploited methods. It holds the advantage of providing a direct, simple, rapid and inexpensive way of characterising varieties (Mutengwa, 2004). Phenotypic characterisation is generally viewed as the best determinant of taxonomic classification and agronomic value of crop plants and this makes it the most classical approach when coming to characterisation (Cholastova and Knotova, 2012). Phenotypic characterisation is also important for processes such as development, production and marketing of varieties. Govindarao (2010) stated that it is important to register a new cultivar that is distinct from other existing cultivars in markets where plant breeder's rights exist due to the fact that different cultivars could be identified based on phenotypic descriptors.

The phenotypic characters that are commonly used to assess genetic variability in soybean include plant morphology, seedling, seed quality and seed 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), Marconato et al. (2016), Ghanbarri et al. (2018), Zeffa et al. (2019) and Ibidunni et al. (2020)). A study was conducted by Hamzekhanlu et al. (2011) studied 34 mutant lines including one control cultivar and detecting variability for number of grains per plant, number of pods per plant, number of leaves per plant, number of nodules per plant, nodule dry weight, 100 seed weight, plant dry weight (shoot dry weight), root dry weight, harvest index and seed yield per plant. The genotypes were then clustered into four groups.

Furthermore, Manjaya and Bapat (2008) carried out a study observing the genetic variation during the characterisation of 55 soybean varieties using phenotypic traits that included: 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. Antalikova et al. (2008) used 52 morphological and agronomic characters to find 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 quite a number of significant differences among all the traits that were assessed.

The Asian Vegetable Research and Development Centre (AVRDC), United States and Pakistan studied the genetic diversity of 92 soybean genotypes and they found high coefficient of variations (CVs) coupled with wide ranges on leaf area (44.8%), number of branches per plant (31.7%), pods per plant (29.5%), 100 seed weight (39.0%) and grain yield per plant (46.6%) (Malik et al., 2011). The findings showed a high level of diversity among the studied genotypes. Interestingly, the genotypes were classified into three distinct groups with the Pakistan germplasm forming its own cluster. Ojo et al. (2012) performed a similar study on 42 genotypes and found seven clusters. The study revealed that the 100 seed weight, number of pods per plant, pod yield per plant and seed yield per plot accounted for the greatest phenotypic variation, implying that there was broad diversity.

Phenotypic traits play a very important role in crop improvement and genetic diversity studies although they may be altered or influenced by the environment. Selection based on these traits is still widely practiced and will continue to play a significant role in estimating diversity among various genotypes using ANOVA in crop research. Results exhibiting high CVs and significant differences present high scope for selection (Malik et al., 2011). 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. The objectives of this study were to determine the presence of genetic diversity among the soybean genotypes using agro-morphological and nutritional quality traits.

4.2. Material and methods

4.2.1 Study site

The study site was described in Chapter 3.2.1 and Table 4.1 shows the metereological data during the study period (October 2017 to May 2018). Whole soybean seed was analysed for nutritional quality using the DA 7250 (Perten Instruments) which checked for the following nutritional quality traits: moisture, ash, fibre, linoleic acid, linolenic acid, palmitic acid, stearic acid, and oleic acid (User Manual DA 7250, 2013).

4.2.2 Experimental layout and management

This study was conducted during the 2017/2018 growing season. The field has a well drained sandy loam soil that had a pH ranging between 5.3 and 5.5. Selective pre-emergence herbicide was applied immediately after planting, subsequently; post-emergence herbicide was applied 31 days after sowing. Thirty soybean genotypes were planted in single rows of 3 m, with intra-row and inter-row spacing of 75 cm and 10 cm, respectively using randomized complete block design, replicated three times. Plants were irrigated to field capacity once or twice a week using sprinklers depending on the soil moisture.

4.3 Data collection

Data were collected on the 4th of January 2018 according to the Standard Key Descriptor Lists for Characterizations for soybean (IBPGR, 1984).

Agronomic attributes

The following agronomic parameters were measured according to soybean descriptor list (IBPGR, 1984):

i. Days to 50 percent flowering

This was calculated as days from planting to when 50 percent of the plants in each plot have flowered.

ii. Days to maturity

This was calculated as days from planting to the day when 90 percent of the pods within the plot have dried.

iii. Plant height

This was measured from the ground surface to the tip of the growing point using meter ruler over five randomly selected plants at maturity and recorded in centimetres.

iv. Number of pod per plant

All the pods per plant were harvested, counted and averaged over three plants.

v. Pod length

The lengths of three randomly selected pods per line were measured using a ruler and average length per pod expressed in centimetres.

vi. Number of seeds per pod

The total number of seed in each pod was counted and averaged over three pods.

vii. Hundred seed weight

Hundred randomly selected good seeds were counted and weighed in grams using a digital scale.

viii. Seed weight

After threshing the dried pods from each net plot, the seeds were weighed using digital weighing scale and expressed in grams.

ix. Nodes at flowering

Number of nodes at 50% flowering stage.

x. Branch number per plant

Number of branches counted from the main stem per plant.

Qualitative attributes

The following qualitative traits were measured according to soybean descriptor list (IBPGR, 1984):

a) *Flower colour was recorded using the Munsell Colour Chart using the following*

1 = White, 2 = Yellow, 3 = Red, 4 = Purple

b) *Stem types were recorded as follows;*

3 = Determinate; 5 = Semi-determinate; 7 = Indeterminate

c) *Leaflet shape was recorded as follows;*

3 = Narrow; 5 = Intermediate; 7 = Broad.

d) *Pubescence was recorded*
As either “present” or “absent” on the pods

e) *Pubescence density was recorded as follows;*

3 = Sparse; 5 = Semi-sparse; 7 = Normal; 9 = Dense (focusing mainly on the stem/leaves).

f) *Pubescence colour were recorded as follows;*

1 = Grey; 2 = Light brown; 3 = Brown (tawny)

g) *Pubescence type was recorded as either*

1 = Erect; 2 = Semi-appressed; 3 = Appressed; 4 = Curly; 5 = Retrorse tip.

h) *Corolla colour was recorded as follows;*

1 = White; 2 = Purple throat; 3 = Purple. Data was collected from 3 to 5 randomly selected plants within each row.

4.4 Data analysis

Analysis of variance was conducted using PROC GLIMMIX, SAS version 9.4 (PROC GLIMMIX SAS Institute 2013). The Tukey's Student Range Test was used to separate means that were significantly different at P=0.05 as described by Steel & Torrie (1980). The principal component analysis (PCA) based on the linear correlation between variables and loading factors was used in multivariate analysis. Agromorphological data were subjected to multivariate data analysis using principal component analysis (PCA – XLSTAT, 2015) to identify and evaluate the groupings between the variables following the description.

Table 4.1 Potchefstroom weather condition during the study period October 2017-June 2018.

Year	Month	Day	Tx	Tn	RHx	RHn	Rs	U2	Rain	ET0
2017	10	Average	26,39	11,36	82,77	27,21	20,22	1,75	1,81	4,13
2017	10	Total	659,82	284,04	2,069.27	680,21	626,95	54,36	56,13	127,94
2017	10	Highest	32,4	17,12	94,02	59,64	30,36	4,29	23,88	6,42
2017	10	Lowest	18,02	5,12	44,21	11,38	8,44	0,79	0	1,62
2017	11	Average	29,12	12,67	78,14	20,46	26,99	1,87	2,31	5,52
2017	11	Total	873,55	380,24	2,344.25	613,84	809,58	56,15	69,34	165,59
2017	11	Highest	34,45	17,29	94,06	50,48	33,49	3,3	18,54	7,03
2017	11	Lowest	17,09	4,53	42,76	8,4	11,35	0,64	0	2,11
2017	12	Average	29,29	15,69	84,56	29,15	24,58	2,02	2,02	5,25
2017	12	Total	908,14	486,26	2,621.50	903,71	761,87	62,69	62,48	162,74
2017	12	Highest	33,39	19,06	96,49	84,34	33,79	4	13,72	7,37
2017	12	Lowest	15,84	10,32	67,17	8,95	3,46	0,54	0	0,81
2018	1	Average	31,04	16,09	76,81	23,64	25,83	1,62	1,52	5,55
2018	1	Total	962,36	498,85	2,381.13	732,82	800,71	50,19	47,24	172,01
2018	1	Highest	36,59	20,25	95,13	50,02	33,25	2,53	12,45	7,15
2018	1	Lowest	24,38	9,31	50,35	9,18	11,75	1,03	0	2,59
2018	2	Average	27,68	15,64	90,59	39,67	19,05	1,38	2,44	3,95
2018	2	Total	775,03	437,8	2,536.52	1,110.88	533,53	38,61	68,33	110,68
2018	2	Highest	31,48	17,72	96,01	61,73	26,11	2,61	14,99	5,6
2018	2	Lowest	20,46	11,75	78,67	25,58	4,75	0,5	0	1,07
2018	3	Average	27,54	14,55	90	36,35	17,37	1,28	1,9	3,56
2018	3	Total	853,61	450,99	2,789.93	1,126.98	538,51	39,55	58,93	110,3
2018	3	Highest	31,08	19,2	96,83	83,29	25,45	2,59	21,84	4,98
2018	3	Lowest	17,55	10,24	77,43	19,5	2	0,61	0	0,54
2018	4	Average	25,33	11,13	91,77	36,63	14,14	0,98	1,19	2,75
2018	4	Total	759,97	333,91	2,753.10	1,098.76	424,29	29,32	35,56	82,47
2018	4	Highest	28,98	16,1	96,92	62,88	19,21	2,3	10,67	3,74
2018	4	Lowest	19,65	5,56	79,96	13,93	4,96	0,4	0	1,03
2018	5	Average	22,78	4,86	87,51	26,38	13,9	1,06	0,36	2,5
2018	5	Total	706,26	150,81	2,712.68	817,74	430,86	32,71	11,18	77,41
2018	5	Highest	26,39	12,08	97,55	56,52	17,73	2,01	9,91	3,38
2018	5	Lowest	16,44	1,3	66,71	13,29	5,77	0,4	0	0,96

Tx = Daily Maximum Temperature °C, Tn=Daily Minimum Temperature °C, Rain=Total Rainfall [Calculated From Hourly Data] mm, Rs=Total Radiation [Calculated From Hourly Data] J/m², U2 =Average Wind Speed [Calculated From Hourly Data] ms, RHx=Daily Maximum Relative Humidity%, RHn=Daily Minimum Relative Humidity%, ET0= Total Relative Evapotranspiration [Calculated From Hourly Data] mm.

4.5 Results and discussion

4.5.1 Agro-morphological diversity

4.5.1.1 Quantitative data

Eleven agro-morphological characters measured were analysed using analysis of variance and the results are shown in Table 4.2.

Table 4.2 Summary of the analysis of variance table of agronomic traits of soybean accessions.

Source	DF	Days to flowering	Plant Height	Nodes at flowering	Branch number	Pod/plant	Pod_T
Accessions	28	**0.0009****	0.2543^{ns}	0.5538^{ns}	0.6500^{ns}	0.4379^{ns}	0.3873^{ns}
Rep	2						
Source	**DF**	**Yield/plant**	**Pod length**	**Seed/pod**	**Seed/plant**	**Moisture**	
Accessions	28	0.2928^{ns}	0.8948^{ns}	**0.0202***	0.3528^{ns}	0.7193^{ns}	
Rep	2						

NS= Not significant. * significant at the P-value ≤0.05, ** significant at the P-value ≤0.001.

No significant differences were found among genotypes on plant height, number of nodes at flowering, number of branches, number of pods per plant, yield per plant, and pod length. However, significant differences occurred on the number of days to 50% flowering (P≤0.001) and number of seed per pod (P<0.05). The early flowering genotypes, BD 1601, B 66 S 37, Oribi, Crawford, B 66 S 41, Egret, B 66 S 387, Dundee and Santa Rosa, took less than 85 days to flower. In contrast, B 66 S 256, Solar 12 and B 66 S 8 took more than 100 days to flower. Lee Ex RHOD, PR 165-52, Hawkeye, Yeluanda, ND 85, Kahala, IBIS 2000 and 61 S 156 genotypes were intermediate and took 86 to 90 days to flower (Figure 4.1).

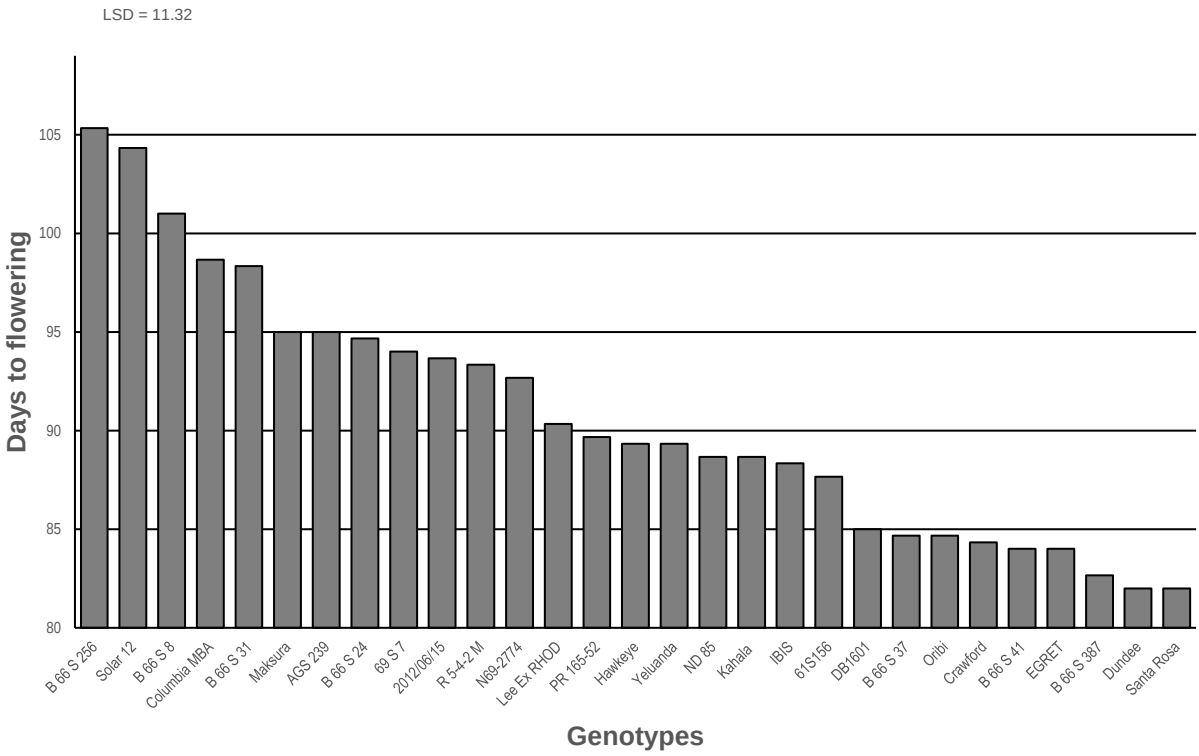


Figure 4.1 Number of days to 50% flowering of 30 genotypes of soybean.

About 63% of the evaluated genotypes had three seeds per pod, only Solar 12 had only two seeds per pod (Figure 4.2) and the remaining genotypes had less than two seeds per pod.

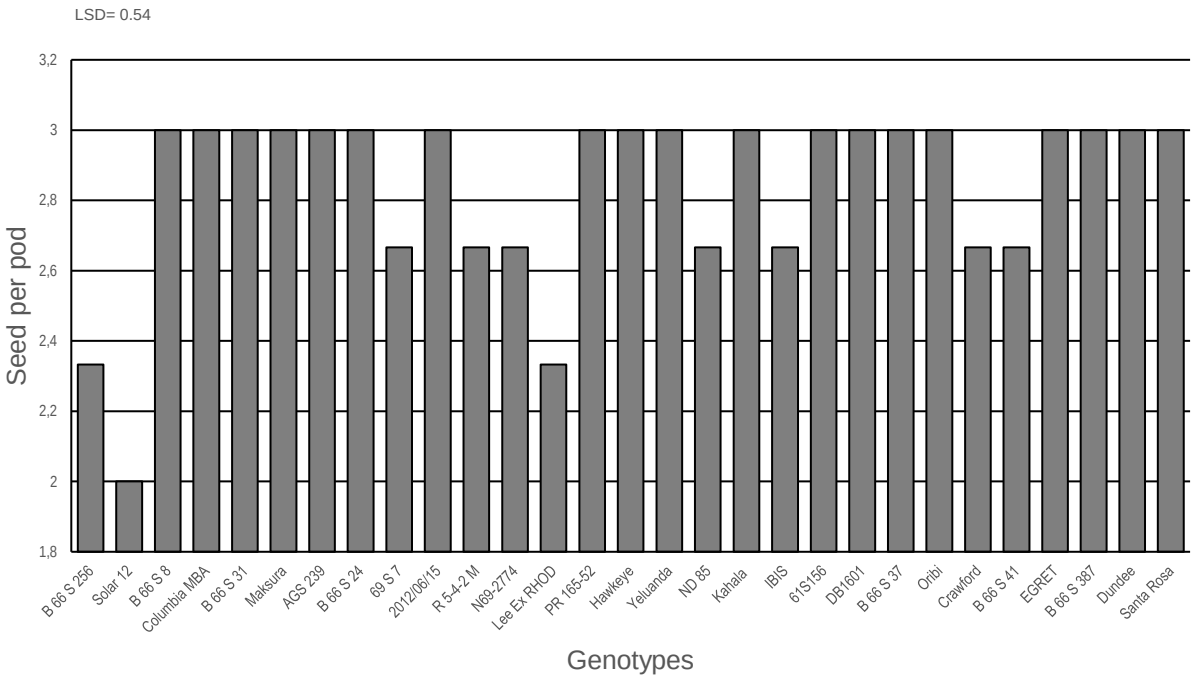


Figure 4.2 Seed number per pod of 30 soybean genotypes evaluated.

4.5.1.2 Qualitative data

The qualitative data of the thirty soybean genotypes grown in 2017/18 season is presented in Table 4.3. All genotypes studied matured with erect pubescence type. The studied soybean genotypes were dominated by 86.7% of dense and 13.3% of normal pubescence density. Most of the genotypes had 60% of the grey and 20% of brown and light brown pubescent colours. The corolla colors varied between purple and white with 73% of the genotypes being purple and the remainder were white. The genotypes showed different leaf shapes including narrow, intermediate and broad leaves with 46.7% of the genotypes having intermediate leaves, 43.3% had narrow leaves whereas 10% of the genotypes had broad leaves.

Table 4.3: Qualitative traits of 30 soybean genotypes recorded in the study.

Genotype	Maturity period	Flower colour	Stem determination	Pubescence	Pubescence type	Pubescence density	Pubescence colour	Corolla colour	Leaflet shape
69 S 7	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Narrow
B 66 S 31	Early	White	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
Lee Ex RHOD	Early	White	Indeterminate	Absent	Erect	Dense	Brown	White	Narrow
Columbia M8A	Early	Purple	Semi-determinate	Absent	Erect	Dense	Brown	Purple	Broad
IBIS	Early	Purple	Indeterminate	Absent	Erect	Normal	Grey	Purple	Narrow
R 5-4-2 M	Early	White	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
Egret	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Narrow
15/06/2012	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
DUNDEE	Early	Purple	Indeterminate	Absent	Erect	Dense	Brown	Purple	Narrow
Solar 12	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
Hawkeye	Early	White	Indeterminate	Absent	Erect	Normal	Light brown	Purple	Narrow
N69-2774	Early	White	Indeterminate	Absent	Erect	Normal	Grey	White	Narrow
Maksura	Early	White	Determinate	Absent	Erect	Dense	Grey	Purple	Narrow
DB 1601	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Broad
Yeluanda	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate

Oribi	Early	White	Determinate	Absent	Erect	Dense	Grey	White	Intermediate
ND 85	Early	White	Determinate	Absent	Erect	Normal	Light brown	White	Narrow
AGS 239	Early	Purple	Determinate	Absent	Erect	Dense	Brown	Purple	Intermediate
61 S 156	Early	White	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
Santa Rosa	Early	Purple	Indeterminate	Absent	Erect	Dense	Light brown	Purple	Narrow
B 66 S 41	Early	Purple	Determinate	Absent	Erect	Dense	Brown	Purple	Narrow
Egret	Early	Purple	Determinate	Absent	Erect	Dense	Grey	Purple	Intermediate
Crawford	Early	Purple	Determinate	Absent	Erect	Dense	Light brown	Purple	Intermediate
B 66 S 37	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
B 66 S 387	Early	White, Purple	Indeterminate	Absent	Erect	Dense	Brown	White	Narrow
B 66 S 24	Early	White	Indeterminate	Absent	Erect	Dense	Grey	White	Intermediate
B 66 S 256	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	Purple	Intermediate
Kahala	Early	White	Semi-determinate	Absent	Erect	Dense	Light brown	Purple	Intermediate
B 66 S 8	Early	Purple	Indeterminate	Absent	Erect	Dense	Grey	White	Narrow
PR 165-52	Early	White	Indeterminate	Absent	Erect	Dense	Light brown	Purple	Narrow

4.5.2 Diversity of nutritional quality traits

The nutritional quality traits of soybean genotypes were analysed using analysis of variance. The results are shown in Table 4.4. Significant differences (P≤0.05) were observed among the soybean genotypes based on ash content and oleic acid. Ash content varied between 4.8 and 6.6% whereas oleic acid ranged between 15.4 and 27.8% (Figure 4.3). The genotypes AGS 239 and Yeluanda had relatively higher ash content. Higher oleic acid contents were recorded for genotypes B 66 S 387, followed by B 66 S 8 and Hawkeye and the lowest was genotype DB 1606 (Figure 4.4).

Table 4.4: Summary of the analysis of variance table of chemical composition of soybean genotypes.

Source	DF	Linolenic acid	Ash	Protein	Oil	Moisture
Rep	2					
Accessions	28	0.8161^{ns}	0.0164*	0.2336^{ns}	0.3215^{ns}	0.7193^{ns}
Source	DF	Oleic acid	Palmitic acid	Stearic acid	Fiber	Linoleic acid
Rep	2					
Accessions	28	0.0292*	0.5141^{ns}	0.1514^{ns}	0.3140^{ns}	0.0955^{ns}

NS= Not significant. * significant at the P-value ≤0.05, ** significant at the P-value ≤0.001

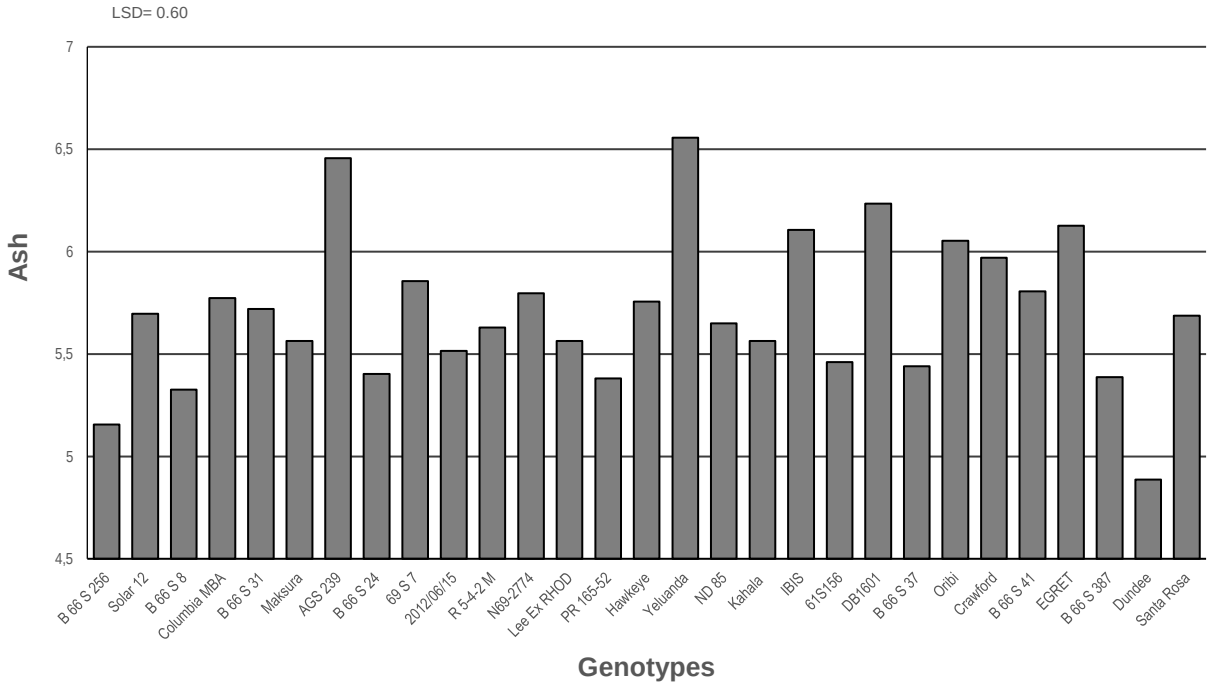


Figure 4.3 Ash content percentage of 30 genotypes of soybean evaluated.

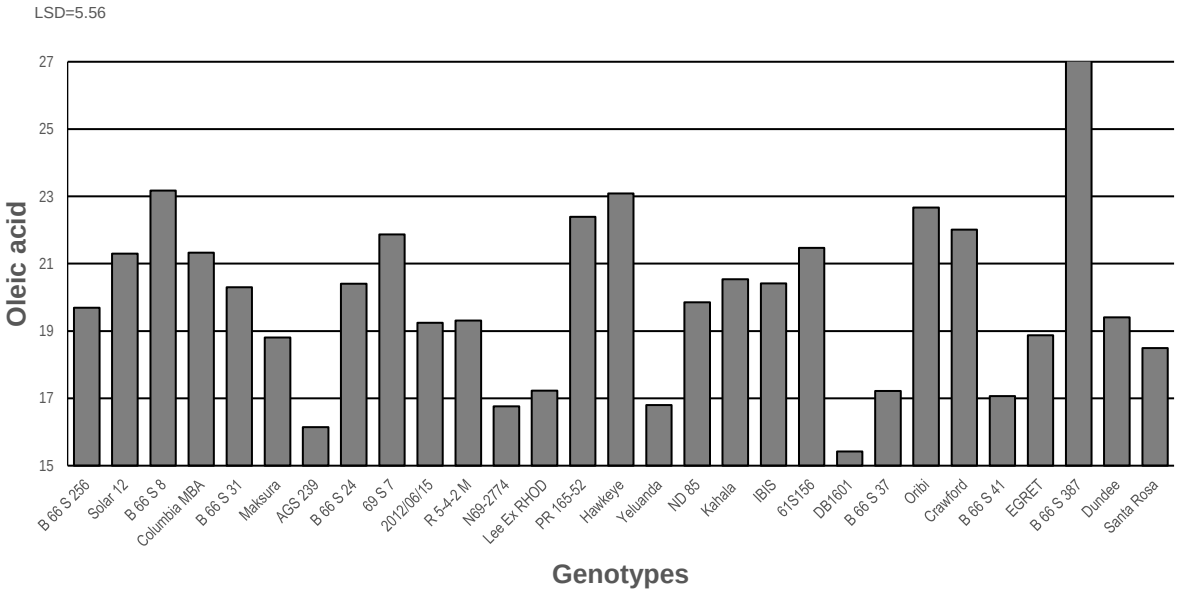


Figure 4.4 Oleic acid content of 30 genotypes of soybean evaluated.

4.6 Principal component analysis

i. Principal component analysis (PCA) of agro-morphological and quality traits of soybean.

Table 4.5 shows the agronomic data that were subjected to a principal component analysis, which revealed that three most important PCs contributed 21.3%, 14.9% and 9.1% to the total variation. The traits that contributed the most variation in the first PC were number of branches per plant, pod number per plant, pod weight before threshing and yield per plant for agromophological traits. In the second PC, the traits that were responsible for contributing the most variation were oil, oleic acid, protein, ash, linoleic acid and stearic acid. The largest contributors to variation in the third PC were moisture content, protein content, days to 50% flowering and plant height.

Table 4.5: Factor loadings of the three PCs based on agronomic and quality traits.

Traits	Factor loadings		
	F1	F2	F3
DFW	-0.048	0.288	-0.499
PHT	-0.278	0.264	-0.450
BNP	0.505	0.023	-0.011
PNP	0.932	-0.114	0.055
PBT	0.941	-0.016	0.045
SNP	0.017	-0.049	0.119
SDP	0.953	-0.147	0.019
YDP	0.937	-0.029	0.010
PDL	-0.022	0.140	-0.066
PDW	0.175	-0.006	-0.390
Protein	0.243	-0.578	-0.537
Oil	-0.096	0.841	0.271
Moisture	0.108	0.346	0.711
Ash	-0.245	-0.580	0.202
Fiber	-0.393	0.379	-0.276
Linoleic acid	-0.318	-0.588	0.234
Linolenic acid	-0.023	-0.300	-0.348

Oleic acid	0.293	0.701	-0.224
Palmitic acid	-0.014	-0.219	0.056
Stearic acid	-0.281	-0.562	0.159
NFW	-0.003	0.066	0.228
Eigenvalue	4.470	3.148	1.920
Variability (%)	**21.287**	**14.992**	**9.142**
Cumulative (%)	21.287	36.279	45.422

Days to 50% flowering=DFW; Plant height=PHT; Number of branches/plant=BNP; Pod number/plant=PNP; Pod weight before threshing=PBT; Seed number per/pod=SNP; Seed number/plant=SDP; Yield/plant (g)=YDP; Pod length (mm)=PDL; Pod width (mm)=PDW; Nodes at flowering=NFW.

ii. Principal component biplots

The principal component biplot (Figure 4.5) grouped the tested soybean accessions into two major groups. The accessions exhibiting high plant height, high number of nodes at flowering, high number of days to 50% flowering, high pod length and high seed number per pod were grouped together. The second group consisted of yield per plant, pod weight before threshing, seed number per plant and pod number per plant. The nutritional quality traits such as linoleic, ash and stearic acid were grouped together and palmitic and linolenic acids were also clusterd together, whereas other quality traits were independent of one another.

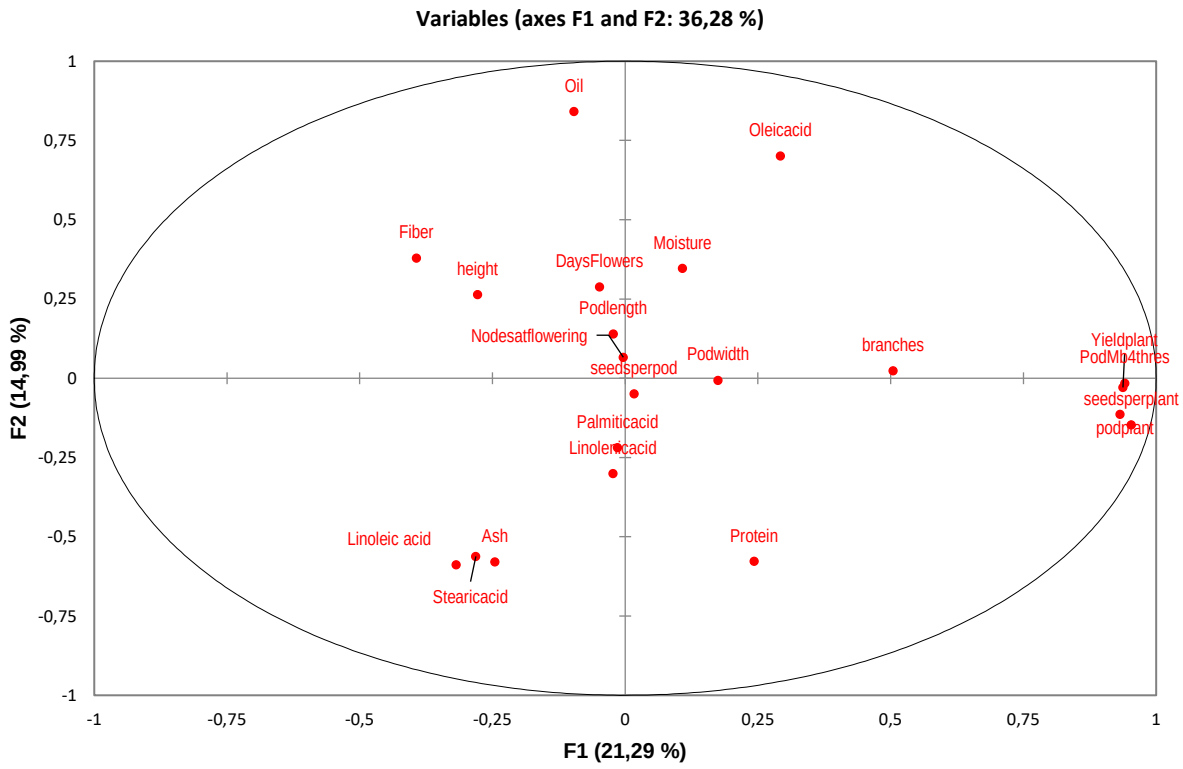
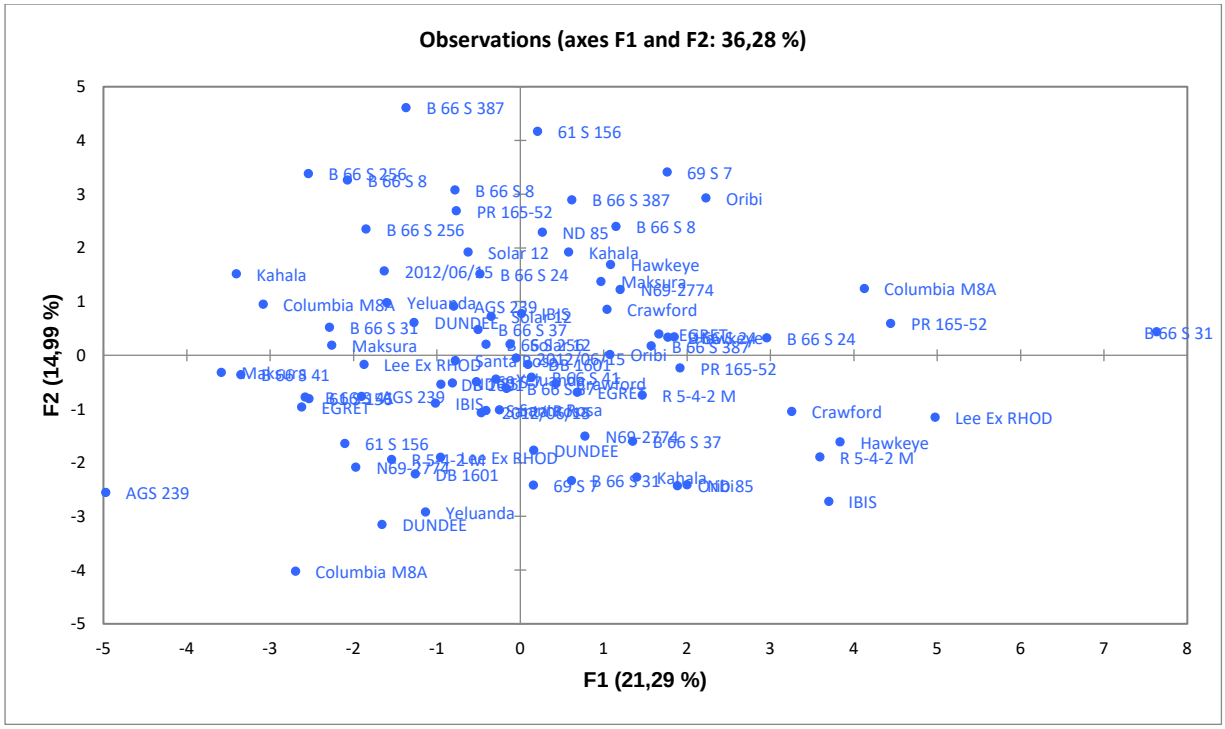


Figure 4.5 A Principal Component Biplot of eleven agro-morphological traits of 30 soybean accessions.



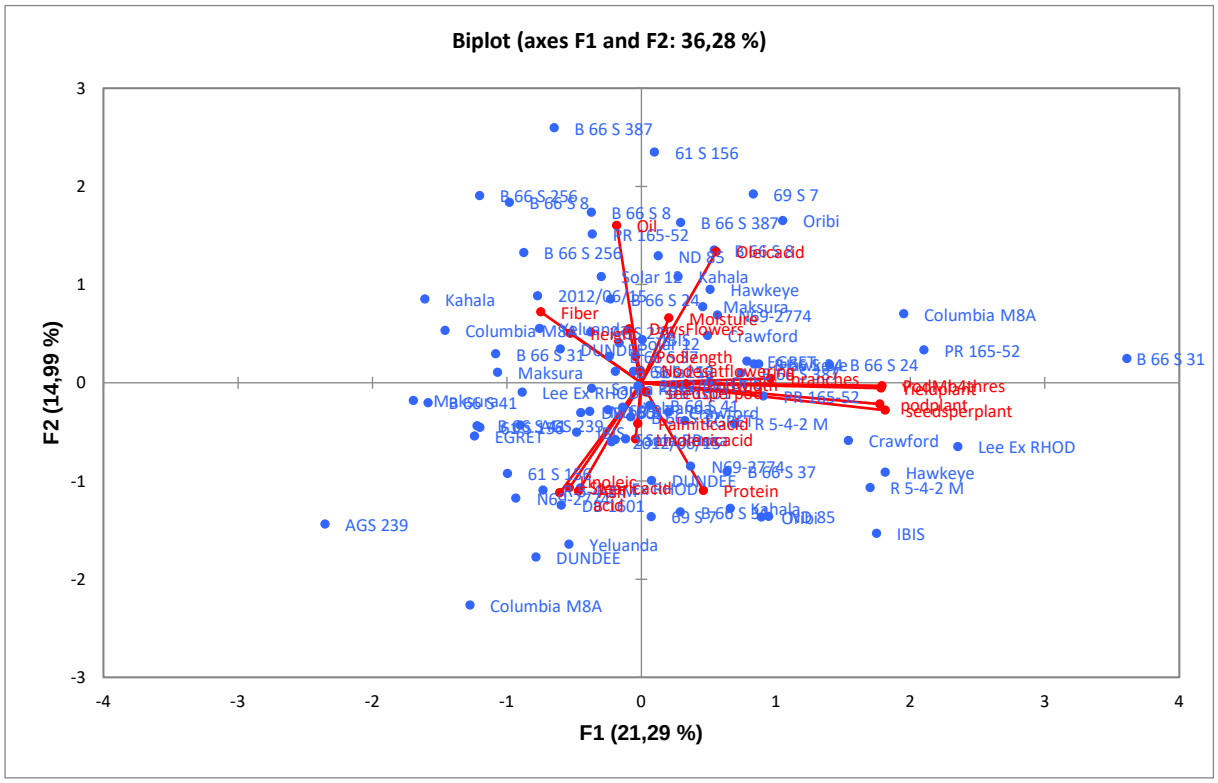


Figure 4.6 A Principal Component Biplot depicting agro-morphological and nutritional quality traits of 30 soybean accessions.

Figure 4.6 showed that number of branches per plant and pod weight before threshing were highly significant and highly correlated with plant height, days to 50% flowering, nodes at flowering, pod length as well as oil, fiber, oleic acid and moisture; whereas they were negatively correlated with pod number per plant, seed per plant, ash, linoleic acid, linolenic acid, palmitic acid, stearic acid and ash. However, was also positively and significantly associated with pod width. Protein was highly significant and positively correlated with palmitic acid, linolenic acid, linoleic acid as well as stearic acid but negatively correlated fiber. It was also significant and positively correlated with seed number per pod. Seed number per pod was highly significant and negatively associated with pod weight per plant, days to 50% flowering, but positively correlated with pod number per plant. Seed number per pod was highly and positively associated with seed number per plant and protein. Pod number per plant was significant and positively correlated with protein. Pod number per plant before threshing was highly and significantly associated with pod width.

ii. Correlation Analysis among Phenotypic and Quality Traits

The phenotypic and quality traits were analysed using correlation coefficients (Table 4.6). The association of the traits were reported based on the 5% significance level. As illustrated in Table 4.6, plant height was positively correlated with days to flowering. Pod number per plant was positively correlated with plant height and branch number per plant. Pod weight before threshing recorded a positive correlation with plant height, branch number per plant and pod number per plant. Seed number per pod was positively associated with plant height, branch number per plant and pod number per plant. Seed yield per plant was positively associated with plant height, branch number per plant, pod number per plant and seed number per pod. Pod width recorded a negative correlation with seed number per pod. Number of nodes at flowering was negatively correlated with palmitic acid. Oil content recorded a positive correlation with days to flowering and and a negative one with protein. Moisture content showed a negative association with protein content and a positive association with with oil content. Ash was positively correlated with plant height, seed yield per plant and oil content, and negatively associated with pod weight before threshing. Fiber recorded a positive correlation with plant height, number of branches per plant, seed number per plant, seed yield per plant and oil content and also showed a negative correlation

with pod number per plant and pod weight before threshing. Linoleic acid was negatively correlated with days to flowering and positively correlated with oil and ash. Linoleic showed a negative association with moisture content and positive association with linoleic acid. Oleic acid correlated positively with seed yield per plant, oil content and ash and negatively with linoleic acid. Palmitic acid correlated positively with ash and fibre and negative with linoleic and oleic acids. Strearic acid showed a positive correlation with oil content and linoleic acid and a negative correlation with with oleic acid. On leaf shape 43.33% were narrow, 46.67 intermediates and 10% were indeterminates. Pubescene density varied at 86.67% for the erect types and 13.33% for the appressed types. Sixty percent of the stems were determinate, 6.67% semi-determinate and 33.33% indeterminate.

Table 4.6: Correlation matrix (Pearson) of eleven agromorphological and ten nutritional quality traits of soybean.

Variables	DFW	PHT	BNP	PNP	PodMb4thres	Seeds perpod	SDP	YDP	PDL	PDW	Protein	Oil	Moisture	Ash	Fiber	Linoleic acid	Linolenicacid	Oleicacid	Palmiticacid	Stearicacid	Nodesat flowering
DFW	1																				
PHT	0,290	1																			
BNP	0,065	0,032	1																		
PNP	-0,014	-0,282	0,391	1																	
PodMb4thres	-0,053	-0,217	0,419	0,906	1																
seedsperpod	-0,158	0,003	-0,003	-0,075	0,033	1															
SDP	-0,109	-0,275	0,469	0,936	0,917	0,045	1														
YDP	-0,027	-0,240	0,401	0,876	0,919	0,061	0,895	1													
PDL	-0,159	0,026	0,084	-0,181	-0,008	0,195	-0,011	0,021	1												
PDW	0,175	0,045	0,001	0,151	0,129	-0,225	0,171	0,171	-0,025	1											
Protein	-0,063	-0,004	0,050	0,180	0,220	0,102	-0,249	-0,229	0,049	0,138	1										
Oil	0,220	0,159	-0,050	-0,123	-0,031	-0,073	0,178	0,058	-0,042	-0,087	-0,742	1									
Moisture	-0,123	-0,120	0,039	0,087	0,153	-0,006	0,026	0,127	-0,061	-0,105	-0,494	0,455	1								
Ash	-0,161	0,227	-0,161	-0,125	-0,256	-0,004	0,141	-0,240	0,160	-0,172	0,077	-0,543	0,015	1							
Fiber	0,199	-0,244	0,212	-0,343	-0,239	0,005	-0,351	-0,293	0,191	0,064	-0,149	-0,406	-0,133	0,159	1						
Linoleic acid	-0,221	0,104	-0,105	-0,150	-0,185	0,007	-0,165	-0,175	0,087	-0,025	0,065	-0,218	-0,087	0,257	-0,096	1					
Linolenicacid	0,130	0,116	0,008	0,025	-0,007	-0,088	0,031	0,052	-0,061	-0,070	0,135	-0,206	-0,306	-0,204	-0,019	-0,211	1				
Oleicacid	0,071	-0,103	-0,078	0,129	0,181	0,002	-0,139	-0,214	0,126	-0,014	-0,117	-0,325	0,066	-0,338	-0,129	-0,690	-0,089	1			
Palmiticacid	0,114	-0,032	-0,027	0,053	-0,034	0,042	-0,005	-0,086	-0,122	-0,152	0,073	-0,155	0,064	-0,289	-0,214	-0,219	0,038	-0,317	1		
Stearicacid	-0,151	0,041	-0,025	-0,176	-0,165	-0,023	-0,138	-0,176	-0,004	0,030	0,132	-0,240	-0,097	-0,151	-0,024	-0,626	0,039	-0,654	-0,081	1	
Nodesatflowering	-0,149	0,067	-0,040	0,027	0,022	0,035	0,008	0,015	0,020	-0,166	-0,142	0,086	-0,030	-0,097	-0,084	0,103	-0,017	0,075	-0,241	0,085	1

Days to 50% flowering=DFW; Plant height=PHT; Number of branches/plant=BNP; Pod number/plant=PNP; Pod weight before threshing=PodMb4thres; Seed number per pod=Seeds perpod; Seed number/plant=SDP; Yield/plant (g)=YDP; Pod length (mm)=PDL; Pod width (mm)=PDW; Nodes at flowering=NFW.

Relationship between the traits

The majority of soybean cultivars that are grown and consumed throughout the world today exhibit yellow or white seed coats, whereas the majority of known accessions of the wild progenitor, *G. soja*, have black or, rarely, brown seed coats. Soybean cultivars should also have chemical composition and qualitative traits that are in need and particularly easy to grow and maintain especially for beneficial aspects either for personal consumption or financial benefits.

The principal component biplot analysis was not only able to identify specific traits contributing to yield but also to identify the out-performing varieties, as this helps to select new lines with these traits. Soybean varieties with higher plant height, more oil content, more branches and pods, grains, fibre, and early flowering days obtained higher yield. Dundee and N69-2774 were associated with high oil content. According to the PCA biplots Lee Ex RHOD, R 5-4-2 M, N69-2774 and DB 1601 were associated with both ash and stearic acid as illustrated on figure 6.4. Crawford, Egret and B 66 S 387 accessions were associated with palmitic acid. Yeluanda,15/06/2012, B 66 S 31 had the highest fibre and height. Maksura, N69-2774 and Crawford were associated with both moisture and days to flowering. B66 S8 had the highest oleic acid content. DB 1601 had the highest number of seeds per pod. PR 165-52 and B 66 S 8 contained the largest oil content. High node set at flowering was observed in 15/06/2012 while high pods per plant were observed in Crawford and Lee Ex RHOD. B 66 S 387 and Oribi had the highest number of branches while PR 165-52 and B 66 S 24 had the highest yields per plant.

Discussion

Genetic diversity of germplasm collections can be determined using a traditional cluster analysis; which is easy and effective according to Belamkar (2011). This type of method studies the genetic diversity by forming core subsets while grouping accessions according to similar characteristics into one homogenous category (Ulaganathan and Nirmalakumari. 2015). Relationships of accessions are grouped according to similar units and thus make it easy to understand them better and easily translatable. According to Dwivedi et al. (2001) and Paterson et al. (1991), the

evaluation of genetic diversity among germplasm is highly required in any hybridization program as it aids in improving the genes of the most diverse lines as well as promoting the use of genetic variations. It is imperative for a breeder to obtain information on genetic diversity and relationships among breeding materials in order to improve the efficiency of a crop, this helps with efficiently managing and conserving germplasm resources. De Chavez et al. (2017) recently studied the diversity among soybean accessions in Philippines reporting a vast diversity. Another study conducted by Khatab et al. (2012) reported the presence of genetic diversity among soybean genotypes using ago-morphological descriptors. Hamzekhanlu et al. (2011) studied 34 mutant lines including one control cultivar and found variability for ten quantitative traits in soybean. A study finding significant differences among all the assessed phenotypic traits was reported by Iqbal et al. (2008).

The assessment of morphological traits with the aim of studying the genetic diversity and classification of existing germplasm material is a traditional method that is low level but can also be a powerful taxonomic tool that can be used to group germplasm prior to their characterisation particularly using more precise marker technologies. Hymowitz (1970) added that the genetic base of soybean cultivars is considered extremely narrow. Li et al. (2020) observed soybean varieties with higher plant height, more nodes of main stem, branches, pods, grains, and 100-grain weight, or longer growth periods may have higher yield. This concurs with Achina et al. (2019) who observed that the significant attributes include the number of pods per plant, seeds per pod, and seed weight, which determine the seed yield. From the research, DB 1601 had the highest number of seeds per pod while while PR 165-52 and B 66 S 24 had the highest yields per plant.

In this study, the nutritional quality traits showed diversity among the studied accessions where the oil and protein contents were within the range recorded in other studies (Shi et al., 2010). The oil content showed a range from 13.16% to 15.41%, while protein ranged between 33.29% and 35.71% with Dundee and N69-2774 having the highest high oil content. In order to select accessions with good qualities for hybridization and conservation, breeders will find the diversity among the accessions very beneficial. In a study conducted by Mazid et al. (2013), it was highlighted that the

analysis of genetic diversity is very crucial for identifying parents while Chowdhury et al. (2002) stated that it helps to achieve long-term selection gain.

The study revealed that traits that contributed to the most variation were number of branches per plant, pod number per plant, pod weight before threshing, seed number per plant, and yield per plant, protein and oleic acid. High number of pods per plant was observed in Crawford and Lee Ex RHOD. High oleic acid content was recorded for genotype B 66 S 387, followed by B 66 S 8 and Hawkeye. All genotypes studied matured early with erect pubescence type. The studied soybean genotypes were dominated by 86.7% of dense and 13.3% of normal pubescence density. Most of the genotypes had 60% of the grey and 20% of brown and light brown pubescent colours. The corolla colors varied between purple and white with 73% of the genotypes being purple and the remainder were white. Breeders in the direct and indirect programmes can be provided with information of the traits that were tested for correlation and traits that are desirable in order to select and use these genotypes with significance concurrently. Ismail and Khalifa (2001), and Kumar and Shukla (2002) also added that yield improvement programs thrive better when a breeder understands the relationship between the component of traits and yield as this will aid in making the best selection of desirable genotypes for these programs.

Plant height, days to flowering, plant branches, pods weight before threshing, and pod length concur with the results reported by a study carried out by Malek et al. (2014) where there were positive significant correlations. Machikowa et al. (2007) argue that early maturing varieties tend to have lower yield while late flowering is associated with higher soybean yield. Therefore, the extension of days to flowering of current early varieties, as well as the development of new varieties with a longer vegetative period, may result in higher yielding varieties. These results mean that when selecting genotypes that are high yielding for a breeding programme, these characters should be given more preference and emphasis as the best selection criteria. Kato et al. (2019) observed that the seed yields of semi-determinate and indeterminate lines were higher than that of the determinate ones; and that of the semi-determinate lines was marginally lower than that of the indeterminate lines. The research also highlighted that lodging score of semi-determinate varieties was smaller than that

of indeterminate varieties because the main stem length of the semi-determinate varieties was shorter than that of the indeterminate varieties. Anand and Torrie (1963) and Arashad et al. (2006) stated that seed yield always showed a positive correlations with other desirable yield traits which indicates that the increase in one trait would result in the increase of the other; that is, simultaneous increase or decrease of both traits would be easy. Soybean genotypes showing high yields indicate a strong positive correlation between seed yield and other traits and would be fairly easy to identify together with high number of pods per plant.

The statistical method that is commonly used in populating genetics in order to identify the genetic distribution and structure across a geographical and ethnic is also known as the principal component analysis (PCA) (McVean, 2009). It is further explained that the reason for this type of analysis is to evaluate each and every variable that forms part of the variation that is available among the studied genotypes (Mofokeng and Mashingaidze, 2018). According to Cruz and Carneiro (2003), this type of method aids in differentiating the important and less important traits within a studied group by grouping them accordingly.

Conclusion

Using agro-morphological and quality traits, some of the most important soybean cultivars were oberseved. These included Dundee and N69-2774 having the highest high oleic acid and oil contents. DB 1601 had the highest number of seeds per pod while while PR 165-52 and B 66 S 24 had the highest yields. There were positive correlations showed between most of the assessed traits in relation to one another and this will be very helpful in assisting in the combined improvement of these traits by selecting ones that were found to have a positive and high correlation, as well as easily measurable phenotypic traits, although most were found to have highly significant and positive correlation with seed yield per plant. The nutritional quality traits also varied significantly among the accessions. In order to achieve a successful selection of parents for breeding and transgressive segregation, studying and knowing the presence of genetic diversity can be very useful. 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 programmes. The specific objectives of the study were successfully accomplished as shown above in the study. Soybean is one of the most important leguminous crops grown globally. Understanding the genetic diversity and its interaction with the environment is of paramount importance in developing cultivars considering farmer's preferred traits. The multivariate analysis was able to reveal the relationship between the genetic diversity, agronomic and nutritional composition of selected soybean genotypes.

4.6 References

Achina, T., Twumasi, P. and Quain, M.D. (2011): Use of morphological amd molecular profiling for the establishment of relatedness between commercially viable and elite soybean. European Journal of Biotechnology and Genetic Engineering 6:1-17.

Aditya, J.P., Bhartiya, P. and Bhartiya, A. (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.

Almeida, R.D., Peluzio, J.M. and Afférri, F.S. (2011): Divergência genética entre cultivares de soja, sob condições de várzea irrigada, no sul do Estado Tocantins, Revista Ciência Agronômica 42:108-115.

Anand, S.C. and Torrie, J.H. (1963): Heritability of yield and other traits and interrelationship among traits in the F3 an F4 generations of three soybean crosses, Crop Science 3:508-511.

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

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

Belamkar, V., Selvaraj, M.G., Ayers, J.L., Payton, P.R., Puppala, N. and Burow, M.D. (2011): A first insight into population structure and linkage disequilibrium in the US peanut minicore collection, Genetica139:411-429.

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

Cholastova, T. and Knotova, D. (2012): Using morphological and microsatellite (SSR) markers to assess the genetic diversity in Alfalfa (*Medicago sativa L.*). World Academy of Science, Engineering and Technology 69:1-7.

Chowdhury, M.A., Vandenberg, B. and Warkentin, T. (2002): Cultivar identification and genetic relationship among selected breeding lines and cultivars in chickpea (*Cicer Arietinum L.*), Euphytica 127:317-325.

Cruz, C.D. and Carneiro, P.C.S. (2003): Biometric models applied to genetic improvement. Viçosa: UFV.

Cunha, M.C.G., Hamawaki, O.T. and Sousa, L.B. (2013): Genetic variability among 79 soybean progenies from ufu breeding program, Bioscience Journal 29:340-349.

Dayaman, V., Senthil, N., Raveendran, M., Nadarajan, N., Shanmugasundaram, P. and Balasubramanian, P. (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 Chavez, H., Borromeo, T., Oswald, B.N. and Gentallan, R. (2017): Phenotypic diversity of soybean (*Glycine max* (L.) Merr.) accessions in the Philippines for utilization. Legume Research-An International Journal 40:9-15.

diversity panel based on agro-morphological traits. Genetics and Molecular Research 15 (4): gmr15048980.

Dwivedi, S.L., Gurtu, S., Chandra, S., Yuejin, W. and Nigam, S.N. (2001): Assessment of genetic diversity among selected groundnut germplasm by RAPD analysis. Plant Breeding 120:345-350.

Genetic Analysis of Genes Controlling Natural Variation of Seed Coat and Flower Colors in Soybean. Journal of Heredity 101:757–768.

Ghanbari, S., Nooshkam, A., Fakheri, B.A., and Mahdinezhad, N., (2018): Assessment of yield and yield component of soybean genotypes (*Glycine Max* L.) in North of Khuzestan. Journal of Crop Science and Biotechnology 21**:**435-441. https://doi.org/10.1007/s12892-018-0023-0.

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

Hamzekhanlu, M.Y., Darbandi, A.I., Beiranvand, P., Hallajian, M.T. and Majdabadi, A. (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.

Hymowitz, T. (1970): On the domestication of the soybean, Economic Botany 24:408-421.

Ibidunni, S.A, Omolayo, J.A., Olabisi, A., and Sunday, O.O. 2020: Agronomic potential and genetic diversity of 43 accession of Tropical soybean (*Glycine max* (L)

Merrill). International Journal of Plant Research 10(2):33-39. doi: 10.5923/j.plant.20201002.02.

International Board for Plant Genetic Resources. (1984): Descriptors for soybean. IBPGR Secretariat, Rome. AGPG: IBPGR/84/183.

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

J. (2010): Jeong, N. and Moon Lee, Kato, S., Sayama, T., Taguchi-Shiobara, F., Kikuchi, A., Ishimoto, M. and Cober, M. (2019): Effect of change from a determinate to a semi- determinate growth habit on the yield and lodging resistance of soybeans in the northeast region of Japan. Breeding Science 69:151-159. doi:10.1270/jsbbs.18112.

Khalifa, M.A., and H.A.K. and Ismail, A.A., (2001): Genetic studies on some yield traits of durum wheat. Asian Journal of Agricultural Science 32:103-129.

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

Kumar, P. and Shukla, R.S. (2002): Genetic analysis for yield and its attributed traits in bread wheat under various situations. Jawaharlal Nehru Krishi Vishwa Vidyalaya Research Journal 36: 95-97.

Li, M., Liu, Y., Jiang, W., Waqng, C., Yang, X., Li, D., Zhang, X., Xu, C., Zhang, Y., Li, W., and Zhao, L. (2020): Identification of traits contributing to high and stable yields in different soybean varieties across three different Chinese latitudes. Frontiers in Plant Science 10:1642. https://doi.org/10.3389/fpls.2019.01642.

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

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

Machikowa, T., Waranyuwat, A. and Laosuwan, P. (2005): Relationships between seed yield and other characters of different maturity types of soybean grown in different environments and levels of fertilizer. Science Asia 31:37-41.

Machikowa, T., Waranyuwat, A., Burton, J.W. and Laosuwan, P. (2007): Yield improvement of early maturing soyabeans by selection for late flowering and early maturity. Science Asia 33:229-234.

Malek, M.A., Raffi, M.Y., Afroz, M.S.S., Nath, U.K., and Mondol, M.M.A. (2014): Morphological characterization and assessment of genetic variability, character association, and divergence in soybean mutants. Scientific World Journal pp. 1-12.

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

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

Marconato, M.B., Pereira, E.M., Silva, F.M., Bizari, E.H., Pinheiro, J.B., Mauro, A.O. and Unêda-Trevisoli, S.H. 2016. Genetic divergence in a soybean (*Glycine max*)

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

Mazid, M.S., Rafii, M.Y., Hanafi, M.M., Rahim, H.A., Shabanimofrad, M. and Latif, M.A. (2013): Agro-morphological characterization and assessment of variability, heritability, genetic advance and divergence in bacterial blight resistant rice genotypes. South African Journal of Botany 86:15-22.

McVean, G. (2009): A genealogical interpretation of principal components analysis, PLoS Genetics 5(10): e1000686. https://doi.org/10.1371/journal.pgen.1000686.

Mofokeng, M.A. and Mashingaidze, K. (2018): Genetic Differentiation of ARC Soybean [*Glycine Max* (L.) Merrill] accessions based on agronomic and nutritional quality traits. Agriculture and Food Sciences Research 5 (1):6-22.

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.

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

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S., and Tanksley, S.D. (1991): Genetics. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations and environments. Genetics 127:181-197.

Ravikumar, M. (1999): Identification of soybean varieties based on seed, seedling and plant characters. University of Agriculture Sciences, Bangalore, Karnataka, India.

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

Shi, A., Chen, P., Zhang, B. and Hou, A. (2010): Genetic diversity and association analysis of protein and oil content in food grade soybeans from Asia and the United States. Plant Breeding 129: 250-256.

Shoemaker, R.C., Polzin, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Kochert, G. and Boerma, H.R. (1996): Genome Duplication in Soybean (*Glycine subgenus soja*). Genetics 144:1329-338.

Steel, R.G.D., and Torrie, J.H. (1980): Principles and Procedures of Statistics. 2nd ed. New York: McGraw-Hill.

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

Thorup, T.A, Tanyolac, B, Livingstone, K.D, Popovsky, S, Paran, I, Jahn, M. (2000): Candidate gene analysis of organ pigmentation loci in the *Solanaceae.* Proceedings of the National Academy of Sciences USA 97: 11192–11197.

Ulaganathan, V. and Nirmalakumari, A. (2015): Finger millet germplasm characterization and evaluation using principal component analysis. Sabrao Journal of Breeding and Genetics 47:79-88.

User Manual DA 7250. (2013): Hägersten, Sweden.

Yang, K., Zeffa, D.M., Perini, L.J., Freiria, G.H., da Silva, G.M., Gomes, G.R., Constantino L.V., Alves, S.M., Koyama, R., and de Oliveira Neto, S.S. 2019:

Genetic variability in food-type soybean accessions assessed by morphoagronomic traits. Genetics and Molecular Research 18 (2): gmr18298.

CHAPTER FIVE

SUMMARY AND RECOMENDATIONS

The analyses for SSR formed two major distinct clusters as the results showed clear separation between the breeding lines and landrace maintained by the Agricultural Research Council. Genotypes that had the lowest dissimilarity index were 69 S 7 (001) and Yeluanda (015), as well as R-5-4-2 M (006) and B 66 S 31 (002); while Santa Rosa (020) and PR 165-52 (030) showed high similarity. A wide genetic diversity was observed with a total of 216 bp alleles with an average of 10.8 alleles per locus detected. The allele numbers ranged from 4 to 22 with an average of 10.8 per locus, whereas the alleles size ranged from 2 to 33 bp. The polymorphic information content ranged from 0.46 to 0.85 (average of 0.716) with heterozygosity values of 0.50 to 0.87 (average 0. 7485).

The principal component analysis (PCA), revealed the three most important PCs, which contributed 21.3%, 14.9% and 9.1% to the total variation. The traits that contributed the most variations were number of branches per plant, pod number per plant, pod weight before threshing, seed number per plant, and yield per plant. The principal component biplot identified the genotypes Dundee and N69 2774 as associated with high oil content while Ex RHOD, R 5-4-2 M, ND69 2774M and DB 1601 were associated with both ash and stearic acid. Crawford, Egret and B66-837 accessions were associated with palmitic acid. DB 1601 had the highest number of seeds per pod. Yeluanda,15/06/2012, B66 S31 had the highest fibre and height. Maksura, ND2692774 and Crawford were associated with both moisture and days to flowering. B66 S8 had the highest oleic acid content. PR 165-52 and B66 S8 contained the largest oil content. High node set at flowering was observed in 15/06/2012 while high pods per plant were observed in Crawford and Lee Ex RHOD. B66 S387 and Oribi had the highest number of branches while PR 165-52 and B 66 S24 had the highest yields per plant. Soybean varieties with higher plant height, more oil content, branches, pods, grains, fibre, branches, and early days to flowering obtained higher yield. The lines that flowered between 80 to 103 days such as Santa Rosa, Dundee, AGS 239 and B 66 S 387 were the most ideal, whereas, B 66 S 31, Columbia M8A,

Solar 12, B 66 S 24 and Kahala were the least. Plant height ranged between 80 cm and 150 cm, making B 66 S 256 and Columbia M8A the tallest genotypes. However, B 66 S 41 proved to be the shortest genotype standing at only 70 cm. One other thing that was evident is that purple dominated for corolla colour with erect pubescence. Genotype ND 85 yielded the highest seed weight (137.7 g) with Maksura yielding less (30 g).

Based on the findings above, the study established the existence of considerable genetic diversity among soybean germplasm maintained by the Agricultural Research Council based on phenotypic, genotypic and nutritional quality traits. The lines identified with superior performance can be selected and used for further quality breeding to achieve profit margins and earnings which farmers can really benefit from. They can also be beneficial for genetic development, enhancement and conservation of breeding populations. For further investigation, it is recommended that genetic diversity studies are proposed to continue enriching it by bringing foreign germplasm. The 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.