GENETIC IMPROVEMENT OF NUTRITIONAL TRAITS AND YIELD OF TROPICAL SOYBEAN IN UGANDA

TONNY OBUA

B. Sc. Agric (Mak); M. Sc. Crop Science (Mak)

2014/HD02/18650U

A THESIS SUBMITTED TO THE DIRECTORATE OF RESEARCH AND GRADUATE TRAINING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DOCTOR OF PHILOSOPHY IN AGRICULTURE DEGREE OF MAKERERE UNIVERSITY

MARCH, 2022

DECLARATION

This thesis is my original work and has not been presented for an award of a degree in any other University

Signed..... Date..24/03/2022.

Tonny Obua

This thesis has been submitted with our approval as University supervisors

Signed..... Date..24/03/22

Prof. Phinehas Tukamuhabwa
Department of Agricultural Production
School of Agricultural Sciences
Makerere University

Signed..... Date..24/03/2022

Dr. Thomas L. Odong
Department of Agricultural Production
School of Agricultural Sciences
Makerere University

i

DEDICATION

To my children Abigail, Christabel, Timothy and Tabitha. May they find this Thesis a great source of inspiration.

ACKNOWLEDGEMENTS

During the undertaking of this research, I had the opportunity to interact with various people who contributed directly and indirectly to the accomplishment of its objectives. I am grateful to my academic supervisors, Prof. Phinehas Tukamuhabwa and Dr. Thomas L. Odong who guided me during my studies. I am also indebted to the Doctoral Committee Members (Assoc. Prof. S. B. Mukasa, Dr. G. Tusiime, Assoc. Prof. Denis Mpairwe, Assoc. Prof. J.B. Tumuhairwe) for their invaluable advice and guidance throughout my studies. Special thanks to Dr. Julius P. Sserumaga, Dr. Stephen Opio, Dr. Oluwaseyi Shorinola and, Mr. Bruno Awio for all the advice they offered to me especially in data analysis.

Sincere gratitude is extended to Dr. Wellington Ekaya, Dr. Josiah Mutuku, Dr. Nasser Yao, Ms. Valerian Aloo, Mr. Fredrick Nganga, Mr. Vincent Njunge, Mrs. Phylis Ochieng, Mrs. Eunice Machuka, Mrs. Lucy Muthui and, all the ABCF fellows for their guidance, encouragement that made my stay at BeCA worthwhile. Gratitude is extended to Makerere University Centre for Soybean Improvement and Development staff particularly Mrs. Mercy Namara Asiimwe, Mr. George Yiga, Mr. Paul Kabayi and, Mr. Yafesi Senkubuge for their unwavering support.

I also thank Makerere University for nominating me to BecA-ILRI Hub in Nairobi for support through the ABCF fellowship program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Soybean constitutes about 40% protein, 20% oil; a highly nutritious legume that has enormous potential to improve dietary quality for people throughout the world. To understand the genetic diversity and population structure of tropical soybean germplasm, 89 genotypes from diverse sources were analyzed using 7,962 SNP markers. The results showed low genetic diversity among the studied germplasm and the polymorphism information content (PIC) was 0.27. The phylogenetic tree and Principal Component Analysis (PCA) both showed that the 89 soybean genotypes were grouped into three major clusters; while population structure grouped the genotypes into two major subpopulations. On the other hand, the average Roger genetic distance within the study population was 0.34.

The variability of different nutritional traits of 52 tropical soybean core collections from diverse sources, showed that traits varied significantly depending on genotypes and country of origin.

Total protein content ranged from 35.07% to 50.4% and genotypes Sline 5.18, BSPS 48A-8 and BSPS 48A-27-1 had the highest protein content of 50.40%, 48.88% and 48.08% respectively. On the other hand, NIIXGC 17.3 and Nam II had the lowest protein content of 30.07% and 35.57% respectively. Total oil content ranged from 14.94% to 23.48% where genotypes G32B, Roan and AGS 338 significantly had the highest oil content among the others while Signal, Maksoy 5N and Sline 16.2 had the lowest. The relative percentage of major fatty acids ranged from 10.58% to 21.18% for palmitic acid (16:0), 4.93% to 16.76% for stearic acid (18:0), 22.69% to 39.95% for oleic acid (18:1), and 30.60% to 51.72% for linoleic acid (18:2). Total oil content varied significantly between origins with genotypes from Seed Co having the highest mean of 20.13% while those from AVRDC had the lowest mean of 18.32%. Genotypes from Uganda had the highest percentages of oleic acid; followed by genotypes from Japan and AVRDC.

The GWAS based on 92 soybean genotypes revealed two significant associations (-log[*P-value*] > 5) with oil content for two SNPs, rs2291820 and rs22918919 on chromosome 7 and 10 respectively. A significant association (-log[*P-value*] > 2.5) with protein content was detected for 3 SNPs, rs 22918920, rs 22918919 and rs 1494480 located on chromosomes 7, 10 and 20 respectively. The study found that both rs2291820 and rs22918919 located on chromosomes 7 and 10 were associated with both oil and protein content.

Genotype NII X GC 20.3 had the highest mean protein content of 43.0%, and BSPS 48A-9-2 and BSPS 48A-28 were superior for mean grain yield (1,207 kg ha^{-1}). Bulindi was the most discriminating and representative test environment for soybean yield. A weak and negative correlation (r = -0.1**, df = 29) was detected between protein content (%) and yield (Kg Ha^{-1}).

The current study identified soybean genotypes that can be used to improve the nutritional traits of soybean in Uganda and across the East African region. The study also reported a low diversity in the studied germplasm pool that can lead to genetic erosion of the existing germplasm pool. Therefore addressing these challenges and developing soybean varieties with the desirable traits, requires diversification of the genetic background of the current germplasm pool by incorporating new genetic backgrounds from other countries. Furthermore, the study identified SNP markers that are associated with both total protein and oil content that will hasten the process for the development of soybean varieties with improved nutritional traits in Uganda and across the East African region. The highest yielding and stable genotypes BSPS 48A-9-2, BSPS 48A-31 and Nam II × GC 44.2 are recommended for further evaluation under farmers' production conditions for selection and release as new soybean varieties in Uganda.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Origin, Distribution and Diversity of Soybean

Soybean (*Glycine max* (L.) Merrill) is an important crop believed to have originated from China. The Yellow River region in China is considered as the origin of soybean based on the existence of a great number of wild relatives and the earliest record of soybean in China (Hymowitz and Kaizuma, 1981; Wang and Wang, 1992). In an attempt to confirm the origin of soybean, Han *et al.* (2015) sequenced more than 50,000 targeted genomic regions of *Glycine max, Glycine soja* and *Glycine gracilis* and observed two insights about the origin of soybean. First, *Glycine gracilis* is a transition species derived from the evolutionary process of domesticated soybean (*Glycine max*) from *Glycine soja.* Secondly, the Huang-Haui valley in Central China, the region between the Yellow river and Huai river is the most likely location where soybean originated from. Furthermore, the accessions from Huang-Haui valley showed greater genetic diversity than those from North eastern and Southern China.

The first domestication of soybean occurred in the eleventh century in China (Lance and Garren, 2005), from where it spread to other parts of the world (Probst and Jude, 1973). From China, the crop was introduced into Europe, America, and later to Africa. The movement of soybean from China to other parts of the world has been confirmed by several studies showing higher diversity of accessions from China than accessions from other countries (Cui *et al.*, 2001; Liu *et al.,* 2017; Mulato *et al.*, 2010; Qiu *et al.*, 1997). Soybean is believed to have reached Africa in the late 1800s starting from Egypt in 1858; followed by Tunisia in 1873 and then Algeria in 1880. In East Africa, soybean was first cultivated in Tanzania in 1907 and later spread to Kenya and Uganda in 1909 and 1913 respectively (Bashaasha, 1992; William and Akiko, 2009). Cultivation of soybean in Uganda started in the central parts of the country and spread to other regions later. Soybean trials started in the late 1930s at Bukalasa and Serere and by the 1940s promising varieties were multiplied for distribution to farmers, with large-scale production initiated in Buganda then to Toro, Bunyoro, Busoga, and finally the Lango region (Rubaihayo and Leakey, 1970).

The process of domestication from *Glycine soja* to *Glycine max* has led to the loss of soybean diversity. A study by Hyten *et al.* (2006) reported a 23% loss in nucleotide diversity and a 21% loss of rare alleles as a result of domestication and intense selection toward the elite varieties. Another study by Zhao *et al.* (2015) reported that genetic diversity was reduced by 37.5% in the early domestication process and 8.3% in later genetic improvements. Previous studies have shown greater diversity in wild relatives of soybean (*Glycine soja* and *Glycine gracilis*) than the domesticated soybean (Han *et al.*, 2015; Keim *et al.*, 1989). Additionally, this observed low diversity could among the domesticated soybean genotypes is due to selections for few traits desired by most soybean breeding programs (Tefera 2011; Tukamuhabwa *et al*., 2019) and sharing of parentage due to germplasm sharing among the different breeding programs (Halewood *et al*. 2020). Genetic diversity in plants is the basis for survival and crop improvement. High genetic diversity in the existing germplasm provides an opportunity for plant breeders to develop new and improved varieties with desirable traits, which include both farmer-preferred traits like high yield potential and, large seed and breeder-preferred traits like pest and disease resistance (Bhandari *et al*. 2017). With so many tropical soybean germplasm that have been collected from different sources, genetic diversity among these germplasm is not known, hence the need for this study.

1.2 Importance of Soybean

Soybean constitutes about 40% protein, 20% oil; a highly nutritious legume that has enormous potential to improve dietary quality for people throughout the world (Liu, 1997). Soybean oil is the most widely consumed vegetable oil worldwide and can be used for processing several food products and animal feeds (Friedman and Brandon, 2001; Singh *et al.*, 2008). Overall, during the last two decades, soybean oil has recorded very fast growth rates in terms of global consumption compared to the other oils and fats (WWF 2014). It is expected that by 2050, production will double (WWF 2014). Additionally, soybean oil can be used in bakery and manufactured foods such as baked bread, crackers, cakes, cookies, and pies, etc. More recently, with the advance of technology, more soybean oil is being utilized for industrial purposes (18%), at the expense of a decreased percent of soybean oil used for baking and frying fats (25%). Some of the newly developed products in the industry using soybean oil as a component include biodiesel**,** polyols (for the production of polyurethane products such as foams and resins), renewable bio-lubricants,

renewable plasticizers and soy-based toner. The utilization of soybean oil in these products is expected to create environmentally friendly alternatives to conventional products and is also very cost-effective.

Around 75% of soybean worldwide is used for animal feed, especially for poultry and pigs (WWF 2014). The remaining soybean produced worldwide is eaten directly, mainly in Asian countries, and a small portion is used in the production of biodiesel (WWF 2014). Studies have shown that soybean and soy-based foods have numerous health benefits such as cholesterol reduction, improved vascular health, preserved bone mineral density, and reduction of menopausal symptoms in humans (Anderson *et al.*, 1999). In animal nutrition, soybean meal is a major source of high-quality protein and amino acids (Drago *et al*. 2011). In 2019, soybean represented 69% of world protein meal consumption and 28% of world vegetable oil consumption (SoyStat, 2019). Soybean improves soil fertility by fixing atmospheric nitrogen through the symbiotic association with rhizobia. The Nitrogen-fixing ability of soybean makes it a good crop for sustainable agricultural systems of tropical Africa that is characterized by infertile soils due to over cultivation and leaching and low fertilizer usage (Bationo *et al*. 2006; Sanginga and Woomer 2009). Additionally, predictions of future land-use allocation and production show that the soybean is poised to dominate future production across Africa (Foyer *et al*. 2019). Soybean has also been used to diversify production and as a rotational crop to improve the productivity of other crops being grown; especially cereals and root crops (Sindelar *et al*. 2015; Varvel and Wilhelm 2003). The dense canopy of soybean helps to maintain or improve soil fertility because it protects soil from the direct rainfall impact and extreme temperature fluctuations. The soybean crop disrupts the life cycle of several pests and diseases of cereals and cassava when grown as an intercrop (Lithourgidis *et al*. 2011; Pandey 1988). Soybean is also used to make numerous industrial products, including vegetable oil, soaps, cosmetics, resins, plastics, ink, crayons, solvents, and clothing (Chen *et al.*, 2012; Guzeler and Yildirim, 2016). The diverse use of soybean implies more diversity is required in the existing germplasm. Therefore this requires understanding the genetic diversity of the existing soybean germplasm.

With all the associated nutritional importance of soybean, little attention has been put on improving the nutritional traits of soybean in Uganda, with more emphasis on agronomic traits

like high yields, promiscuous nodulation, resistance to pests, and diseases. Furthermore, global soybean production is expected to increase rapidly as economic development leads to higher animal protein consumption, especially in developing and emerging countries like Uganda (WWF 2014). In the last decade, the economies of Argentina and Brazil have been greatly supported by soybean. In Argentina, soybean and soy products contributed nearly a third (a USD 20 billion value) of the country's USD 72 billion export value. This was achieved in the last decade when the government of Argentina made deliberate efforts to promote the crop as a food, an industrial and export commodity (Goldsmith 2008; Ridley and Devadoss 2015; Cattelan and Dall'Agnol 2018). Several factories have been established in Uganda that are interested in soybean varieties with improved nutritional traits, especially elevated protein and oil content (Tukamuhabwa *et al*., 2019). This is coupled with increased interest among farmers in using soybean meal in animal feeds (SNV, 2011). Hence the need to understand the nutritional traits among the many soybean germplasm in Uganda that have been collected from various sources.

1.3 Production of Soybean

Over recent decades, soybean production has undergone the greatest expansion of any global crop (Agralytica 2012). The largest expansion of soybean production occurred during the second half of the twentieth century, where production grew tenfold from 27 million tons (MT) in 1962 to 334 MT in 2019 (FAO 2019). It is expected that by 2050, production will double (WWF 2014). The world's leading producer is Brazil which produces about 34% of the world's soybean (114.3 million MT). The USA is second with 29% of soybean produced (96.8 million MT) while Argentina is third with 17% (55.3 million MT) and China, 5% (15.7 million MT). The remaining countries account for 16% (51.6 million MT) of the global soybean output (FAO, 2019). In Africa, total soybean production rose from 1.4 million MT in 2008 to 3.1 million MT in 2019; representing 1.0% of the world's production. The three leading African countries in soybean production are South Africa (1,170,345 MT), Nigeria (630,000 MT), and Zambia (281,389 MT) (FAO, 2019). Uganda is ranked 1st in East Africa with a production of 28,000 MT (FAO, 2019). In the case of Uganda, a report by the Uganda Bureau of Statistics in 2021, showed that the northern region produced 15,729 MT, the eastern region (5,809 MT), the western region (1,886 MT), and the Central region (192 MT). This report also showed that the leading districts in

soybean production in Uganda were Oyam (8,030 MT), Apac (3,225 MT), Tororo (2,180 MT), and Lira (2,045 MT) (UBOS 2021).

1.4 Constraints to soybean production

Although soybean production has increased in Uganda, the yield has remained low; about 1,200 kg ha^{-1}. This is very low compared to yields in other major producers in Africa like South Africa estimated at about 2,000 kg ha^{-1} (FAO 2018). The low yields are attributed to several factors including poor soil fertility, inappropriate management practices, low use of improved varieties, and attack by pests and diseases (Lubungu *et al.*, 2013). In the tropics, several production constraints such as poor adaptation and short seed longevity period have been reported (Lubungu *et al.*, 2013). Soybean rust disease caused by the fungus *Phakopsora pachyrhizi* occurs in all soybean production regions in the world (Rosa *et al*. 2015). Soybean rust is the most destructive foliar disease of soybean and can cause yield losses of over 80% (Rosa *et al*. 2015; Godoy *et al*. 2016). Sources of resistance have been reported and to date, six dominant genes have been identified (Rpp1 – Rpp6) (Rosa *et al*. 2015; Godoy *et al*. 2016). However these genes are not effective against all the populations of *Phakopsora pachyrhizi* and several studies have reported a breakdown in the resistance of soybean genotypes that were initially reported to be resistant to the pathogen. The situation of resistant-soybean genotypes being resistant to a few rust isolates implies that the soybean breeders have to rely on the few soybean genotypes with the specific resistance genes to develop rust-resistant soybean varieties. The breakdown of soybean rust resistance genes has been reported in other countries because of reliance on a few sources of resistance (Chander *et al*. 2019; Tukamuhabwa *et al*., 2019). Reliance on a few sources of resistance narrows genetic variability and limits progress in breeding for other desirable attributes, such as yield (Tukamuhabwa and Maphosa 2011). Therefore there is a need to assess the level of genetic diversity among the existing soybean germplasm to broaden the genetic base.

Groundnut leaf miners that previously were not a major production constraint are now a major challenge to soybean farmers in Uganda. A study by Namara *et al*. (2019) reported percentage grain yield losses caused by groundnut leaf miner on the different soybean varieties ranging from 37.3% to 65.7% and none of the released soybean varieties is resistant to the pest (Tukamuhabwa *et al*., 2019). This is because all six MAKSOY soybean varieties grown by more than 93% of

Ugandan farmers (Tukamuhabwa *et al*. 2016), were derived from only four parents - Nam II, Duiker, GC 00138-29 and TGX 1035-10E (Tukamuhabwa *et al.*, 2019, Appendix 1). Therefore understanding the genetic diversity among the existing soybean germplasm in Uganda is very critical for the development of more superior soybean varieties.

1.5 Soybean Breeding in Uganda

Soybean was introduced in Uganda between 1908 and 1913 (Bashaasha, 1992). Soybean breeding efforts in Uganda can be summarized in four phases;

Phase 1 (1913-1975)

This phase was characterized by evaluation of soybean germplasm imported from other countries and little or no hybridization. Additionally, yield trials were a major component of soybean research that resulted in the release of six varieties (S - 35, Congo 72, Clark 63, Bukalasa 4, No.7, and Kabanyolo 1) in 1975 (Bashaasha 1992).

Phase 2 (1976 - 2000)

In this phase germplasm importation and evalauation continued. Most of the germplasm was imported from IITA and the USA. By around 1990, soybean breeding and agronomic constraints were given first priority. At this moment two broad objectives were identidied. The first was to select and develop soybean varieties of medium maturity (100 -120 days) which are high yielding; non shattering; resistant to lodging, major pests and diseases; free nodulating; with good pod clearance and which can store well for at least 7 to 8 months. The second is to identify appropriate agronomic practices which will maximize the performance potential of commercial soybean production in Uganda. Variety development through hybridization also began within this phase. The outbreak of the soybean rust epidemic in 1996 in Uganda was a key highlight in this phase. The outbreak of soybean rust disease stimulated more soybean research and several crosses were generated that resulted in the release of three varieties (Nam 1, Nam 2, Namsoy 3) (Kawuki *et al.*, 2003; Tukamuhabwa *et al.*, 2019).

Phase 3 (2001 – 2010)

Selection for high yielding and rust-resistant soybean genotypes was key in this phase, that resulted in the release of four varieties (Namsoy 4M, Maksoy 1N, Maksoy 2N, Maksoy 3N)

(Tukamuhabwa *et al.*, 2019). This phase was also characterized by conventional breeding through generation of multiple crosses and several multi locational variety trials across Uganda.

Phase 4 (2011 – 2021)

This phased involves use of the molecular breeding approach. For example, the use of SSR markers to understand rust diversity and pyramid-specific resistance genes in one genotype resulted in the release of three varieties (Maksoy 4N, Maksoy 5N and Maksoy 6N) (Maphosa, 2013). Additionally, this phase also involves the use of SNP-based markers to understand soybean diversity and identify markers associated with nutritional traits (protein and oil content) in soybean (Obua *et al*., 2020a).

Since 2001, a lot of multi environmental trials have been conducted to assess yield stability, different agronomic traits, pest and disease resistance among different breeding lines across different locations in Uganda (Obua *et al.*, 2020b; Mukuze, 2019; Obua, 2013; Tukamuhabwa *et al.*, 2012a; Tukamuhabwa *et al.*, 2012b). However, most of these breeding lines are always progenies generated from crosses of a few selected parents. For example, all the MAKSOY soybean varieties grown by more than 94% of Ugandan farmers are progenies of only four parents, Nam 2, GC 00138-29, Duiker and TGX 1035-10E (Tukamuhabwa *et al*., 2019; Tukamuhabwa *et al*., 2016; Tukamuhabwa and Oloka 2016). This implies that released Ugandan soybean varieties have low genetic diversity. The low genetic diversity among the released varieties can lead to total elimination, especially with changes in climatic factors. For example, several soybean varieties that were initially resistant to soybean rust disease in Uganda succumbed to the disease and have became obsolete. Therefore, this calls for exploring the high diversity among different soybean germplasm to take care of the changes in the industry and climatic factors. While the multi environmental trials conducted in Uganda have emphasized yield stability, pest and disease resistance, and other agronomic traits like resistance to lodging, pod shattering resistance, and high pod clearance, very little has been done to improve nutritional traits of soybean in Uganda. Therefore there is need to have a clear understanding of the behavior of these nutritional traits in terms of profiles in the existing germplasm and genotype by environmental interactions.

1.6 Statement of the problem

Breeding efforts have been made in soybean crop improvement in Uganda since 1976 when crosses between different soybean genotypes were generated that resulted in the release of the current ten commercial soybean varieties. However, these soybean varieties share similar parentage because few parents were used in generating the crosses (Appendix 1). Sharing of similar parentage leads to a narrow genetic base among the soybean varieties and a risk of susceptibility to production constraints like pests and diseases. For example, Nam 1, Nam 2, Nam 3, Namsoy 4M and Maksoy 1N that initially were resistant to soybean rust disease have succumbed to the disease. With the current changes in climate, all these varieties could be wiped out because of the low genetic diversity among them. Therefore there is a need to unlock the genetic diversity potential in the existing Ugandan soybean germplasm, collected from different sources has never been studied.

Over the years, the focus in the Ugandan soybean breeding program has mainly been on the development of soybean varieties with improved yield, disease resistance, and other agronomic traits like resistance to logding, pod shattering resistance, and high pod clearance (Bashaasha 1992; Tukamuhabwa *et al.*, 2019). However, other traits such as nutritional properties have received very little attention. Moreover, the nutritional traits like protein, oil, and fatty acid profiles of the existing soybean germplasm are not known yet this is a requisite for the development of soybean varieties with improved nutritional traits. Hence the need to clearly understand the nutritional traits among the existing soybean germplasm in Uganda. Additionally, the effect of environmental factors on some of the desirable nutritional traits is not known when the same soybean genotypes are planted in different environments. Therefore there is a need to unravel the behavior of nutritional traits in soybean when grown in different environments.

The methodology for assessment of nutritional traits like total protein, total oil and fatty acid profiling are quite expensive and cannot be performed using visual assessment like other agronomic traits; which makes it extremely hard to track these traits in the breeding pipeline. The use of molecular markers is viewed as a cheaper and faster option for nutritional assessment (Sudaric *et al*. 2008). However, the association between molecular markers and the nutritional traits has not been conducted among the Ugandan soybean germplasm collection. Therefore this

study aims to develop molecular markers associated with protein and oil to significantly reduce the time required for the development and release of soybean varieties with improved nutritional traits.

1.7 Study Objectives

1.7.1 Main objective

To develop high yielding soybean varieties with improved nutritional traits to fight food and nutritional insecurity

1.7.2 Specific objectives

Specific objectives were to;

I. Assess genetic diversity and population structure of tropical soybean genotypes using single nucleotide polymorphic markers
II. Determine variability of nutritional traits (total protein, total oil and fatty acid content) in tropical soybean germplasm
III. Identify SNP based markers that are associated with nutritional traits (total protein, total oil and fatty acid)
IV. Determine protein and yield stability and protein-yield relationship in tropical soybean germplasm

1.8 Hypotheses

I. Tropical soybean germplasm have low genetic diversity and indistinct population structure because they share similar parentage
II. Tropical soybean germplasm have diverse levels of nutritional traits (total protein, total oil and oleic fatty acid) because they are from different origins
III. There are several SNP based markers associated with nutritional traits (total protein, total oil and fatty acid) in soybean
IV. Yield and protein content in soybean are sensitive to changes in environmental conditions and the two traits are positively correlated

1.9 Justification

Several efforts have been made to understand the genetic diversity and population structures of soybean germplasm; including *Glycine soja* (Siebold & Zucc.), the wild ancestor to the domesticated soybean (*Glycine max* (L.) Merr.). The findings from these diversity studies would be useful to the soybean breeders in selecting genetically distinct parents and identification of candidate genes for important agricultural traits that could be very important for crop improvement (Kofsky *et al.*, 2018). In Uganda, several soybean genotypes have been collected through the years from different sources that constitute the germplasm base. This germplasm is expected to have high diversity in terms of genetic diversity, population structure, and nutritional traits.

For a successful crop improvement program, a clear understanding of the level of variability of a trait of interest is key because it leads to a more efficient genetic manipulation process and hence the development of better crop varieties. For example, to develop soybean varieties with improved nutritional traits, it is very important to clearly understand the level of variability of these traits among the existing germplasm available in the breeding program. Additionally, the oil extraction plants established in Uganda now require soybean varieties with improved nutritional traits because of drastic changes in consumers' preferences.

Several SNP-based markers have been developed to assess different nutritional traits in soybean. Selection for nutritional traits in soybean is physically extremely challenging compared to the use of SNP-based molecular markers. In the current study, the identification of SNPs associated with nutritional traits is expected to significantly reduce the time for of development of soybean varieties with improved nutritional traits. Additionally, the eventual use of the SNPs in the breeding program greatly increases the efficiency of selection in the breeding pipeline. Therefore a clear understanding of the behavior of the nutritional traits when soybean is grown in different environmental conditions will help the farmers and processors determine the best environmental conditions to obtain maximum nutritional benefits from the soybean.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Assessment of Genetic Diversity and Population in Soybean

Genetic diversity is genetic variability present among individuals within a species while population structure is the presence of differences in allele frequencies between subpopulations within a population. Knowledge of genetic diversity, population structure of a collection of germplasm is very important for crop improvement, effective conservation and utilization of the available genetic resources (Hipparagi *et al.,* 2017). Soybean genetic diversity and relatedness like most plant species can be assessed by the differences in morphological traits, pedigree information, geographic origins, isozymes, and DNA markers. A considerable amount of literature has been published on the use of morphological traits as a technique of assessing genetic diversity and classifying existing soybean genotypes (Malek *et al.*, 2014; Ojo *et al.*, 2012; Perić *et al.*, 2014). This technique has been utilized for the preliminary grouping of germplasm before their characterization using more precise marker techniques. For example, Ningsih *et al.* (2019) used morphological characteristics of pods of 10 soybean germplasm and reported significant differences among the studied soybean genotypes. In another study conducted by Malek *et al.* (2014) to determine genetic variability, character association and genetic diversity of 27 soybean mutants and four mother genotypes, revealed significant differences among the mutants and mothers for nine morphological traits. Most scientific classification of plants still relies on morphological traits (Khalid *et al.*, 2010). Additionally, this technique is cost-effective, easy to score, requires less time and does not require much technical knowledge. The major disadvantages of morphological markers are that they may be limited in number and are influenced by environmental factors and the developmental stage of the plant (Roychowdhury *et al.,* 2014). Uncertain or incomplete data and possible errors in the pedigree information and origins of accessions, and the limitation of data provided by isozymes also limit their utilization in diversity studies (Li and Nelson, 2001; Wang *et al.*, 2006).

Among the different DNA markers, restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) have been used

extensively in soybeans to determine genetic diversity and population structure; each with its advantages and limitations (Chauhan *et al.*, 2015; Doldi *et al.*, 1997; Hipparagi *et al.*, 2017; Kumawat *et al.*, 2015; Li *et al.*, 2010; Liu *et al.*, 2017; Satyavathi *et al.*, 2006; Tantasawat *et al.*, 2011; Torres *et al.*, 2015). However, SSR and SNPs have been widely used in studying genetic diversity and population structure. Many studies have reported average polymorphism information content (PIC) values ranging from low to moderate. For example, Gupta and Manjaya (2017), reported average PIC value of 0.61, Hipparagi *et al.* (2017) 0.36, Liu *et al.* (2017) 0.28, Chauhan *et al.* (2015) 0.57, Kumawat *et al.* (2015) 0.48, Bisen *et al.* (2014) 0.20, Zhang *et al.* (2013) 0.39, Tantasawat *et al.* (2011) 0.60, Mulato *et al.* (2010) 0.63, and Satyavathi *et al.* (2006) 0.44. Being a self-fertilizing crop, soybean is expected to have low heterozygosity than open-pollinated crop species. Therefore most previous studies have reported low heterozygosity; Mulato *et al.* (2010) 0.03, Bisen *et al.* (2014) 0.05, Gupta and Manjaya (2017) 0.06, Liu *et al.* (2017) 0.07, Hipparagi *et al.* (2017) 0.11, and Zhang *et al.* (2013) 0.46. Similarly, genetic diversity observed in several studies is always low; especially in cultivated soybean. Bisen *et al.* (2014) reported genetic diversity of 0.23, Kumawat *et al.* (2015) 0.35, Hipparagi *et al.* (2017) 0.43, Satyavathi *et al.* (2006) 0.56. The low diversity previously reported in soybean is probably because of domestication. Infact Hyten *et al.* (2006) concluded that, during soybean domestication, 50% of the genetic diversity was lost. Many soybean diversity studies have been conducted out of tropical Africa. However, for tropical Africa, very limited research has been conducted to reveal the level of genetic diversity among the existing soybean germplasm. Additionally, the existing soybean germplasm were collected from different origins for different breeding goals, hence possessing varying traits. Therefore, the level of genetic diversity in the existing germplasm in tropical soybean germplasm is not known.

2.2 Nutrient Profile in Soybean

2.2.1 Soybean seed protein content

The soybean is the most important source of high-quality vegetable protein in the world today and has played an increasing role in human and livestock nutrition over the last few decades (Riaz 2001; Singh *et al.* 2008). Soybean protein is an ideal source of some of the essential amino acids used to complement cereal proteins. Soybean protein supplies all 9 essential amino acids

and provides many functional benefits to food processors and for healthy human and livestock nutrition (Lonnerdal, 1994). Goldflus *et al.* (2006) reported the following essential amino acids; Arginine (2.706%), Phenylalanine (1.926 %), Histidine (0.968 %), Ile (1.606%), Leucine (2.838 %), Lysine (2.310%), Methionine (0.405 %), Threonine (1.315 %) and Valine (1.662 %) while assessing different soybean samples collected from different Brazilian states. Similarly, the percentage of these amino acids were within the same range in another study by Carrera *et al.* (2011) who assessed amino acid composition of soybean seeds from 31 multi-environment field trials carried out in Argentina. At present, soybean proteins are more versatile than many other food proteins in various worldwide nutritional programs and the meal provides an important source of protein for animal feeds. Maestri *et al.* (1998) reported that protein content in soybean genotypes ranged from 33.1% to 44.8%. Qin *et al.* (2014) investigated the regional distribution of protein and oil compositions of soybean cultivars in china and found protein content ranging from 31.8% to 49.8%. Additionally, Akpagu *et al*. (2015) investigated the protein content of soybean samples from different states in Nigeria and reported a narrow protein range between 36.74% to 39.23%. Similarly, Ciabotti *et al*. (2016) assessed the chemical composition of soybean grain of varieties and advanced lines with different seed coat colours and reported a narrow protein range of 35.35% to 39.80%. Several factors have been reported to cause the variation of protein content among different soybean genotypes, these include genetic and environmental (Patil *et al*., 2017; Ojo *et al.*, 2002). Some of the environmental factors that have been reported to increase protein content in soybean include high nitrogen available (Vollmann *et al*., 2000), high temperatures (Ojo *et al.,* 2002) and low rainfall (Vollmann *et al*., 2000) during the grain filling stage.

Several methods have been suggested for the quantification of protein content in different food and feed samples. For example, Kjeldahl is a widely used method for determining nitrogen and protein contents in the food and feed industry (VELP Scientifica, 2015; Akpagu Francis et al., 2015). This method involves three major steps; digestion, neutralization and distillation. However, Kjeldhal method assumes that all the Nitrogren in the food/feed sample is protein-bound and the conversion/correction factor of 6.25 is used to determine total protein content. A wide range of other compounds, such as nitrate, ammonia, urea, nucleic acids, free amino acids, chlorophylls and alkaloids contain nitrogen that are not protein-bound (Imafidon and Sosulski,

1990; Maehre et al. 2018). Additionally, the conversion factor used in this method is based on an assumption that the general nitrogen content in food proteins is 16% (Maehre et al. 2018). However, these are quite rough assumptions as the relative nitrogen content varies between amino acids and amino acid composition varies between food proteins (Imafidon and Sosulski, 1990; Maehre et al. 2018). Another Nitrogen-based method for quatitifcation of total protein content in food/feed is Dumas method. In the Dumas method, the nitrogen is liberated in a gaseous form and is determined with a thermal conductivity detector, after removal of carbon dioxide and water aerosols (Maehre et al. 2018; VELP Scientifica, 2013). The Dumas method is fast and does not use chemicals, but is costly to set up and is not very accurate as it does not measure true protein (Hayes, 2020). More recently, UV spectrophotometric methods such as Lowry Methods have been applied as powerful methods for the quantitative detection of proteins in a complex food/feed matrixes, including soybean (Hayes, 2020; Ippoushi *et al.*, 2020; Krager, 2020; Jia *et al.*, 2019; Ippoushi *et al*., 2015; Winther *et al*., 2009).

2.2.2 Soybean seed oil content

Soybean has unparalleled quality oil properties that make it one of the best sources of vegetable oil for different purposes in the world. The profile of soybean oil has been modified depending on the end products of soybean oil to meet the different demands of the markets (Fehr *et al.*, 1992). Soybean oil is very popular because it is rich in Omega 3 and Omega 6. These fatty acids regulate lipid and cholesterol metabolism and improve cardio-muscular functioning in humans. Besides the high content of vitamin B in soybean makes digestion easier and prevents chronic digestion problems and constipation. Thus, refined soybean oil is widely used all over the world as a highly prized vegetable oil. However, the primary limitation of soybean oil is its low oxidative stability which reduces shelf life, flavor and durability at high temperatures (Clemente and Cahoon 2009; Madhujith and Sivakanthan 2018; Medina-juárez *et al*., 2011). To improve oxidative stability and undesirable flavours, soy oil is hydrogenated to reduce double bonds which are sites of oxidative attack that reduce stability, shelf life and increase off-flavor (Berghuis *et al.*, 2010; Pintauro *et al.*, 2005; Singh *et al.*, 2009; Wang *et al.*, 2019).

Ciabotti *et al.* (2019) investigated the chemical composition and lipoxygenase activity of soybean genotypes, with different tegument colours and found oil content ranging from 18.2% to 21.4%. Similarly, Ciabotti *et al.* (2016) reported a narrow range of 18.2% to 19.5% while

studying the chemical composition of soybean genotypes with different seed coat colours. On the other hand, Lee and Chou (2006) studied five black soybean genotypes and found oil values ranging from 10.1% to 21.6%. Additionally, Qin *et al.* (2014) investigated the regional distribution of protein and oil compositions of soybean cultivars in china and found oil content ranging from 14.2% to 22.8%. However, environment, genetics, management, and their interactions have been reported to cause the variability in oil content among different soybean genotypes. For example, Assefa *et al*. (2019) reported a significant increase in oil content when Nitrogen fertilizer was applied during the growing season.

The major extraction methods of total oil quantification in food and feed samples include Soxhlet Extraction Method (Soxhlet, 1879) and Bligh and Dyer (1959). Soxhlet technique is the oldest and the most universally accepted protocol for determining oil and lipid content of seeds, foods, and feeds. The method was developed by von Soxhlet in 1879 and it has been used for more than a century. Numerous studies have supported the fact that Soxhlet is a standard method. Currently, the Soxhlet is the essential reference that the performance of other techniques are compared. Ullah et al. (2011) reported the Soxhlet extraction is the oldest method for performing solvent extraction and it is the official method used by both AOAC and American Oil Chemists' Society (AOCS). However, Bligh and Dyer method of extraction was developed in 1959. This method was developed as a simple, rapid and effective method of determining total oil content in food and feed samples (Iverson et al., 2001; Reis et al., 2013). This method involves the partitioning of lipids into a binary mixture of chloroform and methanol in different ratios.

2.2.3 Soybean fatty acid profile

Elite soybean varieties produce seed that average 8.4% to 17.0% palmitic, 2.8% to 12.6% stearic, 19.2% to 36.8% oleic, 42.1% to 62.6% linoleic and 4.1% to 6.8% linolenic acids (Erickson *et al.*, 1988). However a study by Qin *et al.* (2014) reported small differences in these fatty acid profiles; palmitic (9.26–14.52%), stearic (2.63–7.12%), oleic (16.19–39.88%), linoleic (38.92–58.99%) and linolenic acid (5.11–11.68%). Like most vegetable oil crops, soybeans contain a high level of linoleic acid. Additionally, one major difference between soybean oil and other edible vegetable oils is the high level of linolenic acid, which compromise oil flavor and stability (Wilcox and Cavins 1985). Oil with high oxidative stability is very desirable because when used

increases the shelf-life of food products. Stability of the oil refers to the amount of time before the oil becomes rancid due to oxidation (Mercer *et al.*, 1990). Increasing oleic acid in soybean oil is important because of increased oxidative stability and health benefits. Therefore, for vegetable oil and food production purposes, soybean oil that is high in oleic acid content or low linolenic acid content is preferred. Advances have been made in the development of soybean lines that are inherently high in oleic acid and low in linoleic and linolenic acids to improve the oil quality. Increasing the oleic acid content of soybean oil would decrease the total saturated fatty acid content and increase the oil quality for human consumption (Hayakawa *et al.*, 2000).

Oleic acid is a monounsaturated fatty acid that facilitates improved health and oxidative stability for increased oil shelf life, flavor, durability and cold flow performance (Liu *et al*. 2020; Pham 2011). Increasing oleic acid to more than 60% will improve the edible and industrial applications of soybean oil (Pham 2011). High oleic soy oil reduces the need for hydrogenation and eliminates trans-fats and would improve soy-diesel, lubricants and hydraulic oils (Fehr and Hammond 1998). A diet in which fat consumption is high in oleic acid, like olive oil, is associated with reduced cholesterol, arteriosclerosis, and heart disease. High oleic acid also extends the utility of soybean oil at high cooking temperatures and also increases soybean oil functionality in pharmaceuticals and cosmetics. The success of in the identification and profileing of lipids is critically dependent on the efficiency of the total oil extraction step. Fatty acids methyl esters (FAMES) were determined using Gas Chromatography-Mass Spectrometry (GC – MS). Advances in mass spectrometers including incrased sensitivity, higher mass accuracy, and higher scan speeds have resulted in the increased populiarity of MS as detection tool for fatty acids in recent years (Reis et al., 2013).

2.2.4 Correlations for nutritional traits in Soybean

The development of soybean varieties with improved nutritional traits has been further affected by the negative correlation that exists among the different traits (Maestri *et al.*, 1998; Marega *et al.*, 2001). Therefore during the selection of soybean genotypes for a particular food application or a particular breeding program, it is important to understand the relationships that exist among these quality attributes. Several studies have reported positive and negative correlations that exist among the different nutritional traits in soybean. For example several studies have reported negative correlation between protein and oil content (Chung et al. 2003; Hyten et al. 2004;

Panthee et al. 2005; La et al. 2014). Additionally, significant negative correlation between oleic acid and palmitic acid has been reported in many studies (Ahire 2012; Qin *et al.*, 2014; Rebetzke *et al.*, 2001). Additionally, other studies have reported a positive correlation between palmitic acid and other long-chain fatty acids derived from it such as stearic and oleic fatty acid (Ahire 2012; Bachlava *et al.*, 2008; Rahman *et al.*, 2003; Stoltzfus *et al.*, 2000). Stearic acid is derived from palmitic acid through modification that results in desaturation and elongation during the fatty acid synthesis process. The negative correlation between oleic acid and saturated fatty acids like palmitic coupled with the positive correlation between palmitic acid and stearic acid offers the opportunity to develop soybean varieties with improved oil quality. Hence selection for higher oleic acid will result in lower palmitic and stearic acids hence improvement in the quality of the soybean oil (Pham, 2011; Qin *et al.*, 2014). In soybean, total oil content has been reported to exhibit a strong negative correlation with oleic acid (Bachlava *et al.*, 2008; Rani *et al.*, 2007). The significant negative correlation between oleic acid and total oil suggests that it would be difficult to develop soybean varieties with high oil content as well as high oleic acid. This implies that it is difficult to develop a soybean variety with high oleic acid and oil content using conventional breeding.

2.3 Genetic basis for nutrient content of Soybean

Several studies have reported that most of the nutritional traits in soybean are complex, controlled by multiple genes/ quantitative trait loci (QTLs), and affected by the environment and genotype × environment interaction (Liu *et al*. 2020; Yao *et al*. 2020; Patil *et al*. 2017). The genetic control of seed oil content in soybeans was found to include primarily additive gene action (Wilcox 1989; McKendry *et al*., 1985), but some evidence for epistatic effects have been reported (Wang *et al*. 2020; Zhaoming *et al*., 2017). Heritability values reported for seed oil content ranged from moderate (51%) to high (92%) (Sobko *et al*. 2020; Yao *et al*. 2020; Wiggins 2012; McKendry *et al*., 1985). Additionally, Taliercio *et al*. (2017) reported non-additive changes in gene expression in the F_1 hybrids relative to the parents for seed protein content in soybean seed.

Heritability of soybean oil fatty acids have been reported to range from 65.8% to 91.2% that early generation selection can be effective (Yao *et al*. 2020; Li *et al*. 2018; Li *et al*. 2017;

Bachlava *et al*. 2008; Panthee *et al*., 2006; Gesteira *et al*. 2003). Furthermore, no significant cytoplasmic effects have been reported for stearic acid content (Rahman *et al*., 1997; Bubeck *et al*., 1989) and palmitic acid content (Bubeck *et al*., 1989). In another study to understand the inheritance of elevated palmitic acid in soybean seed oil indicated no maternal effect or dominance for palmitic acid content (Narvel *et al*., 2000). Most of these genetic studies for fatty acids were conducted using mutant soybean genotypes. Therefore improvement of fatty acids in soybean oil is not easy to achieve through conventional breeding.

2.4 Breeding for Improved Nutritional Traits of Soybean

Development of improved soybean varieties has mainly been through conventional breeding that mainly focuses on hybridization, selection, and evaluation of improved lines. Several selection methods have also been used in soybean for improvement of agronomically important traits, including nutritional traits. For example, recurrent method was used for protein (Xu 1990; Brim and Burton 1979). It has been employed for making genetic improvements to oil (Burton, J.M., Brim, C.A. 1981) and oleic fatty acid (Carver et al., 1986; Burton et al., 1983; Burton et al. 2006). Mass selection was used for oil (Burton and Brim 1981). Half-sib family selection was used for oil quality (Carver et al., 1986). Recent biotechnological tools have complemented conventional plant breeding producing an accelerated improvement to soybean. Various biotechnological tools, such as molecular breeding, marker-assisted selection, mutation breeding, genetic transformation have played a major role in developing varieties with improved nutritional traits. The most effective genetic method for increasing oleic fatty acid and decreasing linolenic fatty acid has been through the use of mutagenesis. Examples of such high oleic acid mutants include M23 (Rahman et al., 1994), and KK21 (Anai et al. 2008), S117N (La et al. 2014), 17D (La et al. 2014) and P137R (La et al. 2014). More recently, transgenic soybean with elevated oleic fatty have been developed (Zhang et al. 2014).

2.4 QTL mapping for nutrient content of Soybean

The two main methods for detecting significant marker or quantitative trait loci (QTL) are linkage mapping and genome-wide association study (GWAS). Linkage mapping relies on a population developed specifically to map the trait of interest and uses bi-parental breeding

strategies. This allows the researcher to select parental genotypes and control the relatedness of the population. Several population types and methods can be used for QTL detection with contrasting power of resolution. However, this method of identification of significant markers or QTLs is only effective for the developed mapping population. On the other hand, genome-wide association study (GWAS) is a complementary approach to linkage mapping. While both methods rely on the correlation between DNA marker alleles and the phenotypic expression of a trait of interest, association studies provide relatively higher mapping resolution in terms of defining the genomic position of QTL. The detection of QTL depends on the level of Linkage Disequilibrium (LD) between a causal mutation and physically linked markers. The higher the degree of association between marker alleles and the phenotypic variants, the greater the likelihood that the observed phenotypic trait is linked to the marker. SoyBase (http://soybase.org, accessed 05 April 2021) reported 834 and 594 oil and protein QTLs, respectively, across all the 20 chromosomes.

Recently, GWAS has been widely used for revealing the genetic basis of complex trait variation in soybean (Sonah *et al.*, 2015). For example, a study by Hwang *et al.* (2014) identified 40 SNPs located in 17 different genomic regions in 10 chromosomes for protein content and 25 SNPs located in 13 regions on 12 of the 20 chromosomes for oil content in diverse soybean accessions. Another study that was conducted for protein and oil content on USA soybean accessions identified SNPs with strong signals on chromosomes 20 and 15 (Bandillo *et al*., 2015). The same study further identified three candidate genes for protein and oil content in the chromosome 20 region (Bandillo *et al.*, 2015). Another study discovered 29 SNPs located on ten different chromosomes that are significantly associated with protein, oil and five fatty acids in wild soybean accessions (Leamy *et al.*, 2017). Therefore the application of GWAS in soybean remains to be explored, especially for nutritional traits that are very difficult and expensive to assess. Very little or no research has been conducted in tropical Africa to identify SNP markers that are associated with nutritional traits in soybean.

2.5 Genotype × environment interactions and Stability analysis for yield and nutritional traits in soybean

Genotype × environment interactions (GEI) in soybean like in many other important crops have been widely studied, especially for seed yield. Genotype by environment interaction is the failure of two or more genotypes to achieve the same relative performance in two or more environments (Baker 1988; Bowman 1972). GEI is attributed to differences in sensitivity, which means that a given environmental difference affects some genotypes more than others (Falconer 1989). The presence of significant genotypes by environment interaction for quantitative traits such as seed yield and nutritional traits can reduce the usefulness of subsequent analysis, restrict the significance of inferences that would otherwise be valid, and seriously limit the feasibility of selecting superior genotypes (Flores *et al*., 1998).

One approach to solving such a problem is that most breeders look for varieties that have good mean performance over a wide array of environments. Such an approach is reasonable if there is no GEI, but in most cases there is interaction. Some genotypes can have high performance in few environments and very low performance in other environments, showing better mean performance across environments. But few genotypes may have average performance but remain stable over wider environments. Stability analysis is important for understanding the response of different genotypes to varying environments and for the identification of stable and widely adapted genotypes. Therefore, the performance stability of a genotype across a range of production environments is very important for variety recommendations. Another approach that has been used by several breeders to resolve the challenge of GEI is that of grouping environments into mega-environments by way of genotype response. A mega-environment is defined as a subset of locations that consistently share the best set of genotypes and the regions are relatively homogenous with similar biotic and abiotic stresses and cropping system requirements (Yan *et al*. 2007).

Most of the previous GEI studies have focused on seed yield (Obua *et al*. 2020; Cheelo *et al*., 2017; Mulugeta *et al*. 2013; Tukamuhabwa *et al*.,2012a; Tukamuhabwa *et al*. 2012b), agronomic traits (Zhe *et al*. 2010). However, there has been little research on the GEI of nutritional traits. Therefore, there is a need to investigate the relationship between yield and stability of these nutritional traits across different environments. There is a need to have a clear

understanding of the relationship between grain yield and nutritional traits since it is one of the most important traits in most soybean breeding programs. In an attempt to understand the yield and stability of protein content in soybean, a study by Sommers and Mchale (2015) revealed that soybean genotypes with high protein yield exhibit stable protein yields across variable environments. Another study by Gurmu *et al*. (2010) discovered that Clark-63K had the highest crude protein content and also very stable while IPB-144-81(p) and AFGAT had high crude protein content but very unstable. The same study identified Haddee-1 and Braxton with high oil content and showed stable performance while TGX-297-6f-1 had high oil content but unstable across the environments. A stud by Chaudhary and Wu (2012) based on AMMI analysis, identified HEFTY H15Y12 as the most widely adapted soybean genotype for oil content and grain yield under the conditions of Eastern South Dakota.

A number of statistical approaches have been developed to analyze Multi Environmental Trial (MET) data. One common approach used by plant breeders is the Genotype and genotype-by-environment interaction (GGE) biplot analysis. The method simultaneously displays the genotype main effect (G) and the genotype by environment interaction (GEI), and has been reported to visually address many questions related to multi-environment trial (MET) data (Yan, 2001; Yan *et al.*, 2007; Yan and Kang, 2003). A review by Yan *et al.* (2007) reported that GGE biplot is superior in mega-environment analysis and genotype evaluation. Therefore, the GGE biplot technique has been widely used in soybean to assess GEI. A study conducted in four different locations of Ethiopia for two consecutive years using thirty-two genotypes showed that there was GEI and was crossover in nature (Mulugeta *et al.*, 2013). The same study also identified three genotypes that had both high mean yield and high stability performance across the test environments (Mulugeta *et al.*, 2013). In another study by Adie *et al.* (2014) who evaluated 10 black seeded soybean genotypes in 16 locations revealed that the genotype W9837 × Cikuray-66 was stable and recommended for release as a new high-yielding variety. In Zambia, a MET analysis reported that the best genotype for general adaptability was the variety TGX 1988-22F. This genotype was stable across all the locations with high yields and average stability (Cheelo *et al.*, 2017). For the case of Uganda with quite a diverse agro-ecological zones, no comprehensive multi environmental soybean trial has been conducted in more than six seasons and eight locations. Most of the multi environmental trials have been conducted for three

seasons or less, in a maximum of five locations (Tukamuhabwa *et al.*, 2012a; Tukamuhabwa *et al.*, 2012b). This is the first GEI study in Uganda involving soybean seed yield and nutritional traits.

CHAPTER THREE

3.0 GENETIC DIVERSITY AND POPULATION STRUCTURE OF TROPICAL SOYBEAN (*GLYCINE MAX* (L.) MERRILL)

3.1 INTRODUCTION

Soybean is among the most valuable crops in Uganda and across the East African region due to the high protein content that makes it an important ingredient in the diets of the people and livestock (Tukamuhabwa, 2001). Several soybean processing plants have been established in Uganda and across the East African region with large processing capacities to develop different products from soybean. This new development has motivated the farmers to produce more grains to supply these plants (Tukamuhabwa *et al.*, 2019). The three leading African countries in soybean production are South Africa (1,540,000 MT), Nigeria (758,033 MT), and Zambia (302,720 MT) (FAO 2018a). Uganda is ranked 1st in East Africa, with a production of 29,000 MT (FAO 2018a). Hence soybean production and consumption have led to increased farmers' income, improved food and nutrition security, and poverty eradication at the rural household level (Ssengendo *et al.*, 2010; SNV 2011; Tukamuhabwa and Obua, 2015). Accordingly, soybean has the potential to contribute to poverty alleviation in Uganda and across the East African region.

Despite the contribution of soybean to smallholder farmers in Uganda and across the East African region, the development of new varieties has been hindered by the low genetic diversity of the crop that has been observed by several authors (Kumawat *et al.*, 2015; Torres *et al.*, 2015; Maldonado *et al.*, 2016; Gupta and Manjaya, 2017; Liu *et al.*, 2017). Kumawat *et al.* (2015) investigated the diversity of 82 Indian soybean accessions using SSR markers and identified three major clusters. In another study, Torres *et al.* (2015) found that both Principal Component Analysis (PCA) and STRUCTURE, clustered 191 soybean accessions in Brazil into two groups. Additionally, Gwinner *et al.* (2017) in another study that aimed at understanding the genetic diversity and population structure of 77 commercial soybean varieties in Brazil using 35 SSR markers, reported low genetic diversity in soybean germplasm.

To assess the genetic diversity of soybean and other plants, various methods based on morphological markers, geographic origins, pedigree information, isozymes, and DNA markers have been applied (Dayaman, 2007; Appiah-Kubi, 2012; Ojo *et al.*, 2012; Malek *et al.*, 2014; Villela *et al.*, 2014). The use of morphological traits has remained a powerful taxonomic tool for preliminary grouping of germplasm before their classification using more precise marker techniques. Several studies involving the classification of plants still rely on the use of morphological traits (Khalid *et al.*, 2010). Additionally, the use of morphological markers in classification is easy to score, cheaper and fast. However, the disadvantage of using morphological markers is that it's less robust compared to most molecular markers and outcomes can be easily influenced by environmental factors. In the case of pedigree information, limitations such as uncertain and incomplete data errors are likely, while for isozymes, chances of limited data are more prominent (Li and Nelson, 2001; Wang *et al.*, 2006).

So far, DNA markers remain the most precise method of genetic diversity analysis and have been complemented with morphological trait analysis. Among different DNA markers, random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs) have been widely used in understanding the diversity of soybean; each with its advantages and disadvantages (Doldi *et al.*, 1997; Tantasawat *et al.*, 2011; Ojo *et al.*, 2012; Ren *et al.*, 2013; Singh *et al.*, 2013; Chauhan *et al.*, 2015; Torres *et al.*, 2015; Chen *et al.*, 2017; Hipparagi *et al.*, 2017). SSR markers have been widely used to determine genetic diversity in many crops because they are easy to use, have a reasonably low price, and high level of polymorphism (Vignal *et al.*, 2002). However, recently SNP markers have been widely utilized for the assessment of diversity in plants because they occur much more frequently in the genome than SSR markers, and their genotyping can be easily automated (Mammadov *et al.*, 2012). In the current study, Genotype By Sequencing (GBS) technology was used to study a collection of 89 tropical soybean germplasm collected from different countries. Therefore, the objective of the current study was to understand the genetic diversity and population structure of tropical soybean germplasm using SNP markers. Since the genotypes

included in the current study are parental lines, land races, released varieties, and advanced lines, they are representative of the existing Ugandan germplasm collected from different sources.

3.2 MATERIALS AND METHODS

3.2.1 Plant materials

A total of 89 tropical soybean genotypes were used in this study, 45 from Uganda, 13 from Japan, 6 from the USA, 12 from World Vegetable Centre in Taiwan, and 13 from Seed Co Zimbabwe. The interest was mainly on Ugandan genotypes. Genotypes from the USA, Japan, Taiwan (AVRDC) and Zimbabwe (SeedCo) were added to broaden the genetic base and putting Ugandan genotypes into the global perspective of soybean diversity. Therefore the studied soybean genotypes included collections from different sources and are expected to possess high genetic diversity (Appendix 2).

3.2.2 DNA extraction, Determination of DNA Quality and Quantity

Seeds from each genotype were grown under controlled greenhouse conditions at Biosciences eastern and central Africa - International Livestock Research Institute (BecA - ILRI) Hub, Kenya. Twelve days after germination, one young leaf from one plant from each genotype was harvested and DNA extracted using ZR Plant / Seed DNA MiniPrep^{TM} according to the manufacturer's protocol.

The DNA quality was first checked on 0.8% (w/v) agarose gel in 1 X Tris-acetate EDTA buffer and run at 80V for 45 Minutes. The run gels were photographed using GelDoc-It^{TM} Imager (UVP) and the picture image was interpreted for DNA quality. The DNA was quantified using Thermo Scientific Nanodrop 2000C Spectrophotometer and stored at 4 °C.

3.2.3 SNP Genotyping

Genotyping of the soybean genotypes was conducted by Diversity Arrays Technology (DArTSeq^{TM}) in Australia using the Illumina HiSeq 2500. The samples were genotyped following an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation-sensitive restrictive enzymes before sequencing on next-generation sequencing platforms (Kilian

et al., 2012). The soybean reference genome and annotation were downloaded from ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Gmax. The sequence data generated were then aligned to the soybean reference genome sequence, Soybean_v7, to identify single nucleotide polymorphisms (SNPs) markers.

3.2.4 Data analysis

GBS data for a total of 16,688 SNP loci distributed across the 20 Soybean chromosomes was received from Diversity Arrays Technology (DArTSeq^{TM}), Australia. The genotype data was initially filtered using a minor allele frequency (MAF) of 0.01 and a minimum count of 90% of the sample size using TASSEL v.5.2.43 software (Bradbury *et al.*, 2007), which gave 7,962 polymorphic SNPs. Genetic distance was calculated between a pair of inbred lines in the dataset using the identity by state similarity (IBS) method implemented in TASSELv.5.2.43. A marker-based kinship matrix was then calculated between a pair of inbred lines in the dataset using TASSELv.5.2.43.

Using the model-based clustering approach implemented in the software package STRUCTURE v2.3.4, population structure was estimated (Pritchard *et al.*, 2000). To estimate the posterior probabilities (qK) a 100,000 burn-in period was used, followed by 100,000 iterations; with the hypothetical number of subpopulations (k) ranging from 1 to 10, with 10 replicates for each K. The number of subpopulations was determined when Δk reached its highest value (Evanno *et al.*, 2005). The Delta K was calculated for each value of K using Structure Harvester(Earl and vonHoldt, 2012; Evanno *et al.*, 2005). A line was assigned to a given cluster when the proportion of its genome in the cluster (qK) was higher than a standard threshold value of 70 %. For the chosen optima value of K, membership coefficient matrices of replicates from STRUCTURE were integrated to generate a Q matrix using the software CLUMPP (Jakobsson and Rosenberg, 2007) and the STRUCTURE bar plot was drawn using the DISTRUCT software (Rosenberg, 2004). The principal coordinate analysis was performed based on the genetic distance matrix using the Dissimilarity Analysis and Representation for windows (DARwin) v.6.0.013 (http://darwin.cirad.fr). Analysis of molecular variance (AMOVA) was performed using GenAlEx V6.5 software.

3.3 RESULTS

3.3.1 Genotype Diversity analysis

A total of 16,688 SNP markers were identified in the 89 genotypes of soybean; of which 7,962 were polymorphic and non-redundant, with greater than 5% minor allele frequency (MAF) and missing data lower than 20% were used for subsequent analysis. These 7,962 SNPs detected a total of 15,924 alleles, with each SNP detecting two alleles as expected. The average polymorphism information content (PIC) was 0.27, ranging from 0.01 to 0.50 and heterozygosity ranged from 0.0 to 0.35 for individuals and 0.0 to 0.8 for markers (Figure 1).

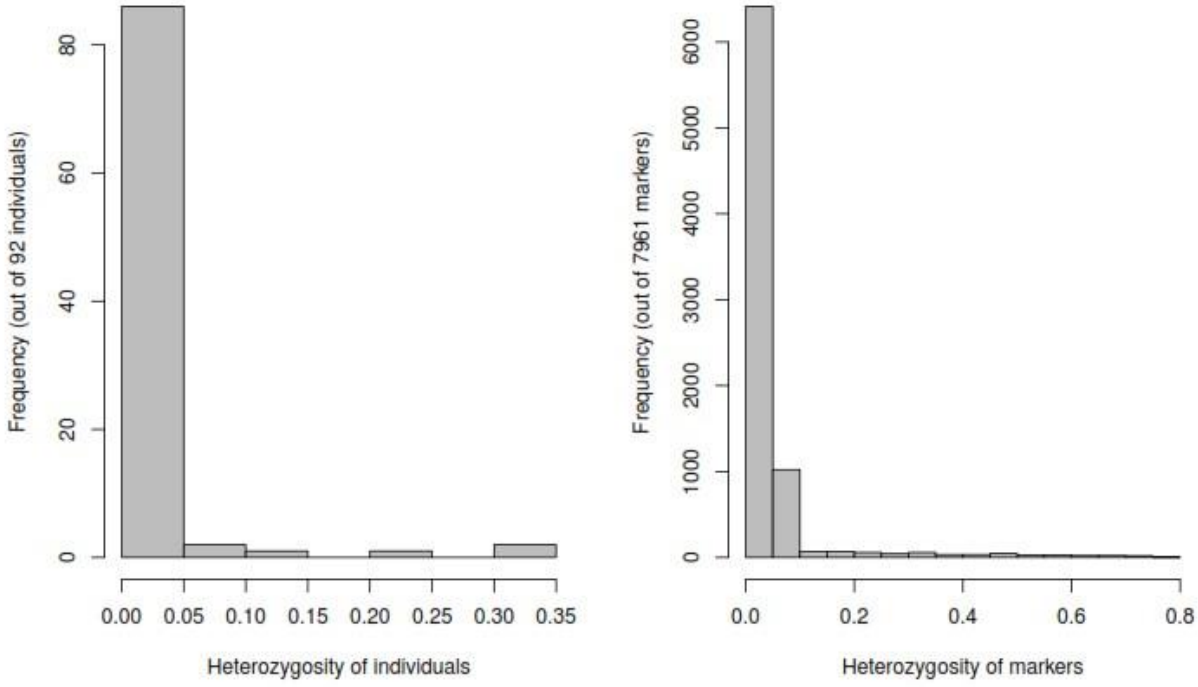


Figure 1: Levels of heterozygosity of individual soybean genotypes and SNPs markers

3.3.2 Genetic distance and relationship

The average Roger genetic distance within the study population was 0.34. From a total of 89 genotypes, 18.1% of the distance values were between 0.0 and 0.05 while 20.7% were between 0.35 and 0.40 (Figure 2). Relative kinship reflects the approximate degree of identity between two given genotypes. For combined analysis of all 89 genotypes, the kinship coefficients ranged from 0 to 1.04, with an overall average of 0.51; only 1.6% of the pairwise kinship estimates had

values of 0.0 – 0.05 while 76.1% had values ranging from 0.5 – 0.550, indicating that most of the genotypes were in one way or another related (Figure 3).

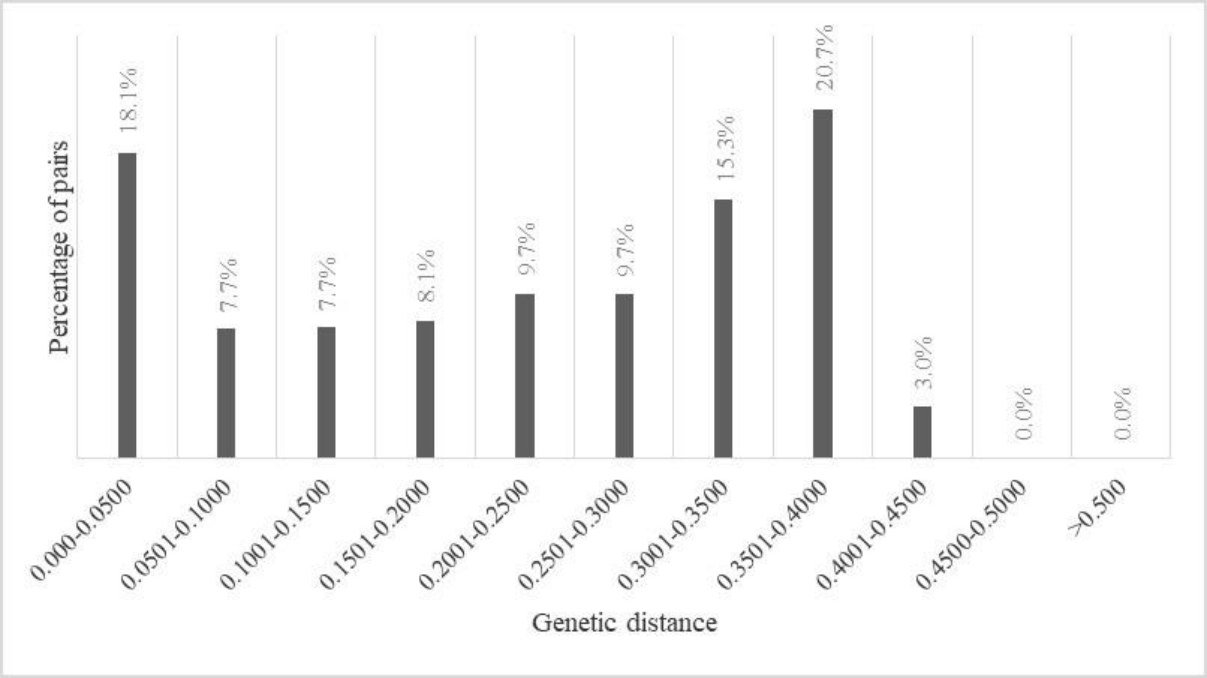


Figure 2: Distribution of pairwise Roger's genetic distance calculated for 89 soybean genotypes

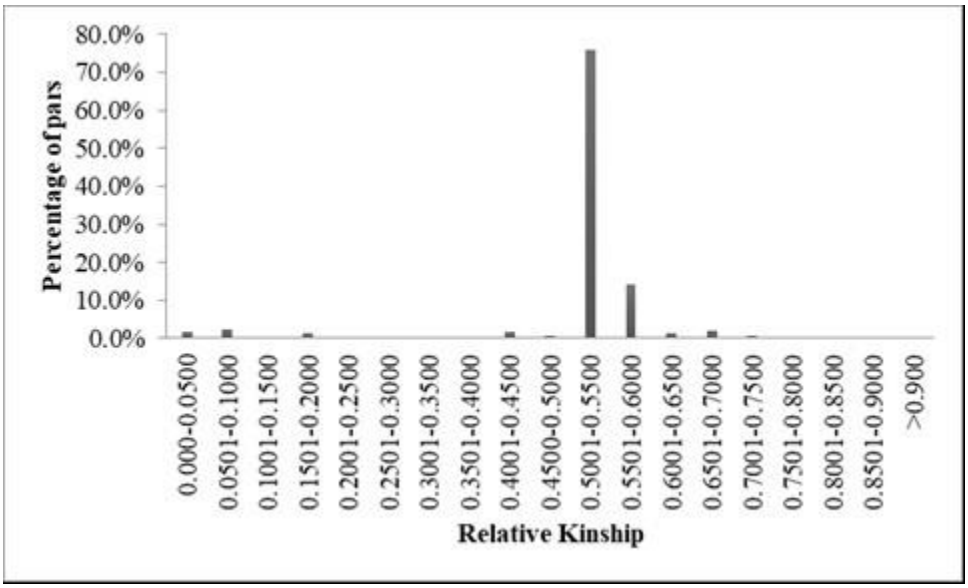


Figure 3: Distribution of pair-wise kinship coefficients among 89 soybean genotypes from different sources based on 7,692SNPs

3.3.3 Population structure analysis

The estimated log probability of the data (LnP(D)) increased continuously with increasing K (number of groups or populations). The ad hoc statistic ΔK showed a higher likelihood value at K = 2 as the highest level of structure (Figure 4). This pattern was also observed in the population structure, where two groups were formed (Figure 5). However, the value at K = 3 was informative than all the ΔK values (Appendix 3)

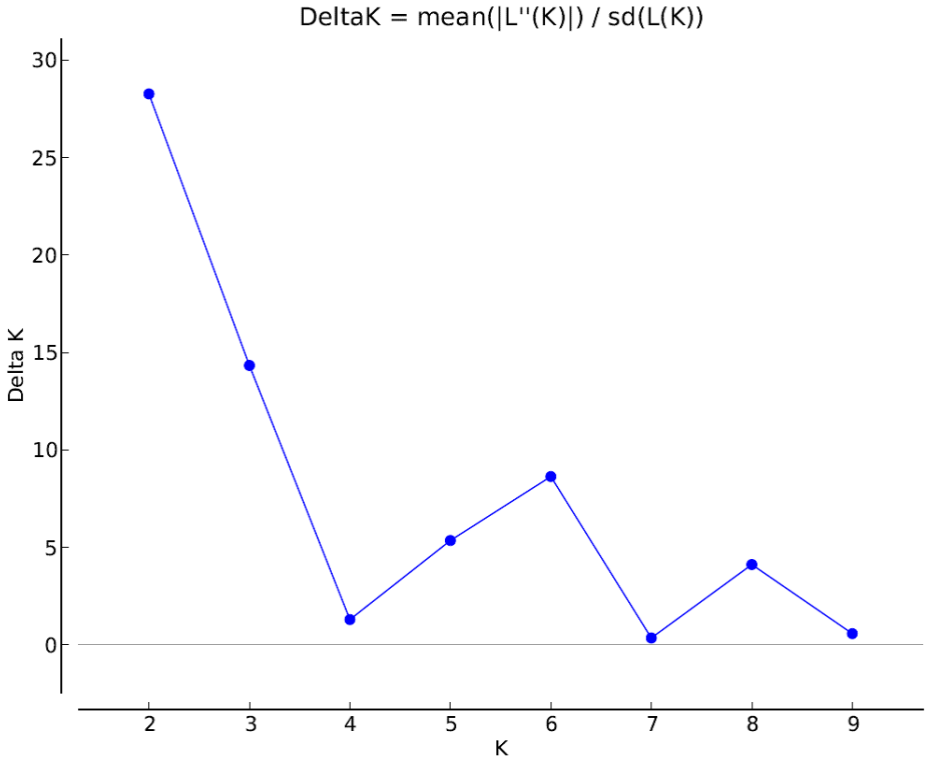


Figure 4: Estimation of the population using LnP(D) derived ΔK with K ranged from 1 to 10 with 7,692SNPs

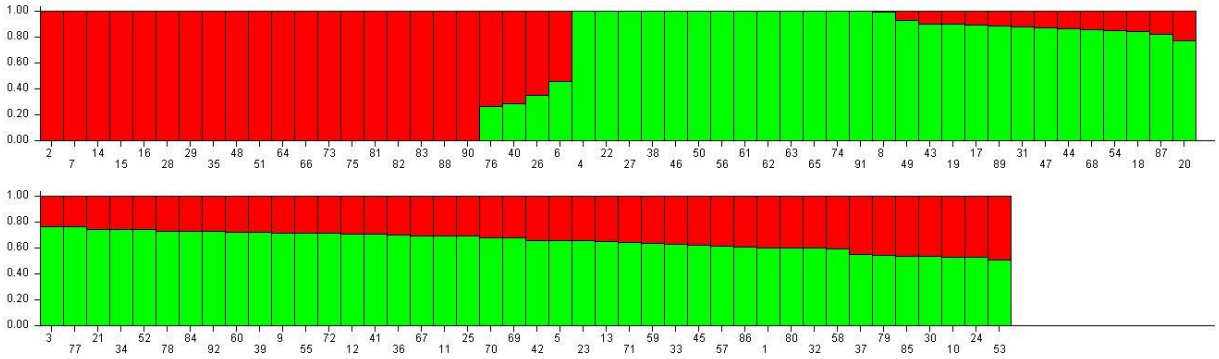


Figure 5: Population structure (K = 2) inferred from STRUCTURE analysis for the 89 soybean genotypes based on 7,692 SNPs

3.3.4 Neighbor-joining Phylogenetic Tree

To validate and gain further insight into the genetic diversity of the soybean germplasm panel used in this study, a phylogenetic tree by neighbor-joining method was generated. The 89 soybean genotypes were classified into three major clusters (Figure 6). The genotypes were clearly separated into three distinct subclusters: There were 40 genotypes in subcluster 1 which included Nam II and GC00138-29 and 13 progenies derived from a cross between these two genotypes. Nam II is a Ugandan variety which is a selection from TGM 79; a variety obtained from IITA while GC00138-29 is a variety from AVRDC in Taiwan. This subcluster also included released varieties in Uganda; Namsoy 3 which is a cross between Kabanyolo 1 and Nam 1 (selection from ICAL 131 from the USA) and Maksoy 5N that is a progeny of Nam II and GC00138-29. The second subcluster had 26 genotypes, among which 13 genotypes were from SeedCo in southern Africa and 8 genotypes from AVRDC, Taiwan. It was surprising that Namsoy 4M, a released Ugandan variety that is a progeny of Nam II and GC00138-29 was clustered in this subcluster. By comparison, the other remaining 23 genotypes belonged to subcluster 3, among which 7 genotypes were progenies from a cross between Duiker and TGx 1835-10E while 9 were from a cross between Duiker and GC00138-29. The subcluster also included released Ugandan varieties; Maksoy 1N (selection from TGx 1835-10E), Maksoy 2N (Duiker X TGx 1835-10E), Maksoy 3N and Maksoy 4N (Duiker X GC00138-29). However, few S lines and AVRDC genotypes were scattered in all three major clusters.

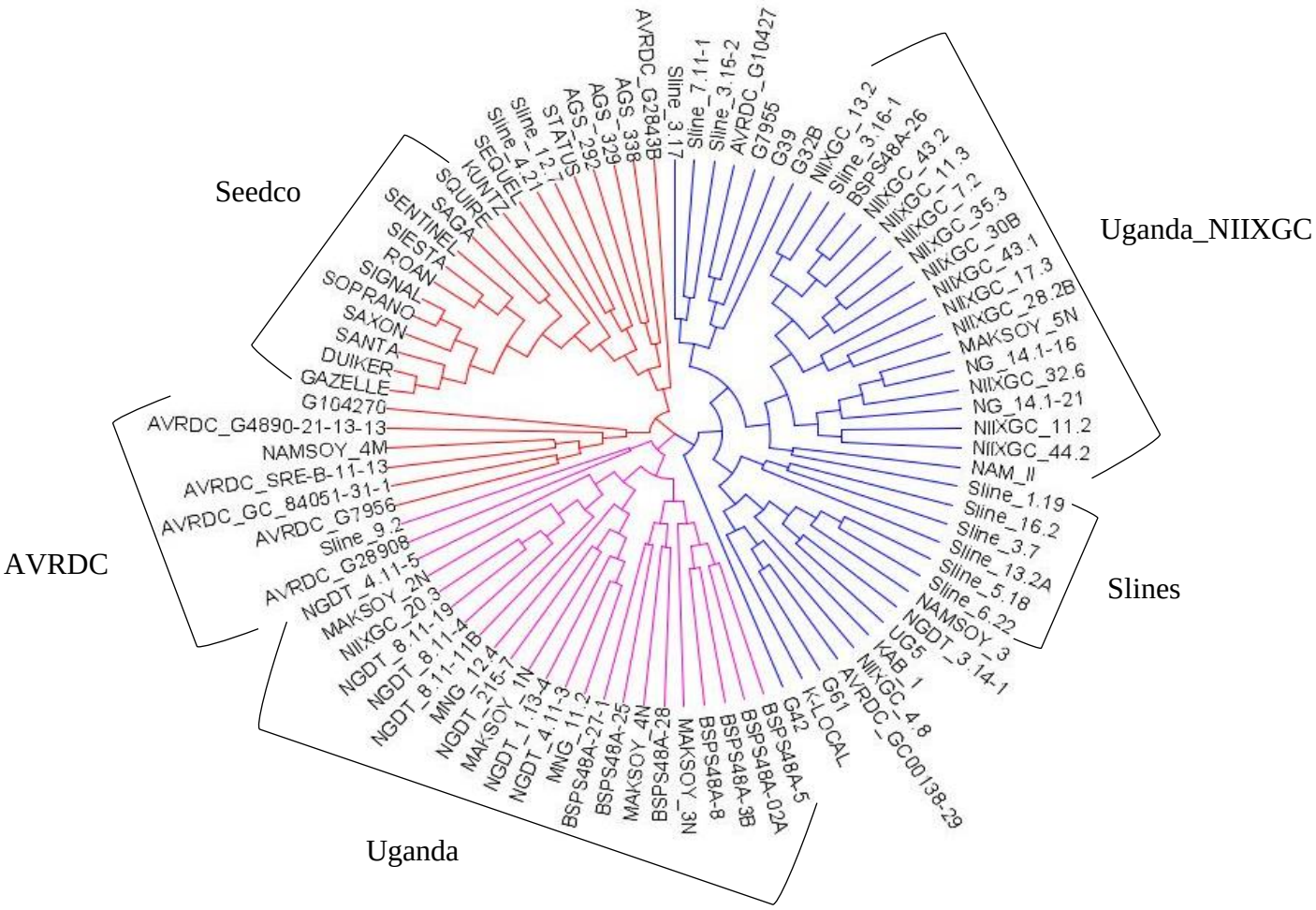


Figure 6: Phylogenetic Tree based on the Neighbor-Joining method showing genetic dissimilarity between soybean genotypes, based on SNP markers

3.3.5 Principal Component Analysis (PCA)

PCA has been proposed as an alternative to population structure analysis for studying population stratification from genotypic data (Patterson *et al.,* 2006). A PCA of the 89 genotypes with the 7,962 SNPs also showed a clear separation of the same three major groups that were identified by the phylogenetic tree (Figure 7).

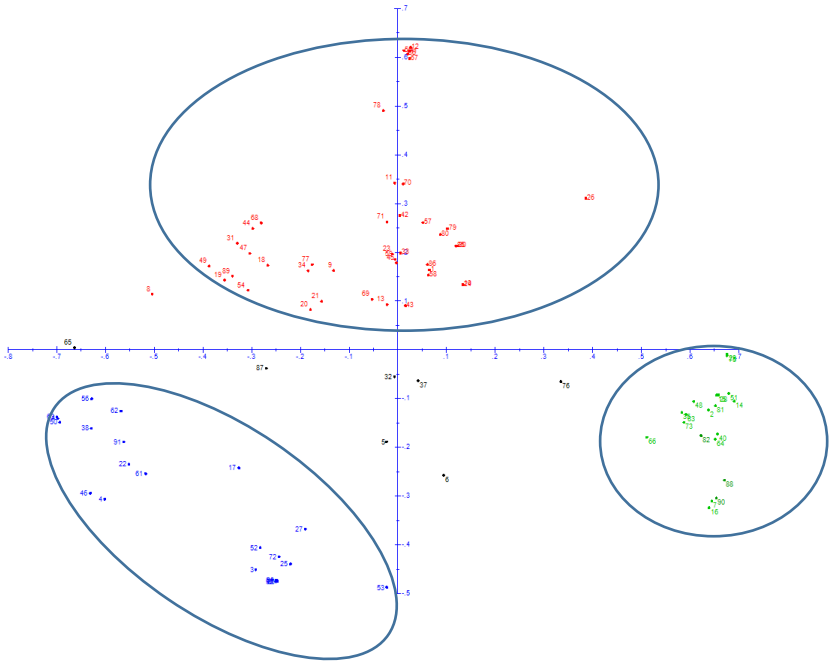


Figure 7: Plot of PC1 (40.6%) and PC2 (18.2%) from principal component analysis based on genetic distance matrix calculated for 89 soybean genotypes genotyped with 7,692SNPs

3.3.6 Analysis of molecular variance

Analysis of molecular variance (AMOVA) among the 89 soybean genotypes indicated that 2% of the variance was due to genetic differentiation among the populations, while 98% of the variance was accounted for by genetic differentiation among individuals within populations.

3.4 DISCUSSION

One of the prerequisites for a successful breeding program is a high level of genetic diversity and understanding of the relationship among the germplasm that is used for the development of new crop varieties. Over the years, most soybean breeding programs have replaced traditional varieties or landraces with more modern varieties with desirable attributes that have led to increased yields. This often results in narrowing the genetic diversity. However, to understand the genetic diversity of tropical soybean genotypes, fairly wider sets of genotypes from different origins were used. The focus of this study was to understand the genetic diversity of soybean germplasm from tropical Africa like Uganda and Zimbabwe. These genotypes included parental lines, landraces, released varieties and advanced lines that were expected to be representative of

the existing germplasm in tropical Africa. However, germplasm from Japan, Taiwan, and the USA were included in the study for comparison of the studied germplasm. In this study, low genetic diversity was observed with the average PIC value of 0.27 obtained, similar to the values reported in previous studies. For example, Bisen *et al*. (2014) reported average PIC value of 0.20, Liu *et al*. (2017) 0.28, Hipparagi *et al.* (2017) 0.36, Zhang *et al.* (2013) 0.39, Satyavathi *et al.* (2006) 0.44, Kumawat *et al.* (2015) 0.48, Chauhan *et al.* (2015) 0.57, Tantasawat *et al.* (2011) 0.60, Gupta and Manjaya (2017) 0.61, Mulato *et al.* (2010) 0.63.

Being a self-fertilizing crop, soybean is expected to have low heterozygosity than open-pollinated crop species as observed in the current study where it ranged from 0.0 to 0.35. Most previous studies have reported low heterozygosity in soybean; Mulato *et al.* (2010) 0.03, Bisen *et al.* (2014) 0.05, Gupta and Manjaya (2017) 0.06, Liu *et al.* (2017) 0.07, Hipparagi *et al.* (2017) 0.11, Zhang *et al.* (2013) 0.46. Similarly, several studies have observed low genetic diversity; especially in cultivated soybean. Bisen *et al.* (2014) reported genetic diversity of 0.23, Kumawat *et al.* (2015) 0.35, Hipparagi *et al.* (2017) 0.43, Satyavathi *et al.* (2006) 0.56. The low diversity previously reported in soybean could be attributed to domestication. Infact Hyten *et al.* (2006) concluded that, during soybean domestication, 50% of the genetic diversity was lost.

Additionally, the low genetic diversity that was observed in this study could have been because the genotypes used were mainly released varieties and advanced breeding lines that share a common pedigree. Most studies that have reported high genetic diversity, have used wild relatives and landraces of soybean (Hao *et al.*, 2012; Li *et al.*, 2010; Zhou *et al.*, 2015). Previous studies that involved improved soybean varieties have observed low genetic diversity (Liu *et al.*, 2017; Maldonado dos Santos *et al.*, 2016). These improved varieties tend to have low genetic diversity because of the high selection pressure subjected to the genotypes during evaluation and selection (Gwinner *et al.*, 2017). This was also confirmed by genetic distance and kinship analysis that showed that the majority of the genotypes in this study are related to each other in one way or another. Similarity of the parentage of the studied genotypes could have also led to the observed low diversity. The studied genotypes especially those from Uganda shared similar parents. For example, sixteen offsprings were derived from Nam 2 X GC00138-29 while sixteen were derived from Duiker X TGx 1835-10E crosses were the majority.

Relatedness of the different genotypes is confirmed by phylogenetic tree and PCA results. The 89 genotypes of the current study were grouped into three major sub-clusters. The first sub-cluster included Nam II and GC00138-29 and 13 progenies derived from a cross between these two genotypes. Nam II is a Ugandan variety which is a selection from TGM 79; a variety obtained from IITA while GC00138-29 is a variety from AVRDC in Taiwan. This sub-cluster also included Maksoy 5N; released in 2013 and NII X GC 44.2 that was released in 2017 as Maksoy 6N that are both progenies of Nam II and GC00138-29 cross. Since TMG 79 and GC00138-29 were introduced to Uganda through Consultative Group for International Agricultural Research (CGIAR) institutions that usually collect germplasm from different countries, so it is easy for such germplasm to share the same geographical origin (Halewood et al. 2020). On the other hand, genotypes from Seed Co and AVRDC, Taiwan were grouped in the second sub-cluster. This implies that soybean varieties from Seed Co share very similar parents and geographical origin with genotypes from AVRDC.

By comparison, the third sub-cluster mainly consisted of progenies from two crosses; Duiker X TGx 1835-10E and Duiker X GC00138-29. The sub-cluster included four released Ugandan varieties; Maksoy 1N (selection from TGx 1835-10E), Maksoy 2N (Duiker X TGx 1835-10E), Maksoy 3N and Maksoy 4N (Duiker X GC00138-29). Duiker originated from Zimbabwe and was used as a female parent during the generation of the two crosses. This implies that four out of the six soybean varieties from Uganda share similar parentage.

3.5 CONCLUSIONS AND RECOMMENDATIONS

The results of the study demonstrated the presence of a low level of heterozygosity within most of the genotypes studied, suggesting that they are fixed lines as a result of constant selection during the breeding process. Furthermore, the genetic diversity among the studied soybean genotypes was low. The low genetic diversity observed in this study can lead to genetic erosion of the existing Ugandan soybean germplasm. Therefore to address these challenges, the development of soybean varieties with the desirable traits, requires diversification of the genetic background of the current breeding population by incorporating new genetic resources from other countries. For example genotypes from AVRDC and Seed Co were clustered in a separate sub-cluster compared to the two sub-clusters that had majorly Ugandan lines. Therefore

germplasm from AVRDC and Seed Co that were clustered separately could be used to broaden the genetic base of Ugandan soybean.

CHAPTER FOUR

4.0 NUTRIENT PROFILING OF TROPICAL SOYBEAN (*GLYCINE MAX* (L.) CORE GERMPLASM

4.1 INTRODUCTION

Soybean (*Glycine max* (L.) Merrill) has become an important ingredient in the diets of both humans and livestock due to its high nutritional value and low cost (Drago *et al.*, 2011; Singh *et al.*, 2008; Friedman and Brandon 2001). The total world soybean production is estimated at 348.7 million metric tons (MT). The world's leading producer is the USA, which produces about 35% of the world's soybean (123.7 million MT). Brazil is second with 34% of soybean produced (117.9 million MT) while Argentina is third with 11% (37.8 million MT), and China is fourth with 4% (14.2 million MT). The remaining countries account for 16% (55.2 million MT) of the global soybean output (FAO 2018). In Africa, total soybean production rose from 1.4 million MT in 2008 to 3.6 million MT in 2018, representing 1.0% of world production. The three leading African countries in soybean production are South Africa (1,316,000 MT), Nigeria (730,000 MT), and Zambia (351,416 MT) (FAO, 2018). Uganda is the leading producer of soybean in East Africa with a production of 29,000 MT (FAO, 2018). In the case of Uganda, a report by the Uganda Bureau of Statistics (2021), recorded that the northern region produced 15,729 MT, the eastern region (5,803 MT), the western region (1,886 MT), and the central region (192 MT). This report also showed that the leading districts in soybean production in Uganda were Oyam (8,030 MT), Apac (3,225 MT), Tororo (2,180 MT), and Lira (2,045 MT) (UBOS 2021).

Soybean grains contain about 40% protein, 20% oil, an optimal supply of protein, and high-calorie value (Singh *et al.*, 2008). Additionally, soybean oil is composed of approximately 16% saturated fatty acids (palmitic [C16:0] and stearic [C18:0]), 24% monounsaturated fatty acids (oleic [C18:1]), and 60% polyunsaturated fatty acids (linoleic [C18:2] and linolenic [C18:3]) (Drago *et al.*, 2011). The processing capacity for soybean in Uganda and across the East African region has significantly increased. The increase in processing capacity has been triggered by the growing interest of the farmers to grow the crop as a main source of cash because of the available superior varieties (Tukamuhabwa *et al.*, 2019). These established processing plants

make different food and feed products. However, most consumers are interested in soy-based food products with improved protein and oil content to meet special food applications (Singh *et al*. 2008; Miladinovic *et al*., 2011). In the past decade, the key focus for most soybean breeding programs in Tropical Africa has been on the improvement of traits such as yield, resistance to pod shattering and lodging, high pod clearance, resistance to pests and diseases. In contrast, the traits related to seed composition have received very little attention (Tukamuhabwa *et al*., 2019; Tukamuhabwa and Oloka, 2016; Bashaasha, 1992).

Moreover, the development of varieties with improved nutritional traits has been further affected by the negative correlation that exists among the different traits. For example, several studies have reported a significant negative correlation between oleic acid and palmitic acid (Ahire, 2012; Qin *et al*., 2014; Rebetzke *et al*., 2001). Similarly, total oil has been reported to exhibit a strong negative correlation with oleic acid (Bachlava *et al*., 2008; Rani *et al*., 2007). Likewise, a negative correlation between oil and protein has also been reported in several studies (Qin *et al*., 2014; Marega *et al*., 2001). Knowledge of the relationship between the different nutritional traits in soybean would be very vital in developing soybean varieties with improved nutritional properties. Therefore the objectives of this study were: (1) to quantify the total protein, total oil and fatty acids of 52 soybean genotypes from different sources, (2) to identify correlations among total protein, total oil content, and fatty acids. Therefore the findings from this study will provide an opportunity to develop soybean varieties with improved nutritional traits.

4.2 MATERIALS AND METHODS

4.2.1 Plant Materials and Sample Preparation

Fifty-two soybean genotypes (Appendix 4) were grown in the first season of 2016 (2016A); at Makerere University Agricultural Research Institute Kabanyolo (MUARIK) which is located in Central Uganda. The 52 soybean genotypes were selected from the core collection of Ugandan germplasm that were part of the 89 genotypes used in the previous study (Chapter 3). However, the basis of selection was for comparison, germplasm from Japan, the USA and SeedCo were included in this study. The 52 soybean genotypes were planted in an alpha lattice design with three replicates. The plot size was 5m long with a spacing of 60 cm × 5 cm. The trials were kept weed-free and three weedings were conducted each season. No agrochemicals were used on the

trials to control pests. The grain from each plot was harvested and sun-dried to below 10% moisture content. The seed samples for each genotype were divided into three replicates and about 300g ground through a 2 mm screen using Cyclotec^{TM} 1093 mill (https://www.fossanalytics.com). The milled samples were stored in the cold room at a temperature of -4^{o}C at BecA – ILRI Hub, Nairobi Kenya, where all the subsequent analysis was performed.

4.2.2 Total Protein Content

Protein concentration was determined using the Modified Folin-Lowry Method (Lowry *et al.*, 1951). One hundred mg of the milled samples were weighed in duplicate into 25 mL culture tubes. Five ml 5% sodium dodecyl sulphate (SDS) was added, vortexed, and left to stand for 2 hours at room temperature and centrifuged at 2000 ppm for 10 min. Fifty µl aliquot was taken and diluted with 950 µl of water and made to 1000 µl in separate culture tubes. One hundred µl aliquot of the diluted extract was taken for analysis. Blank (100 µl of distilled water), standards (0, 100, 200, 300, 400, 500 µl of Bovine serum albumin) and samples (100 µl) were pipetted into glass culture tubes, and all made to 1000 µl.

One ml Reagent A (0.4 volumes of water, 1 volume CTC reagent, 1.6 volumes of 5 % SDS and I volume of 0.8 M NaOH) was then added to each of the tubes and immediately vortexed; at 20 seconds interval. Five hundred µl of Reagent B (1 volume of Folin-Ciocalteu phenol reagent, 2 Aldrich 9252, and 5 volumes of distilled water) was added to each tube and immediately vortexed and left to stand for 30 minutes for color development. An ultraviolet-visible spectrophotometer was used to measure the absorbance of the standards and samples versus the blank at a wavelength setting of 750 nm. The absorbance values of the standards versus their corresponding protein concentrations were plotted to prepare a calibration curve, and the protein concentration of the samples was determined.

The total protein (crude) in the residue was calculated using the formulae:

$$\text{Total Protein (g/100 g)} = \frac{C \text{ X } 100 \text{ XDF}}{10^6 \text{ X W}}$$

Where;

C = Concentration obtained from the calibration in µg/ml
100 = Conversion factor to report results in g/100g
DF = Total dilution factor (1000)
10^6 = Conversion from µg to g
W = Weight of the sample in grams

4.2.3 Total Oil Content and Fatty Acid Profiling

Total oil content was determined for the three replications as in the trial using the chloroform/methanol gravimetric method (Bligh and Dyer, 1959). Two grams of the milled sample for each soybean genotype were weighed in duplicate into 50 mL culture tubes ($\boldsymbol{W_1}$). Thirty-two ml of Clarase solution was added, the tubes were capped and gently shaken until the sample was well mixed with the enzyme solution. The sample was incubated for one hour in a 45^0C water bath while gently mixing by inversion after every 20 minutes. All the extract was transferred to a 250 ml polypropylene bottle, capped and centrifuged at 2000 rpm for 15 minutes to clarify the chloroform. The top aqueous phase was carefully removed and discarded with a tap aspirator pump leaving a 2-4 mm thick layer on the chloroform. A hole was cautiously broken into the surface crust with a glass rod, and 20.0 ml of the chloroform extract was pipetted into a pre-weighed 50 ml beaker ($\boldsymbol{W_2}$). Further, a 20 ml aliquot of the chloroform extract was taken and stored at -20 °C for fatty acids methyl esters (FAMES) analysis with Gas Chromatography-Mass Spectrometry (GC – MS). The solution was evaporated to dryness by leaving it overnight in a fume hood; the beaker was placed in an oven at 102 ^{0}C for 30 minutes, removed, and cooled in an evacuated desiccator for 1 hour. The beaker plus the total oil was weighed on a microbalance to the nearest 0.1 mg ($\boldsymbol{W_3}$).

The total oil in the residue was calculated using the formulae:

$$\text{Total oil (g/100 g)} = \frac{(W_3) - (W_2) \text{ X } 100 \text{ X } 4}{(W_1)}$$

Where; W_1 - Sample weight (g)

W_2 - Weight of beaker

W_3 - Weight of beaker + total oil

FAMES were analyzed using a DB-5 column, on an HP 5890 Series II GC equipped with an HP 7673 autosampler (Hewlett Packard, Sunnydale, CA). Peak areas were recorded using ChemStation software (Hewlett Packard, Sunnydale, CA). The identification of individual FAMES was performed by calculating the Kovats linear retention index. The linear retention index was subsequently compared with obtained values from the National Institute of Standards and Technology (NIST) and Pherobase databases in cases where standards were absent. Quantification of individual FAMES was performed by the area percent method.

4.2.4 Data Analysis

Statistical differences among the different soybean genotypes were determined using ANOVA at the 5 % level ($\alpha = 0.05$) of significance for all the parameters evaluated, using Genstat, 13th Edition. Whenever ANOVA indicated a significant difference, mean separation was done using Fisher's protected LSD. Correlation analysis was performed using the corrplot package for graphical display using the Pearson method in R software.

4.3 RESULTS

4.3.1 Total Protein, Total Oil and Fatty Acids Content of the studied soybean genotypes

Significant differences were found among genotypes ($P = 0.003$) for total protein content (Table 1). Genotypes Sline 5.18, BSPS 48A-8 and BSPS 48A-27-1 had the highest protein content of 50.40%, 48.88% and 48.08% respectively (Table 2). On the other hand, NIIXGC 17.3 and Nam II had the lowest protein content of 30.07% and 35.57% respectively (Table 2). Total oil content was highly significant for both genotypes and origin of the soybean ($P = 0.001$) (Table 1). G32B,

Roan and AGS 338 had the highest oil content of 23.48%, 23.47% and 23.26% respectively while Signal had the lowest content of 14.94% (Table 2).

Significant differences were found among genotypes for the four fatty acids that were detected in this study; oleic fatty acid (P = 0.001), palmitic acid (P = 0.001), stearic acid (P = 0.001) and linoleic acid (P = 0.008) (Table 1). Genotypes BSPS 48A-25, G7955 and BSPS 48A-5 had the highest oleic fatty acid content of 39.95%, 39.95% and 38.66% respectively while NG 14.1-16 had the lowest content of 22.69% (Table 2). Genotypes Nam II, Siesta and Namsoy 3 had the highest palmitic acid content of 21.18%, 20.18% and 19.68% respectively (Table 2). On the other hand, G7955 and K-Local had the lowest palmitic acid content of 10.58% and 10.95% respectively (Table 2). Genotypes NG 14.1-16, AVRDC SRE-B-11-13 and Sequel had the highest stearic acid content of 16.76%, 15.58% and 14.57% respectively while Sline 16.2 had the lowest content of 4.93% (Table 2). Saga had the highest linoleic acid content of 51.72% while Maksoy 5N had the lowest content of 30.73% (Table 2).

Table 1: Analysis of variance for Total Protein, Total Oil and Fatty acids

SOV	d.f.	total protein	total oil	oleic acid	palmitic acid	stearic acid	linoleic acid
Rep	2	2	1.457	2.369	9.156	2.079	31.203
Genotype	51	73.69**	17.826***	40.581***	15.821***	20.495***	34.464***
Residual	108	31.85	2.653	7.848	3.523	3.701	7.811
Total	161			23.204	9.435	11.602	20.831

*, ** and *** indicate significance at $\alpha \leq 0.05$, $\alpha \leq 0.01$, and $\alpha \leq 0.001$ respectively

Table 2: Variation of total protein (%), total oil (%) and fatty acids (%) of soybean genotypes

Genotypes	Total protein	Total oil	Oleic acid	Palmitic acid	Stearic acid	Linoleic acid
AGS 338	37.97	23.26	35.88	11.80	6.54	45.77
AVRDC G2843B	37.32	21.75	24.12	18.21	12.26	45.42
AVRDC G7956	36.88	17.55	33.35	13.06	7.09	46.48
AVRDC GC00138-29	41.43	17.96	32.74	15.20	9.17	42.90
AVRDC GC84051-31-1	42.55	16.46	35.51	18.13	8.32	35.89
AVRDC SRE-B-11-13	43.58	16.23	27.90	19.08	15.58	37.41
BSPS 48A-25	37.11	17.61	39.95	11.65	8.41	39.97

BSPS 48A-27-1	48.08	19.87	37.38	12.08	7.07	43.47
BSPS 48A-3B	37.84	19.30	37.56	11.84	6.54	44.06
BSPS 48A-5	41.37	17.87	38.66	11.78	6.67	42.89
BSPS 48A-8	48.88	17.59	32.64	15.44	13.23	38.54
G32B	44.68	23.48	26.49	16.58	12.15	44.28
G42	43.87	17.19	33.67	14.51	7.56	43.84
G45	36.10	18.27	35.53	11.51	5.84	47.11
G7955	46.06	20.97	38.95	10.58	5.47	45.05
Gazelle	44.57	17.89	34.10	14.68	6.98	44.08
K-Local	39.89	18.94	32.26	10.95	5.17	51.63
Kab 1	39.17	15.59	34.58	11.49	5.36	48.57
Kuntz	39.88	22.24	30.65	16.76	9.01	41.65
Maksoy 2N	43.86	22.58	28.88	18.38	13.57	34.86
Maksoy 4N	45.32	21.34	36.32	13.43	10.00	39.50
Maksoy 5N	40.11	15.27	37.63	17.72	11.46	30.60
MNG 12.4	35.86	19.85	34.49	12.63	8.54	44.36
Nam II	35.57	20.91	23.94	21.18	14.14	40.71
Namsoy 3	40.68	17.54	25.78	19.68	13.13	37.41
NG 14.1-16	44.93	20.20	22.69	19.05	16.76	41.50
NGDT 4.11-4	37.66	21.00	37.04	13.14	8.03	41.79
NGDT 8.11-4	42.11	21.63	36.52	11.35	6.47	45.64
NII X GC 11.2	39.64	18.54	32.51	13.42	7.07	46.97
NII X GC 17.3	30.07	20.53	34.42	11.79	6.47	47.32
NII X GC 20.3	38.90	17.28	31.35	15.99	11.42	41.21
NII X GC 28.2B	38.65	18.26	32.97	13.00	6.04	47.99
NII X GC 30B	42.82	20.58	36.91	12.35	5.80	44.94
NII X GC 32.6	41.01	20.55	37.04	11.57	6.36	45.02
NII X GC 43.2	39.54	19.88	31.95	11.92	6.75	49.38
NII X GC 44.2	38.80	18.27	32.76	11.42	5.64	50.18
NII X GC 7.2	39.46	16.54	34.63	13.37	6.32	45.69
Roan	43.84	23.47	34.32	11.45	6.49	47.71
Saga	42.12	23.07	27.09	11.16	10.07	51.72
Saxon	40.95	17.91	36.95	12.22	6.94	43.90
Sentinel	37.93	15.81	33.50	11.45	5.47	49.57
Sequel	40.27	19.92	26.13	17.42	14.57	40.00
Siesta	43.42	22.79	25.86	20.18	12.62	41.34
Signal	39.17	14.94	32.71	11.93	4.97	50.39
Sline 13.2A	42.19	19.88	31.04	18.29	14.51	35.69
Sline 16.2	39.88	15.35	36.04	11.95	4.93	47.07
Sline 4.21	42.12	19.00	36.26	12.26	5.55	45.98

Sline 5.18	50.4	20.60	30.49	12.90	7.15	48.39
Sline 6.22	44.67	16.90	33.67	12.19	5.45	48.67
Sline 7.11	43.14	20.65	31.56	13.30	6.87	47.08
Soprano	40.11	19.24	35.54	11.84	5.68	46.99
Squire	44.88	22.66	22.82	11.28	11.96	49.44
Mean	40.71	19.44	32.79	13.88	8.50	44.38
LSD	9.33	2.82	3.28	3.32	3.09	3.72
CV%	14.30	9.00	6.10	14.60	22.2	5.10

4.3.2 Germplasm source and Total Protein, Total Oil and Fatty Acids Content of the studied genotypes

Genotypes from Japan had the highest protein content of 43.47%; followed by the USA (42.42%) while genotypes from Uganda had the lowest protein content of 40.19% (Table 3). Genotypes from the USA had the highest oil content of 20.43%; followed by SeedCo (20.11%) and AVRDC had the lowest (18.32%) (Table 3). Soybean genotypes from Uganda had the highest oleic fatty acid content (33.85%); followed by genotypes from Japan (33.17%) while genotypes from SeedCo had the lowest (30.68%) (Table 3). Genotypes from the USA had the highest palmitic acid content of 15.00%; followed by AVRDC (14.71%) and SEEDCO had the lowest (13.23%) (Table 3). Genotypes from AVRDC had the highest stearic acid content of 9.05% while those from Japan were the lowest (7.42%). SEEDCO had the highest content of linoleic acid of 46.82% while those from AVRDC had the lowest content of 43.64% (Table 3).

Table 3: Variation of total protein (%), total 0il (%) and fatty acids (%) of soybean genotypes from different origins

Origin & No. of genotypes		total protein	total oil	oleic acid	palmitic acid	stearic acid	linoleic acid
AVRDC (6)	Mean	41.05 **	18.32 ***	32.40	14.71 *	9.05 ***	43.64
	Range	36.88- 46.06	11.59-23.27	22.94-38.35	11.14-19.12	5.46-15.38	36.59-48.60
	%cv	9.90	6.90	15.80	14.84	14.22	8.16
Japan (6)	Mean	43.47 ***	18.73 ***	33.17 *	13.44	7.42 **	45.52*
	Range	39.88- 50.40	15.35- 20.65	31.13-36.17	11.91-13.31	4.93-7.06	35.69-48.66
	%cv	5.10	5.10	1.16	4.70	4.39	1.03
SEEDCO (10)	Mean	42.18	20.11 ***	30.68 **	13.23 **	8.59 ***	46.82
	Range	37.93- 44.88	14.94- 23.47	22.68-36.79	10.47-20.39	5.02-14.56	40.47-53.19
	%cv	10.90	10.00	8.55	14.93	16.29	8.64
UGANDA	Mean	40.19 ***	19.11 ***	33.85 ***	13.57 ***	8.24 ***	43.94***

(25)							
	Range	30.07- 48.88	15.27-22.58	22.70-39.66	10.93-21.05	4.94-16.75	30.25-51.65
	%cv	16.80	8.70	9.23	10.56	30.00	4.26
USA (5)	Mean	42.42 *	20.43 **	31.69 **	15.00*	8.70 *	43.89
	Range	36.1- 44.68	17.19- 23.48	26.83-35.31	11.66-17.33	5.74-12.02	39.85-47.26
	%cv	10.10	6.90	0.91	20.42	8.83	9.58

*, ** and *** indicate significance at $\alpha \leq 0.05$, $\alpha \leq 0.01$, and $\alpha \leq 0.001$ respectively within each origin of the soybean genotypes

4.3.3 Correlation analysis for total protein, total oil and fatty acids

A total of nine correlations were found to be significant; six negative and three positive correlation (Figure 8). Negative correlations were found between the following pairs; palmitic acid/linoleic acid, stearic acid/linoleic acid, oleic acid/palmitic acid, oleic acid/stearic acid, oleic acid/total oil and oleic acid/total protein. The positive correlation was found between the following pairs; palmitic acid/stearic, stearic acid/total oil and total oil/total protein. The package corrplot revealed that all the six nutritional traits studied approached the normal distribution (Figure 8).

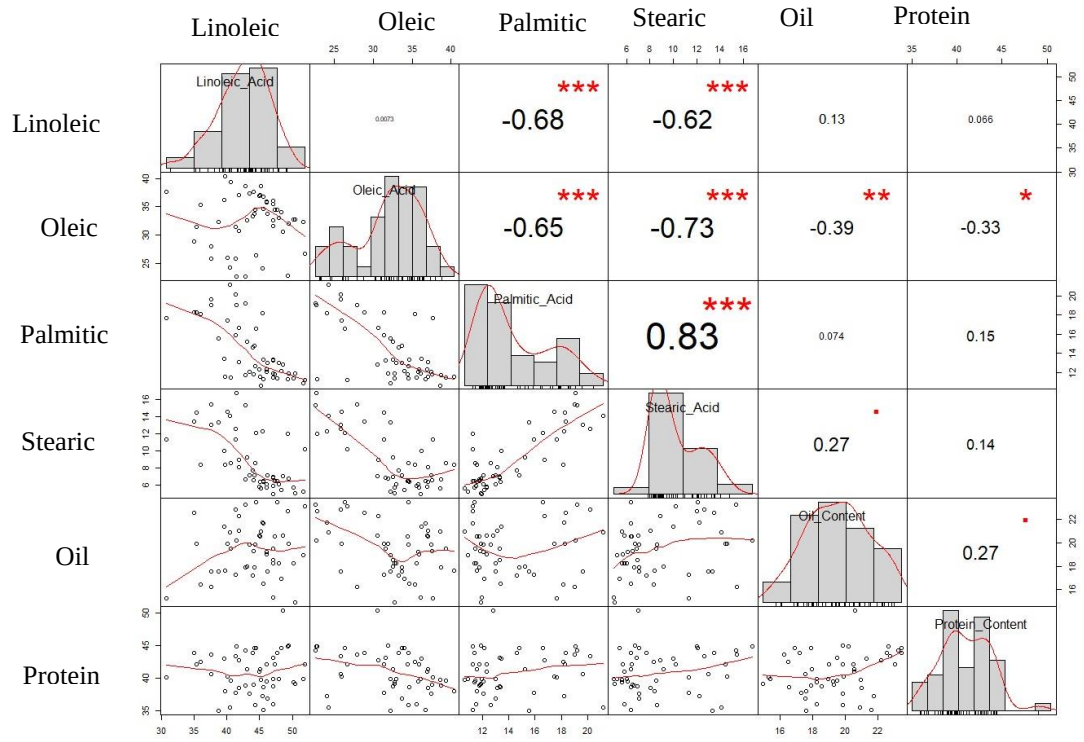


Figure 8: A Corrplot showing distribution and correlation between the different nutritional traits

*, ** and *** indicate significance at $\alpha \leq 0.05$, $\alpha \leq 0.01$, and $\alpha \leq 0.001$ respectively; values without * are not significant at $\alpha < 0.05$.

4.4 DISCUSSION

High protein content has been one of the most key traits for most soybean improvement programs. The results of this study revealed that several soybean genotypes had protein content above the average of most domesticated soybean reported to be 40% (WWF 2014; Patil *et al*. 2018; La *et al*. 2019). For example, two genotypes were identified in this study with high protein content; BSPS 48A-8 (48.88%) and BSPS 48A-27-1 (48.08%). Being elite lines developed in Uganda, these two soybean genotypes are well adapted to Ugandan environments. These genotypes have also gone through a series of selections, therefore, possess most of the desirable traits like high pod clearance, resistance to soybean rust disease, and high yields. These genotypes can be utilized by soybean breeders in Uganda to improve the existing varieties for protein since they are already adapted to Ugandan environments.

Total protein, total oil and fatty acids profiles were assessed from the 52 soybean germplasm collected from different origins in the current study. The protein content of the studied soybean genotypes was in the range of 30.07% to 50.40% and a mean of 40.71% that was in agreement with those reported previously (Maestri *et al*. 1998; Ojo *et al*. 2002; Qin *et al*. 2014; Assefa *et al*. 2019; La *et al*. 2019). However, the mean protein content observed in this study is slightly lower than what La *et al*. (2019) reported (43.89%). This could be because the genotypes studied by La *et al*. (2019) were wild soybean accessions while those in the current study are domesticated soybean. This study also observed that the mean protein content was higher than 35.7% reported by Assefa *et al*. (2019). Such variability between this study and previous works may most likely be due to diverse genotypes used in the different studies.

Wide variability for protein content 20.33% (30.07% - 50.40%) was also observed in the current study compared to previous studies. However, a study by Karr-Lilienthal *et al*. (2004) observed a moderate range of protein content of 12.30% (44.9% - 32.6%). Another study by Arslanoglu *et al*. (2011) reported a smaller range of 9.32% (29.25% - 38.57%). A wider range of total protein content that was observed in the current study could have been as a result of the many test

genotypes and wide collections that were from different origins. For example, the study by Karr-Lilienthal *et al*. (2004) used six genotypes while Arslanoglu *et al*. (2011) used nine genotypes for the compositional analysis, compared to the 52 genotypes used in the current study.

The total oil content of the soybean genotypes in the current study was in the range of 14.94% to 23.48% and a mean of 19.44%. The narrow range of 8.54% for total oil content observed in this study was in agreement with previous studies (Maestri *et al*. 1998; Ojo *et al*. 2002; Qin *et al*. 2014; Sultan *et al*. 2015; La *et al*. 2019). However, the oil content in the current study had a higher mean than the one reported by La *et al*. (2019). This is because most of the genotypes used in the current study were domesticated soybean while those used by La *et al*. (2019) were wild relatives of soybean that usually have low oil content (Patil *et al*. 2018; Yao *et al*. 2020). The small range of total oil content makes it more challenging to improve such a trait compared to traits that have a wide range like total protein content.

Palmitic acid is the predominant saturated fatty acid in soybean oil and most varieties typically contain 11% to 12% of palmitic acid (Burton *et al*. 1994; Cardinal and Burton 2007; Qin *et al*. 2014). To reduce the health risks associated with the consumption of palmitic acid, breeders have developed soybean lines with reduced palmitic acid content below 10% (Burton *et al*. 1994; Cardinal and Burton 2007). The proportion of palmitic acid observed in the current study ranged from 10.58% to 21.18% with a mean of 13.88% that is above 10%. This is not surprising because most of the soybean genotypes are released varieties, not mutants. Most of the soybean breeders develop soybean varieties with palmitic acid below 10% through mutational breeding. Therefore this study shows that the only way to develop soybean varieties with reduced palmitic acid is through mutational breeding (Fehr and Hammond 1998; Lee *et al*. 2007; Ahire 2012).

The high saturated fatty acids (palmitic and stearic) in soybean are also desired for some special food production because saturated fatty acids have no double bonds, hence they resist oxidation and make the oil and foods last longer without any off-flavors. Soybean oil with high proportions of saturated fatty acids has industrial applications for the production of plastic fats like shortening and margarine (Fehr and Hammond 1998; Pham 2011). The proportion of palmitic

acid observed in the current study was in general agreement with previous studies conducted on different soybean genotypes that reported a range of 6.6% to 28.2% (Kumar *et al*. 2006; Ahire 2012). For stearic acid, the range was 4.93% to 22.20% and a mean of 8.50%. The preferred high saturated fatty acids in soybean should be at least about 15% for palmitic acid and 20% for stearic acid (Fehr and Hammond 1998). In the current study, several soybean genotypes showed an elevated proportion of palmitic and stearic acid above 15% and 20% respectively. These genotypes with elevated proportions of palmitic and stearic acids could be used for the development of soybean varieties with elevated levels of saturated fatty acids.

The primary limitation of soybean oil is low oxidative stability which reduces shelf life, flavor and durability at high temperatures; and poor cold flow properties for biodiesel. To improve oxidative stability and undesirable taste, soybean oil is hydrogenated to reduce double bonds which are sites of oxidative attack that reduce stability, shelf life, and increase off-flavors. High oleic soybean oil reduces the need for hydrogenation and eliminates trans-fats that are associated with increased heart diseases in humans. In the current study, oleic fatty acid ranged from 22.69% to 39.95% with a mean of 32.79% as reported in previous studies (Erickson *et al*. 1988; Qin *et al*. 2014; Abdelghany *et al*. 2020). The current study identified mid-oleate soybean genotypes with close to twice the oleic fatty acid content of normal soybean cultivars that have been reported to be at about 23% (Wilson 2004; Alt 2005). The reason for the observation of higher oleic fatty acid content in the current study is not known.

In the current study, genotypes from Japan had the highest mean total protein content of 43.47%. A study conducted by Grieshop and Fahey (2001) who compared the nutritional traits of soybean from Brazil, China and the USA, reported that soybean genotypes from China had the highest protein. Similarly, Karr-Lilienthal *et al.* (2004) also reported that soybean from China had the highest protein content. Therefore in the current study, it is not surprising that soybean genotypes from Japan which is geographically closer to China have high protein content. It may be presumed that the Japanese and Chinese use soybean in similar manner and tend to naturally select soybeans of the same traits. Additionally, the high protein content among genotypes from Japan was because of a deliberate effort to develop soybean varieties with higher protein content

by the government after the second world war; to fight malnutrition (Saito, 1972). Infact protein content has been reported in Japan as the major breeding objective in regard to seed component (Saito, 1972).

In the current study, genotypes from the USA had the second-highest total protein content of 42.42%, the highest oil content of 20.43%, and the highest palmitic acid content of 15.00%. For approximately 40 years, the improvement of seed composition for nutritional traits such as high protein and oil content acid has been one of the soybean breeding priorities in the USA (Anderson *et al.*, 2019). Therefore it is not surprising that germplasm from the USA had the highest oil content and the second-highest protein content in the current study. The high oil content observed in the current study has been reported by previous studies. For example, a study by Baize (1999) compared the nutritional composition of soybean meal from different countries and reported that soybean meal from the USA had the highest oil content compared to those from Brazil and China. These results were similar to another study by Grieshop and Fahey (2001) who compared nutrient compositions of soybeans from the three major soybean-producing countries; and reported that soybean varieties from the United States of America had the highest oil content of 18.70%, followed by Brazil (18.66%) and China was last (17.25%). Although the ancestral lines from Asia and America came from the same regions at approximately the same time, they are genetically distinct (Li *et al.*, 2001). In fact, for nearly 90 years, breeding programs in Asia and America have been selecting for improved varieties that are adapted to their environments using divergent gene pools (Anderson *et al.*, 2019). Hence it is not surprising that the germplasm from the USA behaved differently from the germplasm from Japan that is located in Asia. Additionally, soybean breeding programs in Asia and the USA are among the oldest where more breeding has been conducted to improve nutritional traits compared to the breeding programs in Africa (Uganda and Zimbabwe) where soybean reached much later. These results suggest that soybean breeders in Uganda and across the East African region can access germplasm with high protein from Asia while those with high oil content from the USA to be used as parents in the generation of progenies.

The low oil content among soybean germplasm from AVRDC observed in the current study was not surprising. This is because AVRDC majors in the development of soybean varieties (vegetable and grain soybean) with high protein and low oil being among the major breeding strategies (Djanta *et al.*, 2020). Vegetable soybeans generally have low oil content. For example, Mentreddy *et al.* (2002) reported a range of 5% to 7% on a fresh weight basis while Carson *et al.* (2011) reported a range of 13.4% to 16.8%. The study further revealed that genotypes from Uganda had the highest oleic fatty acid. The observed high oleic fatty acid among genotypes from Uganda could be originating from one of the germplasm used as parents to generate crosses. This is because most of the Ugandan soybean genotypes share similar parentage. There is a need to explore which parental genotypes could be conferring the high oleic fatty acid trait. This finding further suggests that soybean breeders from across tropical Africa can access germplasm with high oleic fatty acid from Uganda.

In the present study, palmitic acid showed a positive correlation with stearic acid that was in agreement with earlier reports (Ahire 2012; Bachlava *et al.*, 2008; Rahman *et al.*, 2003; Stoltzfus *et al.*, 2000). However, for plastic fat to remain stable, it is desirable to have a fatty acid composition of about 15% or more of palmitic acid (Fehr and Hammond, 1998). The use of soybean oil having a relatively high content of saturated fatty acids (palmitic acid and stearic acid) allows the production of more desirable plastic fat. Palmitic acid averages approximately 11% of the total fatty acids whereas stearic acid averages about 4% of total fatty acids present in conventional soybean oil. Additionally, it is highly desirable to produce soybean varieties having elevated palmitic acid and stearic acid contents to produce stable plastic fat. In a separate study by Rahman *et al.* (2003), it was reported that palmitic acid and stearic acid are inherited independently, suggesting that the trait of elevated saturated fatty acids in one genotype can be easily achieved by conventional breeding and selection using different breeding techniques.

A significant negative correlation between oleic acid and palmitic acid observed was also reported by Rebetzke *et al.* (2001). Similarly, Qin *et al.* (2014) reported a significant negative correlation between oleic acid and palmitic acid. Ahire (2012) detected a negative correlation between oleic acid and palmitic acid among selected mutants from soybean variety MACS 450.

In the same way, Alt *et al.* (2005) showed that oleic acid of 88 F2:3 lines was significantly negatively correlated with palmitic acid (r=-0.470). The negative correlation between oleic acid and saturated fatty acids (palmitic and stearic) coupled with the positive correlation between palmitic acid and stearic acid offers the opportunity to develop soybean varieties with improved oil quality. In the present study, results indicated that selection for higher oleic acid will result in lower palmitic and stearic acids leading to improvement in the quality of the soybean oil (Pham 2011; Qin *et al.*, 2014). Oleic fatty acid is a monounsaturated fatty acid that can improve the oil quality and self-life of products processed using such oil.

Total oil in soybean has been reported to exhibit a strong negative correlation with oleic acid that is in agreement with the present study (Bachlava *et al.*, 2008; Rani *et al.*, 2007). The significant negative correlation between oleic acid and total oil suggests that it would be difficult to develop soybean varieties possessing both traits. A similar trend was observed between oleic acid and total protein where there was a negative correlation, suggesting that it is difficult to develop a soybean variety with high oleic acid and protein. This study, therefore, suggests that it would be extremely difficult to develop a soybean variety that has both high oleic and oil or protein content using conventional breeding. This calls for the use of other advanced technologies such as the development of molecular markers associated with a trait of interest to overcome the negative correlation between oleic acid and total oil and total protein content. Through convention breeding, soybean breeders can also explore options of developing different product lines with each developed soybean variety possessing certain unique nutritional traits not all in one variety.

4.5 CONCLUSIONS AND RECOMMENDATIONS

The content, variability, and correlation of different nutritional traits of 52 soybean germplasm from different countries were investigated in the current study, leading to the identification of several genotypes with nutritional traits above average. These genotypes can be used as parents to improve nutritional traits among Ugandan soybean varieties. For example, soybean genotypes with elevated nutritional traits like Sline 5.18 (protein content), G32B (oil content) and BSPS 48A-25 (oleic fatty acid) can be used by breeders to develop soybean varieties with improved nutritional traits.

Soybean genotypes originating from different sources have different nutritional traits like total protein, total oil and fatty acids. For example, soybean genotypes from Japan had the highest total oil content. On the other hand, the highest total oil and palmitic fatty acid were observed among the soybean genotypes from the USA while genotypes from Uganda had the highest oleic fatty acid. Therefore, breeders can seek germplasm with improved total protein content from Japan, total oil content and palmitic fatty acid from the USA and oleic fatty acid from Uganda.

The significant negative correlation of oleic acid with total oil, total protein and palmitic fatty acid indicates that it would be extremely difficult to develop soybean varieties that contain high oleic acid as well as high total oil, total protein and palmitic acid using conventional breeding. There is a need to explore other options of increasing oleic fatty acid content without altering the other nutritional traits such as mutation breeding and genetic transformation. Soybean breeders can also explore options of developing different product lines with each developed soybean variety possessing certain unique nutritional traits not all in one variety using conventional breeding.

CHAPTER FIVE

5.0 GENOME-WIDE ASSOCIATION STUDY FOR NUTRITIONAL TRAITS IN SOYBEAN USING SNP MARKERS

5.1 INTRODUCTION

Soybean (*Glycine max* (L.) Merrill) is among the most valuable crops worldwide due to its numerous uses in human food, animal feed, and industrial uses (Drago *et al.*, 2011; Singh *et al.*, 2008; Stein *et al.*, 2008). Soybean constitutes about 40% protein, 20% oil, with enormous potential to fight malnutrition among people throughout the world (Liu 1997). Soybean can be consumed as vegetable oil or processed into various soybean food products (Friedman and Brandon 2001; Singh *et al.*, 2008). In 2018, soybean represented 70% of world protein meal consumption and 36% of world vegetable oil consumption (SoyStat, 2018). Studies have shown that soybean and soy-based foods have numerous health benefits such as cholesterol reduction, improved vascular health, preserved bone mineral density, and reduction of menopausal symptoms in humans (Anderson *et al.*, 1999). In animal nutrition, soybean meal (SBM) is a major source of high-quality protein (Drago *et al.*, 2011; Stein *et al.*, 2008).

While improving nutritional traits of soybean varieties has been a major objective of many soybean breeding programs for decades, the negative correlations of seed protein content with seed oil content and seed yield have hampered progress (Marega *et al.*, 2001; Maestri *et al.*, 1998; Bachlava *et al.*, 2009; Cardinal and Burton 2007). Most nutritional traits like seed protein and oil content in soybean are quantitatively inherited and usually determined by the interaction of several genes subject to genotype × environment interactions (Arslanoglu *et al.*, 2011; Patil *et al.*, 2018). This nature of inheritance makes it extremely hard to understand the genetic basis of such complex traits (Patil *et al.*, 2018). Many quantitative trait loci (QTLs) associated with protein and oil contents in soybean have been reported in many studies over the past two decades (SoyBase, https://soybase.org/). Most of these QTLs have been identified based on populations derived from crosses of two parents with contrasting seed protein and oil concentration; detected in many different genomic regions throughout the 20 chromosomes of soybean (Cao *et al.*, 2017; Chung *et al.*, 2003; Diers *et al.*, 1992). However, most of these QTLs have low selection

accuracy and have not been used effectively in Marker Assisted Selection (MAS) to breed for high protein and oil content in soybean varieties (Cao *et al.*, 2017; Hwang *et al.*, 2014; Li *et al.*, 2018). Therefore, the identified molecular markers that are associated with QTLs can generally be used only in populations for which the markers were specifically developed and the confidence intervals covered by QTLs are wide; usually 20cM or more (Dias *et al.*, 2017; Hwang *et al.*, 2014; Li *et al.*, 2018).

Several efforts have been made to understand the genetic basis of these complex quantities traits using advanced sequencing technologies like genotype-by-sequencing (GBS) which is based on single nucleotide polymorphisms (SNPs) (Dias *et al.*, 2017; Hwang *et al.*, 2014; Leamy *et al.*, 2017; Li *et al.*, 2018; Sonah *et al.*, 2015). SNPs are abundant in most plant genomes but have previously been extremely costly and time-consuming for application to most plant breeding programs. Advances in sequencing technologies have offered new opportunities for high throughput and low-cost crop genotyping that has provided more for most plant breeding programs to utilize the abundant SNP markers (Kim *et al.,* 2016). Genome-wide association study (GWAS) is an excellent approach to discover genetic factors in a population of non-cross-derived populations and provides higher mapping resolution than conventional QTL mapping. GWAS uses collections of diverse lines that have been genotyped and phenotyped for certain traits of interest and statistical associations between the SNPs and traits are further investigated to identify genomic loci linked with the quantitative trait (Hwang *et al.,* 2014; Korte and Farlow 2013).

A study by Hwang *et al.* (2014) identified 40 SNPs located in 17 different genomic regions in 10 chromosomes for protein content and 25 SNPs located in 13 regions on 12 of the 20 chromosomes for oil content in diverse soybean accessions. Another GWAS study that was conducted for protein and oil content on USA soybean accessions identified SNPs with strong signals on chromosomes 20 and 15. The same study further identified three candidate genes for protein and oil content on the chromosome 20 region (Bandillo *et al.,* 2015). GWAS is useful to identify genes that code for important complex traits in crops such as those with self-pollinating mating systems (Korte and Farlow 2013). Hence the application of GWAS in soybean, a highly

self-pollinated crop with a complex genome structure, remains to be explored, especially for nutritional traits that are very difficult and expensive to assess. Therefore, this study was designed to contribute to the breeding for improved nutritional traits in soybean through the identification of SNP markers and candidate genes associated with high contents of protein and oil.

5.2 MATERIALS AND METHODS

5.2.1 Plant material

A total of 92 soybean genotypes from different sources with varying nutrient profiles (Obua *et al*., 2020a) and genetic diversity (Obua *et al*., 2020b) were used in this study. Forty-seven genotypes were from Uganda, 14 from Japan, 6 from the USA, 12 from The World Vegetable Centre in Taiwan (AVRDC) in Taiwan, and 13 from Seed Co; a seed Company from Zimbabwe. The genotypes from Uganda included landraces, released varieties, and elite lines while those from AVRDC are mainly vegetable soybean that are consumed when the pods are still green. The genotypes from Seed Co (a private seed company based in southern Africa) are all released varieties that are grown by many farmers across Africa. The genotypes used in this study were also used in studies 1 and 2.

5.2.2 Field trials and sample preparation

The soybean genotypes were grown in the first season of 2016 (March - July); planted at Makerere University Agricultural Research Institute Kabanyolo (MUARIK) which is located in Central Uganda. The genotypes were planted in an alpha lattice design with three replicates. The plot size was 5m long with a spacing of 60 cm × 5 cm. The trials were kept weed-free and three weedings were conducted each season. No agrochemicals were used on the trials to control pests. The grain from each plot was harvested and dried to ~ 10% moisture content. The seed samples for each genotype were divided into three replicates and about 300g ground through a 2 mm screen using Cyclotec^{TM} 1093 mill (FOSS, 2011). The milled samples were stored in the cold room at a temperature -4^{o}C at the Nutritional Platform - Biosciences eastern and central Africa - International Livestock Research Institute (BecA - ILRI) Hub, Nairobi Kenya, where all the subsequent chemical analysis was performed.

5.2.3 Measurement of Total Protein Content

Protein concentration for each genotype was determined using the Modified Folin-Lowry Method (Lowry *et al.*, 1951). One hundred mg of the milled samples were weighed in duplicate into 25 mL culture tubes. Five ml 5% sodium dodecyl sulphate (SDS) was added, vortexed, and left to stand for 2 hours at room temperature and centrifuged at 2000 ppm for 10 min. Fifty µl aliquot was taken and diluted with 950 µl of water and made to 1000 µl in separate culture tubes. One hundred µl aliquot of the diluted extract was taken for analysis. Blank (100 µl of distilled water), standards (0, 100, 200, 300, 400, 500 µl of Bovine serum albumin) and samples (100 µl) were pipetted into glass culture tubes, and all made to 1000 µl.

One ml Reagent A (0.4 volumes of water, 1 volume CTC reagent, 1.6 volumes of 5 % SDS and I volume of 0.8 M NaOH) was then added to each of the tubes and immediately vortexed; at 20 seconds interval. Five hundred µl of Reagent B (1 volume of Folin-Ciocalteu phenol reagent, 2 Aldrich 9252, and 5 volumes of distilled water) was added to each tube and immediately vortexed and left to stand for 30 minutes for color development. An ultraviolet-visible spectrophotometer was used to measure the absorbance of the standards and samples versus the blank at a wavelength setting of 750 nm. The absorbance values of the standards versus their corresponding protein concentrations were plotted to prepare a calibration curve, and the protein concentration of the samples was determined.

The total protein (crude) in the residue was calculated using the formulae:

$$\text{Total Protein (g/100 g)} = \frac{C \times 100 \times DF}{10^6 \times W}$$

Where;

C	= Concentration obtained from the calibration in µg/ml
100	= Conversion factor to report results in g/100g
DF	= Total dilution factor (1000)
10^6	= Conversion from µg to g
W	= Weight of the sample in grams

5.2.4 Measurement of Total Oil Content

Total oil content for each genotype was determined for the three replications as in the trial using the chloroform/methanol gravimetric method (Bligh and Dyer, 1959). Two gram of the milled sample for each soybean genotype was weighed in duplicate into 50 mL culture tubes ($\boldsymbol{W_1}$). Thirty-two ml of Clarase solution was added, the tubes were capped and gently shaken until the sample was well mixed with the enzyme solution. The sample was incubated for one hour in a 45^0C water bath while gently mixing by inversion after every 20 minutes. All the extract was transferred to a 250 ml polypropylene bottle, capped and centrifuged at 2000 rpm for 15 minutes to clarify the chloroform. The top aqueous phase was carefully removed and discarded with a tap aspirator pump leaving a 2-4 mm thick layer on the chloroform. A hole was cautiously broken into the surface crust with a glass rod, and 20.0 ml of the chloroform extract was pipetted into a pre-weighed 50 ml beaker ($\boldsymbol{W_2}$). Further, a 20 ml aliquot of the chloroform extract was taken and stored at -20 °C for fatty acids methyl esters (FAMES) analysis with Gas Chromatography-Mass Spectrometry (GC – MS). The solution was evaporated to dryness by leaving it overnight in a fume hood; the beaker was placed in an oven at 102^oC for 30 minutes, removed, and cooled in an evacuated desiccator for 1 hour. The beaker plus the total oil was weighed on a microbalance to the nearest 0.1 mg ($\boldsymbol{W_3}$).

The total oil in the residue was calculated using the formulae:

$$\text{Total oil (g/100 g)} = \frac{(W_3) - (W_2) \text{ X } 100 \text{ X } 4}{(W_1)}$$

Where; W_1 - Sample weight (g)

W_2 - Weight of beaker

W_3 - Weight of beaker + total oil

5.2.5 Quality control for Protein and Oil determination

Both methods for the determination of protein and oil content were validated by running a certified reference sample BCR 708 from the European Commission Joint Research Centre. Additionally, the protein determination method was subjected to inter-laboratory comparison by participating in proficiency testing (PT) 10158 - proximate in soybean meal Sept-Nov 2018 organized by Fapas- Fera Science LTD (Sand Hutton, York, UK) where satisfactory results were obtained with a Z-score of -0.4. Both the certified reference sample and the remnant of the

soybean PT sample were routinely analyzed per batch of samples analyzed each day for both protein and oil content.

5.2.6 DNA Extraction, Determination of DNA Quality and Quantity

Seeds from the 92 genotypes were grown under controlled greenhouse conditions at Biosciences eastern and central Africa - International Livestock Research Institute (BecA - ILRI) Hub, Kenya. Twelve days after germination, one young leaf from one plant from each genotype was harvested and DNA extracted using ZR Plant / Seed DNA MiniPrep™ according to manufacturer's protocol (https://files.zymoresearch.com/protocols/_d6020_quick-dna_plant-seed_miniprep_kit.pdf). The DNA quality was first checked on 0.8% (w/v) agarose gel in 1 X Tris-acetate EDTA buffer and run at 80V for 45 Minutes. The run gels were photographed usil]ng GelDoc-It™ Imager (UVP) and the picture image was interpreted for DNA quality. The DNA was quantified using Thermo Scientific Nanodrop 2000C Spectrophotometer and stored at 4 °C.

5.2.7 SNP Genotyping and Quality Control

The DNA samples were genotyped at Diversity Arrays Technology (DArTSeq™) in Australia using the Illumina HiSeq 2500. High-throughput genotyping was conducted in 96 plex DArTseq protocol. The samples were genotyped following an integrated DArT and genotyping-by-sequencing (GBS) methodology involving complexity reduction of the genomic DNA to remove repetitive sequences using methylation-sensitive restrictive enzymes before sequencing on next-generation sequencing platforms (Kilian *et al.*, 2012). The soybean reference genome and annotation were downloaded from ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Gmax. The sequence data generated were then aligned to the soybean reference genome sequence, Soybean_v7, resulting in a raw dataset of 16,688 SNPs markers. The 16,688 SNP markers were filtered to eliminate SNPs with unknown or multiple chromosome locations. LD pruning with a window size of 50, window increment of 5, LD statistic r^2, r^2 threshold of 0.5 and LD method of composite haplotype method (CHM) algorithm in SVS software. The data was further filtered with a call rate of <0.9, the minor allele frequency of <0.01and Fisher's HWE $p < 0.0001$. Identity by descent (IBD) was further conducted in SVS software to eliminate duplicated genotypes and assess sample contamination. After quality filtering, a total of 4,570 SNP markers

distributed across the 20 soybean chromosomes were used for Principal component analysis (PCA) and Genome-Wide Association (GWAS) analyses.

5.2.8 Statistical Analysis

Statistical differences among the different soybean genotypes were determined using ANOVA at 5% level ($\alpha = 0.05$) of significance for both protein and oil using Genstat, 13th Edition (Payne *et al*. 2010). Pairwise comparison of mean by Least Significant Differences test (LSD) was carried out using Genstat, 13th Edition (Payne *et al*. 2010). Correlation analysis was performed between protein and oil content using the corrplot package for graphical display using the Pearson method in R software.

Principal component analysis (PCA) was used to assess the population structure of the test genotypes using SNP & Variation Suite software. A compressed mixed linear model (MLM) in SNP & Variation Suite was used for the GWAS based on the SNPs from the 92 soybean genotypes. For this study, significance level 0.001 (threshold P-value) was used (Hwang *et al*. 2014; Li *et al*. 2018). The seed protein and oil genomic QTL locations from previous studies were compared with the physical positions of the markers exhibiting significant associations in this study as a means of verifying the identified genomic regions.

To identify candidate genes underlying the association signals detected by GWAS for protein and oil content, the soybean reference genome annotation accessible through ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Gmax was exploited. The physical positions of significant SNPs were searched on the soybean genome browser to discover the relevant genes closest to the SNPs. Annotated functions of the surrounding genes were further investigated from the soybean reference genome.

5.3 RESULTS

5.3.1 Protein and oil phenotypes

Wide and continuous phenotypic distributions were observed for both protein and oil content (Figure 9). Analysis of variance indicated that protein and oil contents were significantly different among the 92 soybean genotypes. A very weak negative correlation ($r = -0.0006$) between seed protein and oil contents was found. The protein content of the 92 soybean

genotypes ranged between 32.7% to 48.1% with a mean of 40.8% and the range of 38-40% had the highest frequency. On the other hand, oil content varied from 12.5 to 25.1 % with a mean of 19.5% and the range of 19-20% had the highest frequency.

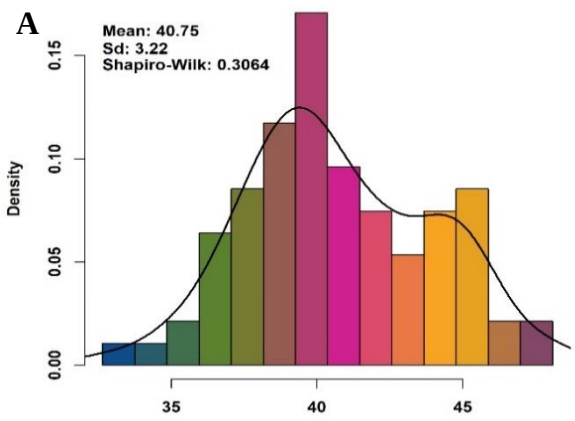


Figure 9a: Frequency Distribution for Seed Protein Content (%) of 92 soybean genotypes

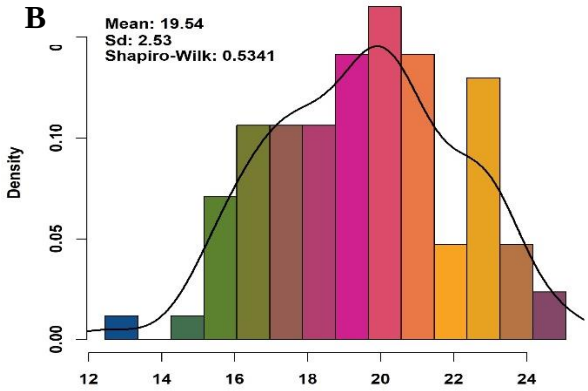


Figure 9b: Frequency Distribution for Seed Oil Content (%) of 92 soybean genotypes

5.3.2 Distribution of SNP Markers

The 92 soybean genotypes were genotyped using SNPs and after being filtered and quality assessed, a total of 4,570 SNP markers were available to construct physical maps. The 4,570 SNP markers were grouped into 20 chromosomes. The physical distance of 20 chromosomes ranged from 37 Mb (Chr. 16) to 62 Mb (Chr. 18). The largest chromosome (Chr. 18) had 338 SNP markers while the smallest chromosome (Chr. 16) had 20 SNP markers. The mean chromosome length was 102.73 Mb and each chromosome contained an average of 103 SNP markers (Figure 10).

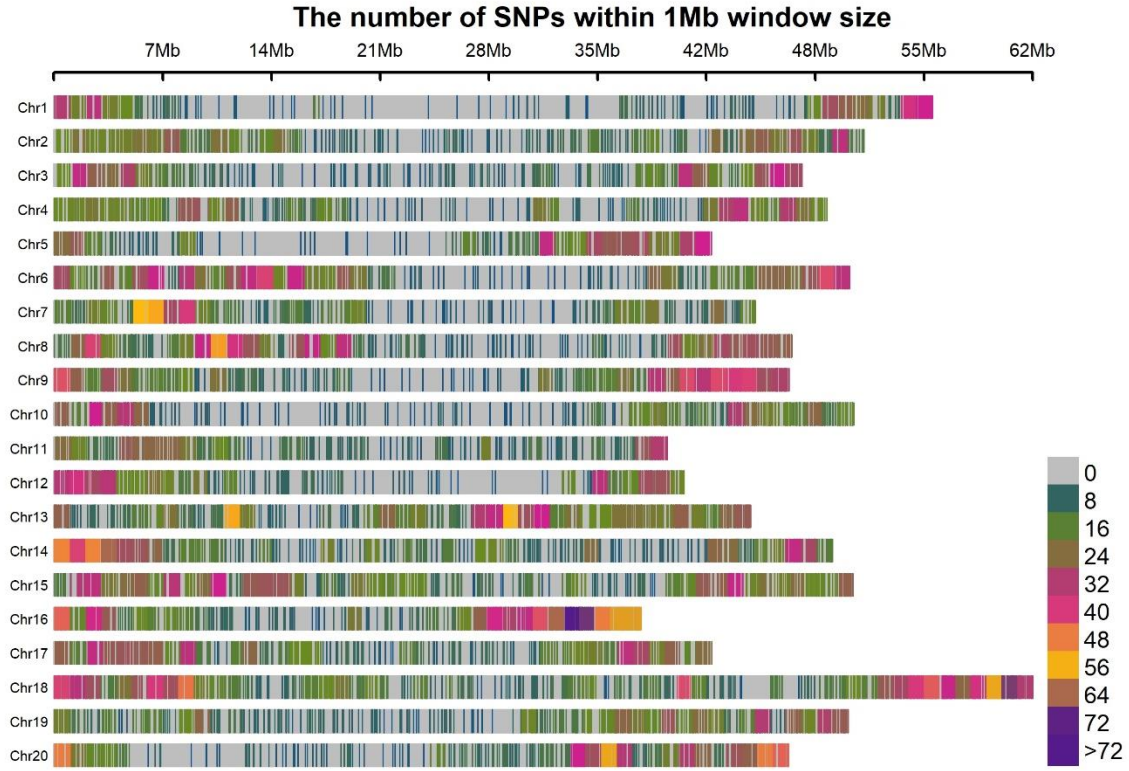


Figure 10: Distribution of SNP Markers across the 20 soybean chromosomes

5.3.3 Principal Component Analysis (PCA)

Principal Component 1 (PC1) explained 5.50% of the variation in the genotypic data, while PC2 explained 4.20% of the variation, respectively. Based on PCA, the soybean genotypes in this study were grouped into three distinctive clusters (Figure 11).

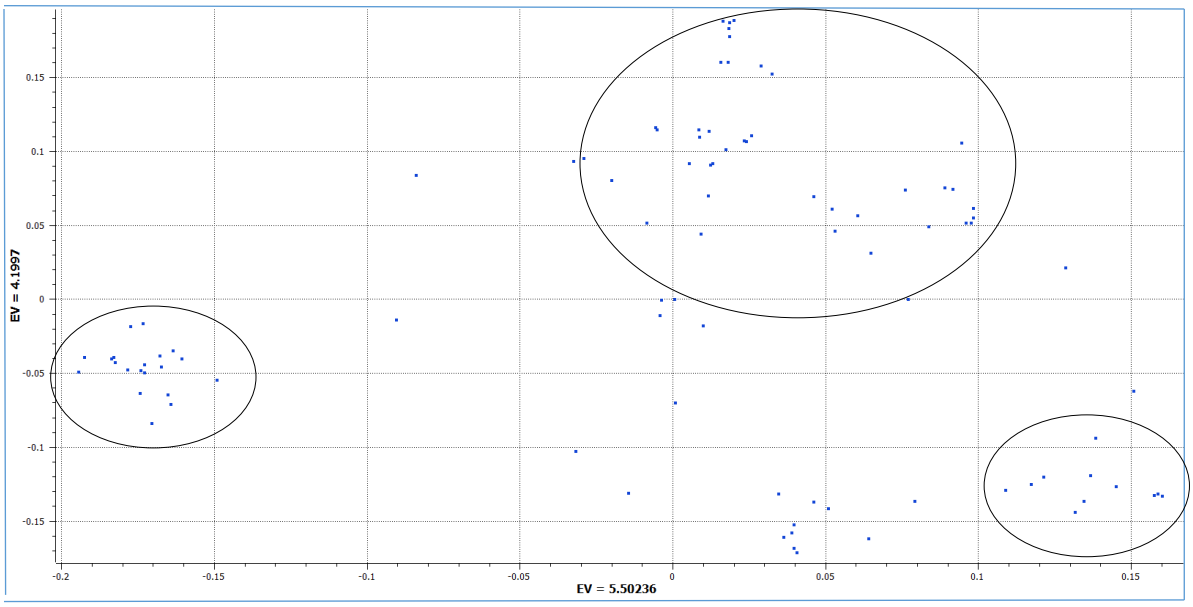


Figure 11: Plot of PC1 (5.50%) and PC2 (4.20%) from principal component analysis for 92 soybean genotypes

5.3.4 GWAS of Total Oil and Protein Contents

A genome-wide association study was performed using the mixed linear model (MLM) that greatly reduced false-positive rates in the data as shown in the quantile-quantile (QQ) plots (Figures 12 and 13). The Manhattan plots of association of oil and protein content among the genotypes are presented in Figures 14 and 15 respectively. The GWAS revealed two significant associations (-log[*P-value*] > 5) with oil content for two SNPs, rs2291820 and rs22918919 on chromosomes 7 and 10 respectively (Figure 14, Table 4). A significant association (-log[*P-value*] > 2.5) with protein content was detected for 3 SNPs, rs 22918920, rs 22918919 and rs 1494480 located on chromosomes 7, 10 and 20 respectively (Figure 15, Table 4). Both rs2291820 and rs22918919 located on chromosomes 7 and 10 respectively were associated with both oil and protein content (Table 4). Notably, three genes (LOC100305928, LOC100816464, and 1-3-1B) were identified within the same region where SNP rs14974480 was mapped on chromosome 20 (Figures 16-18).

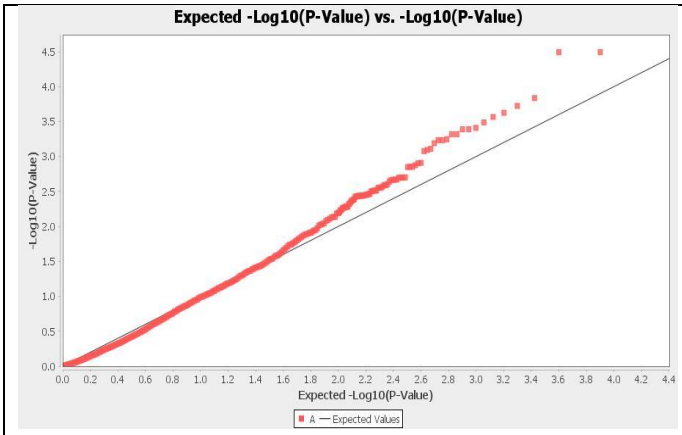


Figure 12: QQ Plot for Total Oil Content

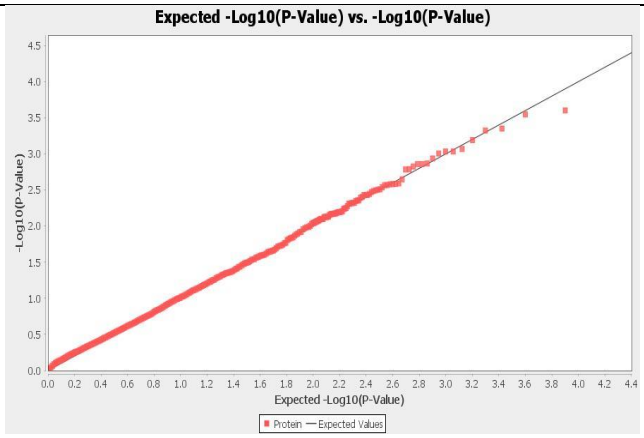


Figure 13: QQ Plot for Total Protein Content

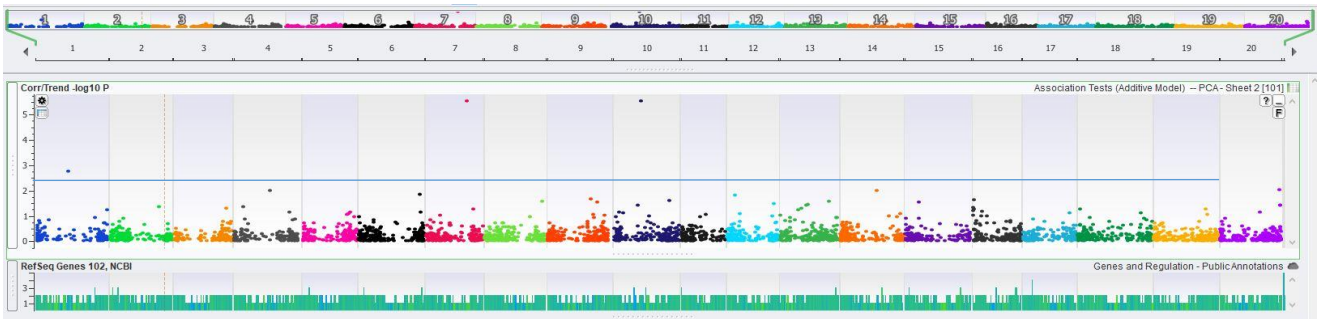


Figure 14: Genome-wide Manhattan plots of associations for oil content. The x-axis indicates the SNPs along with each chromosome; the y-axis is the −log 10 (P-value) for the association.

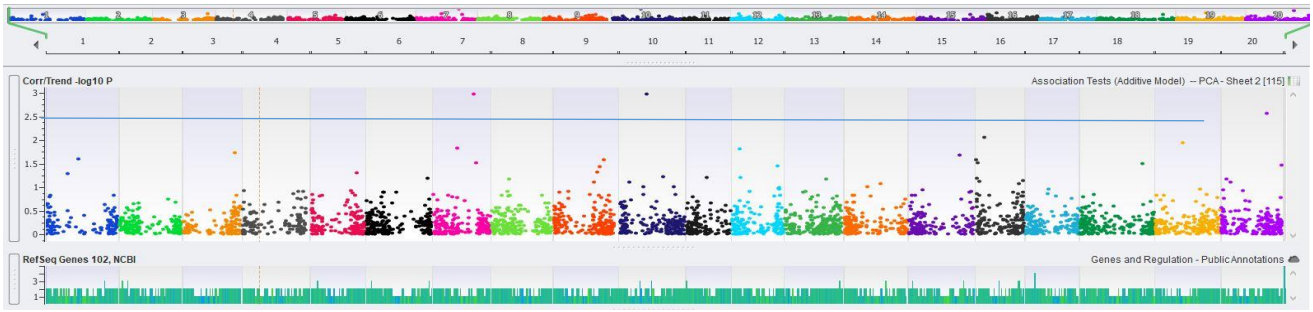


Figure 15: Genome-wide Manhattan plots of associations for protein content. The x-axis indicates the SNPs along with each chromosome; the y-axis is the −log 10 (P-value) for the association.

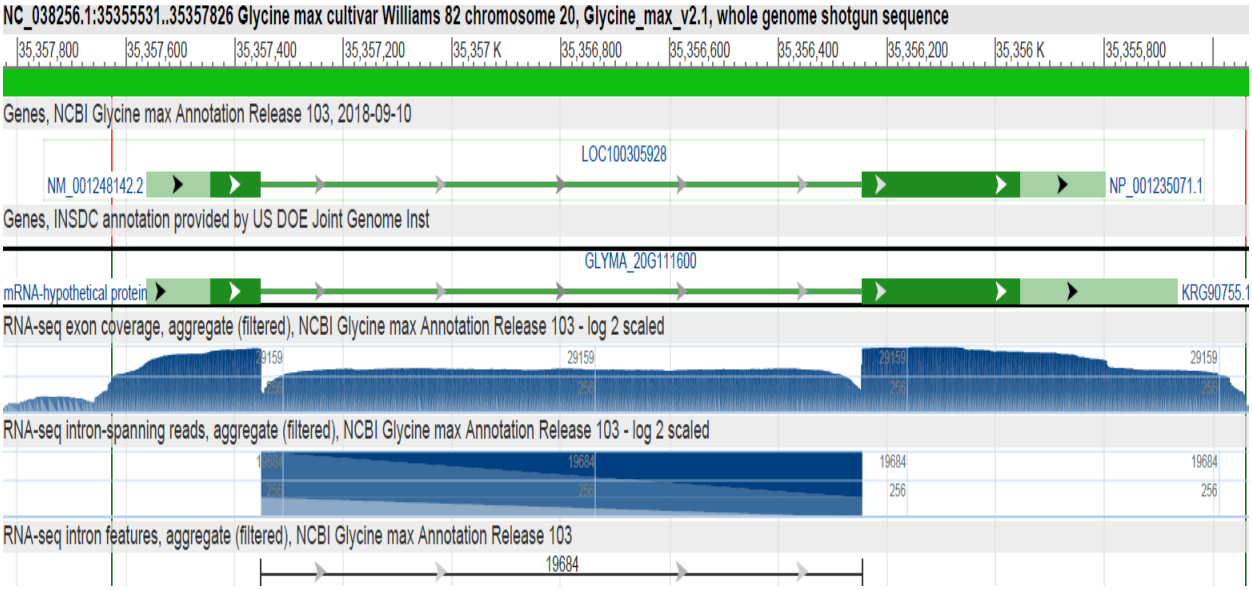


Figure 16: The candidate region of the major seed protein QTL on Gm20 showing LOC100305928 (Glyma.20G111600)

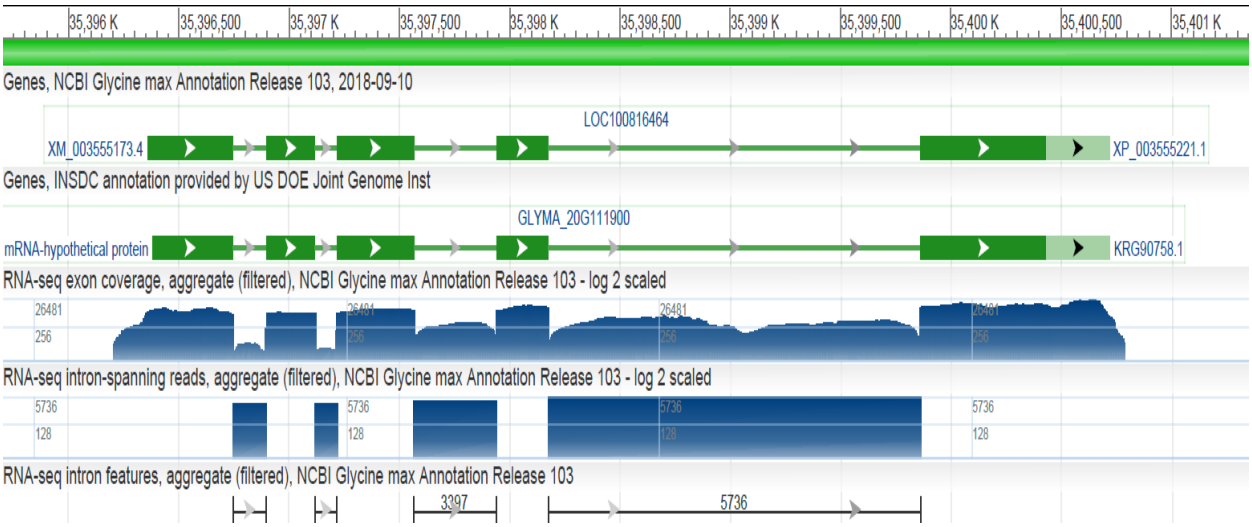


Figure 17: The candidate region of the major seed protein QTL on Gm20 showing LOC100816464 (Glyma.20G111900)

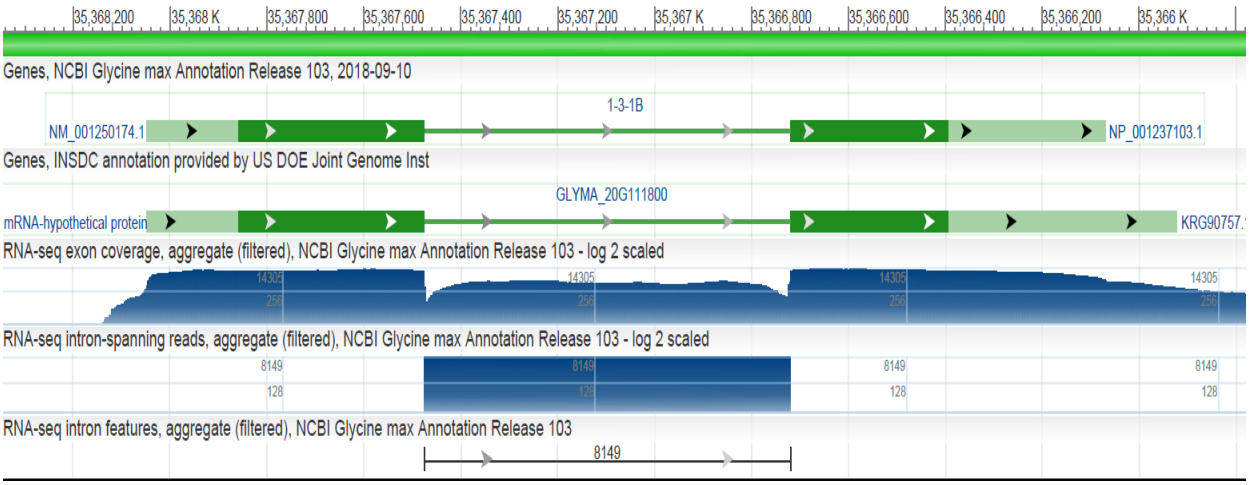


Figure 18: The candidate region of the major seed protein QTL on Gm20 showing 1-3-1B (Glyma.20G111800)

Table 4: SNPs associated with oil and protein content

Trait	SNP	Chr	Physical position	P	-Log_{10}P	Minor Allele	Major Allele
Oil	rs22918920	7	31,715,312	2.90e-06	5.54	A	C
Oil	rs22918919	10	21,100,154	2.90e-06	5.54	A	T
Protein	rs22918920	7	31,715,312	0.001	2.98	A	C
Protein	rs22918919	10	21,100,154	0.001	2.98	A	T
Protein	rs14974480	20	35,378,802	0.003	2.55	T	A

5.4 DISCUSSION

The development of soybean varieties with improved nutritional traits is an important goal of most soybean breeding programs. In this study, the distribution of protein and oil content in evaluated soybean genotypes followed normal distributions (Figs. 9a and 9b). The normal distribution indicates that both traits are quantitatively inherited and therefore, genetic control of these traits is complex, significantly affected by multiple genetic loci, the environment and the interaction between genes and the environment (Akond *et al*. 2014; Li *et al*. 2018). These observations are in agreement with Akond *et al*. (2014) and Eskandari *et al*. (2013) who observed that protein and oil content are controlled by many genes. The quantitative nature of inheritance of protein and oil content possess a big challenge in improving these two traits

because they not only controlled by multiple genes having small or large effects but also are controlled but also influenced by environments (Wang *et al*. 2020; Hyten *et al*. 2004).

Several QTLs associated with protein and oil content have been identified across the 20 chromosomes in numerous studies (http:// www.soybase.org). These QTLs have been effectively mapped using a variety of molecular marker systems including random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLPs) and simple sequences repeats (SSRs) based on linkage analysis, mainly using bi-parental populations (Chung *et al*. 2003; Diers *et al*. 1992). However, several studies have been conducted to map protein and oil QTLs using SNP markers that don't require the development of a mapping population through making crosses between genotypes with contrasting properties and are more accurate (Dias *et al*. 2017; Hwang *et al*. 2014; Li *et al*. 2018). Using a diverse collection of soybean germplasm, the current study identified QTLs associated with protein and oil content using GWAS. Protein and oil contents of 92 soybean accessions were measured and the phenotypic data was used for QTL discovery. Five SNP markers were identified; two associated with oil content (rs22918920 and rs22918919) while three associated with protein content (rs22918920, rs22918919, and rs14974480). Of the two SNPs associated with oil content, rs22918920 located on chromosome 7 collocated with previously reported QTL “seed oil 43-29” (Mao *et al*. 2013). That same SNP was also associated with protein content and collocated with previously reported QTL for protein “cqseed Protein – 006” (Pathan *et al*. 2013). SNP rs14974480 located on chromosome 20; associated with protein content collocated with seven previously reported QTLs for the following "seed protein 30-1" (Tajuddin *et al*. 2003), “seed protein 36-26” (Mao *et al*. 2013), “seed protein 37-8” (Wang *et al*. 2014), “seed protein 10-1” (Sebolt *et al*. 2000), “seed protein 34-11” (Lu *et al*. 2013) “cqseed protein-003” (Nichols *et al*. 2006) and “seed protein 31-1” (Pandurangan *et al*. 2012). The same SNP collocated five previously reported QTLs; “seed oil 34-3” (Tajuddin *et al*. 2003), “seed oil 43-17” (Mao *et al*. 2013), “seed oil 32-3” (Shibata *et al*. 2008), “cqseed oil-004” (Nichols *et al*. 2006) and “seed oil 13-4 (Specht *et al*. 2001). The SNPs that collocated with previously reported QTLs suggest that these SNPs showed reproducibility in different independent experiments. The current study further illustrated that SNP rs22918920 collocated with both protein and oil content related QTLs. This scenario has been reported in previous

studies (Li *et al*. 2018). However, protein and oil content were found to be negatively correlated, that has been reported in many studies (Akond *et al*. 2014; Li *et al*. 2018). Therefore it is essential to clarify the effects of the loci for the two target traits before initiating any soybean marker assisted breeding program.

Many QTLs for protein and oil contents have been mapped particularly on chromosome 20 in many soybean populations (Mao *et al*. 2013; Chung *et al*. 2003; Diers *et al*. 1992; Nichols *et al*. 2006; Sebolt *et al*. 2000; Tajuddin *et al*. 2003; Bolon *et al*. 2010). The high number of QTLs for both protein and oil content detected on chromosome 20 in previous studies as well as in the current study suggests that the candidate genes responsible for these two traits are most likely on chromosome 20. A study by Bolon *et al*. (2010) showed that the QTL was located in an 8.4-Mb region located between 24.5 and 32.9 Mb on chromosome 20. Another study, however, reported that the most significant SNP was identified at 31,972,955 bp on chromosome 20 (Vaughn *et al*. 2014). The current study mapped SNP rs14974480 approximately 2.5 Mb downstream of the region identified by Bolon *et al*. (2010) and 3.5 Mb by (Vaughn *et al*. (2014). The GWAS results in this study support a narrowing of the region of this major seed protein QTL on chromosome 20 to about 2.5 Mbp versus the previously defined region of 8.4 Mbp (Bolon *et al*. 2010). Therefore, there is a need to validate this SNPs marker with more soybean populations. The three genes identified included; LOC100305928 (Glyma.20G111600) that has not been characterized for a specific function (Fig. 6), LOC100816464 (Glyma.20G111900) that has been characterized for cationic amino acid transporter 7 (Fig. 7), and 1-3-1B (Glyma.20G111800) that has not been characterized for a specific function (Fig. 8). Based on its putative function, Glyma.20G111600 gene may be involved in carbon partitioning and regulating protein and oil contents in soybean. There is a need to explore further the new SNPs and genes associated with protein and oil content reported in this study. Two SNP markers rs22918920 and rs22918919, located on chromosomes 7 and 10, respectively, associated with both protein and oil content and were identified for the first time in this study. Therefore, there is a need to validate these two new SNPs markers so that they can be deployed in marker assisted breeding.

5.5 CONCLUSIONS AND RECOMMENDATIONS

Protein and oil content both showed normal distribution. This implies that both traits are quantitatively inherited and determined by the interaction of several genes subject to gene × environment interactions. The 4,570 SNP markers used in this analysis were representative because they were distributed throughout the soybean genome in all 20 chromosomes. Principal Component Analysis grouped the 92 soybean genotypes into three clusters. Additionally, GWAS identified two SNP markers associated with oil content on chromosomes 7 and 10. GWAS also identified 3 SNP markers associated with protein content on chromosomes 7, 10 and 20. The five identified SNP markers mapped in genomic regions containing QTLs previously mapped for protein and oil content could be used for marker-assisted selection of soybean genotypes for improved nutritional traits. Marker-assisted selection would greatly hasten the process required for the development of soybean varieties with improved nutritional properties.

Additionally, two SNP markers (rs22918920 and rs22918919) located on chromosomes 7 and 10 respectively were associated with both oil and protein content. There is a need to further explore these two SNP markers. These markers could simultaneously be used to trace both oil and protein content in the breeding pipeline. Additionally, three genes (LOC100305928, LOC100816464, 1-3-1B) were identified within the same region where SNP marker rs14974480 was mapped on chromosome 20. The three identified candidate genes could be used to develop new soybean varieties by introgressing into farmers' preferred varieties but low in protein and oil content. However, genes LOC100305928 and 1-3-1B have not been characterized for any specific function and there is a need to conduct more association studies using different soybean populations to confirm their functions. The information generated from this study could be used to understand the mode of transmission of the genes associated with protein and oil content in soybean. Further work on molecular cloning and functional analysis of the candidate genes will suggest their roles in protein and oil content regulation in soybean.

CHAPTER SIX

6.0 PROTEIN AND YIELD STABILITY AMONG TROPICAL SOYBEAN GENOTYPES IN SELECTED AGRO-ECOLOGIES IN UGANDA

6.1 INTRODUCTION

Soybean (*Glycine max*) is a vital feed and food resource in the East African region (Tukamuhabwa *et al*., 2016). Over recent decades, soybean production has undergone the greatest expansion of any global crop (Agralytica 2012). The largest expansion of soybean production occurred during the second half of the twentieth century, where production grew tenfold from 27 million tons (M.T.) in 1962 to 334 M.T. in 2019 (FAO 2019). It is expected that by 2050, production will double (WWF 2014). Around 75% of soybean produced worldwide is used for animal feed, especially for poultry and pigs (WWF 2014). The remaining soybean produced is eaten directly and a small portion is used to produce biodiesel (WWF 2014). Soybean grain contains about 40% protein, 20% oil, varied essential amino acids and nutrients, and a high-calorie value (Singh *et al*. 2008). Soybean improves soil fertility through nitrogen fixation and enhanced moisture retention, leading to a more sustainable cropping system (Graham and Vance, 2003). The Nitrogen-fixing ability of soybean makes it a good crop for sustainable agricultural systems of tropical Africa, which is characterized by infertile soils and low fertilizer usage. Therefore the need for Nitrogen fertilizer input is minimized. Though introduced late in Africa, soybean is cultivated as both a food and cash crop (Sanginga *et al*. 1999; Tukamuhabwa *et al*., 2016). Soybean production is expected to increase rapidly as economic development leads to higher animal protein consumption, especially in developing and emerging economies. Therefore, there is a need to promote soybean as a cash and food crop in Tropical Africa.

Makerere University Centre for Soybean Improvement and Development (MAKCSID) has developed several elite soybean genotypes that were evaluated at advanced yield trials (AYT) to assess their adaptability in the major soybean production areas of Uganda (Tukamuhabwa *et al.*, 2011; Tukamuhabwa *et al.*, 2012). However, very little is known about the environments' effect

on nutritional traits, especially the protein content of these soybean genotypes, when grown in different multi-locational field trials. Few studies have assessed the behaviour of protein content of soybean genotypes in a multi-locational trial. Kumar *et al.* (2006) investigated the effect of the environment on the physical properties and biochemical composition of seven Indian soybean cultivars at four growing locations in a multi-locational field trial. They found a significant effect of environment, genotype, and genotype × environment interaction on protein. Similarly, Arslanoglu *et al.* (2011) investigated the effects of genotype and environment interaction on protein and oil content of eight soybean genotypes in eight different environments in Turkey's Middle Black Sea region. They found the effects of the genotype, environment, and their interactions on soybean seeds' protein content to be statistically significant. Several studies have reported that high temperatures and less rainfall during soybean seed development and throughout the whole season tend to increase protein content (Kumar *et al*., 2006; Ojo *et al*., 2002; Piper and Boote 1999).

Yield stability in soybean has been studied more than nutritional traits. A study conducted in four different locations of Ethiopia for two consecutive years using thirty-two genotypes showed a crossover type of GEI for grain yield. The same study also identified three genotypes with high mean yield and high stability performance across the test environments (Mulugeta *et al*. 2013). Adie *et al.* (2014) evaluated 10 black seeded soybean genotypes in 16 locations and found genotype W9837 × Cikuray-66 as most stable and was recommended for release as a new high-yielding variety. In Zambia, a MET analysis reported that the best genotype for general adaptability was the variety TGX 1988-22F. This genotype was stable across all the locations with high yields and average stability (Cheelo *et al*., 2017).

The relationships between protein content and yield of soybean are mainly unknown. Several studies have reported inconsistent relationships between protein content and yield within each genotype when grown in different locations and seasons (Anthony *et al*. 2012; Yin and Vyn 2014). Different studies have reported both positive (Anthony *et al*. 2012; Assefa *et al*. 2019) and negative (Filho *et al*., 2004; Helms and Orf 1998; Wilcox and Shibles 2001) correlations. The quantitative nature of inheritance of protein content and yield in soybean further influences

genotypes' response under different environment and management conditions (Arslanoglu *et al*., 2011; Ojo *et al*., 2002; Piper and Boote 1999; Rotundo and Westgate 2009; Vollmann *et al*. 2000). The inconsistent findings for the protein and yield relationship and the quantitative nature of inheritance of both traits reported in the previous studies make it extremely challenging to improve the two traits in soybean simultaneously. Therefore, the objectives of this study were to; (i) determine protein content and protein stability of 30 elite soybean genotypes in eight locations that represent the major soybean growing areas of Uganda, (ii) assess grain yield performance and stability in soybean, and (iii) determine the relationship between protein and yield in soybean.

6.2 MATERIALS AND METHODS

6.2.1 Experimental Materials

The soybean genotypes included 28 elite breeding lines developed by Makerere University Centre for Soybean Improvement and Development (MAKCSID) and two check varieties (Maksoy 3N and Maksoy 4N) (Table 5). Maksoy 3N is a farmer-preferred variety because of its large seed, high grain yield and high oil content, while Maksoy 4N is another high-yielding soybean variety released in Uganda. All the experimental materials were derived from 3 crosses made at the National Crops Resources Research Institute (NaCRRI) and Makerere University Agricultural Research Institute Kabanyolo (MUARIK) Uganda. The parental lines were Nam 2, a farmers' preferred variety; Duiker introduced from Zimbabwe has desirable farmers' seed traits such as large seed, cream colour, and white helium; GC0038-29 is early maturing and resistant to soybean rust disease. The three bi-parental populations were advanced from F_2 to F_6 generation using a modified single-seed descent (SSD) selection method, where one pod was used instead of a single seed. From F_7 generation, single plant selections were made to identify soybean plants with desirable traits such as high yield, early maturity, resistance to major diseases and insect pests, resistance to lodging, and pod shattering. These selected single plants were planted in single rows, and seed from every row was harvested in isolation and used in a replicated preliminary yield trial at MUARIK in 2013A. The initial yield trial was evaluated in an intermediate yield trial at two locations in 2013B (NaCRRI and MUARIK). The seed from the intermediate yield trial was evaluated further in an advanced yield trial conducted at eight multi-locations that represented Uganda's major soybean growing areas.

Table 5: Description of genotypes used in the study

Genotype	Pedigree	Comment
BSPS 48A-28	GC0038-29 × Duiker	Advanced line
BSPS 48A-9-2	GC0038-29 × Duiker	Advanced line
Nam 2 × GC 44.2	Nam 2 × GC0038-29	Advanced line
BSPS 48A-25	GC0038-29 × Duiker	Advanced line
BSPS 48A-27-1	GC0038-29 × Duiker	Advanced line
BSPS 48A-3B	GC0038-29 × Duiker	Advanced line
Nam 2 × GC 13.2	Nam 2 × GC0038-29	Advanced line
MAKSOY 4N	GC0038-29 × Duiker	Check Variety
BSPS 48A-31	GC0038-29 × Duiker	Advanced line
MAKSOY 3N	GC0038-29 × Duiker	Check Variety
NGDT 8.11-11B	Nam 2 × GC0038-29	Advanced line
BSPS 48A-8	GC0038-29 × Duiker	Advanced line
BSPS 48A-26	GC0038-29 × Duiker	Advanced line
MNG 11.2	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 35.3	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 17.3	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 44.3	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 43.2	Nam 2 × GC0038-29	Advanced line
BSPS 48A-5	GC0038-29 × Duiker	Advanced line
Nam 2 × GC 28.2B	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 11.2	Nam 2 × GC0038-29	Advanced line
NGDT 8.11-4	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 7.2	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 20.3	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 4.8	Nam 2 × GC0038-29	Advanced line
NGDT 8.11-19	Nam 2 × GC0038-29	Advanced line
NGDT 4.11-5	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 30B	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 32.6	Nam 2 × GC0038-29	Advanced line
Nam 2 × GC 43.1	Nam 2 × GC0038-29	Advanced line

Source: MAKCSID

6.2.2 Description of the Test Environments

The grain yield trial was conducted in eight locations, representing Uganda's major soybean growing areas (Table 6). Three locations, NaCRRI, MUARIK and Nakabango are situated in the Lake Victoria Crescent; while Bulindi in the Western Grasslands; Ngetta in the North-Western

savannah grasslands; Iki-iki in the Kyoga plains; Abi in North-Western Farmlands Wooded Savanna and Mubuku in Western Medium High Farmlands. These locations have different climatic conditions that influence soybean yield (Table 6). Mubuku irrigation scheme was selected to assess the adaptability of the soybean genotypes under irrigation conditions. Protein content analysis was conducted in all the above locations except for MUARIK and Bulindi because of limited financial resources.

Table 6: Description of the locations used to evaluate soybean genotypes for yield in six seasons in Uganda

Location	Position	Region	Altitude (masl)	Mean annual temperature (^{o} C)	Mean annual rainfall (mm)
NaCRRI	0^{o}32'N/32^{o}37'E	Central	1,160	22.6	1,400
Nakabango	0^{o}29'N/33°14'E	Eastern	1,210	22.8	1,400
Iki-Iki	1^{o}06'N/34^{o}00'E	Eastern	1,156	24.7	1,200
Ngetta	2°17'N/32°56'E	Northern	1,103	24.7	1,200
Mubuku	0^{o}13'N/30^{o}08'E	Western	1,007	27.8	750
MUARIK	0^{o}28'N/32^{o}36'E	Central	1,180	21.4	1,234
Bulindi	1^{o}41'N/31^{o}42'E	Mid-West	1,122	22.9	1,355
Abi	3^{o}04'N/30^{o}56'E	West Nile	1,214	22.9	1,404

Source: Meteorological station data at the study locations; masl=meters above sea level; Used flood irrigation at Mubuku

6.2.3 Experimental Design, Data Collection and Analysis

The soybean genotypes were planted in a randomized complete block design (RCBD) with three replications. Each genotype was represented by three rows measuring 5 m long with a spacing of 60 cm between rows and 5 cm between plants within a row. The multi-locational trial was conducted for six consecutive seasons; first rains of 2014 (2014 A), second rains of 2014 (2014 B), first rains of 2015 (2015A), second rains of 2015 (2015 B), first rains of 2016 (2016A) and second rains of 2016 (2016 B). The trials were kept weed-free and three weedings were conducted each season. No agrochemicals were used on the trials to control pests. Each genotype was harvested separately at maturity, threshed and corrected to 10% moisture content by sun-drying before determining yield per hectare. Total seed protein content analysis was conducted on soybean genotypes grown in six locations (Abi, Iki Iki, Mubuku, Nakabango, NaCRRI, Ngetta). The protein content was determined for 2016A season, from Biosciences eastern and

central Africa - International Livestock Research Institute (BecA - ILRI) Hub using the Modified Folin-Lowry Method (Lowry *et al*. 1951).

Analysis of variance (ANOVA) for both protein content and grain yield was done separately for each location and combined across locations using package agricolae in R software (R Core Team, 2020). Genotypes were considered as fixed effects and replications and blocks within replications as random effects. For the combined analysis, variances were partitioned into the relevant sources of variation to test for differences among genotypes and the presence of GE. Adjusted protein content and grain yield from ANOVA were subjected to GGE biplot analysis to decompose the GE of each trait to compare genotype stability in performance across the various environments (Yan 2001; Yan and Rajcan 2003). Correlation analysis was performed using the corrplot package for graphical display using the Pearson method in R software.

6.3 RESULTS

6.3.1 Mean Protein Content (%) performance of genotypes across locations

The ANOVA for protein content revealed a highly significant difference (P<0.001) among genotypes, locations and Genotype × Location interaction (Table 7). Nakabango had the highest mean protein content of 42.5%; followed by Mubuku and Iki Iki with a protein content of 41.3% and 41.1% respectively (Table 8). Genotype NII X GC 20.3 had the highest mean protein content of 43.0%, followed by genotypes NII X GC 43.2 (42.5 %), NII X GC 7.2 (42.3%) and BSPS 48A-31 (42.2%) (Table 8). A narrow range in performance of 4.6% between genotype with the highest protein content and that with the lowest protein content was observed across location mean performance.

Table 7: Analysis of variance of 30 soybean genotypes evaluated for protein content (%) in six locations in Uganda in 2016A

SOV	d.f.	s.s.	m.s.	v.r.	F pr.
Genotype	29	1557.65	53.71	3.50	<.001
Location	5	298.79	74.70	4.87	<.001
Genotype × Location	94	3256.39	34.64	2.26	<.001
Reps	2	7.81	3.91	0.25	0.775
Residual	413	6335.51	15.34		
Total	542	11456.15	21.14		

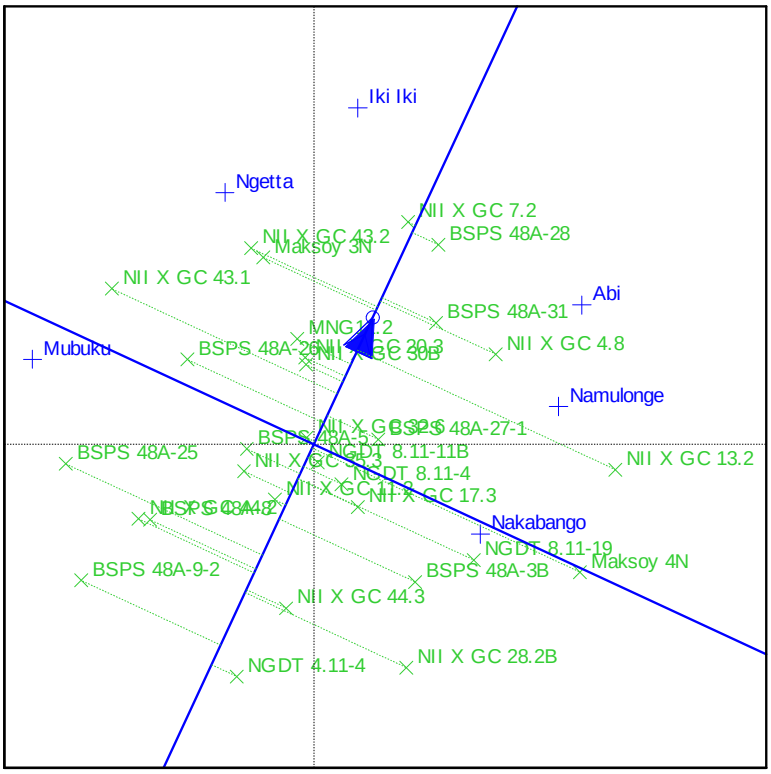
SOV = Source of Variation, d.f. = degrees of freedom, s.s. = sum of squares, m.s. = mean sum of squares, v.r. = variance ratio, F pr. = Probability value

Table 8: Protein Content (%) of 30 soybean genotypes evaluated in six locations in Uganda in 2016A

Genotype	Abi	Iki Iki	Mubuku	Nakabango	NaCRRI	Ngetta	Mean
BSPS 48A-25	34.6		45.2	39.7	39.2	43.1	40.3
BSPS 48A-26	39.5	43.3	42.6	36.4	37.1	40.1	39.8
BSPS 48A-27-1	40.3	40.9	39.3	-	41.7	40.6	40.6
BSPS 48A-28	-	45.1	40.1	42.3	41.0	-	42.1
BSPS 48A-31	43.2	42.4	39.8	44.7	41.2	42.0	42.2
BSPS 48A-3B	42.0	37.5	36.5	41.7	39.8	39.6	39.5
BSPS 48A-5	40.2	40.0	45.2	44.4	38.9		41.7
BSPS 48A-8	38.3	41.0	43.8	40.9	35.1	38.6	39.6
BSPS 48A-9-2	-	37.1	-	42.2	35.8	41.0	39.0
Maksoy 3N	-	45.5	43.6	38.5	37.8	40.2	41.1
Maksoy 4N	42.4	39.5	36.1	49.2	43.4	38.7	41.5
MNG11.2	41.2	42.3	-	40.6	-	41.5	41.4
NGDT 4.11-5	36.7	38.3	43.0	45.5	38.4	37.7	39.9
NGDT 8.11-11B	42.0	41.3	41.4	39.6	38.4	38.5	40.2
NGDT 8.11-19	41.1	39.4	34.9	-	-	-	38.4
NGDT 8.11-4	41.7	41.4	41.9	44.7	38.1	38.7	41.1
NII X GC 11.2	43.9	40.5	46.5	48.6	34.2	38.5	42.0
NII X GC 13.2	44.7	39.7	36.3	47.0	44.7	40.2	42.1
NII X GC 17.3	39.2	39.4	-	-	42.1	40.4	40.3
NII X GC 20.3	41.9		43.7	48.5	37.8	43.1	43.0
NII X GC 28.2B	40.8	38.8	35.9	42.2	38.0	36.6	38.7
NII X GC 30B	43.2	42.6	42.7	43.8	36.3	40.4	41.5
NII X GC 32.6	41.3	42.3	-	35.0	-	38.3	39.2
NII X GC 35.3	40.5	40.2	46.3	42.2	39.5	39.4	41.3
NII X GC 4.8	-	43.5	37.3	43.0	40.9	-	41.2
NII X GC 43.1	-		46.5	36.5	39.6	-	40.9
NII X GC 43.2	40.7	43.5	-	42.4	-	43.2	42.5
NII X GC 44.2	-	38.9	41.2	38.3	35.3	40.7	38.9
NII X GC 44.3	40.5	38.2	42.9	45.3	-	38.4	41.1
NII X GC 7.2	44.1	46.1	40.4	43.4	38.7	41.1	42.3
Mean	41.0	41.1	41.3	42.5	38.9	40.0	40.8
CV	5.6	12.3	8.2	10.9	11.9	7.2	9.2

6.3.2 Stability and Which-Won-Where" Patterns of Genotypes for protein content

The genotype and genotype × Location interaction was visualized in Ranking GGE, "Which-Won-Where" GGE, and the discriminating and representative environment GGE biplot analysis for protein content as represented in Figures 20-22. The GGE biplots explained 60.21% of the total interaction variation, distributed as 34.77% and 25.44% between PC1 and PC2 components, respectively. Genotype NII X GC 7.2 had high protein content and was very stable across the six locations (Figure 19). This genotype was the highest performer as it was furthest from the mean along the "average environment axis" (Yan *et al*. 2007). The biplot showed that most of the genotypes' protein content did not deviate much from the mean (Figure 19).



PC1 - 34.77%

Figure 19: A GGE ranking biplot showing mean performance and stability for protein content of 30 soybean genotypes evaluated in six locations

The Which-Won-Where" GGE biplot gave a fair visual assessment of genotype and genotype × Location interaction for protein content (Figure 20). The six locations were grouped into four mega-environments; the first one included Mubuku and Ngetta, the second one had Abi and NaCRRI, the third and fourth included Iki Iki and Nakabango respectively (Figure 20). Genotypes NII x G C 43.1, NII x G C 13.2, NII x G C 7.2 and Maksoy 4N won in the respective mega-environments (Figure 20).

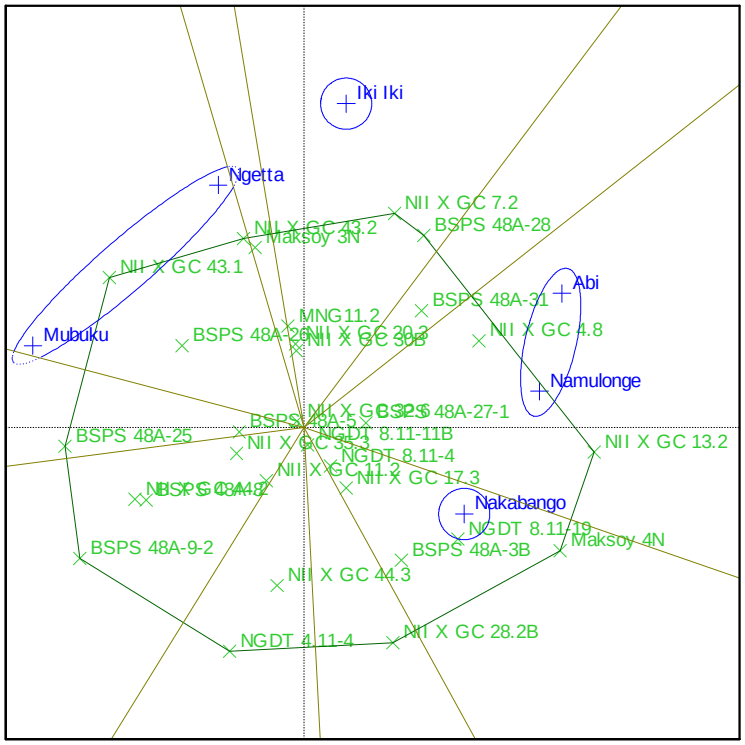
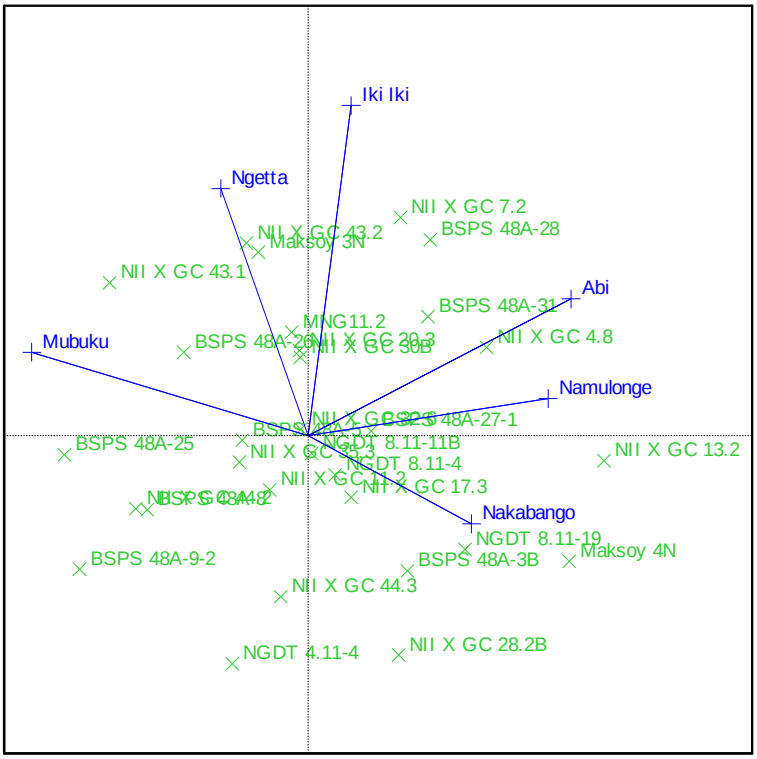


Figure 20: GGE Scatter plot based on symmetrical scaling for the "which-won-where" pattern of 30 soybean genotypes evaluated in six locations

Iki Iki environment was detected as the most discriminating environment than the other locations (Figure 21). This was revealed by the long length of Iki Iki vectors from the origin (Figure 21). NaCRRI was the most representative of the test locations than the rest due to the Average environment axis's small angle (Figure 21).



PC1 - 34.77%

Figure 21: A GGE biplot showing discriminating power and representativeness of test environments involving 30 soybean genotypes evaluated in eight locations and six seasons

6.3.3 Mean Grain Yield performance of genotypes across locations

Results showed significant differences (P<0.001) among genotypes, locations, and interactions; Genotype × Location, Location × Season for seed yield (Table 9). The results also showed that genotypes BSPS 48A-28 and BSPS 48A-9-2 had the highest mean yield of 1,207 kg ha^{-1} (Table 10). The two check varieties Maksoy 3N and Maksoy 4N had yields of 1,131 1,207 kg ha^{-1} and 1,135 1,207 kg ha^{-1} respectively. Bulindi had the highest mean yield of 1,598 kg ha^{-1}; followed by Nakabango and Mubuku with yields of 1,438 kg ha^{-1} and 1,345 kg ha^{-1}, respectively (Table 10).

Table 9: Analysis of variance of 30 soybean genotypes evaluated for grain yield in eight locations and six seasons in Uganda

SOV	d.f	s.s	F value	F Pr.
Genotype	29	19731420	4.3729	1.07E-13
Location	7	1.38E+08	126.761	< 2.2e-16
Season	5	1.44E+08	184.5312	< 2.2e-16
Rep	2	3397014	10.9162	1.92E-05
Genotype × Location	203	53562730	1.6958	2.00E-08
Genotype × Season	145	21094874	0.935	0.6958
Genotype × Location × Season	660	3.96E+08	3.8605	< 2.2e-16
Residuals	2094	3.26E+08		

SOV = Source of Variation, d.f. = degrees of freedom, s.s. = sum of squares, F pr. = Probability value

Table 10: Seed yield (kg ha-1) of 30 soybean genotypes evaluated in eight locations and six seasons in Uganda

Genotype	Abi	Bulindi	Iki Iki	MUARIK	Mubuku	Nakabango	NaCRRI	Ngetta	Mean
BSPS48A-25	1,180	1,674	810	1,024	1,302	1,487	842	1,166	1,186
BSPS48A-26	802	1,393	660	1,002	1,348	1,391	830	1,357	1,098
BSPS48A-27-1	1,013	1,384	669	1,020	1,372	1,628	892	1,260	1,155
BSPS48A-28	1,316	1,709	778	939	1,389	1,496	843	1,187	1,207
BSPS48A-31	790	1,798	580	895	1,259	1,401	1,336	1,000	1,132
BSPS48A-3B	999	1,728	657	936	1,359	1,361	1,048	1,143	1,154
BSPS48A-5	646	1,548	610	973	1,192	1,365	838	1,135	1,038
BSPS48A-8	596	1,610	761	972	1,524	1,483	865	1,017	1,104
BSPS48A-9-2	1,144	1,346	638	936	1,400	1,479	1,509	1,205	1,207
MAKSOY3N	945	1,797	662	1,062	1,335	1,276	783	1,190	1,131
MAKSOY4N	858	1,590	735	1,030	1,316	1,453	807	1,289	1,135
MNG11.2	837	1,814	661	996	1,306	1,390	697	1,067	1,096
Nam II×GC30B	637	1,282	767	837	1,256	1,228	777	1,015	975
NamII×GC11.2	655	1,438	667	806	1,463	1,038	900	1,227	1,024
NamII×GC13.2	892	1,215	649	1,023	1,564	1,536	960	1,316	1,144
NamII×GC17.3	638	901	828	958	1,528	1,460	1,096	1,187	1,074
NamII×GC20.3	619	1,130	863	948	1,414	1,253	822	1,072	1,015
NamII×GC28.2B	861	1,310	705	950	1,177	1,320	782	1,145	1,031
NamII×GC32.6	530	1,180	682	950	1,271	1,206	816	1,160	974
NamII×GC35.3	613	1,465	697	969	1,648	1,294	849	1,074	1,076
NamII×GC4.8	493	1,009	732	1,008	1,312	1,409	944	1,145	1,006
NamII×GC43.1	931	951	649	807	1,157	1,148	751	1,070	933
NamII×GC43.2	783	1,086	729	891	1,572	1,417	819	1,166	1,058
NamII×GC44.2	903	1,652	749	912	1,411	1,649	976	1,325	1,197
NamII×GC44.3	772	743	771	917	1,483	1,080	1,701	1,039	1,063
NamII×GC7.2	651	1,124	707	904	1,520	1,375	852	1,015	1,019
NGDT4.11-5	1,052	1,641	635	884	1,143	1,014	547	945	983
NGDT8.11-11B	952	1,779	681	954	1,339	1,325	848	1,095	1,122
NGDT8.11-19	849	1,763	674	898	1,260	1,051	626	922	1,005
NGDT8.11-5	705	1,733	660	821	1,185	1,245	814	1,018	1,023
Mean	935	1,598	687	981	1,345	1,438	963	1,177	1,141
LSD	504.37	523.26	413.13	249.23	328.40	314.36	485.84	503.19	

Bul = Bulindi, Iki = Iki-Iki, Mub = Mubuku, Nak = Nakabango, Nge = Ngetta

6.3.4 Stability and Which-Won-Where" Patterns of Genotypes for Yield (Kg Ha^{-1})

A ranking GGE biplot analysis showed that genotype BSPS 48A-9-2 was the best performer, though it was relatively unstable across the eight locations (Figure 22). This genotype was the highest performer as it was furthest from the mean along the "average environment axis" (Yan *et al*. 2007). On the other hand, Nam 2 × GC 30B was the most stable genotype yet very low yielding, while Nam 2 × GC 44.3 was the least stable genotype. In comparison, Nam 2 × GC 44.2 was both high-yielding and stable genotype (Figure 22).

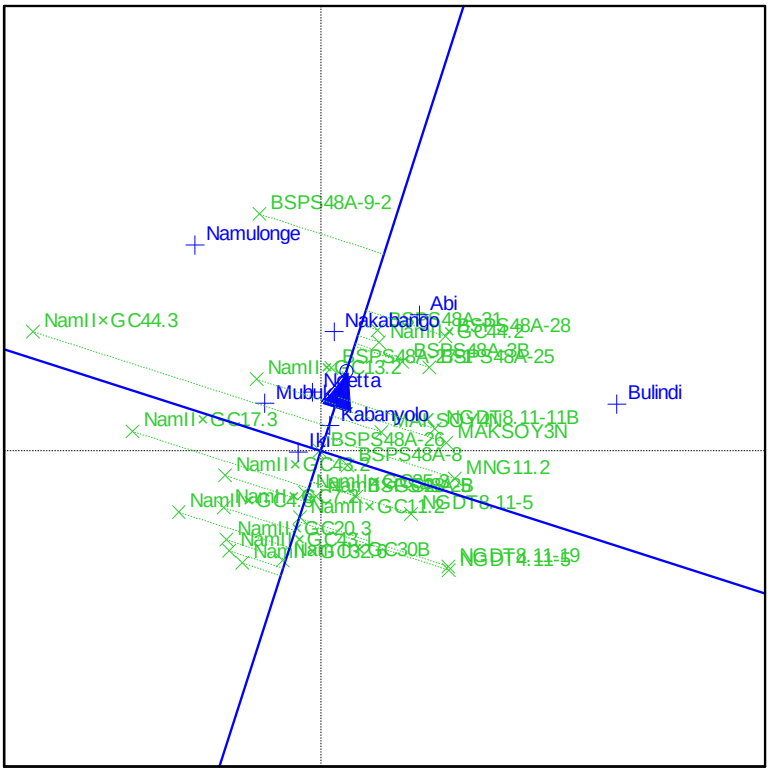


Figure 22: A GGE ranking biplot showing mean performance and stability for seed yield of 30 soybean genotypes evaluated in eight locations and six seasons

The GGE polygon plot gave a fair visual assessment of GEI with both PCA1 and PCA2 explaining about 68% of the total GEI sum of squares (Figure 23). The scatter plot indicated that the eight locations were grouped into three major mega-environments for grain yield. The first mega environment included NaCRRI, Mubuku and Iki-Iki with the best genotype being Nam 2 × GC 44.3. The second mega environment included Ngetta, Nakabango and MUARIK with the

best genotype being BSPS 48A-9-2. The last mega environment included Abi and Bulindi with the best genotype being BSPS 48A-28.

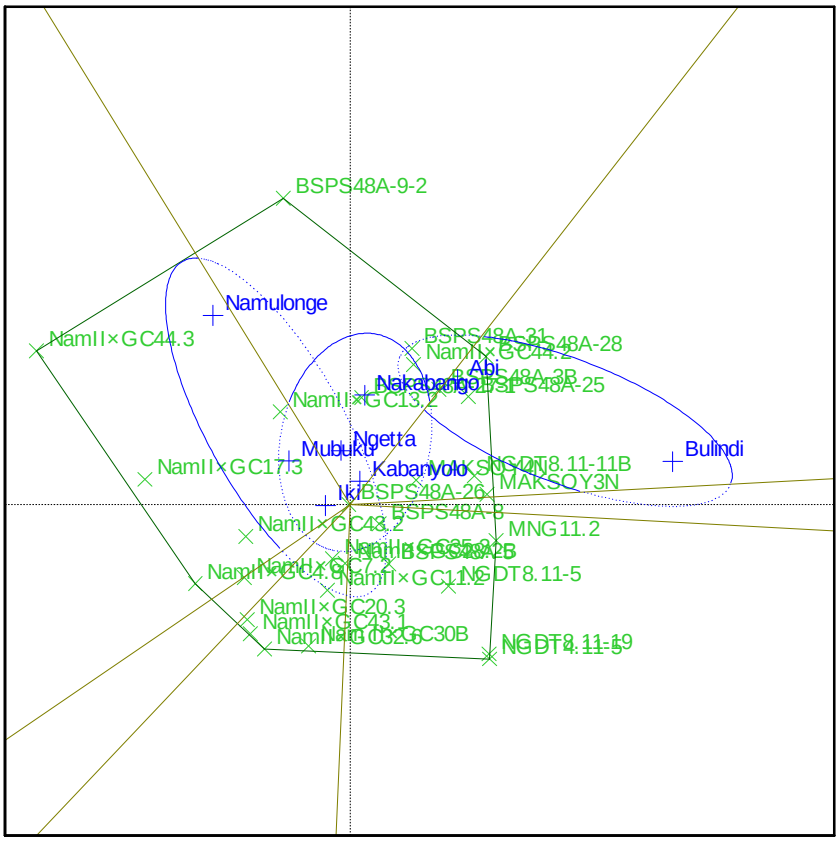


Figure 23: Polygon views of the GGE-biplot based on symmetrical scaling for the "which-won-where" pattern of 30 soybean genotypes evaluated in eight locations and six seasons.

The GGE scatter plot showed that Bulindi was the most discriminating environment, while Iki-Iki was the least of the eight locations (Figure 24). This was revealed by the long and short environment vectors of Bulindi and Iki-Iki, respectively. Bulindi was the most representative of the mega-environment of all the eight test environments than the rest due to the small angle from the Average environment axis (Figure 24).

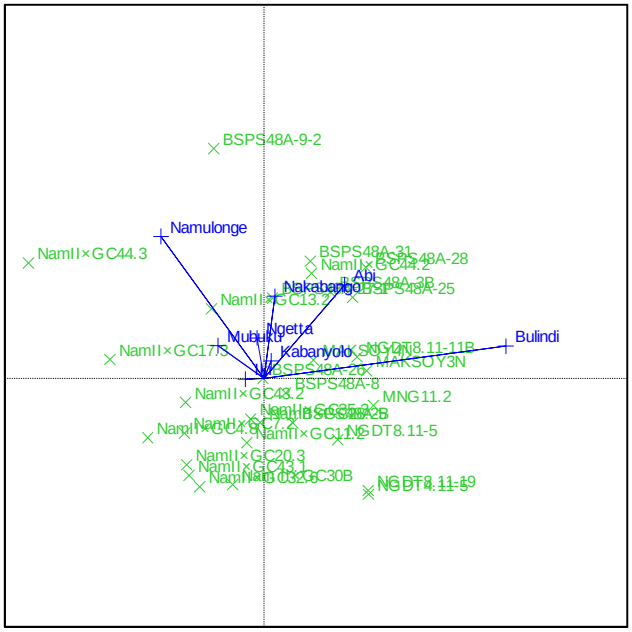


Figure 24: A GGE biplot showing discriminating power and representativeness of test environments involving 30 soybean genotypes evaluated in eight locations and six seasons

6.3.5 Correlation Analysis for Protein Content (%) and Yield (Kg Ha^{-1})

A significant, weak negative correlation (-0.1) was found between protein content (%) and yield (Kg Ha $^{-1}$) (Figure 25). This was reflected in genotype performance. For example, genotype Nam 2 × GC 44.3 had high protein content, while its yield was low. On the other hand, genotypes BSPS 48A-9-2 and Nam 2 × GC 44.2 had higher yields than all the different genotypes but had low protein content.

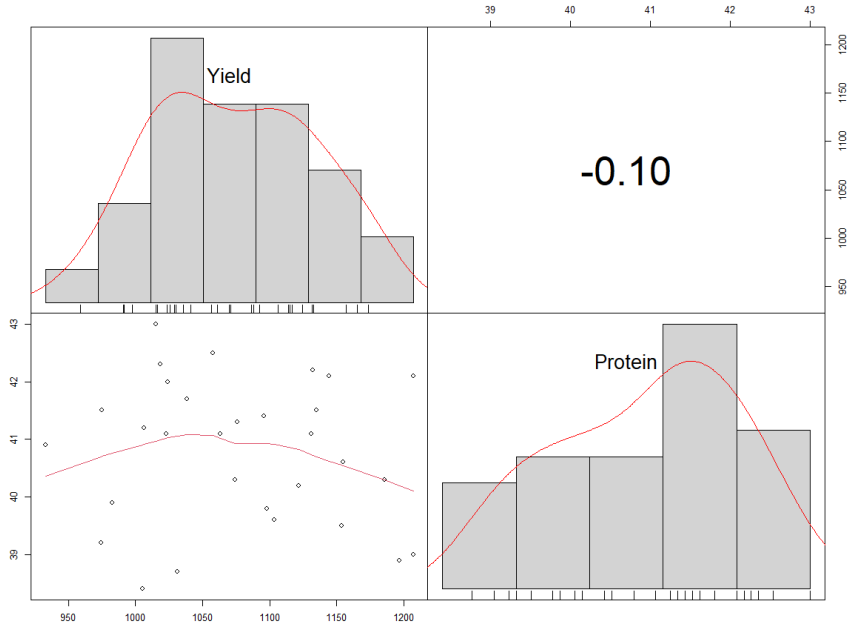


Figure 25: A Corrplot showing distribution and correlation between Protein content and yield

6.4 DISCUSSION

6.4.1 Protein performance and stability

The present study illustrated a narrow protein content range of 4.6% (min, max) between the soybean genotypes with most genotypes having protein content close to the mean value. This is probably because most of the evaluated genotypes are advanced breeding lines that are progenies of two major crosses (Nam 2 × GC and Duiker × GC). However given that the test genotypes were generated from only two major crosses, the range of 4.6% is still remarkable. Therefore, this highly significant difference among genotypes suggests that selection for protein content among the studied soybean genotypes is possible.

The study also showed that the first three soybean genotypes with the highest protein content were progenies of Nam 2 × GC 00138-29 cross. Nam 2 is a Ugandan variety, which was a selection from TGM 79; obtained from IITA. This is because Nam 2 used as a parent to generate the Nam 2 × GC crosses has a high protein content. This is evident by the high protein content in

the commercial Ugandan soybean varieties that are progenies of Nam 2 × GC 00138-29 crosses. For example, Namsoy 4M and Maksoy 6N have 43% and 41% protein content respectively (Tukamuhabwa *et al*., 2019; Tukamuhabwa and Obua, 2015).

The BSPS 48A genotypes are progenies from Duiker × GC 00138-29 crosses and most had low protein content below the average of 41.0%. This is because these genotypes are progenies of Duiker × GC crosses where the primary objective was rust resistance and the seed's beauty (Tukamuhabwa et al., 2019). Duiker originated from Zimbabwe and was used as a female parent during the generation of the crosses because of the beautiful seed colour. On the other hand, GC00138-29 is a variety of AVRDC in Taiwan and was used as a source of rust resistance. The local cheeks included in the experiment (Maksoy 3N and Maksoy 4N) had an average protein content of 41.1% and 40.9% respectively. This is because the varieties developed from a cross between Duiker × GC are good in several traits like high yields, beautiful seed, large-seeded but average in terms of protein content.

The study showed a significant effect of environment, genotype, and genotype × environment interaction on protein content. The significant effect of genotype × environment is indication that protein content in soybean is a quantitative trait as reported by previous studies (Arslanoglu *et al*., 2011; Kumar *et al*., 2006; Ojo *et al*., 2002; Piper and Boote 1999). This finding suggests that changes in environmental conditions resulting from the cultivation of soybean genotypes in different locations can influence protein content. Such observation agrees with previous studies that reported a significant effect of genotype × environmental interaction for protein content in multi-locational field trials (Kumar *et al*., 2006; Ojo *et al*., 2002b; Gurmu *et al*., 2010; Mudenda, 2016). Significant genotype × environmental interaction makes it difficult for breeders to identify the best genotypes during variety evaluations, selection and recommendation. The presence of interactions indicates that environmental conditions significantly contribute to the relative genotype performance during evaluation trials. This coupled with erratic climatic changes, there is a need for soybean breeders to make environment-specific recommendataion inorder to develop soybean varieties that are adapted to specific environments.

6.4.2 Clustering test environments in relation to Protein Content (%)

Nakabango had the highest protein content of 42.5%, followed by Abi ZARDI and Mubuku irrigation scheme. These three locations probably had higher protein content than the other locations because they had higher temperatures and low rainfall during the bean formation stage. Previous studies have shown that high temperatures and low rainfall during the bean formation stage significantly lead to increased protein content (Mudenda, 2016). For example, Ojo *et al*. (2002) reported that locations with higher temperatures had higher protein content than locations with low temperatures. On the other hand, Mudenda (2016) reported a significant negative correlation between soybean protein content and high rainfall.

Four mega-environments for protein content were observed in this study. This implies that different genotypes can be recommended for the different meg-environments. The biplots identified Iki Iki as the most discriminating location for protein content. This observation implies that Iki Iki is the best location for evaluation for protein content in soybean in Uganda. NaCRRI was identified as the most representative location for protein content in Uganda because of the narrow angle between their respective vectors and "average environment axis". This implies that NaCRRI can represent the other test environments used in this study for protein content (Yan, 2001; Yan *et al*. 2007).

6.4.3 Yield performance and stability

The present study showed that several genotypes performed better than Maksoy 3N and Maksoy 4N, which are the most farmer-preferred and high-yielding varieties in Uganda. This study also showed that BSPS 48A-9-2 derived through single plant selection from BSPS 48A had the highest yield level than all the other genotypes. BSPS 48A was released in Uganda as Maksoy 3N as the highest yielding genotype (Tukamuhabwa *et al*., 2012). However, through continuous single plant selection, this variety's yield has been greatly improved and stabilized. BSPS 48A-9-2 had greater yield stability, implying that its yield responds following the prevailing conditions. Such genotypes that display average stability tend to have higher yields when the prevailing conditions like moisture and soil fertility are favorable. Therefore, with increased input use such as fertilizer such genotypes tend to have high yields useful for both smallholder and commercial farmers.

On the other hand, Nam 2 × GC 44.2 was consistent in yield performance irrespective of the prevailing conditions because it displayed narrow adaption (Becker and Leon 1988; Lin *et al*., 1986). Thus genotype Nam 2 × GC 44.2 will have stable yields irrespective of the prevailing conditions. Such genotypes are recommended for low-input farming systems because their performance does not change with the prevailing environmental conditions (Lin *et al*., 1986).

6.4.4 Clustering test environments in relation to Yield (Kg Ha^{-1})

Bulindi had the highest mean seed yield compared to the other seven locations. This was probably because of the soil's high moisture content since this location receives much rainfall for most of the time in the season. This contradicts a previous study by Tukamuhabwa *et al*. (2012a) that reported that NaCRRI had the highest yields due to the high rainfall received through the different seasons. Yet another study by Tukamuhabwa *et al*. (2012b) showed that Nakabango was instead the highest yielding test environment across five locations due to high soil fertility and high amounts of rainfall received. According to Obua (2013), Mubuku was reported as the highest yielding environment compared to the other four test locations because of the available water in the soil through flood irrigation. These results suggest that soil moisture and soil fertility during the cropping season are the significant drivers of soybean seed yield in Uganda. The results further indicate that environmental conditions across Uganda vary significantly in space and time. This is a result of the change in climatic conditions of these locations. This eventually makes variety evaluation and recommendations more challenging. Therefore there is a need for soybean breeders to recommend varieties based on specific adaptability rather than wide adaptability (Silveira et al. 2018).

Three mega-environments were observed in this study. This is contrary to Tukamuhabwa *et al*. (2012b) observations, which showed that Uganda had two mega-environments for soybean seed yield when evaluating 24 soybean genotypes for three seasons in five locations. However, in the current study, evaluations were conducted in six seasons and, eight locations represented the diverse agro-ecological zones than reported by Tukamuhabwa *et al*. (2012b). The three mega-

environments observed in this study suggest that successful soybean breeding and selections must be made in at least each of the selected mega-environments.

Bulindi was the most discriminating test environment for soybean yield in Uganda in this study. This implies that Bulindi provides much information about the differences among the genotypes being evaluated, which was in agreement with a study conducted by Tukamuhabwa *et al*. (2012b) reported Bulindi as the most discriminating test environment. The test genotypes used in the current study are different from those reported by Tukamuhabwa *et al*. (2012b) yet Bulindi was the most discriminating environment. The high discriminative power of Bulindi makes it an excellent location to be used as a primary testing location for differentiating the soybean genotypes for yield, and can be used as a "culling environment" for quick elimination of unstable genotypes during the evaluation process (Yan and Kang 2003). On the other hand, Bulindi was also the most representative environment because it had the smallest angle between its vector and "average-environment axis". This implies that Bulindi can represent the other test environments used in this study for soybean yield.

6.4.5 Correlation between Protein Content (%) and Yield (Kg Ha^{-1})

The study showed a negative correlation between protein content and yield of the studied soybean genotypes. However, it is not clear whether the negative relationship between protein and yield is due to a decrease in protein concentration *per-se* or due to an increase in the composition of other seed components when increasing yield. The negative correlation between protein content and yield of the studied soybean genotypes agrees with previous studies. For example, Helms and Orf (1998) evaluated ten soybean populations in seven environments. They found out that selection among the studied populations for increased protein content led to a decrease in yield by 110 Kg Ha^{-1} while protein content increased by 5 g kg^{-1}. Cober and Voldeng (2000) evaluated single cross and rapid back cross-breeding methods to achieve protein content and high seed yield; the studied populations exhibited very low or no association between the two traits (r = -0.06 to -0.21). Another study by Filho *et al*. (2004) reported a negative and significant correlation between protein content and grain yield when eight populations were evaluated to assess the effect of selection for high protein content and yield.

The negative correlation between protein content and yield implies that selection for high protein content negatively affects the grain yield.

6.5 CONCLUSIONS AND RECOMMENDATIONS

Genotype Nam II × GC 7.2 had high and stable protein content, while its yield was low. The negative correlation between protein content and yield can be overcome by the introgression of high grain yield quantitative traits loci through an appropriate breeding method like recurrent selection or marker-assisted breeding. The breeding method should minimize the loss of high protein content alleles, given the negative correlation between the traits. Furthermore, the soybean breeding program should also explore more new sources of superior alleles for protein content and introgress in the existing soybean germplasm. This is because the studied genotypes in this current study were crosses generated mainly for high yields and soybean rust resistance and not high protein content. Lastly, genotypes BSPS 48A-9-2, BSPS 48A-31 and Nam 2 × GC 44.2 should be further tested under farmers' production conditions for selection and release as new soybean varieties in Uganda high and stable yields. BSPS 48A-9-2 is recommended for high-input farming systems because it had broad stability, while Nam 2 × GC 44.2 for low-input production systems had narrow stability. Therefore, genotypes with broader stability should be recommended for commercial farmers who have access to production inputs, while genotypes with narrow stability should be recommended for resource-poor farmers who have limited access to suitable production technologies. A highly discriminating environment such as Bulindi should be used as a primary location for evaluating or producing soybean genotypes.

CHAPTER SEVEN

7.0 GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

7.1 Genetic Diversity among soybean germplasm

The major objective of the current study was to assess the genetic diversity of tropical soybean germplasm. The results showed low genetic diversity among the tropical germplasm studied. This could be attributed to the genetic status of genotypes used, which were mainly released varieties and advanced breeding lines. The improved varieties tend to have low genetic diversity because of the high selection pressure subjected to the genotypes during evaluation and selection. For example, most released varieties in Uganda share few parents in their pedigree. The low diversity is probably because of the domestication process that eliminates undesirable traits by the farmers.

The observed low diversity in the germplasm pool makes it difficult for soybean breeders to develop new and improved soybean varieties with desirable characteristics, which include both farmer-preferred traits and breeder's preferred traits. Additionally, the low genetic diversity implies that the tropical soybean varieties could be completely wiped out in case of the emergence of a new pathogen or insect pest. This is a sensitive situation for soybeans and could be worsened by the erratic climatic changes that have become more frequent. This observation calls for interventions that will widen the genetic base of Ugandan soybean germplasm collection. In contrast, most studies that have reported high genetic diversity, have used wild relatives and landraces of soybean. Therefore, in the case of Uganda, the soybean breeders would have to seek for wild soybean genotypes to broaden the genetic base of the Ugandan soybean germplasm.

7.2 Total Protein, Total Oil and Fatty Acid Content

Protein and oil content have been reported as the major objective of breeding so far as seed component is concerned in most soybean breeding programs. Although improvement of nutritional traits has not been one of the core breeding objectives for soybean in Uganda, the study identified several soybean genotypes that had elevated content of nutritional traits above the reported/known averages. For example, this study identified several genotypes with high

nutritional traits of economic importance like total protein, total oil and oleic fatty acid. This is a good step towards the development of soybean varieties with elevated nutritional traits in Uganda and across the East African region. The genotypes with elevated nutritional traits identified in the current study could be evaluated in multi-locational yield trials and released as new soybean varieties in Uganda. Some of the Ugandan soybean genotypes that are already adapted to the Ugandan environment also had elevated protein content and oleic fatty acid. Such genotypes with elevated nutritional traits can be used by soybean breeders in Uganda and across the East African region to improve soybean genotypes with low nutritional traits but possessing other superior traits. Additionally, such soybean genotypes with elevated protein and oil content are very important for the large processing plants that are being established in Uganda and across the East African region; for extraction of oil and production of soybean cake for animal feeds. This is expected to greatly improve the livelihood of the farmers from the sale of the soybean grain to these processing plants that have a very large production capacity. Production of the soybean cake is expected to significantly improve the livestock industry in Uganda and across the East African region.

The study also revealed that soybean genotypes from Uganda had the lowest protein content. This is probably because less attention has been given to the improvement of nutritional traits in soybean in Uganda. However, the study revealed that soybean genotypes from Japan had the highest protein content while the USA had the highest oil content and the second-highest protein content. The high protein content among genotypes from Japan was because of a deliberate effort to develop soybean varieties with higher protein content by the Japanese government after the Second World War; to fight malnutrition. Similarly, soybean breeding programs in the USA is among the oldest where more breeding has been conducted to improve nutritional traits compared to the breeding programs in Africa (Uganda and Zimbabwe) where soybean reached much later. The findings of this study imply that soybean breeders in Uganda should put more attention to nutritional traits improvement; especially protein and oil content in the breeding program. Therefore for nutritional improvements of Ugandan soybean germplasm, soybean breeders should seek soybean germplasm from Japan and the USA for protein content and oil content respectively.

Furthermore, the study also revealed that soybean genotypes from Uganda had the highest oleic fatty acid. This is a good prospect for soybean breeding in Uganda. This implies that these genotypes can be used to improve the trait of oleic fatty acid among the released Ugandan soybean varieties. Additionally, these genotypes can also be evaluated in multi-locational trials for other important traits like yield, pests and disease resistance before releasing new varieties in Uganda. Similarly, other soybean breeders around the East African region can use such genotypes to improve their soybean genotypes for the trait of oleic fatty acid.

The distribution of protein and oil content followed a normal distribution curve for all the soybean genotypes that were evaluated, indicating that these two traits are quantitatively inherited and controlled by a complex genetic system that is governed by multiple genetic loci, each with minor effects. The G × E studies done in Uganda for protein content showed that there was a highly significant genotypic difference in protein content among the studied genotypes, suggesting that effective selection for protein content among the studied soybean genotypes is possible. The study also showed a significant effect of environment and genotype × environment interaction on protein content, suggesting that environment has a role to play when selecting for high protein content in soybean. The significant effect of the environment of these nutritional traits implies that it is quite difficult to study or predict these traits in different environments. This has been worsened by frequent erratic climatic changes experienced in the last decade. The effect of climate change has made general variety recommendation for each location more challenging. Therefore varieties should be recommended for adaptation in specific environments (location and seasons) rather than a wide adaptation that most plant breeders have previously used.

7.3 Correlations among nutritional traits and SNP Markers

Correlation analysis has been widely used in crop breeding programs to determine the nature of relationships between different traits that are of economic value to plant breeders. Oleic fatty acid is a monounsaturated fatty acid that can improve the oil quality and shelf-life of products processed using such oil. The significant negative correlation between oleic acid and total oil suggests that it would be difficult to develop soybean varieties possessing both traits. A similar

trend was observed between oleic acid and total protein where there was a negative correlation, suggesting that it is difficult to develop a soybean variety with high oleic acid and protein. The study suggests that it would be extremely difficult to develop a soybean variety that has high oleic and oil or protein content using conventional breeding. The study also showed a negative correlation between protein content and yield of the studied soybean genotypes. The negative correlation between protein content and yield implies that selection for high protein content has a negative effect on grain yield. To overcome the challenge of negative correlations for some of these very important traits in soybean, breeders could consider the development of soybean varieties to serve different purposes. Additionally, the use of advanced breeding approaches such as mutation breeding, gene editing, and genetic transformation could be explored by breeders to overcome the challenge of the negative correlations.

7.4 GWAS and SNP Markers

GWAS has been performed in nearly all economically important crops, including soybean for the selection and improvement of desirable traits. GWAS is a method for the study of associations between a genome-wide set of single-nucleotide polymorphisms (SNPs) and desired phenotypic traits. In the current study, using GWAS, five SNP markers were identified; two associated with oil content (rs22918920 and rs22918919) and three associated with protein content (rs22918920, rs22918919 and rs14974480). Two SNP markers rs22918920 and rs22918919 located on chromosomes 7 and 10 respectively were associated with both protein and oil content and were identified for the first time in this study which points to greater prospects of improving protein and oil content concurrently in the same soybean genotype. There is a need to develop and validate these markers on different sets of soybean populations.

7.5 Protein and Yield Stability

This is the first attempt to assess the effect of the environment on protein content in soybean seed in Uganda. The study showed that several genotypes performed better than Maksoy 3N and Maksoy 4N that are the most farmer-preferred and high-yielding varieties in Uganda, respectively. This is because high yield has been one of the traits that soybean breeders in Uganda have been selecting for in soybean. Additionally, the study identified a few genotypes that had both high yields and high protein content. The identification of these soybean genotypes

having higher yields than the highest yielding soybean variety in Uganda and above the average protein content of 40% is a big step towards the development of improved soybean varieties. There is a need to conduct on-farm evaluation trials and sensory testing with the processors to increase the chances of consumer acceptability of these genotypes with higher yields and protein content; before releasing them for production in Uganda.

A significant effect of the environment was observed for both protein content and yield in the study. Contradiction in performance for both protein content and yield in the G × E study over the years is a clear indication of climate change and its effect on soybean research. Therefore, there is a need for soybean breeders to identify specific environmental factors that affect protein and yield to identify the best genotype for each environment.

7.6 Recommendations

From the curret study, the following recommendations can be drawn;

The soybean breeders should seek for wild soybean genotypes to broaden the genetic base of the Ugandan soybean germplasm whose diversity was observed to be low.

Based on the nutrient profiles of the core collection, the study identified several genotypes with high nutritional traits that could be used by soybean breeders across the East African region for the improvement of nutritional traits in their respective breeding programs. This would greatly boost the soybean value chain that is characterized by the establishment of more plants in East Africa to process soybean grain.

The identified SNP markers that are associated with high total oil and total protein content can now be used by soybean breeders to track these traits in the soybean breeding pipeline. This would significantly hasten the soybean breeding circle in Uganda and the East Africa region. There is a need to validate the two SNPs, rs22918920 and rs22918919 that are associated with both protein and oil content and were identified for the first time in this study.

The soybean genotypes that had high protein content and yields should be subjected to on-farm trials under farmers' production conditions and sensory testing by the processors before variety release.

The G × E for protein content and oil content should be conducted for more seasons to assess the consistency of the performance.

There is a need to identify the specific environmental factors that affect the performance of nutritional traits and yield in soybean.

REFERENCES

Abdelghany, Ahmed M. et al. 2020. Natural Variation in Fatty Acid Composition of Diverse World Soybean Germplasms Grown in China. *Agronomy* 10(1): 1–18.

Adie, M Muchlish, Ayda Krisnawati, and Gatut Wahyu A.S. 2014. Assessment of Genotype × Environment Interactions for Black Soybean Yield Using Ammi and GGE Biplot. *International Journal of Agriculture Innovations and Research* 2(5): 2319–1473.

Agralytica. 2012. *United Soybean Board Connections 2012: Soybean Market Scan*. https://unitedsoybean.org/wp-content/uploads/Soybean-Market-scan-report-final.pdf.

Ahire, Digambar Dadaji. 2012. Correlation Studies and Coefficient of Variation among Fatty Acids and Oil Quality Parameters in Soybean Mutants. *Bioremediation, Biodiversity and Bioavailability* 6: 107–12.

Akond, Masum et al. 2014. Identification of Quantitative Trait Loci (QTL) Underlying Protein , Oil , and Identification of Quantitative Trait Loci (QTL) Underlying Protein , Oil , and Five Major Fatty Acids ' Contents in. (January).

Akpagu Francis, Nnamani Emmanuel V, and Eze Chukwuebuka G I. 2015. Estimation of Protein Content in Soyabeans Consumed In. 10(1): 8–12.

Al., Patcharin Tanya et. 2001. Evaluation of Genetic Diversity among Soybean Genotypes Using SSR and SNP. *Korean J. Crop Sci* (August 2014).

Alt, Jessie L. 2005. Phenotypic and Molecular Analysis of Oleate Content in the Mutant Soybean Line M23. *Crop Science* 45(5): 1997–2000.

Anai, Toyoaki et al. 2008. Two High-Oleic-Acid Soybean Mutants, M23 and KK21, Have Disrupted Microsomal Omega-6 Fatty Acid Desaturase, Encoded by GmFAD2-1a. *Breeding Science* 58(4): 447–52.

Anderson, Edwin J et al. 2019. *Soybean [Glycine Max (L .) Merr .] Breeding : History , Improvement , Production and Future Opportunities*.

Anderson, James W., Belinda M. Smith, and Carla S. Washnock. 1999. Cardiovascular and Renal Benefits of Dry Bean and Soybean Intake. *American Journal of Clinical Nutrition* 70(3 SUPPL.): 464–74.

Anthony, P., G. Malzer, S. Sparrow, and M. Zhang. 2012. Soybean Yield and Quality in Relation to Soil Properties. *Agronomy Journal* (January).

Arslanoglu, Funda, Selim Aytac, and Emel Karaca Oner. 2011. Effect of Genotype and Environment Interaction on Oil and Protein Content of Soybean (Glycine Max (L.) Merrill) Seed. *African Journal of Biotechnology* 10(80): 18409–17.

Assefa, Yared et al. 2019. Assessing Variation in US Soybean Seed Composition (Protein and Oil). 10(March).

Bachlava E and Andrea J. Cardinal. A.J. 2009. Correlation between Temperature and Oleic Acid Seed Content in Three Segregating Soybean Populations. *Crop Science* 49.

Bachlava, Eleni et al. 2008. Heritability of Oleic Acid Content in Soybean Seed Oil and Its Genetic Correlation with Fatty Acid and Agronomic Traits. *Crop Science* 48(5): 1764–72.

Baize, J. 1999. *Global Soybean Meal Analysis Project. Conducted for the Quality Committee of the United Soybean Board: St. Louis, MO.*

Baker, R J. 1988. Tests For Crossover Genotype-Environmental Interactions. *Can. J. Plant Sci.* 68: 405–10.

Bandillo, Nonoy, Diego Jarquin, Qijian Song, et al. 2015. A Population Structure and Genome-Wide Association Analysis on the USDA Soybean Germplasm Collection. *The Plant Genome* 8(3): plantgenome 2015.04.0024.

Bashaasha, Bernard. 1992. Soybean Research in Uganda. (June).

Bationo, a et al. 2006. African Soils: Their Productivity and Profitability of Fertilizer Use. *Background paper for the Africa Fertilizer Summit*: 9–13.

Becker H. C. and Leon J. 1988. Stability Analysis in Plant Breeding. *Plant Breeding* 23: 1–23.

Berghuis, Nila Tanyela et al. 2010. Electrocatalytic Hydrogenation of Soybean Oil Using Hydrogen Transfer Agent. In *Third International Conference on Mathematics and Natural Sciences*, , 848–58.

Bhandari, HR et al. 2017. Assessment of Genetic Diversity in Crop Plants - An Overview. *Advances in Plants & Agriculture Research* 7(3): 279–86.

Bisen, Anchal, Dhirendra Khare, Priya Nair, and Niraj Tripathi. 2014. SSR Analysis of 38 Genotypes of Soybean (Glycine Max (L.) Merr.) Genetic Diversity in India. *Physiology and Molecular Biology of Plants* 21(1): 109–15.

Bligh E.G. and Dyer W.J. 1959. A Lipid Method of Total Lipid Extraction and Purification. *Can. J. Biochem.* 37: 911–17.

Bolon, Yung-tsi et al. 2010. Complementary Genetic and Genomic Approaches Help Characterize the Linkage Group I Seed Protein QTL in Soybean.

Bowman, J C. 1972. Genotype X Environment Interactions. *Ann. Génét Sél anim* 4(1): 117–23.

Brim, C.A., Burton, J.W. 1979. Recurrent Selection in Soybeans II. Selection for Increased Percent Protein in Seeds. *Crop Sci.* 19: 494–498.

Burton, J.M., Brim, C.A. 1981. Recurrent Selection in Soybeans III. Selection for Increased Oil in Seeds. *Crop Sci.* 21: 31–34.

Burton, J. W., R. F. Wilson, and C. A. Brim. 1994. Registration of N79-2077-12 and N87-2122-4, Two Soybean Germplasm Lines with Reduced Palmitic Acid in Seed Oil. *Crop Science* 34(1): 313–313.

Burton, J. W., R. F. Wilson, and C.A. Brim. 1983. Recurrent Selection in Soybeans. IV. Selection for Increased Oleic Acid Percentage in Seed Oil 1'. *Crop Science* 23(4): 744–47.

Burton, J.W., R.F. Wilson, G.J. Rebetzke, and V.R. Pantalone. 2006. Registration of N98–4445A Mid-Oleic Soybean Germplasm Line. *Crop Science* 46(2): 1010–12.

Cao, Yongce et al. 2017. Identification of Major Quantitative Trait Loci for Seed Oil Content in Soybeans by Combining Linkage and Genome-Wide Association Mapping. *Front. Plant Sci.* 8(July): 1–10.

Cardinal, Andrea J., and Joseph W. Burton. 2007. Correlations between Palmitate Content and Agronomic Traits in Soybean Populations Segregating for the Fap1,Fapnc, and Fan Alleles. *Crop Science* 47(5): 1804–12.

Carrera, Constanza Soledad et al. 2011. Amino Acid Composition of Soybean Seeds as Affected by Climatic Variables. *Pesquisa Agropecuaria Brasileira* 46(12): 1579–87.

Carson, Luther C et al. 2011. 'Variety Trials'. 0353(February): 131–35.

Carver, B.F., Burton, J.W., Willson, R.F., Carter, Jr., J.E. 1986. Cumulative Response to Various Recurrent Selections Schemes in Soybean: Oil Quality and Correlated Agronomic Traits. *Crop Sci.* 26: 853–858.

Cattelan, Alexandre José, and Amélio Dall'Agnol. 2018. The Rapid Soybean Growth in Brazil. *OCL - Oilseeds and fats, Crops and Lipids* 25(1): 1–12.

Chander, Subhash et al. 2019. Prospects for Durable Resistance against an Old Soybean Enemy: A Four-Decade Journey from Rpp1 (Resistance to Phakopsora Pachyrhizi) to Rpp7.

Agronomy 9(7).

Chaudhary, Kaushal Raj, and Jixiang Wu. 2012. Stability Analysis for Yield and Seed Quality of Soybean [Glycine Max (L .) Merril] Across Different Environments in Eastern South Dakota.

Chauhan, Devendra K et al. 2015. Molecular Characterization and Genetic Diversity Assessment in Soybean [Glycine Max (L .) Merr .] Varieties Using SSR Markers. 14(October): 504–10.

Cheelo, P, D Lungu, and M Mwala. 2017. GGE Biplot Analysis for Identification of Ideal Soybean [Glycine Max L. Merrill] Test and Production Locations in Zambia. *Journal of Experimental Agriculture International* 15(3): 1–15.

Chen, Kuan I. et al. 2012. Soyfoods and Soybean Products: From Traditional Use to Modern Applications. *Applied Microbiology and Biotechnology* 96(1): 9–22.

Chen, Wu et al. 2017. Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of Ziziphus Jujuba Assessed with Genome-Wide SNPs Developed by Genotyping-by-Sequencing and SSR Markers. 8(April): 1–14.

Chung, J et al. 2003. The Seed Protein, Oil, and Yield QTL on Soybean Linkage Group I. *Crop Sci*.

Ciabotti, S. et al. 2016. Chemical Composition , Protein Profile , and Isoflavones Content in Soybean Genotypes with Different Seed Coat Colors. *International Food Research Journal* 23(2): 621–29.

Ciabotti, Sueli, José Marcos Gontijo Juhász, Ana Cristina Pinto Mandarino, and Elisa Norberto Ferreira Costa, Luciene Lacerda Corrêa, Angelita Duarte Simão, Anderson Assaid Santos. 2019. Chemical Composition and Lipoxygenase Activity of Soybean (Glycine Max L. Merrill.) Genotypes, Specific for Human Consumption, with Different Tegument Colours. *Brazilian Journal of Food Technology*: 1–10.

Clemente, Tom E., and Edgar B. Cahoon. 2009. Soybean Oil: Genetic Approaches for Modification of Functionality and Total Content. *Plant Physiology* 151(3): 1030–40.

Cober, E.R. and Voldeng, H.D. 2000. Developing High-Protein, High-Yield Soybean Populations and Lines. *Crop Science* 40(1): 39–42.

Cui, Z. L., Carter, T. E., Burton, J. W., and Wells, R. Phenotypic Diversity of Modern Chinese

and North American Soybean Cultivars. *Crop Sci.* 41: 1954–1967.

D.S.Falconer. 1989. Introduction to Quantitative Genetic-DS Falconer. 433.

Dias, Douglas Antônio, Leandra Regina, Texeira Polo, and Fabiane Lazzari. 2017. Genome-Wide Association for Mapping QTLs Linked to Protein and Oil Contents in Soybean. (1): 896–904.

Diers, B W, W R Fehr, and R C Shoemaker. 1992. RFLP Analysis of Soybean Seed Protein and Oil Content. *Theor Appl Genet*: 608–12.

Djanta, M. K.A et al. 2020. Utilization and Analysis of Its Adoption in Sub-Saharan Africa. 12(1): 1–12.

Doldi, M. L., J. Vollmann, and T. Lelley. 1997. Genetic Diversity in Soybean as Determined by RAPD and Microsatellite Analysis. *Plant Breeding* 116(4): 331–35.

Drago et al. 2011. Soybean and Nutrition *Use of Soybean in Cereal Based Food Formulation and Development of Nutritionally Improved Foods*.

Erickson, EA, JR Wilcox, and JF Cavins. 1988. Fatty Acid Composition of the Oil in Reciprocal Crosses among Soybean Mutants. *Crop Sci.* 28: 644–46.

Eskandari, Mehrzad, Elroy R. Cober, and Istvan Rajcan. 2013. Genetic Control of Soybean Seed Oil: I. QTL and Genes Associated with Seed Oil Concentration in RIL Populations Derived from Crossing Moderately High-Oil Parents. *Theoretical and Applied Genetics* 126(2): 483–95.

FAO. 2018. Online at (Accessed on 5th June, 2020). *http://www.fao.org/faostat*.

Fehr W.R, Welke G.A, Hammond E.G, Duvick D.N, Cianzio S.R. 1992. Inheritance of Reduced Linolenic Acid Content in Soybean Genotypes A16 and A17. *Crop Sci* 32: 903–6.

Fehr, Walter R, and Earl G Hammond. 1998. Soybean Vegetable Oil Having Elevated Concentrations of Both Palmitic Acid and Stearic Acid.

Filho, Odilon Mello, Carlos Sigueyuki Sediyama, and Newton Piovesan. 2004. Grain Yield and Seed Quality of Soybean Selected for High Protein Content Grain Yield and Seed Quality of Soybean Selected for High Protein Content. 39(August 2015): 445–50.

Flores, F., Moreno, M.T., Cubero, J.I. 1998. A Comparison of Univariate and Multivariate Methods to Analyze G×E Interaction. *Field Crops Research* 56(3): 271–286.

Foyer Christine H et al. 2019. Modelling Predicts That Soybean is Poised to Dominate Crop

Production across Africa. *Plant Cell Environ* 42(1): 373–85.

Friedman and Brandon. 2001. Nutritional and Health Benefits of Soy Proteins. *Journal of Agricultural and Food Chemistry* 49(3): 1069–86.

Goldflus F, Ceccantini M, and Santos M. 2006. Amino Acid Content of Soybean Samples Collected in Different Brazilian States – Harvest 2003 / 2004.

Goldsmith, Peter D. 2008. Soybean Production and Processing in Brazil. *Soybeans: Chemistry, Production, Processing, and Utilization* (October): 773–98.

Graham, Peter H, and Carroll P Vance. 2003. Update on Legume Utilization Legumes : Importance and Constraints to Greater Use. 131(March): 872–77.

Grieshop, Christine M., and George C. Fahey. 2001. Comparison of Quality Characteristics of Soybeans from Brazil, China, and the United States. *Journal of Agricultural and Food Chemistry* 49(5): 2669–73.

Gupta SK, Manjaya JG. 2017. Genetic Diversity and Population Structure of Indian Soybean (Glycine Max (L.) Merr.) as Revealed by Microsatellite Markers. *Physiology and Molecular Biology of Plants* 25(4): 953–64.

Gurmu, F, H Mohammed, and G Alemaw. 2010. Genotype X Environment Interactions and Stability of Soybean for Grain Yield and Nutrition Quality. *African Crop Science Journal* 17(2).

Guzeler, N, and C Yldrm. 2016. The Utilization and Processing of Soybean and Soybean Products. *Ziraat Fakultesi Dergisi, Uludag Universitesi* 30(November): 546–53.

Gwinner, R. et al. 2017. Genetic Diversity in Brazilian Soybean Germplasm. *Crop Breeding and Applied Biotechnology* (17): 373–81.

Halewood, Michael et al. 2020. Germplasm Acquisition and Distribution by Cgiar Genebanks. *Plants* 9(10): 1–31.

Han YP, Zhao X, Liu DY, Li YH, Lightfoot DA, Yang ZJ, Zhao L, Zhou G, Wang, and Huang L ZK. 2016. Domestication Footprints Anchor Genomic Regions of Agronomic Importance in Soybeans. *New Phytologist* Han YP, Zh(209): 871–884.

Hao D, Cheng H, Yin ZT, Cui SY, Zhang D, Wang H, Yu D. 2012. Identification of Single Nucleotide Polymorphisms and Haplotypes Associated with Yield and Yield Components in Soybean (Glycine Max) Landraces across Multiple Environments. *Theor Appl Genet*

124(447–458).

Hayakawa, Kyoko, Yu Yen Linko, and Pekka Linko. 2000. The Role of Trans Fatty Acids in Human Nutrition. *Eur. J. Lipid Sci. Technol.* 52(6–7): 229–35.

Hayes, Maria. 2020. Measuring Protein Content in Food: An Overview of Methods. *Foods* 9(10).

Helms, T.C., and J.H. Orf. 1998. Protein, Oil, and Yield of Soybean Lines Selected for Increased Protein. *Crop Sci.* 38: 707–711.

Hipparagi, Yegappa, Rakesh Singh, Debjani Roy Choudhury, and Veena Gupta. 2017. Genetic Diversity and Population Structure Analysis of Kala Bhat (Glycine Max (L.) Merrill) Genotypes Using SSR Markers. *Hereditas* 154: 9.

Hwang, Eun-young et al. 2014. A Genome-Wide Association Study of Seed Protein and Oil Content in Soybean. *BMC Genomics*: 1–12.

Hwang, Eun Young et al. 2014. A Genome-Wide Association Study of Seed Protein and Oil Content in Soybean. *BMC Genomics* 15(1).

Hymowitz, T., Kaizuma, N. 1981. Soybean Seed Protein Electrophoresis Profiles from 15 Asian Countries or Regions: Hypotheses on Paths of Dissemination of Soybeans from China. *Econ Bot* 25: 10–23.

Hyten, D. L. et al. 2004. Seed Quality QTL in a Prominent Soybean Population. *Theoretical and Applied Genetics* 109(3): 552–61.

Hyten, David L et al. 2004. Molecular Mapping and Identification of Soybean Fatty Acid Modifier Quantitative Trait Loci. 81(12): 1115–18.

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker, and Cregan PB. RC. 2006. Impacts of Genetic Bottlenecks on Soybean Genome Diversity. In *Proceedings of the National Academy of Sciences of the United States of America, 26 Oct 2006*, , 16666–71.

Imafidon, Frank W. Sosulski and Gilbert I. 1990. Amino Acid Composition and Nitrogen-to-Protein Conversion Factors for Animal and Plant Foods. *J. Agric. Food Chem.* 38(6): 1351–1356.

Jia, Hongmin et al. 2019. Quantification of Gly m 5.0101 in Soybean and Soy Products by Liquid Chromatography-Tandem Mass Spectrometry. *Molecules* 24(1): 42–53.

Karr-Lilienthal, Lisa K. et al. 2004. Chemical Composition and Protein Quality Comparisons of

Soybeans and Soybean Meals from Five Leading Soybean-Producing Countries. *Journal of Agricultural and Food Chemistry* 52(20): 6193–99.

Katsunari Ippoushi, Motoe Sasanuma, Hideaki Oike, Masuko Kobori, Mari Maeda-Yamamoto. 2015. Absolute Quantification of Protein NP24 in Tomato Fruit by Liquid Chromatography/Tandem Mass Spectrometry Using Stable Isotope-Labelled Tryptic Peptide Standard. *Food Chemistry* 173: 238–42.

Kawuki, RS, E Adipala, J Lamo, and P Tukamuhabwa. 2003. Responding to the Soybean Rust Epidemic in Sub-Saharan Africa: A Review. *African Crop Science Journal* 11(4).

Keim, P., Shoemaker, R.C. & Palmer, R.G. 1989. Restriction Fragment Length Polymorphism Diversity in Soybean. *Theoret. Appl. Genetics* 77: 786–792.

Khalid, Muhammad et al. 2010. Linkage of Morphological Markers in Brassica. *Pakistan Journal of Botany* 42(5): 2995–3000.

Kilian A. , Wenzl P., Huttner E., Carling J., Xia L., Blois H., Vanessa Caig, Katarzyna Heller-Uszynska, Damian Jaccoud, Colleen Hopper, Malgorzata Aschenbrenner-Kilian, Margaret Evers, Kaiman Peng, Cyril Cayla, Puthick Hok, Grzegorz Uszynski. 2012. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. *In: Pompanon F., Bonin A. (eds) Data Production and Analysis in Population Genomics. Methods in Molecular Biology (Methods and Protocols), vol 888. Humana Press, Totowa, NJ*.

Kim, Changsoo et al. 2016. Application of Genotyping by Sequencing Technology to a Variety of Crop Breeding Programs. *Plant Science* 242: 14–22. http://dx.doi.org/10.1016/j.plantsci.2015.04.016.

Kofsky, Janice, Hengyou Zhang, and Bao Hua Song. 2018. The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (Glycine Soja). *Frontiers in Plant Science* 9(July).

Korte, Arthur, and Ashley Farlow. 2013. The Advantages and Limitations of Trait Analysis with GWAS : A Review. *Plant Methods* 9(1): 1. Plant Methods.

Krager, Jenna. 2020. Assessing the Quantification of Soy Protein in Incurred Matrices Using Targeted LC-MS / MS.

Kumar V, Rani A, Solanki S, S.M.Hussain. 2006. Influence of Growing Environment on the Biochemical Composition and Physical Characteristics of Soybean Seed. *Journal of Food*

Composition and Analysis 19: 188–95.

Kumawat, Giriraj et al. 2015. Molecular Characterization and Genetic Diversity Analysis of Soybean (Glycine Max (L.) Merr.) Germplasm Accessions in India. *Physiology and Molecular Biology of Plants* 21(1): 101–7.

La, Thang et al. 2019. Characterization of Select Wild Soybean Accessions in the USDA Germplasm Collection for Seed Composition and Agronomic Traits. *Crop Science* 59(1): 233–51.

La, Thang C. et al. 2014. Effect of High-Oleic Acid Soybean on Seed Oil, Protein Concentration, and Yield. *Crop Science* 54(5): 2054–62.

Lance, G. and Garren, B. 2005. Origin, History, and Uses of Soybean (Glycine Max).

Leamy, Larry J. et al. 2017. A Genome-Wide Association Study of Seed Composition Traits in Wild Soybean (Glycine Soja). *BMC Genomics* 18(1): 1–15. http://dx.doi.org/10.1186/s12864-016-3397-4.

Lee, I. H. and Chou, C. C. 2006. Distribution Profiles of Isoflavon Isomers in Black Bean Kojis Prepared with Various Filamentous Fungi. *Journal of Agricultural Food Chemistry* 54: 1309–14.

Lee, Jeong-Dong, Kristin D Bilyeu, and James Grover Shannon. 2007. Genetics and Breeding for Modified Fatty Acid Profile in Soybean Seed Oil. *Journal of Crop Science and Biotechnology* 10(4): 201–10.

Li, Z. L., Nelson, R. L. 2001. Genetic Diversity among Soybean Accessions from Three Countries Measured by RAPDs. *Crop Sci.* 41: 1337– 1347.

Li, Bin et al. 2017. High-Resolution Mapping of QTL for Fatty Acid Composition in Soybean Using Specific-Locus Amplified Fragment Sequencing. *Theoretical and Applied Genetics* 130(7): 1467–79.

Li, Dongmei et al. 2019. Genomics Genome-Wide Association Mapping for Seed Protein and Oil Contents Using a Large Panel of Soybean Accessions. *Genomics* (January): 1–6. http://dx.doi.org/10.1016/j.ygeno.2018.01.004.

Li, Ying-hui et al. 2010. Genetic Diversity in Domesticated Soybean (Glycine Max) and its Wild Progenitor (Glycine Soja) for Simple Sequence Repeat and Single-Nucleotide Polymorphism Loci: 242–53.

Li, Ying hui et al. 2018. Genome-Wide Association Mapping of QTL Underlying Seed Oil and Protein Contents of a Diverse Panel of Soybean Accessions. *Plant Science* 266(May): 95–101. https://doi.org/10.1016/j.plantsci.2017.04.013.

Li, Ying Hui et al. 2010. Genetic Diversity in Domesticated Soybean (Glycine Max) and Its Wild Progenitor (Glycine Soja) for Simple Sequence Repeat and Single-Nucleotide Polymorphism Loci. *New Phytologist* 188(1): 242–53.

Lin, C. S., Binns, M. R., Lefkovitch, L. P. 1986. Stability Analysis: Where Do We Stand?. *Crop Science* 26(5): 894–900.

Lithourgidis, A. S., C. A. Dordas, C. A. Damalas, and D. N. Vlachostergios. 2011. Annual Intercrops: An Alternative Pathway for Sustainable Agriculture. *Australian Journal of Crop Science* 5(4): 396–410.

Liu. 1997. Chemistry and Nutritional Value of Soybean Components. In: Soybeans. *Springer, Boston, MA.*: 25–113.

Liu, Xiaoyi et al. 2020. Genome-Wide Association Study Identifies Candidate Genes Related to Oleic Acid Content in Soybean Seeds. *BMC Plant Biology* 20(1): 1–14.

Liu, Zhangxiong et al. 2017. Comparison of Genetic Diversity between Chinese and American Soybean (Glycine Max (L.)) Accessions Revealed by High-Density SNPs. *Frontiers in Plant Science* 8(November).

Lonnerdal B. 1994. Nutritional Aspects of Soy Formulas. *Acta Pediatr Suppl* 402: 105–108.

Lowry, O. H.;, N. J.; Rosebrough, A. L.; Farr, and R. J. Randall. 1951. Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry.* 193(1): 265–275.

Lu, Weiguo et al. 2013. Identification of the Quantitative Trait Loci (QTL) Underlying Water Soluble Protein Content in Soybean. *Theoretical and Applied Genetics* 126(2): 425–33.

Lubungu, Mary, William Burke, and Nicholas J Sitko. 2013. Challenges of Smallholder Soybean Production and Commercialization in Eastern Province of Zambia. (November): 1.

Madhujith, Terrence, and Subajiny Sivakanthan. 2018. Oxidative Stability of Edible Plant Oils. (February).

Maehre, Hanne K. et al. 2018. Protein Determination—Method Matters. *Foods* 7(1).

Maestri, Damián M., Diana O. Labuckas, Carlos A. Guzmán, and Laura M. Giorda. 1998. Correlation between Seed Size, Protein and Oil Contents, and Fatty Acid Composition in

Soybean Genotypes. *Grasas y Aceites* 49(5–6): 450–53.

Maldonado dos Santos, João Vitor et al. 2016. Evaluation of Genetic Variation among Brazilian Soybean Cultivars through Genome Resequencing. *BMC Genomics* 17(1): 1–18. http://dx.doi.org/10.1186/s12864-016-2431-x.

Malek, M A et al. 2014. Morphological Characterization and Assessment of Genetic Variability , Character Association , and Divergence in Soybean Mutants. *The scientific world Journal* 2014(article ID 968796): 12.

Mammadov, Jafar, Rajat Aggarwal, Ramesh Buyyarapu, and Siva Kumpatla. 2012. SNP Markers and Their Impact on Plant Breeding. 2012.

Mao, Tingting et al. 2013. Identification of Quantitative Trait Loci Underlying Seed Protein and Oil Contents of Soybean across Multi-Genetic Backgrounds and Environments. *Plant Breeding* 132(6): 630–41.

Maphosa Mcebisi. 2013. Enhancing Genetic Resistance to Soybean Rust Disease. PhD Thesis. Makerere University Kampala.

Marega, Mario et al. 2001. Relationships among Oil Content, Protein Content and Seed Size in Soybeans. *Brazilian Archives of Biology and Technology* 44(1): 23–32.

McKendry, A.L., McVetty, P.B.E. and Voldeng, H.D. 1985. Inheritance of Seed Protein and Seed Oil Content in Early Maturing Soybean. *Can. J. Genet. Citol.* 27: 603–7.

McKendry, Anne L., Peter B. E. McVetty, and Harvey D. Voldeng. 1985. Inheritance of Seed Protein and Seed Oil Content in Early Maturing Soybean. *Canadian Journal of Genetics and Cytology* 27(5): 603–7.

Medina-juárez, Luis Ángel, and Nohemí Gámez-meza. 2011. Effect of Refining Process and Use of Natural Antioxidants on Soybean Oil.

Mentreddy, S R, A I Mohamed, N Joshee, and A K Yadav. 2002. Edamame : A Nutritious Vegetable Crop. : 432–38.

Mercer, L. C., J. C. Wynne, and C. T. Young. 1990. Inheritance of Fatty Acid Content in Peanut Oil. *Peanut Science* 17(1): 17–21.

Miladinovic, J., Hrustic, M., Vidic, M. 2011. Soybean. Institute of Field and Vegetable Crops, Novi Sad and Sojaprotein, Becej, AMB Graphics, Novi Sad, 510.

Mudenda Hampango. 2016. Stability Analysis For Oil and Protein Content in Soybean (Glycine

Max. L. Merrill) Seed Across Varying Environments. 4(1): 64–75.

Mukuze, Clever. 2019. Genetic Diversity And Genotype × Environment Interaction Of Advanced Elite Soybean Genotypes in Uganda. MSc. Thesis. Makerere University Kampala.

Mulato, Bruno Mello et al. 2010. Genetic Diversity in Soybean Germplasm Identified by SSR and EST-SSR Markers. *Pesquisa Agropecuaria Brasileira* 45(3): 276–83.

Mulugeta, A, S Kidane, S Abadi, and Z Fisha. 2013. GGE Biplots to Analyze Soybean Multi-Environment Yield Trial Data in North Western Ethiopia. *Journal of Plant Breeding and Crop Science* 5(12): 245–54.

Namara, M. et al. 2019. Potential for Yield Loss Reduction and Profitability Assessment of Pesticide Control of Groundnut Leaf Miner among Soybean Genotypes. African Crop Science Journal 27(2): 183.

Nichols, D. M., Glover, K. D., Carlson, S. R., Specht, J. E., & Diers, B. W. 2006. Fine Mapping of a Seed Protein QTL on Soybean Linkage Group I and Its Correlated Effects on Agronomic Traits. *Crop Science,* 46(2): 834–39. https://doi.org/10.2135/cropsci2005.05-0168.

Ningsih F, Zubaidah S, Kuswantoro H. 2019. Diverse Morphological Characteristics of Soybean (Glycine Max L . Merill) Pods and Seeds Germplasm. *IOP Conf. Series: Earth and Environmental Science* 276.

Obua T. 2013. Soybean Rust Diversity and Adaptation of Elite Soybean Lines to the Ugandan Environment. MSc. THesis. Makerere University Kampala.

Obua T. , M Nabasirye , M Namara , G Tusiime , M Maphosa & P Tukamuhabwa (2020) Yield stability of tropical soybean genotypes in selected agro- ecologies in Uganda, South African Journal of Plant and Soil, 37:2, 168-173, DOI: 10.1080/02571862.2019.1678687

Obua, Tonny, Julius Pyton Sserumaga, Stephen O. Opiyo, et al. 2020. Genetic Diversity and Population Structure Analysis of Tropical Soybean (Glycine Max (L.) Merrill) Using Single Nucleotide Polymorphic Markers. *Global Journal of Science Frontier Research; Vol 20, No 6-D (2020) The Global Journal of Science Frontier Research*.

Obua, Tonny, Julius Pyton Sserumaga, Fredrick Nganga, et al. 2020. Nutrient Profiling of Tropical Soybean (Glycine Max) Core Collection. *Global Journal of Science Frontier*

Research; Vol 20, No 7-D (2020) The Global Journal of Science Frontier Research.

Ojo, D. K., Ajayi, A. O., Oduwaye, O. A. 2012. Genetic Relationships among Soybean Accessions Based on Morphological and RAPDs Techniques. *Pertanika J. Trop. Agric. Sci.* 35(2): 237–48.

Ojo, D. K., Ajayi, A. O. and Oduwaye, O. A. 2012. Genetic Relationships among Soybean Accessions Based on Morphological and RAPDs Techniques. *Pertanika J. Trop. Agric. Sci* 35(2): 237–48.

Ojo DK, Adebisi MA, Tijani BO. 2002. Influence of Environment on Protein and Oil Contents of Soybeans Seed (Glycine Max (L.) Merril). *Global J. Agric. Sci.* 1: 27–32.

P. Tukamuhabwa and M. Maphosa. 2011. State of Knowledge on Breeding for Durable Resistance to Soybean Rust Disease in the Developing World.

Pandey, R.K. 1988. *A Farmer's Primer on Growing Soybean on Riceland*. Agriculture, Ecosystems & Environment.

Pandurangan, Sudhakar et al. 2012. Relationship between Asparagine Metabolism and Protein Concentration in Soybean Seed. *Journal of Experimental Botany* 63(8): 3173–84.

Panthee, D. R. et al. 2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. 45(5): 2015–22.

Panthee, D.R., V.R. Pantalone, and A.M. Saxton. 2006. Modifier QTL for Fatty Acid Composition in Soybean Oil. (May): 67–73.

Pathan, Safiullah M. et al. 2013. Genetic Mapping and Confirmation of Quantitative Trait Loci for Seed Protein and Oil Contents and Seed Weight in Soybean. *Crop Science* 53(3): 765–74.

Patil, Gunvant et al. 2017. Molecular Mapping and Genomics of Soybean Seed Protein : A Review and Perspective for the Future. *Theoretical and Applied Genetics* 130(10): 1975–91.

Patil, G., Vuong, Tri D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., Wu, X., Bai, Y., Yungbluth, D., Lu, F., Kumpatla, S., Shannon, J. G., Varshney, Rajeev K., Nguyen, Henry T. 2018. Dissecting Genomic Hotspots Underlying Seed Protein, Oil, and Sucrose Content in an Interspecific Mapping Population of Soybean Using High-Density Linkage Mapping. *Plant Biotechnology Journal* 16(11): 1939–53.

Payne RW, Harding SA, Murray DA, Soutar DM, Baird DB, Glaser AI, Channing IC, Welham SJ, Gilmour AR, Thompson R, Webster R. 2010. The Guide to GenStat Release 13, Part 2: Statistics. *Hemel Hempstead UK: VSN International.*

Perić, Vesna et al. 2014. Genetic Relatedness of Soybean Genotypes Based on Agromorphological Traits and Rapd Markers. *Genetika* 46(3): 839–54.

Pham, AT. 2011. Modification of Fatty Acid Composition in Soybean Seeds To Improve Soybean Oil Quality and Functionality. University of Missouri.

Pham, Toan Duc. 2011. Analyses of Genetic Diversity and Desirable Traits in Sesame (Sesamum Indicum L., Pedaliaceae): Implication for Breeding and Conservation.

Pintauro, P N et al. 2005. Electrochemical Hydrogenation of Soybean Oil with Hydrogen Gas. *Ind. Eng. Chem. Res.*: 6188–95.

Piper, Ernest L., and Kenneth I. Boote. 1999. Temperature and Cultivar Effects on Soybean Seed Oil and Protein Concentrations. *Journal of the American Oil Chemists' Society* 76(10): 1233–41.

Probst, A. H., and Jude, R.W. 1973. Origin, US. History Development and World Distribuation. Incaldwell (Ed.) Pp 1-15. Soybean: Improvement, Production and Uses. *American society of Agronomy*: 1–15.

Qin, Nan et al. 2010. Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection Articles Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection. *Nature Publishing Group* 42(12): 1053–59. http://dx.doi.org/10.1038/ng.715.

Qin, Peiyou et al. 2014. Regional Distribution of Protein and Oil Compositions of Soybean Cultivars in China. *Crop Science* 54(3): 1139–46.

Qiu L., Nelson R. L., Vodkin L. O. 1997. Evaluation of Soybean Germplasm with Random Amplification Polymorphic DNA (RAPD) Markers. *Sci. Agric. Sin.* 23: 408–417.

Rahman, T. Anai, T. Kinoshita, and Y. Takagi. 2003. A Novel Soybean Germplasm with Elevated Saturated Fatty Acids. *Crop Science.*

Rahman Mizanur, Shaikh et al. 1994. High Oleic Acid Mutant in Soybean Induced by X-Ray Irradiation. *Bioscience, Biotechnology, and Biochemistry* 58(6): 1070–72.

Rani et al. 2007. Interrelationship between Oil Content and Fatty Acid Composition in Indian

Soybean (Glycine Max) Cultivars. *The Indian Journal of Agricultural Sciences*: 59–62.

Rebetzke, G. J. et al. 2001. Genetic Background and Environment Influence Palmitate Content of Soybean Seed Oil. *Crop Science* 41(6): 1731–36.

Ren, Jing et al. 2013. Genetic Diversity Revealed by Single Nucleotide Polymorphism Markers in a Worldwide Germplasm Collection of Durum Wheat. (February): 7061–88.

Riaz M.N. 2001. Uses and Benefits of Soy Fiber. *Cereal Foods World* 46(3): 98–100.

Ridley William, and Stephen Devadoss. 2015. Argentina's Agricultural Policies William. *Journal of Chemical Information and Modeling* 53(9): 1689–99.

Rotundo, L, and Mark E Westgate. 2009. Meta-Analysis of Environmental Effects on Soybean Seed Composition Field Crops Research Meta-Analysis of Environmental Effects on Soybean Seed Composition. (January 2009).

Roychowdhury, Rajib et al. 2014. Molecular Marker-Assisted Technologies for Crop Improvement: 241–58.

Rubaihayo, P. R.: Leakey, C. L.A. 1970. Introduction, Testing and Selection of Soya Beans in Uganda. *E. Afr. agric. For. J.* 36: 77–82.

Saito, M. 1972. Breeding Of Soybean in Japan.

Sanginga, N. and Woomer, P.L. 2009. Tropical Soil Biology and Fertility Institute of the International Centre for Tropical Agriculture. Nairobi. *Integrated Soil Fertility Management in Africa: Principles Practices and Development Process.*

Sanginga, P C et al. 1999. Social Impact of Soybean in Nigeria's Southern Guinea Savanna. *Tropical Agriculture*: 34.

Satyavathi, C. Tara et al. 2006. AFLP Analysis of Genetic Diversity in Indian Soybean [Glycine Max (L.) Merr.] Varieties. *Genetic Resources and Crop Evolution* 53(5): 1069–79.

Scientifica, VELP. 2013. Application Note N / N/Protein Determination in Cured Meats According to the Dumas Combustion Method. : 2–4.

Sebolt, A. M., Shoemaker, R. C., & Diers, B. W. 2000. Analysis of a Quantitative Trait Locus Allele from Wild Soybean That Increases Seed Protein Concentration in Soybean. *Crop Science* 40(5): 1438–44.

Shibata, Masayuki et al. 2008. Genetic Relationship between Lipid Content and Linolenic Acid Concentration in Soybean Seeds. *Breeding Science* 58(4): 361–66.

da Silva Gesteira, Abelmon et al. 2003. Biometrical Analyses of Linolenic Acid Content of Soybean Seeds. *Genetics and Molecular Biology* 26(1): 65–68.

Silveira, Daniel Augusto et al. 2018. Adaptability and Stability of Grain Yield in Soybean. *Australian Journal of Crop Science* 12(5): 717–25.

Sindelar, Aaron J. et al. 2015. Long-Term Corn and Soybean Response to Crop Rotation and Tillage. *Agronomy Journal* 107(6): 2241–52.

Singh, Devinder, M E Rezac, and P H Pfromm. 2009. Partial Hydrogenation of Soybean Oil with Minimal Trans Fat Production Using a Pt-Decorated Polymeric Membrane Reactor. 86: 93–101.

Singh, Nivedita et al. 2013. Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. *PLoS ONE* 8(12).

Singh, Preeti, R. Kumar, S. N. Sabapathy, and A. S. Bawa. 2008. Functional and Edible Uses of Soy Protein Products. *Comprehensive Reviews in Food Science and Food Safety* 7(1): 14–28.

SNV. 2011. Increased Competitiveness of the Oilseed Value Chain through Improved Information on the Markets for Soybean in Uganda. Final Report Submitted to SNV Rwenzori Portfolio, Kampala, Uganda.

Sobko, Olena et al. 2020. Environmental Effects on Soybean (Glycine Max (L.) Merr) Production in Central and South Germany. *Agronomy* 10(12): 1847.

Sommers, Darren, and Leah Mchale. 2015. Protein Yield Stability of Soybean Breeding Lines.College of Food, Agricultural and Environemnetal Sciences, The Ohio State University.

Sonah, Humira et al. 2015. Identification of Loci Governing Eight Agronomic Traits Using a GBS-GWAS Approach and Validation by QTL Mapping in Soya Bean. *Plant Biotechnology Journal* 13(2): 211–21.

Specht, J E et al. 2001. Soybean Response to Water : A QTL Analysis of Drought Tolerance. Cell Biology & Molecular Genetics.

Ssengendo, M., Mayende, J., Mubiru, R. 2010. Soybean Fact Sheet: Export Potential of the Soybean Sub-Sector in Uganda. A Survey Carried out by Participants of the COMESA Training Done by International Trade Center (ITC) to Assess the Export Potential of

Soybean Industry. Uganda Export Promotion Boar.

Stein, H. H. et al. 2008. Nutritional Properties and Feeding Values of Soybeans and Their Coproducts. Soybeans: Chemistry, Production, Processing, and Utilization: 613–60.

Stoltzfus, W.R. Fehr, G.A Welke, and D.L. 2000. Relationship of Elevated Palmitate to Soybean Seed Traits. *Crop Science*: 52–54.

Sudaric, A. et al. 2008. Application of Molecular Markers in Parental Selection in Soybean. *Acta Agronomica Hungarica* 56(4): 393–98.

Sultan, Sheikh Mohammad, Nilamani Dikshit, and Umesh J. Vaidya. 2015. Oil Content and Fatty Acid Composition of Soybean (Glysine Max L.) Genotypes Evaluated under Rainfed Conditions of Kashmir Himalayas in India. *Journal of Applied and Natural Science* 7(2): 910–15.

Tajuddin, Teuku, Satoshi Watanabe, Naoki Yamanaka, and Kyuya Harada. 2003. Analysis of Quantitative Trait Loci for Protein and Lipid Contents in Soybean Seeds Using Recombinant Inbred Lines. *Breeding Science* 53(2): 133–40.

Taliercio, Earl, David Eickholt, Rakin Rouf, and Thomas Carter. 2017. Changes in Gene Expression between a Soybean F1 Hybrid and Its Parents Are Associated with Agronomically Valuable Traits. *PLoS ONE* 12(5): 1–18.

Tantasawat, P. et al. 2011a. SSR Analysis of Soybean (Glycine Max (L.) Merr.) Genetic Relationship and Variety Identification in Thailand. *Australian Journal of Crop Science* 5(3): 283–90.

Tantasawat, P et al. 2011b. SSR Analysis of Soybean (Glycine Max (L .) Merr .) Genetic Relationship and Variety Identification in Thailand. 5(3): 283–90.

Tefera, Hailu. 2011. Breeding for Promiscuous Soybeans at IITA. *Soybean - Molecular Aspects of Breeding* (April 2011).

Torres, Adalgisa Ribeiro et al. 2015. Genetic Structure and Diversity of a Soybean Germplasm Considering Biological Nitrogen Fixation and Protein Content. *Scientia Agricola* 72(1): 47–52.

Tukamuhabwa, P., Asiimwe, M., Nabasirye, M., Kabayi, P., And Maphosa, M. 2012. Genotype By Environment Interaction of Advanced Generation. *African Crop Science Journal*, 20(2): 107–15.

Tukamuhabwa P, Obaa B, Obua T, Namara M, Okii D, Kabayi P. 2016. Status of Soybean Production and Impact Indicators of New Soybean Varieties in Uganda. Makerere University Agricultural Research Institute, Kabanyolo (MUARIK), Makerere University, Kampala.

Tukamuhabwa P., Obua T., Namara N. Okii D., Kabayi P., Iga Y. 2019. Soybean Research and Development in Uganda: Highlights 2002-2018. Makerere University, Kampala, Uganda.

Tukamuhabwa P., Obua T. 2015. Production Guide In Uganda. Makerere University, Kampala, Uganda.

Tukamuhabwa P. 2001. Soybean (Glycine Max L.). In Mukiibi, J. B. (Ed.). Agriculture in Uganda. Volume 2, Crops. National Agricultural Research Organization,. In , 133–41.

Tukamuhabwa P and Oloka HK. 2016. Soybean Research and Development in Uganda. A Case of Paradigm Shift in an African University. Makerere University, Kampala, Uganda.

Tukamuhabwa, P, H K Oloka, T Sengooba, and P Kabayi. 2012. Yield Stability of Rust-Resistant Soybean Lines at Four Mid-Altitude Tropical Locations. 183: 1–10.

UBOS. 2021. Https://Www.Ubos.Org. Accessed on 24th August, 2021.

Varvel, G. E., and W. W. Wilhelm. 2003. Soybean Nitrogen Contribution to Corn and Sorghum in Western Corn Belt Rotations. *Agronomy Journal* 95(5): 1220–25.

Vaughn, Justin N. et al. 2014. The Genetic Architecture of Seed Composition in Soybean Is Refined by Genome-Wide Association Scans across Multiple Populations. *G3: Genes, Genomes, Genetics* 4(11): 2283–94.

VELP Scientifica. 2015. N/Protein Determination in Soybean According to the Kjeldahl Method. *VELP Scientifica* 87: 1–4. https://www.velp.com/public/file/14VELPApplicationNoteSoybeansFF-K-001-20152-215467-215470-215475.pdf.

Vignal, A., Milan, D., Cristobal S. M., and Eggen A. 2002. A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics. 34: 275–305.

Villela, O.T et al. 2014. Genetic Divergence of Roundup Ready (RR) Soybean Cultivars Estimated by Phenotypic Characteristics and Molecular Markers. *African Journal of Biotechnology* 13(26): 2613–25. http://academicjournals.org/journal/AJB/article-abstract/4461A2545606.

Vollmann, Johann, Christina N Fritz, Helmut Wagentristl, and Peter Ruckenbauer. 2000. Environmental and Genetic Variation of Soybean Seed Protein Content under Central European Growing Conditions. 1306(February): 1300–1306.

W., Yan. 2001. GGE Biplot—A Windows Application for Graphical Analysis of Multi-Environment Trial Data and Other Types of Two-Way Data. *Agronomy Journal* 93: 1111–1118.

Wang, L.; Wang, J. 1992. Soybean Genetics and Breeding'. In *Beijing, China: Science Press.*

Wang, L. X., Guan, R. X., Liu, Z. X., Chang, R. Z., Qiu, L. J. 2006. Genetic Diversity of Chinese Cultivated Soybean Revealed by SSR Markers. *Crop Sci.* 46: 032–1038.

Wang, Tong et al. 2019. Study on the Electrochemical Hydrogenation of Soybean Oil under H2 Conditions. *Journal of Oleo Science* 68(4): 311–20.

Wang, Xianzhi et al. 2014. Identification and Validation of Quantitative Trait Loci for Seed Yield , Oil and Protein Contents in Two Recombinant Inbred Line Populations of Soybean. (May).

Wang, Yue et al. 2020. Identification of Additive and Epistatic Effects QTLs for Seed Oil Content in Soybean Based on an Integrating Map of Two RIL Populations. : 1–23.

Wiggins, Benjamin T. 2012. Heritability and Genetic Gain of Seed Protein, Oil, and Yield among RIL of Soybean. http://trace.tennessee.edu/utk_gradthes/1248.

Wilcox, J.R. 1989. 'Soybean Protein and Oil Quality. In: World Soybean Research Conference IV, Proceedings. Associatión Argentina de La Soja, Buenos Aires'. : 28-39.

Wilcox, J R, and J F Cavins. 1985. Inheritance of Low Linolenic Acid Content of the Seed Oil of a Mutant in Glycine Max. *Thear Appl Genet* 71: 74–78.

William, S. and Akiko, A. 2009. History of Soybeans and Soyfoods in Africa (1857 - 2009): Extensively Annotated Bibliography and Sourcebook.

Wilson, R.F. 2004. *Seed Composition. In: Boerma, H. and Specht, J.E., Ed., Soybeans: Improvement, Production, and Uses, 3rd Edition, ASA, CSSA, and SSSA, Madison*.

Winther B, Moi P, Nordlund MS, Lunder N, Paus E, Reubsaet JL. 2009. Absolute ProGRP Quantification in a Clinical Relevant Concentration Range Using LC-MS/MS and a Comprehensive Internal Standard. *J Chromatogr B Analyt Technol Biomed Life Sci.* 877(13): 1359–65.

WWF. 2014. WWF Report *The Growth of Soy Impacts and Solutions*. http://issuu.com/wwfsoyreport/docs/wwf_soy_report_final_jan_19/1?e=10667775/6569194 .

Xu, Hongjun. 1990. Recurrent Selection for Maturity, Seed Yield and Percent Seed Protein in Soybeans (Glycine Max (L.) Merr.); Using Genetic Male Sterility. http://dx.doi.org/10.1016/j.jaci.2012.05.050.

Yan W., And, and Kang MS. 2003. YGGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists. *CRC Press, Boca Raton, FL*.

Yan, Weikai et al. 2007. GGE Biplot vs. AMMI Analysis of Genotype-by-Environment Data. *Crop Science* 47(2): 643–55.

Yao, Yanjie et al. 2020. Quantitative Trait Loci Analysis of Seed Oil Content and Composition of Wild and Cultivated Soybean. *BMC Plant Biology* 20(1): 1–13.

Yin, Xinhua, and Tony J Vyn. 2014. Relationships of Isoflavone , Oil , and Protein in Seed with Yield of Soybean. (September 2005): 3–4.

Zhang, Gu wen et al. 2013. Determination of the Genetic Diversity of Vegetable Soybean [Glycine Max (L.) Merr.] Using EST-SSR Markers. *Journal of Zhejiang University. Science. B* 14(4): 279–88.

Zhang, Ling et al. 2014. Changes in Oleic Acid Content of Transgenic Soybeans by Antisense RNA Mediated Posttranscriptional Gene Silencing. *International Journal of Genomics* 2014.

Zhao, Shancen et al. 2015. Impacts of Nucleotide Fixation during Soybean Domestication and Improvement. *BMC Plant Biology* 15(1): 1–12.

Zhaoming, Q., Xiaoying, Z., Huidong, Q. et al. 2017. Identification and Validation of Major QTLs and Epistatic Interactions for Seed Oil Content in Soybeans under Multiple Environments Based on a High-Density Map. *Euphytica* 213.

Zhe, Yan, Joseph G. Lauer, Roger Borges, and Natalia de Leon. 2010. Effects of Genotype × Environment Interaction on Agronomic Traits in Soybean. *Crop Science* 50(2): 696–702.

Zhou, Zhengkui et al. 2015. Resource Resequencing 302 Wild and Cultivated Accessions Identifies Genes Related to Domestication and Improvement in Soybean. *nature biotechnology* 33(4).

Appendix 1: Soybean varieties that have been released in Uganda and their Pedigree

Variety Name	Year of Release	Pedigree
Bukalasa 4	1967	-
S 38	1968	-
Congo 72	1969	-
Kabanyolo 1	1971	-
Nam 1	1991	ICAL 131 (USA)
Nam 2	1994	TGM 79 (IITA)
Namsoy 3	2000	Kabanyolo 1 X Nam 1
Namsoy 4M	2004	Nam 2 X GC 00138-29
Maksoy 1N	2004	TGX 1035-10E
Maksoy 2N	2008	Duiker X TGX 1835-10E
Maksoy 3N	2010	Duiker X GC 00138-29
Maksoy 4N	2013	Duiker X GC 00138-29
Maksoy 5N	2013	Nam 2 X GC 00138-29
Maksoy 6N	2017	Nam 2 X GC 00138-29

Source: VIN for NARO released varieties

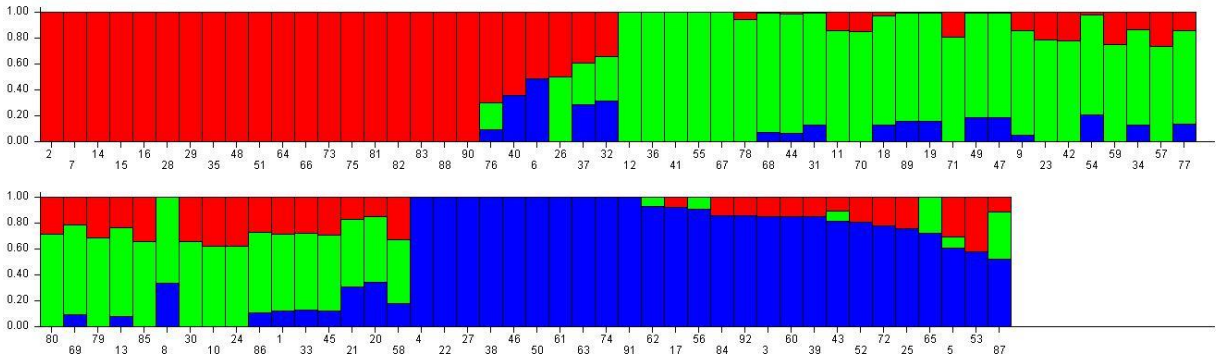
Appendix 2: The Tropical soybean genotypes used in understanding genetic diversity and population structure analysis

Genotype Name	Origin
Sline 3.17	Japan
AVRDC G10427	AVRDC
BSPS 48A-02A	Uganda
Namsoy 4M	Uganda
Santa	SeedCo
MNG 12.4	Uganda
AVRDC G7956	AVRDC
Maksoy 1N	Uganda
AVRDC GC00138-29	AVRDC
Sline 6.22	Japan
Maksoy 3N	Uganda
Duiker	SeedCo
G16	USA
Sline 3.16-1	Japan
AVRDC GC84051-31-1	AVRDC
Roan	SeedCo
Sline 7.11	Japan

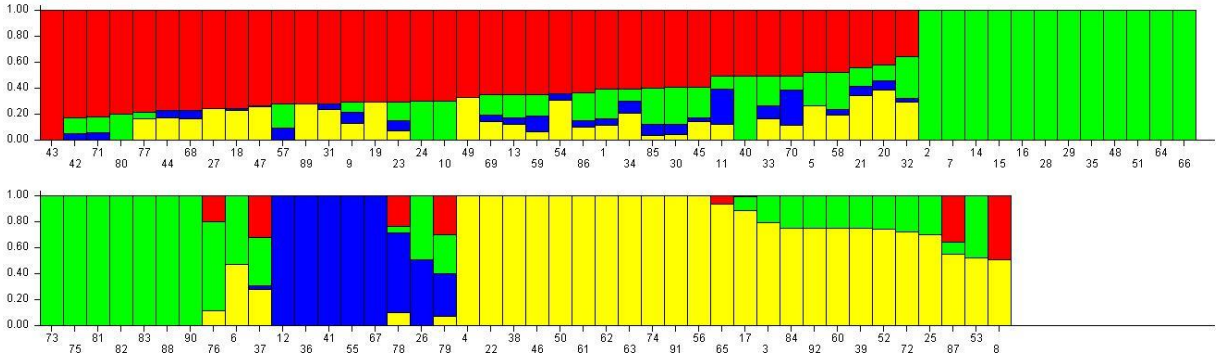
AGS 292	AVRDC
Siesta	SeedCo
Nam II	Uganda
Sline 1.19	Japan
Maksoy 5N	Uganda
Saxon	SeedCo
Soprano	SeedCo
G104270	AVRDC
Namsoy 3	Uganda
UG 5	Uganda
Status	SeedCo
AGS 338	AVRDC
Gazelle	SeedCo
AVRDC G28908	AVRDC
AVRDC SRE-B-11-13	AVRDC
Signal	SeedCo
Maksoy 2N	Uganda
Squire	SeedCo
Kab 1	Uganda
NG 14.1-21	Uganda
Sequel	SeedCo
Sline 9.2	Japan
G42	USA
AVRDC G4890-21-13-13	AVRDC
AVRDC G2843B	AVRDC
G7955	USA
K-Local	Uganda
Maksoy 4N	Uganda
Sline 4.21	Japan
NGDT 1.13-4	Uganda
Sentinel	SeedCo
NGDT 4.11-3	Uganda
Sline 13.2A	Japan
Sline 3.7	Japan
AGS 329	AVRDC
NG 14.1-16	Uganda
Sline 16.2	Japan
G32B	USA
Sline 3.16-2	Japan
Saga	SeedCo
G39	USA
Kuntz	USA

NII X GC 11.3	Uganda
NGDT 2.15-7	Uganda
Sline 5.18	Japan
Sline 12.7	Japan
NGDT 3.14-1	Uganda
NII X GC 30B	Uganda
BSPS 48A-8	Uganda
NII X GC 17.3	Uganda
BSPS 48A-27-1	Uganda
NII X GC 13.2	Uganda
NII X GC 7.2	Uganda
NII X GC 4.8	Uganda
NGDT 8.11-11B	Uganda
NGDT 8.11-4	Uganda
NGDT 4.11-4	Uganda
NII X GC 43.1	Uganda
NII X GC 28.2B	Uganda
BSPS 48A-5	Uganda
BSPS 48A-26	Uganda
BSPS 48A-28	Uganda
NII X GC 35.3	Uganda
NGDT 8.11-19	Uganda
NII X GC 20.3	Uganda
BSPS 48A-25	Uganda
MNG11.2	Uganda
NII X GC 32.6	Uganda
NII X GC 43.2	Uganda
NII X GC 11.2	Uganda
BSPS 48A-3B	Uganda
NII X GC 44.2	Uganda

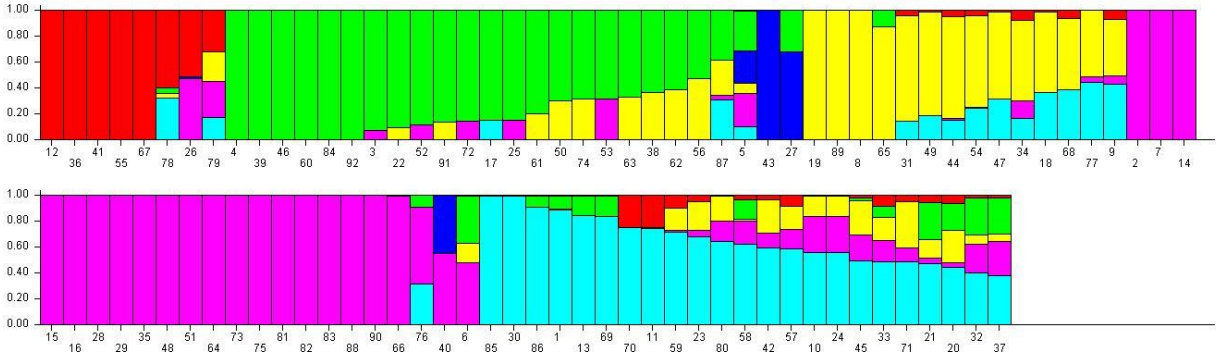
Appendix 3: Population structure inferred from STRUCTURE analysis for 89 soybean genotypes showing different K values



K = 3



K = 4



K = 6

Appendix 4: The Tropical soybean core collection used in nutrient profiling

Genotype	Origin
AGS 338	AVRDC
AVRDC G10427	AVRDC
AVRDC G2843B	AVRDC
AVRDC G7956	AVRDC
AVRDC GC00138-29	AVRDC
AVRDC GC84051-31-1	AVRDC
AVRDC SRE-B-11-13	AVRDC
BSPS 48A-25	Uganda
BSPS 48A-27-1	Uganda
BSPS 48A-3B	Uganda
BSPS 48A-5	Uganda
BSPS 48A-8	Uganda
Maksoy 2N	Uganda
Maksoy 4N	Uganda
Maksoy 5N	Uganda
MNG 12.4	Uganda
Nam II	Uganda
Namsoy 3	Uganda
NG 14.1-16	Uganda
NGDT 4.11-4	Uganda
NGDT 8.11-4	Uganda
NII X GC 11.2	Uganda
NII X GC 17.3	Uganda
NII X GC 20.3	Uganda
NII X GC 28.2B	Uganda
NII X GC 30B	Uganda
NII X GC 32.6	Uganda
NII X GC 43.1	Uganda
NII X GC 43.2	Uganda
NII X GC 44.2	Uganda
NII X GC 7.2	Uganda
K-Local	Uganda
Roan	SEEDCO
Saga	SEEDCO
Sequel	SEEDCO
Siesta	SEEDCO
Signal	SEEDCO
Gazelle	SEEDCO
Squire	SEEDCO

G104270	USA
G32B	USA
G42	USA
G45	USA
G7955	USA
Kuntz	USA
Sline 13.2A	Japan
Sline 16.2	Japan
Sline 4.21	Japan
Sline 5.18	Japan
Sline 6.22	Japan
Sline 7.11	Japan