GENETIC CHARACTERIZATION AND ESTIMATION OF GENOTYPE BY ENVIRONMENT INTERACTION OF ETHIOPIAN SESAME (*SESAMUM INDICUM L.*) GERMPLASM

JULY 2021

ADDIS ABABA, ETHIOPIA

GENETIC CHARACTERIZATION AND ESTIMATION OF GENOTYPE BY ENVIRONMENT INTERACTION OF ETHIOPIAN SESAME (*SESAMUM INDICUM L.*) GERMPLASM

A THESIS SUBMITTED

TO
DEPARTMENT OF MICROBIAL, CELLULAR AND MOLECULAR BIOLOGY
COLLEGE OF NATURAL SCIENCES

ADDIS ABABA UNIVERSITY

BY
TEWODROS TESFAYE NEGASH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PhD) (APPLIED GENETICS)

UNIVERSITY SUPERVISOR: DR. KASSAHUN TESFAYE
CO-SUPERVISORS: DR. GEMECHU KENENI AND
DR. TESFAHUN ALEMU

JULY 2021

ADDIS ABABA, ETHIOPIA

Dedication

This thesis is dedicated to the everlasting love and memories of my hero and beloved mother, Yeshalem Yemataw, who, to me, never died but simply departed.

Declaration

I, the undersigned, declare that the thesis hereby submitted for the Degree of Doctor of Philosophy (Ph.D.) in Applied Genetics to the Department of Microbial, Cellular, and Molecular Biology, Addis Ababa University is my own work and has not previously been submitted at another university. The materials obtained from other sources have been duly acknowledged in the thesis.

Signed on _____ July 2021, Department of Microbial, Cellular, and Molecular Biology, College of Natural Sciences, Addis Ababa University, Arat Kilo.

PhD candidate Signature: _____

Tewodros Tesfaye

University Supervisor Signature: _____

Kessahun Tesfaye (Ph.D.)

Acknowledgments

First of all, I would like to thank my major supervisor, Dr. Kassahun Tesfaye, and co-supervisors Dr. Gemechu Keneni, and Dr. Kifle Dagn for their unreserved support starting from conceiving and designing the study to the final write-up of the thesis. Thanks for your support and encouragement through the ups and downs of my time that made me reap the seeds of my hard work. Thank you for putting your trust in me and all your exemplary deeds that made me feel lucky for having you as my supervisor. Unfortunately, on process of my research, we lost our model instructor and my co-supervisor Dr. Kifle Dagn, and I wish peace in rest. My heartfelt gratitude would go to Dr. Tesfahun Alemu for their invaluable assistance and support during data analysis, write-up of the thesis, and editing of the manuscript for publication. This research work also could have never been realized without the initiative and consistent support of the whole community of the Gondar agricultural research center. My special gratitude extends to the former center directors Dr. Abirham Abiyu, and Tsedalu Jenberu, and present director Asfaw Azanaw for their indispensable effort and providing me basic facilitation during field and laboratory works. My special thanks also go to Daniel Admasu, Antenh Adebabay, Fentahun Biset, Getachew Asmare, Tesfaye Gorgi, Asnakew Takele, Asaye Birhanu, Mekides Abebe,Girmachew Wubie, Angoch Amsal, Robel Ayelign, Kibra, for their support in the fieldwork and experimentation directly or indirectly.

In addition, this research work also could have never been realized without the initiative and consistent support of the community of the Benefit sesame business network (Benefit SBN). My special gratitude extends to the Benefit-SBN coordinator Ted Schrader,

Benefit-SBN manager Dr. Geremew Terefe, Assistance manager Anteneh Mekuria for their indispensable effort and financial support at field and laboratory work starting from my work up to full write-up of the thesis. I am grateful to the staff of ILRI BecA, Nairobi, IGSS members for providing me access and materials during Genotyping and sequencing laboratory process and training. A special gratitude I give to Dr. Catharine Ziyomo, Mercy Chepngetich, and Jackline Chepkoech for their support and facilitation during the molecular laboratory work.

I am also extending my thanks to Humera, shire, worer, Assosa, Pawe, Areka, Bako Agricultural research centers, and Haromay University for conducting my research work in their research trial sites and provide the germplasm that available in their hands. A special gratitude I give to Fiseha Baraki (National Sesame research coordinator), Yirga Belay, Daniel Endale, Tesfahun Mekonit, Yasin , Mohamed, Dr. Chemeda , Amare kebede, Dr. Abush (national low land oil research coordinator) for their close follow up of my research trial from planting up to harvesting. My special thanks also go Ethiopian Biodiversity Institute (EBI) for providing me the germplasm. My special gratitude extends to Dr. Asnake Fikre, Dr. Firew Mekibib, Dr. Mulugeta Atinafu, Dr. Baye Berihun, Dr. Admas Alemu, Dr. Yihyis Erezene, Dr. Yonas Moges, Dr. Selamawit ketema, Dr. Adane Leta, Dr. Getu Abebe, Dr. Muhamed Abate and Mr. Atikilt Abera for their technical support.

This study was financed by the Benefit SBN project (SBN), Amhara Agricultural Research Institute (ARARI), and the Department of Microbial, Cellular and Molecular Biology (MCMB) of Addis Ababa University (AAU). I take this opportunity to thank all institutions.

My sincere thanks also go to Dr. Biru Yitaferu, Former Director General (DG) of ARARI, and Dr. Tilay Tekleweld Director General (DG) of ARARI for their facilitation and coordination of the research work. Dr. Dereje Beyene, Dr. Gurja Belay, Dr. Adey, Kidist Teferedegn, and Ayele Gucho of AAU are highly acknowledged for their facilitation in logistics and enrollment during each semester at the University. I thank all my instructors Prof. Endashaw Bekele, and Dr. Tileye Feyissa, Dr. Dereje Beyene, Dr. Adey for your irreplaceable mentor and advice.

Any special words of gratitude and thanks cannot express my inherent indebtedness, I owe to my mother, Yeshalem Yemataw, my father, Tesfaye Negash, and my aunts Asegedech Yemataw, and Amaledech Yemataw, for their love, prayers, caring, and sacrifices for educating and preparing me for my future. Unfortunately, my father and mother were not fortunate enough to live a little more to see their ambitions being realized with the eventual success of their son. I wish to dedicate this thesis to their immortal love and memories which always reside deep into my heart. My brothers and sisters, Eden Endalikachew, Bethlehem Mesfin, Dr. Baye Berihun, Shiferahu Andargi, Fasil Adane, Hewan Baye, Ananya Baye, Amisayaw Tareqegn, and my friend Tewodros Alene have all played their shares in boosting the morale that propelled me to series of successes in my education. Finally, I would like to express my gratitude with affection to my wife Ms. Mulugonder Nigru for her cheer, love, and patience to carry all burden of caring for our two children alone. I owe my indebtedness and darling to my daughters Yanet Tewodros, and Elidana Tewodros are always the backbone of my courage and reasons for my hard work. Above all, I thank my Almighty God and St. Mary for making it all possible.

Table of Contents

List of Tables

List of Figures

List of Appendices

Acronyms

AEC	Average Tester Coordinator
AEC	Average Environment Coordination
AFLP	Amplified Fragment Length Polymorphism
AMMI	Additive Main Effects & Multiplicative Interactions
AEA	Average Environment Axis
AMOVA	Analysis of Molecular Variance
ANOVA	Analysis of Variance
ASV	The AMMI's Stability Value
BecA	Biosciences East and Central Africa
BLUP	Best linear unbiased prediction
BS	Broad Sense
CSA	Central Statistical Agency
DArT	Diversity Array technology
EBI	Ethiopian Biodiversity Institute
ECV	Environmental Coefficient of variability
EIAR	Ethiopian Institute of Agricultural Research
ERCA	Ethiopian Revenue and Customs Authority
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organization of United Nations
FAOSTAT	FAO Statistics
GAM	Genetic Advance as Percent of Mean
GBS	Genotype By Sequencing

GCV	Genetic Coefficient of Variation
GD	Genetic Diversity
G x E	Genotype by Environment Interaction
GGE	Genotype plus Genotype by Environment Interaction
GWAS	Genome Wide Association Study
Ha	Hectare
He	Expected Heterozygosity
ILRI	International Livestock Research Institute
InDels	Insertions and Deletions
IPGRI	International Plant Genetic Resources Institute
IPCA	Interaction Principal Component Analysis Axis
ISSR	Inter-simple Sequence Repeat
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MAS	Marker Assisted Selection
MCMC	Markov Chain Monte Carlo
MoARD	Ministry of Agriculture and Rural Development
NBPGR	National Bureau of Plant Genetics Resources
NGS	Next Generation Sequencing
PARC	Pawe Agricultural Research Center
PAR	Photosynthetically Active Radiation
PCA	Principal Component Analysis
PCV	Phenotypic Coefficient of Variation

PCR	Polymerase Chain Reaction
PIC	Polymorphic Information Content
QTL	Quantitative Trait Loci
RAD	Restriction-site Associated DNA
RAPD	Random Amplified Polymorphic DNA
RCBD	Randomized Complete Block Design
RE	Restriction Enzymes
RTA	Real-time Analysis
SAS	Statistical Analysis System
SLAF-seq	The specific-locus Amplified Fragment Sequencing
SNP	Single Nucleotide Polymorphism
SRAP	Sequence-related Amplified Polymorphisms
SSR	Simple Sequence Repeats/Microsatellites

Abstract

Sesame is one of the major oil crops that has great economic importance for the country. In Ethiopia, sesame is among the foremost important oil crops both in terms of area coverage and total national annual production. However the crop suffers from low productivity due to biotic and abiotic stresses. Therefore, the present study conducted in different sets to generate information that can be used to design the future breeding program of sesame in the country. The first set of experiments was to study the morphological and molecular genetic diversity of the sesame germplasm collected from Ethiopia and other countries (Asian and other African countries). The same genotypes planted at three locations for phenotyping and genotyping using the two high throughput diversity array technology (DArT) markers (silicoDArT and SNP). Further to understand the impact of different putative genes Genome-wide association study of yield-related traits using 2997 SNPs in two environments was performed. The second set of experiments was conducted in 19 environments to assess the performance and stability of sesame varieties, and to characterize sesame growing environments in Ethiopia. Based on morphological characterization, genotypes showed wide variability for most morphological traits, except for plant growth type, leaf glands, anther filament color, and anther connective tip gland. High heritability combined with high genetic advance was recorded for plant height, primary branch, days to flower initiation, days to 50% flowering, pod bearing zone, seed yield per plant, and bacterial blight reaction indicating the potential of improving the population through a direct selection for these traits. Grain yield showed a significant and positive genotypic correlation with plant height, the number of capsules per plant, and pod bearing zone, the magnitudes of the positive genetic correlation suggest that the selection by those characters produces a significant increase in grain yield. Genetic divergence using Mahalanobis D^2 statistics was computed, and the genotype lines were grouped into six different clusters. Clustering was not associated with the geographical distribution; instead, genotypes were grouped mainly based on morphological differences. The maximum inter-cluster distance was observed between clusters IV and VI (D^2 =342.56, followed by clusters I, and VI (D^2 =217.9783). Maximum genetic recombination and variation in the subsequent generation are expected from crosses that involve parents from the clusters characterized by

maximum distances. The genetic diversity analysis showed that the average nucleotide diversity of the panel was 0.14. Considering the genotypes based on their geographical origin, Africa collections (0.21) as a whole without Ethiopian collection was more diverse than Asia and when further portioned Africa, North Africa (0.23) collection was more diverse than others, but at the continent level, Asia (0.17) was more diverse than Africa (0.14). The genetic distance among the sesame populations ranged from 0.015 to 0.394, with an average of 0.165. The structure analysis divided the panel into four hypothetical ancestral populations and 21 genotypes were clustered as an admixture. Under Genome-wide association study (GWAS) a total of 21 significant SNPs with 7 yield-related traits in two environments were identified and, these explaining the phenotypic variation ranged from 7.02 (DF) to 16.11% (CAPL), with an average of 9.76%, suggesting a moderate contribution to the traits. All significant loci found in LG 2, 6, and 11 associated with capsule length except one associated with the physiological period. The significant loci found in LG 3, 7, and 8 associated with a physiological period (Days to flower initiation, Days to 50% flowering, and Days to physiological maturity) except one associated with capsule length. Dissecting genetic control of flowering time and maturity is importance to foster sesame breeding and to develop new varieties able to adapt to changing climatic conditions. Indeed, flowering time and maturity strongly affect yield and plant adaptation ability. Since several favorable alleles detected in this study have not yet been intensively selected, our GWAS results will assist in incorporating further useful alleles into the elite sesame germplasm for a seed yield increase in the future. Based on the genotype x environment interaction study the test locations were divided into six groups. Humera, Banat, and Tach Armacho were highly discriminating and representative in the first, the second, and the third group respectively, and were identified as a core test site in that group. While Alemaya, Worer, and Mender67 were identified as the only test site in groups four, five, and six. The core testing sites identified would be used to facilitate the identification of superior sesame varieties and to reduce testing costs in the country. Environment Tach Armacho in 2017/18 and 2018/19 were close to the ideal environment. The GGE biplot analysis identified genotype G2 (setit-1) as the "ideal" genotype and among the highest mean seed yield. Setit-1 considered the most stable across variable environments.

Key Words: AMMI, DArTseq, Genetic diversity, GGE-biplot, GWAS, ideal genotype, Population structure, significant loci

Chapter 1. General Introduction

1.1. Background and Justification

Sesame (*Sesamum indicum L.,* 2n = 26) is a member of the Pedaliaceae family, and one of the ancient oil crops that are grown widely in tropical and subtropical areas (Bedigian and Harlan, 1986; Ashri, 1998). Archeological findings revealed that the cultivated sesame traces back its progenitor to the wild populations native to South Asia (Fuller, 2003). The sesame has been cultivated in South Asia since the time of the Harappan civilization from where it was spread west to Mesopotamia before 2000 B.C. (Fuller, 2003). Others believed that the crop was first cultivated in Africa and later taken to India (Simmonds and Purseglove, 1969; Alegbejo *et al.*, 2003). Still, others proposed that sesame was the main oil crop grown by the Indus Valley Civilization from where it was likely transferred to Mesopotamia around 2500 B.C (Tunde-Akintude, 2012).

Sesame is produced in different parts of the world for various purposes but more than 96% of the world's sesame seed production is covered by Africa and Asia. India, China, Burma (Myanmar), Sudan, Nigeria, The United Republic of Tanzania, Ethiopia, and Uganda being the key contributors (FAO, 2017). Sesame seeds are good sources of fat, protein, carbohydrates, fiber, and essential minerals. Seeds are chemically composed of 44–57% oil, 18–25% protein, and 13–14% carbohydrates (Borchani *et al.*, 2010). Sesame also referred to as "queen of oilseeds", is employed in sweets such as sesame bars and dessert, and bakery products or milled to get high-grade edible oil (Bedigian, 2004).

In Ethiopia, sesame is among the foremost important oil crops both in terms of area coverage and total national annual production (CSA, 2019). However, the farm level productivity of sesame in Ethiopia is very low (0.7 t ha^{-1}) (CSA, 2019) compared to the

genetic potential of the crop yield of 2 t ha^{-1} (Mkamilo, and Bedigian, 2007) and other countries like Egypt (1.357 t ha^{-1}), Nigeria (0.818 t ha^{-1}), Tanzania (0.723 t ha^{-1}), and China (1.618 t ha^{-1}) (FAOSTAT, 2019). Improved varieties released in Ethiopia are reported to yields ranging from 0.3 to 1.3 t ha^{-1} under rainfed and 1 to 2.4 t ha^{-1} under irrigation on research stations (Gebremichael, 2017).

The Low sesame productivity in Ethiopia has been attributed to lack of varietal replacement, low yielding varieties, significant yield loss during threshing, indeterminate growth, uneven ripening of capsules and, lack of improved varieties tolerant to biotic and, abiotic stresses (Ashri, 2007; Lakhanpaul *et al*., 2012). The low seed yield of sesame is a consequence of a lack of breeding attention (Ashri, 2007). Sesame production is also limited by lack of uniform maturity of capsules, and seed shattering (Langham and Wiemers, 2002).

Sesame production is also highly affected by the genotype x environment interaction. Breeders in any crop evaluate their genotypes or elite lines across locations to select the best genotypes for wide and specific adaptation. Thinking of it different researchers conducted multi environment-trials in sesame and have found out that sesame grain yield and other agronomic and adaptive traits are responsive to different growing conditions, implying the presence of genotype by environment interactions (Badu-Apraku, 2003; Moghaddam *et al*., 2009). Abate, (2015) also reported the effect of genotype by environment interaction (G x E) on the performance of grain yield and other traits on released sesame varieties in Ethiopia. But still, the multi-environment trial in the sesame population in Ethiopia is limited and did not cover the major sesame producing areas that showed the importance of studying the effect and magnitude of G x E in sesame

genotypes including more locations and generate to determine the stable genotypes and conducting environment analysis.

Ethiopia is considered one of the centers of genetic diversity of sesame crop and has an immense wealth of genetic diversity in the germplasm collections that can be exploited through genetic improvement in the breeding program (IBC, 2012). This fact also describes from the time of N.I. Vavilov (Vavilov, 1951) based on morphological level studies (Sileshi, 2008; Gidey *et al*., 2012; Teklu and Kebede, 2014; Abate and Mekbib, 2016). The effective utilization of these collections requires a systematic genetic characterization, assessment of genetic diversity, and identification of potential putative genes that can be used in the breeding program. Despite the huge amount of locally collected and introduced germplasm held in the Ethiopian gene bank, and in breeders' stock, its morphological and molecular characterization is limited.

This showed the importance of characterizing the sesame germplasms found in the gene bank for its use in the breeding program. This analysis can be done using morphological, biochemical, and molecular markers. Very few studies on molecular characterization of sesame were reported that used a limited number of populations and not advanced molecular marker types.

Studies have been employed on molecular markers to assess genetic diversity using various types of markers (Gebremichael and Parzies, 2011; Abate and Mekbib, 2015b; Dagmawi, *et al*., 2015). Among these markers Amplified Fragment Length Polymorphism (AFLP) (Laurentin and Karlovsky, 2006), sequence-related amplified polymorphisms (Zhang *et al.*, 2010, 2012), Inter-Simple Sequence Repeat (Kumar *et al*., 2012), simple sequence repeats (Park *et al*., 2011), expressed sequence tag (Farshadfar

and Farshadfar, 2008; Zhang *et al.*, 2012), and insertions and deletions (Wu *et al.*, 2014) have been used elsewhere for the analysis of germplasm genetic diversity.

More recently high-thorough put marker systems particularly single-nucleotide polymorphisms and Diversity Arrays Technology (DArT) markers have become the genetic markers of choice for genetic analyses including characterization of germplasm because of the efficiency, low cost, speed, and abundance in the genome (Gupta *et al*., 2001; Gupta *et al*., 2008; Wei *et al.*, 2015; Cui *et al.*, 2017; Basak *et al*, 2019). SilicoDArT markers are dominant microarray markers and scored for the ‘’presence’’ or ‘’absence’’ of a single allele, whereas DArTseq based SNPs are co-dominant markers, both of them being successfully applied in genetic diversity (Wenzl *et al.*, 2004; Yang *et al.*, 2006; Bolibok-Bragoszewska *et al.*, 2009; Sánchez-sevilla *et al*., 2015; Tang *et al.*, 2015) and population structure (Matthies *et al.*, 2012; Laidò *et al.*, 2013) studies of several crop species.

Even if sesame is the most important oil crop in the world limited information is available on the genetic diversity using advanced molecular markers (Bedigian and Harlan, 1986; Bhat *et al*., 1999; Bedigian, 2010). This called the importance of developing research activities to assess the genetic diversity of the sesame germplasm collections available using modern molecular markers techniques and generate information that may help to exploit it in the future breeding program.

Besides studying the genetic diversity, identifying the markers associated with the specific trait of interest will be crucial and helpful to accelerate the genetic gain through breeding. Genome-wide association mapping study (GWAS) is a method taking advantage of both the wide phenotypic variation and the high number of historical

recombination events in natural populations, to identify the markers associated with a specific trait of interest and employed in various crop species (Guo *et al.*, 2013; Huang and Han, 2014). The application of GWAS analysis in sesame was reported by Wei *et al.*, (2015) using 705 diverse sesame germplasm accessions on 56 agronomic traits for the first time. But in contrast to its importance, the GWAS analysis reports are limited to African sesame natural populations.

In the current study, 300 genotypes comprised of 209 Ethiopian landraces, 75 exotic collections, and 16 released varieties that were genotyped using SNPs markers under the DArTseq platforms were evaluated. The same genotypes were also phenotyped for different agronomic and morphological descriptors to undertake genetic diversity and GWAS analysis. In a separate experiment elite and released sesame genotypes from the Ethiopian breeding program were evaluated across locations to determine the genotype x environment interaction and its stability.

Therefore, the study designed to answer the following major objectives;-

1.2. General Objective

Evaluate the performance of Ethiopian sesame collections in Ethiopia and its genetic diversity based on morphological traits and SNP markers, as well as identifying SNP markers associated with important agronomic traits.

1.3. Specific Objectives

- To determine the level of morphological variation and association of traits in sesame germplasm of Ethiopia

- To estimate genetic variability of sesame germplasm using agro-morphological traits and SNPs markers
- To identify highly diverse germplasm for the purposes of broadening the genetic bases of sesame germplasm grown in Ethiopia.
- To identify SNPs markers associated with economically important traits in sesame that can be used in MAS (Marker-Assisted Selection) in the future breeding program.
- To evaluate the performance and stability of sesame landraces in different growing environments, and characterize sesame growing environments in Ethiopia.

Chapter 2. Literature Review

2.1. Taxonomy and Biology of sesame

Sesame belongs to the family Pedaliaceae and genus *Sesamum*. The genus *Sesamum* comprises 20 accepted species native to Africa and Asia (Bedigian, 2015), but only *S.indicum* has been recognized as a cultivated species within the family. The family pedaliaceae encompasses a superior ovary, usually two-celled, which is totally or partially divided by false septa, each compartment having one to many ovules attached to a central placenta (Demetrios, 1980). Morinaga et al.(1929) reported the chromosome number of *S. indicum as* 2n=26. Sesame has an erect growth habit that may reach a height of 2m depending on the cultivar and growing conditions. Some cultivars are highly branched, while others grow relatively unbranched. The leaves are variable in size and shape, opposite or occasionally alternate. Flowers are zygomorphic, solitary, occasionally growing as two or three together, generally pale or rose-colored born within the axils of the leaves and on the upper stem or branches. The fruit is a deeply grooved capsule that contains 50 to 100 or more seeds depending on the environment and cultivar. The seed varies in color. The color of the testa are often black, white, golden, brown, or grey, but also dark grey and very dark brown seeds can be found. The weight of 1000 seeds is usually between 2 and 4g. Under optimum conditions, the crop produces an extensively much-branched root system. The growth of sesame is indeterminate, that is the plant continues to produce leaves, flowers, and capsules as long as weather permits (Weiss, 1971; Bedigian, 2015). Cultivated sesame has been described as a self-pollinated species. However, varying degrees of natural crossing to the extent of two to 48% are reported (Gebremichael and Parzies, 2011). Sesame is an herbaceous annual plant requiring 80-130 days to achieve physiological maturity. It requires 25 ^{0}C to 27 ^{0}C for rapid

germination, initial growth, and flower formation while temperature below 18 ^{0}C after germination restricts growth and high temperature (> 40 ^{0}C) during flowering reduces fertilization. Sesame is very drought resistant/tolerant, due in part to an extensive root system but it requires adequate moisture for germination and early growth. It is extensively susceptible to waterlogging and heavy continuous rains at all stages of development (Weiss, 1971; Seegler, 1983; Ashri, 1998).

2.2. History of sesame cultivation and utilization

Sesame (*Sesamum indicum* L.), is an erect annual herb commonly referred to as sesamum, benniseed, or simsim. It's one of the oldest and most traditional oilseed crops, valued for its high-quality seed oil. Based on recent archeological findings, sesame cultivation was derived from wild populations native to South Asia, and its cultivation was established in South Asia from the time of the Harappan civilization and spread west to Mesopotamia before 2000 B.C. (Fuller, 2003). Despite other claims, it was first cultivated in Africa and later taken to India at a very early date (Simmonds and Purseglove, 1969; Alegbejo *et al.*, 2003). Tunde-Akintude *et al*. (2012) suggested that sesame was the main oil crop grown by the Indus Valley Civilization and was likely transferred to Mesopotamia around 2500 B.C. The Assyrians used its oil for various purposes like food, salves (ointments), and medicine, while Hindus believed it to be sacred. Sesame is additionally referred to as the "queen of oilseeds," regardless it is an ''orphan'' crop. Little research into sesame has been undertaken and, hence, it's not a crop mandated by any international crop research institute (Bedigian and Harlan, 1986; Bhat, *et al*., 1999), despite being cultivated in both tropical and temperate zones of Africa, Asia, Latin America, and some parts of the southern United States (IPGRI and

NBPGR, 2004; Bedigian, 2010). Sesame is adaptable to a variety of soil types, although it performs well in well-drained, fertile soils of medium texture (typically sandy loam) at neutral pH. Generally, sesame is a short-day plant that will grow in long-day areas. Depending upon the light intensity and day period in various regions, sesame has produced genotypes with different photoperiod requirements (Ashri A, 2007).

Sesame seeds have both nutritional and medicinal value because they are rich in fat, protein, carbohydrates, fiber, and essential minerals. They are utilized in sweets like sesame bars and halva (dessert), and in bakery products or milled to induce high-grade edible oil (Dorothea Bedigian, 2004). Seeds are chemically composed of 44–57% oil, 18–25% protein, and 13–14% carbohydrates (Borchani *et al.*, 2010). The oil of Sesame is known for its stability as a result of its resistance to oxidative rancidity after long exposure to air (Global Agri Systems, 2010). Generally, the oil contains 35% monounsaturated fatty acids and 44% polyunsaturated fatty acids (Hansen, 2011). The oil of Sesame has significant resistance against oxidation as a result of it containing endogenous antioxidants including lignins and tocopherols (Elleuch *et al.*, 2007; Lee *et al*., 2008). There are two types of lignins, (i) sesamin and (ii) sesamolin, in sesame oil. Sesamolin is converted to sesamol after roasting. The molecular structure of sesamol consists of phenolic and benzodioxide groups. The phenolic group is responsible for antioxidant activities in many natural products, while the benzodioxide group is involved in anticancer and antioxidant activities. Recent research into sesame showed that it contains immunoglobulin E (IgE)–mediated food allergens (Pastorello *et al*., 2001; Dalal *et al.*, 2002; Agne *et al*., 2003). The preponderance of allergy to sesame seed is associated with its wider use in baked and fast food products.

2.3. Sesame production and constraints

The world sesame production is about 6.5 million metric tons (MT) behind soybean, groundnut, cottonseed, sunflower, linseed, and rapeseed, within the amount of world oilseed production. The average sesame productivity of the world's top producing countries within 10 years (2010–2019) is given in Figure 2.1. However, the data for Sudan and South Sudan is o*ffi*cially recorded in 2012 by Food and Agriculture Organization Statistical Databases. Myanmar, India, and Sudan are the highest producers among the countries. Average sesame yield is found to be highest in China (1.6 t ha^{-1}) followed by Nigeria (0.8 t ha^{-1}, Tanzania (0.72 t ha^{-1}), and Ethiopia (0.7 t ha^{-}) (*FAOSTAT*, 2019).

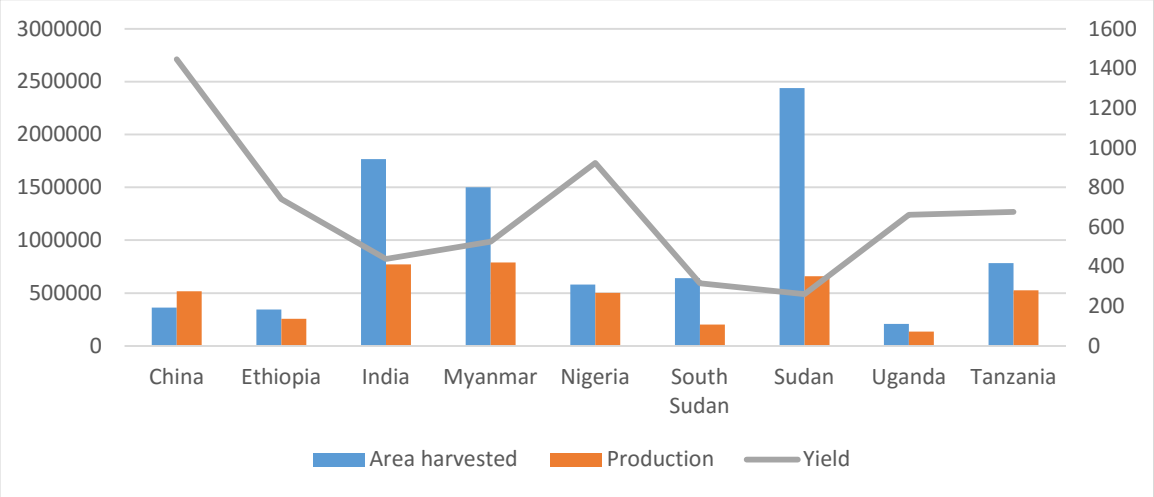


Figure 0-1 Trend of sesame production values in top producing countries during the last twenty years (Average 2010–2019)

Source: *FAOSTAT* (2019)

In 2019, 6.5 million MT was grown worldwide on 12.8 million hectares (ha) with an average yield of 0.5 t ha^{-1} (Table 2.1). Asia and Africa produce nearly 95% of the world's

supply of sesame. Sudan devotes the largest acreage however, it has lowest records of yield per hectare. Sudan produces nearly 18.23% of the world's sesame crop, followed by Myanmar at 11.37% and India at 10.52% (Table 2.1).

Table 0.1. Area, production, and yield of sesame within the selected countries and also the world in 2019.

	Area (ha)	Production (MT ha −1)	Yield (kg ha−1)	% of world production
Top producing countries				
China	289801	469104	1619	7.2
Ethiopia	375120	262654	700	4.0
India	1419970	689310	485	10.5
Myanmar	1505163	744498	495	11.4
Nigeria	586539	480000	818	7.3
South Sudan	607226	208109	343	3.2
Sudan	4243680	1210000	285	18.5
Uganda	212000	144000	679	2.2
Tanzania	940000	680000	723	10.4
Regions				
Asia	3661172	2247431	614	34.3134
Africa	8737270	3998148	458	61.0430
America	423271	304107	719	4.6430
Europe	39	39	1000	0.0006
World	12821752	6549725	511	

Source: (*Food and Agriculture Organization Statistical Databases (FAOSTAT)*, 2019).

Ethiopia is one of the top ten sesame producing countries and ranks the six largest in cultivated area and the seventh-largest in production. In 2019, there were about 543,236 sesame growers most of which were smallholder farmers with an average acreage of 0.7 t ha−1 and produced 262,654.189 tons of sesame. Besides smallholders, there are a limited number of investors or large commercial farmers (having more than 100 ha). The share of the latter is less than 2% (Wijnands and Biersteker, 2007).

Sesame seed is the second-largest agricultural export earner for Ethiopia, next to coffee, involving a number of small-holder farmers in its production throughout the country (CSA, 2019). In the 2018 production year, about 474,747 smallholder farmers participate in sesame production; while in the year 2019 the number of participants has increased to 543,236 farmers (CSA, 2019). This indicates that sesame sector has the potential to involve more smallholders under its production, revealing the need to link them to domestic and international markets. Besides, in Ethiopia produces a large variety of sesame seeds can be produced, among which the Humera, Gondor, and Wellega types are well known in the world markets. On one hand, the Humera and Metema sesame seeds are suitable for bakery and confectionary purposes due to their white color, sweet taste, and aroma. On the other hand, the high oil content of the Wellega sesame gives it a major competitive advantage for edible oil production (USAID, 2010). According to different reports, sesame seed is an important export crop in Ethiopia and the country has a substantial role in the global sesame trade. It is the third world exporter of the commodity after India and Sudan (Dawit, 2010). In this regard, in the last few years, sesame production and marketing have confirmed highly significant growth. In 1997, the total area under sesame production was about 64,000 ha (Aysheshm, 2007). In the 2019 cropping season, the total area under sesame production reaches 375,119.95 hectares and about 262,654.189 tons of sesame seed has produced in the country (CSA, 2019).

Despite the potential for increasing the production and productivity of sesame, there are also a number of constraints inhibiting sesame production and productivity. Aysheshm, (2007); Kefyalew, (2012) identified the following major constraints in sesame production and marketing in Ethiopia. Among Problems related to sesame production, the

productivity of the crop is highly dependent on the amount and distribution of rainfall. Problems related to rainfall intensity and incidence of crop pests and diseases are affecting production and productivity. The other is the lack of input supply and extension services, Though efforts were made to improve sesame seed varieties through agricultural research, the inability to provide market demanded improved and appropriate varieties is still an outstanding problem. Sesame productivity is low due to the low level of extension service. Harvesting requires a high demand of the labor force. In the major producing areas, there is a shortage of daily laborers during peak periods due to different health hazards. Illegal trading of farm oxen has also resulted in a shortage of plow animals. The absence of modern warehouses in the nearby areas has resulted in the mishandling of output. Producers are unable to build their own storage devices due to tenure insecurity. This in turn resulted in the absence of soil and water conservation measures and planting permanent crops.

Shortly summarized as follows:-

- **Production and productivity challenges**

- Lack of high yielding and improved varieties for different agro-ecologies with selected agronomic traits viz. non-shattering, diseases/pests resistance
- Low level of improved input utilization;
- High postharvest loss;
- Highly dependent on rainfall;

- **Infrastructural challenges**

- Insufficient or limited rural feeder roads and transportation;
- Very limited modern warehouses and facilities; Limited modern marketing centers /primary, secondary and terminal markets/ especially for other oilseeds except for sesame

- **Skill and knowledge challenges**

- Insufficient post-harvest technologies and know-how;
- Luck of advanced value addition technologies
- Low level of modern packaging and processing skill
- Low level of information networking and processing
- Lack of collaboration between breeding institutes and food industries

- **Competitiveness challenges**

- Price fluctuation /Price volatility
- Construct default (both sides-exporter and buyer)
- Low Negotiation capacity of exporters

2.4. Sesame breeding in Ethiopia

Sesame research in Ethiopia started in the late l960s by the Ethiopian Institute of Agricultural Research (EIAR) at Werer Agricultural Research Center (WARC) under irrigation with landraces and exotic germplasms. The objective of sesame breeding in Ethiopia is to develop the potential by creating cultivars which meet the demands of the sesame growers, processor, and consumers. The improvement was targeted to the higher seed yield and oil content. Ethiopian sesame breeders have come to rely on breeding methods involving the use of collection, introduction, selection (mass and pure line), hybridization (followed by pedigree, and bulk method of breeding), and recently induced mutation (Gebremichael, 2017). A crossing program was started at Werer Agricultural Research Center in 1983 with the main aim to incorporate high seed and oil yield, drought and disease resistance (bacterial blight), partial-shattering, and determinate flowering characters. Subsequent segregating populations were handled using pedigree and bulk methods of breeding (WARC, 2006). An induced mutation program using chemical mutagens is in progress for the generation of variability for some important traits such as shattering resistance since sufficient variability in the existing germplasm is lacking (WARC, 2015). Generally, sesame breeding research in Ethiopia can be divided into three inter-related periods: First period (the late 1960s to 1979) dealt with the collection, introduction, characterization, and evaluation of sesame germplasm for identifying suitable and best adaptable sesame cultivars for the potential areas, sesame breeding relying entirely upon pure-line and/or mass selection from the local germplasm and introduction. Second period (1980–2007) characterized by the incorporation of the crossing program into the already pre-existing breeding methods. Third period (2008 to

present) marked by the initiation of molecular approaches including analyses of molecular genetic diversity, development of molecular markers, and initiation of induced mutation techniques particularly for pod-shattering (Gebremichael and Parzies, 2011; Dagmawi *et al*., 2015). There are several sesame varieties released in Ethiopia. Some of them are under production and trial for various agronomic parameters under different climatic and soil conditions. They are differently stable for their oil and/or seed at different agro-ecologies (Daba *et al*., 2014). There are some improved varieties of sesame released in Ethiopia at different places and times. Here below are some varieties released from 2010 up to 2017 by different regional and national research centers and Haromaya University (Table 2.2).

Table 0.2. Some improved varieties of sesame released by the different regional and national research centers and Haromaya University from 2010 to 2017

Varietal name	Year of release	Pedigree name	Altitude (a.m.s.l)	Average grain yield (kg/ha)		Oil content (%)	Seed color	Maintainer/released
				On farm	On station			
Obsa	2010	EW004	1250-1650	868.8	1069	51.55	White	Bako ARC
Dicho	2010	EW015	1250-1650	810.6	1063	53.8	White-tan	Bako ARC
Barsan	2010	ACC-00016	500-700		960 irrigate	45.9	Brown	Gode ARC
Lidan	2010	ACC-00044(2)	500-700		1080 irrigate	46.9	Brown	Gode ARC
Humera-1	2011	ACC.038sel.1	600-1110		590-900	54.56	White	Humera ARC
Setit-1	2011	col sel p#1	600-800		620-1000	52.54	White	Humera ARC
ACC-00047	2013	ACC00047	_		700-800	50.4	White	Sirinka ARC
Chalesa	2013	EW023(2)	1350-1650	975-1200	1050-1480	51	light White	Bako ARC
Dangur	2015	EW013(8)	_		750	56.7	Gray	Pawe ARC
Benshangul-1	2016	WW-001(6)	740-1280	400	435-836	54.1	White	Assosa ARC
Setit-2	2016	J-03	600-1028	800	913	53.77	White	Humera ARC
Gonder-1	2016	ACC.ba002	760-1022		500-900	50	White	Gonder ARC
BaHnecho	2016	ACC-EW-012(5)	560-1650	800-1200	1200	52	White	Haromaya University
BaHzeyit	2016	ACC-EW-023(1)	560-1650	1000-1300	1300	56	light gray	Haromaya University
Waliin	2017	BG-004-1	1250-1450	950-1100	1000-1380	_	light White	Bako ARC

Source: MoARD,crop variety register book 2010-2017

2.5. Genetic Diversity in Ethiopian Sesame

The Genetic diversity of Ethiopian sesame has been studied using morphological and molecular markers. Agro morphological marker has been a primary tool to estimate genetic differences among Ethiopian sesame genotypes (Sileshi, 2008; Gidey *et al*., 2012; Yirgalem, 2013; Abate and Mekbib, 2016).

Gidey *et al*. (2012) studied the genetic diversity of morphological traits of 81 Ethiopian sesame accessions. The result indicated that the accessions were grouped into six clusters of different sizes. There was no definite relationship between geographic and genetic diversity as overlapping was encountered in clustering patterns among accessions from different parts of the country.

Abate and Mekbib, (2016) also studied the genetic diversity based on agro morphological traits of 49 sesame genotypes collected from low-altitude areas of Ethiopia. The Analysis of variance revealed a significant difference among the genotypes for each character except for primary branches. A higher genotypic variance was observed for seed yield and the number of capsules/plants. D^2 values and clustering patterns revealed the existence of a considerable amount of genetic diversity among the studied genotypes. Genotypes from different regions were closely related while genotypes from the same region were grouped in different clusters, indicating the absence of a relationship between genetic diversity and geographic diversity. Under their study, they observed wide range of variation for most of the traits cluster mean values.

Information on the use of different and recent molecular markers for the characterization of genetic diversity in Ethiopian sesame is limited. Molecular markers such as Simple

sequence repeat (SSR) markers (Gebremichael and Parzies , 2011), inter simple sequence repeat (ISSR) markers (Alemu and Tesfaye, 2013; Teshome, Kassahun, and Bekele, 2015), and Random Amplified Polymorphic DNA (RAPD) markers (Abate and Mekbib, 2015b) were applied to Ethiopian sesame genetic diversity studies.

Gebremichael and Parzies, (2011), conducted research on the genetic diversity of Ethiopian sesame using ten Simple Sequence Repeats (SSRs) markers with 50 sesame populations representing the existing Ethiopian collections. The result indicates that the number of alleles ranged from 6 to 17, with an average of 12.1 per locus. The Polymorphism Information Contents of the markers ranged from 0.393 to 0.820. The observed landraces showed higher Nei's average genetic diversity (He=0.377) than the cultivars (He=0.305). The fixation index values for the landraces and cultivars were 0.427 and 0.26, respectively, indicating that genetic divergence between populations was smaller than genetic divergence within. The dendrogram based on Modified Rogers' Distances (MRD) and a principal coordinate analysis based on a simple dissimilarity matrix, points towards some grouping based on geographical origin. Overall, Ethiopian sesame is genetically diverse, a factor that needs to be considered when planning conservation strategies or in breeding programs.

Alemu and Tesfaye, (2013) reported research on the Genetic diversity of Ethiopian sesame using six farmers' cultivars and varieties of sesame from north-western Ethiopia each consisting of ten individual samples using four inter simple sequence repeat (ISSR) markers. The four ISSR primers produced 37 amplification products of which 36 bands (97.30%) exhibited polymorphism. The result of the analysis of molecular variance (AMOVA) revealed (a slightly) higher percentage of genetic variation among the

cultivars and varieties (54.37%) than individuals within each cultivar and variety populations (45.63%). The unweighted pair-group method with arithmetic averages (UPGMA) cluster analysis grouped most of the studied samples to their population. The result indicated the presence of high genetic diversity in the study area that deserves formulation of breeding strategy for improving the productivity of the crop in the country (Alemu and Tesfaye, 2013).

(Abate and Mekbib, 2015b) also reported the genetic diversity of 128 sesame (Sesamum indicum L.) genotypes using 10 geographically distinct populations in Ethiopia assessed at the DNA level using RAPD analysis. Based on the result Percent of polymorphic loci (P%), the number of different (Na) and effective (Ne) alleles along with Shannon information index (I) and Nei's gene diversity (He) values suggested that populations from Afar (cultivars) and AM-NSh were found to be the least diverse while the population of Oromia germplasm was the most diverse of all populations.

Based on The UPGMA clustering the genotypes clustered into 3 major groups and 11 subgroups. Generally, both clustering and PCoA patterns revealed that most genotypes located geographically far apart were found to cluster in the same group, while those genotypes from the same origin dispersed. The results indicated that the RAPD method revealed a high level of genetic variation among sesame genotypes.

In addition, Teshome et al. (2015) reported the genetic diversity study of 120 (82 Ethiopian and 38 exotic) sesame accessions using six ISSR primers. The result indicates the presence of higher polymorphism among accessions collected from Ethiopia (75.85) than the exotic accessions (65.52). The average gene diversity relative to the overall population was 0.24. Samples from Wellega were the most diverse, with a gene diversity

value of 0.26 followed by samples from Shewa (0.20), and Tigray (0.20). Samples from Sudan (0.12), and Gojam (0.10) were the least diverse. For the overall population, the Inter-population genetic distance (D) ranged from 0.031 to 0.165. From the exotic accession, samples of South East Asia are distantly related to most of the Ethiopian accessions. Unweighted pair group method with arithmetic mean analysis (UPGMA) of Ethiopian sesame populations revealed two major groups and three outliers (Cultivated, Wellega and, Illubabore).

2.6. Genotyping-by-Sequencing (GBS)

Generally, the use of molecular marker technology in sesame breeding is a new occurrence. However, there have been efforts on the use of molecular techniques to complement and hasten conventional sesame breeding.

The advances of next generation sequencing driven the costs of DNA sequencing down to the point that GBS is now feasible for large genome species and high diversity (Elshire *et al.*, 2011). GBS is simple highly multiplexed system for constructing reduced representation libraries for the Illumina NGS next generation sequencing platform developed in the Buckler lab (Elshire *et al.*, 2011). It generates large numbers of SNPs for use in genetic analyses and genotyping (Beissinger *et al.*, 2013). The main components of this system include reduced sample handling, low cost, fewer PCR and purification steps, no reference sequence limits, efficient barcoding, no size fractionation, cost-effective, and easiness to scale up (Davey *et al.*, 2011). GBS is becoming increasingly important as unique tool for genomics-assisted breeding in a range of plant species.

GBS combined with genome-independent imputation provides an efficient and a simple method for the construction of a genetic map in any pseudo-testcross progeny (Ward1 *et al.*, 2013). GBS method offers a greatly simplified library production procedure more amenable to use on large numbers of lines/individuals (Elshire *et al.*, 2011). A two-enzyme (*PstI/Msp*I) GBS protocol, which provides a greater degree of complexity reduction and uniform library for sequencing than the original protocol using *Ape*KI, has been developed and applied to both wheat and barley (Poland and Rife, 2012). Sonah *et al.* (2013) reported a modified library preparation protocol, in which selective amplification is used to increase both their depth of coverage and the number of SNPs called, resulting in high efficiency to allow an important reduction in per-sample cost.

Two different GBS strategies have been developed with the Ion PGM system (Poland *et al.*, 2012). (A) Restriction enzyme digestion, in which no specific SNPs have been identified and ideal for discovering new markers for marker assisted selection programs. The complexity of the genome under this approach is reduced by digesting the DNA with one or two selected restriction enzymes prior to the ligation of the adapters. (B) Multiplex enrichment PCR, in which a set of SNPs has been defined for a section of the genome. This approach uses PCR primers designed to amplify the areas of interest.

2.6.1. Application of GBS in plant breeding

Genotyping-by-sequencing is an ideal platform for studies ranging from single-gene markers to whole-genome profiling (Poland and Rife, 2012). Among the most important of applications of GBS is the field of plant breeding. GBS gives low-cost and a rapid tool to genotype breeding populations, allowing plant breeders to implement genomic molecular marker discovery, genetic linkage analysis, diversity study, Genome-Wide

Association Mapping (GWAS), and genomic selection (GS) under plant breeding programs. One of the advantages of the GBS method is no need for a priori knowledge of the species genomes as it has been shown to be robust across a range of species and SNP discovery and genotyping are completed together (Poland and Rife, 2012; Narum, *et al.*, 2013).

As GWAS requires 100s of 1000s to millions of markers to generate sufficient coverage, and information, the emergence of NGS technologies has greatly improved such marker resolution (Edwards and Batley, 2010). Recently, GBS through the NGS approach has been used to resequencing collections of recombinant inbred lines (RILs) to analyze and map various traits of interest in specific breeding programs (Deschamps *et al*., 2012). More crops, such as barley, wheat, maize, rice, potato, sesame, and cassava have been optimized by GBS for the low-cost, efficient, and large scales of genome sequencing (Poland and Rife, 2012; van Poecke *et al.*, 2013). GBS enabled the identification of more than 15,000 SNP loci in the sesame genome, proving superiority over SLAFseq and RADtaq sequencing approaches in high-throughput polymorphism discovery (Uncu *et al.*, 2016).

2.6.2. Genome-wide Association Mapping (GWAS)

Genome-wide association studies have been widely used to identify QTL underlying quantitative traits in humans and animals and have recently also become a popular method of mapping quantitative traits loci (QTL) in plants. Association mapping identifies QTL based on the historic recombination in a panel of diverse germplasm by the presence of linkage disequilibrium (LD) between SNPs and QTL, i.e., the non-

random association of alleles (Zhou *et al.*, 1998; Zhu *et al.*, 2008). A high- density marker panel that covers the genome is required in order to monitor the density of recombination breakpoints in the population (Flint-Garcia, *et al*., 2003; Zhu *et al*., 2008).

GWAS is most commonly performed in diversity panels, this indicates, collections of unrelated diverse germplasm, that is important to maximize the diversity of haplotypes, and alleles (Huang *et al*., 2010; Famoso *et al.*, 2011; Huang *et al*., 2012; Huang and Han, 2014; Yang *et al.,* 2014). While this is advantageous in terms of identifying candidate genes, and novel QTL that underlie agronomic traits of interest, it also requires that any identified QTL be validated in a breeding population before they can be used for genomics-assisted selection. For this reason, it is of interest to perform GWAS in a population of adapted lines. QTL identified in this way could be more directly utilized for marker-assisted selection (MAS) and/or genomic selection in applied breeding programs (Zhang *et al*., 2014).

Genomic selection (GS), or Genome-wide prediction, indicates the process of using genome-wide DNA markers to estimate which individuals in a breeding population are most valuable as parents of the next generation of offspring. GS takes the same inputs as GWAS, a phenotype dataset and genotype dataset on a population of lines of interest to plant breeders (Heffner, *et al*., 2009; Lorenz *et al*., 2011; Zhang *et al.*, 2014). As such, it is possible to perform GS and GWAS on the same population, where all that is needed in additional computation analysis. Such an undertaking has clear advantages. The genetic architecture revealed by association mapping can be used to inform the GS models—for example, if highly significant SNPs are revealed by a GWAS, these SNPs could be fit as

fixed effects in a GS model (Bernardo, 2014; Zhang *et al.*, 2014), and experimenting with different types of genomic selection statistical methods (i.e., linear versus non-linear, additive versus non-additive) can corroborate inferences about the genetic architecture of a trait.

Chapter 3. Morphological Characteristics, Genetic Variability and Diversity of Ethiopian sesame Genotypes

African Crop Science Journal, Vol. 29, No. 1, pp. 59 - 77
Printed in Uganda. All rights reserved

ISSN 1021-9730/2021 $4.00

DOI: https://dx.doi.org/10.4314/acsj.v29i1.5

MORPHOLOGICAL CHARACTERISTICS AND GENETIC DIVERSITY OF ETHIOPIAN SESAME GENOTYPES

T. TESFAYE[1,2], K. TESFAYE[2], G. KENENI[3] and T. ALEMU[3]

[1]Amhara Agricultural Research Institute, Gondar Research Center, P. O. Box 1337, Gondar, Ethiopia
[2]Addis Ababa University, College of Natural Sciences, Department of Microbial, Cellular and Molecular Biology, P. O. Box 1176, Addis Ababa, Ethiopia
[3]Ethiopian Institute of Agricultural Research, Holeta Research Center, P. O. Box 2003, Holeta, Ethiopia
Corresponding author: tesfaye.tewodros@yahoo.com

(*Received 14 September 2020; accepted 4 January 2021*)

ABSTRACT

Sesame (*Sesamum indicum* L.) is produced worldwide, although more than 96% of the world sesame seed is produced in Africa and Asia. The objective of this study was to determine morphological properties and identify the genetic diversity of cultivated sesame genotypes grown in different parts of Ethiopia. Three hundred sesame genotypes collected from diverse ecologies of Ethiopia and introduced from different African and Asian countries, were used in this study. Genotypes showed wide variability for most morphological traits, except for plant growth type, leaf glands, anther filament colour, anther connective tip gland, and anthocyanin colouration of the capsule. Genetic divergence using Mahalanobis D2 statistics was computed, and the genotype lines were grouped into six different clusters. Clustering was not associated with the geographical distribution; instead genotypes were grouped mainly based on morphological differences. The lowest divergence was noticed between cluster I and V (10.06). Maximum inter-cluster distance was observed between clusters IV and VI (D2 =342.56, followed by clusters I and VI (D2 =217.9783), and III with IV (D2 =190.8707). Maximum genetic recombination and variation in the subsequent generation, is expected from crosses that involve parents from the clusters characterised by maximum distances. Thus, maximum distances or varation could maximise opportunities for transgressive segregation, since unrelated genotypes would contribute unique desirable alleles at different loci.

Key Words: Inter-cluster distance, Mahalanobis D2 statistics, transgressive segregation

Published; - T. TESFAYE, K. TESFAYE, G. KENENI, and T. ALEMU 2021

Abstract

Sesame (Sesamum indicum L.) is produced worldwide, although more than 96% of the world sesame seed is produced in Africa and Asia. The objective of this study was to determine morphological properties and association of traits, to estimate genetic variability of sesame genotypes for agro-morphological traits, and to identify the genetic diversity of cultivated sesame genotypes grown in different parts of Ethiopia. Three hundred sesame genotypes collected from diverse ecologies of Ethiopia and introduced from different African and Asian countries, were used in this study. Genotypes showed wide variability for most morphological traits, except for plant growth type, leaf glands, anther filament colour, anther connective tip gland, and anthocyanin colouration of the capsule. High heritability combined with high genetic advance was recorded for plant height, primary branch, petiole length of top leaf, days to flower initiation, days to 50% flowering, pod bearing zone, seed yield per plant, and bacterial blight reaction indicating the potential of improving the population through a direct selection for these traits. Grain yield showed a significant and positive genotypic correlation with plant height, a number of capsule per plant, and pod bearing zone, the magnitudes of the positive genetic correlation suggest that the selection by those characters produces a significant increase in grain yield. Genetic divergence using Mahalanobis D^2 statistics was computed, and the genotype lines were grouped into six different clusters. Clustering was not associated with the geographical distribution; instead genotypes were grouped mainly based on morphological differences. The lowest divergence was noticed between cluster I and V (10.06). Maximum inter-cluster distance was observed between clusters IV and VI (D^2 =342.56, followed by clusters I and VI (D^2 =217.9783), and III with IV (D^2 =190.8707). Maximum genetic recombination and variation in the subsequent generation, is expected from crosses that involve parents from the clusters characterized by maximum distances. Thus, maximum distances or variation could maximize opportunities for transgressive segregation, since unrelated genotypes would contribute unique desirable alleles at different loci.

Keywords: *Genetic advance, heritability, Traits diversity, and variability*

INTRODUCTION

Sesame (*Sesamum indicum* L.) is a major oil seed crop worldwide, with more than 96% of seed production accounted for by Africa and Asia *(FAOSTAT*, 2019). Sesame seeds are richly endowed with oil (44-57%), protein (18- 25%) and carbohydrates (13-14%) (Borchani *et al.*, 2010). Cultivated sesame has been described as a self-pollinated species; however, varying degrees of natural crossing to the extent of 2 to 48% have been reported (Gebremichael and Parzies, 2011). In Ethiopia, sesame accounts for about 44% of the total acreage and 34% of gross production of major oilseeds cultivated in the country. It grows in almost all regions of the country with altitudes less than 2000 m above sea level (Woldemariam, 1985; Adefris *et al*., 2011). Farm level national average productivity in Ethiopia is lower (0.7 t ha-1) (CSA, 2019) than the potential yield of 2 t ha-1 (Mkamilo and Bedigian, 2007) Improved varieties released in Ethiopia reportedly yield 0.3 to 1.3 t ha-1 under rainfed, and 1 to 2.4 t ha-1 under irrigated conditions (Gebremichael, 2017). Sesame is one of the major indigenous oilseeds, displaying considerable diversity in Ethiopia (IBC, 2012). Moreover genetic diversity in crop plants is essential to sustain high productivity (Rabbani *et al*., 2010). In addition to select the desirable genotype and to plan the breeding program, Knowing the nature and extent of the genetic variability with the information on the association of plant character with grain yield is great importance.

A large number of Ethiopian sesame germplasm, locally collected and introduced, are held in the gene bank and in breeders' stock. There have been efforts on the use of agro morphological characterization, since the agro morphological marker has been a primary tool for estimating genetic differences among Ethiopian sesame genotypes (Teklu and

Kebede, 2014; Abate and Mekbib, 2016). Genetic diversity with agro-morphological characterization and eco-geographic distribution; and microcenters of the diversity have a number of limitations. Despite the huge amount of locally collected and introduced germplasm held in the Ethiopian gene bank, and in breeders' stock, sesame morphological characterization has only been done on a limited number of genotypes. The objective of this study was to determine morphological properties, to assess the genetic variability of sesame genotypes for agro-morphological traits, and to estimate the extent of correlation between pairs of yield and yield related traits, and to compare the direct and indirect effects, to identify and cluster the Ethiopian sesame genotypes of different origins into similarity groups and assess the extent and pattern of diversity of the genotypes.

MATERIALS AND METHODS

Materials used

A total of 300 sesame genotypes, comprising 225 local Ethiopian collections, including 16 released varieties and 75 exotic collections received from the Ethiopian Biodiversity Institute (EBI), and different federal and regional research center of Ethiopia, were used in this study. The collections were mainly from different regions of Ethiopia, and different countries of Africa and Asia. The number of sesame genotypes used in this study and their countries of origin are listed in Table 3.1. The genotypes were planted at the Metema trial site (120 39'N, 360 17' E) in the 2017/18 cropping season, and this was repeated in the 2018/19 cropping season at Metema and Tach Armacho trial (13088'N, 370 43'E) sites. Metema is located at 760 meters above sea level and receives 1030.2 mm

of rainfall per *annum* and its soil is a Vertisol. Tach Armacho is located at 1022 meters above sea level and receives 970.88 mm of rainfall per *annum*, also with a Vertisol. The study was laid out in an alpha lattice design, with each plot consisting of two rows of 4 m length with a spacing of 40 cm between rows and 10 cm between plants. Up to 65 kg ha-1 of Urea, with two times split application, was applied manually based on unpublished site specific recommendations. Thinning and hand weeding were carried out. All quantitative traits (plant height, primary branch, secondary branch, length of basal leaf, width of basal leaf, length of middle leaf, width of middle leaf, length of top leaf, width of top leaf, petiole length of basal leaf, petiole length at middle (mid-level/mid-height) leaf, petiole length of top leaf, days to flower initiation, days to 50% flowering, number of capsules per plant, mean capsule length, mean capsule width, mean capsule thickness, seeds per capsule, 1000-seed weight, days to physiological maturity, pod bearing zone, seed yield, bacterial blight) and qualitative traits plant growth parameters (plant growth habit, root system, main stem colour, stem hairiness, shape of hair, stem shape in cross section, stem fasciation, stem branching, branching pattern, leaf colour, leaf hairiness, leaf arrangement, middle leaf shape, top leaf shape, basal leaf profile, basal leaf margin, lobe incision of basal leaf, leaf glands, leaf angle to main stem, petiole colour, petiole hairiness, shape of petiole hair, number of flowers per leaf axil, extra-floral nectary development, extra-floral nectar colour, calyx tip colour, calyx hairiness, shape of calyx hair, corolla hairiness, shape of corolla hair, exterior corolla colour, interior corolla colour, corolla interior pigmentation, lower lip colour, absence/presence of foveolar, anther filament colour, anther connective tip gland, style length, number of carpels per capsule, Bicarpellate capsule shape, capsule arrangement, capsule hairiness, shape of

capsule hair, anthocyanin colouration of capsule, colour of dry capsules, capsule dehiscence at ripening, type of capsule beak, and thickness of capsule mesocarp data) were recorded according to the sesame descriptors list of IPGRI and NBPGR, (2004). All measurements were done after tagging five randomly selected plant in each plot. Yield, growth period, plant and capsule variation and disease reactions, were recorded using standard procedures (IPGRI and NBPGR, 2004). The early Flowering dates were recorded as the number of days from sowing to observation of the first flower on 50% of individuals from the total. Flower and leaf-related traits (i.e. days to flower initiation, days to 50% flowering and days to physiological maturity) were observed and measured in the full-bloom stage. After harvesting manually, yield-related traits including number of carpels per capsule, seeds per capsule, 1000-seed weight and seed yield were measured in the laboratory. Seed yield was collected per plot and later converted into metric tonnes per hectare.

Table 0.1. List of origin countries and number of sesame genotypes used for this study

Collection from Asia	China	Japan	Israel	Philippines			
Number of genotypes	2	1	3	1			
Collection from different Africa countries without Ethiopia	North Africa	South Africa		West Africa	East Africa		
	Egypt	Zambia	Zimbabwe	Burkina Faso	Kenya	Somalia	Sudan
Number of genotypes	27	3	15	17	3	1	2
Collection from different regions of Ethiopia	Amhara	Benshangul - Gumz	Improved	Oromia	SNNP	Tigray	
Number of genotypes	56	38	16	52	3	60	

Data Analysis

Analysis of variance

Data obtained from the different environments were analyzed separately, and thereafter combined for after the error homogeneity test, conducted between testing locations using SAS computer software (SAS, 2002). Statistically significant genotype means were separated using LSmeans package of SAS at $P < 0.05$ level.

Estimation of phenotypic and genotypic variances

Components of variance, σ^2g=genotypic variance, σ^2p=phenotypic variance, and σ^2e=error variances were calculated according to Burton and Devane (1953); Wricke and Weber (1986).

Environmental variance (σ^2e) = MSE/. Where, MSE=error mean square and r = number of replications.

Genotypic variance $(\sigma^2 g) = \frac{MSG - MSgl}{rl}$

Where, σ^2g=Genotypic variance, MSG=genotype mean square, MSgl = mean square of genotype by location interaction, r = number of replication and l = number of locations. Phenotypic variance (σ^2p) =$\sigma^2g+\sigma^2gl/1+\sigma^2e/rl$ Where, σ^2g is genotypic variance, σ^2gl is genotype by location interactions, σ^2e=Environmental variance and r is the number of replications.

Estimation of the genotypic and phenotypic coefficient of variability

Genotypic and phenotypic coefficients of variability: Genotypic and phenotypic coefficients of variability were computed according to Burton and Devane (1953).

Genotypic coefficient of variability $(\mathrm{GCV}) = \frac{\sqrt{\sigma^2 g}}{\mathrm{X}} \times 100$

Phenotypic coefficient of variability $(\mathrm{PCV}) = \frac{\sqrt{\sigma^2 p}}{\mathrm{X}} \times 100$

Environmental coefficient of variability $(\mathrm{ECV}) = \frac{\sqrt{\sigma^2 e}}{\mathrm{X}} \times 100$

Where, $\sigma^2 g$ = Genotypic variance, $\sigma^2 p$ = Phenotypic variance, and $\sigma^2 e$ = Environmental variance, X = General mean of character.

The PCV and GCV values are ranked as low, medium, and high, as follows: 0-10%- Low; 10-20%- Moderate; >20%-High (Sivasubramanian, 1973).

Estimation of heritability and genetic advance

Heritability: Broad sense heritability was estimated based on the ratio of genotypic variance to the phenotypic variance and was expressed in percentage (Falconer, and Mackay, 1996).

$$h^2(\mathrm{BS})\% = \frac{V_g}{V_P} \times 100$$

Where h^2 (BS) % =heritability in broad sense; Vg = Genotypic variance; Vp = Phenotypic variance. Heritability values are categorized as low, moderate, and high (Robinson *et al,*, 1949) as follows: 0-30%: Low; 30-60%: Moderate; 60% and above: High.

Genetic advance: The extent of genetic advance expected by selecting a certain proportion of the superior progeny was calculated by using the following formula given by Robinsonet *et al*.(1949).

Genetic advance $(GA) = k.\sigma p.h^2$

A) =k. σp.h^2 Where: k=selection intensity at 5% (k=2.06); σp = Phenotypic standard deviation;

h^2 (BS) % = Heritability in broad sense.

Genetic advance expressed as a percentage over Mean: $GAM(\%) = \frac{GA}{\times}$

Where: GAM (%) = Genetic advance over mean; x = General means of the character

Johnson *et al*. (1955) classified **genetic advance** as percentage of mean (GAM); values from 0-10% are low, 10-20% are moderate and 20% and above are high.

Correlation studies

Character associations among the measured traits were estimated at a genotypic, phenotypic, and environmental level based on genotypic, phenotypic, and environmental covariance, respectively. The analysis of covariance between all possible pairs followed the same form as the variance analysis. The mean product expectation of the covariance analysis is analogous to the mean square expectation of the analysis of variances. Thus,

the estimate of the genotypic and phenotypic covariance components between two traits (σg 1.2, σp1.2) was derived in the same fashion as those for the corresponding variance components. These covariance components were substituted in the following formula to calculate the genotypic and phenotypic correlation as described by Sharma, (1998).

$$\text{Genotypic r} = \frac{\text{COV}_{g12}}{\sqrt{(\sigma^2 g_1).\sigma^2 g_2}}$$

Where Cov_{g12} is the genotypic covariance between two traits, $\sigma^2 g_1$ is the genotypic variance of the first trait, and $\sigma^2 g_2$ is the genotypic variance of the second trait; and

$$\text{Phenotypic r} = \frac{\text{COV}_{P12}}{\sqrt{(\sigma^2 P_1).\sigma^2.P_2}}$$

Where, Cov_{p12} is the phenotypic covariance of the progeny means between the two traits, and $\sigma^2 p_1$ and $\sigma^2 p_2$ are the phenotypic variance for each trait. The calculated phenotypic correlation values were tested for its significance using the t-test:

$$t = \frac{r_P}{SE(r_p)}$$

Where, r_p = Phenotypic correlation; SE (r_p) = Standard error of phenotypic correlation was obtained using in the following procedure (Sharma, 1998)

$$SE(r_p) = \sqrt{\frac{(1 - r_p^2)}{(n-2)}}$$

Where, n is the number of genotypes tested, r_p is the phenotypic correlation coefficient. The coefficients of correlations at genotypic levels were tested for their significance using the formula described by Robertson, (1959) as indicated below:

$$t = \frac{r_{gxy}}{SEr_{gxy}}$$

The calculated "t" value was compared with the tabulated "t" value at (n-2) degree of freedom at a 5% level of significance. Where, n = number of genotypes

$$SEr_{gxy} = \sqrt{\frac{\left(1 - r_{gxy}^2\right)}{\left(2H_X . H_y\right)}}$$

Where, H_x =Heritability of trait x, H_y= Heritability of trait y

Multivariate Analyses

Multivariate analyses, including Cluster Analysis and Principal Component Analysis, were used to group the 300 sesame genotypes into respective categories. The principal components analyses were meant to identify large contributing traits to the total variation among the populations. Hierarchal clustering of accessions based on the Average Linkage Method were performed using JMP SAS software (JMP®, 2002) to group sesame genotypes. Statistics, pseudo F statistic and pseudo t^2 statistic generated by SAS were examined to decide the number of optimum clusters.

Divergence analysis.

The patterns of distribution of morphological variation were analyzed using Mahalanobis Generalized Distances (D^2). The D^2 were applied to estimate the distances between and within clusters, using the SAS computer software package as per the following formula: D^2ij = (Xi-Xj)' S-1(Xi-Xj)

Where: D^2ij is the distance between class i and j; Xi and Xj are the vector means of the traits for the ith and jth groups, and S-1 is the inverse of the pooled covariance matrix. The D^2 analysis was based on the mean values of all morphological traits across locations. The D^2 values obtained for pairs of clusters were considered as the calculated values of Chi-square (χ^2) and were tested for significance at P< 5% against the tabulated value of χ^2 for 'P' degree of freedom, where P is the number of parameters considered (Singh., 1985).

Principal Component Analysis.

Principal components based on correlation matrix, and Euclidian distances were calculated using PAST software (Hammer *et al*., 2001). One of the major reasons that analyses of principal components on correlation matrix was done to standardize each variate (by subtracting its mean and dividing by its standard deviation), which is useful as the parameters considered in this study did not share a common scale of measurement. Principal components having Eigen values >1 were considered as significant and presented in the results.

RESULTS

Morphological Characteristics

All genotypes showed wide ranges of variation for most of the morphological traits studied; except plant growth type, leaf glands, anther filament colour, anther connective tip gland and anthocyanin colouration of capsule (Table 3.2). All tested genotypes consisted of the plants with shattering and indeterminate growth habits. There was only white anther filament colour and absence of leaf glands and anther connective tip gland. Most of the genotypes had one flower per axil (95.64%). Of the 300 sesame genotypes, only 13 that were introduced from one of the African countries, Egypt had three flowers per axil. There were large variations in stem, leaf, calyx, and capsule and corolla hairiness among the accessions (Table 3.2). Most of the genotypes showed glabrous (hair absent) and weak or sparse hairiness; and short and straight shape of hair of stem, petiole and capsule. Strong or profuse hair was observed in a few genotypes (Table 3.2); on the other hand, uniqueness of hairiness and shape of hair of stem, petiole and capsule were recorded on certain genotypes that were introduced from Egypt. There were large variations in sesame branching across all genotypes (Table 3.2), with 88.09% for Ternate, 1.36% for Alternate, 3.4% for Opposite and 7.48% as mixed branching patterns. All 21 genotypes that recorded mixed branching pattern were introduced from different African countries; and out of 21 germplasm, 13 were from Egypt alone. Some genotypes had Tetracarpellate capsule structure. (Table 3.2).

Table 0.2. Predominant morphological characters for the sesame genotypes used in a study in Ethiopia

Plant growth type	100% Indeterminate				
Plant growth habit	0.3% Prostrate	0.3% Semi-erect	99.32% erect		
Root system	Shallow fibrous	94.01% Deep thin taproot	5.63% Tuberous thick taproot		
Main stem color	0.36% Green	50% Yellow	1.07% Purplish green	48.20 Purple	0.36% Other
Stem hairiness	0.35% Glabrous (hair absent)	96.52% Weak or sparse	2.09% Medium	1.05% Strong or profuse	
Shape of hair	97.19% Short and straight	2.10% Medium and straight	0.7% Long and bent		
Stem shape in cross section	0.7% Round	99.3% Square			
Stem fasciation	99.3% Absent	0.7% Present			
Stem branching	3.4% Opposite	1.36% Alternate	88.09% Ternate	7.48% Mixed	
Branching pattern	4.60% Non-branching	16.31% Basal branching	31.91% Top branching	47.16% Other	
Leaf color	40.88% Green	18.24% Green with yellowish cast	40.20% Green with blue-gray cast	0.67% Green with purple cast	
Leaf hairiness	82.37% Glabrous (hair absent)	17.62% Weak or sparse			
Leaf arrangement	4.73% Opposite	18.92% Alternate	8.11% Ternate	68.24% Mixed	
Leaf shape (middle)	Linear	31.9% Lanceolate	0.34% Elliptic	67.69% Ovate	
Leaf shape (Top)	30.45% Linear	69.55% Lanceolate	Elliptic	Ovate	
Basal leaf profile	13.17% Flat	4.73% Cup shaped (concave)	82.09% Reverse cup shaped (convex)		
Basal leaf margin	5.76% Entire	93.22% Serrate	1.02% Dentate		
Lobe incision of basal leaf	6.84% Absent (leaf entire)	2.74% Weak	26.71% Medium	63.7% Strong (three or more lobes)	
Leaf glands	100% Absent				
Leaf angle to main stem	16.55% Acute (<90°)	83.11% Horizontal (=90°)			
Petiole color	21.33% Green	33.92% Greenish	40.56% Purple	4.19% Pink	

		purple			
Petiole hairiness	2.09% Glabrous (hair absent)	96.15% Weak or sparse	1.39% Medium	0.35% Strong or profuse	
Shape of petiole hair	98.21% Short and straight	1.43% Medium and straight	0.36% Long and bent		
Number of flowers per leaf axil	95.64% One	4.36% More than one			
Extra-floral nectary development	0.34% Rudimentary	59.12% Small	39.18% Medium	1.35% Large	
Extra-floral nectar color	0.34% Light yellow	99.66% Yellow			
Calyx tip color	3.31% Green	96.69% Purple			
Calyx hairiness	0.83% Glabrous (hair absent)	97.93% Weak or sparse	1.24% Medium		
Shape of calyx hair	99.58% Short and straight	0.42% Medium and straight			
Corolla hairiness	35.37% Weak or sparse	37.75%Medium	26.87% Strong or profuse		
Shape of corolla hair	38.09% Short and straight	40.47% Medium and straight	21.43% Long and bent		
Exterior corolla color	6.78% White	70.85% White with pink shading	22.37% White with deep pink shading		
Interior corolla color	13.65% White	60.75% White with pink shading	25.59% White with deep pink shading		
Corolla interior pigmentation	4.41% Absent	31.86% Pigmented throughout	49.83% Pigmentation along the lip region of corolla tube	4.40% Pigmentation in the supra foveolate region	9.49% Pigmentation in the infra foveolate region

Lower lip color	16.27% Colorless	83.73% Colored	
Absence/presence of foveolar	99.66% Absent	0.34% Present	
Anther filament color	100% White		
Anther connective tip gland	100% Absent		
Style length	1% Short (stigma terminating below the position of anthers)	96.99% Medium (stigma position at anther's level)	2% Long (stigma protruding outside the position of anthers)
Number of carpels per capsule	98.99% Bicarpellate	0.66% tricarpellate	0.33% Tetracarpellate
Bicarpellate capsule shape	24.49% Narrow oblong	75.51% Broad oblong	
Capsule arrangement	95.65% Monocapsular	4.35% Multicapsular	
Capsule hairiness	97.32% Weak or sparse	2.67% Medium	
Shape of capsule hair	96.98% Short and straight	3.02% Medium and straight	
Anthocyanin coloration of capsule	100% Present		
Color of dry capsules	29.89% Straw/yellow	69.56% Brown/tan	0.54% Purple
Capsule dehiscence at ripening	15.05% Partially shattering	84.95% Completely shattering	
Type of capsule beak	99.33% Curved	0.67% Cleft	
Thickness of capsule mesocarp	98.91% Thin	1.09% Thick	

Yield and yield components

There were significant differences (P< 0.05) among the study genotypes for yield and yield components; with the exception of number of secondary branches, petiole length at middle leaf, mean capsule width and mean capsule thickness (Table 3.3). The maximum values for days to flower initiation, days to 50% flowering and days to maturity were 72, 82 and 128 days, respectively. The highest seed yield was found in 9026 genotypes collected from Benshangul-Gumz Region. The average seed yield was 0.52 t ha-1 with a range of 0.034 and 1.24 t ha-1. The mean and range values of all quantitative characters are presented in Table 3.3.

Table 0.3. Range of variation and F- value of analysis of variance for quantitative characters in sesame genotypes

Character	Min	Max	Mean	Range	F Value
PTH	47.36	181.32	136.81	133.96	2.56***
PBR	0.12	6.50	3.50	6.38	2.16***
SBR	0	2.2	0.37	2.2	1
LBL	5.51	13.72	9.91	8.21	1.34**
WBL	2.51	9.52	6.76	7.01	1.54***
LML	5.49	12.68	9.30	7.19	1.84***
WML	2.00	6.44	3.53	4.44	1.42***
LTL	3.67	7.25	5.17	3.58	1.36***
WTL	0.46	7.25	0.72	6.79	1.31**
PLBL	2.84	8.44	5.89	5.6	1.46***
PLML	0.95	4.67	2.67	3.72	1.04
PLTL	0.24	1.44	0.47	1.2	1.89***
DFI	30	72	49.28	42	1.37***
DF	35	82	53.76	47	1.35**
NCPP	7.14	57.06	34.33	49.92	2.29***
CAPL	2.01	3.81	2.83	1.8	2.47***
CW	0.55	1.01	0.71	0.46	1.15
CT	0.33	0.74	0.50	0.41	1.06
SPC	48.44	82.40	67.36	33.96	1.4***
TSW	1.66	3.33	2.25	1.67	2.32***
DM	79	128	103.57	49	1.73***
PBZ	20.56	83.46	45.24	62.9	2.67***
YLD	0.034	1.24	0.520	1.204	3.95***
BBL	8.64	69.10	21.30	60.46	1.54***

PTH = plant height in centimeter; PBR = primary branch; LBL = length of basal leaf; WBL = width of basal leaf ; LML = length of middle leaf; WML = width of middle leaf ; LTL= length of top leaf ; WTL= width of top leaf; PLBL = petiole length of basal leaf; PLTL = petiole length of top leaf ; DFI = days to flower initiation; DF = days to 50% flowering ; COL= corolla length; LLL = length of longest lip; NCPP = Number of capsules per plant; CAPL = capsule length; SPC = seeds per capsule; TSW = 1000 seed weight in gram; DM = days to maturity; PBZ = Pod bearing zone; YLD = yield in tonnes per hectare; BBL = bacterial blight reaction. ** Significant at P<0.01; *** Significant at P <0.001

Mean, Coefficients of variation, heritability and genetic advance for quantitative traits in sesame germplasm

Different parameters such as the mean performance, phenotypic variance, genotypic variance, phenotypic coefficients of variation (PCV), genotypic coefficients of variation (GCV), heritability estimates, and predicted genetic advance over mean for all the characters are presented in Table 3.4 and some parameters are also presented in **Figure**

3.1. According to Sivasubramanian, (1973) PCV and GCV values greater than 20% are regarded as High, whereas a value less than and values between 10% and 20% considered low and medium, respectively. The traits;- number of primary branches per plant, Petiole length of top leaf, seed yield per hectare, and bacterial blight showed high PCV and GCV estimates while Width of top leaf and number of capsules per plant showed high PCV and moderate GCV.

Plant height, Days to flower initiation, days to 50% flowering, 1000 seed weight, and pod bearing zone showed moderate PCV and GCV. While the width of basal and middle leaf, length of middle leaf, Petiole length of basal leaf, and capsule length showed moderate PCV and low GCV. The remaining traits such as length of basal leaf, length of top leaf, Corolla length, Length of the longest lip, Seeds per capsule, and days to maturity recorded low PCV and GCV

Table 0.4. Estimation of variance component

Traits	Mean	Env	rep	ENV	trt	ENV.trt	Error	σ2g	√ σ2g	σ2P	√ σ2p	σ2e	√ σ2e	GCV (%)	PCV (%)	H2b (%)	k=2.06	GA	GAM
PTH	131.32	3	2	238.30	408.80	40.30	159.40	408.80	20.22	448.80	21.18	159.40	12.63	15.40	16.13	91.09	2.06	39.75	30.27
PBR	3.23	3	2	0.39	0.47	0.14	0.96	0.47	0.69	0.68	0.82	0.96	0.98	21.29	25.55	69.41	2.06	1.18	36.53
SBR	0.32	3	2	0.17	0.01	0.16	0.26	0.01	0.10	0.11	0.33	0.26	0.51	31.18	102.00	9.34	2.06	0.06	19.63
LBL	10.26	3	2	10.98	0.11	1.12	2.55	0.11	0.33	0.90	0.95	2.55	1.60	3.20	9.27	11.94	2.06	0.23	2.28
WBL	6.71	3	2	2.24	0.19	0.79	2.15	0.19	0.44	0.81	0.90	2.15	1.47	6.51	13.44	23.49	2.06	0.44	6.51
LML	9.53	3	2	95.87	0.35	1.42	3.23	0.35	0.59	1.36	1.17	3.23	1.80	6.21	12.25	25.70	2.06	0.62	6.49
WML	3.83	3	2	0.71	0.12	0.18	2.19	0.12	0.35	0.55	0.74	2.19	1.48	9.16	19.33	22.44	2.06	0.34	8.93
LTL	5.29	3	2	-0.02	0.13	-0.01	0.69	0.13	0.36	0.24	0.49	0.69	0.83	6.88	9.34	54.27	2.06	0.55	10.44
WTL	0.83	3	2	0.01	0.01	0.02	0.17	0.01	0.11	0.05	0.22	0.17	0.42	12.81	26.10	24.10	2.06	0.11	12.96
PLBL	5.84	3	2	18.17	0.11	0.77	2.29	0.11	0.33	0.75	0.87	2.29	1.51	5.65	14.81	14.55	2.06	0.26	4.44
PLTL	0.54	3	2	0.01	0.02	0.01	0.05	0.02	0.13	0.03	0.16	0.05	0.22	23.91	30.57	61.19	2.06	0.21	38.53
DFI	47.34	3	2	2.17	30.43	3.34	2.92	30.43	5.52	32.03	5.66	2.92	1.71	11.65	11.96	95.01	2.06	11.08	23.40
DF	51.69	3	2	2.70	37.01	2.83	3.08	37.01	6.08	38.46	6.20	3.08	1.75	11.77	12.00	96.22	2.06	12.29	23.78
COL	14.11	3	2	354.04	0.02	0.49	1.90	0.02	0.14	0.50	0.71	1.90	1.38	1.00	5.00	4.01	2.06	0.06	0.41
LLL	8.48	3	2	121.24	0.04	0.28	1.44	0.04	0.19	0.37	0.61	1.44	1.20	2.30	7.19	10.23	2.06	0.13	1.51
NCPP	33.97	3	2	144.94	38.83	121.45	88.32	38.83	6.23	94.03	9.70	88.32	9.40	18.34	28.54	41.29	2.06	8.25	24.28
CAPL	2.86	3	2	0.00	0.08	0.02	0.06	0.08	0.28	0.10	0.31	0.06	0.24	9.84	10.84	82.49	2.06	0.53	18.42
CW	0.72	3	2	0.00	0.00	0.00	0.01	0.00	0.04	0.00	0.05	0.01	0.08	5.47	7.48	53.38	2.06	0.06	8.23
CT	0.51	3	2	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.04	0.00	0.07	5.18	7.53	47.35	2.06	0.04	7.35
SPC	68.43	3	2	5.30	15.32	8.49	46.60	15.32	3.91	25.92	5.09	46.60	6.83	5.72	7.44	59.11	2.06	6.20	9.06
TSW	2.32	3	2	0.02	0.06	0.05	0.04	0.06	0.25	0.09	0.29	0.04	0.21	10.89	12.67	73.90	2.06	0.45	19.29
DM	99.81	3	2	20.19	73.60	8.21	26.14	73.60	8.58	80.69	8.98	26.14	5.11	8.59	9.00	91.21	2.06	16.88	16.91
PBZ	46.66	3	2	-4027.72	55.74	30.48	66.41	55.74	7.47	76.97	8.77	66.41	8.15	16.00	18.80	72.42	2.06	13.09	28.05
YLD	544.48	3	2	8338.0	45747.0	45543.0	21234.0	45747.0	213.9	64467.0	253.9	21234.0	145.7	39.3	46.6	71.0	2.1	371.2	68.17
BBL	22.22	3	2	4.84	60.70	0.24	22.32	60.70	7.79	64.50	8.03	22.32	4.72	35.06	36.14	94.11	2.06	15.57	70.06

PTH= plant height; PBR= primary branch; LBL= length of basal leaf; WBL=width of basal leaf ; LML= length of middle leaf; WML= width of middle leaf ; LTL=length of top leaf ; WTL= width of top leaf; PLBL= petiole length of basal leaf; PLTL= petiole length of top leaf ; DFI= days to flower initiation; DF= days to 50% flowering ; COL= corolla length; LLL=length of longest lip; NCPP = Number of capsules per plant; CAPL= capsule length; SPC= seeds per capsule; TSW= 1000 seed weight; DM= days to maturity; PBZ= Pod bearing zone; YLD= yield; BBL= bacterial blight reaction.

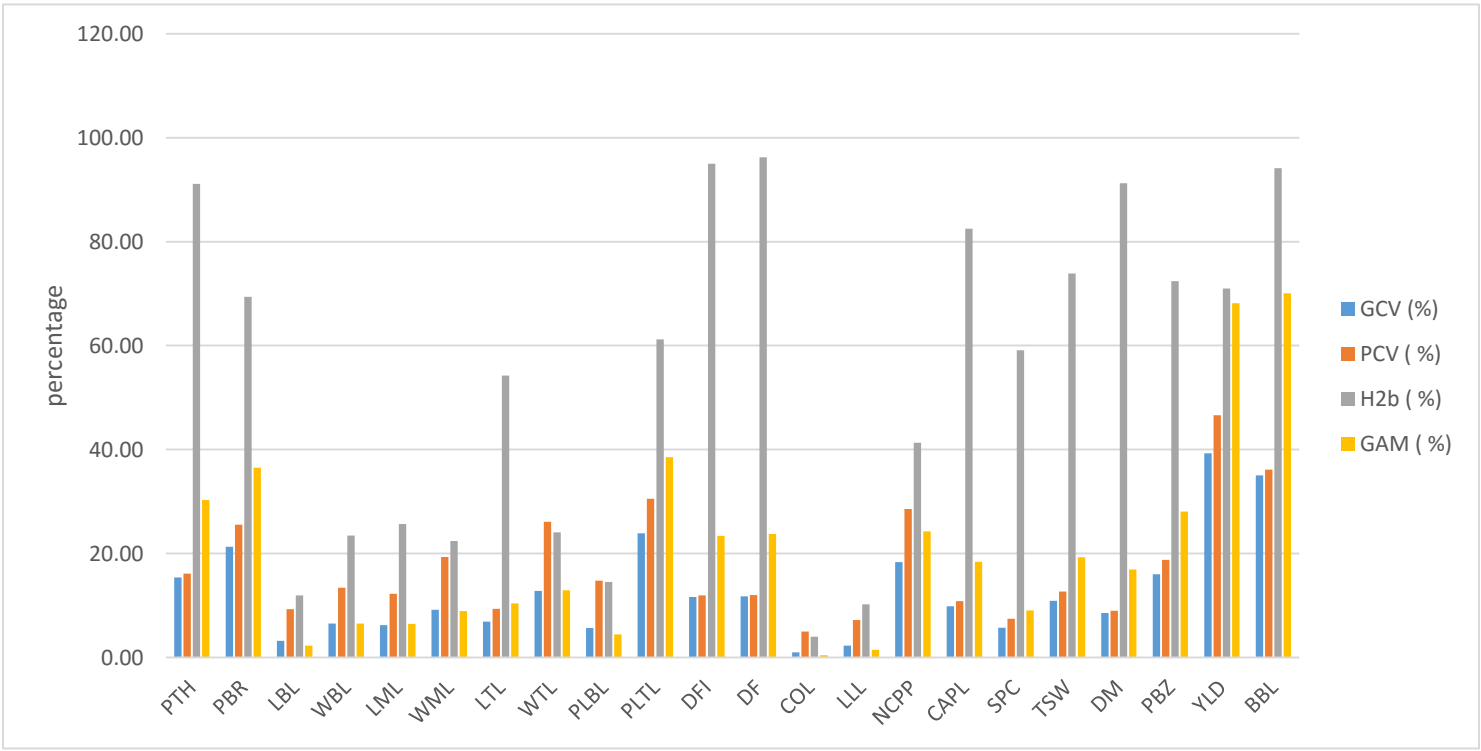


Figure 0-1. Parameters estimation of variability for yield and yield components in sesame

PTH= plant height; PBR= primary branch; LBL= length of basal leaf; WBL=width of basal leaf ; LML= length of middle leaf; WML= width of middle leaf ; LTL=length of top leaf ; WTL= width of top leaf; PLBL= petiole length of basal leaf; PLTL= petiole length of top leaf ; DFI= days to flower initiation; DF= days to 50% flowering ; COL= corolla length; LLL=length of longest lip; NCPP = Number of capsules per plant; CAPL= capsule length; SPC= seeds per capsule; TSW= 1000 seed weight; DM= days to maturity; PBZ= Pod bearing zone; YLD= yield; BBL= bacterial blight reaction.

Heritability and Genetic Advance

According to Robinson (1949), heritability values greater than 60% are very high, values from 30-60% are medium and values less than 30% are low. The heritability in broad sense values ranged from **4.01** (corolla length) to **96.22** (days to 50% flowering). High heritability estimates were observed for plant height, primary branch, Petiole length of top leaf, days to flower initiation, days to 50% flowering, capsule length, 1000 seed weight, days to maturity, pod bearing zone, seed yield, and bacterial blight reaction (Table 3.4). Heritability estimates obtained from traits; - length of top leaf, number of capsules per plant, capsules width, capsules thickness, and seeds per capsule were moderate, while estimated from the length of basal leaf, the width of basal leaf, length of middle leaf, the width of middle and top leaf, petiole length of basal leaf, corolla length and length of the longest lip heritability were low**.**

According to Johnson *et al*. (1955) **genetic advance** as percentage of mean (GAM); values from 0-10% are low, 10-20% are moderate and 20% and above are high. The range of genetic advance as a percent of mean ranged from 1.5% for length of the longest lip to 70.06% for bacterial blight. High heritability combined with high genetic advance (as per cent of the mean) was observed for plant height, primary branch, petiole length of top leaf, days to flower initiation, days to 50% flowering, pod bearing zone, seed yield per plant and bacterial blight reaction (Table 3.4).

Table 0.5. Phenotypic Correlation Coefficients of the 20 parameter sesame germplasms / Pr > |r|

TRAITS	PTH	PBR	LBL	WBL	LML	WML	LTL	WTL	PLBL	PLTL	DFI	DF	NCPP	CAPL	SPC	TSW	DM	PBZ	YLD
PBR	0.50**																		
LBL	0.35**	0.13*																	
WBL	0.44**	0.30**	0.83**																
LML	0.49**	0.08 ns	0.52**	0.48**															
WML	0.25**	0.10 ns	0.42**	0.48**	0.56**														
LTL	0.05 ns	-0.27**	0.31**	0.16**	0.47**	0.25**													
WTL	-0.44**	-0.36**	-0.18**	-0.28**	-0.09 ns	0.00 ns	0.32**												
PLBL	0.45**	0.28**	0.78**	0.84**	0.53**	0.37**	0.17**	-0.23**											
PLTL	-0.60**	-0.52**	-0.30**	-0.44**	-0.22**	-0.14*	0.21**	0.63**	-0.38**										
DFI	0.59**	0.56**	0.25**	0.38**	0.04 ns	0.26**	-0.25**	-0.46**	0.21**	-0.57**									
DF	0.60**	0.59**	0.19**	0.35**	0.01 ns	0.22**	-0.28**	-0.46**	0.20**	-0.58**	0.94**								
NCPP	0.60**	0.35**	0.30**	0.34**	0.48**	0.21**	0.03 ns	-0.20**	0.44**	-0.39**	0.18**	0.18**							
CAPL	0.05 ns	-0.31**	0.46**	0.30**	0.38**	0.26**	0.46**	-0.03 ns	0.26**	-0.07 ns	-0.14*	-0.19**	0.06 ns						
SPC	0.11 ns	-0.12 ns	0.23**	0.25**	0.28**	0.25**	0.21**	-0.14*	0.23**	-0.13*	-0.04 ns	-0.04 ns	0.22**	0.45**					
TSW	0.23**	0.08 ns	0.40**	0.41**	0.31**	0.22**	0.24**	-0.10 ns	0.35**	-0.27**	0.12*	0.07 ns	0.03 ns	0.30**	0.04 ns				
DM	0.65**	0.63**	0.20**	0.38**	0.08 ns	0.23**	-0.28**	-0.46**	0.27**	-0.59**	0.87**	0.87**	0.31**	-0.24**	0.02 ns	0.14*			
PBZ	0.52**	-0.07 ns	0.31**	0.23**	0.58**	0.12*	0.30**	-0.09 ns	0.37**	-0.17**	-0.17**	-0.19**	0.67**	0.25**	0.26**	0.10 ns	-0.06 ns		
YLD	0.50**	0.04 ns	0.31**	0.30**	0.52**	0.20**	0.19**	-0.11 ns	0.44**	-0.22**	-0.07 ns	-0.09 ns	0.67**	0.12*	0.22**	0.18**	0.07 ns	0.73**	
BBL	-0.59**	-0.23**	-0.38**	-0.48**	-0.42**	-0.31**	0.01 ns	0.31**	-0.51**	0.48**	-0.32**	-0.31**	-0.69**	-0.07 ns	-0.29**	-0.09 ns	-0.43**	-0.59**	-0.66**

PTH= plant height; PBR= primary branch; LBL= length of basal leaf; WBL=width of basal leaf ; LML= length of middle leaf; WML= width of middle leaf ; LTL=length of top leaf ; WTL= width of top leaf; PLBL= petiole length of basal leaf; PLTL= petiole length of top leaf ; DFI= days to flower initiation; DF= days to 50% flowering ; NCPP = Number of capsules per plant; CAPL= capsule length; SPC= seeds per capsule; TSW= 1000 seed weight; DM= days to maturity; PBZ= Pod bearing zone; YLD= yield; BBL= bacterial blight reaction. * Significant at $P \leq 0.05$; ** Significant at $P \leq 0.01$ and ns non-significance

Table 0.6. Genotypic correlations coefficients of the 20 parameters sesame germplasms / Pr > |r|

TRAITS	PTH	PBR	LBL	WBL	LML	WML	LTL	WTL	PLBL	PLTL	DFI	DF	NCPP	CAPL	SPC	TSW	DM	PBZ	YLD
PBR	0.51**																		
LBL	0.4**	-0.75**																	
WBL	0.62**	0.33**	1.00**																
LML	0.53**	-0.46**	1.00**	0.75**															
WML	0.05^{ns}	-0.08^{ns}	1.00**	0.88**	0.28**														
LTL	-0.09^{ns}	-0.75**	1.00**	0.39**	0.93**	0.16**													
WTL	-1**	-0.72**	0.65**	-0.12*	0.35**	-0.19**	0.57**												
PLBL	0.8**	0.2**	0.58**	0.93**	1.00**	0.06^{ns}	0.37**	0.36**											
PLTL	-0.77**	-0.75**	-0.64**	-0.83**	-0.48**	-0.82**	0.23**	0.90**	-0.76**										
DFI	0.69**	0.79**	0.79**	0.83**	-0.20**	0.35**	-0.48**	-1.00**	0.47**	-0.83**									
DF	0.72**	0.86**	0.97**	1.00**	-0.21**	0.35**	-0.53**	-1.00**	0.72**	-0.84**	1.00**								
NCPP	0.73**	0.26**	0.00^{ns}	0.20**	0.50**	-0.48**	-0.21**	-0.41**	0.81**	-0.60**	0.10^{ns}	0.25**							
CAPL	-0.04^{ns}	-0.55**	1.00**	0.77**	0.67**	0.66**	0.79**	0.02^{ns}	0.69**	0.05^{ns}	-0.17**	-0.18**	-0.30**						
SPC	-0.07^{ns}	-0.27**	0.82**	0.71**	0.53**	0.64**	0.38**	-0.41**	0.55**	-0.25**	-0.10^{ns}	-0.07^{ns}	-0.10^{ns}	0.54**					
TSW	0.16**	-0.04^{ns}	1.00**	1.00**	0.67**	0.49**	0.54**	0.18**	1.00**	-0.13*	0.05^{ns}	-0.03^{ns}	-0.43**	0.36**	0.04^{ns}				
DM	0.73**	0.92**	0.02^{ns}	0.64**	-0.17**	0.41**	-0.59**	-1.00**	0.39**	-0.75**	0.96**	1.00**	0.33**	-0.34**	0.01^{ns}	0.01^{ns}			
PBZ	0.45**	-0.33**	0.57**	0.29**	0.77**	-0.32**	0.33**	0.10^{ns}	0.79**	-0.09^{ns}	-0.28**	-0.31**	0.81**	0.14*	0.11^{ns}	-0.09^{ns}	-0.16**		
YLD	0.46**	-0.15**	0.53**	0.30**	0.81**	-0.14*	0.16**	-0.05^{ns}	0.90**	-0.14*	-0.23**	-0.22**	0.90**	-0.01^{ns}	0.06^{ns}	0.04^{ns}	-0.03^{ns}	0.86**	
BBL	-0.66**	-0.34**	-1.00**	-1.00**	-1.00**	-1.00**	-0.07^{ns}	0.56**	-1.00**	0.70**	-0.35**	-0.34**	-1.00**	-0.10^{ns}	-0.53**	-0.21**	-0.4**8	-0.77**	-0.89**

PTH= plant height; PBR= primary branch; LBL= length of basal leaf; WBL=width of basal leaf ; LML= length of middle leaf; WML= width of middle leaf ; LTL=length of top leaf ; WTL= width of top leaf; PLBL= petiole length of basal leaf; PLTL= petiole length of top leaf ; DFI= days to flower initiation; DF= days to 50% flowering ; NCPP = Number of capsules per plant; CAPL= capsule length; SPC= seeds per capsule; TSW= 1000 seed weight; DM= days to maturity; PBZ= Pod bearing zone; YLD= yield; BBL= bacterial blight reaction. * Significant at $P \leq 0.05$; ** Significant at $P \leq 0.01$ and ns non-significance

The Correlation coefficient between yield and yield component characters in sesame germplasm

The genotypic, and phenotypic correlation coefficients for yield and yield-related traits are presented in Tables 3.5 and 3.6.

Correlations of grain yield with other traits

In the present study, the result of phenotypic correlation analysis revealed that grain yield had a significant and positive correlation with plant height, a length of basal leaf, a width of basal leaf, a length of middle leaf, a width of middle leaf, a length of top leaf, petiole length of basal leaf, number of capsule per plant, capsule length, seed per capsule, 1000 seed weight, and pod bearing zone, and had a negative, and significant correlation with petiole length of top leaf and bacterial blight reaction.

The result of genotypic correlation analysis revealed that grain yield had a significant and positive genotypic correlation with plant height, a length of basal leaf, a width of basal leaf, a length of middle leaf, a length of top leaf, petiole length of basal leaf, number of capsule per plant, pod bearing zone, and positively but not significantly correlated with seed per capsule, and 1000 seed weight, and had a negative significant correlation with primary branch, a width of middle leaf, a width of top leaf , petiole length of top leaf, days to flower initiation, days to 50% flowering, capsule length, and bacterial blight reaction.

At genotypic, and phenotypic levels plant height had a significant and positive correlation with Primary branch, a length of basal leaf, a width of basal leaf, length of middle leaf, width of middle leaf, petiole length of basal leaf, days to flower initiation, days to 50%

flowering, number of capsule per plant, 1000 seed weight, days to maturity, pod bearing zone, and a significant negative correlation with a width of top leaf, petiole length of basal leaf, seed per capsule, and bacterial blight reaction.

The number of capsules per plant had a significant and positive correlation with plant height, primary branch, width of basal leaf, length of middle leaf, petiole length of basal leaf, days to 50% flowering, days to maturity, and pod bearing zone but had a significant, and negative correlation with width of top leaf, petiole length of top leaf, and bacterial blight reaction.

Pod-bearing zone had a significant and positive correlation with plant height, length of basal leaf, width of basal leaf, length of middle leaf, length of top leaf, petiole length of basal leaf, number of capsule per plant, and capsule length, and had a significant, and negative correlation with primary branch, days to flower initiation, days to 50% flowering, days to maturity, and bacterial blight reaction.

Cluster analysis.

The 300 genotypes were grouped into 6 clusters based on 22 sesame parameters (Figure 3.2). Five of the six clusters comprised of more than one germplasm; whereas one cluster was singleton (containing single accession). Cluster 1 contained 115 genotypes (38.3%) out of 300; and was followed by clusters II, V, IV and III containing 107, 36, 31 and 10 accessions, respectively. Whereas the cluster VI contained only one genotype (0.33%). Cluster I constituted the largest number of germplasm, mainly collected from the different regions of Ethiopia 73(64.34 %); namely Amhara (n=36), Benshangul-Gumz (11), Oromia (4), Tigray (22) and 1 improved variety, 35 (30.43%). Those registered as

introduced from different African countries included Burkina Faso (8), Egypt (10), Kenya (1), Somalia (1), Sudan (1), Zambia (1) and Zimbabwe (13). The remaining 6 genotypes were introduced from different Asian countries. Most genotypes of Cluster II were collected from the different regions of Ethiopia 93(86%); - Amhara (n=9), Benshangul-Gumz (21), Oromia (35), SNNP (2), Tigray (26) and 10 improved varieties. Only 4 genotypes were introduced from three African countries; namely Burkina Faso (2), Egypt (1) and Kenya (1) belonged to Cluster II. All genotypes of Cluster III were collected from three different regions of Ethiopia; namely Amhara (n=4), Oromia (3), Tigray (2), and 1 improved variety. Most genotypes of Cluster IV were collected from different countries of Africa 22 (70.96%) (Burkina Faso (6), Egypt (12), Sudan (1), Zambia (2), and Zimbabwe (2); and 9 genotypes were collected from three regions of Ethiopia (Oromia (4), SNNP (1) and Tigray (4)).

Cluster V consisted of genotypes mainly collected from different regions of Ethiopia 29 (80.55 (Amhara (n=7), Benshangul-Gumz (5), Oromia (6) and Tigray (6); and 4 were improved varieties. Only 7 genotypes were introduced from three African countries; namely Burkina Faso (2), Egypt (4) and Kenya (1). The remaining one genotypes was from Israel. Cluster VI consist only one genotype collected from Benshangul-Gumz from Ethiopia (Table 3.7).

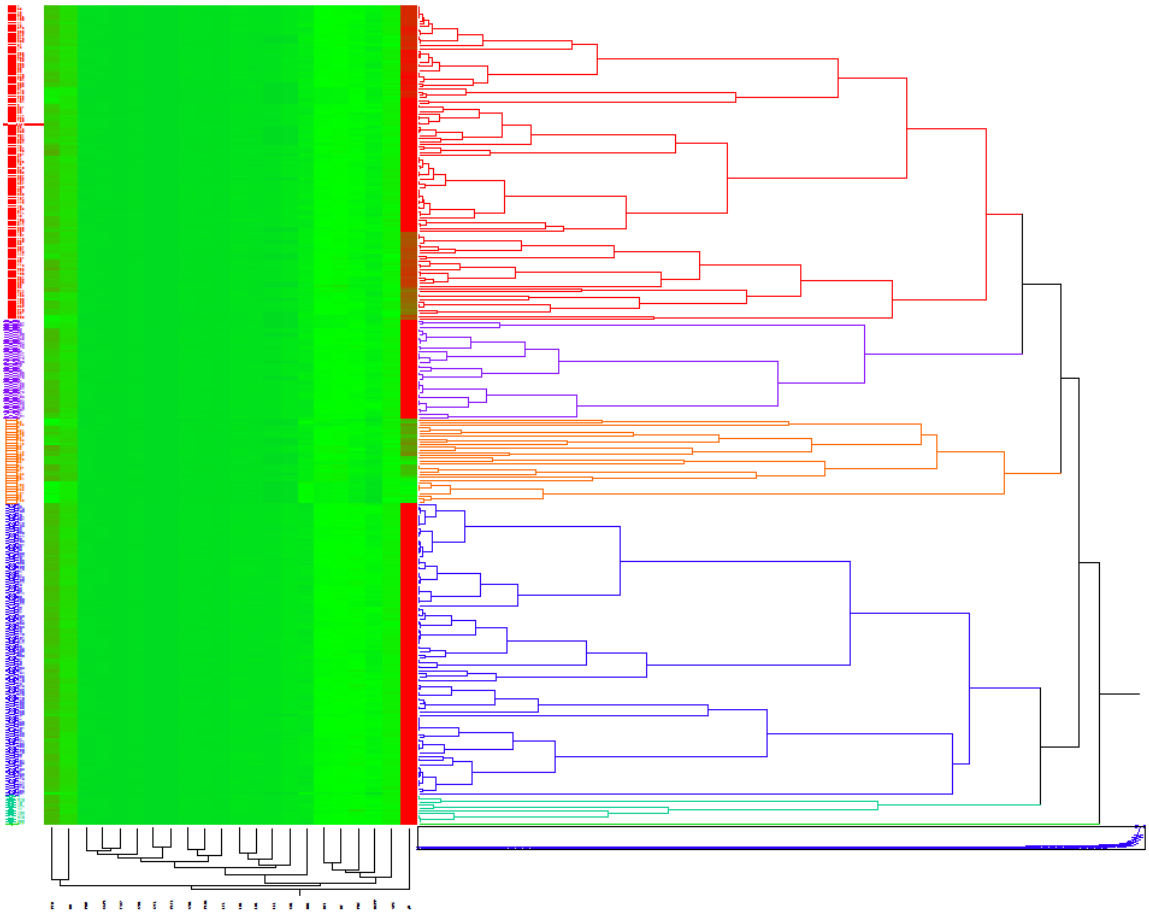


Figure 0-2. Dendrogram of 300 sesame genotypes based on average linkage hierarchical cluster analysis between groups of sesame in Ethiopia

Table 0.7 Grouping of the 300 sesame genotypes into different diversity clusters

Cluster number	Number & % of accessions	Accessions grouped	Origins
I	115 (38.3%)	9015, 9691, 28300, 235905, 241338, Acc 051-02-sel 1 (2), Adebay, ACC-203-030, Manasiibuu -60 (mandil), 238269, 241341, 9019, 28306, 28324, 203600, 205197, 241319, Mayweyni, Maydelye-04, ACC-203-336-2, GOBYO-82-3, B-02, 241308, 235769, BOUNJA FILWAHA-2, BAHA ZEYT, 241305, 212992, 205194, 205188.2, 9020, 9027, 215816, ACC-0065, ACC-NO-026, Baunja fiyel kolet, Ass-acc-3, Hirhir nugar 1st set-2, 9687, 18759, 241342, ACC-EW-011(2), ACC-NO-041, 207957, Mazegabu-01, 205195, 203104, 28328, 9693, 9021, ACC-205-184, Wuhdet, 241322, 234016, 17713, 9695, ACC-EWO12(16), Dansh-01, Bawal-02, TATI, 241332, 241318, 235404, 205190, 9024, 17712, 28308, 28320, 28327, 205191, 216733, 237523, 241317, 241344, ADI, Banat-04, Maykeyh-01, ACC-EW-017(4), ACC ≠031, 9022, 28304, 28322, 203640, 9017, 241330, Turka, Maydelye-02, ACC-202-889, 241343, 28318, 205185, 28302, 28311, 241314, 203608, 28303, 28309, 17693, 28323, Maydelye-03, 241315, 205186, 28321, 9686, 23546, 28307, 28319, 234017, ACC-BG001, ACC-EW-006, 241340, 28301, 28317, 28325, China FAO (ACC - 68 -542), 205188, Maydelye-01, 9696, 28305, 28330, 203595, ENSHANGUL- 1, 207758, 9025, Ass-acc-64, 23547, 227890, 227936, 227891, USR - 82 # 171 NS, 227875, 205185.2, 227893, 227944, 227942, 205197.2, 227883, 227939, X - 30/40 # 403, 227945, 227876, 205189, 227874, 205190.2, 227941, 205186.2, 227888, 203614, 227938, 227879, Ying White – 2, 203640.2, 227884, HB-38FAM-2BAR GRAY, 203603, 203630, 227885, 227880, 215013, 227943, 203615, 203610, 227882, 227881, BAR – 002, HB - 22 - FAM (1- 4), HB - 49 FAM - 2 – 2, Clusu - Acc – 2, JAPAN-651, SPS - SIK - # 811, SSBS - (9 - 2) -3, Acc - 203 - 336 – 2, Acc - 203 - 336 – 4, Acc - 203 – 612, Acc - 210 - 986 – 1, Acc - 210 - 991 – 4, Tmax, 203637, 216896, 203102, 203633, 9242	Amhara, Benshangul-Gumz, Gambella Oromia, Tigray, IMPROVED varieties, China, Egypt, Burkina-Faso, Israel ,Japan, Kenya, Philipins, Somalia, Sudan, Zimbabwe, Zambia
II	107 (35.6%)	17704, Lugudi-01, DICHO , NAJIJO-70(GABAA KAISAA),23555, 9692, 17703, 241334, OBSA, Selam, Mana siba-64(mandil), 23565, GULISO-71(GOSOLA), Ass-acc-128, Backer, SETIT-1, 238270, 203639, 241295, 238271, Najjoo-68	Amhara, Benshangul-Gumz, Gambella,

		(gabaa kamijaa), MAN-66(MATIL), NAJOO-73 GIIBII-27(TOLE), Bawal-03, Adebay-02, 9023, 17696, 17706, CHALESA, Banat-01, Local-90-1, Ass-acc-74, ACC-BG007, 17700, Manasbu-62(mandii), Hirhir adigoshu sel-8,BAHA NECHO, SETIT-2, 241323, 17708, 202514, Ebantuu-11 (qeloo), Ass-acc-85, HUMERA-1, 23564, Manjibu-63(mandii), MANSIBU-61(MANDII), 202513, Maygaba, MECHAL , 205196.2, 17710, 9690, 17697, 17705, 241309,241328, Bawal-01, Ass-acc-71, Ass-acc-36, Ass-acc-127, GULISO-72(GOOGOLA), 241326, 9688, 17694, 17701, 17709, 28313, 28329, 205196, DANGUR, Shelela-02, Ass-acc-25, NAJOO-69 (GABAA KAMISA), 23559, 241316, 227937, 17707, Nigari-02, 222876, 28310, 9697, Ass-acc-11, 208888, 23557, 17695, 9694, 17698, 241325, ABASENA, Lugudi-02, 9689, 17699, Ass-acc-105, 241324, HUARC-4, x-3014≠401, 23556, Man-67 (mandii), Najoo-74, 203099, 28326, 17702, Najjoo-77,Najjoo-76, 23561	Oromia, SNNP Tigray, Kenya, Egypt, Burkina-Faso
III	10 (3.3%)	28316, 241321, Shelela-01, 28315, 28314, GONDAR-1, Shelela-03, 17711, 19043, Najjoo-75	Amhara, Oromia, Tigray
IV	31(10.3%)	227890, 227936, 235905, 227875, 205185.2, ACC-203-336-2, 212992, 227944, 205188.2, 227942, 18759, 205197.2, 216896, ACC-NO-041, 227939, 203104, 227945, 203102, 227876, 205190.2, 203633, 227941, 203614, 227879, 237523, 215013, 227943, 205188, 9242, Acc - 210 - 986 – 1, 207758	Oromia, SNNP, Tigray Burkina-Faso, Egypt, Zimbabwe, Zambia, Sudan
V	36 (12%)	227891, Acc-051-02-sel 1 (2), Manasiibuu -60 (mandil), 9019, 28306, BAHA ZEYT, ACC-NO-026, Ass-acc-3, 9687, 207957, 205195, X - 30/40 # 403, 17713, 9695, Bawal-02, 17712, 205191, ADI, Banat-04, ACC-EW-017(4), ACC-202-889, 28311, 227880, 17693, Maydelye-03, 28321, 23546, 28319, 227881, ACC-EW-006, 28317, 28325, HB - 49 FAM - 2 – 2, SSBS - (9 - 2) -3, BENSHANGUL- 1, Ass-acc-64	Amhara, Benshangul-Gumz, Gambella Oromia, TIGRAY ,Burkina-Faso, Egypt , Kenya, Israel
VI	1 (0.33%)	9026	Benshangul-Gumz

Table 0.8. Cluster means for parameters in sesame genotypes used in a study of sesame germplasm conducted in Ethiopia

Parameters	CLUSTER						
	I	II	III	IV	V	VI	GM
PTH	125.6	142.48	149.8	105.1	133.47	144.32	133.46
PBR	3.35	3.17	3.26	3.13	2.99	6.08	3.66
LBL	10.31	10.63	11.24	8.37	10.31	10.52	10.23
WBL	6.76	6.96	7.35	5.40	6.76	6.61	6.64
LML	9.43	9.96	10.85	7.78	9.71	9.42	9.52
WML	3.94	3.80	3.87	3.40	3.90	4.16	3.85
LTL	5.32	5.30	5.50	4.96	5.39	5.24	5.28
WTL	0.80	0.82	0.83	0.99	0.84	0.96	0.87
PLBL	5.81	6.20	6.61	4.30	5.98	6.52	5.90
PLTL	0.51	0.52	0.57	0.69	0.57	0.43	0.55
DFI	48.21	46.54	45.07	49.38	45.77	48.00	47.16
DF	52.81	50.74	49.03	53.10	50.42	53.33	51.57
COL	13.17	16.45	16.27	8.21	14.58	16.37	14.18
LLL	7.89	9.85	10.12	4.85	8.90	10.07	8.62
NCPP	31.47	38.85	45.20	21.71	34.41	51.36	37.17
CAPL	2.94	2.85	2.95	2.63	2.86	2.38	2.77
SPC	68.92	69.36	68.14	63.30	68.74	63.80	67.04
TSW	2.33	2.30	2.45	2.17	2.43	2.62	2.38
DM	100.46	99.54	98.46	100.75	97.85	110.00	101.18
PBZ	42.79	52.95	58.27	33.62	48.32	45.48	46.91
YLD	0.393	0.771	0.973	0.132	0.569	1.24	0.679
BBL	24.34	17.08	18.10	34.46	21.73	9.14	20.81

PTH = plant height in centimeter; PBR = primary branch; LBL = length of basal leaf; WBL = width of basal leaf ; LML = length of middle leaf; WML = width of middle leaf ; LTL = length of top leaf ; WTL = width of top leaf; PLBL = petiole length of basal leaf; PLTL = petiole length of top leaf ; DFI = days to flower initiation; DF = days to 50% flowering ; COL = corolla length; LLL = length of longest lip; NCPP = Number of capsules per plant; CAPL = capsule length; SPC = seeds per capsule; TSW = 1000 seed weight; DM = days to maturity; PBZ = pod bearing zone; YLD = yield in tonnes per hectare; BBL = bacterial blight reaction

Cluster mean analysis.

The clusters for different traits indicated wide variations for all the characters considered (Table 3.8). The highest values for grain yield were recorded from Cluster VI (1.24 t ha-1) and the lowest from Custer IV (0.132 t ha-1). The next highest yield to Cluster VI

were recorded by Cluster III (0.973 t ha-1). Clusters means for all characters in sesame genotypes presented in Table 3.8.

Divergence analysis.

Pair wise generalized squared distances (D^2) among six clusters are presented in Table 3.9. There were 15 possible pair wise genetic distances between any two clusters. Among these, only 4 genetic distances (between clusters I and IV, I and V, II and III, II and V) were not significant ($P>0.05$); but the remaining genetic distances were highly significant ($P<0.01$).The maximum distance was found between clusters IV and VI (D^2 =342.56), and the distances between I, IV and V clusters with cluster VI and cluster II and III with IV were maximum and very highly significant ($P<0.01$).

Table 0.9. Intra- (bolded diagonals) and inter-cluster distance between sesame genotypes categorized into clusters

	Generalized Squared Distance to class					
From class	I	II	III	IV	V	VI
I						
II	41.14002					
III	97.63437	14.69811				
IV	19.77445	108.4294	190.8707			
V	10.05937	12.56597	48.69146	52.35292		
VI	217.9783	88.42169	50.93601	342.5608	149.602	

X^2 = 38.932 at 1% probability level and X^2 = 32.671 at 5% probability level

Principal Component Analysis (PCA).

The first five principal components were found to be significant (Eigen value greater than 1) and accounted for about 75.4% of the total variation (Table 3.10). The first PCA component explained 34.8% of the total variance; while the first and second PCA components accounted for 54.6% of the variation (Figure 3.3). Parameters that contributed relatively more with an Eigen Vector value for the first PCA were plant height, length and width of basal leaf, petiole length of basal leaf, number of capsule per plant and grain yield. Most of the variations attributed to the second PCA were contributed by length of middle and top leaf, petiole length of top leaf and pod bearing zone. The third PCA explained about 9.5% of the variation traits, such as width of middle leaf, capsule length and bacterial blight reaction contributed much of its variation (Figure 3.3). The fourth PCA were explained by corolla length, length of the longest lip and seed per capsule; while the fifth PCA explained by 1000 seed weight.

Table 0.10. Eigen vectors, explained variance and Eigen values of the first significant Principal components for parameters of sesame genotypes

Parameter	Eigen Vectors				
	PCA1	PCA2	PCA3	PCA4	PCA5
Plant height	0.304	-0.09	-0.124	0.002	0.068
Primary branch	0.147	-0.292	-0.048	-0.251	0.123
Length of basal leaf	0.296	0.117	0.209	-0.001	0.108
Width of basal leaf	0.294	0.028	0.229	-0.121	0.074
Length of middle leaf	0.222	0.241	0.09	-0.281	-0.06
Width of middle leaf	0.121	0.069	0.296	-0.407	-0.424
Length of top leaf	0.035	0.304	0.217	-0.179	-0.009
Width of top leaf	-0.19	0.201	0.035	-0.316	-0.055
Petiole length of basal leaf	0.276	0.082	0.107	-0.206	0.156
Petiole length of top leaf	-0.236	0.21	0.011	-0.153	-0.136
Days to flower initiation	0.156	-0.368	0.15	-0.027	-0.171
Days to 50% flowering	0.162	-0.384	0.122	-0.068	-0.121
Corolla length	0.277	0.07	0.006	0.354	0.012
Length of longest lip	0.278	0.069	-0.013	0.351	0.001
Number of capsules per plant	0.222	0.07	-0.399	-0.18	-0.096
Capsule length	0.092	0.261	0.331	0.265	-0.107
Seeds per capsule	0.143	0.152	0.099	0.318	-0.521
1000 seed weight	0.118	0.073	0.323	-0.005	0.589
Days to maturity	0.173	-0.371	0.018	-0.101	-0.083
Pod bearing zone	0.178	0.269	-0.347	-0.054	0.055
Grain yield	0.209	0.205	-0.351	-0.092	0.093
Bacterial blight	-0.278	-0.017	0.258	0.045	0.18
Eigenvalue	7.672	4.3453	2.1004	1.3935	1.0867
Explained variance %	34.9	19.8	9.5	6.3	4.9
Cumulative variance	34.9	54.6	64.2	70.5	75.4

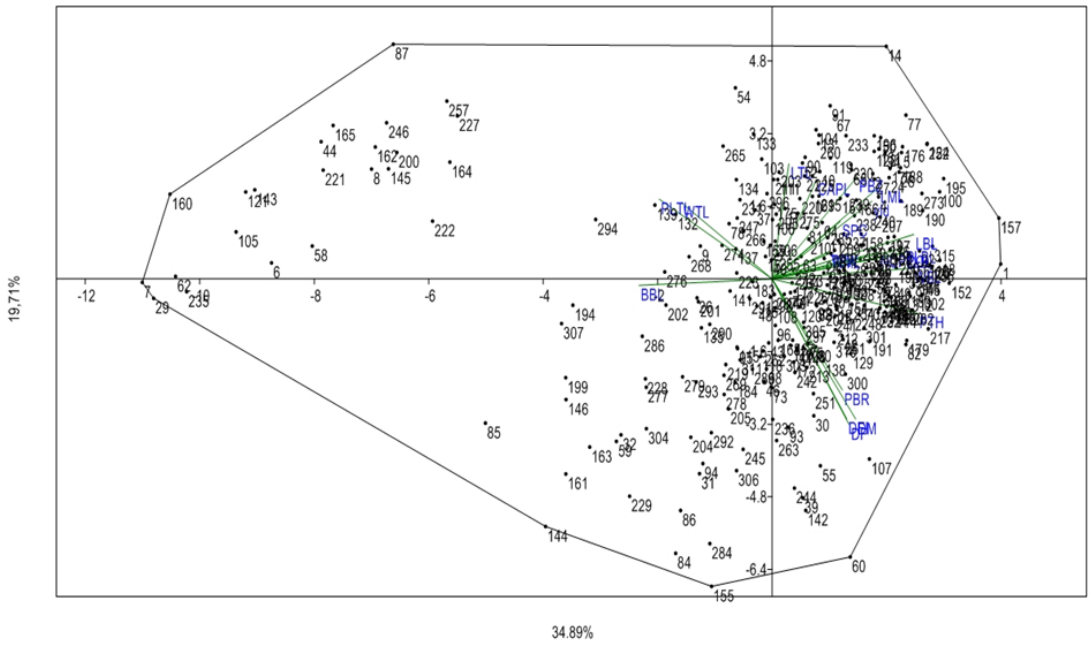


Figure 0-3. Biplot of PC1 and PC2 showing the overlay of 300 genotypes and the 22 studied parameters.

Note: - The black color represents the germplasms (n=300) and the blue color represents the parameters under the study. PCA= Principal component analysis

DISCUSSION

Morphological characteristics

There was a high genetic variability among the tested genotypes for most of the evaluated traits that showed the potential of these genotypes for the sesame breeding program to improve grain yield through direct selection and crossing among variable potential genotypes (Table 3.2). Among the evaluated traits, grain yield showed high variabilities that can be used for direct selection for release and producing potentially high yielder segregation populations through crossing form different origins. Similar results were reported on 10 morphological and 11 agronomic descriptors of sesame (Uzun and Çaðýrgan, 2006; Furat and Uzun, 2010). Number of flowers per leaf axil (number of capsule per plant) is one of the important characters for plant breeding programs since it provides sufficient seed per cross for subsequent evaluation activities. In our study, the 13 sesame genotypes introduced from Egypt gave up to three flowers per leaf axil that showed the potential of these genotypes in the future sesame breeding programs in Ethiopia. The importance of this trait as a yield contributing character was emphasized by Osman, (1989); Ibrahim *et al*. (1983). Large variations were observed in stem, leaf, calyx capsule and Corolla hairiness among the genotypes; ranging from glabrous to strong or profuse (Table 3.2). Uniqueness on the hairiness and shape of hair of stem, petiole and capsule were recorded on a certain number of genotypes that were introduced from Egypt. Since hairiness is a typical character of sesame and can be seen in many parts of the plant such as stem, leaf, corolla, and capsule (Weiss, 1983), and the trait has advantages for preventing insect pests and diseases of sesame, it also has positive effect in terms of preventing water loss, and is a desirable trait in breeding. Having this

variation is important for further breeding program. Some genotypes had Tetracarpellate capsule structure (Table 3.2). This type of capsule potentially enlarges the space for more seeds to fit and a structural modification converting the two extra-floral nectaries to capsules (Bedigian, 2003). However, plants with Tetracarpellate capsules exhibit flatness on the stem, which is rarely seen in Bicarpellate sesame genotypes.

Yield and yield components.

The analysis of variance revealed statistically significant differences at 0.1 and 0.5% probability level among three hundred sesame genotypes for yield and yield component characters (Table 3.2), except for a certain characters (secondary branch, petiole length at middle (mid-level/midheight) leaf, mean capsule width and mean capsule thickness). This indicates the existence of substantial genetic variation among genotypes in all characters. This result agrees with the observation of Saha *et al*. (2012). Genotype 203637 had the maximum values for flower initiation (72 days), days to 50% flowering (82 days), and days to maturity (128 days); while several early maturing germplasms existed in the collection (Table 3.3). These germplasms demonstrated the valuable resource available in the germplasm pool to be used in the breeding programs to develop sesame genotypes adapted to different environments, as well as for studies on thermos and photo-period sensitivity (Suddihiyam *et al*., 1992; Aziz ur Rehman *et al*., 2009). The mean seed yield was 0.52 t ha-1, with a range of 0.034 and 1.24 t ha-1. The mean and the ranges of the genotype values revealed a large genetic diversity useful for the development of varieties. The genotypes with a wide range of variation for agronomic characters had potential to determine the best genotypes for different environments.

Mean, Coefficients of variation, heritability and genetic advance for quantitative traits in sesame germplasm

In all studied characters the phenotypic coefficient of variation (PCV) was greater than the genotypic coefficient of variation (GCV), this indicated that the environment had an important role in the expression of these characters. A Similar type of results is reported by (Ismaila, 2012; Gidey *et al*., 2013; Abate and Mekbib, 2015a; Abate *et al*., 2015a).

Number of primary branches per plant, Petiole length of top leaf, seed yield per hectare, and bacterial blight showed high PCV and GCV estimates while Width of top leaf and number of capsules per plant showed high PCV and moderate GCV. This value provides enough scope for selection and diverse germplasm to improve these characters in the breeding program. The high PCV and GCV for number of primary and secondary branches per plant, number of capsules per plant, and seed yield per hectare is in confirmation with the recent report of Aye *et al.*, (2018) and earlier reports of (Shadakshari *et al.*, 1995; Alegbejo *et al.*, 2003; Solanki and Gupta, 2003; Sumathi and Muralidharan, 2010). A High coefficient of variation for the number of capsules per plant was also reported by (Anitha., 2000). High coefficient of variation for the number of capsules per plant and seed yield per hectare and medium genotypic coefficients of variation (GCV) and high phenotypic coefficients of variation (PCV) for the number of primary branch per plant was reported by Desawi *et al*. (2014).

Plant height, days to flower initiation, days to 50% flowering, 1000 seed weight, and pod bearing zone showed moderate PCV and GCV. Moderate coefficient of variation for Plant height and 1000 seed weight were reported by Aye *et al*. (2018). Low coefficient

of variation for days to maturity was reported by (Shadakshari *et al*., 1995; Krishnaiah *et al*., 2002; Sudhakar *et al*., 2007; Aye *et al*., 2018). The estimate of PCV and GCV values gives only the extent of variability existing in traits. But when it combined with estimates of heritability and genetic advance it would give a better idea about the possible gains of selection in the breeding program.

Heritability and Genetic Advance

According to Robinson (1949), High heritability estimates were observed for plant height, primary branch, Petiole length of top leaf, days to flower initiation, days to 50% flowering, capsule length, 1000 seed weight, days to maturity, pod bearing zone, seed yield, and bacterial blight reaction (table 4). High Heritability for days to 50% flowering (>80%), plant height (90.04%) was reported by Teklu and Kebede, (2014).

Moderate Heritability estimates were obtained from traits; - length of top leaf, number of capsules per plant, capsules width, capsules thickness, and seeds per capsule. Moderate to high heritability for a trait indicates the traits are least influenced by the environmental effects, therefore, the direct selection of the trait for improvement will be effective.

Genetic advance as per cent of mean (GA) is a more reliable index for understanding the effectiveness of selection in improving the traits because the estimates are derived by the involvement of heritability, phenotypic standard deviation, and intensity of selection. Thus, genetic advance along with heritability provides a clear picture regarding the effectiveness of selection for improving the plant characters. Estimates of heritability and genetic advance in combination are more important for selection than heritability alone.

The range of genetic advance as a percent of mean ranged from 1.5% for a length of the longest lip to 70.06% for bacterial blight.

High heritability combined with high genetic advance (as per cent of the mean) was observed for plant height, primary branch, petiole length of top leaf, days to flower initiation, days to 50% flowering, pod bearing zone, seed yield per plant and bacterial blight reaction. This indicates the lesser influence of environment in the expression of these characters and the prevalence of additive gene action in their inheritance and direct phenotypic selection may be effective. Aye *et al*. (2018); Bharathi *et al*. (2004); Furat and Uzun, (2010); Krishnaiah *et al*. (2002); Reddyet *et al*. (2001) reported high heritability combined with high genetic advance (as per cent of the mean) for the number of primary branches per plant, number of capsules per plant and seed yield per plant.

The Correlation coefficient between yield and yield component characters in sesame germplasm

Correlation coefficient analysis measures the mutual relationship between various characters and is used to determine the component character on which selection can be done for improvement in yield. Yield is a complex quantitative trait, greatly influenced by environmental fluctuations. A study of nature and degree of association of component characters with yield will help determine the character that can be used as selection criteria to improve grain yield. The analysis of the relationship among characters and their association with seed yield is essential to establish selection criteria (Singh *et al*., 1990).

Correlations of grain yield with other traits

In the present study, the result of phenotypic correlation analysis revealed that grain yield had a significant and positive correlation with a length of basal leaf, a width of basal leaf, a width of middle leaf, a length of top leaf, petiole length of basal leaf, capsule length, seed per capsule and 1000 seed weight, and had a negative and significant correlation with petiole length of top leaf.

The result of genotypic correlation analysis revealed that grain yield had a significant and positive genotypic correlation with plant height, a length of basal leaf, a width of basal leaf, a length of middle leaf, a length of top leaf, petiole length of basal leaf, number of capsule per plant, pod bearing zone, and positively but not significantly correlated with seed per capsule and 1000 seed weight. The magnitudes of the positive genetic correlation suggest that the selection by those characters produces a significant increase in grain yield.

In agreement with our finding Karuppaiyan, and Ramasamy, (2000), and Singh *et al*. (2000) reported a significant correlation among plant height, and number of capsules with grain yield. Similarly Kathiresan, (2000) showed that the dry matter accumulation and number of capsules per plant contributed positively and significantly to seed yield.

The magnitudes of the positive phenotypic and genetic correlation suggest that the selection by those characters produces a significant increase in grain yield. The phenotypic correlation (r_p), and genotypic correlation (r_g) between the variables show a smaller phenotypic correlation than the genetic correlation, which implies the environmental effect suppressed the association at phenotypic levels.

There is an effect of the environmental factors and/or of the non-additive factors that negatively affect the level of real association between the characters under study. This results in agreement with the work of Espitia and Aramendiz, (2008); Zaman *et al*. (2011).

Cluster analysis

The 300 sesame genotypes were grouped into 6 clusters containing significantly different numbers of sesame genotype, which ranged from 1 to 115. Cluster I and V contained genotypes originating from two different continent countries. Even from the African continent (East, West, South and North Africa), from this we observed a close relationship between genotype from East Africa (majority from different regions of Ethiopia), South Africa, North Africa and West Africa to the genotype from Asia. Clusters II and IV contained genotypes from different countries of Africa that included different regions of Ethiopia. Genotypes from the same origin were not all grouped into the same cluster. This close genetic relationship observed might be due to the introduction of sesame into many countries and material exchange from widely separated locations (Kim *et al*., 2002). Moreover, the exchange of plant materials between Asia and East Africa dated back to a long time ago and still occurs (Zohary *et al*., 2012), with a steady increase in annual exportation of raw sesame seeds mainly for industrial applications. The likelihood of crossover events between materials from different locations grown within the same area is high, knowing that cross-pollination in sesame has been reported to occur at a frequency between 5 and 60% (Wei *et al.*, 2014). This crossing could explain the similarity of genotypes from the eastern part of Africa and Asia. Similar patterns have been observed by other researchers (Kim *et al*., 2002;

Laurentin and Karlovsky, 2006; Cho *et al.*, 2011). Cluster III contained genotypes originating from the same country, but from different three regions (Figure 3.2). This result indicates the possibility of exchange of seeds, and seed trade between farmers, and gene flow across boundaries of those areas (Forsberg *et al.*, 2015) . The last cluster contained one genotype from Benshangul-Gumz. This might indicate that the genotype from Benshangul-Gumz was more diverse than others. The distribution and pattern of genotype, over significantly different clusters from Africa and Asia major geographic origins, would suggest future collections of local genotypes in those geographic regions is importance, for future national and international collection mission in sesame.

Cluster mean analysis.

The cluster mean for different traits (Table 3.8), suggested a wide range of variation for all the characters under the study. Based on Clusters VI and III, the highest mean values for grain yield were recorded from genotypes originating from Ethiopia; but based on clusters IV and I, the lowest mean value for grain yield were recorded from most genotypes introduced from different African and Asian countries. Several factors may contribute to this observation, including climatic reasons such as temperature (day/night), day length, light intensity, precipitation, altitude and latitude. Photosynthesis is influenced by various biotic and abiotic stresses during grain filling; therefore, decrease or increase photosynthesis capacity is a major limiting factor for yield and all yield components (Beheshti and fard, 2010). There are reasons to believe that the less performing germplasm from different African and Asia countries were not adapted to the field conditions, and therefore, gave low yields in this study. Basu *et al.* (2009) also

reported seed yield to be a complex trait governed by polygene, and therefore is influenced more by environmental factors. Our observation is in agreement with previous studies, in which sesame was shown to be highly sensitive to day length since it is a short day plant (Narayanan and Reddy, 1982). Suddihiyam *et al*., (1992) also reported about the significant interactions temperature and day length had on the flowering rate. According to them, yield depends on the interactions of different climatic parameters such as solar radiation, temperature, humidity relative to photosynthetically active radiation (PAR) (Beech and Ashri, 1985; Nath *et al*., 2001).Yadav, *et al*. (1988) also reported close correlations between PAR absorption and yield in sesame. The germplasm listed in the clusters III and VI were found valuable to be selected for use in breeding strategy to improve high yielding sesame genotypes.

Divergence analysis.

The maximum distance among tested sesame genotypes lies between clusters IV and VI (D^2 =342.56). Maximum genetic recombination and variation in the subsequent generation, is expected from crosses that involved parents from the clusters characterized by maximum distances. Thus, it could maximize opportunities for transgressive segregation, since a higher probability that unrelated genotype would contribute unique desirable alleles at different loci. Genetic distance, as a good indicator of transgression and hybridization, has been reported by several authors on many crops (Mulugeta and Hussein, 2013; Pickup *et al.*, 2013). Hence, the attempt to cluster Ethiopian sesame genotypes using multivariate analyses, in the present study, is a significant precursor to initiating sesame breeding program. However, the selection of parents for a particular

cross should also consider the special advantages of each cluster and accession within a cluster, depending on specific objectives of hybridization programs. Members within a cluster are assumed to be more closely related, in terms of trait under consideration than with members in different clusters (Million, 2012; Habtamu and Million, 2013). This indicates that superior hybrids or recombinants can be realized by mating between the lines of these clusters in a definite fashion. Crossing between genotypes belonging to the same cluster might not be expected to yield desirable segregates. This approach is, however, based on the assumption that suitable parents for crossing may be showing greater amount of genetic divergence. Further research on these selected germplasm will save a lot of time for the breeder in future.

Principal Component Analysis (PCA).

Principal components analyses in this study showed that the first five PCAs explained about 75.4% of the variation (Table 3.10). The amount of explained variance by the first PCA and parameters that contributed relatively more, clearly indicated that grain yield and architectural traits of sesame are important traits that could be considered for sesame breeding and selection. Principal components analyses results indicated that the genotype lines studied had a considerable level of variability that could be exploited in future breeding programs. Hybridization between genetically diverse genotypes in sesame to generate promising breeding material has been suggested by Alarmelu and Ramanathan, (1998).

Chapter 4. Genotype × Environment Interactions for Released sesame Varieties across Different Agroecology of Ethiopia

ABSTRACT

Sesame is an important oil crop in Ethiopia in terms of both area coverage and production and it is used as a sources of foreign currency for the national economy. However, the productivity is low because of lack of detail information about genotypes, environment and their interaction. Sixteen nationally and regionally released sesame varieties were evaluated at eleven locations in Ethiopia for two years (i) to determine grain yield performance, stability and adaptability, (ii) to determine the representativeness and discriminating ability of the test locations and (iii) identify core testing sites for selection of superior sesame varieties The experiment in each location was arranged in a randomized complete block design with three replications. Parameters observed included grain yield and yield components. The yield data were analyzed using AMMI and GGE biplot and the yield components data were analyzed using analysis of variance. Genotype, environment and genotype × environment interactions were significant ($p < 0.01$) for grain yield and most other traits measured. In the present study, Grain yield of the sesame ranged from 0.42 t ha-1 for G6 (ADI) to 0.65 t ha-1 for G9 (Obsa). Test locations were divided into six groups; E2 and E7 the first group; E3, E5 and E15 the second and E4, E6, E8, E9, E10, E11, E12, E14, E16 and E17 the third group; and E1, E18 and E19 standing alone as the fourth, fifth and sixth group respectively. E7, E13 and E14 were highly discriminating and representative in the first, the second and in the third group respectively and were identified as a core test site in that group. While E1, E18 and E19 were identified as the only test site in group four, five and six. The core testing sites identified would be used to facilitate the identification of superior sesame varieties to reduce testing cost in the country. Among all Ethiopian representative sesame sites, Environment E16 and E17 (Tach Armacho in 2017/18 and 2018/19) were close to the ideal environment, therefore, it should be regarded as the most suitable to select widely adapted genotypes. The GGE biplot analysis identified genotype G2 (setit-1) as the "ideal" genotype and among the highest mean seed yield. G2 (Setit-1) considered the most stable across variable environments. Humera-1 was plotted near to the ideal genotype considered as desirable genotype.

Key words: - genotype× environment interactions, GGE biplot, ideal genotype, ideal environments,

INTRODUCTION

Sesame (*Sesamum indicum* L.) is one of the oldest and most important oil seed crop widely grown in tropical and subtropical regions around the world and is cultivated for its oil-rich seeds, which grow in pods (Weiss, 1983). Sesame is an herbaceous annual plant requiring 80 to 130 days to attain physiological maturity. It requires 25^{o}c to 27°c for rapid germination, initial growth, and flower formation while temperature below 18^{o}c after germination restricts growth and high temperature (>40^{o}c) during flowering reduces fertilization. Sesame is very drought resistant/tolerant, due in part to an extensive root system but it requires adequate moisture for germination and early growth. It is extensively susceptible to waterlogging and heavy continuous rains at all stages of development (Weiss, 1971; Seegler, 1983; Ashri, 1998).

Sesame is an important oil crop in Ethiopia in terms of both area coverage and production (CSA, 2019). The objective of sesame breeding in Ethiopia is to develop this potential by creating cultivars which meet the demands of the sesame growers, processor, and consumer.

Sesame improvement research in Ethiopia was started in the late l960s by Institute of Agricultural Research (IAR), known at present as the Ethiopian Institute of Agricultural Research (EIAR) at Werer agricultural research center (WARC) under irrigation with landraces and exotic germplasms. As a result, more than twenty varieties were released by national and regional research centers. As sesame is a short day plant and sensitive to light, heat, and moisture stress the yield is not stable (Abate *et al*., 2015b) and no detailed multi-environment evaluation of Ethiopian sesame has been undertaken so far.

In plant breeding programs, genotypes are evaluated in multi-environment trials (METs) by testing their performance across environments and selecting the best genotypes in specific environments. However, the selection of superior genotypes in multi-environment trials usually results in genotype-by-environment interactions that often complicate the interpretation of results obtained and reduce efficiency in selecting the best genotypes (Annicchiaricoand Perenzin, 1994). This interaction is due to the changes in genotype's relative performance across environments, as a result of differential responses of the genotypes to various abiotic and biotic factors (Dixonand Nukenine, 1997). Hence, a significant Genotype by Environment interaction (GEI) for a quantitative trait like grain yield can complicate the identification of superior genotypes for both improved crop development and new crop introduction.

Analyzing the magnitude of GEI by proper techniques rather than neglecting them is useful for exploiting the opportunities and or limiting the disadvantages that these effects may cause. Several statistical models have been proposed for studying the GEI effect and exploiting its advantage. The two frequently used statistical analyses are the additive main effects and multiplicative interaction (AMMI) model, the genotype main effect, and the genotype x environment interaction effect (GGE) model (Gauch, 2006). AMMI model combines the analysis of variance, genotype and environment main effects with principal component analysis of GEI into a unified approach (Gauch and Zobel, 1996). However, the GGE biplot method, which is always close to the best AMMI model in most cases (Ma *et al.*, 2004), was developed to use some of the functions of these methods jointly. Purchase *et al.* (2000) developed a quantitative stability value known as the AMMI stability value (ASV) to rank genotypes through the AMMI model. The

developed ASV was considered to be the most appropriate single method to describe the stability of genotypes. Gruneberg *et al.* (2005) (Grüneberg *et al.*, 2005) showed that AMMI, as a multivariate tool was highly effective for the analysis of multi-environment trials (MET). The GGE- methodology, which is composed of two concepts- the biplot concept (Gabriel, 1971) and the GGE concept (Yan *et al.*, 2000) was used to visually analyze the multi-environment yield trial (MEYTs) data. The GGE concept is based on the understanding of genotype by environment interaction (GE) and genotype (G) and they are the two sources of variation that are relevant to genotype evaluation and that they must be considered simultaneously (Yan, 2002). The GGE-biplot model provides breeders with a more complete and visual evaluation of all aspects of the data by creating a biplot that simultaneously represents mean performance and stability as well as identifying mega environments (Yanand Kang, 2003; Ding *et al*., 2007).

The difference of AMMI from GGE is that GGE biplot analysis is based on environment centered PCA whereas AMMI analysis is based on double centered PCA. For the research purpose of gaining accuracy AMMI and GGE are still equally useful (Gauch *et al*., 2008). Samonte *et al.* (2005) found the AMMI and GGE biplot analyses complementing one another.

The main objectives of this study were therefore,

- To determine grain yield performance, stability and adaptability,
- To determine the representativeness and discriminating ability of the test locations and
- To identify core testing sites for selection of superior sesame varieties.

MATERIAL AND METHOD

The experiment was conducted at Alemaya, Areka, Assosa, Banat, Humera, Kebebew, Metema, Pawe, Shiraro, Tach Armacho, Worer and Mender67 trial sites in Ethiopia (Table 4.1). Sixteen released sesame varieties, denoted as G1, G2, G3, G4, G6, G7, G8, G9, G10, G11, G12, G13, G14, G15 and G16, were evaluated (Table 4.2).

The plant materials were planted across locations during 2017/18 and 2018/19 cropping season. The location by year combinations would have totaled twenty four environments, but the second year data of four location were lost due to security and environmental condition, then nineteen environment data were used for analysis. The experiment was laid out in a completely randomized block design with three replications. Each experimental plot had five rows of 5 m length with 0.4 m between rows ($10m^2$). Each plot had a total area of 10 m^2 with of five rows and $6m^2$ net plot areas with three harvestable rows. Each experimental plot received fertilizer based on site specific recommendation. All field management practices were done equally and properly as per the recommendations to each environment.

All quantitative; plant height, primary branch, days to 50% flowering, number of capsules per plant, Seeds per capsule, 1000-seed weight, days to physiological maturity, pod bearing zone, seed yield, bacterial blight were recorded according to the sesame descriptors list of IPGRI and NBPGR, (2004). All measurements were done after tagging five randomly selected plant in each plot. Yield, growth period, and disease reactions were recorded using standard procedures (IPGRI and NBPGR, 2004). The early flowering date was recorded as the number of days from sowing to observation of the first flower on 50% of the individuals. Flower related traits (days to 50% flowering and

days to physiological maturity) were observed and measured in the full-bloom stage. After harvesting manually, yield-related traits (Seeds per capsule, 1000-seed weight seed yield and oil content) were measured in the laboratory. Seed yield was collected per plot and later converted into metric tonnes per hectare.

Table 0.1. Description of twelve locations used for evaluation of sesame genotypes

Location	Soil type	Temperature (mean)	Rainfall (mm)	Latitude	Longitude	Altitude a.m.s.l
Alemaya	Sandy loam		671	9°08'	42°21'	1650
Areka	Sandy-clay		1456	090 19’N	330 23E	1600
Assosa	loams and black clay		900 - 1300	10°02.922’N	34° 33.8’E	1650
Banat				13°49.686'N	36° 30.691E	635
Humera		18.8-37.6	576.4	14°15'N	36°37'E	609
Kebebew				13°36'50.593"N	36° 41'26.398E	689
Metema			1030.2	120 39’N	360 17’ E	760
Pawe	Nitisol		1586	110 18’N	360 24’E	1100
Shiraro		18.8-34.9	676.7	14°24.175'N	37°45.914'E	1018
Tach Armacho			970.88	13088’N	370 43’E	1022
Worer	Fluvisol and Vertisol soil	19.5-32.5.0 C	450	90 60’ N	400 9’ E	740
Mender67	Clay loam			1433505	203088	676m

Table 0.2. Description of 16 released sesame genotypes evaluated in 12 locations during the 2017/18 and 2018/19 cropping season

Genotype	Source	Year of Release	Seed color	Oil (%)	Maturity date
HUMERA-1	HUMERA ARC	2011	White	54-56	90-110
SETIT-1	HUMERA ARC	2011	White	52-54	80-90
SETIT-2	HUMERA ARC	2016	White	53.77	80-87
HUARC-4	HUMERA ARC	2017			
ABASENA	WORER ARC	1990	White	40.6-48.7	103-120
ADI	WORER ARC	1993	White	40.20-57.7	85-91
TATI	WORER ARC	2000	Light grey	47.48- 48.71	111-115
Acc-051-02-sel 1 (2)	WORER ARC	2017			
OBSA	BAKO ARC	2010	White – tan	51.55	120-137
CHALESA	BAKO ARC	2013			
DANGUR	PAWE ARC	2015	Grey	56.7	124
GONDAR-1	GONDAR ARC	2016	White	50	101
MECHAL	SIRINKA ARC	2013	White	50.4	105-120
BENSHANGUL- 1	ASSOSA ARC	2016	White	54	90-115
BAHA NECHO	ALEMAYA ARC	2016	White	52	114-129
BAHA ZEYT	ALEMAYA ARC	2016	Light grey	56	113-134

Table 0.3. Genotypes and environments and their codes

No	Environments	Env. code	No	Genotype	Genotype code
1	Alemaya 2017/18	E1	1	HUMERA-1	G1
2	Areka 2017/18	E2	2	SETIT-1	G2
3	Areka 2018/19	E3	3	SETIT-2	G3
4	Assosa 2017/18	E4	4	HUARC-4	G4
5	Banat 2017/18	E5	5	ABASENA	G5
6	Banat 2018/2019	E6	6	ADI	G6
7	Humera 2017/18	E7	7	TATI	G7
8	Humera 2018/19	E8	8	Acc-051-02-sel 1 (2)	G8
9	Kebebew 2017/18	E9	9	OBSA	G9
10	Kebebew 2018/19	E10	10	CHALESA	G10
11	Metema 2017/18	E11	11	DANGUR	G11
12	Metema 2018/19	E12	12	GONDAR-1	G12
13	Pawe 2017/18	E13	13	MECHAL	G13
14	Shiraro 2017/18	E14	14	BENSHANGUL- 1	G14
15	Shiraro 2018/19	E15	15	BAHA NECHO	G15
16	Tach Armacho 2017/18	E16	16	BAHA ZEYT	G16
17	Tach Armacho 2018/19	E17			
18	Worer 2017/18	E18			
19	Mender67 2017/18	E19			

Data Analysis

First analysis of variance was made for each of the environments to know the existence of genetic variability among experimental genotypes and to verify homogeneity of the error variances. Homogeneity of residual variances was tested prior to a combined analysis using Bartlett's test (Steel and Torrie, 1980).

The combined analysis of variance of the locations were performed, to identify the possible interactions of genotypes with environments. Analysis of variance for each environment, combined analysis of variance over environments and GGE biplot analysis was computed using GenStat 18th edition (2015) (VSN International, 2015) and SAS.9.0. Adaptability and stability analyses were done using the multivariate AMMI and GGE-biplot methods after the significance of the GxE interaction was determined.

AMMI and GGE-biplot analysis

The AMMI model, which combines the standard analysis of variance with principal component analysis (Zobel *et al*., 1988), was used to estimate the magnitude of G x E interaction. Bartlett's test (Steel and Torrie, 1980) indicated heterogeneity error variance for the trait seed yield in each of the 19 environments. The AMMI analysis and the IPCA were performed using Genstat 18th edition (2015)(VSN International, 2015). The AMMI's stability value (ASV) was calculated to rank genotypes in terms of yield stability using the formula suggested by Purchase *et al*. 2000) as shown below.

$$ASV = \sqrt{\left[\left[\frac{SSIPCA1}{SSIPC2}\right] * [IPCA1score]\right]^2 + (IPCA2score)^2}$$

Where: SS= sum of squares, IPCA1= Interaction principal component analysis axis one, IPCA2= Interaction principal component analysis axis two.

In general, an absolute stability value (ASV) was determined using a procedure that combines IPCA1 and IPCA 2.

The GGE-biplot shows the first two principal components (PC1 and PC2, also referred to as primary and secondary effects, respectively) derived from subjecting environmental centered yield data (yield variation due to GGE) to singular value decomposition (Yan *et al*., 2000). For raw data of seed yield biplots of the first two principal components were constructed using GenStat 18th edition (VSN International, 2015) and GEA-R used to illustrate the relation among genotypes, environments and between the genotypes and environments. In the present study, genotype-focused scaling was used to compare

genotypes, while environment focused, scaling was used to compare environments. Furthermore, symmetric scaling was preferred in visualizing the which–won-where pattern of the multi environment trial yield data (Yan and Kang, 2003).

RESULTS AND DISCUSSION

Analysis of variance and mean values

The combined ANOVA across different environment showed that mean squares for environment were significant ($p < 0.01$) for grain yield and all other measured traits of the genotypes studied (Table 4.4). Genotypes (G) mean squares were significant ($p < 0.05$) for grain yield and all other traits measured except for pod per plant. Genotype by environment interaction were significant for grain yield and for most other traits measured except for pod per plant and pod length (Table 4.4). The significant mean squares recorded in the current study for genotypes for most traits, except for pod per plants indicated that the genotypes responded differently to the test environments and thus, call for the need to identify high-yielding and stable genotypes across the locations (Badu-Apraku *et al.*, 2003; Moghaddam *et al*., 2009). Also, the highly significant genotype by environment interaction (GEI) for grain yield of the released sesame varieties seeks to justify the need for the testing of the genotypes in multiple locations over years before recommendation.

The mean value of 16 released varieties for recorded quantitative traits were listed in Table 4.5. Grain yield of the genotypes ranged from 0.42 t ha-1 for G6 (ADI) to 0.65t ha-1 for G9 (Obsa).

Table 0.4. Mean squares for grain yield and other agronomic traits of released sesame varieties evaluated across nineteen environments in 2017/18 and 2018/19 cropping season

	Mean squares				
SOURCE	Environment (E)	Genotype (G)	Rep(E)	G*E	Error
DF	18	15	38	270	570
Grain Yield	2384879.29**	203750.52**	49773.41**	69906.03**	25224.84
% CTV	64.31	4.58	2.83	28.28	4.58
Days to flowering	1038.05**	431.84**	11.09**	12.98**	4.74
Days to maturity	4900.32**	405.92**	26.34**	44.8**	6.98
Number of Primary branch	27.95**	15**	3.15**	1.41	1.08
Plant height	28986.25**	3404.55**	246.41**	178.25**	124.18
Pod bearing zone	8628.52**	1006.53**	255.62**	123.17**	73.06
Pod per plant	10006.96**	387.22	534.56**	244.35	206.87
Pod length	4.17**	0.85**	0.1	0.08	0.06
Seed per pod	7535.99**	420.00**	108.38	109.66**	60.56
Bacterial blight	94.54**	9.13**	0.54	0.8**	0.29
Thousand seed weight	6.19**	1.1**	0.16	0.24**	0.14
Oil content	307.47**	20.71**	5.12**	8.41**	2.15

Notes: ns = not significant; %CTV = percentage contribution to total variation; *Significant at 0.05 probability level. **Significant at 0.01 probability level.

Table 0.5. Combined Grain yield and other agronomic traits of released sesame varieties evaluated across nineteen environment in 2017/18 and 2018/19 cropping season.

TRT	DF	DM	PBPP	PLTH	PBZ	PPP	PL	SPP	BBL	TSW	YLD	OC
G1	46	100	2.83	117.16	61.79	41.51	2.64	55.5	3.61	2.85	625.77	52.78
G2	46	101	2.8	119.31	63.69	45.29	2.68	56.74	3.42	2.91	626.2	52.26
G3	47	98	2.91	119.89	63	41.77	2.98	59.96	3.91	2.62	586.26	51.93
G4	47	97	2.91	117.59	63.79	45.76	2.62	57.12	3.97	2.52	566.54	51.33
G5	54	105	3.83	133.63	57.04	43.88	2.56	62.18	2.85	2.48	554.7	51.1
G6	53	105	3.29	139.55	56.92	42.28	2.57	50.65	3.67	2.7	418.46	50.06
G7	53	104	4.33	123.41	49.12	42.48	2.58	56.68	3.88	2.52	490.62	50.19
G8	52	104	4.06	130.25	52.31	44.87	2.44	56.06	3.61	2.66	491	50.65
G9	52	105	4.06	130.25	60.1	49.62	2.46	61.2	2.64	2.32	654.66	50.78
G10	52	106	3.93	132.15	61.05	45.03	2.43	62.97	2.64	2.36	559.72	51.03
G11	53	106	3.74	134.57	58.31	44.19	2.46	59.9	2.55	2.47	611.72	51.43
G12	53	104	4	133.45	57.26	47.2	2.42	59.25	2.67	2.46	570.17	50.84
G13	49	103	3.3	133.3	64.59	48.48	2.57	59.47	2.97	2.5	551.26	51.34
G14	53	105	4.11	139.97	58.21	49.93	2.5	61.2	2.61	2.65	567.41	50.8
G15	53	105	3.96	135.21	57.88	45.34	2.48	59.39	2.91	2.58	555.17	51.4
G16	53	105	3.98	136.24	57.02	46.06	2.54	57.7	3	2.46	508.22	51.2
MEAN	51.1	103.2	3.63	129.74	58.88	45.23	2.56	58.5	3.18	2.57	558.62	51.19
LSD	1.48	1.86	0.69	7.18	5.51	9.27	0.18	5.65	0.45	0.28	102.39	1.1
CV	4.26	2.56	28.59	8.59	14.52	31.8	9.64	13.3	16.8	14.35	28.43	2.86

AMMI Analysis

The AMMI analyses of variance showed that grain yield was significantly ($P<0.01$) influenced by environment, genotype, and genotype-environment interaction (GEI) (Table 4.6). The significant effect of GEI on seed yield implied differential responses of the genotypes across the environments. This suggestion is consistent with that of (Daba *et al*., 2015) who found similar results in sesame. Significant GEI complicates selection since the variety with the highest mean yield may not be the best genetically (Signor *et al*., 2001). In the present study environment, GEI, and genotype explained 64.31%, 28.28%, and 4.58% of the total variation, respectively (Table 4.6). The Genotype was

small in variation among them, whereas environment (E) and genotype by Environment Interaction (GEI) explained most of the variations. This indicated that Environment (E) and genotypes by Environment Interaction (GEI) are both important in governing for expression of this trait (Mulugeta *et al*., 2014).

The magnitude of the environment was 14 times greater than the genotype, implying that most of the variation in seed yield was due to the environment. This indicates that the locations were diverse and large influence of environment on yield performance of sesame genotypes across all locations. Similar result was reported on sesame (Abate, 2015; Misganaw *et al*., 2015; Belay, 2016; Belay, 2018).

A highly significant GEI indicates the necessity for further analysis for yield stability and GEI must be considered in genotype evaluation and that GGE biplot analysis would be essential to reach meaningful conclusions about the genotypes (Yan, 2014).

The AMMI analysis partitioned the sum of squares of GEI into nine interaction principal component axes (IPCA), of which the first four IPCA were significant (Table 4.6). The results from the AMMI model showed that, the first IPCA captured 42.16% of the interaction sum of squares. Similarly, the second and the third (IPCA2 and IPCA3) explained 18.25% and 9.98% of the GEI sum of squares, respectively. The sum of squares for the first four IPCAs cumulatively contributed to 79.2 % of the total GEI. In this line, Zobel *et al.* (1988); Crossa *et al*. (1990) proposed that two interaction principal component axes for AMMI model were sufficient for a predictive model. Other interaction principal component axes captured were mostly non-predictive random variation and did not fit to predict validation observations. Therefore, in general, the model chosen by predictive criterion consists of two IPCA (Sanni *et al.*, 2009).

Table 0.6. Analysis of variance (ANOVA) for grain yield (2017/18 and 2018/19)

Sources	DF	SS	MS	Total variation explained (%)	(%) G x E Explained	Cumulative (%)
Total	911	81128268	89054			
Genotypes	15	3056256	203750***	4.58		
Environments	18	42927829	2384879***	64.31		
Reps within Env.	38	1891390	49773***	2.83		
Interactions	270	18874631	69906***	28.28		
IPCA 1	32	7958420	248701***		42.16	42.16
IPCA 2	30	3444904	114830***		18.25	60.41
IPCA 3	28	1884065	67288***		9.98	70.39
IPCA 4	26	1663195	63969***		8.81	79.2
IPCA 5	24	876788	36533		4.65	83.85
IPCA 6	22	712870	32403		3.78	87.63
IPCA 7	20	654044	32702		3.47	91.1
IPCA 8	18	562442	31247		2.98	94.08
IPCA 9	16	411881	25743		2.18	96.26
Residuals	54	706022	13074		3.74	100
Error	570	14378162	25225			

Note: Grand mean = 558.62; R-squared = 0.8228; C.V. = 28.43%; **P<0.01; *** P<0. 001; IPCA=Interaction principal component axis.

Purchase, (1997) reported that the IPCA scores of genotypes in the AMMI analysis are an indication of the stability of a genotype over environments. The greater the absolute value IPCA1 scores, the more specifically adapted a genotype is to a particular environment. The more IPCA2 scores approximate to zero, the more stable or adapted the genotype is over all environments sampled (Gauch and Kang, 1996; Ferney *et al.*, 2007). The genotype G2 (SETIT-1) and G13 (MECHAL) showed the lowest absolute scores for the IPCA1 and they were the most stable followed by G16 (BAHA ZEYT) (Table 4.7). The more the IPCA2 score approximates to zero in absolute terms, the more stable or adapted the genotype is over all the environments sampled (Alberts, 2004). When IPCA2 was considered, G16 (BAHA ZEYT) was the most stable followed by G12 (GONDAR-1). Stability rank of genotypes varied for IPC1 to IPC2. This means that the two IPCA have

different values and meanings. Therefore, the other option is to calculate ASV to get estimated value between IPCA1 and IPCA2 scores as ASV was reported to produce a balance measurement between the two IPCA scores (Purchase, 1997). In the present study, Genotype G13 (MECHAL), G16 (BAHA ZEYT), G14 (BENSHANGUL- 1) and G2 (SETIT-1) were found to be stable (Table 4.7). Although MECHAL and BAHA ZEYT were the first and second stable genotypes for ASV, it was ranked 12th and 13th for mean seed yield. As per the value of ASV the most unstable genotypes were G7 (TATI), G8 (Acc-051-02-sel 1 (2)) and G6 (ADI). It is to note that a genotype with low ASV values is considered more stable than a genotype with high ASV (Purchase, 1997).

Table 0.7. Mean yield (kg ha-1) rank, IPCA1 and 2 scores and ASV sesame genotypes tested across 19 environments.

No	Genotype	Yield	Rank	IPCA1	IPCA2	ASV	Rank
1	HUMERA-1	625.77	3	-7.12468	-11.0876	19.85	5
2	SETIT-1	626.2	2	-0.86882	-11.3408	11.52	4
3	SETIT-2	586.26	5	-11.2291	-12.0946	28.62	12
4	HUARC-4	566.54	8	-9.76351	-14.8257	26.99	11
5	ABASENA	554.7	11	11.0488	1.35129	25.56	9
6	ADI	418.46	16	-12.6691	9.12935	30.66	14
7	TATI	490.62	15	-16.7354	12.215	40.55	16
8	Acc-051-02-sel 1 (2)	491	14	-13.9062	11.3544	34.07	15
9	OBSA	654.66	1	12.558	-0.36402	29.01	13
10	CHALESA	559.72	9	8.2858	8.33943	20.88	6
11	DANGUR	611.72	4	10.1487	1.19988	23.48	7
12	GONDAR-1	570.17	6	10.7665	-0.2075	24.87	8
13	MECHAL	551.26	12	1.81811	3.44992	5.44	1
14	BENSHANGUL- 1	567.41	7	3.79687	2.74372	9.19	3
15	BAHA NECHO	555.17	10	11.4389	0.23806	26.43	10
16	BAHA ZEYT	508.22	13	2.43527	-0.10081	5.63	2

Where: IPCA1= Interaction principal component analysis axis one; IPCA2= Interaction principal component analysis axis two; ASV = AMMI stability value

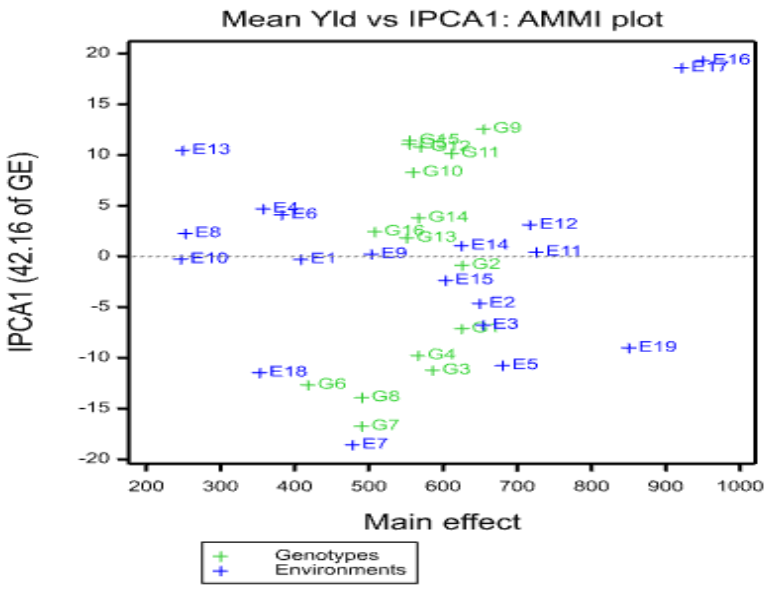


Figure 0-1**.** A biplot of sesame yield environmental means (kg ha^{-1}) vs IPCA1 for 16 sesame genotypes in the nineteen environments.

GGE biplot analysis of grain yield response and stability

Which Won Where/What

The principal component axis 1 (PC1) accounted for 41.56% of total variation and principal component axis 2 (PC2) also accounted for 21.76%. Cumulatively, these two principal components explained 63.32% of the total variation for grain yield (Figures 4.2 and 4.3). The polygon in Figure 4.2 is formed by connecting the markers of the genotypes that are farthest away from the biplot origin, such that all other genotypes are contained in the polygon. The polygon view also contains a set of lines perpendicular to each side of the polygon. These perpendicular lines divide the biplot into several sectors. The vertex genotype in each sector represents the highest yielding genotype (the winning genotype) in the environment that falls within that particular sector (Yan and Tinker,

2005; Yan *et al.*, 2010). There are nine sectors in (Figure 4.2) with genotypes G1, G2, G3, G4, G6, G7, G9 and G10 as the vertex genotype. Environment E2 and E7 fell in the sector in which genotypes 3 and 4 were the vertices genotypes implying that, genotypes 3 and 4 were the best genotypes for E2 and E7. Genotypes 1 and 2 were the highest yielding genotypes at E3, E5 and E15. Genotypes G9 was the highest yielding in E4, E6, E8, E9, E10, E11, E12, E14, E16 and E17. Genotypes G10 was best in E1 while genotypes G6 and G7 were the highest yielding in E19. No genotypes fell as a vertices and notably in the mega environment that environment 18 classified. Genotypes within the polygon, notably genotypes G5, G8, G11, G12, G13, G14, G15 and G16 were less responsive than the vertex genotypes. The ninety environment fell into six sectors. This pattern suggests that the target environment may consist of six different testing environments and that different genotypes should be selected and deployed for each.

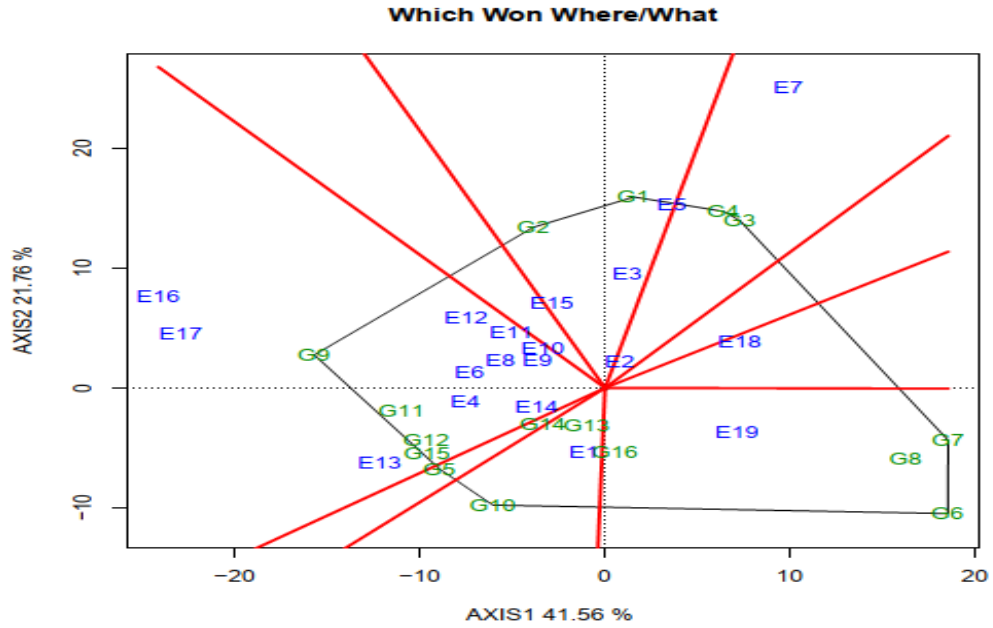


Figure 0-2. A "Which won where/what" of genotype × environment biplot of the best 16 genotypes evaluated under rain-fed and irrigated conditions at 19 environments

Mean vs. Stability

Using the stability GGE biplot of grain yield for the best 16 national and regional released sesame varieties as shown in Figure 4.3. This figure is the average-environment coordination (AEC) view of the GGE biplot. The single arrowed line is the AEC abscissa (or AEA) and points to higher mean yield across environments. The mean yield of the genotypes is estimated by the projections of their markers on the average-tester axis. The genotypes were ranked along the average-tester axis (ATC abscissa), thus, ‘G9’ had the highest mean yield, followed by ‘G2’, ‘G1’, etc., whereas ‘G6’, ‘G7’ and ‘G8’ had the lowest mean yield.

The AEC ordinate passes the plot origin and is perpendicular to the AEC abscissa and points to greater variability (poorer stability) in either direction. The greater the absolute length of the projection of a genotypes, the less stable it is. Thus, ‘G3’, ‘G4’ and ‘G7’ were highly unstable, whereas ‘G13’ ‘G14’ ‘G16’ were highly stable.

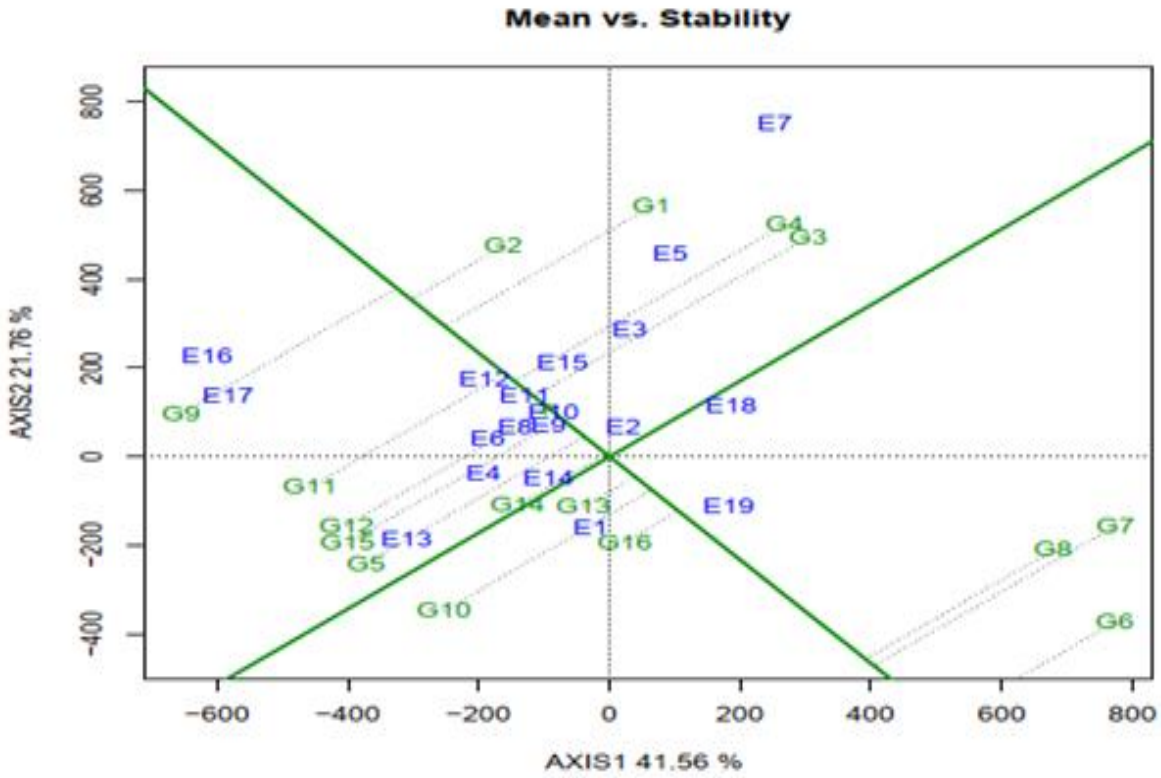


Figure 0-3. Average environment coordination (AEC) views of the GGE-biplot based on environment-focused scaling for the means performance and stability of genotypes.

Relationship among environments

The nineteen environment showed significant variation. Each environment provided different conditions for expression of yield for the tested genotypes. The vector view of the GGE biplot in Figure 4.4 shows the interrelationships between the test environments. The lines connecting the biplot origin and the markers for the environments are termed environment vectors. The cosine of the angle between the environmental vectors indicates the correlation coefficient between them (Yan, 2002). The smaller the angle, the more highly correlated the environments (Yan and Kang, 2003; Yan and Tinker, 2006). When the angle between two environments is less than 90°, the correlation coefficient

between them is positive (Yan, 2002). There is therefore a positive correlation between E2, E3, E5, E7 and E18 and between E6, E8, E9, E10, E11, E12, E15, E16 and E17 as well as among E1, E4, E13 and E14. From the vector view of the GGE biplot in Figure 4.4 the nineteen environments in this study were clustered into four groups: E2, E3, E5, E7 and E18 constitute the first group, then E6, E8, E9, E10, E11, E12, E15, E16 and E17 the second group, E1, E4, E13 and E14 constitute the third group whiles E19 stands alone as the fourth group.

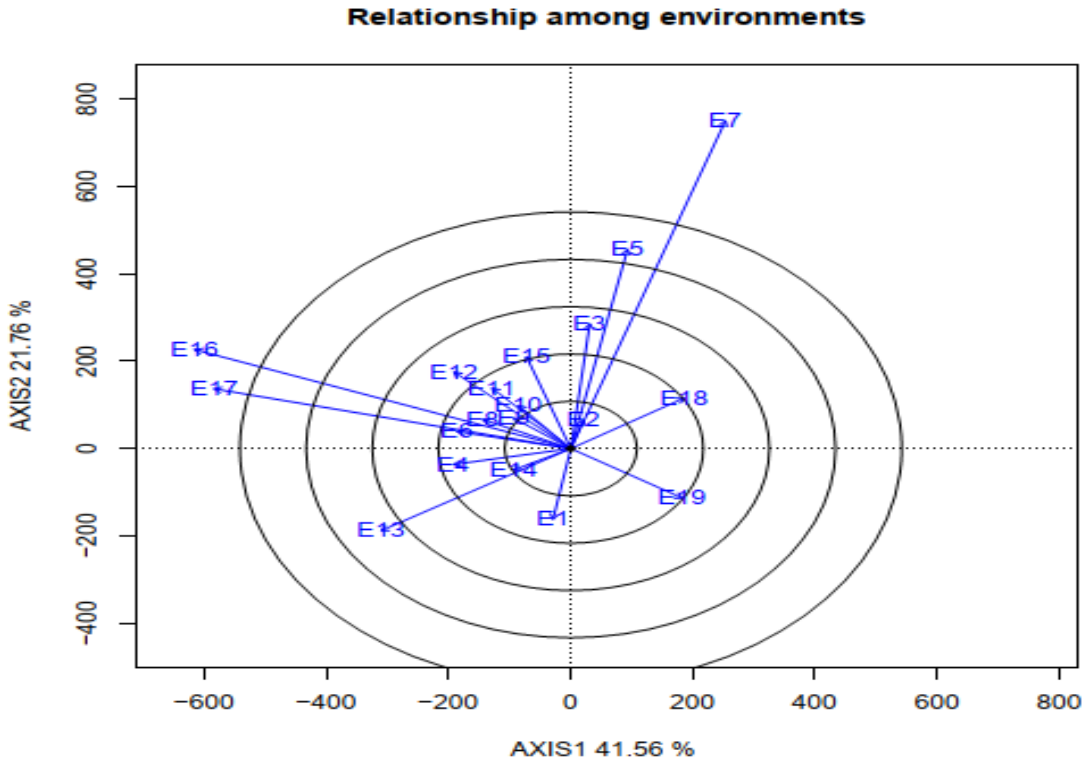


Figure 0-4. Vector view of the GGE biplot showing relationships among locations.

Ranking test locations based on both discriminating ability and representativeness

Figure 4.5 represents the discriminating ability and representativeness of the test environment. The ability of an environment to identify an ideal test environment is

referred to as the discriminating power of an environment (Badu-Apraku *et al.*, 2012) and the distance between the markers of the environment to the biplot origin, is a measure of its discriminating ability (Frutos *et al*., 2014). The ability of a test environment to represent the mega-environment is referred to as the representativeness (Badu-Apraku *et al.*, 2012) and the magnitude of the projection from the marker of the environment onto the average environment coordinate (AEC) axis is the measurement of its representativeness (Frutos *et al*., 2014). The small circle is the average-environment axis (AEA), and according to Yan and Tinker (2005) the arrow pointing to it is used to indicate the direction of the AEA. Test environments having large angles with the AEA are less representative of the mega environment than those having small angles with it. Environments with longer vectors are more informative compared to those with shorter vectors and offer more information about the genotypes. Environments with shorter vectors could therefore be excluded when choosing test environments since they offer little or no information about the genotypes. Test environments with shorter environmental vectors indicate weak correlation with test environments with longer vectors. Consequently, the short-vector environments E1, E18 and E19 may be regarded as independent research environments and may be treated as unique and, therefore, essential research environments. In contrast, the long-vector test environments E5, E7, E13, E16 and E17 were more powerful in discriminating among the cultivars, E7, E16 and E17 being the most discriminative environments. Test environments with long vectors and small angles with the AEC abscissa are ideal for selecting superior genotypes while test environments with long vectors and large angles with the AEC abscissa are useful in culling unstable genotypes (Yan *et al.*, 2007, 2010). E7, E16 and E17 had long

vectors and large angles with the AEC abscissa suggesting that they may not be used in selecting superior genotypes, but may be used in culling unstable genotypes. Furthermore, the distance between two test environments measures their dissimilarity in discriminating the genotypes and presence of close associations among test environments suggest that the same information can be obtained from the fewer environments and this will reduce the testing cost. In the present study E2 and E7 and testing environment E9, E10, E11, and E12 and similarly, test environment E4, E13 and E14 as well as test environment E3 and E5 were closely associated, confirming that these environment produced similar information about the genotypes and thus implying that a promising genotypes in this study selected in one of these environment will also be suitable for production in the other environment. Hence, testing environment can be dropped in this case. The discriminativeness versus representativeness biplot (Figure 4.5) strongly suggests, E7 from E2, and E12 among environments E9, E10, and E11, and E13 from Environments E4, E13 and E14 as well as E5 from Environment E3 were better discriminating and representative environments.

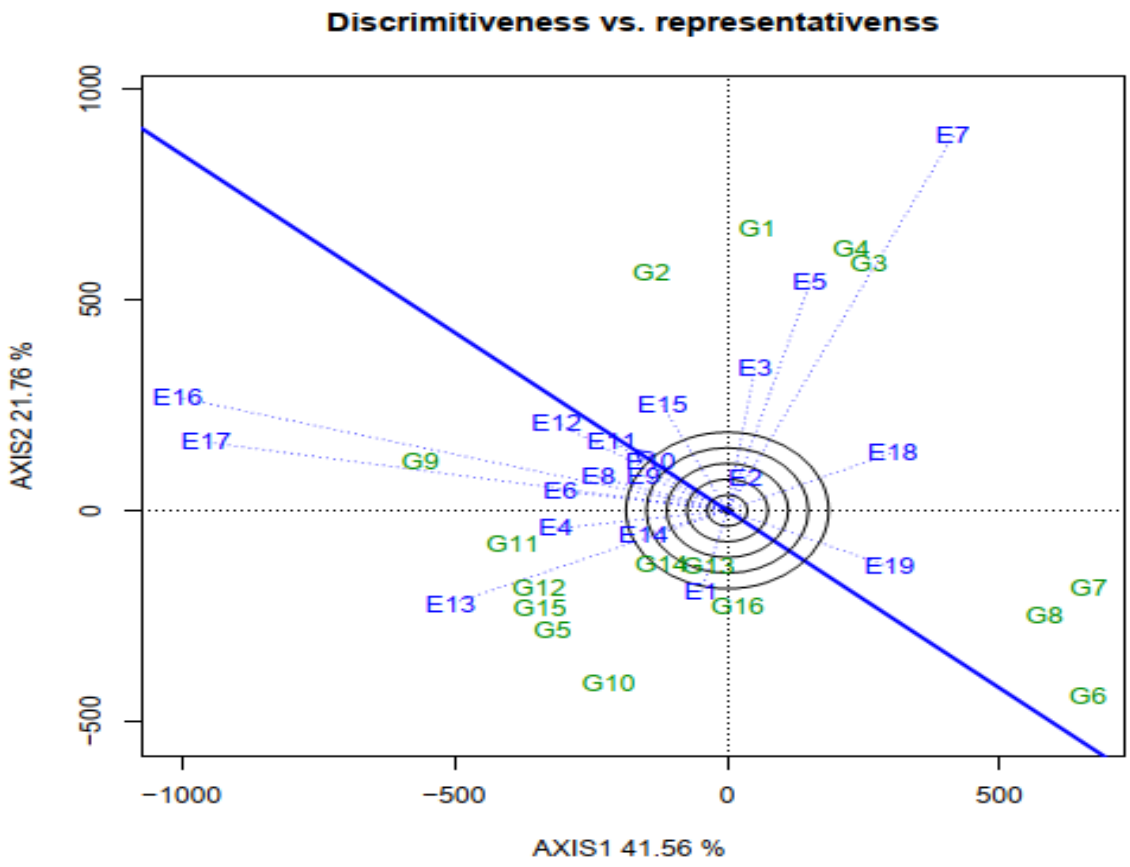


Figure 0-5. The discriminating ability and representativeness of the test environments

Evaluation of Genotypes Based on the Ideal Genotype

An ideal genotype has the highest mean seed yield and is stable across locations (Farshadfar *et al.*, 2012). The ideal genotype is located in the first concentric circle in the biplot. Desirable genotypes are those located close to the ideal genotype. Thus, starting from the middle concentric circle pointed with arrow concentric circles was drawn to help visualize the distance between genotypes and the ideal genotype (Yan and Tinker, 2006). The ideal genotype can be used as a benchmark for selection. Genotypes that are

far away from the ideal genotype can be rejected in early breeding cycles while genotypes that are close to it can be considered in further tests (Yan and Kang, 2003). From this study G2 (setit-1) was the "ideal" genotype and the highest mean seed yield. G2 (Setit-1) considered the most stable across variable environments. Genotypes closer to the ideal genotype were the stable ones, while genotypes far from the ideal genotypes were the unstable. Humera-1 was plotted to the ideal genotype considered as desirable genotype, while G6 (ADI), G7 (TATI) and G8 (Acc-051-02-sel 1 (2)) were low yielding genotypes associated with genotypic instability (Figure 4.6). Genotype is more desirable if it is located closer to the ideal genotype.

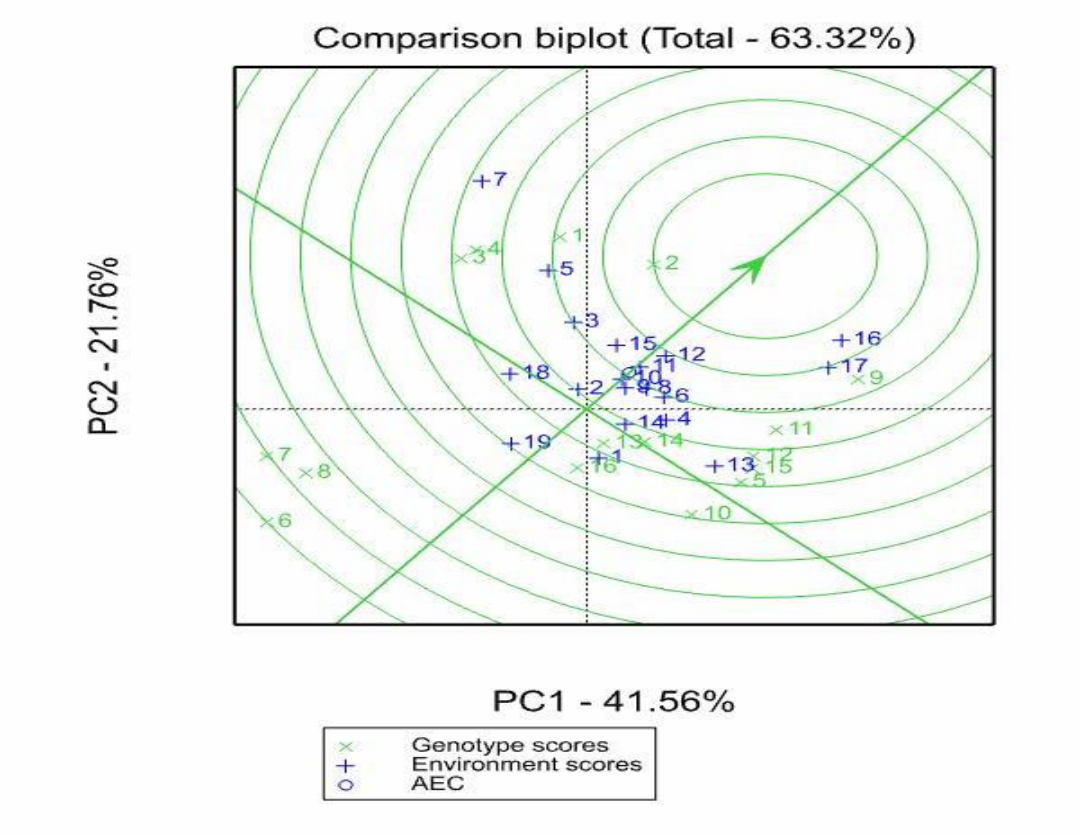


Figure 0-6. GGE-biplot showing the "ideal" genotype.

Evaluation of Environments Based on the Ideal Environment

The ideal environment is representative and has the highest discriminating power (Yan and Tinker, 2006). Similarly to the ideal genotype, the ideal environment is located in the first or near to the first concentric circle in the environment focused biplot, and desirable environments are close to the ideal environment. Nearest to the first concentric circle, Environment E16 and E17 were close to the ideal environment (Figure 4.7), therefore, it should be regarded as the most suitable to select widely adapted genotypes. E19, E18, and E7 were far from the ideal environment and considered as unstable.

Based on the distance from the concentric circle in the environment focused biplot the order of ranking the other test Environment were as follows: E16, E17, E12, E11, E15, E6, E8, E10, E9, E4, E3, E13, E5, E14, E2, E1, E7, E18, and E19.

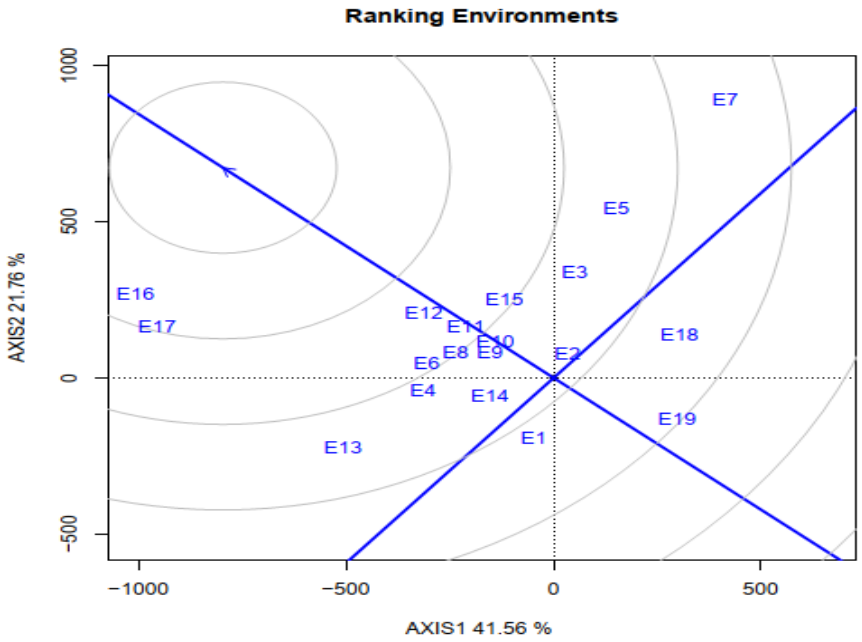


Figure 0-7. GGE - biplot showing the “ideal” environment

Chapter 5. Genetic Diversity, Population Structure Analysis Using Ultra-High Throughput Diversity Array Technology (DArTseq) in Different Origin Sesame (*Sesamum indicum L*)

Submitted and available on web as preprint

ABSTRACT

Sesame is an important oil crop widely cultivated in Africa and Asia continent. Characterization of genetic diversity and population structure of sesame genotypes in these continents can be used to designing breeding methods. In the present study, 300 sesame genotypes comprising 209 Ethiopian landraces, and 75 exotic collections from different African and Asia countries, and 16 improved varieties provided from Ethiopian Biodiversity Institute and research centers were used. The panel was genotyped using two high-throughput diversity array technology (DArT) markers (silicoDArT and SNP). Both markers were used to identify the genetic diversity and population structure of sesame germplasm. A total of 6115 silicoDArT and 6474 SNP markers were reported, of which 5002 silicoDArT and 4638 SNP markers were screened with quality control parameters. The average polymorphic information content values of silicoDArT and SNP markers were 0.07 and 0.08, respectively. For further analysis, the allele frequency for each SNP site was calculated and filtered with MAF < 0.01 and left 2997 high-quality SNPs evenly distributed across the whole genome that could be used for subsequent analysis. All genotypes used in this study were descended from eight geographical origins. The genetic diversity analysis showed that the average nucleotide diversity of the panel was 0.14. Considering the genotypes based on their geographical origin, Africa collections (0.21) as a whole without Ethiopian collection was more diverse than Asia and when further portioned Africa, North Africa (0.23) collection was more diverse than others, but at the continent level, Asia (0.17) was more diverse than Africa (0.14). The genetic distance among the sesame populations was ranged from 0.015 to 0.394, with an average of 0.165. The sesame populations was clustered into four groups. The structure analysis divided the panel into four hypothetical ancestral populations and 21 genotypes were clustered as an admixture. These indicates genotypes from the same origin didn't classify properly on the premise of the country of origin. The genetic diversity and population structure revealed in this study should guide the future research work to design association studies and the systematic utilization of the genetic variation characterizing the sesame panel.

Keywords: *DArTseq, silicoDArT, SNP, Genetic diversity, and Population structure*

INTRODUCTION

Sesame (*Sesamum indicum L*., 2n = 26), is a member of the Pedaliaceae family, and one of the ancient oil crops is grown widely in tropical and subtropical areas since the time of long past (Bedigian and Harlan, 1986; Ashri, 1998). Archeological findings revealed that the cultivated sesame traces back its progenitor to the wild populations native to South Asia (Fuller, 2003). The sesame has been cultivated in South Asia since the time of the Harappan civilization from where it was spread west to Mesopotamia before 2000 B.C. (Fuller, 2003). Others believed that the crop was first cultivated in Africa and later taken to India at a very early date (Simmonds and Purseglove, 1969; Alegbejo *et al.*, 2003). Still, others proposed that sesame was the main oil crop grown by the Indus Valley Civilization from where it was likely transferred to the Mesopotamia around 2500 B.C (Tunde-Akintude *et al*., 2012).

Sesame is produced in different parts of the world for various purposes but more than 96% of the world sesame seed production is covered by Africa and Asia (FAO, 2017). Sesame seeds are good sources of fat, protein, carbohydrates, fiber, and essential minerals. Seeds are chemically composed of 44–57% oil, 18–25% protein, and 13–14% carbohydrates (Borchani *et al.,* 2010). Sesame, also referred to as "queen of oilseeds", is employed in sweets such as sesame bars and halva (dessert), and bakery products or milled to get high-grade edible oil (Bedigian, 2004).

In Ethiopia, sesame is among the foremost important oil crops both in terms of area coverage and total national annual production (CSA, 2019). However the farm level productivity of sesame in Ethiopia is low (0.7t ha^{-1}) (CSA, 2019) compared to the genetic potential of the crop yield of 2 t ha^{-1} (Mkamilo and Bedigian, 2007) and other

countries like Egypt (1.29 t ha^{-1}), Nigeria (1.1 t ha^{-1}) Tanzania (1 t ha^{-1}), and China (1.4 t ha^{-1}) (Sharaby and Butovchenko, 2019). Improved varieties released in Ethiopia are reported to yields ranging from 0.3 to 1.3 t ha^{-1} under rain fed and 1 to 2.4 t ha^{-1} under irrigation on research stations (Gebremichael, 2017)

Ethiopia is considered one of the centers of genetic diversity of sesame crop and has an immense wealth of genetic diversity in the germplasm collections that can be exploited through genetic improvement in the breeding program (IBC, 2012). This fact also describes from the time of N.I. Vavilov (Vavilov, 1951) based on morphological level studies (Sileshi, 2008; Gidey *et al*., 2012; Yirgalem, 2013;Teklu and Kebede, 2014; Abate and Mekbib, 2016). The effective utilization of these collections requires a systematic genetic characterization, assessment of genetic diversity, and identification of potential putative genes that can be used in the breeding program. In Ethiopia despite the availability of a large number of local collections in gene banks, the molecular characterization to study the genetic diversity of the sesame population is limited (Gebremichael, 2017). This showed the importance of characterizing the available sesame population in Ethiopia and use it in the improvement of production and productivity in the country in the breeding program.

The genetic diversity can be assessed using morphological, biochemical, and molecular markers. Several studies employ molecular markers to assess genetic diversity using various types of markers (Gebremichael and Parzies, 2011; Abate and Mekbib, 2015b; Dagmawi *et al*., 2015). Among these markers Amplified Fragment Length Polymorphism (Laurentin and Karlovsky, 2006), Sequence-related Amplified Polymorphisms (Zhang *et al.*, 2010, 2012), Inter simple Sequence Repeat (Kumar *et al.*, 2012), Simple Sequence

Repeats (Park *et al*., 2011), expressed sequence tag (Farshadfar and Farshadfar, 2008; Zhang *et al.*, 2012), and insertions and deletions (Wu *et al.*, 2014) have been used elsewhere for the analysis of germplasm genetic diversity.

More recently high thorough put marker systems particularly single-nucleotide polymorphisms and Diversity Arrays Technology (DArT) markers have become the genetic markers of choice for genetic analyses including characterization of germplasm because of the efficiency, low cost, speed, and abundance in the genome (Gupta *et al*., 2001; Gupta *et al*., 2008; Wei *et al.*, 2015; Cui *et al.*, 2017; Basak *et al.*, 2019). SilicoDArT markers are dominant microarray markers and scored for the presence or absence of a single allele, whereas DArTseq based SNPs are co-dominant markers, both of them being successfully applied in genetic diversity (Wenzl *et al.*, 2004; Yang *et al.*, 2006; Bolibok-Bragoszewska *et al.*, 2009; Sánchez-sevilla *et al*., 2015; Tang *et al.*, 2015) and population structure (Matthies *et al.*, 2012; Laidò *et al.*, 2013) study of several crop species.

Even with all these technologies available to study genetic diversity and the importance of sesame crop as nutritional and economic importance in several parts of the world little research work was done at the national and international levels (Bedigian and Harlan, 1986; Bhat *et al*., 1999; Bedigian, 2010). This showed the importance of developing research activities to assess the genetic diversity of the sesame germplasm collections available in the Ethiopian Biodiversity institute and farmers' hands to exploit it in the future breeding program to solve the sesame production constraint in the country.

Therefore, this research work was done to study the genetic diversity of the sesame landraces, introduced and released varieties in the breeding program using SNPs markers

developed using the DArT platforms with the objective of 1) assessing the genetic diversity of the existing sesame populations (landraces, introduced and released varieties) in Ethiopia, 2) understand the population structure of the sesame population and define direction how the available diversity can be exploited in the sesame breeding program in the country.

MATERIAL AND METHODS

Plant materials

A total of 300 genotypes comprised of 209 Ethiopian landraces, 75 exotic collections, and 16 released varieties were used in this study. The Ethiopian landraces were collected from areas between 1931 and 2008 a.m.s.l and distributed in Amhara (56), Benshangul-Gumz (BG) (38), Oromia (52), SNNP (3), and Tigray (60) regions. In this study, 16 varieties released between 1942 and 2014 were also included. The introduced germplasms were obtained from North Africa (27), South Africa (18), West Africa (17), and East Africa (6) without including the Ethiopian collection and Asia (7). The germplasms were kindly provided by the Ethiopian Biodiversity Institute (EBI) and regional and federal research centers and the geographical location of the collections is presented in (**Figure 5.1** and Appendix table 1).

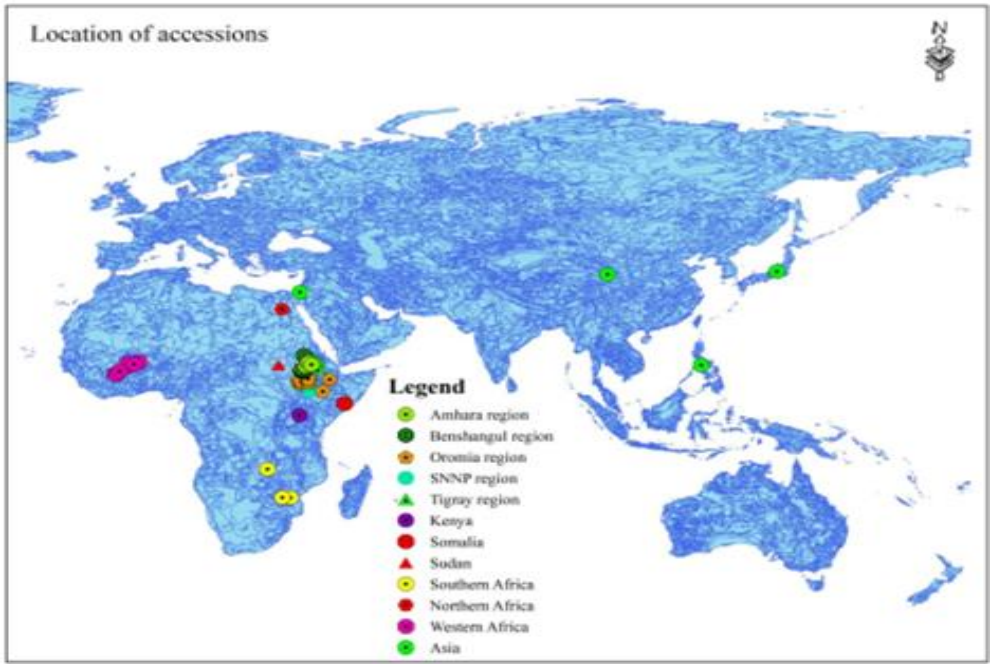


Figure 0-1. Map of collection areas of sesame genotypes

DNA extraction

The DNA of the sesame genotypes is extracted from the seed obtained in the previous harvest and the seed was crushed using Geno/Grinder to produce the power for DNA extraction. Then 800µl Lysis buffer was added to the sample of each genotype powder for the tan bead DNA extraction process. Lysis buffer is a solution for the purpose of breaking open cells for use in molecular biology experiments that analyze the labile macromolecules of the cells (e.g. western blot for protein, or for DNA extraction). Most lysis buffers contain buffering salts (e.g. Tris-HCL) and ionic salts (e.g. NaCl) to regulate the PH and osmolarity of the lysate. Those samples were incubated for 1 hour at 65°c then centrifuge for 5 minutes to remove plant tissue debris. The lysate was taken and load on column #1 and the nucleic acid of the samples was extracted with Automated Nucleic acid Extractor (Maelstrom series). During the process, the silicon dioxide layer coated on the magnetic beads adsorb nucleic acid from samples, remove contaminants with wash Buffer, and elute purified genomic DNA by Elution Buffer. At the end of the program, collected Nucleic acid was found in column #6 with a clean tube. DNA quality was evaluated on 0.8% agarose gels and it was adjusted to 50 ng/µl for GBS analysis.

Genotype by sequencing (GBS) library preparation and sequencing

DArTseq combines genome complexity reduction methods and next-generation sequencing platforms (Courtois *et al.*, 2013; Cruz *et al*., 2013; Raman *et al.*, 2014; Kilian *et al*., 2016). Therefore, DArTseq represents a new implementation of the sequencing of complexity-reduced representations (Huang *et al.*, 2014) and more recent applications of this concept on the next-generation sequencing platforms (Sonah *et al.*, 2013; Bastien *et*

al., 2014). DArTseq libraries (96-plex) were prepared for the 300 accessions using 50 ng of DNA each. Briefly, DNA samples were digested individually with *Pst*I- *Mse*I restriction enzymes. In this technology, the *Pst*I-based complexity reduction method (Wenzl *et al.*, 2004) was applied for the enrichment of genomic representation with single-copy sequences. This method involved the digestion of DNA samples with a rare cutting enzyme *Pst*I, paired with a set of secondary frequently cutting restriction enzyme *Mse*I, ligation with site-specific adapters, and amplification of adapter-ligated fragments. Post digestion with a *Pst*I-*Mse*I pair, a *Pst*I overhang compatible oligonucleotide adapter (5′-CAC GAT GGA TCC AGT GCA-3′ annealed with 5′ -CTG GAT CCA TCG TGC A-3′) was ligated, and the adapter-ligated fragments were amplified in adherence to the prescribed standard procedures (Wenzl *et al.*, 2004). To develop SNP and silicoDArT markers, the DArTseq technology was optimized using replacing a single *PstI*-compatible adapter with two different adapters corresponding to two different restriction enzymes (RE) overhangs. The *PstI*-compatible adapter was designed to include Illumina flowcell attachment sequence, sequencing primer sequence, and staggered varying length barcode regions. The reverse adapter contained the flowcell attachment region and *MseI*-compatible overhang sequence. Only "mixed fragments" (*PstI–MseI*) were effectively amplified in 30 rounds of PCR using the following reaction conditions: 1 min at 94°C for initial denaturation; 30 cycles each consisting of 20 s at 94°C for denaturation, 30 s at 58°C for annealing and 45 s at 72°C for the extension, and finally a 7 min extension step at 72°C. Total genomic DNA was extracted by adhering to the modified CTAB protocol (DoyleDoyle, 1990; Baloch *et al*., 2016) as following the procedures described by Kilian *et al.* (2012b).

Post-PCR, cluster generation was carried out in cBOT (Illumina) according to the procedures described by the manufacturer. Briefly: 10nM DNA of each library is denatured, diluted in hybridization buffer, loaded into the machine, and clusters are generated in the flow cell by cBOT with use of the set cBOT reagents (Bridge Amplification). During cluster generation, the molecules of each library were attached to the flow cell surface and amplified to form clonal clusters.

Next-generation sequencing technology was implemented using the sequencer HiSeq2500 (Illumina, USA) to detect SNPs and silicoDArT markers. The flow cell with clusters generated in the previous step (cBOT) is loaded to the HiSeq 2500 together with the sequencing reagents. HiSeq 2500 performed sequencing according to user-selected sequencing parameters. All amplicons were sequenced in a single lane. The single-read sequencing was run for 77 cycles.

Real-time Analysis (RTA) happened simultaneously to the sequencing run and RTA data were outputted to a server. The main sequence outputted data were base calling files *.bcl files. These files were the input files for downstream data conversion. The primary workflow was a custom build software for downstream processing of *.bcl files. The first step was a conversion of *.bcl files which was done by Illumina bc12fastq software embedded in the primary workflow, the second step performed two functions at the same time: first using target definition from DArTdb the software splits sequencing reads according to the barcode sequence (demultiplexing), Secondly, it removed reads below quality filters. Two quality filters were applied: more stringent for barcode sequence and less stringent for the remaining part of the sequencing read.

Finally fold compression of the sequence tags was copied to DArTdb (Diversity Arrays Technology data base, Australia) for permanent storage. We extracted compressed sequence tags from DArTdb and load them to DArTsoft14 for marker data extraction. DArTsoft14 extracts two types of marker data: SNPs and SilicoDArTs. SilicoDArTs represent dominant markers and is scored in a binary format "1"= Presence and "0"= Absence of restriction fragment with the marker sequence in the genomic representation of the sample. **"-"** represents calls with non-zero counts but too low to score confidently as "1" (often representing heterozygotes). Single Nucleotide Polymorphism (SNPs) can be defined as a variation in the base composition of a single nucleotide position within a specific locus of a single chromosome of the haploid set. In standard format, SNPs markers were presented for reference and SNP alleles for each marker and genotype. This format of SNPs can be converted to other formats if required. The report was prepared as binary or read counts file, or both depending on the order specifications. Two technical replicates of the DNA samples of each of 21 accessions were genotyped to calculate the reproducibility of the marker data. Thereafter, the SNPs and SilicoDArTs obtained were run against the sesame reference genome database

(https://www.ncbi.nlm.nih.gov/genome/?term=sesame) to understand on which chromosomes of sesame the SNPs and SilicoDArTs were located. . In addition the updated genome assemble were consider for linkage group (Wang *et al.*, 2016). Genotyping of the materials was carried out at the Biosciences Eastern and Central Africa- International Livestock Research Institute (BecA- ILRI) in Kenya.

Quality analysis of marker data

The markers were tested for reproducibility (%), call rate (%), polymorphism information content (PIC), one ratio, and minor allele frequency (MAF). Scoring of reproducibility involved the proportion of technical replicate assay pairs for which the marker score exhibited consistency. The call rate determined the success of reading the marker sequence across the samples and was estimated from the percentage of samples for which the score was either '0' or '1'. PIC is the degree of diversity of the marker in the population and showed the usefulness of the marker for linkage analysis. One ratio constitutes the proportion of the samples for which genotype scores equaled '1'.

Data analysis

DArTseq markers were mapped using the consensus map version 4.0 (www.diversityarrays.com) developed by DArT Pty. Ltd., Australia, and the updated genome assembly and annotation issued from the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, available online at https://www.ncbi.nlm.nih.gov/genome/?term=sesame and (Wang et al., 2016)

Data cleaning and genetic diversity analysis:

DArTseq raw data were filtered according to markers criterion; minor allele frequency > 0.01% and missing data ≤ 25%. The summary statistics of the filtered DArTseq markers such as the expected heterozygosity (He) or genetic diversity (GD), minor allele frequency (MAF), and the polymorphic information content (PIC), were calculated using Power Marker v 3.25 (Liu and Muse, 2005). PIC was estimated based on the probability

of finding polymorphisms between any two random samples while gene diversity defined as the probability of two randomly chosen alleles from the population is different.

Genetic diversity analysis

The genetic distance between the tested 300 sesame genotypes and between the population were estimated using the Euclidean and Nei's genetic distance (Nei, 1972) using the R software KD compute plugin system.

Then the clustering analysis was done based on the Euclidean genetic distance and the ward D^2 agglomeration (Ward, 1963) method according to the KD compute plugin system and similarly the dendrogram was developed. In line with it, the principal component analysis was done to assess the distribution of the genotypes in biplot using the KD compute plugin system.

Allele No, Gene Diversity, Heterozygosity, PIC and Major Allele Frequency were estimated using Power Marker genetic analysis package (version 3.25); (Liu and Muse, 2005). The software Arlequin V3.5 (Excoffier and Lischer, 2010) was used to calculate the genetic variation between and within geographical groups with an analysis of molecular variance (AMOVA).

Population structure

The Bayesian clustering analysis approach was used to analyze the genetic structure of the sesame genotypes using STRUCTURE v.2.3.4 (Pritchard *et al*., 2000) software. For the analysis five individual Markov Chain Monte Carlo (MCMC) simulations were conducted for each K-value from 1 to 11 with a burnin length of 50,000, followed by

100,000 iterations. The admixture model was applied without using any prior population information and correlated allele frequencies were also employed.

The structure results were subsequently analyzed by the structure harvester application (Earl and vonHoldt, 2012) (*http://taylor0.biology.ucla.edu/structureHarvester/*) to identify a distinct peak in the change of likelihood (ΔK) at the true value of K. CLUMPAK: "a program for identifying clustering modes and packaging population structure inferences across K" (CLUMPAK server) was used. Each sesame accession was then assigned to a cluster (k) based on a membership coefficient determined by STRUCTURE V2.3.4, The cut-off probability for assignment to a cluster was 0.50 for the clusters.

RESULTS

Marker discovery by DArTseq

A total of 6115 polymorphic silicoDArT and 6474 SNP markers were generated of which 5065 silicoDArT and 5821 SNP were aligned with the Reference sesame genome obtained from https://www.ncbi.nlm.nih.gov/genome/?term=sesame and updating genome assembly and annotation available at Wang *et al.* (2016) and the remaining were scaffold, and unknown markers (Table 5.1).

Table 0.1. Distribution of DArTseq markers on different sesame chromosomes.

Chromosome no	Chromosome size (Mb) *	Number of silicoDArT markers	silicoDArT marker/Mbp	No of SNP markers	SNP marker/Mbp
1	20.26	391	19.35	539	26.60
2	18.42	407	22.10	483	26.22
3	25.85	545	21.08	733	28.36
4	20.59	321	15.59	378	18.36
5	16.58	397	23.94	381	22.98
6	25.97	643	24.80	662	25.49
7	16.76	234	13.90	238	14.20
8	26.18	492	18.79	539	20.59
9	22.85	505	22.10	533	23.33
10	19.49	332	17.03	387	19.86
11	14.05	301	21.42	399	28.40
12	16.28	255	15.60	315	19.35
13	16.47	242	14.69	234	14.21
Scaffold	-	724	-	305	-
Unknown	-	326	-	348	-
Total	259.73	6115	23.54	6474	24.93

*Indicates the chromosomal size taken from the reference genome published by Wang *et al.* (2016)

After removing those markers that had missing rates >0.25, and one ratio ≤ 0.5 a set of 5002 silicoDArT and 4638 SNPs were generated. For further analysis the allele frequency for each SNP site was calculated and filtered with minor allele frequency (MAF): the MAF of the SNPs varied from 0 to 49.6%, with an average of 5.1%, and ~61.29% of the SNPs had a low frequency (MAF < 0.05) across the 300 accessions. After excluding the SNPs with a MAF < 0.01, 2997 were left (~64.61%) high-quality SNPs evenly distributed across the whole genome that could be used for subsequent analysis.

The aligned markers were distributed on all 13 chromosomes of sesame with an average of 389.62 silicoDArT and 447.7 SNP markers per chromosomes. The maximum number of silicoDArT (643) and SNP (733) were found on chromosome 6 and 3 respectively; (**Figure 5.2, Figure 5.3 and Table 5. 1**).

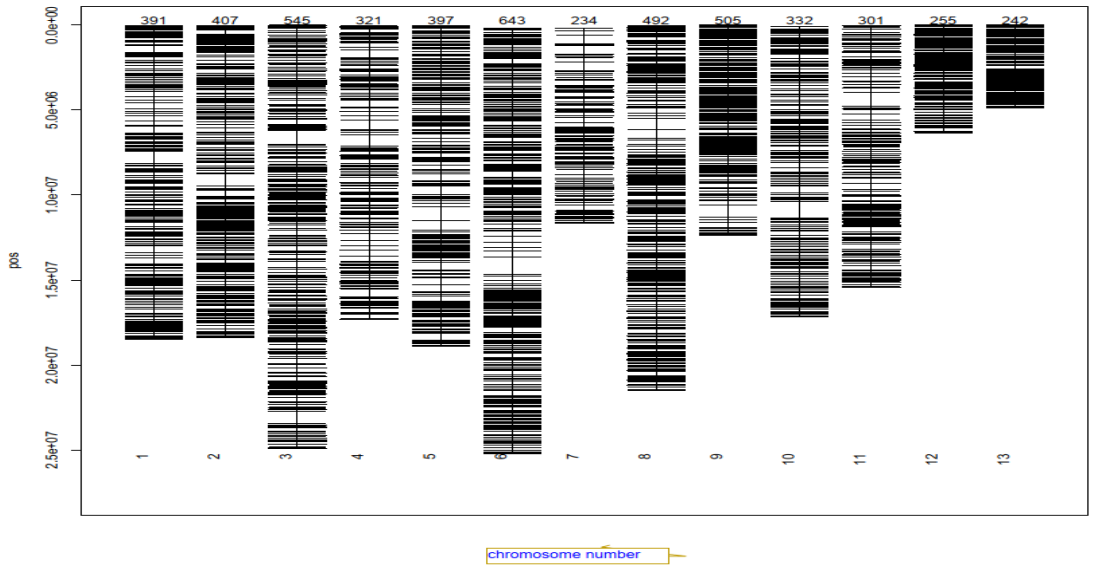


Figure 0-2. Distribution of DArTseq silicoDArT markers on different chromosomes of sesame

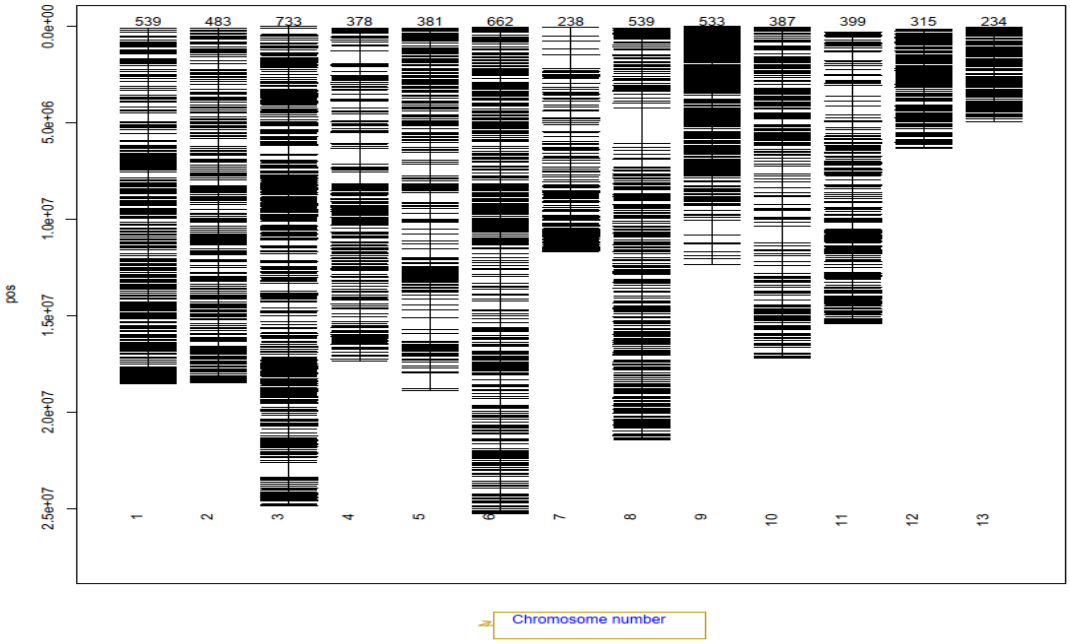


Figure 0-3. Distribution of DArTseq SNP markers on different chromosomes of sesame

Analysis of genetic diversity

The number of accessions, number of alleles, genetic diversity, heterozygosity, and the polymorphism information content (PIC), and major allele frequency of the eight populations (Africa, Amhara, Asia, Benshangul-Gumz (BG), Improved, Oromia, Southern Nations, Nationalities, and People's Region (SNNP), and Tigray respectively) are shown in Table 5.2.

The genetic diversity study showed the sesame germplasm from Asia are more diverse than Africa (Table 5.2) as expressed with high PIC (0.14), Heterozygosity (0.19), and Genetic diversity index (0.17). In contrast the number of germplasm used in this study from Asia continent in only 7 as compared to the African continent (293). That indicates the potential of the Asian germplasms in the future sesame breeding program in Ethiopia.

The comparison of the genetic diversity among the different African regions showed the North African showed high genetic variability followed by South Africans in contrast to the low genetic variability registered in East Africa (Table 5.2). In Africa, the highest number of genotypes were considered from Ethiopia/East Africa but still show the lowest genetic diversity. When we compare the regions in Ethiopia, the maximum genetic diversity was observed in Tigray region (PIC = 0.12, He = 0.12, GD = 0.13) followed by Amhara region (PIC = 0.09, He = 0.09, GD = 0.10). Unexpected result observed is the relatively high genetic diversity among the improved and released varieties of sesame in Ethiopia as compared to the landrace collected from different regions of the country (Table 5.2).

Table 0.2. Summary of the genetic diversity of the 300 sesame accessions based on their different geographical regions

Group	No of accessions	Allele No	Gene Diversity (GD)	Heterozygosity	PIC	Major Allele Frequency	GD with 3% Missing and 0.05 MAF
Introduced from Asia	7	1.6	0.17	0.19	0.14	0.88	0.24
Introduced from different Africa countries except Ethiopia	68	1.97	0.21	0.15	0.18	0.86	0.26
Amhara	56	1.82	0.10	0.09	0.09	0.94	0.15
BG	38	1.68	0.07	0.07	0.06	0.96	0.11
Improved	16	1.59	0.12	0.12	0.10	0.93	0.18
Oromia	52	1.7	0.06	0.06	0.06	0.96	0.096
SNNP	3	1.24	0.08	0.08	0.06	0.95	0.10
Tigray	60	1.85	0.13	0.12	0.12	0.92	0.19
Continents							
Asia	7	1.6	0.17	0.19	0.14	0.88	0.24
Africa	293	2.0	0.14	0.10	0.12	0.92	0.18
By partition Africa							
E. Africa/ Ethiopia	231	1.97	0.10	0.09	0.09	0.94	0.15
N. Africa	27	1.86	0.23	0.19	0.19	0.83	0.26
S. Africa	18	1.75	0.15	0.13	0.13	0.9	0.18
W. Africa	17	1.7	0.14	0.11	0.12	0.9	0.18
Mean	300	2.0	0.14	0.11	0.12	0.92	0.19

PIC= polymorphic information content

Genetic relationships among Germplasm

The Nei's genetic distance estimate among the 300 sesame germplasms evaluated in this study showed the highest value between "Najjo-68 (gabaa kamijaa)" and "17712"

landraces from Oromia regions of Ethiopia. This shows the possibility of identifying some divergent genotypes within Ethiopian sesame populations. When we consider the eight populations evaluated, the maximum genetic distance was observed between the Asia and SNNP regions of Ethiopia (0.073) followed by Asia with Oromia region (0.057) (Table 3). Whereas the minimum genetic distance was observed between Oromia with BG regions of Ethiopia (0.004).

Table 0.3. Pairwise population Nei's genetic distance showing the magnitude of genetic differentiation between sesame populations from different sources

	AFRICA	AMHARA	ASIA	BG	IMPROVED	OROMIA	SNNP	TIGRAY
AFRICA	0.000							
AMHARA	0.021	0.000						
ASIA	0.040	0.052	0.000					
BG	0.020	0.006	0.056	0.000				
IMPROVED	0.022	0.012	0.037	0.015	0.000			
OROMIA	0.021	0.010	0.057	0.004	0.014	0.000		
SNNP	0.030	0.020	0.073	0.016	0.028	0.013	0.000	
TIGRAY	0.024	0.006	0.047	0.015	0.009	0.017	0.027	0.000

Cluster analysis of the 300 germplasms derived from the eight different geographical origins was performed using the Ward D agglomeration based on the Euclidean genetic distance and clustered into four major cluster groups (Figure 5.4).

The first cluster comprised the majority of the germplasms from different countries of Africa (28), all accessions that were introduced from Asia (7) and the different regions of Ethiopia, Amhara (8), Benshangul-Gumz (4), Oromia (10), SNNP (1), Tigray (12), and 7 improved varieties. The second cluster constitutes the highest number of accessions that were collected from the different regions of Ethiopia, Amhara (40), Benshangul-Gumz

(34), Oromia (41), SNNP (2), Tigray (23), and 4 improved varieties, the remaining 13 accessions were introduced from different Africa countries. The third cluster is comprised of majority from the Tigray region (25) and a small number from Amhara (n=8), Oromia (1), and 5 Improved varieties, the remaining 4 accessions were introduced from different countries of Africa. Cluster 4 comprised all the accessions that were introduced from only one African country, Egypt (23) only (**Figure 5.4**). The clustering did not follow any country origin and pedigree information but only the Egypt collections clustered in a separate group (Cluster IV).

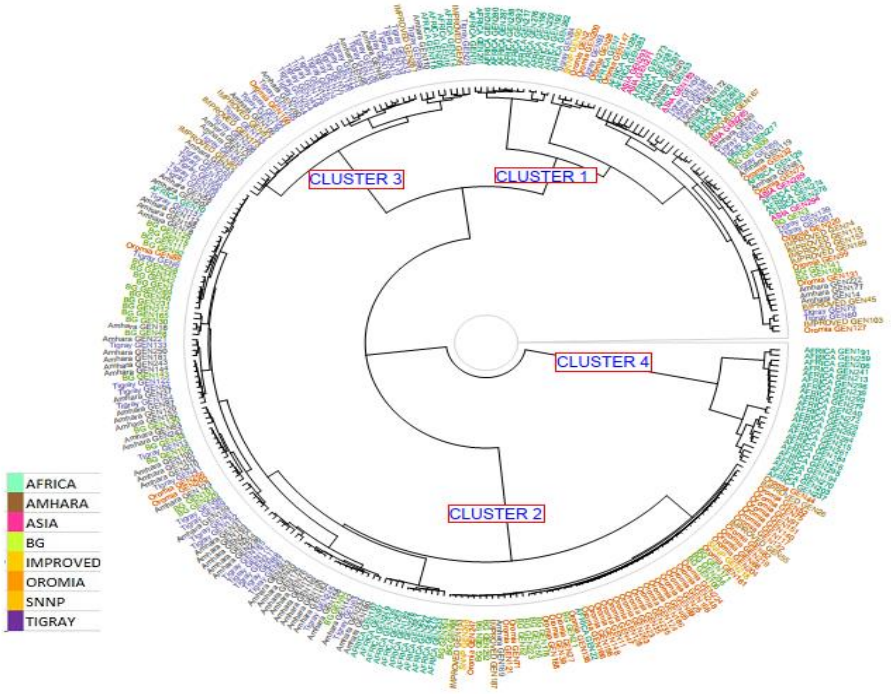


Figure 0-4. UPGMA dendrogram showing the genetic relationships among 300 sesame genotypes grouped into four distinct clusters.

Note;- Each geographical origin represented by a colored; green for Africa, teal for Amhara, red violet for Asia, BG: yellow green for Benshangul-Gumz; orange for Improved, red orange for Oromia, yellow orange for Southern Nations, Nationalities, and People's Region; blue violet for Tigray.

An analysis of molecular variance (AMOVA)

Analysis of molecular variance (AMOVA) among the 300 sesame germplasms based on eight geographical origins indicated that 8.31% of the variance was differentiation among the populations, 15.24% of the variance was accounted by genetic differentiation among individuals within populations, while the remaining 76.44% of the variance was due to the differences within individuals (**Table 5.4**).

While in terms of continents 11.49% of the total molecular variation observed due to differentiation between Asia and Africa, 19.45% of the variance was accounted by genetic differentiation among individuals within continents, while the remaining 69.06% of the variance was due to the differences within individuals (**Table 5.4**).

When we see further, In terms of population subdivision with different regions of Africa and Asia 22.17% of the total molecular variation observed was due to differentiation between different regions of Africa and Asia, 10.69% of the variance was accounted by genetic differentiation among individuals within different regions of Africa and Asia, while the remaining 67.12% of the variance was due to the differences within individuals (Table 5.4).

Table 0.4. Analysis of molecular variance (AMOVA) among and within sesame subpopulations

Source of variation	df	SS	Variance components	Percentage variation
Based on different 8 Geographical origins				
Among populations	7	9643.614	17.41032	8.31858
Among individuals Within populations	292	61461.430	31.89832	15.24089
Within individuals	300	46195.500	159.98566	76.44052
Total	599	117300.544	209.29431	
Based on the Continents				
Among continents	1	941.728	26.62029	11.49050
Among individuals Within continents	298	70163.316	45.06621	19.45258
Within individuals	300	46195.500	159.98566	69.05692
Total	599	117300.544	231.67217	
Based on Different regions of Africa and Asia				
Among populations	4	12564.718	52.85675	22.17834
Among individuals Within populations	295	58540.327	25.48356	10.69273
Within individuals	300	46195.500	159.98566	67.12892
Total	599	117300.544	238.32598	

df = degree of freedom, SS = Sum of Square

Population structure

The Bayesian model-based cluster analysis was conducted using the STRUCTURE program. The appropriate number of cluster K was determined, As *K* changed from 1 to

11 by inferring on Delta K of Evanno *et al*. (2005), the log-likelihood value [LnP(D)] increased continuously and inflection was evident when *K* increased numerically from 1 to 4 (Fig 5.5A). Thus, the most likely numerical value of *K* was 4. The number of hypothetical ancestral populations (*K*) was further validated by the second-order statistics of ΔK. The ΔK value showed a peak at *K* = 4 (Fig. 5.5B), which supported the classification of the panel into four major hypothetical ancestral populations (Figure 5.5C). The genetic diversity within each population was explained through the estimation of the expected heterozygosity, which varied from 0.06 (POP2) to 0.31 (POP4). The genetic divergence among the populations revealed by Nei's net nucleotide distance (D) indicated that a higher distance between POP3 and POP4 (0.22) and the genetic distance observed between POP1 and POP2 (D = 0.09) was the least among the pairs of populations. Mean fixation index of sub-populations ranged from 0.39 (POP4) to 0.77 (POP2) (Table 5.5).

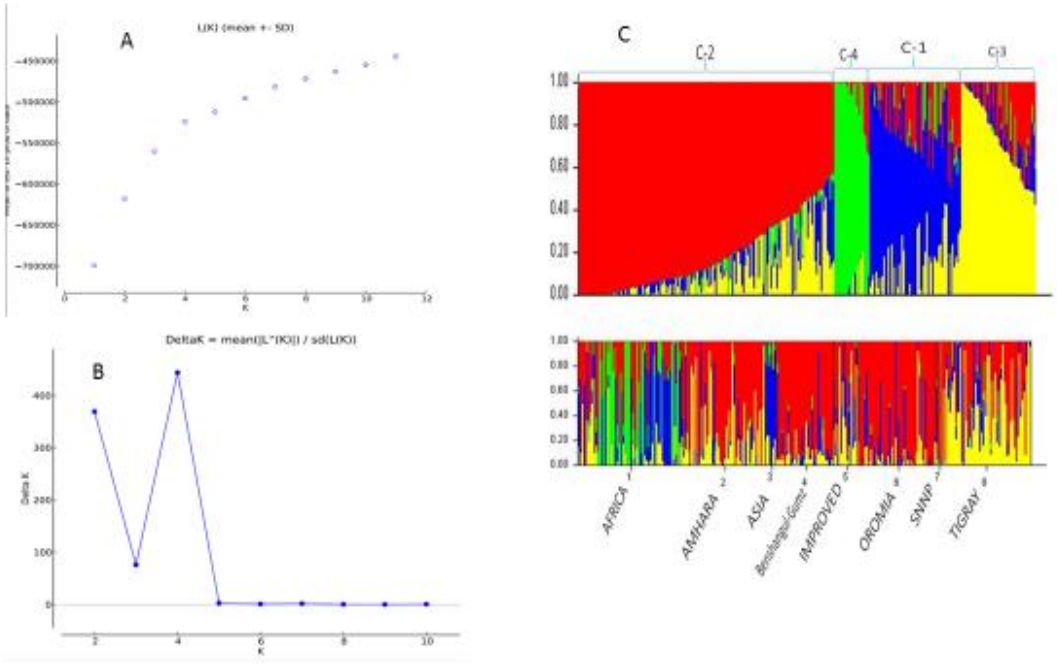


Figure 0-5. Analysis of the population structure of the 300 sesame accessions using STRUCTURE.

Note; - (A) Estimated LnP(D) of possible clusters (K) from 1 to 12; (B) ΔK based on the rate of change of LnP(D) between successive K; (C) population structure based on K = 4 and each individual is represented by a vertical bar partitioned into four colored segments, with their respective lengths representing the proportion of the individual's genome in a given group.

Table 0.5. Genetic divergence among (net nucleotide distance) and within (expected heterozygosity) population, proportion of membership, and mean value of Fst observed from the study of the population structure of 300 sesame accessions and genotypes using DArTseq-SNP markers.

Population	Net nucleotide distance			Expected Heterozygosity	% of Membership	Mean Fixation Index (Fst)
	pop2	pop3	pop4			
pop1	0.09	0.13	0.19	0.22	0.24	0.45
pop2		0.11	0.17	0.06	0.50	0.77
pop3			0.22	0.18	0.19	0.57
pop4				0.31	0.07	0.39

POP1, POP2, POP3 and POP4 represents the four sub-populations that grouped based on structure software

According to the structure analysis, based on the probability of membership threshold of 50%, 54, 159, 43, and 23 genotypes were respectively assigned into the four hypothetical ancestral populations, Pop 1, Pop 2, Pop 3, and Pop 4, while the remaining 21 accessions showed admixture among other subgroup (Appendix table 2).

Most accessions of Pop 1 introduced from different countries of Africa (27), Asia countries (7), while 18 accessions from Ethiopia, Amhara (n =4), Benshangul-Gumz (2), Oromia (5), Tigray (7) and 2 improved varieties. The accessions and genotypes of Pop 2 constitute the largest that was mainly collected from the different regions of Ethiopia, Amhara (n=40), Benshangul-Gumz (35), Oromia (42), SNNP (2), Tigray (20), and 7 Improved varieties, the remaining 13 accessions were introduced from different Africa countries. The accessions of Pop 3 comprised mainly from three regions of Ethiopia, Amhara (n=9), Oromia (1), Tigray (26), and 4 Improved varieties, the remaining 3 accessions were introduced from different countries of Africa. Pop 4 introduced from one

of the African countries Egypt (23) only. For the Mixed group, 19 accessions were collected from different regions of Ethiopia and 2 accessions from two Africa countries.

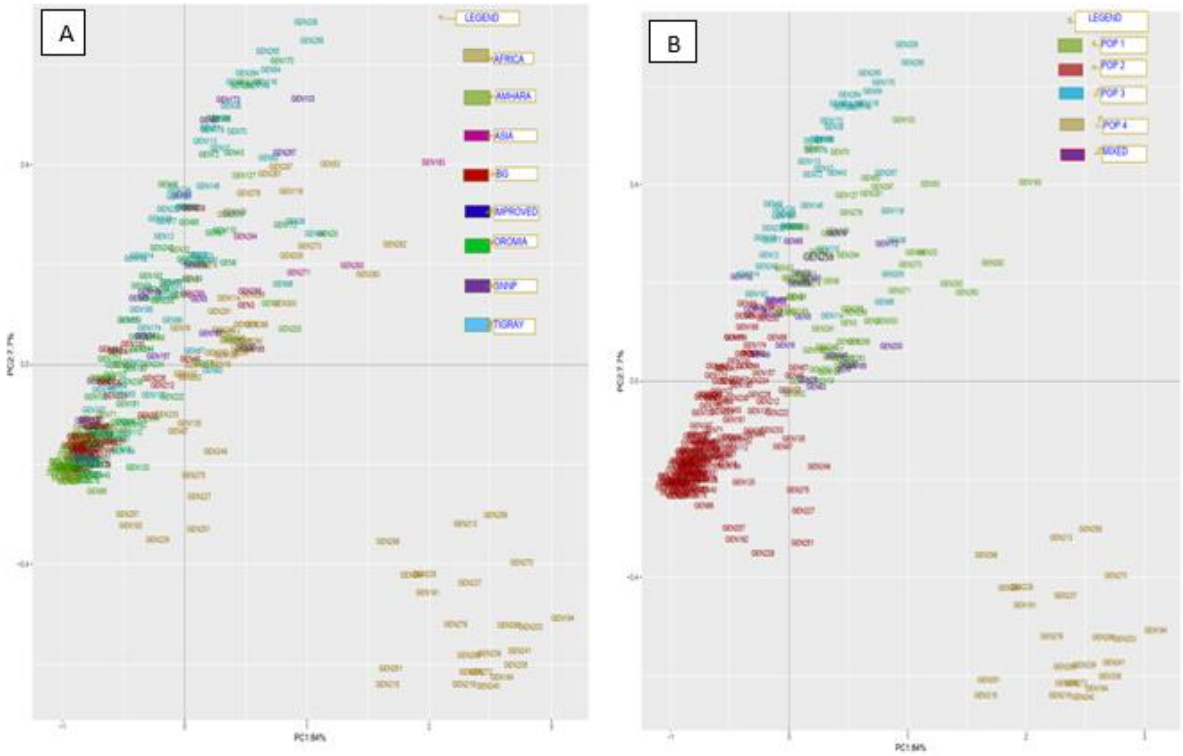


Figure 0-6. Principal components analysis (PCA) of the population for 300 sesame accessions based on 2997 single-nucleotide polymorphisms (SNPs).

Note; - Each individual is represented by GEN number, with its symbol color corresponding to the assigned subgroup classification. (A) PCA plots of the sesame germplasm collection based upon their geographic origins. (B) PCA plots of the same sesame germplasm collection but now based on hypothetical ancestral populations as identified by STRUCTURE.

The Principal component analysis (PCA) based on DArTseq - SNP markers revealed four distinct groups of sesame genotypes and the first two principal components, accounting for 93.7% of total variation (Figure 5.6A). PC1 explained 84% of the genetic variation found, while PC2 explained 7.7% of the variation, respectively. However, some

intermediate lines (admixture) made the grouping less than clear-cut. When considering these intermediate lines, the panel could be neatly divided into four clusters (**Figure 5.6B**) corresponding to the four hypothetical ancestral populations as inferred by using the STRUCTURE result.

DISCUSSION

In our study, the silicoDArT and SNP markers distributed across the genomes and provide better genome coverage that is correlated with gene density (Dierig *et al*., 2009; Kilian *et al.*, 2012a). This showed both; - silicoDArT and SNP markers used in our study are better suited for genetic diversity studies, association/linkage mapping, and/or sequence-based physical mapping in sesame. In comparison with the other existing marker technologies like microsatellite markers, DArT markers are pertinent to high-throughput work and have merits in terms of cost-effectiveness and time aspect (Kilian *et al.*, 2003) Since it provides high information with low cost.

Then, 2997 SNP markers were filtered with a call rate of 75%, and those having >0.01 minor allele frequency were used for the analysis, the proportion of rare SNPs (i.e., MAF < 0.05) we examined amounted to ~61.29% %, which was similar to those reported for the genomes of sesame (Cui *et al.*, 2017) a high proportion of rare SNPs In our study may be caused due to the broad genome coverage obtained by GBS technology that is less prone to bias than low-coverage sequencing data (Wei *et al.*, 2012) and the collection of large numbers of less abundant sesame landraces by the Ethiopian Biodiversity Institute. However, studies showed that the rare SNPs might also have control over the expression of a particular phenotype (Song *et al.*, 2015). Providing that the number of individuals

with a specific genotype will be very small, the effect of rare alleles on genome mapping could extend beyond the effect of just small population sizes. In such cases, increasing the number of individuals with rare alleles could improve the ability to check these rare alleles.

The average value of genetic diversity (0.14) was lower in the present study than in the earlier reports for the sesame collections analyzed with SNPs markers (Zhang *et al.*, 2010; Park *et al.*, 2011; Cui *et al.*, 2017; Basak *et al*., 2019) and SSR markers (Wei *et al.*, 2014; Dossa *et al.*, 2016). In contrast to the current study , high genetic diversity (0.19) was reported in sesame using different marker types (Park *et al*., 2011; Cui *et al.*, 2017). The broad range of variability among collections might be due to the composition of genetic resources under study (such as landraces, advanced breeding lines, cultivars, etc.), data filtering methods, sampling approaches, and the number of markers (Dossa *et al.*, 2016). The type of marker also influences the estimated genetic diversity where high and low genetic diversity was estimated using SSR and SNPs markers. However, the accurate consideration of genetic diversity reflected the number of loci instead of the number of alleles (Cui *et al.*, 2017).

The genetic diversity observed in Asia (0.17) sesame population is higher than Africa sesame population (0.14), even if the number of genotypes in Asian population much lower than the African populations. Laurentin and Karlovsky, (2006) reported high genetic diversity in sesame accessions collected from Asia that support our research finding.

The other result was obtained when we study the genetic diversity by considering different regions of Africa separately and Asian collections. The study showed the high

genetic diversity obtained from the North African sesame collections (0.23) that may be because this part of the region is drier than the other part of the region and the adverse environmental situation in the North African region may result in an adaptation change that reflected in the genetic diversity. The low genetic diversity of East Africa (Ethiopian) collection as compared to other regions of Africa can be attributed to the favorable and high rainfall situation suitable for sesame production. In addition, this finding showed the importance of broadening the genetic base of the sesame population in Ethiopia through hybridization with North Africa and Asia populations and introduction to maintain its market share of the quality sesame at the international level. The genetic distance matrix among the sesame populations from 8 geographical origins ranged from 0.015 to 0.394, with an average of 0.165 and populations clustered into four groups. The clustering analysis did not classify the sesame population according to the country of origins and showed high level of admixture except genotype obtained from Egypt. Similar results were reported previously indifferent sesame germplasm (Ercan *et al*., 2004; Ali, 2007; Cho *et al.*, 2011; Basak *et al*., 2019). The high level of admixture among the sesame genotypes from different countries and regions of the world showed the high level gene flow from country to country through migration, trade and seed exchange for cultivation. In line with our finding by Laurentin and Karlovsky, (2006) found no association between genetic diversity and accession origin, and they proposed that ecological and geographical factors have not played a significant role in the evolution of sesame. The AMOVA analysis in our study also supported the possibility of high rates of gene flow between regions, with a low level of genetic differentiation that reach only 8.3% and

11.49% of the total variation for different geographical regions of sesame panels and continents (Africa and Asia) (**Table 5.4**).

The clustering analysis grouped the sesame genotypes collected from different regions of Ethiopia into Cluster 2 and 3 that showed a high-level of germplasm exchange among different regions of Ethiopia consequently resulted in a high level of admixture. The other probable consequence may be the similar genotypes of sesame were dispersed to nearby countries by human activities.

The grouping of Sesame populations from Africa and Asia in Cluster I indicated a high level of genetic material exchange among the continents (Kim *et al.*, 2002) and identical establishment stock. Evidence also showed the long history of germplasm exchange among the Asia and East African countries and the trend is also continuing today through export for industrial applications (Zohary *et al*., 2012).

Besides, the high level of crossover event was observed in the sesame populations when different genotypes from different locations grown in the same location due to the high level of cross-pollination observed ranged from 5 to 60% (Wei *et al.*, 2014). This crossing could result the similarity of accessions from the eastern part of Africa and Asia. Similar patterns have also been observed by other researchers (Kim *et al.*, 2002; Laurentin and Karlovsky, 2006; Cho *et al.*, 2011).

Population Structure analysis of the sesame population

The knowledge of the population structure in the breeding population is very important and avoid any spurious associations (Flint-Garcia, 2005). Due to that different researchers conducted the assessment of population structure in sesame to understand the

structure and design future breeding programs. Ali *et al*. (2007) evaluated sesame genotypes from different parts of the world and clustered them into two major groups that aligned with their geographical origin in contrast to our finding. Similar to the above study 705 sesame accessions from different origins clustered into two cluster groups employing a neighbor-joining tree (Wei *et al.*, 2015). The population structure study of sesame genotypes were conducted by various researchers and obtained information that helps for the breeding and conservation activity of the crop (Cui *et al.*, 2017) . In agreement to our finding the germplasm collected from Asia, Europe, America, and Africa clustered into three groups without considering their geographical origin (Basak *et al.*, 2019).

In our study, the sesame populations were clustered into four cluster groups with the membership coefficient value >50%, and the clustering pattern did not follow the geographical origin of the germplasm. Among the tested genotypes 21 accessions showed the variable percent of admixture and did not cluster in any group. This evidence support the above genetic diversity study analysis and confirmed the high level of gene flow among the germplasm of the populations of different origins. The occurrence of some admixed/hybrid and introgressed hybrid genotypes indicated frequent hybridization and introgression events due to high level of cross pollination. Although the extent and significance of natural hybridization/introgression are unclear (Jarvis *et al*., 1998), new gene combinations between domestic cultivars and their wild or weedy relatives are important for the evolution of domesticated plant species (Jarvis *et al*., 1999).

The population genetic structure reflects interactions among species about their long-term evolutionary history, mutation and recombination, genetic drift, reproductive system,

gene flow, and natural selection (Slatkin, 1987; Schaal *et al.*, 1998). This was also observed in our study where the genetic structure analysis showed the genetic relationship and the possible center of genetic diversity of sesame which is the Asian region. The understanding of the extent and structure of the genetic diversity of a crop could be a prerequisite for the conservation and efficient use of the germplasm available for breeding (Mangini *et al.*, 2013). Our findings will also give information to guide our future genetic conservation strategy in sesame in Ethiopia to utilize and conserve efficiently with minimum sample high genetic variability. The various approaches (STRUCTURE, PCA, and the clustering) used in our study provide valuable information to understand the genetic diversity of the sesame population found in Ethiopia and what action will be needed to conserve and utilize the existing sesame population in the breeding program of the country.

According to the AMOVA results, 8.3% and 11.49% of the marker variation was explained among the population from different geographical regions of the sesame panel that showed a high level of differentiation among these populations. This may be resulted due to the different agro ecological conditions impacted the genetic change in African and Asian populations. The high value of within variability indicates the high level of genetic heterozygosity individuals within the population as expected in most studies.

In this study, most collections (225) were from Ethiopia and a specific collection was from West, South, and North Africa and seven collections were from 4 Asia countries. The Ethiopian sesame has useful characteristics, and often branded as ‘Humera’, ‘Gondar’ and ‘Wellega’ types, well known in the world market for their white color,

sweet taste and aroma. Even if the Ethiopian sesame populations showed unique quality characters, in the current study they clustered with germplasms of different origin. The Humera and Gondar sesame seeds are suitable for bakery and confectionary purposes and the high oil content of the Wellega sesame seed gives a major advantage for edible oil production (Wijnands and Biersteker, 2007). Collections that were introduced from a different regions of Africa and Asia showed some degree of genetic relationship and differentiation among genetic resources of Ethiopian collection. The incorporation of these germplasm in Ethiopian breeding population (genetic pool) will broaden genetic diversity that can be used as a source of favorable alleles for agronomic traits in the breeding program. (Wang *et al.,* 2019).The SNPs obtained from this collection could benefit future breeding and association mapping work in sesame.

This study showed the genetic relationship among and within the sesame populations from different origins that may be a valuable source of parental material in the future crossing program of sesame breeding. This study was also able to identify genetically divergent germplasm like Njjoo-68 and 17712 that can be used in the future crossing program to develop high yielding sesame varieties to improve the production and productivity in the country.

Chapter 6. Genome-Wide Association Study of Yield Related Traits of Sesame (*Sesamum Indicum L.)*

ABSTRACT

Sesame (Sesamum indicum L.) is an important oil crop in Ethiopia in terms of both area coverage and production. However, the farm level national average productivity is low due to biotic and abiotic stress. Therefore, dissecting the genetic basis of seed yield-related traits in sesame breeding programs is instrumental to develop stable and high-yielding sesame varieties. Genome-wide association study (GWAS) using 2997 SNPs were performed with an association mapping panel comprising a diverse 300 sesame germplasm lines after testing the panel in two environments. In total, 21 significant SNPs ($P < 1.67\times10^{-5}$) were detected for 7 yield-related traits (days to flower initiation, days to 50% flowering (DF), days to physiological maturity (DM), capsule length (CAPL), seeds per capsule (SPC), 1000-seed weight (TSW) and bacterial blight (BBL) explaining the phenotypic variation ranged from 7.02 (DF) to 16.11% (CAPL), with an average of 9.76%. All significant loci found on the sesame linkage groups (LG) 2, 6 and 11 was associated with capsule length trait except one that was associated with physiological period. The majority of significant loci detected on LG 3, 7 and 8 were associated with traits related to physiological periods (i.e., days to flower initiation, days to 50% flowering and days to physiological maturity) except the only single QTL that was associated with capsule length. Dissecting genetic control of flowering time and maturity has a pivotal contribution to foster sesame breeding and to develop new varieties able to adapt to changing climatic conditions. The QTLs detected for flowering time and maturity that strongly affects yield and plant adaptation ability would play an important role for future marker-assisted selection in sesame. Since several novel favorable alleles have been detected in this study that are not yet been intensively selected previously, our GWAS results will assist in incorporating these alleles into the elite sesame germplasm through marker-assisted selection.

Key words*;* alleles, GWAS, QTL, significant loci, yield-related traits

INTRODUCTION

Sesame (*Sesamum indicum* L.) is a major oil seed crop worldwide, with more than 95% of seed production accounted for by Africa and Asia *(FAOSTAT*, 2019). It plays a very significant role in preserving food and nutritional security as well as livelihood improvement in developing regions of the world. The cultivation of sesame offers two main advantages: it is a very rewarding crop because it can provide optimum yield and generate incomes in marginal areas where many other crops cannot grow (Langham, 2007; Dossa *et al.*, 2017). Over the last decade, the production of sesame seeds has doubled and the growing area has extended to more than 50 countries in the world, showing an ever-increasing interest in this crop. However, sesame has a very low seed yield capacity compared to other oilseed crops (Akhtar *et al.*, 2009). According to FAO, (2019), the average seed yield of sesame was only 511 kg/ha, ranked as the second lowest among the major oil crops. When we consider Africa's sesame productivity, it gives lower yield than other sesame producing continents including Asia, America and Europe (*FAOSTAT*, 2019).

In Ethiopia, sesame is an important oil crop in terms of both area coverage and production (CSA, 2019), it accounts for about 44% of the total acreage and 34% of gross production of major oilseeds cultivated in the country. It grows in almost all regions of the country with altitudes less than 2000 meter above sea level (Yebiyo, 1985; Adefris *et al*., 2011). However, the farm level national average productivity is lower in the country (0.7 t ha-1) (CSA, 2019) than the potential yield of 2 t ha-1 (Mkamilo *et al*., 2007). Improved varieties released in Ethiopia reportedly yield 0.3 to 1.3 t ha-1 under rainfed, and 1 to 2.4 t ha-1 under irrigated conditions (Gebremichael, 2017).

Therefore, understanding the genetic basis of seed yield-related traits and applying the skill acquired in sesame breeding programs would be instrumental in developing stable high-yielding sesame varieties. Since the grain yield traits are inherited quantitatively and governed by multiple genes sensitive to the environment, QTL- mapping is needed to dissect the genetics of these traits (Morrell *et al*., 2012). The high- density SNP markers based genetic mapping have been proved to be a very effective and important approach for QTLs detection in rice (Li *et al.*, 2011; Yu *et al.*, 2011; Marathi *et al.*, 2012; Gao *et al.*, 2013) and other crops (Chutimanitsakun *et al.*, 2011; Pfender, *et al*., 2011; Xu *et al.*, 2013).

Recently, the genotyping load and cost has been significantly reduced by the next-generation sequencing (NGS) technologies (Davey *et al.*, 2011). Several high-density SNP markers have been developed and a large number of quantitative trait loci (QTLs) were then detected with bi-parental linkage-based QTL mapping for agronomical important traits in sesame (Zhang *et al.*, 2013, 2016; Wu *et al.*, 2014; Wang *et al.*, 2016; Mei *et al.*, 2017). However, QTL mapping efforts using the segregated progeny of a bi-parental cross only enable the detection of a subset of loci/alleles within the crop, and offer limited resolution owing to the small number of informative recombination events between linked genetic loci (Nordborg and Weigel, 2008). As an alternative approach to traditional QTL analysis, the genome-wide association study (GWAS), taking advantage of both the wide phenotypic variation and the high number of historical recombination events in natural populations, has been used for dissecting complex traits in crop species (Guo *et al.*, 2013; Huang and Han, 2014). However, GWAS analysis for economically relevant traits in sesame is still limited. Wei *et al.* (2015) re-sequenced 705 diverse

sesame germplasm accessions and performed a comprehensive GWAS on 56 agronomic traits for the first time. From this panel, Dossa *et al.*, (2019) used 400 accessions and performed a large-scale GWAS analysis for five drought tolerance related traits. As to our knowledge, no GWAS study has been conducted for any traits using the African sesame natural populations.

The objective of the study was to employ the GWAS analysis approach to comprehensively decipher the genetic basis of 11 seed yield-related traits in natural population of African sesame and unlock potential alleles and genes for seed yield improvement based on a large and diverse sample phenotype in two environments.

MATERIALS AND METHODS

Plant Materials

A total of 300 sesame genotypes, comprising 225 local Ethiopian collections, including 16 released varieties and 75 exotic collections provided by the Ethiopian Biodiversity Institute (EBI), and different federal and regional research Centre of Ethiopia were used in this study. The collections were originated mainly from different regions of Ethiopia, and different countries of Africa and Asia.

Field Growth Conditions

The genotypes were planted at the Metema trial site (120 39'N, 360 17' E) in the 2017/18 and 2018/19 cropping season. Metema is located at 760 meters above sea level and receives 1030.2 mm of rainfall per annum and its soil is a vertisol. The study was laid out in an alpha lattice design, with each plot consisting of two rows of 4 m length with a spacing of 40 cm between rows and 10 cm between plants. Up to 65 kg ha-1 of urea, with

two times split application, was applied manually based on unpublished site specific recommendations. All quantitative traits including plant height, primary branch, secondary branch, days to flower initiation, days to 50% flowering, number of capsules per plant, seeds per capsule, 1000-seed weight, days to physiological maturity, pod bearing zone, seed yield, bacterial blight and qualitative were recorded according to the sesame descriptors list of IPGRI and NBPGR, (2004).

Measurement of yield related traits and statistical analysis

All measurements were done after tagging five randomly selected plant in each plot. Yield, growth period, plant and capsule variation and disease reactions, were recorded using standard procedures (IPGRI and NBPGR, 2004). The early flowering dates were recorded as the number of days from sowing to observation of the first flower on 50% of the individuals. Flower and leaf-related traits (days to flower initiation, days to 50% flowering and days to physiological maturity) were observed and measured in the full-bloom stage. After harvesting manually, yield-related traits (Number of carpels per capsule, Seeds per capsule, 1000-seed weight and seed yield) were measured in the laboratory. Seed yield was collected per plot and later converted into metric tonnes per hectare.

Statistical Analysis

Analysis of variance (ANOVA) and broad-sense heritability (h^2) of yield and yield related traits of 300 sesame genotypes were conducted using META-R software (Alvarado *et al*., 2016).

Broad-sense heritability (h^2) of yield and yield-related traits were estimated based on the ratio of genotypic variance to the phenotypic variance and was expressed in percentage (Falconer *et al*., 1996). It was calculated with the mean values of each experiment among the two replications according to the formula:

$$h^2(BS)\% = \frac{V_g}{V_p} \times 100b$$

Where h^2 (BS) % =heritability in broad sense; Vg=Genotypic variance; Vp = Phenotypic variance. Heritability values are categorized as low, moderate and high (Robinson *et al*., 1949) as follows: 0-30% as low; 30-60% as moderate; and 60% and above as high. The coefficient of variance (CV) was calculated for all yield and yield related traits.

Correlation among the seed yield related traits was estimated by Pearson's method at a significance level of $p < 0.05$ using R software. For the correlation analysis, we used the best linear unbiased prediction (BLUP) values of phenotype data generated from each environments by using software Meta-R (Alvarado *et al.*, 2016).

Linkage disequilibrium analysis

Linkage disequilibrium (LD) r^2 values between pairwise intra-chromosomal SNPs among the sesame accessions was performed using the markers data generated from Tassel V5.2.60 (Bradbury *et al*., 2007). The LD decay was then estimated as a function of physical genetic distance (Mbp) between pairs of SNP markers across the genome and graphical LD decay was imputed by GAPIT R package (Lipka *et al*., 2012).

Marker-trait association analysis

In this study, a set of 2997 SNP markers with a minor allele frequency (MAF) ≥ 0.05 was used for the marker-trait association (MTA) analysis. Kinship matrix and marker-trait association analysis were performed using mixed linear models (MLM) (Yu *et al.*, 2006) implemented in Tassel 5.0 software (Bradbury *et al*., 2007). Population structure (Q) and kinship similarity matrix were used to correct for any false positives. The uniform Bonferroni threshold (negative log (0.05/*n)*) was used for the significance of associations between SNPs and traits, where *n* was the total number of SNP markers in the association analysis (Xu *et al*., 2016). In this study, the threshold was 4.48 ($-\log_{10}(0.05/n) = -\log_{10}(0.05/2997) \approx 4.48$). Manhattan and QQ plots were drawn using Tassel 5.0 software and GAPIT R package.

RESULT

Phenotypic variations of sesame yield and yield related traits

The ANOVA results indicated the presence of highly significant variation among genotypes for all yield and yield-related traits across Metema trial site in the two consecutive years (Table 6.1 and 6.2). Most yield and yield related traits showed high level of broad-sense heritability (h^2), except primary branch and number of capsule per plant in the first year and number of capsule per plant and seed per capsule in the second year that showed a moderate and low level of heritability (Table 6.1 and 6. 2).

Table 0.1. ANOVA and heritability results for the yield and yield related traits of 300 sesame genotype measured in 2017/18 cropping season at Metema.

Statistic	PTH	PBR	DFI	DF	NCPP	CAPL	SPC	TSW	DM	PBZ	YLD	BBL
h2 (%)	72.14	39.37	94.06	94.73	35.15	89.24	69.66	80.69	93.54	68.03	75.51	74.44
Mean	154.03	3.18	47.23	51.40	29.06	2.86	71.43	2.35	95.84	46.00	499.67	20.96
CV (%)	6.61	31.27	3.63	3.45	28.27	7.04	8.18	9.07	2.68	16.76	26.41	22.50
p Genotype	***	***	***	***	***	***	***	***	***	***	***	***

* P < 0.05; *** P < 0.001

Table 0.2. ANOVA and heritability results for the yield and yield related traits of 300 sesame genotype measured in 2018/19 cropping season at Metema.

Statistic	PTH	PBR	DFI	DF	NCPP	CAPL	SPC	TSW	DM	PBZ	YLD	BBL
h2 (%)	82.39	63.59	87.26	68.89	51.21	66.08	22.39	78.43	81.95	59.80	84.63	84.53
Mean	126.43	2.80	48.53	54.05	31.00	2.86	67.60	2.44	102.87	44.86	574.68	23.17
CV (%)	10.77	33.03	7.15	8.98	29.37	9.59	10.49	7.96	6.44	17.49	25.59	19.43
p Genotype	***	***	***	***	***	***	*	***	***	***	***	***

* P < 0.05; *** P < 0.001

Correlation among yield related traits

Several strong correlations were observed between yield and yield related traits (Figure 6.1). Highly significant positive correlations were detected for DFI vs. DF, DFI vs. DM and DF vs. DM in the two consecutive cropping seasons. Strong positive correlations were observed between yield with number of capsule per plant (NCPP) and pod bearing zone (PBZ) and negative correlations were observed between yield and DFI, DF and BBL traits in the two consecutive cropping season.

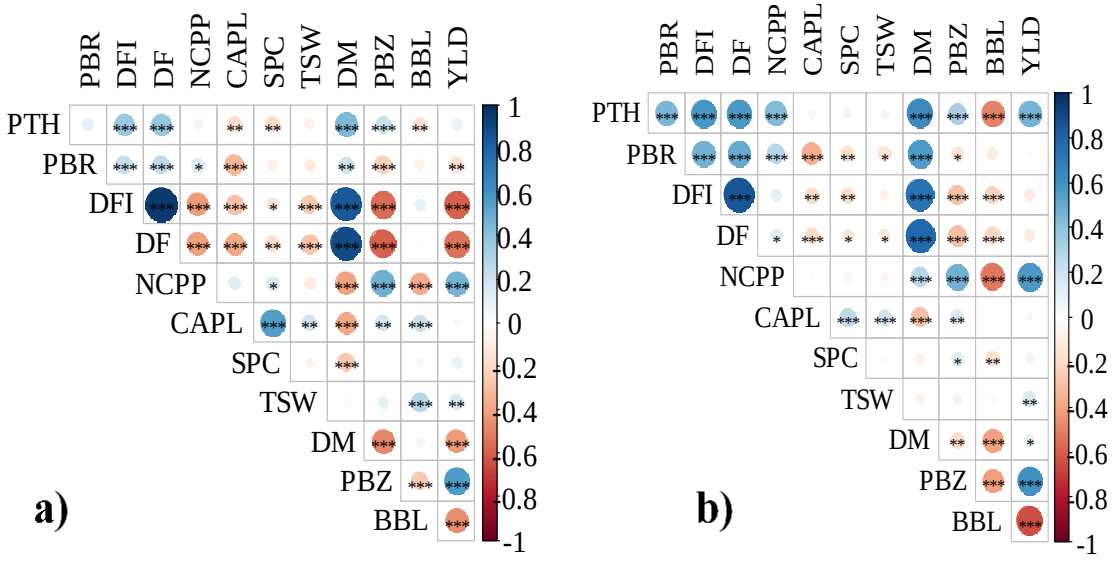


Figure 0-1. Correlation coefficient and level of significant for yield and yield related traits of 300 sesame accessions

Note; - a) At Metema 2017/18 cropping season b) At Metema 2018/19 cropping season

Population structure and Kinship similarity Matrix

The Bayesian model-based cluster analysis was conducted using the STRUCTURE program. The appropriate number of cluster K was determined, As *K* changed from 1 to 11 by inferring the Delta K values according to Evanno *et al*. (2005). The number of hypothetical ancestral populations (*K*) was validated by the second-order statistics of ΔK. The ΔK value showed a peak at $K = 4$, indicating that four is the optimal number of subgroups in the panel (Figure 6.2B). Both the kinship and population structure analysis clearly indicated the four groups (Figure 6.2A and 6.2B).

According to the structure analysis, based on the probability of membership threshold of 50%, 54, 159, 43, and 23 genotypes were assigned into the four hypothetical ancestral populations, Pop 1, Pop 2, Pop 3, and Pop 4 respectively. The remaining 21 accessions showed admixture among the created groups (Figure 6.2B). The clustering pattern did not follow the geographical origin of the germplasm. This evidence support the above genetic diversity study analysis and confirmed the high level of gene flow among the germplasm of the populations of different origins.

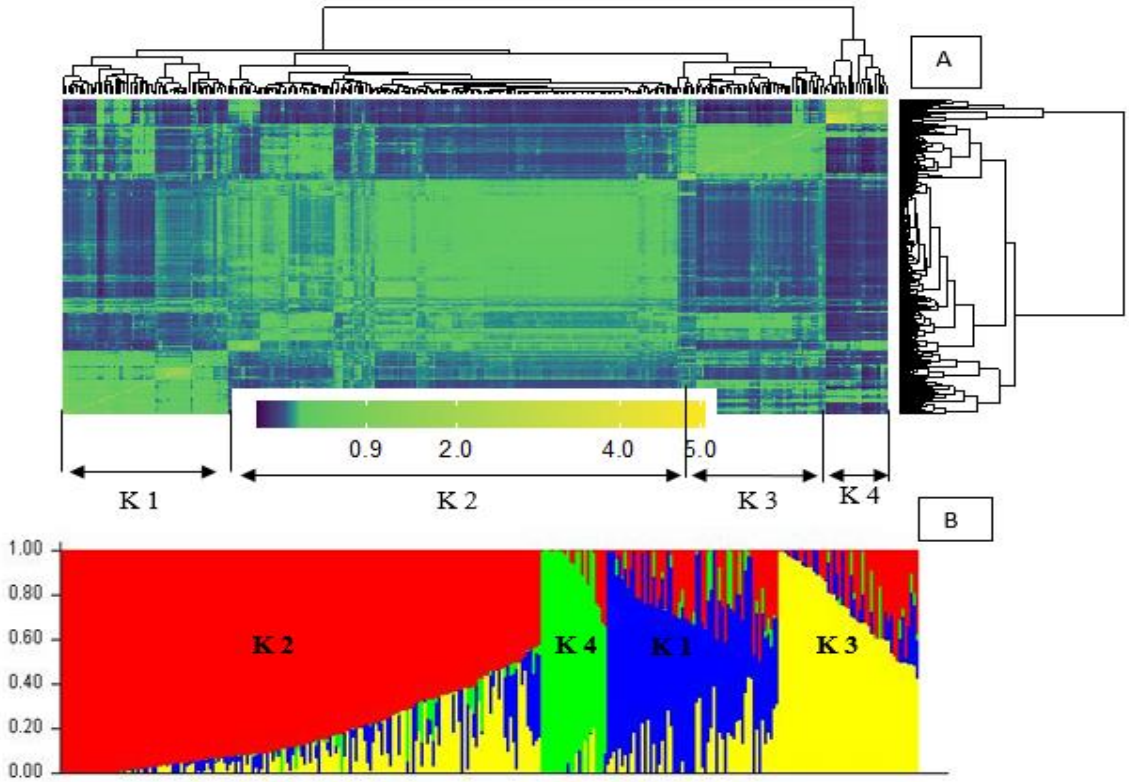


Figure 0-2. Population structure and kinship-matrix similarity analysis for 300 sesame genotypes.

Note; - Heat-map clustering results based on the kinship matrix from tag-SNP ($r^2 = 1$) by identity-by-state (IBS) algorithm (a). Population structure plot and K1, K2, K3, and k4 represents subpopulations 1, 2, 3 and 4, respectively (b). Accessions arrangement was based on the order of heat-map kinship result. The color represents the membership of each accession in the STRUCTURE-inferred subpopulations. The color of the legend indicates the level of kinship similarity of the heat-map

Linkage disequilibrium

Linkage disequilibrium (LD) was analyzed for the 2997 SNP markers across the sesame genome with 100 sliding window size. Out of the 93079 pairs of comparisons between SNP markers, only 15444 (17.24%) show highly significant LD (pDiseq < 0.001). Mean R^2 value was 0.099.

Linkage disequilibrium-decay (LDD) across thirteen chromosomes was determined using the entire set of 2997 DArTseq markers. The LDD was plotted as LD (R^2) between the

adjacent pair of markers on the Y-axis against the distance in base pairs (bp) on the X-axis (Figure 6.3). The R^2 threshold level was set to 0.2 and observed rapid LDD across the sesame genome with an average LDD of 1639.247kb (1639247bp). Among the thirteen chromosomes, the shortest LDD was observed in chromosome 4 with 113.52 kb (113520bp, R^2 = 0.2) and the longest LDD was observed in chromosome 11 with 6027.618kb (6027618bp, R^2 = 0.2). The overall LD decay in sesame germplasm was relatively high and few markers showed $R^2 \geq 0.8$. Nevertheless 2429 pairs of markers showed complete LD (R^2 = 1). High LD values were observed over a short physical genetic distance which decayed rapidly, with very low decrease at longer distances between markers across the genome (Figure 6.3).

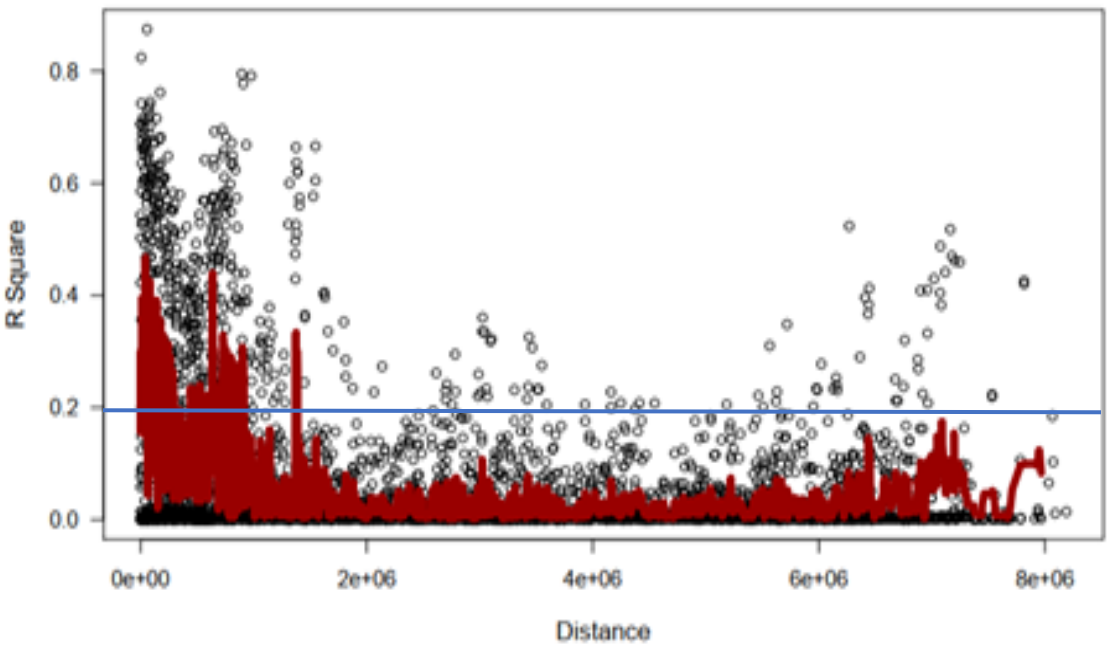


Figure 0-3. Linkage disequilibrium (LD)-measured R^2 plotted vs. the physical map (bp) between pairs of DArTseq-SNP markers in a panel of 300 sesame genotypes

Genome-wide association analysis for sesame yield and yield related traits

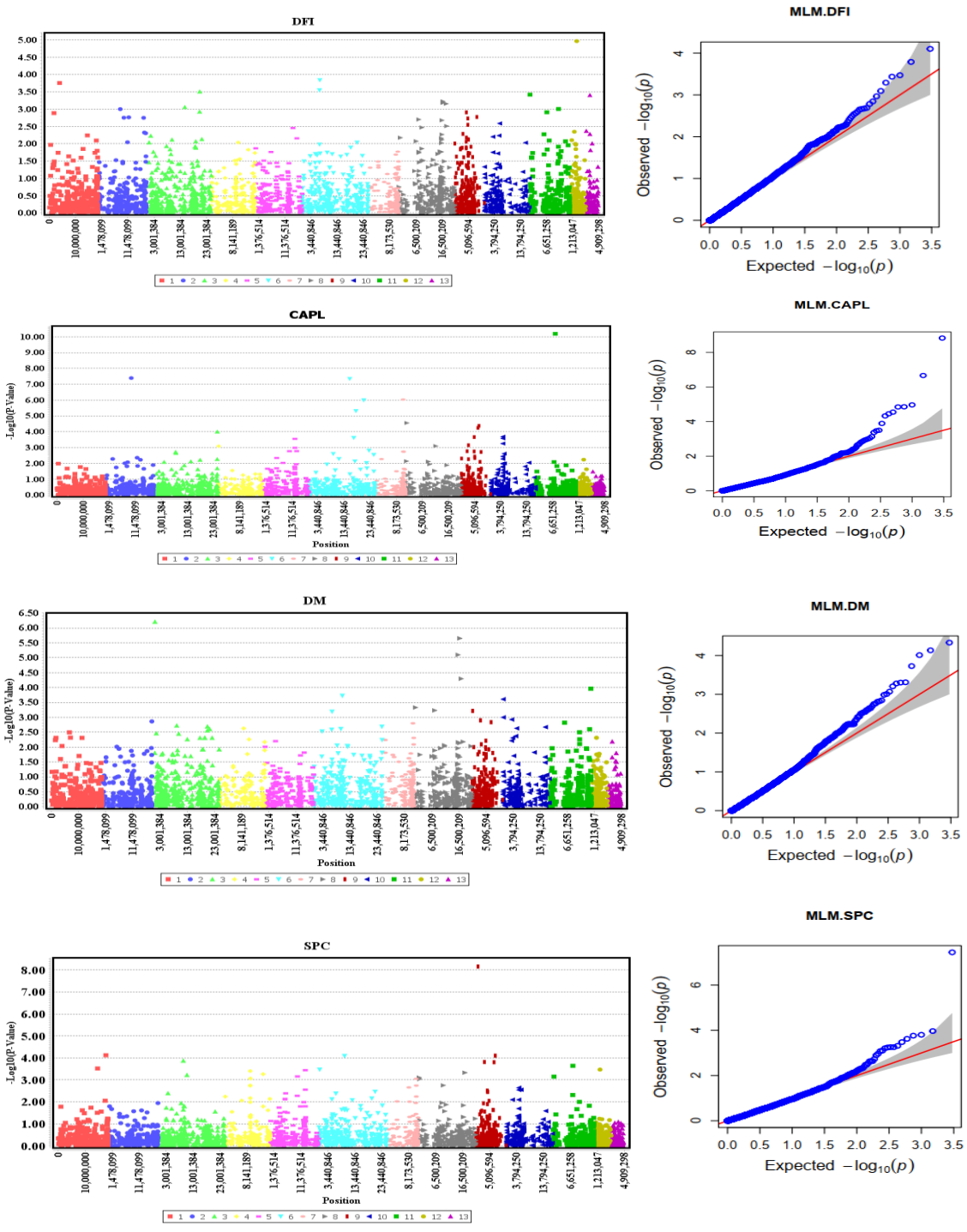
To uncover the genotypic variation of yield related traits in sesame, 2997 SNPs markers that were imputed with LD KNNi imputation were used for marker-traits association (MTA) analyses. The mixed linear model with population structure and kinship matrix was chosen for MTA analysis, as the quantile-quantile (Q-Q) plot showed that the observed MTA P-values were close to the expected distribution (Figure 6.4). Using a uniform Bonferroni threshold for significance of $P < 1.67\times10^{-5}$, 21 statistically significant SNPs were associated with the 7 traits out of which 11 were identified at Metema trial site in 2017-18 cropping season and 10 were found in 2018-19 cropping season. Significant loci were found on 10 linkage groups (LG) or chromosome of the genome, justifying the complex genetic architecture of the seed yield in sesame. The highest number of significant loci (4) was detected on the LG6, while the LG 2, 3, 4, and 11 harbored only one significant loci (Table 6.3, and Figure 6.4).

Among them 2 significant markers were found at chromosome 1 associated with trait thousand seed weight. Only one significant marker was found on chromosomes 2, 3, 4 and 11 associated with traits including capsule length, days to physiological maturity, bacterial blight and capsule length respectively. Four significant markers were found at chromosome 6 and 3 associated with capsule lengths and the remaining one associated with days to flowering initiation. Three significant markers were found on chromosome 7 and each of them associated with trait capsule length, days to flowering initiation and days to 50% flowering. On chromosome 8 three significant markers were found associated with two traits (i.e. days to flowering initiation and days to physiological maturity), and one of the significant marker was repeatable in the two environments.

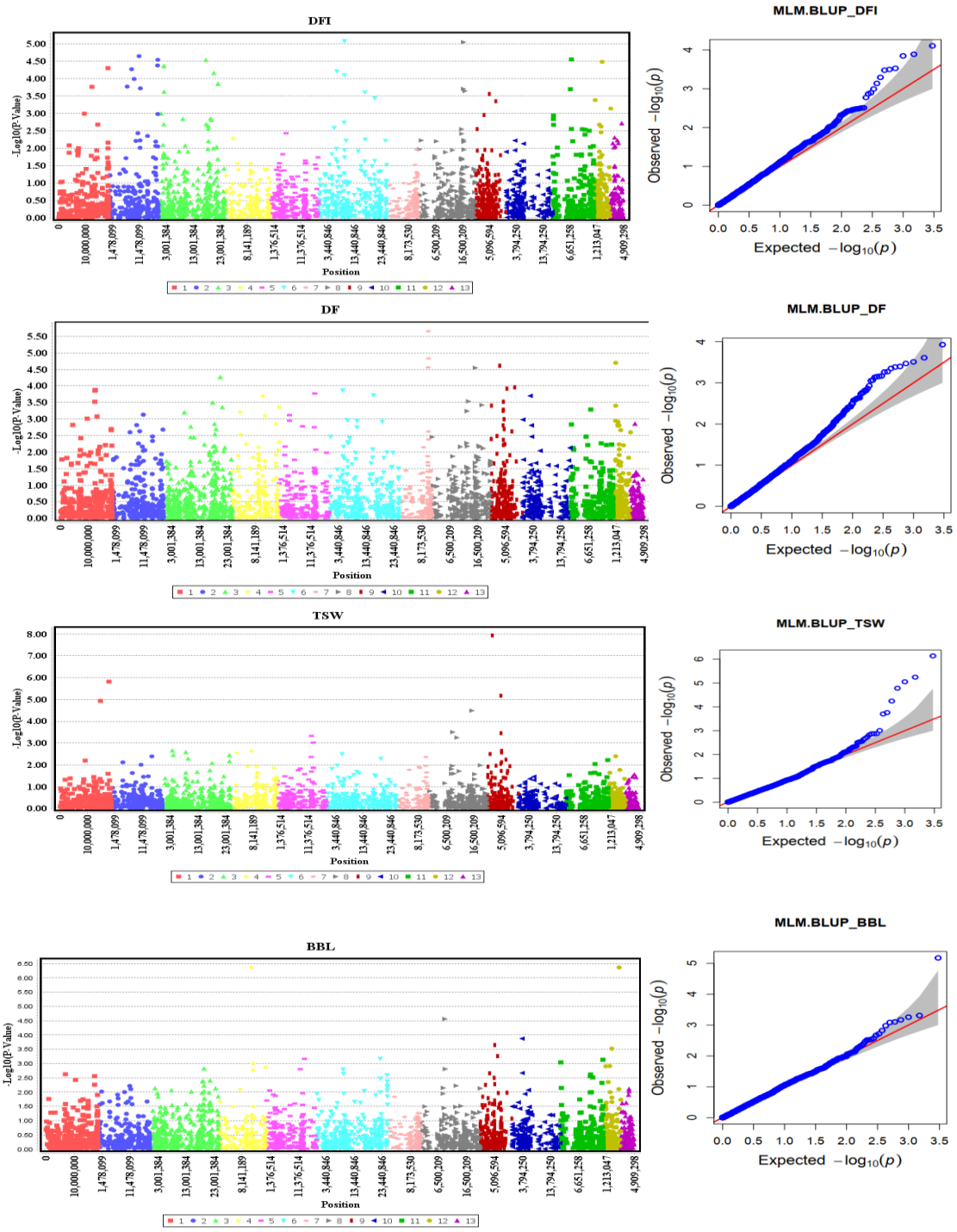
Three markers were found on chromosome 9 and of which two of them were associated with thousand seed weight and the other one associated with seed per capsule. On chromosome 12, two significant markers were found associated with days to flowering initiation and bacterial blight. In general, 6, 1, 3, 2, 3, 4, 2 significant markers associated with traits capsule length, seed per capsule, Days to flowering initiation, Days to 50% flowering, days to physiological maturity, thousand seed weight, and bacterial blight respectively.

For each linkage group, the physical distance (bp) at which the LD decayed to the critical r^2 (0.2) value was used to determine the confidence interval for declaring distinct QTL. So, based on ± of the average LDD of 1639.247kb (1639247bp) with significance marker, the three significant markers found on chromosome 7, and 8 and the two significant markers on chromosome 9 were set on the same confidence interval, and those markers are considered on the same QTL. The significant markers that form the same QTL in chromosome 7, 8 and 9 were detected in the two environments and these QTL are repeatable.

The phenotypic variation explained by the lead loci ranged from 7.02 (DF) to 16.11% (CAPL), with an average phenotype variation explained (PEV) of 9.76% (Table 6.3). Manhattan and QQ plots for the studied traits with the associated with significant markers are available in figure 6.4 and a map based on physical base pair distance of 21 markers (associated with studied traits) is constructed and given on Figure 6.5.


DFI
MLM.DFI
CAPL
MLM.CAPL
DM
MLM.DM
SPC
MLM.SPC
-Log10(P-Value)
Position
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

a)

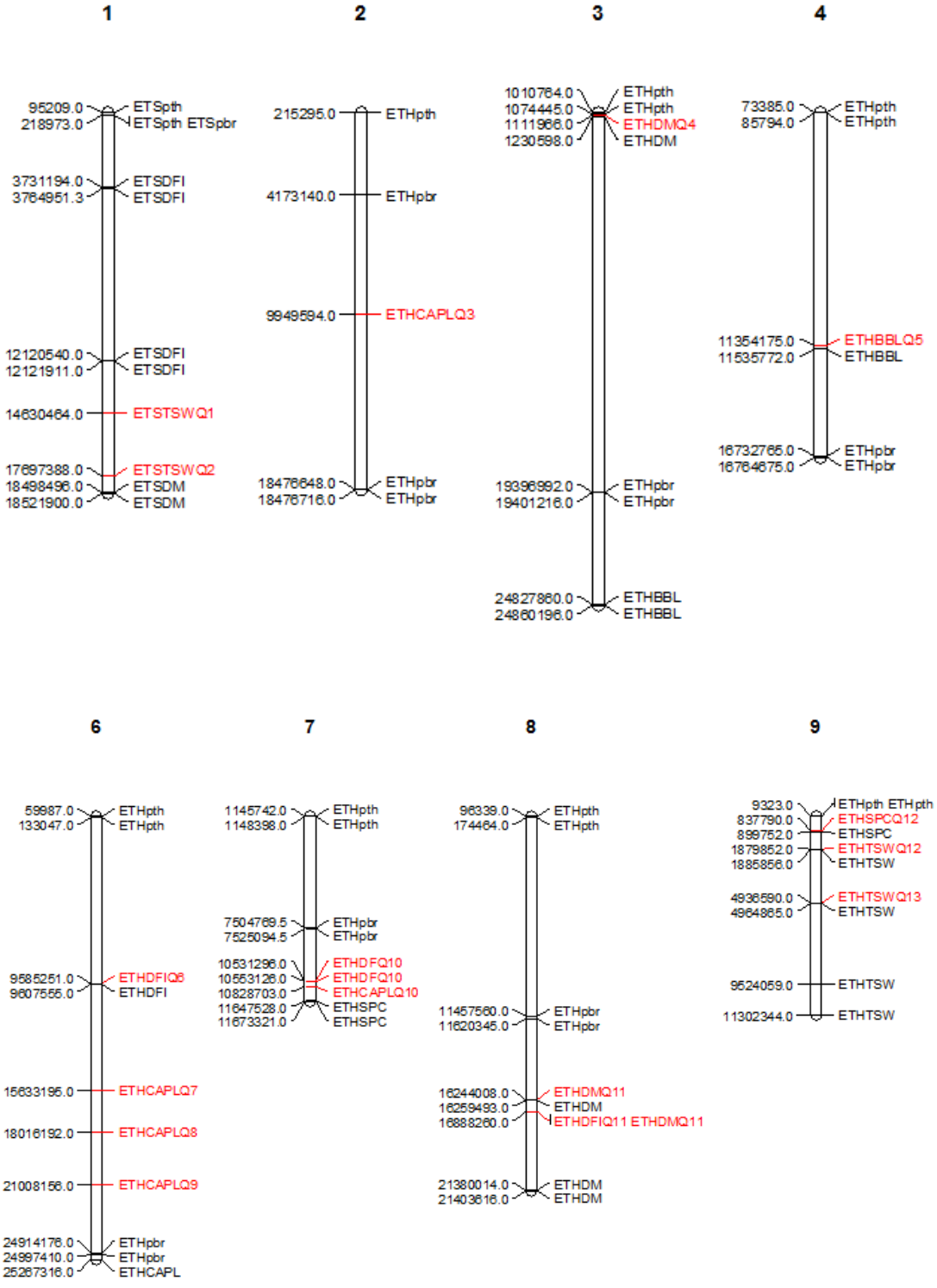


b)

Figure 0-4. Manhattan and quantile–quantile (Q–Q) plots of DFI, CAPL, DM and TSW at Metema in 2017/18 (a) and DFI, DF, TSW and BBL at Metema in 2018/19 (b). In the Manhattan plot, SNPs are ordered by physical position and grouped by chromosome.

Table 0.3. List of major and nominal QTLs for Yield related traits identified in 300 sesame accessions.

QTL	Trait	Marker	Chr.	Position (bp)	p-value	-lg p	r^2	Allele (SNP base)	Effect	Allele count (No.)	Allele (SNP base)	Effect	Allele count (No.)
ETSSWQTL1	TSW	MARKER236	1	17697389	1.52E-06	5.82	8.7	C	0.36	3	T	-0.12	265
ETSTSWQTL2	TSW	MARKER203	1	14630464	1.16E-05	4.94	7.43	A	0.38	1	G	-0.26	284
ETSCAPLQTL3	CAPL	MARKER396	2	9949594	4.09E-08	7.39	11.7	G	-0.86	239	R	0	3
ETSDMQTL4	DM	MARKER536	3	1111966	6.29E-07	6.2	10.5	A	12.53	9	T	-1.47	227
ETSBBLQTL5	BBL	MARKER1054	4	11354175	4.33E-07	6.36	8.68	C	-16.2	289	T	3.97	2
ETSCAPLQTL6	CAPL	MARKER1533	6	15633195	4.55E-08	7.34	13	A	-0.07	230	C	0.8	2
ETSCAPLQTL7	CAPL	MARKER1591	6	21008156	1.02E-06	5.99	10.8	A	0	237	C	1.01	2
ETSCAPLQTL8	CAPL	MARKER1561	6	18016193	4.93E-06	5.31	8.25	C	0.93	3	G	0	239
ETSDFIQTL9	DFI	MARKER1460	6	9585251	8.31E-06	5.08	7.26	A	11.41	4	T	-1.46	282
ETSCAPLQTL10	CAPL	MARKER1736	7	10828703	9.29E-07	6.03	10.9	A	0.05	7	C	-0.39	214
ETSDFQTL10	DF	MARKER1727	7	10531296	2.21E-06	5.66	8.16	A	-2.65	259	G	-2.4	6
ETSDFQTL10	DF	MARKER1729	7	10553126	1.48E-05	4.83	7.02	A	-2.7	3	C	-4.21	276
ETSDMQTL11	DM	MARKER1997	8	16888261	2.25E-06	5.65	9.62	C	16.1	2	G	0.67	230
ETSDMQTL11	DM	MARKER1969	8	16244008	8.04E-06	5.09	8.72	A	9.87	4	C	-0.15	215
ETSDFIQTL11	DFI	MARKER1997	8	16888261	8.90E-06	5.05	7.23	C	8.81	4	G	1.24	274
ETSSPCQTL12	SPC	MARKER2090	9	837790	6.93E-09	8.16	13.9	A	5.34	241	C	-15.3	1
ETSTSWQTL12	TSW	MARKER2112	9	1879852	1.16E-08	7.94	11.7	C	-0.24	280	T	0.12	7
ETSTSWQTL13	TSW	MARKER2207	9	4936590	6.70E-06	5.17	7.77	A	-0.21	269	G	0.24	6
ETSCAPLQTL14	CAPL	MARKER2598	11	7594736	6.55E-11	10.18	16.1	A	-1.43	241	G	0	1
ETSDFIQTL15	DFI	MARKER2814	12	2686837	1.09E-05	4.96	8.95	C	-2.24	229	T	8.64	3
ETSBBLQTL16	BBL	MARKER2876	12	5649256	4.35E-07	6.36	8.67	A	-3.73	6	T	-19.8	290


1
95209.0 ETSpth
218973.0 ETSpth ETSpbr
3731194.0 ETSDFI
3764951.3 ETSDFI
12120540.0 ETSDFI
12121911.0 ETSDFI
14630464.0 ETSTSWQ1
17697388.0 ETSTSWQ2
18498496.0 ETSDM
18521900.0 ETSDM
2
215295.0 ETHpth
4173140.0 ETHpbr
9949594.0 ETHCAPLQ3
18476648.0 ETHpbr
18476716.0 ETHpbr
3
10110764.0 ETHpth
10744445.0 ETHpth
11111966.0 ETHDMQ4
12305698.0 ETHDM
19396992.0 ETHpbr
19401216.0 ETHpbr
24827860.0 ETHBBL
24860196.0 ETHBBL
4
73385.0 ETHpth
85794.0 ETHpth
11354175.0 ETHBBLQ5
11535772.0 ETHBBL
16732765.0 ETHpbr
16764875.0 ETHpbr
6
59987.0 ETHpth
133047.0 ETHpth
9585251.0 ETHDFIQ6
9607555.0 ETHDFI
15633195.0 ETHCAPLQ7
18016192.0 ETHCAPLQ8
21008156.0 ETHCAPLQ9
24914176.0 ETHpbr
24997410.0 ETHpbr
25267316.0 ETHCAPL
7
1145742.0 ETHpth
1148398.0 ETHpth
7504769.5 ETHpbr
7525094.5 ETHpbr
10531296.0 ETHDFQ10
10553126.0 ETHDFQ10
10828703.0 ETHCAPLQ10
11647528.0 ETHSPC
11673321.0 ETHSPC
8
96339.0 ETHpth
174464.0 ETHpth
11457560.0 ETHpbr
11620345.0 ETHpbr
16244008.0 ETHDMQ11
16259493.0 ETHDM
16888260.0 ETHDFIQ11 ETHDMQ11
21380014.0 ETHDM
21403616.0 ETHDM
9
9323.0 ETHpth ETHpth
837790.0 ETHSPCQ12
899752.0 ETHSPC
1879852.0 ETHTSWQ12
1885856.0 ETHTSW
4936590.0 ETHTSWQ13
4964865.0 ETHTSW
9524059.0 ETHTSW
11302344.0 ETHTSW

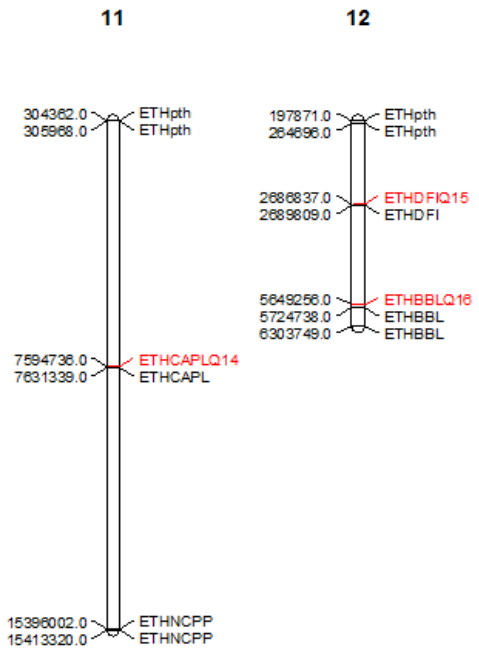


Figure 0-5. Genomic location of the 21 significant marker trait association for yield-related traits in sesame.

Note;- QTLs were named as follow: ETS- Refers Ethiopian sesame and the letter after ETS indicate the traits name (PTH; - Plant height, DFI; - Days to flower initiation, DF; - Days to 50% flowering, DM; - Days to physiological maturity, CAPL; - Capsule length, SPC;- seed per capsule, NCPP;- Number of capsule per plant, TSW;- Thousand seed weight, BBL;- Bacterial blight, and QTL with number to indicate the markers with in interval and the same QTL. Numbers to the left of each LG are marker positions (bp) and the red color shows the significant markers.

DISCUSSION

The GWAS approach is recognized as a powerful tool to reconnect traits back to the underlying genetics and offers higher resolution than classical linkage mapping (Korte and Ashley, 2013). In this study, we observed a high genetic variation for the studied yield related traits, indicating that our association panel harbors a large diversity necessary for genome wide association studies (GWAS). Among studied traits, strong positive correlations were observed between yield with number of capsule per plant (NCPP) and pod bearing zone (PBZ). Several authors have reported traits that contribute to the seed yield formation in sesame with similar results. For instance, several of these studies speculated capsule number per plant is a primary determinant for high seed yield in sesame (Biabani *et al*., 2008; Akhtar *et al.*, 2009; Wu *et al.*, 2014; Monpara and Khairnar, 2016; Shim *et al*., 2016; Zhou *et al.*, 2018).

GWAS results revealed certain significant loci associated with 7 yield related traits, all significant loci found in LG/chromosome 2, 6 and 11 associated with capsule length except one associated with physiological period. The significant loci found in LG/chromosome 3, 7 and 8 associated with physiological period (i.e. days to flower initiation, days to 50% flowering and days to physiological maturity) except one was associated with capsule length and similarly.

Flowering time is also an important trait for adaptation of crops to different agro-climatic conditions that significantly affects the yield. Two candidate genes at flowering-time loci *SiDOG1* (*SIN_1022538*) and *SiIAA14* (*SIN_1021838*) have been discovered (Wei *et al*., 2013). Sesame's indeterminate growth habit is one of the reasons of its low yielding capacity compared to other oilseed crops (Yoland Uzun, 2012). Recently, the gene *SiDt*

(*DS899s00170.023*) was detected as a target gene for conferring the determinate trait in sesame cultivar (Zhang *et al.,* 2016). So, dissecting genetic control of flowering time and maturity is of pivotal importance to foster sesame breeding and to develop new varieties able to adapt to changing climatic conditions. Indeed, flowering time and maturity strongly affects yield and plant adaptation ability.

The other significance loci found in LG 1 and 9 associated with thousand seed weight. This indicate those who are interested in specific traits can focus on these chromosome number that associated with specific traits.

The phenotypic variation explained by the lead loci ranged from 7.02 to 16.11%, with an average phenotype variation explained (PEV) of 13.34%, suggesting a moderate contribution to the traits.

Previously, two studies were performed on the genetics of the sesame seed yield by employing the linkage mapping approach (Wu *et al.*, 2014; Zhou *et al.*, 2018). Comparing our results with the previous QTLs, we couldn't detect any overlapping loci associated for similar traits. This suggests that the QTLs detected in this study could potentially be novel.

Based our result, associated SNPs with traits associated SNPs with traits of economic importance could potentially be used as candidates for further functional analysis using the transgenic approach (Chowdhury *et al*., 2017) and genome-editing technologies such as CRISPR/Cas system. Meanwhile, the peak loci or markers with high effect could be transformed into allele-specific markers or KASP markers for applications of marker-assisted selection in breeding programs to design sesame varieties with improved seed

yield. This can be, inter alia, related to the use of elite cultivars. For instance, the highest yielder in our panel have, on average, 53 capsules on the main stem, which is approximately 4.5 times than the capsule number in the lowest yielder landraces, and thus have a higher yield potential. Since several favorable alleles detected in this study have not yet been intensively selected, our GWAS results will undoubtedly assist in incorporating further useful alleles into the elite sesame germplasm for a seed yield increase in the future.

Chapter 7. General Conclusion and Recommendations

Sesame is one of the most ancient oil crops and it is grown widely in both tropical and subtropical areas. In Ethiopia, sesame is among the foremost important oil crops both in terms of area coverage and total national annual production. It is the main exported product after coffee. Despite the potential and importance of the crop, the farm level productivity of sesame in Ethiopia remained very low compared to the genetic potential of the crop yield, due to different biotic and abiotic stresses.

To address these problems, setting up a practical sesame breeding program targeting major production constraints is quite essential and necessary. The present study attempted to avail various genetic and/or breeding ingredients which are fundamental to realize sesame improvement in the country. Based on germplasm available at gene bank and in the hand of breeders, the experience of research centers, and the availability of trial sites, certain activities were implemented from conventional method, morphological characterization, and genotype by environment interaction including the modern genome-wide association study, to address the constraints and to have short term and long term solution with future sesame breeding direction. In this study, 300 genotypes comprised of 209 Ethiopian landraces, 75 exotic collections, and 16 released varieties were phenotyped and genotyped using two high-throughput diversity array technology (DArT) markers to characterize the genotypes and 16 released varieties across different agroecology of Ethiopia.

In the morphological characterization, a considerable level of variability within the Ethiopian germplasm was observed, as that of the Africa and Asian collections that can be exploited in future breeding programs. Plant height, primary branch, petiole length of top leaf, days to flower initiation, days to 50% flowering, pod bearing zone, seed yield

per plant, and bacterial blight reaction showed high heritability along with high genetic advance. Therefore, progeny selection will be effective to improve these characters. Capsule length, 1000 seed weight, days to maturity, showed high heritability with moderate genetic advance. These characters can be exploited through hybridization. Correlation coefficient analysis measures the mutual relationship between various characters and is used to determine the component character on which selection can be done for improvement in yield. Grain yield showed a significant and positive genotypic correlation with plant height, a number of capsule per plant, and pod bearing zone, the magnitudes of the positive genetic correlation suggest that the selection by those characters produces a significant increase in grain yield.

Based on the quantitative data, 300 genotypes are grouped into six clusters, with clusters I and II accounting for the largest number of genotypes. Greater genetic divergence is present between clusters IV and VI; followed by distances between clusters I and VI; indicating that best recombinants can be realized by mating between the lines of these clusters in a definite fashion. Their high yield potential can subsequently be combined with improvements of other traits such as plant height, number of capsules per plant, oil content, resistance to pests, and disease. The genotype originally collected from four regions of Ethiopia (Amhara, Benshangul-Gumz, Oromia, and Tigray), especially genotype categorized under clusters VI and III, were found interesting and could be candidates for potential immediate breeding sources due to their high seed yield.

The results of this study provided a better understanding of sesame populations in different ecological regions. Therefore, the wise use of results obtained in this study

would facilitate the improvement of sesame varieties through breeding and the ex situ conservation of sesame genetic resources.

The molecular characterization showed the effectiveness of DArTseq in characterizing the genetic diversity and population structure of sesame collection in Ethiopia. The gene diversity study showed the Asian population of sesame is more divergent than the sesame populations from Africa that indicate the importance of further collection from the Asian continent to enrich the genetic pool of the sesame in the breeding program. Even if Ethiopian sesame has useful characteristics such as aroma and tests, it has low genetic diversity as compared to the population obtained from other regions of the world. This tells us the breeding program should focus on obtaining more sesame accessions from Asia and North Africa to enrich the genetic pool and conducting further crossing programs to introgress the favorable genes from the imported germplasm. This study also supports the idea; ecological and geographical factors are less effective in the evolution of sesame. This finding guides for the systematic utilization and conservation of the genetic resource of sesame and underlines the requirement of conducting a further collection of sesame genotypes from all sesame growing regions of the world. Therefore, our next objective is to identify sesame genotypes with desirable traits and to conduct association mapping on the economically important traits in sesame to implement markers assisted selection (MAS) in the sesame breeding program in the country to increase the efficiency of the breeding program.

Based on the advantage of both morphological and molecular data, a large-scale GWAS were performed on yield related traits of African sesame natural populations, and a total of 21 significant SNPs with 7 yield-related traits were identified in two environments.

The favorable alleles detected in this study have not yet been intensively characterized, These SNPs will play an important role in understanding the genetic basis of yield related traits in sesame and to use as potential candidates for further functional analysis using the transgenic approach (Chowdhury *et al*., 2017) and genome-editing technologies using CRISPR/Cas system. So, our GWAS results will undoubtedly assist in incorporating further useful alleles into the elite sesame germplasm for a seed yield increase in the future.

The G X E analysis results suggested sesame producing environment in Ethiopia classified into six mega-environments and that different genotypes should be selected and deployed for each. Overall 'G9' (Obsa) had the highest mean yield, followed by 'G2' (Setit-1) and 'G1' (Humera-1), whereas 'G6' (ADI), 'G7' (TATI), and 'G8' (Acc-051-02-sel 1 (2)) had the lowest mean yield. Thus, 'G3' (SETIT-2), 'G4' (HUARC-4), and 'G7' (TATI) were highly unstable, whereas 'G13' (MECHAL) 'G14' (BENSHANGUL-1) 'G16' (BAHA ZEYT) were highly stable genotypes. Among tested genotypes, G2 (Setit-1) was the "ideal" genotype and one of the highest mean seed yields. G2 (Setit-1) considered the most stable across variable environments Hence, it is the most desirable genotype to be directly recommended for farmers' use and/or to be used as source material for future sesame breeding that targets high yielding and stable genotypes. Based on the test environment, Environment E16 (Tach Armacho 2017/18) and E17 (Tach Armacho 2018/19) were close to the ideal environment (Figure 4.7), therefore, Location Tach Armacho should be regarded as the most suitable to select widely adapted genotypes. The findings of this study provide useful information for sesame breeding and commercialization in Ethiopia.

REFERENCES

Aaron J. Lorenz, Shiaoman Chao, F. G. A., Elliot L. Heffner,§ Takeshi Hayashi, H. I. and Kevin P. Smith, k Mark E. Sorrells, and J.-L. J. (2011) 'Genomic Selection in Plant Breeding: Knowledge and Prospects', in *Advance in Agronomy*. doi: 10.1016/B978-0-12-385531-2.00001-3.

Abate, Mohammed, Firew Mekbib, Amsalu Ayana, and M. N. (2015a) 'Genetic Variability and Association of Traits in Mid- altitude Sesame (Sesamum indicum L .) Germplasm of Ethiopia', *Journal of Experimental Agriculture International*, 9(3), pp. 1–14. doi: 10.9734/AJEA/2015/18483.

Abate, Mohammed, Firew Mekbib, Amsalu Ayana, and M. N. (2015b) 'Genotype x Environment and Stability Analysis of Oil Content in Sesame (Sesamum indicum L.) Evaluated Across Diverse Agro-ecologies of the Awash Valleys in Ethiopia', *American Journal of Experimental Agriculture*, 9(2), pp. 1–12. doi: 10.9734/ajea/2015/18482.

Abate, M. (2015) 'Genotype X Environment Analysis for seed yield and its components in sesame (Sesamum Indicum L.) evaluated across diverse agro-ecologies of the Awash valleys in Ethiopia', *journal of Advanced Studies in Agricultural, Biological and Environmental Sciences (JABE) www.jabe.in*, 2(4), pp. 1–14.

Abate, M. and Mekbib, F. (2015a) 'Assesment of Genetic Variability and Character Association in Ethiopian Low-Altitude Sesame (Sesamum Indicum L.) Genotypes', *Journal of Advanced Studies in Agricultural, Biological and Environmental Sciences*, 2(3), pp. 55–66. doi: 10.2133/JABE/2015/2015.

Abate, M. and Mekbib, F. (2015b) 'Assessment of Genetic Diversity in Ethiopian Sesame

(Sesamum indicum L.) Germplasm using Random Amplified Polymorphic DNA (RAPD) Markers Council for Innovative Research', (November).

Abate, M. and Mekbib, F. (2016) 'Study on genetic divergence in low-altitude sesame (Sesamum indicum L.) germplasm of Ethiopia based on agro morphological traits.', *Journal of Advanced Studies in Agricultural, Biological and Environmental Sciences*, (August 2015). doi: 10.3234/JABE/2015/2016.

Adefris, T., Tadele, A., & Tesfaye, B. (2011) 'Sesame cultivation and use in Ethiopia', *In. Bedigian D (ed) Sesame: the genus Sesamum. Medicinal and Aromatic Plants Industrial Profiles.*

Agne, P.S.E., Rance, F. and Bidat, E. (2003) 'Sesame seed allergy', *REVUE FRANCAISE D ALLERGOLOGIE ET D IMMUNOLOGIE CLINIQUE*, 4(8), pp. 507–516.

Akhtar, K. P. *et al.* (2009) 'Sesame phyllody disease: Its symptomatology, etiology, and transmission in Pakistan', *Turkish Journal of Agriculture and Forestry*, 33(5), pp. 477–486. doi: 10.3906/tar-0901-23.

Alan Robertson (1959) 'The Sampling Variance of the Genetic Correlation Coefficient', *International Biometric Society*, 15(3), pp. 469–485.

Alarmelu, S., & Ramanathan, T. (1998) 'Genetic Divergence and Heterosis in sesame (Sesamum indicum L)', *Journal of oilseeds research*, 15, pp. 25–31.

Alegbejo, M. D. *et al.* (2003) 'Sesame: A potential industrial and export oilseed crop in Nigeria', *Journal of Sustainable Agriculture*. doi: 10.1300/J064v23n01_05.

Alemu, Dawit, and G. W. M. (2010) *Sesame traders and the ECX: An overview with focus on transaction costs and risks, VC4PPD report 8*.

Alemu, A. and Tesfaye, K. (2013) 'Genetic Distance of Sesame (Sesamum indicum L .)

Cultivars and Varieties from Northwestern Ethiopia using Inter Simple Sequence Repeat Markers Genetic Distance of Sesame (Sesamum indicum L .) Cultivars and Varieties from Northwestern Ethiopia using Int', (November 2015).

Ali, G. M., Yasumoto, S., & Seki-Katsuta, M. (2007) 'Assessment of genetic diversity in sesame (Sesamum indicum L.) detected by amplified fragment length polymorphism markers', *Electronic Journal of Biotechnology*, 10(1), pp. 12–23.

Ali, G. M. (2007) 'Assessment of genetic diversity in sesame (Sesamum indicum L .) detected by Amplified Fragment Length Polymorphism markers', 10(1), pp. 1–12. doi: 10.2225/vol10-issue1-fulltext-16.

Alvarado, G. *et al.* (2016) 'META-R (Multi Environment Trial Analysis with R for Windows)'.

Anitha, V. Y., K. S. and J. G. (2000) 'Studies on variability and genetic advance for certain characters in mutant populations of sesame (Sesamum indicum L.)', *Sesame and Safflower Newsetter*, 15, pp. 39–43.

Annicchiarico, P., Perenzin, M. (1994) 'Adaptation Patterns and Definition of Macro-environments for Selection and Recommendation of Commonwheat Genotypes in Italy', *Plant Breeding*, 113, pp. 197–205.

Ashri, A. (1998) 'Sesame breeding', *Plant Breeding Reviews*, 16, pp. 179–228.

Ashri A (2007) 'Sesame (Sesamum indicum L.)', *In: Singh RJ (ed) Genetic resources chromosome engineering, and crop improvement, oilseed crops. CRC Press, Taylor and Francis Group, Boca Raton, US*, pp. 231–289.

Aye, M., Khaing, T. and Hom, N. (2018) 'MORPHOLOGICAL CHARACTERIZATION AND GENETIC DIVERGENCE IN MYANMAR

SESAME (Sesamum indicum L.) GERMPLASM.', *International Journal of Advanced Research*. doi: 10.21474/ijar01/6858.

Badu-Apraku, B., Abamu, F. J., Menkir, A., Fakorede, M. A. B., & and Obeng-Antwi, K. (2003) 'Genotype by environment interactions in the regional early variety trials in West and Central Africa', *Maydica*, 48, pp. 93–104.

Badu-Apraku, B. *et al.* (2012) 'Performance of extra-early maize cultivars based on GGE biplot and AMMI analysis', *Journal of Agricultural Science*, 150(4), pp. 473–483. doi: 10.1017/S0021859611000761.

Baloch, F. S., Alsaleh, A. and Andeden, E. E. (2016) 'High levels of segregation distortion in the molecular linkage map of bread wheat representing the West Asia and North Africa region', pp. 352–364. doi: 10.3906/tar-1508-27.

Basak, M., Uzun, B. and Id, E. Y. (2019) 'Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq', pp. 1–15. doi: 10.1371/journal.pone.0223757.

Bastien, M., Sonah, H. and Belzile, F. (2014) ' Genome Wide Association Mapping of Sclerotinia sclerotiorum Resistance in Soybean with a Genotyping-by-Sequencing Approach ', *The Plant Genome*. doi: 10.3835/plantgenome2013.10.0030.

Basu, S. K. *et al.* (2009) 'Effects of genotype and environment on seed and forage yield in fenugreek (Trigonella foenum-graecum L.) grown in western Canada', *Australian Journal of Crop Science*, 3(6), pp. 305–314.

Bedigian, Dorothea (2010) 'Characterization of sesame (Sesamum indicum L .) germplasm : a critique', pp. 641–647. doi: 10.1007/s10722-010-9552-x.

Bedigian, D (2010) ‘Introduction and early use of sesame in America: medicine, cookery and folkways.’, *Sesame: The Genus Sesamum, Medicinal and aromatic plants–Industrial profiles series, Boca Raton, Florida, CRC Press, Taylor and Francis Group*, pp. 389–421.

Bedigian, D. and Harlan, J. R. (1986) ‘Evidence for cultivation of sesame in the ancient world’, *Economic Botany*. doi: 10.1007/BF02859136.

Bedigian D (2015) ‘Systematics and evolution in Sesamum L.(Pedaliaceae)’, *part 1: Evidence regarding the origin of sesame and its closest relatives. Webbia*, 70(1), pp. 1–42.

Beech, D. F., & Ashri, A. (1985) ‘Sesame research possibilities for yield improvement. In: Ashri A (ed) Sesame and safflower: Status and potential’, *FAO plant production and protection, Rome*.

Beheshti, A. R. and fard, B. B. (2010) ‘Dry matter accumulation and remobilization in grain sorghum genotypes (Sorghum bicolor L. Moench) under drought stress’, *Australian Journal of Crop Science*, 4(3), pp. 185–189.

Belay, F. (2018) ‘Evaluation of the Performance of Some White Seeded Sesame (Sesamum Indicum L .) Genotypes Using GGE Biplot in Northern’, 8(11), pp. 60–68.

Bernardo, R. (2014) ‘Genomewide selection when major genes are known’, *Crop Science*, 54(1), pp. 68–75.

Bhat, K. V., Babrekar, P. P. and Lakhanpaul, S. (1999) ‘Study of genetic diversity in Indian and exotic sesame (Sesamum indicum L.) germplasm using random amplified polymorphic DNA (RAPD) markers’, *Euphytica*. doi:

10.1023/A:1003724732323.

Biabani, A.R. and Pakniyat, H. (2008) ‘Evaluation of seed yield-related characters in sesame (Sesamum indicum L.) using factor and path analysis’, *Pakistan journal of biological sciences: PJBS*, 11(8), pp. 1157–1160.

Bolibok-Bragoszewska, H. *et al.* (2009) ‘DArT markers for the rye genome - genetic diversity and mapping’, *BMC Genomics*. doi: 10.1186/1471-2164-10-578.

Borchani, C. *et al.* (2010) ‘Chemical Characteristics and Oxidative Stability of Sesame Seed , Sesame Paste , and Olive Oils’, *Journal of Agricultural Science and Technology*, 12(5), pp. 585–596.

Bradbury, P.J., Zhang, Z., Kroon, D.E. and Casstevens, T.M., R. and B. E. (2007) ‘TASSEL: Software for association mapping of complex traits in diverse samples’, *Bioinformatics*, 23, p. 26332635.

Burton Glenn W. Devane E. H. (1953) ‘Estimating Heritability in Tall Fescue (Festuca Arundinacea) from Replicated Clonal Material’, *Agronomy journal*, 45(10), pp. 478–481. doi: https://doi.org/10.2134/agronj1953.00021962004500100005xCitations: 12.

Cho, Y. Il *et al.* (2011) ‘Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers’, *Genes and Genomics*. doi: 10.1007/s13258-010-0130-6.

Chowdhury, S., Basu, A. and Kundu, S. (2017) ‘Overexpression of a new osmotin-like protein gene (sindOLP) confers tolerance against biotic and abiotic stresses in sesame’, *Frontiers in Plant Science*, 8(March), pp. 1–16. doi: 10.3389/fpls.2017.00410.

Chutimanitsakun, Y. *et al.* (2011) 'Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley', *BMC Genomics*, 12, pp. 1–13. doi: 10.1186/1471-2164-12-4.

Courtois, B. *et al.* (2013) 'Genome-wide association mapping of root traits in a japonica rice panel', *PLoS ONE*. doi: 10.1371/journal.pone.0078037.

Crossa, J., Gauch, H. G. and Zobel, R. W. (1990) 'Additive Main Effects and Multiplicative Interaction Analysis of Two International Maize Cultivar Trials', *Crop Science*, 30(3), pp. 493–500. doi: 10.2135/cropsci1990.0011183x003000030003x.

Cruz, V. M. V., Kilian, A. and Dierig, D. A. (2013) 'Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop Lesquerella and Related Species', *PLoS ONE*. doi: 10.1371/journal.pone.0064062.

CSA (2019) *THE FEDERAL DEMOCRATIC REPUBLIC OF ETHIOPIA CENTRAL STATISTICAL AGENCY AGRICULTURAL SAMPLE SURVEY, CROP AND LIVESTOCK PRODUCT UTILIZATION (PRIVATE PEASANT HOLDINGS, MEHER SEASON) ADDIS*.

CSA (Central Statistical Agency) (2018) 'Sample survey area and production of major crops (private peasant holdings, meher season)', I, pp. 1–58.

Cui, C. *et al.* (2017) 'Genetic Diversity , Population Structure , and Linkage Disequilibrium of an Association-Mapping Panel Revealed by Genome-Wide SNP Markers in Sesame', 8(July), pp. 1–10. doi: 10.3389/fpls.2017.01189.

D.Bharathi, V.Thirumala Rao*, Y.Chandra Mohan, D. B. and V. V. (2004) 'Studies on

genetic variability in sesame (Sesamum indicum L.)', *Annals of Agri Bio Research*, 9(1), pp. 7–11.

Daba, Chemeda, Amsalu Ayana, Habtamu Zeleke, and A. W. (2014) 'Association of Stability Parameters and Yield Stability of Sesame (Sesamum indicum L .) Genotypes in Western Ethiopia', *East African Journal of Sciences*, 8(2014), pp. 125–134.

Daba, Chemeda, Amsalu Ayana, Habtamu Zeleke, and A. W. (2015) 'Genotype x Environment interactions for seed yield in Sesame in western Ethiopia', *East African Journal of Sciences*, 9(2), pp. 85–96.

Dagmawi, T. W., Kassahun, T. and Endashaw, B. (2015) 'Genetic diversity of sesame germplasm collection (SESAMUM INDICUM L.): implication for conservation, improvement and use', *International Journal of Biotechnology and Molecular Biology Research*, 6(2), pp. 7–18. doi: 10.5897/ijbmbr2014.0219.

Dalal, Ilan, Inga Binson, Ram Reifen, Z. Amitai, Tt Shohat, S. Rahmani, A. Levine, A. Ballin, and E. S. (2002) 'Food allergy is a matter of geography after all: Sesame as a major cause of severe IgE-mediated food allergic reactions among infants and young children in Israel', *Allergy: European Journal of Allergy and Clinical Immunology*, 57(4), pp. 362–365. doi: 10.1034/j.1398-9995.2002.1s3412.x.

Daniel Endale Gebremichael and Heiko K. Parzies (2011) 'Genetic variability among landraces of sesame in Ethiopia', *African Crop Science Journal*, 19(June), pp. 1–13. doi: 10.4314/acsj.v19i1.68656.

Davey, J. W. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics*. Nature Publishing Group,

12(7), pp. 499–510. doi: 10.1038/nrg3012.

Demetrios, M. (1980) ‘Hybridization of crop plants (No. BOOK)’, *American Society of Agronomy and Crop Science Society of America*.

Desawi Hdru Teklu and Sentayehu Alamerew Kebede, D. E. G. (2014) ‘Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes’, *Asian Journal of Agricultural Research*, 8(4), pp. 181–194. doi: 10.3923/ajar.2014.181.194.

Deschamps, S., Llaca, V. and May, G. D. (2012) ‘Genotyping-by-sequencing in plants’, *Biology*, 1(3), pp. 460–483. doi: 10.3390/biology1030460.

Dierig, D., and Ray, D. T. (2009) ‘New crops breeding: Lesquerella.’, *In Oil Crops*, pp. 507–516.

Ding, M., Tier, B. and Yan, W. (2007) ‘Application of GGE- biplot analysis to evaluate genotypes (G) environment (E) and G x E interaction on P.radiata.’, in *Case study. Australian Forest Genetics Conference, 11-14 April 2007*.

Dixon, A.G.O., Nukenine, E.N. (1997) ‘Statistical analysis of cassava yield trials with the additive main effects and multiplicative interaction (AMMI) model’, *African Journal of Root and Tuber Crops*, 3, pp. 46–50.

Dorothea Bedigian (2004) ‘History and Lore of sesame in Southwest Asia’, *Economic Botany*. doi: 10.1663/0013-0001(2004)058.

Dossa, K. *et al.* (2016) ‘Analysis of Genetic Diversity and Population Structure of Sesame Accessions from Africa and Asia as Major Centers of Its Cultivation’, pp. 1–13. doi: 10.3390/genes7040014.

Dossa, K. *et al.* (2017) ‘The emerging oilseed crop sesamum indicum enters the “Omics”

era', *Frontiers in Plant Science*. doi: 10.3389/fpls.2017.01154.

Dossa, K. *et al.* (2019) 'The genetic basis of drought tolerance in the high oil crop Sesamum indicum', *Plant Biotechnology Journal*, 17(9), pp. 1788–1803. doi: 10.1111/pbi.13100.

'DoyleDoyle_Focus_1990_CTAB.pdf' (no date).

Earl, D. A. and vonHoldt, B. M. (2012) 'STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method', *Conservation Genetics Resources*. doi: 10.1007/s12686-011-9548-7.

Edwards, D. and Batley, J. (2010) 'Plant genome sequencing: Applications for crop improvement', *Plant Biotechnology Journal*, 8(1), pp. 2–9. doi: 10.1111/j.1467-7652.2009.00459.x.

Elleuch, M. *et al.* (2007) 'Quality characteristics of sesame seeds and by-products', *Food Chemistry*, 103(2), pp. 641–650. doi: 10.1016/j.foodchem.2006.09.008.

Elshire, R. J. *et al.* (2011) 'A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species', *PLoS ONE*, 6(5), pp. 1–10. doi: 10.1371/journal.pone.0019379.

Ercan, A. G., Taskin, M. and Turgut, K. (2004) 'Analysis of genetic diversity in Turkish sesame (Sesamum indicum L .) populations using RAPD markers ⋆ Analysis of genetic diversity in Turkish sesame (Sesamum indicum L .) populations using RAPD markers ☙', (September). doi: 10.1023/B.

Espitia M, Aramendiz H, C. J. (2008) 'Correlacionesyanálisisdesendero en algodónGossypium hirsutum L. en el Caribe colombiano.', *Rev. Fac. Nac. Agron*, 61(1), pp. 4325–4335.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study', *Molecular Ecology*. doi: 10.1111/j.1365-294X.2005.02553.x.

Excoffier, L. and Lischer, H. E. L. (2010) 'Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.', *Evolutionary Bioinformatics Online*. doi: 10.1016/j.ympev.2005.04.031.

Falconer, D. S., & Mackay, T. F. C. (1996) 'Introduction to quantitative genetics', *Longman. Essex, England*.

Famoso, A. N. *et al.* (2011) 'Genetic architecture of aluminum tolerance in rice (oryza sativa) determined through genome-wide association analysis and qtl mapping', *PLoS Genetics*, 7(8). doi: 10.1371/journal.pgen.1002221.

Farshadfar, E. *et al.* (2012) 'GGE biplot analysis of genotype × environment interaction in wheat-barley disomic addition lines', *Australian Journal of Crop Science*, 6(6), pp. 1074–1079.

Farshadfar, M. and Farshadfar, E. (2008) 'Genetic variability and path analysis of chickpea (Cicer arientinum L.) landraces and lines', *Journal of Applied Sciences*, 8(21), pp. 3951–3956. doi: 10.3923/jas.2008.3951.3956.

Ferney, G. H. *et al.* (2007) 'Evaluation of Grain Yield Stability, Reliability and Cultivar Recommendations in Spring Wheat (Triticum Aestivum L.) From Kazakhstan and Siberia', *Journal of Central European Agriculture*, 7(4), pp. 649–660. doi: 10.5513/jcea.v7i4.407.

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et

al. (2005) ‘Maize association population: a high-resolution platform for quantitative trait locus dissection’, *Plant J*, 44, 1054–1, pp. 1054–1064. doi: doi: 10.1111/j.1365-313X.2005.02591.x.

Flint-Garcia, S. A., Thornsberry, J. M. and Edward IV, S. B. (2003) ‘Structure of Linkage Disequilibrium in Plants’, *Annual Review of Plant Biology*, 54, pp. 357–374. doi: 10.1146/annurev.arplant.54.031902.134907.

Food and Agriculture Organization of the United Nations (FAO) (2017) *FAOSTAT. Food and agriculture data, FAOSTAT Data- base Gateway*.

Food and Agriculture Organization Statistical Databases (FAOSTAT) (2017).

Food and Agriculture Organization Statistical Databases (FAOSTAT) (2019).

Forsberg, N. E. G. *et al.* (2015) ‘Farmers without borders - genetic structuring in century old barley (Hordeum vulgare)’, *Heredity*. Nature Publishing Group, 114(2), pp. 195–206. doi: 10.1038/hdy.2014.83.

Frutos, E., Galindo, M. P. and Leiva, V. (2014) ‘An interactive biplot implementation in R for modeling genotype-by-environment interaction’, *Stochastic Environmental Research and Risk Assessment*, 28(7), pp. 1629–1641. doi: 10.1007/s00477-013-0821-z.

Fuller, D. Q. (2003) ‘Further Evidence on the Prehistory of Sesame’, *Asian Agri-History*, 7, pp. 127–137.

Furat, S. and Uzun, B. (2010) ‘The use of agro-morphological characters for the assessment of genetic diversity in sesame (Sesamum indicum L.)’, *Plant OMICS*, 3(3), pp. 85–91.

Gabriel, K. R. (1971) ‘The Biplot Graphic Display of Matrices with Application to

Principal Component Analysis', *Biometrika*, 58(3), p. 453. doi: 10.2307/2334381.

Gao, Z. Y. *et al.* (2013) 'Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), pp. 14492–14497. doi: 10.1073/pnas.1306579110.

Gauch, H. G., & Kang, M. S. (1996) *AMMI analysis of yield trials. In: Genotype by environment interaction*. Kang & Gauch (Eds). CRC Press, Boca Raton, New York.

Gauch, H. G. and Zobel, R. W. (1996) 'AMMI analysis of yield trials. In: Genotype by environment interaction', *Kang & Gauch (Eds). CRC Press, Boca Raton, New York.*

Gauch, H. G. (2006) 'Statistical Analysis of Yield Trials by AMMI and GGE', *Crop Science*, 46, p. 1488.

Gauch, H. G., Piepho, H. P. and Annicchiarico, P. (2008) 'Statistical analysis of yield trials by AMMI and GGE: Further considerations', *Crop Science*, 48(3), pp. 866–889. doi: 10.2135/cropsci2007.09.0513.

Gebremichael, D. E. (2017) 'Sesame (Sesamum indicum L .) Breeding in Ethiopia', *International Journal of Novel Research in Life Sciences*, 4(1), pp. 1–11. Available at: file:///home/rnsarma/Downloads/SesameBreedinginEthiopia-924.pdf.

Gidey, Y. T., Kebede, S. A., & Gashawbeza, G. T. (2012) 'Extent and pattern of genetic diversity for morpho-agronomic traits in Ethiopian sesame landraces (Sesamum indicum L.)', *Asian Journal of Agricultural Research*, 6(3), pp. 118–128.

Gidey, Y. T., Kebede, S. A., & Gashawbeza, G. T. (2013) 'Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes', *International Journal of Plant Breeding and Genetics*, 7(1), pp. 21–34.

Global Agri Systems (2010) 'Dehulled and roasted sesame seed oil processing unit'.

Grüneberg, W. J. *et al.* (2005) 'Genotype x environment interactions for a diverse set of sweetpotato clones evaluated across varying ecogeographic conditions in Peru', *Crop Science*, 45(6), pp. 2160–2171. doi: 10.2135/cropsci2003.0533.

Guo, B. *et al.* (2013) 'Family-based association mapping in crop species', *Theoretical and Applied Genetics*, 126(6), pp. 1419–1430. doi: 10.1007/s00122-013-2100-2.

Gupta, P. K., Roy, J. K. and Prasad, M. (2001) 'Single nucleotide polymorphisms : A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants', (February).

Gupta, P. K., Rustgi, S. and Mir, R. R. (2008) 'Array-based high-throughput DNA markers for crop improvement', *Heredity*. doi: 10.1038/hdy.2008.35.

Habtamu Seboka and Million Fikreselassie (2013) 'Multivariate analysis of some Ethiopian field pea (Pisum sativum L.) genotypes', *International Journal of Genetics and Molecular Biology*, 5(6), pp. 78–87. doi: 10.5897/ijgmb2013.0080.

Hansen, R. (2011) 'Sesame profile'.

Heffner, E.L., Sorrells, M.E. and Jannink, J.L. (2009) 'Genomic selection for crop improvement', *Crop Science*, 49(1), pp. 1–12.

Huang, X. H., Wei, X. H., Sang, T., Zhao, Q. A., Feng, Q., Zhao, Y., et al. (2010) 'Genome-wide association studies of 14 agronomic traits in rice landraces', *Nat.*

Genet., pp. 42, 961–976. doi: 10.1038/ng.695.

Huang, X. H., Zhao, Y., Wei, X. H., Li, C. Y., Wang, A. H., Zhao, Q., et al. (2012) 'Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm', *Nat. Genet.*, pp. 44, 32–39. doi: 10.1038/ng.1018.

Huang, X. and Han, B. (2014) 'Natural variations and genome-wide association studies in crop plants', *Annual Review of Plant Biology*, 65, pp. 531–551. doi: 10.1146/annurev-arplant-050213-035715.

Huang, Y. F. *et al.* (2014) 'Using Genotyping-By-Sequencing (GBS) for genomic discovery in cultivated oat', *PLoS ONE*. doi: 10.1371/journal.pone.0102448.

Institute of Biodiversity Conservation (IBC) (2012) *Ethiopia: Third country report on the state of plant genetic resources for food and agriculture.* Addis Ababa, Ethiopia.

IPGRI and NBPGR (International Plant Genetic Resources Institute and National Bureau of Plant Genetic Resources) (2004) *Descriptors for sesame (Sesamum spp.)*, *Current Sociology*. National Bureau of Plant Genetic Resources, New Delhi, India. doi: 10.1177/001139216901700106.

Ismaila A, U. A. (2012) 'Genetic Variability for Yield and Yield Components in Sesame (Sesamum indicum L.)', *International Journal of Science and Research (IJSR)*, 3(9), pp. 63–66.

Jarvis, D. I., & Hodgkin, T. (1998) 'Wild relatives and crop cultivars: conserving the connection', in *In The Proceedings of an International Symposium on in situ conservation of plant genetic diversity*. George Allen & Unwin, London, UK, p. (pp. 163-179).

Jarvis, D. I., & Hodgkin, T. (1999) (1999) 'Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems', *Molecular Ecology*, 8, pp. S159–S173.

JMP® (2002) 'User's Guide, Version 5'. SAS Institute Inc., Cary, NC, USA.

Johnson, Herbert W., H. F. Robinson, and R. E. C. (1955) 'Estimates of genetic and environmental variability in soybeans', *Agronomy journal*, 47(7), pp. 314–318.

Judson A Ward1*, J. B. *et al.* (2013) 'Saturated linkage map construction in Rubus idaeus using genotyping by sequencing and genome-independent imputation', *BMC Genomics*, 14(2), p. 160108. doi: doi: 10.1186/1471-2164-14-2.

Karuppaiyan, R., & Ramasamy, P. (2000) 'Cause and effect relationship between seed yield and its components in sesame', *Madras Agricultural Journal*, 87(1/3), pp. 74–76.

Kathiresan G, and G. P. (2000) 'Studies on seed yield-contributing characters in sesame', *Sesame and Safflower Newsletter*, 15, pp. 29–32.

Kefyalew, G. (2012) 'Analysis of smallholder farmer's participation in production and marketing of export potential crops: The case of sesame in diga district, East Wollega zone of Oromia Regional State', (November 2012). doi: 10.13140/RG.2.1.4777.2406.

Kilian, A. *et al.* (2003) 'The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement', *Proceedings of the International Congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution", 27-31 May*, (May 2003), pp. 443–461.

Kilian, A. *et al.* (2012a) 'Diversity arrays technology: A generic genome profiling

technology on open platforms', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-870-2_5.

Kilian, A. *et al.* (2012b) 'Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms BT - Data Production and Analysis in Population Genomics: Methods and Protocols', in *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-870-2_5.

Kilian, A., Sanewski, G. and Ko, L. (2016) 'The application of DArTseq technology to pineapple', in *Acta Horticulturae*. doi: 10.17660/ActaHortic.2016.1111.27.

Kim, D.H., Zur, G., Danin-Poleg, Y., Lee, S.W., Shim, K.B., Kang, C.W. and Kashi, Y. (2002) 'Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats', *Plant Breeding*, 121(3), pp. 259-262.

Kim, D. H. *et al.* (2002) 'Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats', *Plant Breeding*, 121(3), pp. 259-262. doi: 10.1046/j.1439-0523.2002.00700.x.

Kindie Aysheshm (2007) *Sesame market chain analysis: The case of Metema woreda, North Gondar zone, Amhara national Regional state*. Haramaya University.

Korte, A. and Ashley, F. (2013) 'The advantages and limitations of trait analysis with GWAS : a review Self-fertilisation makes Arabidopsis particularly well suited to GWAS', *Plant methods*, 9(1), p. 29.

Krishnaiah, G., Reddy, K. R., & Sekhar, M. R. (2002) 'Variability studies in sesame', *CROP RESEARCH-HISAR*, 24(3), pp. 501–504.

Kumar, H., Kaur, G. and Banga, S. (2012) 'Molecular Characterization and Assessment of Genetic Diversity in Sesame (Sesamum indicum L.) Germplasm Collection

Using ISSR Markers', *Journal of Crop Improvement*. doi: 10.1080/15427528.2012.660563.

Laidò, G. *et al.* (2013) 'Genetic Diversity and Population Structure of Tetraploid Wheats (Triticum turgidum L.) Estimated by SSR, DArT and Pedigree Data', *PLoS ONE*. doi: 10.1371/journal.pone.0067280.

Lakhanpaul, S., Singh, V., Kumar, S., Bhardwaj, D. and Bhat, K. V. (2012) 'Sesame: overcoming the abiotic stresses in the queen of oilseed crops', *Improving crop resistance to abiotic stress*, pp. 1251–1283.

Langham, D.R. and Wiemers, T. (2002) 'Progress in mechanizing sesame in the US through breeding. In Trends in new crops and new uses', in *Proceedings of the Fifth National Symposium, Atlanta, Georgia, USA, 10-13 November*, pp. 157–173.

Langham, R. (2007) 'Phenology of Sesame', *New crops and new uses*, pp. 144–182.

Laurentin, H. E. and Karlovsky, P. (2006) 'Genetic relationship and diversity in a sesame (Sesamum indicum L .) germplasm collection using amplified fragment length', *BMC Genetics*, 10, pp. 1–10. doi: 10.1186/1471-2156-7-10.

Lee, J., Lee, Y. and Choe, E. (2008) 'Effects of sesamol, sesamin, and sesamolin extracted from roasted sesame oil on the thermal oxidation of methyl linoleate', *LWT - Food Science and Technology*, 41(10), pp. 1871–1875. doi: 10.1016/j.lwt.2007.11.019.

Li, Y. *et al.* (2011) 'Natural variation in GS5 plays an important role in regulating grain size and yield in rice', *Nature Genetics*, 43(12), pp. 1266–1269. doi: 10.1038/ng.977.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z. (2012) 'GAPIT: genome association and prediction integrated tool', *Bioinformatics*, 8(18), pp. 2397–2399.

Liu, J. & Muse, S. V. (2005) 'PowerMarker: Integrated analysis environment for genetic marker data', *Bioinformatics*.

Ma, B. L. *et al.* (2004) 'Graphic Analysis of Genotype, Environment, Nitrogen Fertilizer, and Their Interactions on Spring Wheat Yield', *Agronomy Journal*, 96(March), pp. 169–180.

Mangini, G. *et al.* (2013) 'Genetic Diversity and Population Structure of Tetraploid Wheats (Triticum turgidum L .) Estimated by SSR , DArT and Pedigree Data', 8(6). doi: 10.1371/journal.pone.0067280.

Marathi, B. *et al.* (2012) 'QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (Oryza sativa L.)', *BMC Plant Biology*, 12. doi: 10.1186/1471-2229-12-137.

Matthies, I. E. *et al.* (2012) 'Population structure revealed by different marker types (SSR or DArT) has an impact on the results of genome-wide association mapping in European barley cultivars', *Molecular Breeding*. doi: 10.1007/s11032-011-9678-3.

Mei, H. *et al.* (2017) 'High-Density Genetic Map Construction and Gene Mapping of Basal Branching Habit and Flowers per Leaf Axil in Sesame Plant Materials and Trait Investigation', 8(April), pp. 1–10. doi: 10.3389/fpls.2017.00636.

Million, F. (2012) 'Variability, heritability and association of some morpho-agronomic traits in field pea (Pisum sativum L.) genotypes', *Pakistan Journal of Biological*

Sciences, 15(8), pp. 358–366. doi: 10.3923/pjbs.2012.358.366.

Misganaw, M., Mekbib, F. and Wakjira, A. (2015) 'Genotype x environment interaction on sesame (Sesamum indicum L.) seed yield', *African Journal of Agricultural Research*, 10(21), pp. 2226–2239. doi: 10.5897/AJAR2015.9748.

Mkamilo, G. S. and Bedigian, D. (2007) 'In PROTA (Plant Resources of Tropical Africa/Ressources végétales de l'Afrique tropicale), Wageningen, Netherlands'.

MoARD (Ministry of Agriculture and Rural Development) (no date) *Crop Variety Register Book. Animal and Plant health Regulatory Directorate. Addis Abeba, Ethiopia 2010-2017*.

Moghaddam, M. J., & Pourdad, S. S. (2009) 'Comparison of parametric and non-parametric methods for analyzing genotype × environment interactions in sunflower (Carthamus tinctorius L.)', *The Journal of Agricultural Science,* 147, pp. 601–612. doi: https://doi.org/10.1017/S0021859609990050.

Monpara, B. and Khairnar, S. (2016) 'Heritability and predicted gain from selection in components of crop duration and seed yield in sesame (*Sesamum indicum* L.)', *Plant Gene and Trait*, 7(2), pp. 1–5. doi: 10.5376/pgt.2016.07.0002.

Morinaga, T., Fukushima, E., Kano, T., Maruyama, Y., & Yamasaki, Y. (1929) 'Chromosome numbers of cultivated plants II', *Shokubutsugaku Zasshi*, 43(515), pp. 589–594.

Morrell, P. L., Buckler, E. S. and Ross-Ibarra, J. (2012) 'Crop genomics: Advances and applications', *Nature Reviews Genetics*. Nature Publishing Group, 13(2), pp. 85–96. doi: 10.1038/nrg3097.

Mulugeta Atnaf Tiruneh, Hussein Mohammed Ali and, H. Z. (2013) 'Estimation of Better

Parent and Economic Heterosis for Yield and Associated Traits in Common Beans', *Journal of Applied Biosciences,* 71(1), pp. 5706– 5714. doi: 10.4314/jab.v71i1.98814.

Mulugeta, G., Dessalegn, Y. and Genet, T. (2014) 'Genotype-environment interaction and correlation of some stability parameters of total starch yield in potato in Amhara region, Ethiopia', *Journal of Plant Breeding and Crop Science,* 6(3), pp. 31–40. doi: 10.5897/jpbcs2013.0426.

Narayanan, A., & Reddy, K. B. (1982) 'Growth, development and yield of sesame (Sesamum indicum L.) cultivars', *Field crops research*, 5, pp. 217–224.

Nath, R., Chakraborty, P. K., & Chakraborty, A. (2001) 'Effect of climatic variation on yield of sesame (Sesamum indicum L.) at different dates of sowing', *Journal of Agronomy and Crop Science*, 186(2), pp. 97–102.

Nei, M. (1972) 'Genetic distance between populations', *The American Naturalist*, 106(949), pp. 283–292.

Nordborg, M. and Weigel, D. (2008) 'Next-generation genetics in plants', *Nature*, 456(7223), pp. 720–723. doi: 10.1038/nature07629.

Park, Y. C. J., Ra, C. L. W. and Lee, J. C. J. (2011) 'Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L .) using microsatellite markers', *Genes & Genomics*, 33(April), pp. 187–195. doi: 10.1007/s13258-010-0130-6.

Pastorello, E.A., Pravettoni, V., Trambaioli, C., Pompei, C., Brenna, O., Farioli, L. and Conti, A. (2001) 'Lipid transfer proteins and 2S albumins as allergens', *Allergy*, 56, pp. 45–47.

Pfender, W.F., Saha, M.C., Johnson, E.A. and Slabaugh, M. B. (2011) 'Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in Lolium perenne', *Theoretical and Applied Genetics*, 122(8), pp. 1467–1480.

Pickup, M. *et al.* (2013) 'Source population characteristics affect heterosis following genetic rescue of fragmented plant populations', *Proceedings of the Royal Society B: Biological Sciences*, 280(1750). doi: 10.1098/rspb.2012.2058.

van Poecke, R. M. P. *et al.* (2013) 'Sequence-based SNP genotyping in durum wheat', *Plant Biotechnology Journal*, 11(7), pp. 809–817. doi: 10.1111/pbi.12072.

Poland, J. A. *et al.* (2012) 'Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach', *PLoS ONE*, 7(2). doi: 10.1371/journal.pone.0032253.

Poland, J. A. and Rife, T. W. (2012) 'Genotyping-by-Sequencing for Plant Breeding and Genetics', *The Plant Genome*. doi: 10.3835/plantgenome2012.05.0005.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) *Structure Software for Population Genetics Inference (V.2.3.4), Pritchard Lab, Stanford University*. doi: V2.3.4, Jul 2012.

Purchase, J. (1997) 'Parametric analysis to describe genotype by environment interaction and yield stability in winter wheat', *PhD Thesis*, p. 83.

Purchase, J. L., Hatting, H. and van Deventer, C. S. (2000) 'Genotype × environment interaction of winter wheat (Triticum aestivum L.) in South Africa: II. Stability analysis of yield performance', *South African Journal of Plant and Soil*, 17(3), pp. 101–107. doi: 10.1080/02571862.2000.10634878.

Rabbani, M. A., Masood, M. and Shinwari, Z. K. (2010) 'GENETIC ANALYSIS OF BASMATI AND NON-BASMATI PAKISTANI RICE (ORYZA SATIVA L .) CULTIVARS USING MICROSATELLITE MARKERS', (August).

Raman, H. *et al.* (2014) 'Genome-wide delineation of natural variation for pod shatter resistance in Brassica napus', *PLoS ONE*. doi: 10.1371/journal.pone.0101673.

Reddy, P. A. V., Sekhar, M. R., Ranganatha, A. R. G., & Dhanraj, A. (2001) 'Genetic variability and heritability for seed yield and its components in sesame (Sesamum indicum L.)', *Journal of Oilseeds Research*, 18(2), pp. 173–175.

Robinson, H. F., Comstock, R. E., & Harvey, P. H. (1949) 'Estimates of Heritability and the Degree of Dominance in Corn 1', *Agronomy journal*, 41(8), pp. 353–359.

Samonte, S. O. P. B. *et al.* (2005) 'Targeting cultivars onto rice growing environments using AMM and SREG GGE biplot analyses', *Crop Science*, 45(6), pp. 2414–2424. doi: 10.2135/cropsci2004.0627.

Sánchez-sevilla, J. F., Horvath, A. and Botella, M. A. (2015) 'Diversity Arrays Technology (DArT) Marker Platforms for Diversity Analysis and Diversity Arrays Technology (DArT) Marker Platforms for Diversity Analysis and Linkage Mapping in a Complex Crop , the Octoploid Cultivated Strawberry (Fragaria × ananassa', (December). doi: 10.1371/journal.pone.0144960.

Sanni, K. A. *et al.* (2009) 'Additive main effects and multiplicative interactions analysis of grain yield performances in rice genotypes across environments', *Asian Journal of Plant Sciences*, 8(1), pp. 48–53. doi: 10.3923/ajps.2009.48.53.

SAS (2002) 'System Analysis Software. Version 9.1.2.', *SAS Institute INC. Cary, North Carolina, USA*.

Schaal, B. A. *et al.* (1998) 'Phylogeographic studies in plants: Problems and prospects', *Molecular Ecology*. doi: 10.1046/j.1365-294x.1998.00318.x.

Seegler, C. J. . (1983) *Oil plants in Ethiopia,their taxonomy and agricultural significance, Oil Plants in Ethiopia, Their Taxonomy and Agricultural Significance*.

Shadakshari, Y. G., Virupakshappa, K., & Shivashankar, G. (1995) 'Genetic variability studies in germplasm collection of sesame (Sesamum indicum L.)', *Mysore Agricultural Research Journal*, 29, pp. 133–137.

Sharaby, N. and Butovchenko, A. (2019) 'Cultivation technology of sesame seeds and its production in the world and in Egypt', *IOP Conference Series: Earth and Environmental Science*, 403(1). doi: 10.1088/1755-1315/403/1/012093.

Sharma, J. R. (1998) 'Statistical and Biometrical Techniques in Plant Breeding. New Age International (P) Limited Publishers, New Delhi', p. 432.

Shawn R. Narum, C. Alex Buerkle, John W. Daveys, M. R. M., And and Hohenlohe, P. A. (2013) 'Genotyping-by-sequencing in ecological and conservation genomics', *Mol Ecol*, 22(11), pp. 2841–2847. doi: 10.1111/mec.12350.

Shim, K.B., Shin, S.H., Shon, J.Y., Kang, S.G., Yang, W.H. and Heu, S. G. (2016) 'Classification of a collection of sesame germplasm using multivariate analysis', *Journal of crop science and biotechnology*, 19(2), pp. 151–155.

Signor, C. E. L., Dousse, S., Lorgeou, J., Denis, J. B., Bonhomme, R., Carolo, P., & Charcosset, A. (2001) 'Interpretation of genotype× environment interactions for early maize hybrids over 12 years', *Crop Science*, 41(3), pp. 663–669.

Sileshi AM (2008) *Genetic Divergence and Correlation Study in Sesame*

(Sesamumindicum L.) Genotypes. M.Sc thesis. Addis Ababa University, Ethiopia.

Simmonds, N. W. and Purseglove, J. W. (1969) 'Tropical Crops. Dicotyledons.', *The Journal of Ecology*. doi: 10.2307/2258508.

Singh, H. C., Nagaich, V. P., & Singh, S. K. (2000) 'Genetic variability for dry matter production in sesame (Sesamum indicum L.)', *Annals of Agricultural Research*, 21(3), pp. 323–327.

Singh, K.B., Bejiga, G. and Malhotra, R. . (1990) 'Associations of some characters with seed yield in chickpea collections', *Euphytica*, 49(1), pp. 83–88.

Singh, RK., Chaudhary, B. (1985) 'Biometrical methods in quantitative genetic analysis', *Kalyani Publishers, New Delhi*.

Sivasubramanian S, M. M. (1973) 'Heterosis and inbreeding depression in rice', *Madras Agric. J*, 60(7), pp. 1139–1140.

Slatkin, M. (1987) 'Gene flow and the geographic structure of natural populations', *Science*. doi: 10.1126/science.3576198.

Solanki, Z. S., & Gupta, D. (2003) 'Variability and character association among quantitative characters of sesame', *JOURNAL OF OILSEEDS RESEARCH*, 20, pp. 276–277.

Sonah, H. *et al.* (2013) 'An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping', *PLoS ONE*. doi: 10.1371/journal.pone.0054603.

Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., et al. (2015) 'Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice', *Proc. Natl. Acad. Sci. U.S.A*, pp.

112, 76–81. doi: 10.1073/pnas.1421127112.

Steel, R. G. and Torrie, J. H. (1980) 'Principles and procedures of statistics', *McGraw-Hili, New York.*

Suddihiyam, P., Steer, B. T., & Turner, D. W. (1992) 'The flowering of sesame (Sesamum indicum L.) in response to temperature and photoperiod', *Australian Journal of Agricultural Research*, 43(5), pp. 1101–1116.

Sudhakar, N., Sridevi, O., & Salimath, P. M. (2007) 'Variability and character association analysis in sesame, Sesamum indicum L', *JOURNAL OF OILSEEDS RESEARCH*, 24(1), p. 56.

Sumathi, P. and Muralidharan, V. (2010) 'Analysis of genetic variability, association and path analysis in the hybrids of sesame (Sesamum indicum L)', *Tropical Agricultural Research and Extension*, 13(3), p. 63. doi: 10.4038/tare.v13i3.3140.

Tang, J. *et al.* (2015) 'DArT-based characterisation of genetic diversity in a Miscanthus collection from Poland', *Planta*. doi: 10.1007/s00425-015-2335-z.

Teshome, D., Kassahun, T. and Bekele, E. (2015) 'Genetic diversity of sesame germplasm collection (SESAMUM INDICUM L .): implication for conservation , improvement and use', (February). doi: 10.5897/IJBMBR2014.0219.

Timothy M Beissinger*,§, Candice N Hirsch†,**, R. S. S. *et al.* (2013) 'Marker density and read-depth for genotyping populations using genotyping-by-sequencing', *Genetics*, 193, pp. 1073–1081. doi: 10.1534/genetics.112.147710.

Tunde-Akintude, T. ., Oke M, O. and Akintunde B.O (2012) *Sesame Seed, oil seeds*. Edited by Dr.Uduak G Akpan. In Tech.

Uncu, A. O. *et al.* (2016) 'High-throughput single nucleotide polymorphism (SNP)

identification and mapping in the sesame (Sesamum indicum L.) genome with genotyping by sequencing (GBS) analysis', *Molecular Breeding*, 36(12). doi: 10.1007/s11032-016-0604-6.

'USAID-Ethiopia Agribusiness and Trade Expansion; Market Bulletin' (2010), No. 9.

Vavilov, N. I. (1951) 'The origin, variation, immunity and breeding of cultivated plants', 72, p. 482.

VSN International (2015) 'Genstat Reference Manual (Release 18), Part 1 Summary', *VSN International, Hemel Hempstead, UK*.

Wang, L. *et al.* (2016) 'Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map', *BMC Genomics*. BMC Genomics, pp. 1–13. doi: 10.1186/s12864-015-2316-4.

Wang, Y. *et al.* (2019) 'Collection and Evaluation of Genetic Diversity and Population Structure of Potato Landraces and Varieties in China', 10(February), pp. 0–10. doi: 10.3389/fpls.2019.00139.

WARC (Werer Agricultural Research Center) (2006) *Progress report. Lowland Oilcrops Research Division, Werer, Ethiopia*.

WARC (Werer Agricultural Research Center) (2015) 'Progress report. Lowland Oilcrops Research Division, Werer Ethiopia'.

Ward, J. H. (1963) 'Hierarchical Grouping to Optimize an Objective Function', *Journal of the American Statistical Association*. doi: 10.1080/01621459.1963.10500845.

Wei, W., Zhang, Y., Lv, H., Li, D., Wang, L., and Zhang, X. (2013) 'Association analysis for quality traits in a diverse panel of Chinese sesame (Sesamum indicum L.) germplasm', *J. Integr. Plant Biol*, 55, pp. 745–758. doi: 10.1111/jipb.1204.

Wei, X. *et al.* (2012) 'Domestication and geographic origin of Oryza sativa in China: Insights from multilocus analysis of nucleotide variation of O. sativa and O. rufipogon', *Molecular Ecology*. doi: 10.1111/j.1365-294X.2012.05748.x.

Wei, X. *et al.* (2014) 'Development of Simple Sequence Repeat (SSR) Markers of Sesame (Sesamum indicum) from a Genome Survey', *Molecules*, 19, pp. 5150–5162. doi: 10.3390/molecules19045150.

Wei, X. *et al.* (2015) 'Genetic discovery for oil production and quality in sesame', *Nature Communications*. Nature Publishing Group, (May). doi: 10.1038/ncomms9609.

Weiss, E. A. (1971) 'Castor, sesame and safflower', *Ed. Leo nard Hill (Londres)*, pp. 45–90.

Weiss, E. A. (1983) 'sesame', *Oil seed crops. Longman, London*, pp. 282–340.

Wenzl, P. *et al.* (2004) 'Diversity Arrays Technology (DArT) for whole-genome profiling of barley', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0401076101.

Wijnands J, Biersteker J, and H. R. (2007) 'Oilseeds Business Opportunities in Ethiopia Survey report, Ministry of Agriculture, Nature and Food Quality', *The Netherlands, The Hague*, pp. 8–20.

Wricke G. Weber WE (1986) 'Quantitative genetics and selection in plant breeding', *Walter de Gruyter and Co. Berlin, Germany*. Available at: https://books.google.com/books?hl=en&lr=&id=9GdvJJnt3CgC&oi=fnd&pg=PR5&dq=Wricke+G,+Weber+WE+(1986)+Quantitative+genetics+and+selection+in+plant+breeding.+Walter+de+Gruyter+and+Co.+Berlin,+Germany&ots=w30K7L-5WI&sig=gs35N9nxceQovAtiNStq4YGHDnA.

Wu, K. *et al.* (2014) 'High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (Sesamum indicum L .) based on RAD-Seq techonology', pp. 1–14.

Xu, P. *et al.* (2013) 'QTL mapping and epistatic interaction analysis in asparagus bean for several characterized and novel horticulturally important traits', *BMC Genetics*, 14. doi: 10.1186/1471-2156-14-4.

Yadav, A. K., Thakral, S. K., Hari, S., Phool, S., & Tadav, T. P. (1988) 'Effect of morpho-physiological attributes on PAR absorption and seed yield in sesame (Sesamum indicum L.).', *J. Oilseeds Research*, 5, pp. 155–158.

Yan, W., & Kang, M. S. (2003) 'GGE biplot analysis: A graphical tool for breeders, geneticists, and agronomists', *Boca Raton, FL: CRC Press*, pp. 63–88.

Yan, W., Hunt, A. L., Sheng, Q. and Szlavnics, Z. (2000) 'Cultivar evaluation and mega-environment investigation based on GGE- biplot', *Crop Science*, 40, pp. 596–605.

Yan, W. (2002) 'Singular-value partitioning in biplot analysis of multienvironment trial data', *Agronomy Journal*, 94(5), pp. 990–996. doi: 10.2134/agronj2002.0990.

Yan, W. *et al.* (2007) 'GGE biplot vs. AMMI analysis of genotype-by-environment data', *Crop Science*, 47(2), pp. 643–655. doi: 10.2135/cropsci2006.06.0374.

Yan, W. *et al.* (2010) 'Identifying essential test locations for oat breeding in Eastern Canada', *Crop Science*, 50(2), pp. 504–515. doi: 10.2135/cropsci2009.03.0133.

Yan, W. (2014) 'Genotype-by-Location two-way data analysis. In Crop Variety Trials: Data Management and Analysis', *(Eds W. Yan). New York: John Wiley & Sons, INC*, pp. 133–162.

Yan, W. and Tinker, N. A. (2005) 'An integrated biplot analysis system for displaying,

interpreting, and exploring genotype x environment interaction', *Crop Science*, 45(3), pp. 1004–1016. doi: 10.2135/cropsci2004.0076.

Yan, W. and Tinker, N. A. (2006) 'Biplot analysis of multi-environment trial data: Principles and applications', *Canadian Journal of Plant Science*, 86(3), pp. 623–645. doi: 10.4141/P05-169.

Yang, S. *et al.* (2006) 'Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology', *Theoretical and Applied Genetics*. doi: 10.1007/s00122-006-0317-z.

Yang, W. *et al.* (2014) 'Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice', *Nature Communications*. Nature Publishing Group, 5, pp. 1–9. doi: 10.1038/ncomms6087.

Yebiyo Woldemariam (1985) 'Sesame Adaptation test in different agro ecological zones of Ethiopia.', *In: Abbas Omran (ed.) Oil crops : Sesame and Safflower. Proceedings of the Second Oil Crops Network Workshop held in Hyderabad,India. IDRC, Canada India.*, pp. 162–167.

Yirga Belay (2016) *Genotype x Environment Interaction and Yield Stability of White Seeded Sesame (Sesamumindicum L.) Genotypes in Northern Ethiopia*. Haramaya University, Haramaya, Ethiopia.

Yirgalem TG, S. A. and G. T. (2013) 'Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes', *International Journal of Plant Breeding and Genetics*, 7(1), pp. 21–34.

Yol, E., and Uzun, B. (2012) 'Geographical patterns of sesame (Sesamum indicum L.)

accessions grown under Mediterranean environmental conditions, and establishment of a core collection', *Crop Sci.*, 52, pp. 2206–2214. doi: 10.2135/cropsci2011.07.0355.

Yu, H. *et al.* (2011) 'Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers', *PLoS ONE*, 6(3). doi: 10.1371/journal.pone.0017595.

Yu, J. *et al.* (2006) 'A unified mixed-model method for association mapping that accounts for multiple levels of relatedness', *Nature Genetics*, 38(2), pp. 203–208. doi: 10.1038/ng1702.

Zaman MA, Tuhina-Khatun, M, Ullah MZ, M. and M, and A. K. (2011) 'Genetic Variability and Path Analysis of Groundnut (Arachis hypogaea L.)', *The Agriculturists*, 9(1–2), pp. 29–36. doi: 10.3329/agric.v9i1-2.9476.

Zhang, H. *et al.* (2016) 'Ultra-dense SNP genetic map construction and identification of SiDt gene controlling the determinate growth habit in Sesamum indicum L.', *Scientific Reports*. Nature Publishing Group, 6(July), pp. 1–13. doi: 10.1038/srep31556.

Zhang, Y. *et al.* (2012) 'Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection', *BMC Genetics*. BMC Genetics, 13(1), p. 1. doi: 10.1186/1471-2156-13-102.

Zhang, Y. *et al.* (2013) 'Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing', (2).

Zhang, Y. X. *et al.* (2010) 'Analysis of genetic diversity among indigenous landraces

from sesame (Sesamum indicum L.) core collection in China as revealed by SRAP and SSR markers', *Genes and Genomics*. doi: 10.1007/s13258-009-0888-6.

Zhang, Z. *et al.* (2014) 'Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies', 9(3), pp. 1–12. doi: 10.1371/journal.pone.0093017.

Zhou, Q. *et al.* (1998) 'Genetics and Analysis of Quantitative Traits', in *Sunderland, MA: Sinauer Associates Inc.*.

Zhou, R. *et al.* (2018) 'Genome-Wide Association Studies of 39 Seed Yield-Related Traits in Sesame (Sesamum indicum L .)', *INTERNATIONAL JOURNAL OF MOLECULAR SCIENCE*, 19. doi: 10.3390/ijms19092794.

Zhu, C. *et al.* (2008) 'Status and Prospects of Association Mapping in Plants', pp. 5–20. doi: 10.3835/plantgenome2008.02.0089.

Zobel, R. W., Wright, M. J. and Gauch, H. G. (1988) 'Statistical analysis of yield trial', *Agronomy journal*, 80, pp. 388–393.

Zohary, D., Hopf, M. and Weiss, E. (2012) *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*, *OXFORD*. doi: 10.1093/acprof:osobl/9780199549061.001.0001.

APPENDICS

Appendix 1: Passport description of tested Genotypes.

Molecular lab code	EBI and Research center code	Region/country	Zone	District	Latitude	Altitude	Longitude
GEN1	9015	Benshangul-Gumz	METEKEL	GUBA	07-46-19-N	855	12-46-19-E
GEN2	216896	Oromia	ARSSI	MERTI		1570	
GEN3	ACC-BG007	Benshangul-Gumz					
GEN4	241323	Tigray	MIRABAWI	KAFTA HUMERA		600	
GEN5	Acc-051-02-sel 1 (2)	Benshangul-Gumz					
GEN6	238269	Tigray	MIRABAWI	KAFTA HUMERA	14-02-00-N	820	36-05-00-E
GEN7	203600	Zimbabwe			20-18-00-S	700	32-49-00-E
GEN8	ACC-203-336-2	Zimbabwe					
GEN9	BOUNJA FILWAHA-2	Amhara	SEMEN GONDARA	METEMA			
GEN10	205194	Borkina Faso			12-29-00-N	700	22-31-00-E
GEN11	Najjoo-68 (gabaa kamijaa)	Oromia	West Wellega	Mana sibu			
GEN12	Bawal-03	Tigray	MIRABAWI	k/humera	14.05933	631	36.56701
GEN13	9691	Tigray	MIRABAWI	KAFTA HUMERA	14-06-69-N	799	36-57-04-E
GEN14	241342	Amhara	SEMEN GONDAR	METEMA		600	
GEN15	207957	Gambella	ZONE 1	GAMBELA	08-15-00-N		34-30-00-E
GEN16	205195	Borkina Faso			12-29-00-	700	22-31-00-E

					N		
GEN17	Lugudi-01	Tigray	MIRABAWI	k/humera	13.98959	639	36.39773
GEN18	241341	Amhara	SEMEN GONDAR	QUARA		600	
GEN19	205197	Borkina Faso					
GEN20	GOBYO-82-3	Amhara	METEMA				
GEN21	Ass-acc-128	Benshangul-Gumz	Assosa	Oda			
GEN22	203639	Kenya			00-18-00-S		34-46-00-E
GEN23	MAN-66(MATIL)	Oromia	West Wellega	Mana sibu			
GEN24	Adebay-02	Tigray	MIRABAWI	k/humera	14.19388	660	36.76001
GEN25	17704	Oromia	MIRAB WELLEGA	ANFILO	08-25-05-N	1517	34-35-07-E
GEN26	CHALESA	Oromia	East Wellega				
GEN27	17700	Oromia	MISRAK WELLEGA	HARO LIMO	09-46-50-N	2190	36-10-38-E
GEN28	203104	Oromia	JIMMA	SOKORU		1400	
GEN29	Shelela-01	Tigray	MIRABAWI	k/humera	14.13313	634	36.59771
GEN30	23555	Benshangul-Gumz	METEKEL	GUBA	11-08-08-N	554	35-10-45-E
GEN31	241319	Amhara	SEMEN GONDAR	ADDI ARKAY			
GEN32	B-02	Oromia	Illibabur	Bedele			
GEN33	Backer	Tigray	MIRABAWI	k/humera	13.96909	779	36.85813
GEN34	9020	Benshangul-Gumz	METEKEL	GUBA	11-01-6 -N	524	35-38-28-E
GEN35	NAJOO-73	Oromia	West Wellega	Nejo			
GEN36	Hirhir nugar 1st set-2	Tigray	MIRABAWI	k/humera	13.87835	707	36.62495
GEN37	28300	Amhara	SEMEN GONDAR	Kora	08-14-61-N	708	13-37-36-E
GEN38	Banat-01	Tigray	MIRABAWI	k/humera	13.78043	593	36.48987
GEN39	Manasbu-62(mandii)	Oromia	West Wellega	Mana sibu			

GEN40	28328	Amhara	SEMEN GONDAR	T/Armacho	13-15-07-N	1010	14-40-91-E
GEN41	Adebay	Tigray	MIRABAWI	k/humera	14.19388	660	36.76001
GEN42	9019	Benshangul-Gumz	METEKEL	GUBA	11-01-17-N	518	35-39-32-E
GEN43	241334	Amhara	SEMEN GONDAR	METEMA		750	
GEN44	Mana siba-64(mandil)	Oromia	West Wellega	Mana sibu			
GEN45	BAHA ZEYT	Oromia	East Wellega				
GEN46	9027	Benshangul-Gumz	METEKEL	GUBA	11-10-45-N	665	35-40-58-E
GEN47	GIIBII-27(TOLE)	Oromia	West Wellega	Gimbi			
GEN48	9023	Benshangul-Gumz	KAMASHI	SIRBAABAY	11-12-20-N	552	35-03-7 –E
GEN49	28316	Amhara	SEMEN GONDAR	Metema	01-91-68-N	872	14-52-47-E
GEN50	Local-90-1	Tigray	MIRABAWI	k/humera	13.90069	669	36.53748
GEN51	Hirhir adigoshu sel-8	Tigray	MIRABAWI	k/humera	14.16507	1132	37.30969
GEN52	17708	Oromia	MIRAB WELLEGA	BEGI	09-14-46-N	1793	34-32-53-E
GEN53	ACC-203-030	Zimbabwe					
GEN54	9692	Tigray	MIRABAWI	KAFTA HUMERA	14-12-88-N	652	36-45-36-E
GEN55	OBSA	Oromia	East Wellega				
GEN56	241308	Tigray	MIRABAWI	TAHTAY ADIYABO		1130	
GEN57	SETIT-1	Tigray	HUMERA				
GEN58	241295	Amhara	OROMIYA	BATI		1380	
GEN59	ACC-0065	Tigray					
GEN60	9687	Tigray	MIRABAWI	SHIRARO/BADME	14-30-49-N	1047	37-47-17-E
GEN61	235905	Tigray	MIRABAWI	KAFTA HUMERA	14-18-00-N	820	37-24-00-E

GEN62	Ass-acc-74	Benshangul-Gumz	Kemashi	Kamashi			
GEN63	Mazegabu-01	Tigray	MIRABAWI	Welkait	13.53501°	874	37.07295°
GEN64	9693	Tigray	MIRABAWI	KAFTA HUMERA	13-52-43-N	812	36-52-84-E
GEN65	DICHO	Oromia	East wellega				
GEN66	17703	Oromia	MISRAK WELLEGA	GIDA AYANA	09-28-19-N	1365	36-32-04-E
GEN67	Selam	Tigray	MIRABAWI	Tsegede	13.50741°	835	36.37011°
GEN68	23565	Benshangul-Gumz	METEKEL	GUBA	11-16-05-N	858	35-17-17-E
GEN69	241305	Tigray	MIRABAWI	TAHTAY ADIYABO		1200	
GEN70	ACC-NO-026	Tigray	HUMERA				
GEN71	17696	Oromia	MISRAK WELLEGA	DIGA	09-02-49-N	1345	36-15-57-E
GEN72	241321	Amhara	SEMEN GONDAR	SANJA		1100	
GEN73	ACC-EW-011(2)	Oromia	East Wellega				
GEN74	BAHA NECHO	Oromia	East Wellega				
GEN75	NAJIJO-70(GABAA KAISAA)	Oromia	West Wellega	Nejo			
GEN76	28306	Amhara	SEMEN GONDAR	metema	02-25-64-N	760	14-08-03-E
GEN77	Mayweyni	Tigray	MIRABAWI	k/humera	13.72471	940	36.67827
GEN78	235769	Amhara	SEMEN GONDAR	METEMA	12-46-00-N	900	15-10-00-E
GEN79	238270	Tigray	MIRABAWI	KAFTA HUMERA	14-02-00-N	710	36-05-00-E
GEN80	238271	Tigray	MIRABAWI	KAFTA HUMERA	14-02-00-N	710	36-05-00-E
GEN81	Baunja fiyel kolet	Amhara	SEMEN GONDARA	METEMA			
GEN82	17706	Oromia	Kelem WELLEGA	YAMALAGI WALAL	08-44-50-N	1425	34-54-02-E

GEN83	241338	Amhara	SEMEN GONDAR	QUARA		600	
GEN84	ACC-NO-041	Tigray					
GEN85	SETIT-2	Tigray	HUMERA				
GEN86	Manasiibuu -60 (mandil)	Oromia	West Wellega	Mana sibu			
GEN87	28324	Amhara	SEMEN GONDAR	T/Armacho	03-09-77-N	900	14-54-64-E
GEN88	Maydelye-04	Tigray	MIRABAWI	Tsegede	-	-	-
GEN89	GULISO-71(GOSOLA)	Oromia	West Wellega	Nejo			
GEN90	212992	SNNP	SEMEN OMO	HUMBO	37-48-00-N	1630	06-43-00-E
GEN91	215816	Oromia	MISRAK WELLEGA	DIGA LEKA	08-15-00-N	1435	34-35-0 –E
GEN92	Ass-acc-3	Benshangul-Gumz	Assosa	Oda			
GEN93	18759	Tigray	MEHAKELEGNAW	TANGU ABERGELE	13-33-27-N	1709	39-01-47-E
GEN94	9021	Benshangul-Gumz	METEKEL	GUBA	11-01-6 -N	524	35-38-28-E
GEN95	28329	Amhara	SEMEN GONDAR	T/Armacho	03-14-65-N	987	14-44-83-E
GEN96	Shelela-02	Tigray	MIRABAWI	k/humera	14.18328	616	36.64878
GEN97	205185	Borkina Faso			13-05-00-N		01-05-00-E
GEN98	234016	Tigray	MIRABAWI	TAHTAY ADIYABO		1090	
GEN99	ACC-EWO12(16)	Oromia	East Wellega				
GEN100	241332	Amhara	SEMEN GONDAR	METEMA		750	
GEN101	9024	Benshangul-Gumz	METEKEL	GUBA	15-09-58-N	664	35-40-44-E
GEN102	28327	Amhara	SEMEN GONDAR	T/Armacho	03-13-37-N	957	14-47-02-E
GEN103	ADI	Benshangul-Gumz					
GEN104	Ass-acc-127	Benshangul-Gumz		Belojigenfoy			

GEN105	9022	Benshangul-Gumz	METEKEL	GUBA	11-15-16-N		35-17- 9-E
GEN106	202514	Benshangul-Gumz	ASOSA	ASOSA		1000	
GEN107	203640	Zimbabwe					
GEN108	Ass-acc-25	Benshangul-Gumz	Kemashi	Kamashi			
GEN109	241316	Amhara	SEMEN GONDAR	DEBARK			
GEN110	28315	Amhara	SEMEN GONDAR	Metema	01-91-85-N	722	14-64-10-E
GEN111	202513	Benshangul-Gumz	ASOSA	ASOSA		1000	
GEN112	241318	Amhara	SEMEN GONDAR	ADDI ARKAY			
GEN113	9690	Tigray	MIRABAWI	KAFTA HUMERA	14-04-00-N	933	37-02-41-E
GEN114	205191	Borkina Faso			13-20-00-N	800	22-53-00-E
GEN115	GONDAR-1	Oromia					
GEN116	9688	Tigray	MIRABAWI	KAFTA HUMERA	14-06-86-N	906	37-14-02-E
GEN117	Ebantuu-11 (qeloo)	Oromia	East Wellega	Ebantu			
GEN118	205196	Borkina Faso			10-25-00-S		04-58-00-E
GEN119	ACC-202-889	Amhara	Wello				
GEN120	28302	Amhara	SEMEN GONDAR	Kora	08-17-61-N	708	13-37-36-E
GEN121	17713	Oromia	MISRAK WELLEGA	NUNU QUMBA	08-49-42-N	1461	36-45-57-E
GEN122	Dansh-01	Tigray	MIRABAWI	Tsegede	13.55139	773	36.96491
GEN123	235404	Amara	SEMEN GONDAR	METEMA	12-40-00-N	1900	37-09-00-E
GEN124	17697	Oromia	MISRAK WELLEGA	DIGA	09-03-53-N	1360	36-13-30-E
GEN125	216733	Gambella	ZONE 2	ABOBO	07-56-00-N	600	34-41-00-E

GEN126	Banat-04	Tigray	MIRABAWI	k/humera	13.78043	593	36.48987
GEN127	ACC-EW-017(4)	Oromia	East Wellega				
GEN128	17694	Oromia	MISRAK WELLEGA	GITO GIDA	09-23-28-N	1352	36-29-55-E
GEN129	ACC-205-184	Borkina Faso					
GEN130	9017	Benshangul-Gumz	METEKEL	GUBA	11-20-10-N	778	35-39-21-E
GEN131	NAJOO-69 (GABAA KAMISA)	Oromia	West Wellega	Nejo			
GEN132	17707	Oromia	Kelem WELLEGA	DALE WABERA	08-51-09-N	1515	35-06-35-E
GEN133	9695	Tigray	MIRABAWI	TSEGEDE (WELKAIT)	13-44-95-N	770	36-56-29-E
GEN134	Bawal-02	Tigray	MIRABAWI	k/humera	14.0652	631	36.57556
GEN135	205190	Borkina Faso					
GEN136	17705	Oromia	Kelem WELLEGA	HAWA GELAN	08-39-24-N	1346	35-00-55-E
GEN137	237523	Tigray	MIRABAWI	TSELEMTI	11-57-00-N	900	38-41-00-E
GEN138	Bawal-01	Tigray	MIRABAWI	k/humera	14.07907	621	36.57278
GEN139	ACC ≠031	Tigray	HUMERA				
GEN140	17701	Oromia	MISRAK WELLEGA	IBANTU	10-05-59-N	2055	36-27-56-E
GEN141	Ass-acc-85	Benshangul-Gumz	Kemashi	Kamashi			
GEN142	241330	Amhara	SEMEN GONDAR	METEMA		750	
GEN143	23559	Benshangul-Gumz	METEKEL	GUBA	11-24-46-N	709	35-20-17-E
GEN144	28311	Amhara	SEMEN GONDAR	Metema	01-94-77-N	864	14-11-47-E
GEN145	23564	Benshangul-Gumz	METEKEL	GUBA	11-12-30-N	705	35-22-21-E
GEN146	Maygaba	Tigray	MIRABAWI	Welkait	13.77776	896	37.68234

GEN147	203102	Oromia	JIMMA	SOKORU		1110	
GEN148	17712	Oromia	MISRAK WELLEGA	JIMA ARJO	08-42-27-N	1522	36-26-41-E
GEN149	241309	Tigray	MIRABAWI	TAHTAY ADIYABO		1130	
GEN150	Shelela-03	Tigray	MIRABAWI	k/humera	14.19347	617	36.63557
GEN151	17709	Oromia	MISRAK WELLEGA	GOBU SEYO	09-01-04-N	1590	37-00-59-E
GEN152	Wuhdet	Tigray	MIRABAWI	k/humera	14.05508	900	37.17873
GEN153	DANGUR	Oromia	East Wellega				
GEN154	241343	Amhara	SEMEN GONDAR	CHILGA		980	
GEN155	241314	Amhara	SEMEN GONDAR	ADDI ARKAY		1460	
GEN156	Manjibu-63(mandii)	Oromia	West Wellega	Mana sibu			
GEN157	MECHAL	Oromia					
GEN158	28308	Amhara	SEMEN GONDAR	Metema	02-25-39-N	791	17-07-69-E
GEN159	241317	Amhara	SEMEN GONDAR	DEBARK			
GEN160	Maykeyh-01	Tigray	MIRABAWI	k/humera	14.03516	882	37.48927
GEN161	GULISO-72(GOOGOLA)	Oromia	West Wellega	Nejo			
GEN162	28304	Amhara	SEMEN GONDAR	Metema	02-04-50-N	730	14-23-17-E
GEN163	241322	Amhara	SEMEN GONDAR	SANJA		1090	
GEN164	Turka	Tigray	MIRABAWI	k/humera	13.79681	774	36.92588
GEN165	9026	Benshangul-Gumz	METEKEL	GUBA	11-07-1 - N	665	35-40-6 –E
GEN166	MANSIBU-61(MANDII)	Oromia	West Wellega	Mana sibu			
GEN167	TATI	Oromia	Hararghe - Babile				
GEN168	17710	Oromia	MISRAK WELLEGA	WAMA HAGELO	08-51-04-N	1582	36-48-19-E
GEN169	28314	Amhara	SEMEN GONDAR	Metema	01-91-49-	719	14-63-69-E

					N		
GEN170	241328	Amhara	SEMEN GONDAR	SANJA		1230	
GEN171	Ass-acc-71	Benshangul-Gumz	Kemashi	Agalo meti			
GEN172	28313	Amhara	SEMEN GONDAR	Metema	02-03-78-N	693	14-32-49-E
GEN173	HUMERA-1	Tigray	HUMERA				
GEN174	Maydelye-02	Tigray	MIRABAWI	Tsegede	13.41058°	889	36.97653°
GEN175	28318	Amhara	SEMEN GONDAR	Metema	02-10-13-N	744	14-25-70-E
GEN176	28320	Amhara	SEMEN GONDAR	Metema	02-11-20-N	744	14-25-70-E
GEN177	241344	Amhara	SEMEN GONDAR	CHILGA		980	
GEN178	Ass-acc-36	Benshangul-Gumz	Kemashi	Kamashi			
GEN179	241326	Tigray	MIRABAWI	KAFTA HUMERA		650	
GEN180	28322	Amhara	SEMEN GONDAR	T/Armacho	02-98-97-N	946	14-61-01-E
GEN181	28303	Amhara	SEMEN GONDAR	Kora	08-22-58-N	602	13-88-91-E
GEN182	205188.2	Borkina Faso					
GEN183	X - 30/40 # 403	Israel					
GEN184	227941	EGY					
GEN185	Maydelye-03	Tigray	MIRABAWI	Tsegede	13.43625°	850	36.97694°
GEN186	17695	Oromia	MISRAK WELLEGA	GITO GIDA	08-26-17-N	1346	36-34-51-E
GEN187	ABASENA	Oromia	East Wellega	Gota Gida			
GEN188	17699	Oromia	MISRAK WELLEGA	WAYU TUQA	08-52-41-N	1755	36-41-15-E
GEN189	BENSHANGUL- 1	Oromia	west Wellega				
GEN190	28326	Amhara	SEMEN GONDAR	T/Armacho	03-06-76-N	1175	14-34-14-E
GEN191	227890	EGY					

GEN192	28309	Amhara	SEMEN GONDAR	Metema	00-20-11-N	744	14-25-70-E
GEN193	203637	ZAM					
GEN194	227945	EGY					
GEN195	205186.2	Borkina Faso					
GEN196	Ass-acc-11	Benshangul-Gumz	Kemashi	Yaso			
GEN197	9686	Tigray	MIRABAWI	ASEGEDE	13-58-98-N	1806	38-11-47-E
GEN198	234017	Tigray	MIRABAWI	TAHTAY ADIYABO		1090	
GEN199	19043	Oromia	BALE	DELO MENA	06-24-39-N	1300	39-50-31-E
GEN200	9242	Oromia	MISRAK HARERGE	GORO GUTU	09-23-53-N	1620	41-16-55-E
GEN201	x-3014≠401	Tigray	MIRABAWI		-	-	-
GEN202	17702	Oromia	MISRAK WELLEGA	GIDA AYANA	09-36-08-N	1361	36-38-49-E
GEN203	227936	EGY					
GEN204	17693	Oromia	MISRAK WELLEGA	GITO GIDA	09-21-51-N	1346	36-30-48-E
GEN205	205196.2	Borkina Faso					
GEN206	227888	EGY					
GEN207	241315	Amhara	SEMEN GONDAR	DEBARK			
GEN208	9694	Tigray	MIRABAWI	TSEGEDE (WELKAIT)	13-44-95-N	773	36-56-29-E
GEN209	ACC-BG001	Benshangul-Gumz					
GEN210	28317	Amhara	SEMEN GONDAR	Metema	01-91-68-N	872	14-52-47-E
GEN211	9696	Tigray	MIRABAWI	KAFTA HUMERA	13-33-65-N	768	36-58-58-E
GEN212	9025	Benshangul-Gumz	METEKEL	GUBA	11-09-48-N	667	35-40-37-E

GEN213	227891	EGY					
GEN214	Nigari-02	Tigray	MIRABAWI	k/humera	13.87835	707	36.62495
GEN215	227942	EGY					
GEN216	227876	EGY					
GEN217	203614	Zimbabwe					
GEN218	208888	SNNP	SEMEN OMO	KUCHA	06-25-00-N		37-05-00-E
GEN219	17698	Oromia	MISRAK WELLEGA	SASIGA	09-21-28-N	1385	36-23-36-E
GEN220	ACC-EW-006	Oromia	East Wellega				
GEN221	28325	Amhara	SEMEN GONDAR	T/Armacho	03-07-95-N	884	14-58-75-E
GEN222	28305	Amhara	SEMEN GONDAR	Metema	02-04-45-N	729	14-23-25-E
GEN223	23556	Benshangul-Gumz	METEKEL	GUBA	11-06-81-N	601	35-25-60-E
GEN224	Najjoo-77	Oromia	West Wellega	Kiltu- kara			
GEN225	USR - 82 # 171 NS	Israel					
GEN226	222876	Gambella	ZONE 1	GAMBELA			
GEN227	205197.2	Borkina Faso			10-37-00-S		05-07-00-E
GEN228	205189	Borkina Faso					
GEN229	227938	EGY					
GEN230	205186	Borkina Faso			11-18-00-S		04-14-00-E
GEN231	23546	Benshangul-Gumz	METEKEL	GUBA	11-41-06-N	599	35-24-08-E
GEN232	Lugudi-02	Tigray	MIRABAWI	k/humera	13.98959	639	36.39773
GEN233	205188	Borkina Faso			13-24-00-S		00-08-00-E
GEN234	28330	Amhara	SEMEN GONDAR	T/Armacho	03-12-45-	956	44-73-16-E

					N		
GEN235	Man-67 (mandii)	Oromia	West Wellega	Mana sibu			
GEN236	Ass-acc-64	Benshangul-Gumz	Kemashi	Kamashi			
GEN237	227875	EGY					
GEN238	28323	Amhara	SEMEN GONDAR	T/Armacho	03-02-08-N	877	14-61-43-E
GEN239	227883	EGY					
GEN240	227874	EGY					
GEN241	227879	EGY					
GEN242	28321	Amhara	SEMEN GONDAR	T/Armacho	02-97-10-N	983	14-58-88-E
GEN243	28307	Amhara	SEMEN GONDAR	Metema	02-26-01-N	793	14-07-93-E
GEN244	241340	Amhara	SEMEN GONDAR	QUARA		600	
GEN245	Maydelye-01	Tigray	MIRABAWI	Tsegede	13.46912°	864	36.96353°
GEN246	203595	Zimbabwe			20-15-00-S	1100	31-03-00-E
GEN247	Najoo-74	Oromia	West Wellega	Nejo			
GEN248	Najjoo-76	Oromia	West Wellega	Nejo			
GEN249	205185.2	Borkina Faso					
GEN250	28310	Amhara	SEMEN GONDAR	Metema	02-00-58-N	784	14-12-21-E
GEN251	205190.2	Borkina Faso			13-35-00-N		02-25-00-E
GEN252	23557	Benshangul-Gumz	METEKEL	GUBA	11-16-28-N	837	35-22-12-E
GEN253	28319	Amhara	SEMEN GONDAR	Metema	02-11-20-N	744	14-25-70-E
GEN254	28301	Amhara	SEMEN GONDAR	Kora	08-14-16-N	721	13-37-25-E
GEN255	Ass-acc-105	Benshangul-Gumz		Sambas			

GEN256	241324	Tigray	MIRABAWI	KAFTA HUMERA		700	
GEN257	207758	Africa					
GEN258	23547	Benshangul-Gumz	METEKEL	GUBA	11-07-40-N	604	05-28-02-E
GEN259	227893	EGY					
GEN260	9697	Amhara	SEMEN GONDAR	LAY ARMACHO	12-57-03-N	1081	37-19-29-E
GEN261	227939	EGY					
GEN262	203633	ZAM					
GEN263	17711	Oromia	MISRAK WELLEGA	LEQA DULCHA	08-55-31-N	1869	36-35-27-E
GEN264	241325	Tigray	MIRABAWI	KAFTA HUMERA		650	
GEN265	9689	Tigray	MIRABAWI	KAFTA HUMERA	14-03-10-N	893	37-10-51-E
GEN266	Najjoo-75	Oromia	West Wellega	Nejo			
GEN267	HUARC-4	Tigray	HUMERA				
GEN268	203099	SNNP	GURAGE	GORO		1240	
GEN269	23561	Benshangul-Gumz	METEKEL	GUBA	11-35-11-N	652	35-08-32-E
GEN270	227944	EGY					
GEN271	Ying White - 2	China					
GEN272	227880	EGY					
GEN273	BAR - 002	Somalia					
GEN274	Acc - 203 - 336 - 2	Zimbabwe					
GEN275	203640.2	Zimbabwe					
GEN276	203608	Zimbabwe					
GEN277	HB - 22 - FAM (1- 4)	EGY					
GEN278	Acc - 203 - 336 - 4	Zimbabwe					
GEN279	227884	EGY					
GEN280	215013	ZAM					

GEN281	HB - 49 FAM - 2 - 2	EGY
GEN282	Acc - 203 - 612	Zimbabwe
GEN283	HB-38FAM-2BAR GRAY	EGY
GEN284	227943	EGY
GEN285	China FAO (ACC - 68 - 542)	China
GEN286	Acc - 210 - 986 - 1	Sudan
GEN287	203603	Zimbabwe
GEN288	203615	Zimbabwe
GEN289	Clusu - Acc - 2	Philipins
GEN290	Acc - 210 - 991 - 4	Sudan
GEN291	203630	Zimbabwe
GEN292	203610	Zimbabwe
GEN293	JAPAN-651	Japan
GEN294	Tmax	Israel
GEN295	227937	EGY
GEN296	227882	EGY
GEN297	SPS - SIK - # 811	Kenya
GEN298	227885	EGY
GEN299	227881	EGY
GEN300	SSBS - (9 - 2) -3	Kenya

Appendix 2: population structure of 300 sesame genotypes with their cluster category with 50% levels of membership

GEN	cluster	Collection region	pop1	pop2	pop3	pop4
GEN283	clu1	AFRICA	0.575	0	0.166	0.259
GEN282	clu1	AFRICA	0.577	0	0.184	0.239
GEN293	clu1	ASIA	0.633	0	0.14	0.227
GEN271	clu1	ASIA	0.567	0.003	0.242	0.188
GEN183	clu1	ASIA	0.622	0	0.202	0.176
GEN300	clu1	AFRICA	0.599	0.074	0.18	0.147
GEN20	clu1	Amhara	0.749	0	0.133	0.118
GEN53	clu1	AFRICA	0.556	0	0.357	0.087
GEN273	clu1	AFRICA	0.838	0	0.082	0.08
GEN3	clu1	BG	0.695	0.17	0.056	0.079
GEN290	clu1	AFRICA	0.698	0.231	0.001	0.07
GEN285	clu1	ASIA	0.893	0.036	0	0.07
GEN2	clu1	Oromia	0.519	0.222	0.198	0.061
GEN167	clu1	IMPROVED	0.668	0.271	0.001	0.059
GEN294	clu1	ASIA	0.723	0.008	0.211	0.058
GEN9	clu1	Amhara	0.732	0.095	0.142	0.031
GEN50	clu1	Tigray	0.549	0.001	0.429	0.022
GEN297	clu1	AFRICA	0.664	0	0.318	0.017
GEN286	clu1	AFRICA	0.843	0.115	0.032	0.009
GEN280	clu1	AFRICA	0.732	0.259	0.001	0.009
GEN281	clu1	AFRICA	0.66	0	0.338	0.002
GEN201	clu1	Tigray	0.575	0.194	0.23	0.001
GEN51	clu1	Tigray	0.551	0.021	0.428	0
GEN103	clu1	IMPROVED	0.615	0	0.385	0
GEN70	clu1	Tigray	0.72	0.001	0.279	0
GEN225	clu1	ASIA	0.759	0.001	0.24	0
GEN127	clu1	Oromia	0.772	0.005	0.223	0
GEN278	clu1	AFRICA	0.82	0.001	0.179	0
GEN32	clu1	Oromia	0.592	0.253	0.154	0
GEN139	clu1	Tigray	0.574	0.276	0.15	0
GEN220	clu1	Oromia	0.559	0.31	0.131	0
GEN84	clu1	Tigray	0.879	0.011	0.109	0
GEN129	clu1	AFRICA	0.903	0.002	0.095	0
GEN209	clu1	BG	0.834	0.079	0.087	0
GEN81	clu1	Amhara	0.931	0.003	0.066	0
GEN274	clu1	AFRICA	0.743	0.218	0.038	0
GEN277	clu1	AFRICA	0.96	0.001	0.038	0
GEN73	clu1	Oromia	0.656	0.321	0.023	0
GEN276	clu1	AFRICA	0.756	0.238	0.006	0

GEN8	clu1	AFRICA	0.692	0.305	0.003	0
GEN291	clu1	AFRICA	0.675	0.323	0.001	0
GEN262	clu1	AFRICA	0.507	0.493	0	0
GEN193	clu1	AFRICA	0.531	0.469	0	0
GEN230	clu1	AFRICA	0.59	0.41	0	0
GEN195	clu1	AFRICA	0.607	0.393	0	0
GEN19	clu1	AFRICA	0.647	0.352	0	0
GEN287	clu1	AFRICA	0.688	0.312	0	0
GEN217	clu1	AFRICA	0.708	0.291	0	0
GEN246	clu1	AFRICA	0.714	0.286	0	0
GEN59	clu1	Tigray	0.748	0.252	0	0
GEN292	clu1	AFRICA	0.765	0.234	0	0
GEN289	clu1	ASIA	0.786	0.213	0	0
GEN288	clu1	AFRICA	0.791	0.209	0	0
GEN119	clu1	Amhara	0.838	0.162	0	0
GEN249	clu2	AFRICA	0.022	0.573	0.173	0.233
GEN251	clu2	AFRICA	0.056	0.714	0.002	0.229
GEN227	clu2	AFRICA	0.11	0.672	0.031	0.187
GEN275	clu2	AFRICA	0.114	0.663	0.036	0.186
GEN228	clu2	AFRICA	0	0.825	0	0.174
GEN97	clu2	AFRICA	0.004	0.68	0.191	0.125
GEN135	clu2	AFRICA	0.11	0.612	0.162	0.116
GEN182	clu2	AFRICA	0	0.891	0	0.108
GEN257	clu2	AFRICA	0	0.878	0.017	0.104
GEN107	clu2	AFRICA	0.001	0.577	0.324	0.099
GEN233	clu2	AFRICA	0.067	0.664	0.174	0.095
GEN226	clu2	BG	0.004	0.616	0.288	0.092
GEN44	clu2	Oromia	0.117	0.665	0.129	0.088
GEN222	clu2	Amhara	0.296	0.616	0.019	0.069
GEN120	clu2	Amhara	0	0.824	0.115	0.06
GEN48	clu2	BG	0.184	0.508	0.257	0.051
GEN122	clu2	Tigray	0.036	0.826	0.09	0.048
GEN67	clu2	Tigray	0.308	0.505	0.14	0.046
GEN157	clu2	IMPROVED	0.277	0.525	0.161	0.038
GEN164	clu2	Tigray	0	0.902	0.075	0.022
GEN41	clu2	Tigray	0.112	0.658	0.209	0.02
GEN10	clu2	AFRICA	0.033	0.842	0.106	0.019
GEN37	clu2	Amhara	0.001	0.876	0.106	0.018
GEN88	clu2	Tigray	0.176	0.54	0.267	0.017
GEN18	clu2	Amhara	0.058	0.854	0.072	0.016
GEN125	clu2	BG	0.007	0.928	0.049	0.015
GEN30	clu2	BG	0.183	0.722	0.081	0.014
GEN86	clu2	Oromia	0.003	0.983	0	0.013

GEN142	clu2	Amhara	0.001	0.846	0.143	0.011
GEN137	clu2	Tigray	0.177	0.633	0.181	0.01
GEN223	clu2	BG	0.142	0.766	0.083	0.01
GEN112	clu2	Amhara	0.197	0.779	0.016	0.009
GEN111	clu2	BG	0.077	0.913	0.002	0.009
GEN64	clu2	Tigray	0.001	0.772	0.219	0.008
GEN42	clu2	BG	0.001	0.994	0.001	0.005
GEN75	clu2	Oromia	0.004	0.905	0.088	0.004
GEN58	clu2	Amhara	0.101	0.879	0.015	0.004
GEN174	clu2	Tigray	0	0.647	0.351	0.002
GEN212	clu2	BG	0.273	0.609	0.116	0.002
GEN268	clu2	SNNP	0.046	0.942	0.009	0.002
GEN155	clu2	Amhara	0.046	0.952	0	0.002
GEN133	clu2	Tigray	0.001	0.617	0.381	0.001
GEN130	clu2	BG	0.001	0.643	0.356	0.001
GEN169	clu2	Amhara	0.038	0.798	0.163	0.001
GEN231	clu2	BG	0.053	0.874	0.072	0.001
GEN151	clu2	Oromia	0.07	0.886	0.044	0.001
GEN161	clu2	Oromia	0.042	0.918	0.039	0.001
GEN56	clu2	Tigray	0.002	0.972	0.025	0.001
GEN140	clu2	Oromia	0.062	0.912	0.025	0.001
GEN13	clu2	Tigray	0.001	0.979	0.02	0.001
GEN132	clu2	Oromia	0	0.999	0	0.001
GEN224	clu2	Oromia	0.009	0.99	0	0.001
GEN83	clu2	Amhara	0.005	0.53	0.465	0
GEN254	clu2	Amhara	0	0.54	0.459	0
GEN163	clu2	Amhara	0.001	0.561	0.438	0
GEN250	clu2	Amhara	0.004	0.56	0.436	0
GEN211	clu2	Tigray	0.069	0.54	0.391	0
GEN185	clu2	Tigray	0	0.63	0.369	0
GEN61	clu2	Tigray	0	0.636	0.364	0
GEN221	clu2	Amhara	0	0.643	0.357	0
GEN45	clu2	IMPROVED	0.139	0.53	0.33	0
GEN159	clu2	Amhara	0.016	0.666	0.317	0
GEN101	clu2	BG	0	0.711	0.289	0
GEN244	clu2	Amhara	0	0.717	0.282	0
GEN210	clu2	Amhara	0.001	0.718	0.281	0
GEN31	clu2	Amhara	0.001	0.73	0.269	0
GEN123	clu2	Amhara	0	0.741	0.258	0
GEN102	clu2	Amhara	0.001	0.759	0.24	0
GEN14	clu2	Amhara	0.147	0.621	0.232	0
GEN202	clu2	Oromia	0.007	0.786	0.207	0
GEN181	clu2	Amhara	0.001	0.796	0.203	0

GEN266	clu2	Oromia	0.043	0.767	0.19	0
GEN238	clu2	Amhara	0	0.813	0.186	0
GEN100	clu2	Amhara	0.001	0.829	0.171	0
GEN141	clu2	BG	0.171	0.683	0.146	0
GEN63	clu2	Tigray	0.073	0.782	0.145	0
GEN121	clu2	Oromia	0.008	0.849	0.143	0
GEN197	clu2	Tigray	0.001	0.858	0.141	0
GEN187	clu2	IMPROVED	0.09	0.772	0.139	0
GEN196	clu2	BG	0.001	0.869	0.13	0
GEN234	clu2	Amhara	0.218	0.657	0.125	0
GEN188	clu2	Oromia	0.001	0.888	0.11	0
GEN74	clu2	IMPROVED	0.387	0.516	0.097	0
GEN105	clu2	BG	0.001	0.908	0.09	0
GEN62	clu2	BG	0.159	0.751	0.09	0
GEN46	clu2	BG	0.039	0.878	0.083	0
GEN245	clu2	Tigray	0	0.918	0.081	0
GEN66	clu2	Oromia	0	0.919	0.08	0
GEN91	clu2	Oromia	0.03	0.891	0.078	0
GEN160	clu2	Tigray	0.112	0.814	0.073	0
GEN26	clu2	IMPROVED	0.095	0.834	0.072	0
GEN176	clu2	Amhara	0.001	0.928	0.071	0
GEN143	clu2	BG	0.043	0.902	0.055	0
GEN162	clu2	Amhara	0.148	0.797	0.055	0
GEN218	clu2	SNNP	0.026	0.92	0.054	0
GEN94	clu2	BG	0.02	0.928	0.051	0
GEN158	clu2	Amhara	0.026	0.923	0.051	0
GEN71	clu2	Oromia	0.17	0.779	0.051	0
GEN204	clu2	Oromia	0.016	0.934	0.05	0
GEN269	clu2	BG	0.01	0.944	0.046	0
GEN87	clu2	Amhara	0.094	0.86	0.046	0
GEN15	clu2	BG	0.003	0.952	0.045	0
GEN145	clu2	BG	0.029	0.926	0.045	0
GEN252	clu2	BG	0.085	0.873	0.041	0
GEN236	clu2	BG	0.013	0.946	0.04	0
GEN153	clu2	IMPROVED	0.036	0.923	0.04	0
GEN34	clu2	BG	0.001	0.96	0.039	0
GEN21	clu2	BG	0.001	0.962	0.037	0
GEN144	clu2	Amhara	0	0.967	0.033	0
GEN1	clu2	BG	0.001	0.966	0.033	0
GEN39	clu2	Oromia	0.006	0.961	0.033	0
GEN180	clu2	Amhara	0.225	0.743	0.032	0
GEN198	clu2	Tigray	0.018	0.952	0.029	0
GEN92	clu2	BG	0.048	0.93	0.022	0

GEN104	clu2	BG	0.077	0.907	0.017	0
GEN6	clu2	Tigray	0	0.985	0.015	0
GEN168	clu2	Oromia	0.002	0.984	0.014	0
GEN27	clu2	Oromia	0.051	0.936	0.013	0
GEN78	clu2	Amhara	0.067	0.925	0.008	0
GEN154	clu2	Amhara	0.001	0.994	0.005	0
GEN247	clu2	Oromia	0.067	0.93	0.003	0
GEN186	clu2	Oromia	0.013	0.985	0.002	0
GEN117	clu2	Oromia	0.024	0.974	0.002	0
GEN128	clu2	Oromia	0	0.999	0.001	0
GEN40	clu2	Amhara	0.016	0.983	0.001	0
GEN106	clu2	BG	0.024	0.975	0.001	0
GEN65	clu2	Oromia	0.026	0.973	0.001	0
GEN68	clu2	BG	0.057	0.942	0.001	0
GEN178	clu2	BG	0.079	0.92	0.001	0
GEN207	clu2	Amhara	0.2	0.799	0.001	0
GEN23	clu2	Oromia	0	1	0	0
GEN148	clu2	Oromia	0	1	0	0
GEN166	clu2	Oromia	0	1	0	0
GEN22	clu2	AFRICA	0.001	0.999	0	0
GEN35	clu2	Oromia	0.001	0.999	0	0
GEN47	clu2	Oromia	0.001	0.999	0	0
GEN263	clu2	Oromia	0.001	0.999	0	0
GEN11	clu2	Oromia	0	0.999	0	0
GEN25	clu2	Oromia	0	0.999	0	0
GEN76	clu2	Amhara	0	0.999	0	0
GEN156	clu2	Oromia	0	0.999	0	0
GEN175	clu2	Amhara	0	0.999	0	0
GEN190	clu2	Amhara	0	0.999	0	0
GEN219	clu2	Oromia	0	0.999	0	0
GEN248	clu2	Oromia	0	0.999	0	0
GEN253	clu2	Amhara	0	0.999	0	0
GEN124	clu2	Oromia	0.01	0.99	0	0
GEN235	clu2	Oromia	0.019	0.98	0	0
GEN82	clu2	Oromia	0.03	0.97	0	0
GEN171	clu2	BG	0.043	0.957	0	0
GEN89	clu2	Oromia	0.043	0.956	0	0
GEN136	clu2	Oromia	0.046	0.953	0	0
GEN258	clu2	BG	0.06	0.94	0	0
GEN243	clu2	Amhara	0.08	0.92	0	0
GEN55	clu2	IMPROVED	0.089	0.91	0	0
GEN255	clu2	BG	0.092	0.908	0	0
GEN52	clu2	Oromia	0.147	0.853	0	0

GEN108	clu2	BG	0.298	0.701	0	0
GEN131	clu2	Oromia	0.38	0.62	0	0
GEN96	clu3	Tigray	0.004	0.16	0.603	0.233
GEN36	clu3	Tigray	0.132	0.001	0.673	0.194
GEN205	clu3	AFRICA	0.102	0.085	0.624	0.189
GEN118	clu3	AFRICA	0.189	0.001	0.705	0.105
GEN114	clu3	AFRICA	0.105	0.307	0.5	0.088
GEN267	clu3	IMPROVED	0.137	0.001	0.812	0.05
GEN110	clu3	Amhara	0.001	0.206	0.757	0.035
GEN146	clu3	Tigray	0.094	0.185	0.694	0.027
GEN12	clu3	Tigray	0.006	0.372	0.61	0.013
GEN43	clu3	Amhara	0.003	0.192	0.804	0.001
GEN126	clu3	Tigray	0.001	0.318	0.68	0.001
GEN4	clu3	Tigray	0	0.001	0.999	0
GEN264	clu3	Tigray	0.001	0	0.999	0
GEN170	clu3	Amhara	0.01	0.001	0.99	0
GEN260	clu3	Amhara	0	0.023	0.976	0
GEN149	clu3	Tigray	0.028	0.001	0.972	0
GEN265	clu3	Tigray	0.04	0	0.96	0
GEN54	clu3	Tigray	0.045	0	0.954	0
GEN57	clu3	IMPROVED	0	0.048	0.952	0
GEN60	clu3	Tigray	0.018	0.053	0.929	0
GEN199	clu3	Oromia	0.067	0.005	0.928	0
GEN173	clu3	IMPROVED	0.002	0.072	0.926	0
GEN109	clu3	Amhara	0	0.075	0.924	0
GEN208	clu3	Tigray	0.099	0	0.901	0
GEN116	clu3	Tigray	0.108	0	0.891	0
GEN113	clu3	Tigray	0.001	0.109	0.89	0
GEN72	clu3	Amhara	0.001	0.123	0.876	0
GEN179	clu3	Tigray	0.009	0.13	0.862	0
GEN256	clu3	Tigray	0.196	0	0.804	0
GEN17	clu3	Tigray	0.126	0.1	0.774	0
GEN38	clu3	Tigray	0.25	0.001	0.749	0
GEN134	clu3	Tigray	0.001	0.254	0.745	0
GEN49	clu3	Amhara	0.001	0.262	0.737	0
GEN85	clu3	IMPROVED	0.056	0.269	0.675	0
GEN150	clu3	Tigray	0.001	0.328	0.672	0
GEN77	clu3	Tigray	0.001	0.339	0.66	0
GEN242	clu3	Amhara	0.001	0.391	0.608	0
GEN138	clu3	Tigray	0	0.399	0.6	0
GEN33	clu3	Tigray	0.126	0.276	0.598	0
GEN232	clu3	Tigray	0.092	0.315	0.594	0

GEN24	clu3	Tigray	0.174	0.286	0.54	0
GEN214	clu3	Tigray	0	0.476	0.523	0
GEN192	clu3	Amhara	0	0.496	0.503	0
GEN203	clu4	AFRICA	0.001	0	0	0.999
GEN184	clu4	AFRICA	0	0	0	0.999
GEN194	clu4	AFRICA	0	0	0	0.999
GEN206	clu4	AFRICA	0	0	0	0.999
GEN239	clu4	AFRICA	0	0	0	0.999
GEN241	clu4	AFRICA	0	0	0	0.999
GEN240	clu4	AFRICA	0	0.001	0	0.998
GEN299	clu4	AFRICA	0	0.002	0.001	0.997
GEN295	clu4	AFRICA	0.001	0.015	0	0.984
GEN296	clu4	AFRICA	0.046	0	0	0.953
GEN216	clu4	AFRICA	0	0.05	0	0.949
GEN272	clu4	AFRICA	0	0.058	0	0.941
GEN237	clu4	AFRICA	0.001	0	0.086	0.913
GEN270	clu4	AFRICA	0.102	0	0.002	0.896
GEN279	clu4	AFRICA	0.001	0.109	0	0.889
GEN259	clu4	AFRICA	0.001	0	0.145	0.855
GEN191	clu4	AFRICA	0.001	0.152	0.007	0.84
GEN213	clu4	AFRICA	0.009	0	0.173	0.818
GEN284	clu4	AFRICA	0	0.007	0.206	0.786
GEN215	clu4	AFRICA	0	0.24	0	0.759
GEN229	clu4	AFRICA	0.001	0.256	0	0.743
GEN261	clu4	AFRICA	0	0.322	0	0.678
GEN298	clu4	AFRICA	0.087	0.339	0	0.574
GEN152	mix	Tigray	0	0.5	0.5	0
GEN165	mix	BG	0.322	0.434	0.123	0.121
GEN200	mix	Oromia	0.483	0.163	0.237	0.117
GEN172	mix	Amhara	0.325	0.103	0.489	0.084
GEN90	mix	SNNP	0.441	0.299	0.183	0.077
GEN7	mix	AFRICA	0.421	0.355	0.151	0.073
GEN79	mix	Tigray	0.269	0.192	0.481	0.058
GEN93	mix	Tigray	0.372	0.426	0.163	0.04
GEN80	mix	Tigray	0.212	0.258	0.495	0.035
GEN99	mix	Oromia	0.451	0.499	0.019	0.032
GEN147	mix	Oromia	0.48	0.358	0.134	0.028

GEN28	mix	Oromia	0.393	0.429	0.151	0.027
GEN29	mix	Tigray	0.176	0.373	0.426	0.025
GEN5	mix	IMPROVED	0.405	0.283	0.293	0.018
GEN115	mix	IMPROVED	0.385	0.301	0.312	0.002
GEN16	mix	AFRICA	0.027	0.478	0.494	0
GEN95	mix	Amhara	0.174	0.335	0.491	0
GEN69	mix	Tigray	0.062	0.475	0.463	0
GEN98	mix	Tigray	0.118	0.489	0.393	0
GEN177	mix	Amhara	0.34	0.448	0.212	0
GEN189	mix	IMPROVED	0.367	0.445	0.188	0