Analysis of Genetic Diversity and Population Structure of Sesame (*Sesamum indicium* L.) Germplasm Using Inter Simple Sequence Repeats (ISSRs) Markers

Elsa Nigusie Roba

A Thesis submitted to the Department of Applied Biology,

School of Applied Natural Science

Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science in Applied Biology (Specialization in Biotechnology)

Office of Graduate Studies

Adama Science and Technology University

May, 2024
Adama, Ethiopia

Analysis of Genetic Diversity and Population Structure of Sesame (*Sesamum indicum* L.) Using Inter Simple Sequence Repeats (ISSRs) Markers

Elsa Nigusie Roba

Advisor:

Geleta Dugassa (PhD)

Co-Advisors:

Tilahun Mekonnen (PhD)

Kassahun Tesfaye (PhD)

A Thesis Submitted to the department of Applied Biology

School of Applied Natural Science

Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science in Applied Biology (Specialization in Biotechnology)

Office of Graduate Studies

Adama Science and Technology University

May, 2024

Adama, Ethiopia

Declaration

I hereby declare that this MSc. Thesis entitled “Analysis of Genetic Diversity and Population Structure of Sesame (*Sesamum indicum* L.) Germplasm Using Inter Simple Sequence Repeats (ISSRs) Markers” is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

_____	_____	_____
Name of the student	Signature	Date

Recommendation

I/we, the advisor(s) of this thesis, hereby certify that I/we have read the revised version of the thesis entitled "Analysis of Genetic Diversity and Population Structure of Sesame (*Sesamum indicum* L.) Germplasm Using Inter Simple Sequence Repeats (SSRs) Markers" prepared under my/our guidance by Elsa Nigusie submitted in partial fulfillment of the requirements for the degree of Master of Science in Applied Biology (Specialization in Biotechnology). Therefore, I/we recommend the submission of revised version of the thesis to the department following the applicable procedures.

_____	_____	_____
Major Advisor	Signature	Date
_____	_____	_____
Co-advisor	Signature	Date
_____	_____	_____
Co-advisor	Signature	Date

ACKNOWLEDGEMENT

First of all, I would like to thank the Almighty GOD for His mercy in my life. I would like to thank my research main advisor, Dr. Geleta Dugassa, for his skillful guidance, encouragement up to the accomplishment of this thesis. I would like to express my sincere thanks to my co-advisor, Dr. Kassahun Tesfaye, Director General of Ethiopian Biotechnology Institute (EBTi) for allowing me to work in his Plant Genetics Research Laboratory at College of Natural and Computational Sciences, Addis Ababa University

My special appreciation goes to my co-advisor Dr. Tilahun Mekonnen for his constructive guidance, endless and technical support during the laboratory work, interpretation of the results and review of the thesis. So many people have made meaningful contribution to achieve my goal. I am thankful to all friends who I have met along the way even if your names do not appear here.

I would like to thank Adama Science and Technology University for the support by facilitating condition for continuity of my post graduate study. I would also like to thank Addis Ababa University, Institute of Biotechnology for providing the molecular laboratory facilities to conduct this research, and I would like to thank Dr. Genet Atsebeha and Mr. Melaku Tesfaye for helping me through laboratory work and analysis software's guidance.

Last but not least, I would like to pass my heartfelt thanks to my family without their support, encouragement and unconditional love, this work would not be have been possible.

Contents

Declaration I
Recommendation II
ACKNOWLEDGEMENT III
Abstract IX
Chapter One 1
Introduction 1
1.2. Objectives 3
1.2.1. General objective: 3
1.2.2. Specific objectives: 3
Chapter Two 4
Literature Review 4
2.1. Origin and distribution of *Sesamum indicium* L. 4
2.2. Genomics of sesame 4
2.3 Taxonomy 6
2.4. Uses and nutritional composition 6
2.5. Production and productivity 7
2.5.1. Sesame production in Ethiopia 7
2.6. Production challenges 8
2.6.1 Prevailing challenges for sesame production in Ethiopia 8
2.7. Genetic diversity studies 10
2.7.1. Types of molecular markers 11
2.7.2. Breeding goals 13
Chapter Three 14
Materials and methods 14
3.1. Plant material 14
3.2. Genomic DNA extraction 14
3.3. ISSR genotyping and data scoring 14
3.4 Statistical analysis 16
Chapter Four 17
Results 17
4.1.1. ISSR primers and their banding patterns 17
4.1.2. ISSR based genetic diversity in sesame germplasms 18

4.1.3. Genetic variability within and among populations 19
4.1.4. Banding patterns across populations 20
4.1.5. Genetic relationships between the populations 21
4.1.6. Genetic relationship within and among populations 22
4.1.7. Cluster analysis using PCoA and UPGMA 23
4.2. Population genetic structure analysis 25
Chapter Five 27
Discussion 27
5.1. Polymorphism level of the ISSR markers 27
5.2. Population genetic diversity 27
5.3. Population genetic structure 28
Chapter Six 30
Conclusion and recommendation 30
6.1. Conclusion 30
6.2. Recommendation 30
Chapter Seven 31
References 31
Chapter Eight 35
Appendices 35

Lists of table

TABLE PAGE

Table 1. Sesame production by region in Ethiopia 2014/15……..…… 8

Table 2. Reproducible and polymorphic inter simple sequence repeat (ISS) R primers with their primer motif, annealing temperature (°C), amplification patterns, fragment size and score-able bands……19

Table 3. In formativeness and other genetic diversity summary statistics of inter simple sequence repeat (ISSR) loci..... 20

Table 4. Allelic patterns and diversity indices across six Sesame populations estimated from 10 inter simple sequence repeat (ISSR) markers 21

Table 5. Pairwise Nei's measures of genetic distance……...22

Table 6. Population genetic differentiation measured by PhiPT (below the diagonal) between the Six Sesame populations with p values (above the diagonal). .....

Table 7. AMOVA showing the partitioning of genetic variation within and among populations of sesame..... 24

List of Figures and Illustration

FIGURE PAGE

Figure 1 Global scenario of total oil consumption and sesame production……7

Figure 2 Banding pattern of inter simple sequence repeat (ISSR) primer UBC 864 in 18 Sesame genotypes……...17

Figure 3 Band pattern across populations…….22

Figure 4**.** Analysis of molecular variation of pie chart for 93 *Sesamum indicium* L. of six populations……23

Figure 5 Principal coordinate analysis (PCoA) of the 93 *Sesamum indicium* L. germplasms as revealed by 10 inter simple sequence repeat (ISSR) markers……..24

Figure 6 UPGMA tree showing relatedness among the 93 Sesame germplasms……25

Figure 7UPGMA dendrogram showing genetic relationships among the six populations of 93 bread sesame germplasm based on nei's unbiased genetic distances……..26

FIGURE 8 Population structure of 93 sesame germplasms representing six populations of breeding material…….27

List of Acronyms

AMOVA	=	Analysis of molecular variance
EBI	=	Ethiopian biodiversity institute
He	=	Expected heterozygosity
Ho	=	Observed heterozygosity
Na	=	Number of alleles
NJ	=	Neighbor joining
PCoA	=	Principal coordinate analysis
PIC	=	Polymorphic information content
STR	=	Short tandem repeat
ISSR	=	Inter Simple sequence repeat
UPGMA	=	Unweighted pair-group method using arithmetic averages

Abstract

Sesame is one of the most important Oily crops to sustain food security worldwide. Sesamum indicium L. takes the lead in production and economic significance among other species of Sesame. Despite many genetic diversity studies of Sesamum indicium L. using varieties of techniques in Ethiopia, there is still a need for continual researching of the species in order to meet the high demand rising with population. One way of exploring potential sesame germplasms with wide genetic diversity is engaging different breeding materials in the research. So far to the best extent of available data, genetic diversity of Sesamum indicium .L has not been done with ISSR markers in Ethiopia. Therefore, the present study was targeted to investigate the extent of genetic diversity and population structure of 93 Sesamum indicium .L using 10 ISSR markers. Modified CTAB method was used for DNA extraction. The quality and quantity of the extracted DNA was checked by agarose gel electrophoresis and Nano drop spectrophotometer, respectively. PCR was performed using 10 ISSR primers. Different software applications were employed to score and analyse the data output. The ISSR markers used were highly polymorphic, resulting in 204 alleles. The polymorphic information content of the loci ranged from 14 to 27 with an overall mean of 20. The population showed high gene diversity (0.23), expected heterozygosity (0.28), Shannon's information index (0.36) and percent of polymorphic loci (84.80%). Analysis of molecular variance (AMOVA) revealed the presence of a genetic differentiation (PhiPT:0.124) in which 88 % of the total genetic variation was accounted by the within populations variation, leaving 12% for the differentiation among populations genetic variation. Population of Amhara and Oromia showed the highest pairwise genetic differentiation and Nei's genetic distance while the other population resulted in the moderate pairwise genetic differentiation and Nei's genetic distance. Clustering did not sharply grouped the sesame *according to their source of breeding material likely due to the presence of high gene flow (Nm =3.75). STRUCTURE analysis confirmed two sub-groups with greater degree of genetic admixture. All the sesame populations showed high gene diversity indicating the relevance of the respective breeding material for studies related to sesame genetic analysis. In this study, collections showed relatively high gene diversity, indicating that these breeding materials could be potential sesame sources, and hence future studies and conservation plans shall largely include these breeding materials.*

Key words: Sesame; Genetic diversity; ISSR marker; Population structure

Chapter One

Introduction

Sesame (2n = 26) *Sesamum indicium* L. is the most important ancient oilseed crop from the genus *Sesamum*, family Pedaliaceae, and order Tubiflorae. The crop is grown on more than 9 million ha of land worldwide, mostly in tropical and subtropical regions, with annual yields of about 5.5 million t of seeds (Basak, 2019). For over 4000 years, sesame has been used for oil and food (Weiss, 1971), and it has been referred to as "the queen of the oilseeds" (Bedigian & Harlan, 1986).

Sesame seeds have an oil content of 50-60%, which is higher than that of major oil seeds such as peanut, rapeseed, soybean, and sunflower (Uzun, 2003). The oil profile is high in unsaturated fatty acids (about 80%) and low in saturated fatty acids, with stearic (C18:0) and palmitic (C16:0) acids dominating (Uzun, 2008).

Global sesame consumption is steadily increasing mainly due to changing consumer's consumption patterns and increasing health awareness. Nowadays, the consumers mostly prefer the high nutritive value products. Consequently, the demand of sesame seeds is higher since it has several nutritional characteristics such as vitamins, minerals, fiber, healthy fat, and protein. About 70% of the world's sesame seed is used to produce oil and meal (Haritha, 2016).

Total annual oil and food consumption is about 65% and 35%, respectively (Morris, 2002). Tanzania is the highest sesame seed consuming country (21% based on tonnes), followed by China (19%), Sudan (9%), Myanmar, India, Ethiopia, and Nigeria (6% each) with almost 74% of the global consumption. The consumption of Tanzania, Sudan, and Myanmar was 30.8, 17.6, and 10.1 kg per year respectively in 2016 (Myint, 2020). The world sesame production is about 5,532,000 metric tons (MT) behind soybean, groundnut, cottonseed, sunflower, linseed, and rapeseed, in the quantity of world oilseed production. It is predicted that vegetable oil consumption will be doubled, and sesame oil consumption may be 100 million MT by 2030 (Troncoso-Ponce, 2011). Therefore, the demand for genetic studies of oil-rich crops will increase to improve oil content.

Sesame is one of the oldest oilseeds and is widely cultivated in both tropical and subtropical areas (Bedigian, 1986). The crop is drought tolerant. Sesame grows well in a variety of soil

types, but prefers deep, well-drained, fertile sandy loams. Sesame grows well in Ethiopia's semiarid regions of Amhara, Tigray, Benshangul Gumuz, and Somali, Oromia and Southern Nations nationalities and peoples' regions' lowlands (Terefe, 2012).

Despite rising global demand and prices for sesame, productivity in various countries, including Ethiopia, is declining, Because of residual wild traits such as capsule shattering, sesame has a low yield when compared to other commercial oil crops (Uzune, 2003), indeterminate growth habit (Çağırgan, 2006), nonsynchronous maturity, low environmental adaptability (Ram *et al*., 2006), and susceptibly to phyllody disease (Ikten *et al*., 2014) Additional yield constraints arise as a result of agricultural issues such as the use of mixed local landraces, a low harvest index, insufficient cultural practices, poor crop rotations, and a lack of advanced breeding lines and high-yielding cultivars (Ashri & Singh, 2007). Understanding and exploiting plant genetic diversity is thus critical in sesame genetic and breeding research (Basak, 2019).

Genetic diversity and relationships among sesame landraces are critical for crop improvement and maintenance. The study of genetic diversity helps to understand the relationships among the study materials, identify genotypes candidates for selection and hybridization, select and combine alleles, exploiting heterosis, improvement of nutritional quality and adaptation to changes. Using morphological and agronomic characteristics, isozyme analysis, and molecular marker analysis, genetic diversity in various crop species can be determined (Yepuri *et al*., 2013).

However, the application of morphological and agronomic characteristics is associated with a strong influence from environmental factors and is thus dependent on cultivation conditions. Because they are not affected by the environment, molecular markers overcome this limitation. Molecular markers have also enabled detailed characterizations of germplasm in order to distinguish sesame genotypes and identify genetic diversity (Nimmakayala *et al*., 2014).

Different molecular markers, such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) (Laurentin & Karlovsky, 2006), inter-simple sequence repeats (ISSR) (Dagmawi *et al*., 2015), genomic simple sequence repeats (SSR) (Wei *et al.*, 2014) .

ISSR are among those extensively used in plant research, particularly for the evaluation of genetic variability in Sesame (Rafq, 2019). They are useful genetic tools to assess genetic variation both within and among populations, being highly polymorphic, more accessible projects, not requiring prior sequence information, and high reproducibility (Cara, 2005). However, their application to address Sesame genetic diversity in Ethiopia is limited. Scientific experiment-based knowledge about the genetic diversity of sesame (*Sesamum indicum* L.) could be used for further planning of selection and hybridization to respond to sesame yield limitations in Ethiopia.

Therefore, the present study was designed with the following general and specific objectives.

1.2. Objectives

1.2.1. General objective:

- To investigate the extent and pattern of genetic variability and population structure of (*Sesamum indicum* L.) in Ethiopia using ISSR markers.

1.2.2. Specific objectives:

1. To assess the genetic diversity of Sesame (*Sesamum indicum* L.) using ISSR markers.
2. To determine the population structure of Sesame (*Sesamum indicum* L.) Landraces of Ethiopia evaluated from ISSR markers.

Chapter Two

Literature Review

2.1. Origin and distribution of *Sesamum indicium* L.

Its origin has been to be in East Africa (Ethiopia) (Hiltebrandt ,1932) considered Africa to be the original home of sesame because it is home to a large number of wild species. The presence of weedy or wild forms of sesame *(S. alatum)*2n=26 and (*S. latifolium*, 2n=32) in Ethiopia indicates that it is indigenous and considered the centre of origin for sesame, with high genetic diversity serving as resources for crop improvement (Gebremichael & Parzies, 2011).

Considerations of a plant species' origin involve several questions, including the time and place of origin, the species from which it evolved, and its mode of evolution from the ancestral species .Agriculture is thought to have begun when early man abandoned his nomadic habits in favour of a sedentary lifestyle, resulting in a shift in his habits from food gathering to food production. This is thought to have occurred around 10,000 years ago. Though we have no idea about the chronological order of plant evolution, cereals are thought to be the first group of plants to be domesticated, with flax following closely behind. Burkill (Burkill, 1953) has proposed that the sequence was pulses, short-lived greens, oil seeds, root crops, herbaceous fruits, fibre and dry plants, woody plants, and industrial plants at the time. Coconut and sesame are thought to be the oldest oil-producing plants known to man. Sesame has even been proposed as the world's oldest cultivated oil crop (Nayar & Mehra, 1970).

Sesame was one of the plants introduced from Africa to India in this manner, according to him. Anderson has supported this thesis on entirely different grounds (Anderson, 1960), it has also been rather severely criticized by most students of the African scene (J. R. Harlan, 1967, personal communication) and especially by Baker (Baker, 1962), who argued in favour of Wingley's theory that most of the "Sudanese complex" of plants were brought into West Africa from eastern Sudan and western Ethiopia However, there is evidence for the ancient African origin of some cultivated species that have been in Indian culture since very early times. A good example is the finger millet, Eleusine corocana, for which evidence of African origin has now been presented (Mehra, 1963).

2.2. Genomics of sesame

The sesame genome size is about 350 Mb and it is largely unexplored. Dinucleotide repeat motifs are the most common (84.24%), followed by 13.53% trinucleotide,

1.65%tetranucleotide, 0.3% Penta nucleotide and 0.28% hex nucleotide motifs in the sesame genome sequence (Wei *et al*., 2014) .

Candidate genes represent loci that encode components of metabolic or signalling pathways known to be related to the corresponding phenotypes or based on expression profile. Using whole sesame genome analysis, (Supriya & Bhat, 2018) The sesame genome revealed 8244 functional genes, 58 transcription factors, and 25,069 transposable elements. Numerous loci associated with various agronomic traits such as plant height (Xia *et al*., 2013) ,disease resistance (YanXin *et al*., (2012) ,drought toleranceLi *et al*., (2013), waterlogging tolerance (Y Zhang *et al*., 2014). seed coat colour (Karthik *et al*., 2014) and oil and protein content (Karthik *et al*., 2014) had been identified in sesame land races (cvs. Baizhima and Mishuozhima).

By genome comparison with Zhongzhi 13, next generation sequencing (NGS) revealed a total of 1,332,025 SNPs (single nucleotide polymorphisms) and 506,245 indels (insertion-deletions) (white seeded). Indels are more polymorphic than SSRs, but in terms of deciphering genetic diversity, they are comparable (Wu *et al*. 2014). The seed colour ranges from black to intermediate to white. Black seeded varieties typically contain less oil but more protein and lignin (H. Zhang *et al*., 2013) than black-coated seeds. The polyphenol oxidase (PPO) gene produces black pigments is the key regulatory gene of sesame seed coat colour (Wei *et al*., 2017) .

Sesame was domesticated 5000 years ago, then dispersed to very different environments along trade routes, where it was subjected to natural selection to adapt to new conditions (Wood, 2003). Some variations (indels) may have resulted from natural selection during the domestication of land races. Rare SNPs and indels can be used to aid genetic improvement and delineate the status of variation across genomes (Ohmido, 2000). The regions saturated with high SNPs and indels contain genes with elevated localized mutation rates or recombination hot spots(Lercher & Hurst, 2002). High stringency filtering resulted in the identification of 420 SNPs distributed among 13 linkage groups. These markers will be useful to tag QTLs of metabolic and agronomic traits (Uncu, 2016).

2.3 Taxonomy

The detailed taxonomic hierarchy of the present-day cultivated sesame (Sesamum indicum L.) is as follows:

Kingdom: Plantae – Plant kingdom

Sub-kingdom: Viridiplantae

Infrakingdom: Streptophyta

Superdivision: Embryophyta

Division: Tracheophyta

Sub-division: Spermatophytina

Class: Magnoliopsida

Superorder: Asteranae

Order: Lamiales

Family: Pedaliaceae

Genus: Sesamum (Tripathy, 2019)

2.4. Uses and Nutritional Composition

Sesame seeds and oil have traditionally been used as key ingredients in food products, salad preparation, pharmaceuticals, and the production of margarine, confectioneries, cookies, cake, breads, cosmetics (skin softener and in massage), antibacterial mouthwash, and perfume (Bedigian, 2003). At mealtime, sesame seeds are mixed with cereals, rice, noodles, and other dishes as a condiment. The seeds are frequently combined with warm jiggery and sugar to make sweet balls that are consumed as a snack. Locally, low-grade oil is used in soaps, paints, varnishes, insecticides, and lubricants, and as source of illuminants and biodiesel (Ahmad, 2009). After oil extraction, the meal contains 30-35% protein, making it an excellent feed for poultry and livestock. Sesame oil is used for both dietary (edible oil) and therapeutic purposes. Sesame oil is also used as a laxative, a solvent for intramuscular injections, and an Ayurvedic

Medicine ingredient. Sesame oil is also used as a coating on stored grains to keep weevils at bay.

Sesame seeds are rich in oil and protein, and have high dietary energy value 5730 kcal/kg. Chemical composition of seed reveals that it contains 49.7% oil (fat), 17.7% protein and 23.4% carbohydrate .Oleic acid (43%), linoleic acid (35%), palmitic acid (11%) and stearic

acid (7%) are the most abundant fatty acids present in sesame oil, which together contribute 96% of total fatty acid content (Elleuch, 2007) 2. Iron, magnesium, manganese, copper, and calcium are abundant in seeds, as are vitamins B1 (thiamine) and E. (tocopherol). Because of its delicate flavour and stability, as well as its high quality cooking value, sesame fat is preferred in the food industry. Sesamin and sesaminol lignans are found in the nonglycerol fraction of sesame oil and contribute to its oxidative stability and antioxidative properties (W.-H. Wu, 2007).

2.5. Production and productivity

The majority of sesame cultivation (90%) occurs in Asia and Africa. In 2016, global sesame production was estimated to be 6.1 million mt, with Tanzania, Myanmar, India, China, and Sudan leading the way (Tripathy *et al*., 2019)(Fig. 2). India ranked third in sesame production (866 million mt) with the largest area (1947 million ha) among different countries in 2015–2016(Tripathy *et al*., 2019). Despite its unique position, productivity in India is extremely low (413 kg/ha), below the world average yield (535 kg/ha), and about one-third the productivity China1234kg/ha.

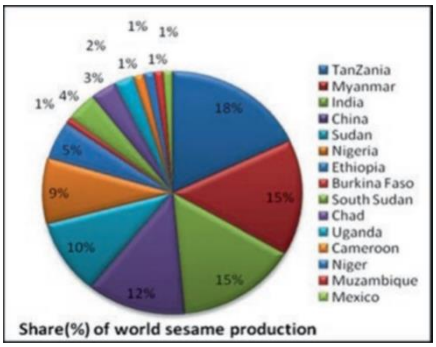


Fig. 1 Global scenario of total oil consumption and sesame production (Tripathy *et al*., 2019)

2.5.1 Sesame production in Ethiopia

Oil seeds are an important part of Ethiopia's rural and national economies. The oilseeds sector is one of the fastest growing; it is the second largest export and incoming earning sector in Ethiopia, employing over three million small, medium, and large-scale Ethiopians (Gelalcha, 2009)). Sesame production in the world is estimated to be 2.25 million tons, with 70% consumed in the producing countries, and an annual trading volume of 600,000 mt valued at US$500 million (Haile, 2009). Ethiopia is a major sesame growing and exporting country in Africa, exporting massive quantities to the global market. It is a major crop in

Tigray, Amhara, and Somalia, as well as in some areas of Southern nations, nationalities, and peoples (SNNP) (Girmay, 2018).

Table 1. Sesame Production by region in Ethiopia 2014/15 (Girmay, 2018)

Region	Area (Ha)	Production (ton)	Productivity (ton/ha)
Tigray	120,855.45	850,45.133	0.704
Amhara	169,988.58	1,122,09.218	0.66
Oromia	82,018.04	602,76.539	0.735
BenshanulGumuz	40,722.60	278,09.345	0.68
SNNP	6,365.70	31,65.097	0.497

2.6. Production challenges

2.6.1 Prevailing challenges for sesame production in Ethiopia

Drought Sesame production takes place in rain-fed conditions. However, sesame production and productivity are hampered by climatic variability. Prolonged dry spells leading to drought are common in Ethiopia's dry land areas. Soil moisture stress affects sesame crop establishment, particularly during the early growth stages, followed by the mid to late growth stages, resulting in significant yield loss. As a result, increasing productivity and production in Ethiopia is a major challenge(Girmay, 2018).

2.6.1.1. Weeds: Broad leaf weeds and grass weeds from the Gramineae family have an impact on sesame growth and development in Ethiopia by competing for growth resources such as light, moisture, and nutrients in the soil. Weeds from 31 families have been identified in the sesame growing areas of Ethiopia, with grasses, sedges, bind weeds, and broadleaf weeds predominating in various regions throughout the country (Terefe, 2012). Sesame seedlings are sensitive to weed competition during the early growth stages; so sesame seedlings have to be hand weeded at least two times at 10-15 and 30-45 days after emergence. Tillage, pre and post-emergence herbicide application, and growing conditions can all affect sesame yield loss. as stated by (Terefe, 2012) yield loss due to weed infestation in Metekel, Humera, and Werer under irrigation recorded 42, 83 and 92% respectively.

2.6.1.2. Insect pests: Webworm (Antigstra catalaunalis), seed bug (Elasmolomus sordidus), gall midge (Asphondilia sesami), green vegetable bug (Nezara viridula), grasshoppers, African bollworm (Helicoverpa armigera) and crickets are the major pests of sesame in Ethiopia (Terefe, 2012). According to (Ali, 2015) sesame seed bug is very damaging

postharvest pest sucking the oil content which reduce the oil content. Insect pests such as termites, aphids, jassides, whitefly, moths (warehouses) and red flour beetle affect sesame in Ethiopia(Girmay, 2018).

2.6.1.3. Diseases: In Ethiopia, several diseases affect the sesame crop. The major diseases of sesame in Ethiopia are bacterial blight (Xanthomonase sesami), leaf spot (Pseudomonas sesami), phyllody (Phytoplasma transmitted by vector) causing deformation of leaves and flowers, and wilting (Fusarium oxysporium f. sesami). Aside from weeds, insect pests, and diseases, the following factors are regarded as obstacles to sesame production and low productivity in Ethiopia(Girmay, 2018).

2.6.1.4. Drought: The unpredictable onset and cessation of rainfall in Ethiopia has an impact on sesame production, productivity, and quality. Despite the fact that sesame is a drought-tolerant crop, prolonged dry spells during its early growth stages have an impact on growth and development. Rainfall cessation during the mid-growth stages (flower initiation and grain filling) discourages pollination and seed set. El Nino effects in 2015(Gitima & Mersha, 2020) affect sesame production in Ethiopia, with complete crop loss in Humora vicinity.

2.6.1.5. Low productivity of sesame varieties: Sesame is naturally a low yielder, but productivity of sesame varieties in Ethiopia is very low when compared to other countries'(Girmay, 2018).

2.6.1.6. Market fluctuation (domestic and world market)

2.6.1.7. Traditional production system: Farmers' and investors' poor management practices Farmers use traditional plowing systems, such as oxen, donkeys, or camels, which are labour intensive and result in high production costs, shallow till, and difficulty with row plantation and fertilizer application. Sesame research receives less attention than other crops such as maize and wheat, despite the fact that it is a major export commodity second only to coffee (Girmay, 2018).

2.6.1.8. Lack of improved technologies (planting, harvester): The vast majority of sesame growers are farmers who cannot afford modern planting, harvesting, and threshing equipment(Girmay, 2018).

2.6.1.9. Lack of improved facility, Poor fertilizer response of sesame crop

2.6.1.10. Shattering: sesame seed capsules when the seeds reach maturity and harvesting is delayed, they crack and shed. A significant amount of sesame yield is lost due to shattering, even when harvested and bundled locally as 'Hilla.' A good remedy is to collect harvest on a smooth floor or on plastic sheets (Girmay, 2018).

2.6.1.11. Smallholder farming: Different land holdings in Ethiopia are responsible for sesame production in different areas. Big investors own hundreds of hectares, whereas small scale farmers own less than ten hectares, with pieces of land in different locations incurring extra production costs and uneven crop management in some areas. Small-scale farming combined with a backward production system resulted in low sesame production productivity. Sesame productivity is less than 10Qt/ha in most farmer-managed areas. Investors use an extensive production system rather than intensive production, which produces poor results regardless of field size. as stated by(Girmay, 2018) Traditional tillage practices used in small-scale farming in Ethiopia resulted in low sesame production and productivity. In general, the challenges for sesame production are as follows (USAID, 2012);

2.6.1.12. Inability to supply and/or finance inputs to sesame out-growers

2.6.1.13. Smallholder farmers' and cooperatives' low capacity and productivity

2.6.1.14. Environmental factors: Drought and excessive rainfall, which cause water logging, are major environmental factors causing crop failure and/or productivity. For instance, hulling companies. and Inflationary pressures are high (Girmay, 2018).

2.7. Genetic diversity studies

In any crop breeding program, adequate genetic variability is required for selection to be effective in producing the desired genotype. The genetic variability in available Sesamum germplasm is generally limited. Domestication of sesame has a long history. Some plant types have adapted to survival through natural selection as a result of continuous cultivation in rain-fed marginal lands. As a result, most of the valuable genes associated with high yield have been eroded during sesame evolution. Furthermore, modern plant breeding has presumably narrowed the genetic basis of cultivated sesame by limiting the use of land races (as parents). As a result, broad-based genetic resources: Wild distant relatives and closely related species (a), local land races (b), obsolete breeding lines including mutants (c), and newly developed improved varieties (d). Knowledge of genetic variation, inheritance patterns, and the interrelationship of plant traits is required prior to using germplasm effectively in any genetic improvement program (Ganesh & Thangavelu, 1995).

Characterization of cultivars for genetic diversity using molecular markers, on the other hand, is extremely valuable for assisting phylogenetic relationships, parental line selection, population structure, and allele distribution among germplasm lines, as well as designing breeding strategies. At the moment, molecular techniques such as isozymes are being used (Isshiki & Umezaki, 1997) and amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs), inter simple sequence repeats (ISSRs), and random amplified polymorphic DNA (RAPD) profiles, which are all widely used to study genetic diversity in sesame (Abdellatef, 2008);(Ercan, 2004);(Kim *et al*., 2002);(Kumar, 2012); (Pham *et al.*, 2009).

Scientists and plant breeders use morphological and molecular markers to aid in screening and selection (Pham *et al*., 2011) of germplasm and their use in breeding program. However, little progress has been made in this regard in sesame (Venkataramana Bhat, 1999).Development of genetic maps (Dossa, 2017) and availability of few closely-linked molecular markers (using the AFLP technique) have made it possible for reliable screening of sesame germplasm. (Uzun, 2003) detected closely-linked molecular marker for the closed-capsule mutant trait and it helped to introgress shattering resistance in sesame.

2.7.1. Types of molecular Markers

DNA/molecular markers are one of the two classifications of genetic markers along with classical markers which include types like morphological, cytological and biochemical markers. These genetic markers are a gene with a known chromosol location which control specific gene or trait (Nadeem, 2018).

Several PCR based molecular markers namely random amplified polymorphic DNA(RAPD), amplified fragment length polymorphism(AFLP), sequence-characterized amplified region, inter simple sequence repeats(ISSR), simple sequence repeats(SSR) and single-nucleotide polymorphism(SNP), etc. have been developed and can be successfully used for genetic diversity studies in plants (Nadeema, 2018). The assessment of genetic variation among Sesame germplasm using different molecular markers has the capacity to show the presence of high genetic diversity (Mondini, 2009).

Evaluation of genetic diversity and determination of inter- and intra-species genetic relationships are possible through the application of molecular marker to recognize polymorphic nucleotide sequences dispersed throughout the genome (Khaled *et al*., 2015).

Molecular markers could be different from one to other types as they exhibit difference in technical principle they have, type of inheritance, reproducibility extent, amount of polymorphism and in their expense (Khaled, 2015).

Data generation efficiency and genome area coverage in a study are some of factors which vary among different molecular marker techniques (Darvishzadeh, 2010).The target crop and the issue are stand points for selecting a type of marker tool to be used for study. For instance, random amplified polymorphic DNA (RAPD) markers are known for their simplicity, cost efficiency, fast polymorphism assessment, no prior information of DNA sequences being required and extensive coverage of the intact genome being possible. Molecular markers with low reproducibility will not be selected to be used and other better markers will be prioritized (Madhumati, 2014).

The amplified fragment length polymorphism (AFLP) marker system, SSR markers which require the DNA sequence's prior information for its analysis and ISSR, dominant marker system are chosen over RAPD. Inter-simple sequence repeat markers (ISSR) primers can produce more reproducible and reliable band patterns due to high annealing temperature and extended sequence in comparison to RAPD markers (Yuan, 2014).

Inter-simple sequence repeat markers are comparable to the SSR system in reproducibility and used for distinguishing DNA on the basis of single base variation or insertions and deletions {Wang, 2002 #99}. The major limitations with some of these methods is low reproducibility of RAPD, high cost of AFLP and need to know the flanking sequences to design specific primers for SSR markers. ISSR markers overcome most of these limitations (Madhumati, 2014).

2.7.1.1. Inter Simple Sequence Repeats (ISSR)

Inter simple sequence repeat (ISSRs) are DNA fragments of about 100-3000 bp positioned between two identical, adjacent, oppositely oriented microsatellite regions (Nadeem, 2018). Amplification of ISSRs by PCR is done using microsatellite which is Variable Tandem Repeat bases are ranging from 2-5 as primers. These primers could be either unanchored or anchored at 3' or 5'with 1-4 degenerate bases extending to flanking sequences. However primers anchored at 3' are preferred as the specifically anneal to DNA sequence for amplification generating clear bands wit out smear (Pradeep Reddy, 2002).

Primers used in ISSR PCR based method are longer being 16-25 which allows higher annealing temperature for highest specificity (Pradeep Reddy, 2002). About 10-60 fragments

from multiple loci are generated simultaneously, separated by gel electrophoresis and scored as the presence or absence of fragments of particular size (Nadeem, 2018).

ISSR PCR base technique is known for its reproducibility which is possibly due to longer primer is used compared to RAPD. Most of the benefits of AFLP and SSR are combined along the universality of RAPD. SSR shows considerable polymorphism in its sequence as it has higher evolutionary change within them. There are several mechanisms for the creation of variation in SSR sequences including slippage of Polymerase during DNA replication, its failure to correct mismatches, nature and sequence (motif) of in the primer {Pradeep Reddy, 2002).

Application of ISSR PCR in plant improvement is wide but mainly includes genomic finger printing, genetic diversity and phylogenetic analysis, genome mapping, gene tagging, marker assisted marker, SSR motif frequency and natural population studies (Wang, 2002).

ISSR markers is advantageous due to there is no prior knowledge of sequence is required for designing primer, it is randomly distributed throughout the genome, produce highly reproducible, environmentally stable, less DNA is adequate to conduct PCR reaction and technically easy to work with. Unfortunately it is dominant marker, low reproducible and locus nonspecific (Nadeem, 2018).

2.7.2. Breeding Goals

The primary goals of sesame breeding are improved seed retention in the capsule, increased oil content, uniform maturity, and disease resistance. For greater clarity, the following breeding objectives should be highlighted: (a) Determinate plant type, higher number of leaf axils for more capsule bearing/plant, high yield, and performance stability; (b) High oil content, improved fatty acid composition, and other quality traits; (c) Low or zero anti-nutritional factors (oxalic and phytic acids) for value addition;(d) CMS line development for ease of hybrid seed production; (e) Resistance to biotic stress: insect pests (leaf eating caterpillar, gall fly), and diseases such as phyllody (virus, mycoplasma), bacterial leaf spot (Pseudomonas sesami), powdery mildew, wilt, and leaf curl; (f) Semi-indehiscence of capsules; (g) Resistance to abiotic stresses: drought, waterlogging, and Applications in crop improvement, conservation, and so on. Status (morphology, molecular, both locally and globally) -Include the majority of diversity parameters (Singh & Jauhar, 2006).

Chapter Three

Materials and methods

3.1. Plant material

In this research, a total of 93 *Sesamum indicium* L. germplasms obtained from EBI were used. From93 genotypes, 17 were from Tigray, 6 from SNNP, 26 from the Oromia, 34 from the Amhara, 13 from Benishan Gumuz (B. Gumuz), and the remaining 3 genotypes were from Gambella. Accordingly, the research materials were grouped into 6 populations (Tigray, SNNP, Oromia, Amhara, B.Gumuz, and Gambella) based on their origin.

3.2. Genomic DNA extraction

For genotyping, high molecular weight genomic DNA (gDNA) was extracted from 3-week-old silica gel dried leaf samples pooled from five plants per using a modified cetyltrimethyl ammonium bromide (CTAB) procedure (Contains 2 % CTAB, 20 mM EDTA·Na_2·$2H_2O$, 1.4 M NaCl and 100 mM Tris, pH 8) (Ogbonnaya, 2001: Reitsma, 2017). The quantity and quality of the extracted DNA were checked using Nano drop spectrophotometry and electrophoresis on a 1% agarose gel, respectively.

3.3. ISSR genotyping and data scoring

Thirty ISSR primers ranging from 16 to 25 base pairs (bp) in length and representing di, tri, and tetramer repetitions were tested for polymorphism and reproducibility. Out of the 30 primers examined, 10 reproducible ones with distinct banding patterns were used. These were dissolved in 10 pmol/l concentration of sterilized, double-distilled water. For optimum amplification, several amounts of template DNA, dNTPs, and Polymerase chain (PCR) was carried out in a total volume of 20 µL containing 2.5 µL of 10×PCR buffer, 2.5 µL MgCl2 (25 mM), 0.5 µL dNTP mixture (10 mM), 0.5 µL primer (15 pmol/µL), 0.3 µL Taq DNA polymerase (1 U/µL), 2 µL of template DNA (80 ng/µL) and 11.7µL double-distilled water. Optimization of the PCR condition (annealing temperature) was performed on a Biometra 2003 T3 thermocycler (USA). The PCR amplification program involved an initial denaturation at 94 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 1 min, optimized primer annealing conditions (48–51 °C for 1 min), and primer extension at 72 °C for 2 min, followed by a final extension step of 10 min at 72 °C. Amplification products were fractionated by loading 10 µl amplification product of each sample with 2 µl 6X loading dye on 1.67% agarose gel electrophoresis supplemented with ethidium bromide in 1×TBE buffer

at 70 V for 2 h. The gel was visualized under UV light and subsequently photographed using a BIO-RAD Gel Doc TM EZ Imaging System. A 100-base-pair DNA ladder was used to estimate the amplification size.

Y = (C,T); R = (A, G); H = (A, C, T); B = (C, G, T); V = (A, C, G); D = (A, G, T) b

Annealing temperature in °C (AT °C), amplification product (AP), The italicized primers were the 10 most informative for *Sesamum indicum* .L

Primer	Annealing temperature (°C)	Amplification pattern	Fragment size (bp)	Scora-ble bands
UBC 809	51	Reproducible and polymorphic	250-1000	14
UBC 811	51	Reproducible and polymorphic	200-800	17
UBC 815	51	Reproducible and polymorphic	150-1500	16
UBC 825	50	Reproducible and polymorphic	150-1400	17
UBC 834	52	Reproducible and polymorphic	200-1000	20
UBC 854	51	Reproducible and polymorphic	350-1500	24
UBC 864	51	Reproducible and polymorphic	200-1900	27
UBC 872	50	Reproducible and polymorphic	200-3000	22
UBC 873	48	Reproducible and polymorphic	150-1400	29
UBC 880	51	Reproducible and polymorphic	200-1500	19

For scoring purpose, an ISSR marker was used as dominant marker, and hence, each locus was considered as a bi-allelic locus with one amplifiable and one null allele. Scoring was performed manually for each primer based on presence (1) and absence (0), and each band was regarded as a locus. The band scoring resulted in a “0” and “1” data matrix that was used for further analysis.

3.4 Statistical analysis

Different statistical software packages were used to research genetic diversity and population structure of the bread wheat germplasms. Locus based diversity parameters including number of observed number of alleles (Na), effective number of alleles (Ne), Shannon's information statistic (I), Nei's genetic diversity (h), total genetic diversity (HT), genetic differentiation (HS), genetic differentiation statistics by locus (Gst), estimate of the number of migrants (gene flow) (Nm) and percent polymorphic loci (PPL) were computed using POPGENE version 1.31 software (Yeh et al. 1999). Within population genetic diversity parameters such as number of observed alleles (Na), Effective number of alleles (Ne), Shannon's Information index (I), Nei's genetic diversity (h) and Percent of polymorphic Loci (PPL) across loci were computed using GenAlEx ver. 6.501 software(Nei, 1983 : Peakall, 2006). The polymorphism information content (PIC) was computed using the formula: PIC=2f (1−f) (Botstein, 1980) where f is the frequency of occurrence of polymorphic bands in different primers.

The same software was used to compute the analysis of molecular variance (AMOVA) and the estimation of the variance components. Clustering was carried out based on the unweighted pair group method with arithmetic mean (UPGMA) using molecular evolutionary genetics analysis (MEGA) software (Pravin, 2018). Bayesian model-based clustering with presumed K populations was employed for 93 sesame germplasm using STRUCTURE, version 2.3.4 (Pritchard, 2000; Hubisz, 2009). The run parameters were 80,000 iterations of burn-in with 80,000 Monte Carlo Markov Chain (MCMC) iterations. Using web-based STRUCTURE HARVESTER (Earl and VonHoldt 2011), (http://taylor0.biology.ucla.edu K values from 1 to 10 were used with 10 independent runs for each K. A bar plot for the optimum K was determined using the Clumpak beta version (Kopelman, 2015).

Chapter Four

Results

4.1.1. ISSR primers and their banding patterns

The 10 primers used revealed polymorphism ranging from 150 to 2000 bp (the 100 bp DNA Ladder consists of 13 double standard DNA fragments that range from 100-3000 bp in size) in the amplified DNA fragments. Figure 2 presents a sample of score-able polymorphic bands that were generated by primer UBC 864. Among the tested ISSR markers, 10 primers that are relatively highly polymorphic, reproducible with clear bands and an unambiguous amplification, generated a total of 204 bands (Table 2). The number of score-able bands for individual primers ranged from 14 to 29, with an average number of bands of 19.1 per primer. The highest number of bands was generated by primer UBC 873 with 29 bands, followed by primer UBC 864 with 27 bands. Whereas, the lowest amplification was recorded for primer 809, which generated 14 bands (Table 2).

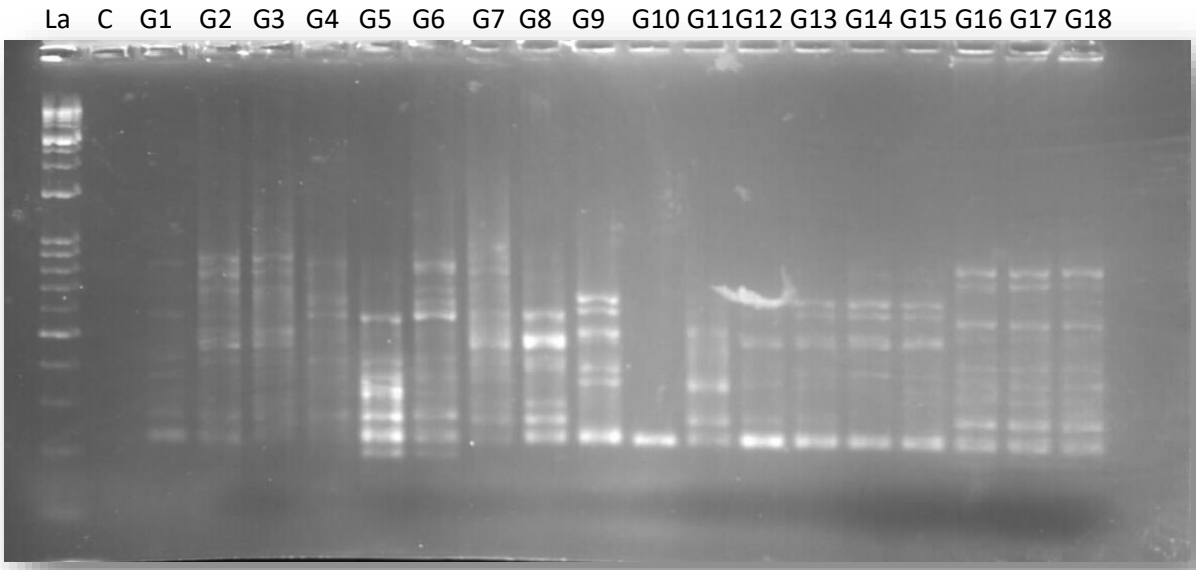


Fig. 2 Banding pattern of inter simple sequence repeat (ISSR) primer UBC 864 in 18 Sesame genotypes: La: DNA ladder, G1: genotype 1, G2: genotype 2, G3: genotype 3, G4: genotype 4, G5: genotype 5, G6: genotype 6, G7: genotype 7, G8: genotype 8, G9: genotype 9, G10: genotype 10, G11: genotype 11, G12: genotype 12, G13: genotype 13, G14: genotype 14, G15: genotype 15, G16: genotype 16, G17: genotype 17, G18: genotype 18.

Table 2 Reproducible and polymorphic inter simple sequence repeat (ISS) R primers with their primer motif, annealing temperature (°C), amplification patterns, fragment size and score-able bands

Primer	Annealing temperature (°C)	Amplification pattern	Fragment size (bp)	Scora-ble bands
UBC 809	51	Reproducible and polymorphic	250-1000	14
UBC 811	51	Reproducible and polymorphic	200-800	17
UBC 815	51	Reproducible and polymorphic	150-1500	16
UBC 825	50	Reproducible and polymorphic	150-1400	17
UBC 834	52	Reproducible and polymorphic	200-1000	20
UBC 854	51	Reproducible and polymorphic	350-1500	24
UBC 864	51	Reproducible and polymorphic	200-1900	27
UBC 872	50	Reproducible and polymorphic	200-3000	22
UBC 873	48	Reproducible and polymorphic	150-1400	29
UBC 880	51	Reproducible and polymorphic	200-1500	19

4.1.2. ISSR based Genetic diversity in sesame germplasms

The analysis revealed that the ISSR markers used were highly polymorphic and informative with number of alleles range from 1.00–1.60 with an average of average 1.35, effective number of alleles range from 1.16–1.42 with an average value of 1.30, Shannon's Information index range from 0.19–0.36 with an average value of 0.28, Nei's gene diversity range from 0.12–0.24 with an average value of 1.95, and gene flow range from 2.55–4.99 with an average value of 3.75 (Table 3). The highest number of allele (Na=1.60), effective number of allele (Ne=1.42), gene diversity [(h=0.24 and Shannon's information index (I=0.36)] were obtained for primer 834, while primer UBC-854 resulted in least values for the corresponding genetic diversity indices (Table 3).The locus polymorphic information content ranged from 0.24 (UBC-854) to 0.38 (UBC-834) with an overall mean value of 0.32 (Table 3). Markers with a PIC value between 0.25 and 0.5 were considered moderately informative, and less than 0.25 were considered less informative. Accordingly, 90% of the used ISSR markers showed moderate PIC. Moreover, 70% of the primers (7) demonstrated PIC values greater than and equal the average value (0.32).

Table 3 In formativeness and other genetic diversity summary statistics of inter simple sequence repeat (ISSR) loci

Primers	Amplicon size range(bp)	Na	Ne	I	H	Ht	Hs	Gst	Nm	PIC
UBC 809	250-1000	1.39	1.37	0.36	0.23	0.31	0.24	0.18	2.92	0.32
UBC 811	200-800	1.26	1.30	0.28	0.18	0.30	0.20	0.15	3.22	0.30
UBC 815	150-1500	1.57	1.30	0.33	0.20	0.34	0.29	0.15	4.59	0.34
UBC 825	150-1400	1.45	1.30	0.31	0.19	0.33	0.28	0.16	2.94	0.36
UBC 834	200-1000	1.60	1.42	0.36	0.24	0.34	0.31	0.14	4.99	0.38
UBC 854	350-1500	1.00	1.16	0.19	0.12	0.20	0.15	0.26	2.98	0.24
UBC 864	200-1900	1.30	1.31	0.29	0.19	0.32	0.23	0.22	4.56	0.32
UBC 872	200-3000	1.46	1.34	0.32	0.21	0.33	0.27	0.17	4.15	0.34
UBC 873	150-1400	1.04	1.18	0.19	0.19	0.22	0.16	0.24	4.61	0.26
UBC 880	200-1500	1.46	1.33	0.20	0.20	0.36	0.26	0.24	2.55	0.37
Maximum	200-1000	1.60	1.42	0.36	0.24	0.34	0.31	0.14	4.99	0.38
Minimum	350-1500	1.00	1.16	0.19	0.12	0.20	0.15	0.26	2.98	0.24
Mean (average)		1.35	1.30	0.28	1.95	0.30	0.23	0.24	3.75	0.32

Na: Observed number of alleles; Ne: Effective number of alleles; I: Shannon's Information index; h: Nei's gene diversity; HT: Total genetic diversity HS: genetic differentiation; Gst: Genetic differentiation statistics by locus; Nm: estimate of the number of migrants (gene flow) from Gst where Nm=0.5(1−Gst)/Gst, PIC: polymorphic information content

4.1.3. Genetic variability within and among populations

The average within population genetic diversity estimates for the varying population sizes ranging from 3 to 28 were: observed number of alleles 1.32, effective alleles of 1.29, Shannon's information index of 0.28, Nei's gene diversity value of 0.18, unbiased expected heterozygosity of 0.19 and percentage of polymorphic loci of 64.05% (Table 4). The Amhara population showed the highest observed number of alleles, effective number of alleles, Nei's gene diversity, Shannon's information index, and PPL while Oromia and Tigray populations ranked second and third, respectively. The highest and the lowest Nei's gene diversity were observed in the populations of Amhara and Gambella, respectively. Sixty seven percent of the research populations displayed a genetic diversity greater than mean value (h=0.19) and the PPL per population was between 13.73% (Gambella) and 84.80% (Amhara) with an

overall mean of 64.05%. A summary of the different genetic diversity estimates over all loci across populations is presented in Table 4.

Table 4 Allelic patterns and diversity indices across six Sesame populations estimated from 10 inter simple sequence repeat (ISSR) markers

Pop	N	Na	Ne	I	H	UHe	PPL (%)
Tigray	17	1.66	1.34	0.35	0.22	0.22	82.35
Oromia	26	1.69	1.34	0.35	0.23	0.23	84.31
Amhara	28	1.70	1.35	0.36	0.23	0.28	84.80
B.Gumuz	13	1.40	1.34	0.32	0.21	0.22	70.10
Snnp	6	1.03	1.30	0.26	0.17	0.19	49.02
Gambella	3	0.47	1.08	0.08	0.05	0.06	13.73
Maximum	28	1.70	1.35	0.36	0.23	0.28	84.80
Minimum	3	0.47	1.08	0.08	0.05	0.06	13.73
Mean(average)	15	1.32	1.29	0.29	0.19	0.19	64.05

N: sample size, Na: Observed number of alleles, p & q: Estimated Allele Frequency, Ne: Effective number of alleles, h: Nei's gene diversity, I: Shannon's Information index uHe: Expected and Unbiased Expected Heterozygosity and PPL: percentage of polymorphic loci. Na: No. of Different Alleles, Ne: No. of Effective Alleles: 1/(p^2+q^2), I: Shannon's Information Index: −1* (p * Ln (p)+q * Ln(q)), He=Expected Heterozygosity: 2 * p * q, uHe: Unbiased Expected Heterozygosity: (2N/(2N−1)) * He Where for Diploid Binary data and assuming Hardy–Weinberg Equilibrium, q: (1−Band Freq.) ^0.5 and p: 1−q Population abbreviation is: B.GUMUZ, Benishangul-Gumuz

4.1.4. Banding patterns across populations

In the banding pattern analysis of 93 germplasms that were grouped into six populations, Amhara and Oromia represented the highest number of bands (173), followed by Tigray with 170 while Gambella comprises the lowest number of bands (68). The highest mean heterogeneity (0.225) was observed in Tigray, followed by Oromia and Amhara with mean heterogeneities of 0.223 and 0.217, respectively. Private bands used for distinguishing populations from each other were not identified; all populations had no private band (Fig. 3)

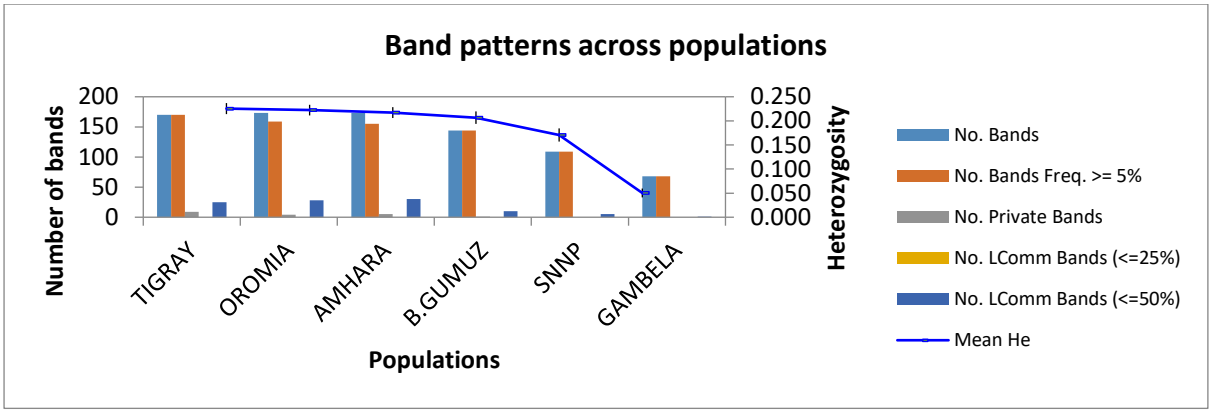


Fig. 3 Band pattern across populations, (Population abbreviation is: B.Gumuz Benishangul-Gumuz)

4.1.5. Genetic relationships between the populations

Estimates using genetic distance in Fig. 5 showed the highest genetic distance of the genetic distance between Tigray and Oromia is 0.037, which is statistically significant, suggesting moderate genetic differentiation between these two populations. The genetic distance between Tigray and Gambella is 0.158, also statistically significant, indicating substantial genetic differentiation between these populations (Table 5).

Table 5 Pairwise Nei's measures of genetic distance

	TIGRAY	OROMIA	AMHARA	B. GUMUZ	SNNP	GAMBELA
TIGRAY	0.000	***				
OROMIA	0.037	0.000	***			
AMHARA	0.036	0.023	0.000	***		
B. GUMUZ	0.047	0.049	0.032	0.000	***	
SNNP	0.060	0.066	0.066	0.061	0.000	***
GAMBELLA	0.158	0.185	0.166	0.169	0.131	0.000

Population abbreviation is: B. GUMUZ Benishangul-Gumuz

The pairwise coefficient of genetic differentiation between the populations ranged from 0.056 (B.Gumuz and Amhara) to 0.953 (between Amhara and Oromia). The highest and statistically significant genetic differentiation, (PhiPT=0.953) was observed between populations of Oromia, and Amhara. The genetic differences between populations of Tigray and B.Gumuz, Tigray and SNNP, Tigray and Gambella, Oromia and B.Gumuz, Oromia and SNNP, Oromia and Gambella, Amhara and B.Gumuz, Amhara and SNNP, Amhara and Gambella, B.Gumuz and SNNP, B.Gumuz and Gambella, Gambella and SNNP were statistically significant although slightly lower compared to some other pairs, still indicating substantial genetic divergence ($p>0.05$) (Table 6).

Table 6 Population genetic differentiation measured by PhiPT (below the diagonal) between the Six Sesame populations with p values (above the diagonal)

Pop	TIGRAY	OROMIA	AMHARA	B.GUMUZ	SNNP	GAMBELA
TIGRAY	****	0.9380	0.9361	0.9238	0.9340	0.8545
OROMIA	0.0640	****	0.9525	0.9148	0.9162	0.8250
AMHARA	0.0660	0.0487	****	0.9452	0.9174	0.8507
B.GUMUZ	0.0793	0.0890	0.0563	****	0.9294	0.8454
SNNP	0.0683	0.0875	0.0862	0.0732	****	0.8885
GAMBELA	0.1573	0.1924	0.1617	0.1680	0.1183	****

Population abbreviation is: B. GUMUZ Benishangul-Gumuz

4.1.6. Genetic relationship within and among populations

Analysis of molecular variance (AMOVA) revealed that 88% of genetic variation occurred with in populations (Table 6, Figure 4). The variability among populations only accounted for 12% of the total genetic diversity. The overall genetic differentiation coefficient among the populations was moderate (PhiPT = 0.124, p = 0.001) with high gene flow (6.2). Probability, P (rand >= data), for PhiPT is based on standard permutation across the full data set.

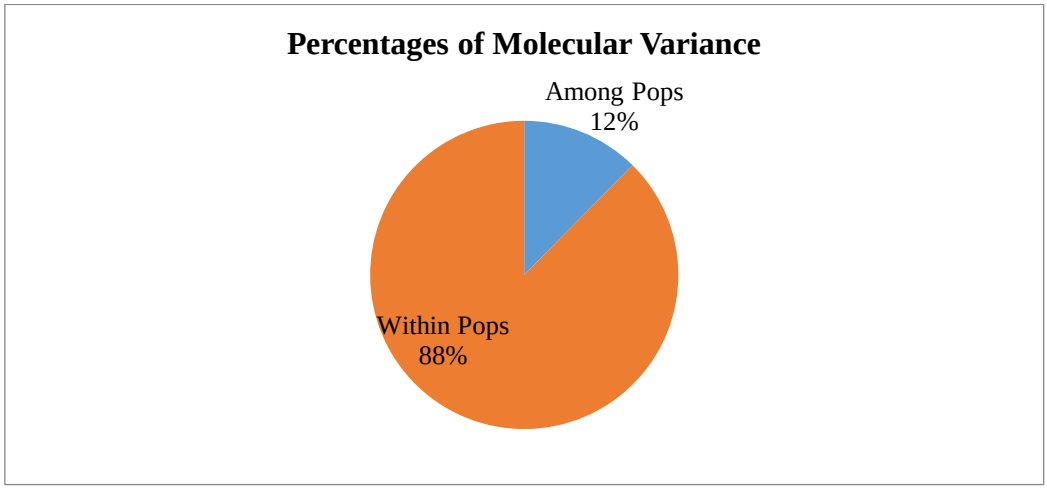


Fig 4. Analysis of molecular variation of pie chart for 93 *Sesamum indicum* L. of six populations

Table 7 AMOVA showing the partitioning of genetic variation within and among populations of Sesame

Source of variation	Df	SS	MS	Est.Var.	PV %	PhiPT	P value
Among Pops	5	449.491	89.898	4.190	12%	0.124	p<0.001
Within Pops	87	2579.326	29.647	29.647	88%		p<0.001
Total	92	3028.817		33.838	100%		

df degree of freedom, SS sum of squares, MS mean square, Est. Var. estimated variation, PV percent of variation

4.1.7. Cluster analysis using PCoA and UPGMA

PCoA is a technique frequently used in multivariate statistics to display the pattern of genetic structure and similarly to determine the amounts of variance described per component and cumulatively (Mekonnen *et al.* 2020). In the current research, PCoA explained by each axis shows some degree of segregation of different populations but almost all three axis have mixed population of others. So, at axis 1, the cumulative percentage is 5.73%, as only axis 1 has been considered. At axis 2, the cumulative percentage is 10.64%, which includes the variation explained by both axis 1 and axis 2. At axis 3, the cumulative percentage is 14.67%, which includes the variation explained by axis 1, axis 2, and axis 3 (Fig. 5).

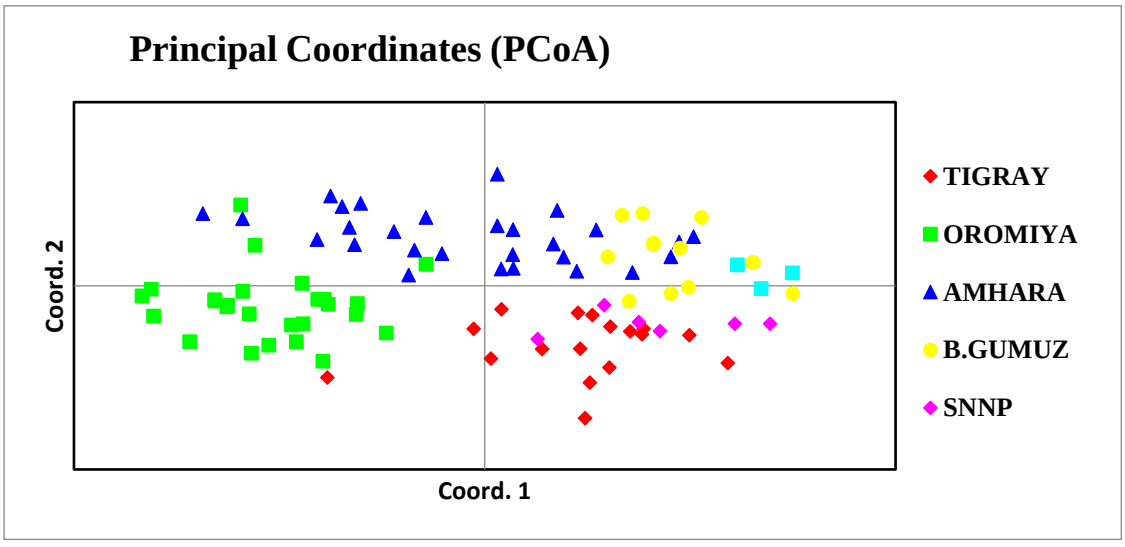


Fig. 5 Principal coordinate analysis (PCoA) of the 93 *Sesamum indiculasim* L. germplasms as revealed by 10 inter simple sequence repeat (ISSR) markers, samples coded with the same symbol and colour belongs to the same population. Population abbreviation: B. Gumuz Benishangul-Gumuz

The dendrogram generated using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) grouped the 93 Sesame genotypes based on the 204 bands(on the pair-wise comparison values of Nei's genetic distance) obtained from 10 ISSR primers the six populations into five major clusters (C1, C2, C3, C4 and C5), each of which was further grouped into two sub-clusters. From ninety three, 40 (43%) of the germplasms were assigned to C5 followed by C4 and C3, which were composed of 33 (35.5%) and 16 (17.2%) individuals. The lowest numbers (2 or 2.15%) of germplasm were assigned to cluster C1 (Fig. 6). none of the major clusters were composed of individuals from a single population, indicating the presence of high genetic admixture. The dendrogram generated using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) also grouped the six populations into five major clusters (Fig. 6)

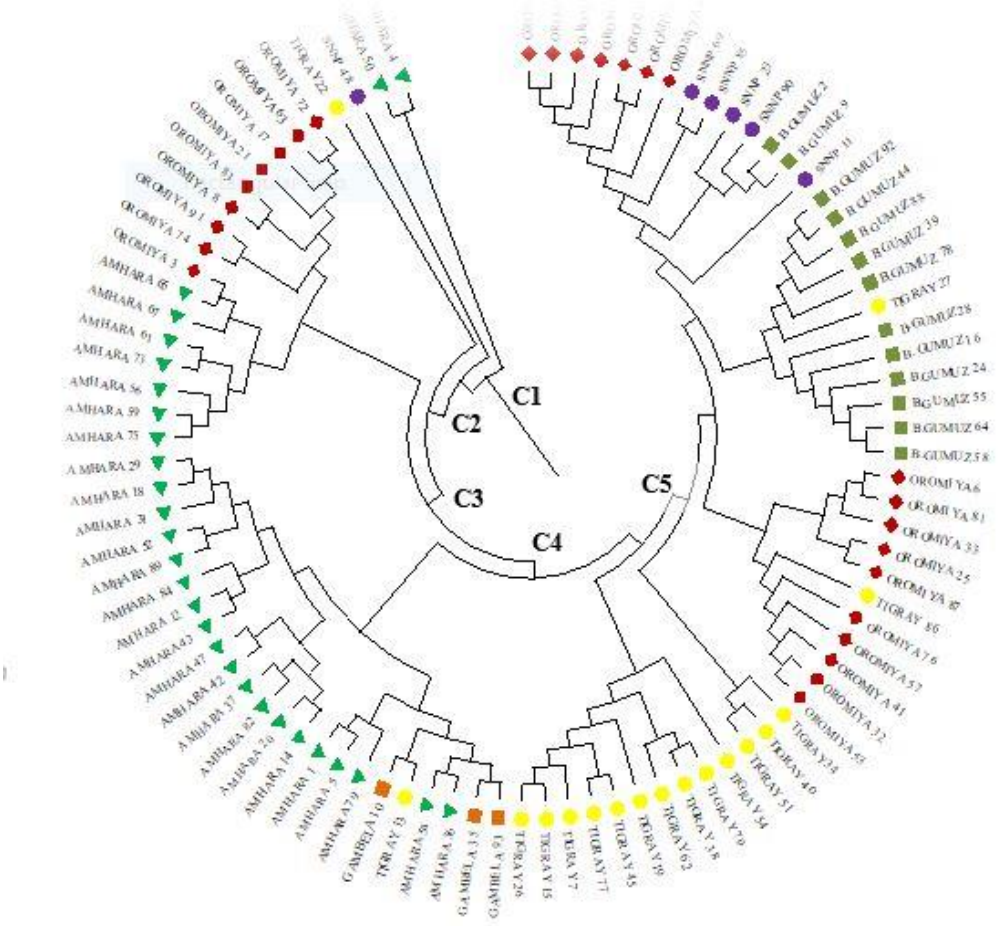
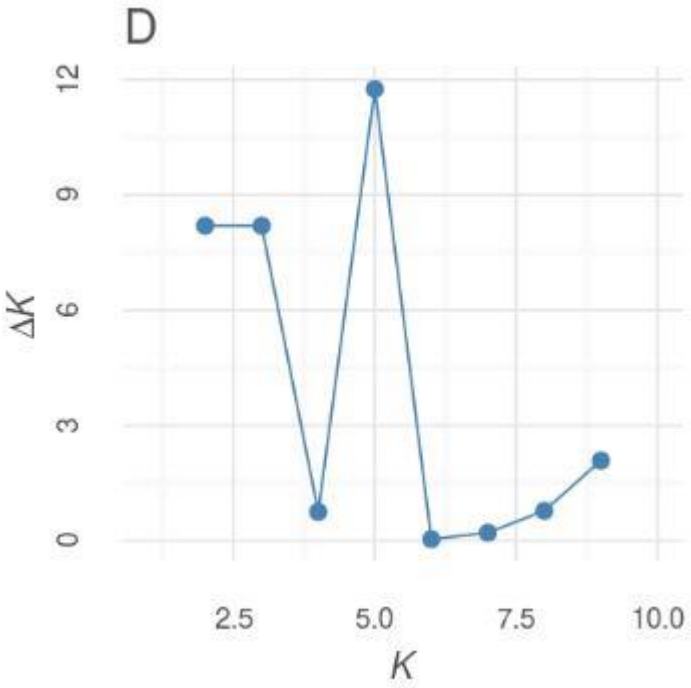


Fig 6. UPGMA tree showing relatedness among the 93 Sesame germplasms.

4.2. Population genetic structure analysis

It is known that testing for population structure is important while conducting association investigations and drawing relationships between markers. Accordingly, sesame germplasm can be efficiently categorized using population structure analyses, and maximum membership probability in STRUCTURE. The NJ, UPGMA clustering, and PCoA grouped the sesame populations, indicating the presence of high genetic intermixing among the Six Sesame populations. The Bayesian model-based population structure of the 93 Sesame genotypes was applied using Structure v2.3.4 software, and the K value was tested for 10 runs and used to estimate the number of clusters of the isolates. The output of the structure harvester revealed that the delta K (ΔK) values reached a sharp peak at K=5 (Fig. 7A). The Clumpak result (bar plot) detected a greater degree of genetic admixture between the three subpopulations (Fig. 7).



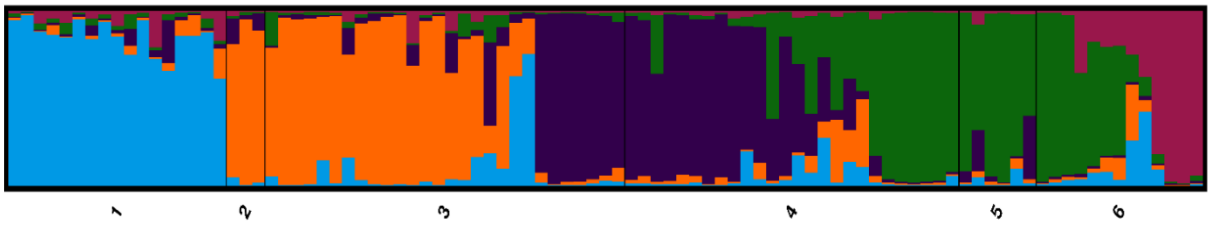


Fig 8. Population structure of 93 sesame germplasms representing six populations of breeding material.(a) Best delta K value estimated using the method of Evanno *et al.*, (2005); and (b) Estimated population structure for K = 5 according to breeding material. The color codes represent subpopulations : subpopulation I blue, subpopulation II orange, subpopulation III green, subpopulation IV purple and subpopulation V brown): the x-axis represents individual samples and different segments of each vertical line show extent of admixture in an individual genotype. Pop 1(Tigray), Pop 2(Oromia), Pop 3(Amhara), Pop 4(B.Gumuz),Pop 5(SNNP),Pop 6(Gambella).

Chapter Five

Discussion

Genetic diversity in Sesame has been increasingly narrowed down due to different reasons, modern breeding practices being the major one. For the improvement of existing germplasm genetic diversity investigations are important (Birindelli, 2017).In the current research, the genetic diversity and population structure analysis of 93 sesame germplasms were conducted so that they could be used in the selection of parents with desirable traits for forthcoming breeding programs and also for sustainable use of this valuable genetic resource.

5.1. Polymorphism level of the ISSR markers

In this work, the polymorphic information content of the used ISSR loci found across populations range from 0.24 to 0.38, with an overall mean of 0.32. The moderate PIC observed in most of the ISSR primers could be attributed to the diverse nature of the sesame germplasm and/or the highly in formativeness of the ISSR markers used. The most informative primer in the present study was UBC 834 with PIC value of 0.38 which is similar to previous result reported by (El-Sherbeny, 2020) on Wheat genotypes and the most informative primer on his work was primer UBC 834 with PIC value 0.19. The polymorphism information content (PIC) ranged from 0 to 0.5, with higher value indicating greater in formativeness.

5.2. Population genetic diversity

Worldwide sesame production is facing challenges due to the impact of abiotic and biotic stresses. Consequently, it is important to look for germplasm with the highest genetic diversity of important traits through research for effective conservation and improvement of existing germplasm. Sesame germplasms are thus used in many sesame breeding programs because they have a unique potential and diversity of essential genes that influence both biotic and abiotic stressors (Manickavelu, 2006; Gillespie, 2022).

It could be useful for exploring genetic constituents and identifying loci, which are then used to improve sesame performance in breeding programs. In the present work, genetic diversity in the studied populations was determined, and gene diversity and Shannon's information index were found to be 0.19 and 0.29, respectively. The levels of gene diversity and Shannon's information index obtained in the current research differed from those reported by

(Hsissou, 2021), most likely due to substantial gene flow, which reduced the amount of variation among groups.

In this study inter population genetic distance for the whole populations ranged from 0.056 to 0.953 from all six populations comparatively samples from Amhara, Tigray and Oromia , hence, for selection based population improvement, these populations have high potential as compared to populations with lower genetic diversity like Benishan Gumuz and Gambella. This might be due to high number of samples for the first three populations Amhara Oromia and Tigray and low number of samples from Benishan Gumuz and Gambella.

In the present study there was moderate level of genetic differentiation (Gst=0.24) among all sesame genotypes. In other study (Dagmawi, 2015) using ISSR marker also found Gst=0.25 among Ethiopian sesame accessions. Among the six studied sesame populations, the Amhara population showed the highest observed number of alleles (Na=1.70), effective number of alleles (Ne=1.35), Nei's gene diversity (h=0.36), Shannon's information index (I=0.28) and PPL (84.80%). This could likely be due to its highest population sample size as compared to the other populations. This implies that Amhara populations could be a good source of valuable alleles for sesame improvement as compared to other populations.

5.3. Population genetic structure

The information gained from population structure help with germplasm management and conservation strategies, as well as the choice of suitable parents in sesame breeding programs. In this study Analysis of molecular variance (AMOVA) revealed high (88%) genetic diversity within population than among populations (12%), likely due to high sexual recombination within the population and high gene flow among populations. The high gene flow (Nm=5.34) could be attributed to the genetic diversity differences within and among populations. Basically, a high level of genetic differentiation among populations is inversely proportional to gene flow (Aldrich, 1992). Genetic diversity is considerably influenced by gene flow, which encompasses several mechanisms of gene exchange among populations (Slatkin, 1987).

This could be a consequence of exchange of genetic materials among the neighbouring farmers as well as traders in the region. The human factor has been previously shown to be responsible for the lack of correlation between genetic and geographical distance (Stankiewicz, 2001). The increasing demand for sesame by exporters in Ethiopia, the movement of traders and exchange of germplasm across different regions could be a possible

explanation for the spreading of sesame seeds. In the present study, pollen movement among populations, seed exchange practices between communities, and common marketing as well as socio-economic conditions may have facilitated the high gene flow (Nm=4.99) among populations, leading to lower among-population genetic differentiation (Fst: 0.14).

Principal coordinate analysis and UPGMA also confirmed the presence of higher genetic variation at the population level than among populations, where the individuals of the different populations failed to form distinct clusters (Geetharamani, 2019), rather were mixed up along the axis. PCoA resulted in five clusters, where none of the clusters were composed of germplasm entirely from a particular population. This impliesthe existence of a significant intermixing of genetic backgrounds among the populations. Likewise, STRUCTURE analysis supported PCoA, and signifying the presence of high genetic relationships among the studied Sesame populations, likely due to the presence of higher gene flow. The STRUCTURE analysis detected five sub-groups ($\Delta K=5$), with a greater degree of genetic admixture. In cluster I, interestingly only one pair of accessions from Amhara (Amhara50 and Amhara04) could not be distinguished (genetic similarity of 1.00), and can be considered as duplicates for the purpose of shortlisting germplasm for conservation. (Haritha, 2016) studied on *Oryza rufipogon* and found out In cluster II, interestingly only one pair of accessions from India (IRGC104425 and IRGC104639) could not be distinguished (genetic similarity of 1.00), and can be considered as duplicates for the purpose of shortlisting germplasm for conservation.

Chapter Six

Conclusion and Recommendation

6.1. Conclusion

In the present work, genetic diversity and population structure of 93 of sesame genotypes were profiled, examined and analysed using highly informative ISSR markers to generate valuable information for sesame breeding programs and conservation purposes. The analysis revealed that markers used were entirely highly informative, and hence, very suitable tools to describe genetic diversity and population structure of sesame germplasms.

The research also revealed how the sesame genotypes are structured, and their potential to contribute to the sesame improvement programs through selection breeding, and also designing proper management strategies. Among the six studied populations, Amhara showed relatively higher Shannon's Information Index (I=0.3) and Ne's gene diversity, implying that it could serve as a good source of desirable genes for sesame improvement program. UPGMA based clustering, PCoA analysis, and STRUCTURE analysis inadequately grouped the populations, confrming high genetic intermixing and high gene flow among populations. The STRUCTURE analysis confirmed the studied sesame populations shared genetic background that originated from five sub-populations.

In this work the marker with high PIC value was identified it was marker UBC 834 (0.38) and also two divergent genotypes was identified in the present work at cluster I (Amhara 50 and Amhara 04) . So Identifying which accessions are divergent and and understanding the geographical distribution will provide useful information and abundant source genes for Sesame breeding program to make new cultivators ,adapting to a wide range of ecological regions. For example in cluster I there were two divergent Sesame genotypes they could be target for selection and hybridization and also for conservation purposes.

6.2. Recommendation

We recommend further research using high-density marker that includes large sets of sesame genotypes and more number of markers to generate a better insight from the whole nation to disclose the genetic structure of sesame in the country, and also for development of best-performing cultivars.

Chapter Seven

References

Abdellatef. (2008). Study of genetic diversity in Sudanese sesame (Sesamum indicum L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *African Journal of Biotechnology, 7*(24).

Ahmad. (2009). Environment-friendly renewable energy from sesame biodiesel. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 32*(2), 189-196.

Ali, A. A. F. (2015). *Biology and damage inflicted by the sesame seed bug, Elasmolomus sordidus Hemiptera (lygaeidae.* west kordfen unversity.

Anderson, E. (1960). The evolution of domestication. *S. Tax, S., The Evolution of Man. Man, Culture, and Society. Vol. II of Evolution after Darwin, Chicago and London (The University of Chicago Press) 1960, pp. 67-84.*

Ashri, A., & Singh, R. (2007). Genetic resources, chromosome engineering, and crop improvement. *Journal of Plant Breeding Revolution, 4*, 231-289.

Baker, H. (1962). Comments on the thesis that there was a major centre of plant domestication near the headwaters of the River Niger. *The Journal of African History, 3*(2), 229-233.

Basak. (2019). Genetic diversity and population structure of the Mediterranean sesame core collection with use of genome-wide SNPs developed by double digest RAD-Seq. *PLoS One, 14*(10), e0223757.

Bedigian, D. (2003). Evolution of sesame revisited: domestication, diversity and prospects. *Genetic Resources and Crop Evolution, 50*(7), 779-787.

Bedigian, D., & Harlan, J. R. (1986). Evidence for cultivation of sesame in the ancient world. *Economic botany, 40*(2), 137-154.

Burkill, I. H. (1953). *The Hooker Lecture: Habits of Man and the Origins of the Cultivated Plants of the Old World.* Paper presented at the Proceedings of the Linnean Society of London.

Çağırgan, M. İ. (2006). Selection and morphological characterization of induced determinate mutants in sesame. *Field crops research, 96*(1), 19-24.

Dagmawi, T. W., Kassahun, T., & 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), 7-18.

Dossa. (2017). Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: SisatBase. *Frontiers in plant science, 8*, 1470.

Dossa, K., Wei, X., Zhang, Y., Fonceka, D., Yang, W., Diouf, D., . . . Zhang, X. (2016). Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. *Genes, 7*(4), 14.

Elleuch. (2007). Quality characteristics of sesame seeds and by-products. *Food Chemistry, 103*(2), 641-650.

Ercan. (2004). Analysis of genetic diversity in Turkish sesame (*Sesamum indicum* L.) populations using RAPD markers⋆. *Genetic Resources and Crop Evolution, 51*(6), 599-607.

Ford-Lloyd, B., & Jackson, M. (1986). *Plant genetic resources: an introduction to their conservation and use*: Edward Arnold.

Ganesh, S., & Thangavelu, S. (1995). Genetic divergence in sesame (*Sesamum indicum* .L). *Madras Agricultural Journal, 82*, 263-265.

Gebremichael, D. E., & Parzies, H. K. (2011). Genetic variability among landraces of sesame in Ethiopia. *African Crop Science Journal, 19*(1).

Gelalcha, S. D. (2009). Sesame trade arrangements, costs and risks in Ethiopia. *Ministry of Foreign Affairs, The Haque*.

Girmay, A. B. (2018). Sesame production, challenges and opportunities in Ethiopia. *Agricultural Research & Technology: Open Access Journal, 15*(5), 555972.

Gitima, G., & Mersha, M. (2020). The Impacts of El-Niño-Southern Oscillation (ENSO) on Agriculture and Coping Strategies in Rural Communities of Ethiopia: Systematic Review Article. *Asian Journal of Geographical Research*, 56-69.

Ikten, C., Catal, M., Yol, E., Ustun, R., Furat, S., Toker, C., & Uzun, B. (2014). Molecular identification, characterization and transmission of phytoplasmas associated with sesame phyllody in Turkey. *European Journal of Plant Pathology, 139*(1), 217-229.

Ishaq. (2007). *Germplasm conservation and its impact on crop improvement in Nigeria.* Paper presented at the Plant genetic resources and food security in West and Central Africa. Regional Conference, Ibadan, Nigeria, 26-30 April, 2004.

Isshiki, S., & Umezaki, T. (1997). Genetic variations of isozymes in cultivated sesame (*Sesamum indicum* L.). *Euphytica, 93*(3), 375-377.

Karthik, L., Kumar, G., Keswani, T., Bhattacharyya, A., Chandar, S. S., & Bhaskara Rao, K. (2014). Protease inhibitors from marine actinobacteria as a potential source for antimalarial compound. *PLoS One, 9*(3), e90972.

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

Kumar. (2012). Molecular characterization and assessment of genetic diversity in sesame (*Sesamum indicum* L.) germplasm collection using ISSR markers. *Journal of Crop Improvement, 26*(4), 540-557.

Laurentin, H. E., & Karlovsky, P. (2006). Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC genetics, 7*(1), 1-10.

Lercher, M. J., & Hurst, L. D. (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in genetics, 18*(7), 337-340.

Li, D., Liu, W., Zhang, Y., Wang, L., Wei, W., Gao, Y., . . . Zhang, X. (2013). Identification method of drought tolerance and association mapping for sesame (*Sesamum indicum* L.). *Acta Agronomica Sinica, 39*(8), 1425-1433.

Mahajan, R., Bisht, I., & Dhillon, B. (2007). Establishment of a core collection of world sesame (*Sesamum indicum* L.) germplasm accessions. *Sabrao Journal of Breeding and Genetics, 39*(1), 53-64.

Mehra, K. (1963). Consideration of the African origin of Eleusine coracana (L.) Gaertn. *Curr. Sci., 32*, 300-301.

Nayar, N. M., & Mehra, K. (1970). Sesame: Its uses, botany, cytogenetics, and origin. *Economic botany*, 20-31.

Nimmakayala, P., Levi, A., Abburi, L., Abburi, V. L., Tomason, Y. R., Saminathan, T., . . . Wehner, T. C. (2014). Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC genomics, 15*(1), 1-15.

Ohmido. (2000). Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in indica and japonica rice. *Molecular and General Genetics MGG, 263*(3), 388-394.

Pham, T. D., Bui, T. M., Werlemark, G., Bui, T. C., Merker, A., & Carlsson, A. S. (2009). A study of genetic diversity of sesame (*Sesamum indicum* L.) in Vietnam and Cambodia estimated by RAPD markers. *Genetic Resources and Crop Evolution, 56*(5), 679-690.

Pham, T. D., Geleta, M., Bui, T. M., Bui, T. C., Merker, A., & Carlsson, A. S. (2011). Comparative analysis of genetic diversity of sesame (*Sesamum indicum* L.) from Vietnam and Cambodia using agro-morphological and molecular markers. *Hereditas, 148*(1), 28-35.

Ram, S. G., Sundaravelpandian, K., Kumar, M., Vinod, K., Bapu, J., & Raveendran, T. (2006). Pollen–pistil interaction in the inter-specific crosses of Sesamum sp. *Euphytica, 152*(3), 379-385.

Singh, R. J., & Jauhar, P. P. (2006). *Genetic Resources, Chromosome Engineering, and Crop Improvement: Cereals, Volume 2*: CRC press.

Supriya, P., & Bhat, K. (2018). Genome-wide identification of genes, transcription factors and transposable elements in sesame (*Sesamum indicum* L.). *Int J Curr Microbiol App Sci, 7*(2), 2362-2366.

Terefe. (2012). Sesame production manual. *Ethiopia: Ethiopian Institute of Agricultural Research Embassy of the Kingdom of the Netherlands*.

Tripathy, S. K., Kar, J., & Sahu, D. (2019). Advances in sesame (Sesamum indicum L.) breeding *Advances in plant breeding strategies: industrial and food crops* (pp. 577-635): Springer.

Uncu. (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), 1-12.

Uzun. (2003). Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. *Plant Breeding, 122*(1), 95-97.

Uzun. (2008). Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (Sesamum indicum L.). *Journal of the American Oil Chemists' Society, 85*(12), 1135-1142.

Uzune. (2003). Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. *Plant Breeding, 122*(1), 95-97.

Venkataramana Bhat. (1999). Study of genetic diversity in Indian and exotic sesame (Sesamum indicum L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica, 110*(1), 21-34.

Wei, X., Gong, H., Yu, J., Liu, P., Wang, L., Zhang, Y., & Zhang, X. (2017). SesameFG: an integrated database for the functional genomics of sesame. *Scientific Reports, 7*(1), 1-10.

Wei, X., Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X., . . . Zhang, X. (2014). Development of simple sequence repeat (SSR) markers of sesame (Sesamum indicum) from a genome survey. *Molecules, 19*(4), 5150-5162.

Weiss, E. (1971). Castor, sesame and safflower. Barnes and Noble. *Inc., New York, 529*.

Wu, K., Yang, M., Liu, H., Tao, Y., Mei, J., & Zhao, Y. (2014). Genetic analysis and molecular characterization of Chinese sesame (Sesamum indicum L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. *BMC genetics, 15*(1), 1-15.

Wu, W.-H. (2007). The contents of lignans in commercial sesame oils of Taiwan and their changes during heating. *Food Chemistry, 104*(1), 341-344.

Xia, D., Lin-hai, W., Yan-Xin, Z., Dong-hua, L., Yuan, G., Wen-liang, W., . . . Xiu-rong, Z. (2013). Genetic variation and associated mapping for traits related to plant height constitutions in core collections of sesame (Sesamum indicum L.). *CHINESE JOURNAL OF OIL CROP SCIENCES, 35*(3), 262.

Yepuri, V., Surapaneni, M., Kola, V. S. R., Vemireddy, L., Jyothi, B., Dineshkumar, V., . . . Siddiq, E. (2013). Assessment of genetic diversity in sesame (Sesamum indicum L.) genotypes, using EST-derived SSR markers. *Journal of Crop Science and Biotechnology*, *16*(2), 93-103.

Zhang, H., Miao, H., Wei, L., Li, C., Zhao, R., & Wang, C. (2013). Genetic analysis and QTL mapping of seed coat color in sesame (Sesamum indicum L.). *PLoS One, 8*(5), e63898.

Zhang, Y., Wang, L., Li, D., Gao, Y., Lu, H., & Zhang, X. (2014). Mapping of sesame waterlogging tolerance QTL and identification of excellent waterlogging tolerant germplasm. *Sci. Agric. Sin*, *47*, 422-430.

Zhang, Y., Wang, L., Li, D., Wei, W., Gao, Y., & Zhang, X. (2012). Association mapping of sesame (Sesamum indicum L.) resistance to Macrophomina phaseolina and identification of resistant accessions. *Scientia Agricultura Sinica, 45*(13), 2580-2591.

Chapter Eight

Appendices

Appendix 1. DNA extraction protocol (Modified CTAB method) described by Tilahun Mekonnen *et al*., 2017.

1. Introduction

This protocol describes a modified CTAB method for the isolation of DNA from plant material.

A maximum of 24 samples can be isolated at once using this protocol.

2. Materials

CTAB	-	1 x 700 µl per sample
ß-Mercapto-Ethanol		
Chloroform		
Isopropanol	-	100 %, 4°C
Ethanol	-	100 % and 70% in sterile ddH2O
TE	-	1x, p.a. grade
NH4Ac	-	7.5 M Solution, sterile
NaAc	-	3 M Solution, sterile
Water bath	-	65°C
Centrifuge	-	4°C
Cut 1000 µl tips	-	Sterile

3. Procedure

1. Poor CTAB solution (700 µl per sample) in a 1.5ml-tube and add 0.2 vol % Mercaptoethanol

(Use fume hood!). Mercapto-ethanol is stored at 4°C.

2. Aliquot CTAB in 1.5 ml Eppendorf-caps and warm in water bath up to 65°C.

3. Weigh in 100 mg fresh leave material (50mg dry material) per sample. Pulverise thoroughly using a Geno grinder machine.

4. Add 700 µl of warm CTAB solution to the powdered sample (open the caps carefully), dissolve the powder and incubate the sample for 30 minutes at 65°C.

6. Centrifuge for 5 minutes at 15000 rpm.

7. Transfer the supernatant (only clear liquid) in a new Eppendorf-cap. Use blue pipette tips which are cut.

8. Add new CTAB solution (700 µl) to the tissue pellet and stir slightly with a new 1000 µl pipette tip, incubate 30 min at 65°C. Step 6 and 7 are repeated. The same is carried out for a third extraction. Each fraction proceeds with step 9 and is treated separately.

9. Add 600 µl Chloroform to the cap with supernatant and shake carefully a few times upside down. This chloroform step should be carried out immediately.

10. Shake the samples thoroughly by turning inversing the Eppendorf caps for approximately 5 minutes. (Longer incubation is possible)

11. Centrifuge for 5 min at 15000 rpm.

12. Transfer the supernatant (only clear liquid) in a new Eppendorf-cap. Use blue pipette tips which are cut. Work carefully; do not transfer suspended matter (normally the Chloroform Is covered by a thin layer of fine sediment material). Chloroform has to be disposed of in a special waste bottle.

13. Repeat the chloroform extraction (step 9-12) to make sure that all impurities are removed, and then proceed with step 14.

14. Add cooled Iso-propanol (4°C), approximately 2/3 of the solution volume. Shake carefully by inversing the Eppendorf cap. In most cases DNA becomes visible as white threads. Freeze for more than 2 h at -20°C.

15. Centrifuge 10 min at 15000 rpm.

16. Aspirate liquid using yellow tips (without touching pellet!). If pellet is solid enough the larger part of the liquid may be poured out. (Alternatively add TE and proceed with Qiagen kit)

17. Add 200 µl Ethanol 70 % to the pellet. Rinse the inner cap surface by turning the cap.

18. Centrifuge for 10 min at 15000 rpm in a cooled centrifuge.

19. Aspirate Ethanol using yellow tips. Dry the DNA-pellet at room temperature. (Usually 15 min are sufficient; after drying no liquid drops are to be seen)

20. Dissolve pellet in 100 µl TE (1x, p.a. grade) and store at 4°C.

21. Add/ cooled 7.5 M NH4Ac-solution (4°C, half of the solution volume). Mix carefully.

22. Add cool Ethanol 100 % (double of the solution volume). Mix carefully. Freeze for more than 2 hr at -20°C.

23. Centrifuge 30 min at 15000 rpm. Aspirate fluid carefully.

24. Add 200 µl Ethanol 70%. Rinse the inner cap surface by turning the cap.

25. Centrifuge 10 min. at 15000 rpm. Aspirate liquid and dry pellet at room temperature.

    Dissolve the pellet in 100 µl TE (1x, p.a. grade)

26. Repeat steps 21 to 24 with 3 M Na-AC-solution (4°C, half the volume) then proceed with Step 27.

27. Centrifuge 10 min. at 15000 rpm. Aspirate liquid and dry pellet at room temperature. Dissolve the pellet in 100 µl TE (1x, p.a. grade) and store at -20^{o}c for further use.