GENETIC DIVERSITY AND POPULATION STRUCTURE ANALYSIS OF ETHIOPIAN SORGHUM [*SORGHUM BICOLOR* (L.) MOENCH] LANDRACES USING SIMPLE SEQUENCE REPEAT (SSR) MARKERS

ABDULNASIR KEDIR MUHAMED

A THESIS SUBMITTED TO THE DEPARTMENT OF APPLIED BIOLOGY
SCHOOL OF APPLIED NATURAL SCIENCE

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOTECHNOLOGY

OFFICE OF GRADUATE STUDIES

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

JANUARY, 2022

ADAMA, ETHIOPIA

Genetic Diversity and Population Structure Analysis of Ethiopian Sorghum [*Sorghum bicolor* (L.) Moench] Landraces Using Simple Sequence Repeat (SSR) Markers

Abdulnasir Kedir Muhamed

Advisor: Mulugeta Kebede (PhD, ASTU)

Co-advisor: Tileye Feyissa (Prof., AAU)

A Thesis Submitted to the Department of Applied Biology
School of Applied Natural Science

Presented in Partial Fulfilment of the Requirements for the Degree of Master of Science in Biotechnology

Office of Graduate Studies

Adama Science and Technology University

January, 2021

Adama, Ethiopia

Declaration

I hereby declare that this Master Thesis entitled “Genetic Diversity and Population Structure Analysis of Ethiopian Sorghum [*Sorghum bicolor* (L.) Moench] Landraces Using Simple Sequence Repeat Markers” is my original work. 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

we, the advisor(s) of this thesis, hereby certify that I/we have read the revised version of the thesis entitled "Genetic Diversity and Population Structure Analysis of Ethiopian Sorghum [*Sorghum bicolor* (L.) Moench] Landraces Using Simple Sequence Repeat (SSR) Markers" submitted in partial fulfillment of the requirements for the degree of Master of Science in Applied Biology (Biotechnology) prepared by **Abdulnasir Kedir Muhamed, ID No: A/PR15437/10,** was conducted under our direct supervision. This is therefore to kindly request the Department of Applied Biology at School of Applied Natural Science, Adama Science and Technology University to officially approve the thesis for open defense following the applicable procedures.

_____	_____	_____
Name of the major Advisor	Signature	Date
_____	_____	_____
Name of the Co-advisor	Signature	Date

Approval page of M.Sc. Thesis

I/we, the advisors of the thesis entitled "Genetic Diversity and Population Structure Analysis of Ethiopian Sorghum [*Sorghum bicolor* (L.) Moench] Landraces Using Simple Sequence Repeat (SSR) Markers" and developed by **Abdulnasir Kedir**, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Name of the major Advisor	Signature	Date
Name of the Co-advisor	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by **Abdulnasir Kedir** have read and evaluated the thesis entitled "Genetic Diversity and Population Structure Analysis of Ethiopian Sorghum [*Sorghum bicolor* (L.) Moench] Landraces Using Simple Sequence Repeat (SSR) Markers" and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Biotechnology.

Chairperson	Signature	Date
Internal examiner	Signature	Date
External examiner	Signature	Date

Final approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department head	Signature	Date
School Dean	Signature	Date
Office of Postgraduate Studies, Dean	Signature	Date

Acknowledgments

First and foremost, I want to give my thanks to Allah (The Almighty) Alhamdulillah Rabil Alamiin! Who gave me the power, strength, and ability to carry out the thesis through all the difficulties surrounding me. I would like to thank Adama Science and Technology University (ASTU), School of Applied Natural Sciences, department of biology for the complete support provided and full concern shown during the study period. I would also like to thank Addis Ababa University for allowing me to use the Plant Genetics Research molecular laboratory and Ethiopian Biodiversity Institute for providing the Sorghum seeds for the study.

I am also very grateful to my major advisor, Dr. Mulugeta Kebede for his guidance, advice, kindness, patience, and responsibility in all of my works till accomplishment. My deepest thanks go to my co-advisor, Prof. Teleye Feyissa for giving me the chance to work in the SIDA Sorghum project, and also for his valuable inputs, advice, guidance, and enthusiastic encouragement from beginning to the end of this thesis. My warm gratitude also goes to Dr. Kassahun Tesfaye, Director General of Ethiopian Biotechnology Institute (EBTi) for allowing me to work in his Plant Genetics Research Laboratory at the Institute of Biotechnology of Addis Ababa University (AAU) with all the required resources and facilities.

Special thanks also go to Dr. Tilahun Mekonnen, Mr. Temesgen Dagnaw, and Mr. Muluken Birara for their close supervision during the laboratory work, utilization of software, data analysis, and interpretation of the results. I am very thankful to Dr. Tilahun Mekonnen for his critical comments and constructive suggestion during the review process. I would like to thank Dr. Gamado Dale, Dr. Ahmed Hussen, Dr. Tesfaye Alemu, Dr. Bullo Neda, Dr. Demisachew Guade, Dr. T/Haimanot H/ Selassie, Mss. Hawi Nigussie, Mr. Behailu Mulugeta, Mss. Workiya Ahmed, Mss. Leyla Nasir, Mr. Mefewus Asfaw, Mss. Gete Eshetu, Mr. Abdisa Barisoo, Mr. Abdisa Kedir, Mr. Juneydin Safi, Mr. Homa Feyisa, Mr. Mubarek Tamene, Mr. Suleyman Bushura, Mr. Firaol Ofga, and all friends around me for their continued support and encouragement during my stay in the laboratory.

I would like to extend my heartfelt and limitless thanks to my family, my late father Haji Kedir Muhamed (Haji Kedir Bombaasoo) who was a father of all kinds, my mother Zahra Haji Muhamud Sheikh Husen (Karara), and my late grandmother Nuritu Sixamoo (Aallaa), to whom I want to dedicate this thesis. Once again, I would like to give a lovely thanks to my mother for her

remarkable and infinite love, encouragement, sacrifice, prayer, support, advice, and always wished to see this day. In addition, my brother Kemal Haji Kedir, and my sisters Rehima Haji Kedir, Hawa Haji Kedir, and Zebiba Haji Kedir, and Gemmechu Korro, Fiixaa Gammachu, Iftu Gammmachu, without their support, the completion of this work might not be possible. I am also thankful to Fayisa Edao for his special support and all his encouragement throughout my career to achieve my goal.

Finally, I am thankful to Sheikh Mohammed-Siraj Mohammed-Amiin, my second mother Amatulla sheikh Muhammed (Ummi), Derartu Samuel, Muhammed Ali (Haji Ware), Sheikh Abdulkadir Muhamed, Sheikh Hamid Muhamed, and all other people who have made meaningful contributions to complete this work. I would like to say thank you to all of those whose names may not have been listed. May Allah bless you all!

TABLE OF CONTENTS

Contents **Pages**

LIST OF TABLES ..... **ix**

LIST OF FIGURES ..... **x**

LIST OF APPENDICES ..... **xi**

LIST OF ACRONYMS ..... **xii**

ABSTRACT ..... **xiii**

1. INTRODUCTION **1**

1.1. Background of the study 1

1.2. Statement of the problem 4

1.3. Objectives 5

1.3.1. General objective 5

1.3.2. Specific objectives 5

1.4. Significance of the Study 5

2. LITERATURE REVIEW **6**

2.1. Origin, domestication, and distribution of sorghum 6

2.2. Taxonomic classification and races 6

2.3. Agroecology of sorghum 8

2.4. Sorghum production, constraints, and utilization 8

2.5. Genetic diversity of sorghum 10

2.6. Marker systems used in crop genetic diversity study 11

2.6.1. Morphological markers 11

2.6.2. Biochemical markers 12

2.6.3. Molecular markers 12

3. MATERIALS AND METHODS **14**

3.1. Plant materials 14

3.2. DNA extraction 15

3.3. Genotyping 15

3.4. Data scoring and statistical analysis 18
4. RESULTS 19
4.1. Microsatellite markers based polymorphism 19
4.2. Genetic diversity indices among populations and regions 20
4.3. Genetic Distance, differentiation, and gene flow between regions and populations 23
4.4. Analysis of molecular variance 26
4.5. Cluster analysis 27
4.6. Principal coordinates analysis 29
4.7. Population structure 30
5. DISCUSSION 32
5.1. Polymorphism level of the microsatellite markers 32
5.2. The genetic diversity of populations and regions 33
5.3. Genetic distance, differentiation, and gene flow 34
5.4. Genetic relationship 35
5.5. Population structure 35
6. CONCLUSION AND RECOMMENDATIONS 37
6.1. CONCLUSION 37
6.2. RECOMMENDATIONS 38
7. REFERENCES 39
8. APPENDICES 52

LIST OF TABLES

Tables **Pages**

Table 1. Taxonomic classification of S. bicolor (Source: CABI) 7

Table 2. Top 10 sorghum producer countries in the world (Source: FAOSTAT, 2021) 9

Table 3. Sorghum producer regions in Ethiopia (CSA, 2020) 9

Table 4. Primer's sequence, repeat motif, chromosome number, annealing temperatures, and the range of molecular weight detected using 11 microsatellite markers of sorghum in Ethiopia 16

Table 5. The PCR cycles and annealing temperature used to amplify the 11 SSR markers in 77 landraces of sorghum 17

Table 6. Summary of genetic diversity indices for 11 microsatellite markers across seventy-seven sorghum landraces 20

Table 7. Allelic patterns and diversity indices across the regions averaged over the 11 SSR loci 22

Table 8. Allelic patterns and diversity indices across populations averaged over the 11 SSR loci 22

Table 9. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among the five regions of Ethiopia 23

Table 10. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among the ten populations in Ethiopia 24

Table 11. Estimate of genetic differentiation measured by Fst (below the diagonal) among five regions 25

Table 12. Estimate of genetic differentiation measured by Fst (below the diagonal) among ten populations 25

Table 13. Analysis of molecular variance (AMOVA) showing the partitioning of genetic variation within and among populations, and regions of Ethiopian sorghum landraces 26

Table 14. Number of genotypes contributed by populations to the clusters 28

LIST OF FIGURES

Figures **Pages**

Figure 1. Sorghum basic races by their basic spikelet types. 7

Figure 2. Map of Ethiopia illustrating the collection region and sites of sorghum landraces 14

Figure 3. Generated electrophoretic bands of SSR markers of sorghum landraces collected from different parts of Ethiopia……...19

Figure 4. Neighbor-Joining tree showing relatedness among the 77 sorghum landraces of regions and populations. 27

Figure 5. The Neighbor-Joining- based clustering showing genetic relationships among regions (a) and populations (b). 29

Figure 6. Principal coordinate analysis (PCoA) of 77 Ethiopian sorghum landraces based on regions (a) and populations (b) using 11 SSR markers. 30

Figure 7. Population structure of 77 Ethiopian sorghum landraces using the maximum peak of K values based on Evanno *et al*. (2005) prediction among the studied regions (a) and populations (b). 31

Figure 8. The population structure bar graphs of the studied regions (a), populations (b) of the 77 sorghum landraces of Ethiopia. 31

LIST OF APPENDICES

Appendices **Pages**

Appendix 1. List of sorghum landraces used in this study. 52

Appendix 2. Some of the sorghum landraces (Seeds) used for planting in this study. 55

Appendix 3. Sorghum seeds while growing in the greenhouse. 55

Appendix 4. Leaf samples dried in plastic bags filled with silica gels. 55

Appendix 5. Plant DNA extraction by using CTAB immersed leaf samples procedure and chemical used..... 56

Appendix 6. Materials and chemicals used for DNA extraction. 58

Appendix 7. The result of genomic DNA obtained via quantification and qualification of Nanodrop. 58

Appendix 8. Electrophoretic bands of SSR markers..... 59

Appendix 9. Representative pictures to give a further illustration of all practical works beginning from sample collection to the end of the laboratory session 61

Appendix 10. The image that shows how ARLEQUIN calculates for: variance among populations (AP), variance within populations (WP), Total (T), sum of squares (SSD). 62

Appendix 11. Identification of the true number of clusters (K) based on delta k (ΔK) of the structure harvester (a=region and b=population). 62

LIST OF ACRONYMS

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
CSA	Central Statistics Agency
CTAB	Cetyl-trimethyl Ammonium Bromide
EBI	Ethiopian Biodiversity Institute
EDTA	Ethylene Diamine Tetra Acetic acid
FAO	Food and Agriculture Organization of the United Nations
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
ISSR	Inter Simple Sequence Repeat
NJ	Neighbor-Joining
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SNP	Single Nucleotide Polymorphism
SNNP	Southern Nations, Nationalities, and Peoples
SSR	Simple Sequence Repeat
TAE	Tris-Acetate-EDTA

ABSTRACT

As developing effective conservation and management strategies relies on the genetic diversity of Sorghum, the study especially in a country like Ethiopia, which is the center of origin with a wide range of cultivation, is vital. Therefore, the present study has targeted to assess the genetic diversity and population structure of Sorghum ***[***Sorghum bicolor *(L.) Moench****]*** *landraces using simple sequence repeat (SSR) markers. Seventy-seven sorghum landraces representing five regions and ten populations were analyzed using 11 SSR markers. The analysis of diversity indices identifies a total of 49 alleles with an average of 4.5 alleles per locus. Different diversity parameters were computed, and the average Shannon's diversity index and expected heterozygosity of the 11 loci were 0.61 and 0.36, respectively. The polymorphic information content (PIC) of the loci ranged from 0.23 to 0.67 with an overall mean of 0.48. The study revealed high gene diversity within the populations ranging from 0.25 to 0.72 with an overall mean of 0.53. The genetic distance and differentiation between possible pairs of regions and populations were low to moderate. Hierarchical AMOVA revealed that 82% and 79% of the total variation accounted for variation within regions and populations, respectively, leaving only 18% and 21% variations among regions and populations, respectively. The unweighted Neighbour-Joining based clustering, PCoA, and population structure analysis failed to group landraces according to their collection region and population, indicating the presence of high gene flow (Nm =1.16 or 0.95) between regions and population, respectively. Structure analysis confirmed two sub-groups with a greater degree of genetic admixture for both regions and populations. Overall, the genetic markers used were highly informative and useful tools to uncover the genetic structure of sorghum. The study also revealed the presence of high genetic diversity in Ethiopian sorghum landraces, indicating the potential to improve through breeding, and the need to design appropriate management strategies to maintain the available genetic resources.*

Keywords/Phrases: *Ethiopia, Genetic diversity, Population structure, Sorghum bicolor, SSR*

1. INTRODUCTION

1.1. Background of the study

Sorghum **[***Sorghum bicolor* (L.) Moench**]** which is mostly a self-pollinated diploid species (Poehlman and Sleper, 1979), with a chromosome number of 2n = 2x = 20, a monocotyledon C4 *grass* plant of tropical origin and the fifth most important cereal crop in the world after maize, rice, wheat, and barley in terms of both production and harvested area (FAOSTAT, 2019). It belongs to the grass family *Poaceae* (formerly *Gramineae*) and the *Andropogoneae* tribe (De Wet, 1978; Clayton and Renvoize, 1986). The origin and domestication of cultivated sorghum were thought to be in northeastern Africa, most likely near the Ethiopian-Sudanese border (Doggett, 1988; Harlan, 1992; Dillon *et al*., 2007).

Sorghum has a lot of genetic variation that allows it to adapt to a wide range of photoperiod and temperature conditions (Craufurd *et al*., 1999). The tolerance of sorghum to biotic and abiotic stress made the crop being cultivated worldwide, largely in the arid and semi-arid areas (Vara and Staggenborg, 2009; Mohammed, 2017; Kidanemaryam Wagaw, 2019).

Globally, sorghum is the most important type of grain after wheat, rice, maize, and barley. It is grown for food, feed, and bioenergetic raw materials (Hariprasanna, 2016). In the developed world, the majority of sorghum is used for livestock feed, particularly in the United States, Mexico, Australia, and South Africa (Mann *et al*., 1983). However, it is the main cereal food for over 750 million people living in Africa, Asia, and Latin America, especially in the semi-arid tropics (FAO, 2011).

In recent years, Africa has become a leading sorghum producer, contributing more than 28 million tons of grain annually (FAOSTAT, 2021). Ethiopia is the third sorghum producer after the United States, which is the world's leading sorghum producer, and Nigeria (FAOSTAT, 2021). In Ethiopia, the produced grain is used for the sale and around 74% is consumed at the household level (Kidanemaryam Wagaw, 2019).

In Ethiopia, sorghum is a major staple food crop ranking third after teff (*Eragrostis tef*) and maize (*Zea mays*) in area coverage and fourth after maize, teff, and wheat in total production (CSA,

2020). It is cultivated in nearly every part of the country, and in 2019 about 1.8 million ha of land was grown with sorghum having a total production of 5.27 million tons (FAOSTAT, 2019). Although the current average national sorghum productivity in Ethiopia is 2.88 t ha^{-1} is significantly higher than that of the global average productivity (1.44 t ha^{-1}) (FAOSTAT, 2019).

Low soil fertility, drought, insect shoot fly, insect stem borers, quelea birds, and *Striga* are some of the main challenges to sorghum cultivation (Wortmann *et al.,* 2006). Nonetheless, the magnitude varies among the region. They are responsible for significant sorghum grain yield losses in Ethiopia and elsewhere in the world.

To alleviate the production challenges and to feed the ever-increasing population, improving Sorghum production is always important. Knowledge of the genetic structure of the crop is a prerequisite for improving traits through breeding. Nowadays, genetic markers are widely used to uncover the genetic structure of crop plants including sorghum.

Assessing genetic diversity is an important step in any crop improvement program. Knowledge of the genetic diversity of a crop generally helps breeders to select desirable parents for breeding programs to improve agronomic traits of yield, stress resistance, and tolerance (Allard, 1999). Therefore, genetic diversity study is essential for the development of improved varieties (Yaqoob *et al.,* 2015), which many researchers used various techniques of a marker system for the study of sorghum genetic diversity.

The genetic diversity within and among populations could be explored through morphological, biochemical (allozyme), and DNA (molecular) marker analysis (Nemera Geleta and Labuschagne, 2005; Mehmood *et al*., 2008; Govindaraj *et al.,* 2015). The genetic diversity analysis based on morphological characteristics such as flower color, seed shape, boom habits, and pigmentation is strongly affected by environmental factors, shows less polymorphism, and requires more time and resources for evaluation in the field. Their developmental and tissue specificity made biochemical markers less informative in a breeding program. Molecular markers have several advantages over morphological markers, in that, they are phenotypically neutral, occur throughout the genome, are not influenced by environments, and their expression does not depend on the age of the plant (Ashutosh, 2013).

So far, various marker systems like random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), inter simple sequence repeat (ISSR), microsatellite or simple sequence repeat (SSR), and single nucleotide polymorphism (SNP), and-the-like were developed and broadly used to disclose the genetic structure and in analyzing the genetic diversity of sorghum (Ruiz-Chután *et al*., 2019; Zhu *et al*., 2020).

Microsatellite marker is the preferred marker system for many genomic and molecular breeding applications of sorghum due to its technical simplicity, high performance, the potential for automation, and co-dominant nature (Missihoun *et al*., 2015). Due to these reasons, microsatellite markers have a great potential to detect genetic diversity and relationships in organisms. The genetic diversity of sorghum has been studied using SSR markers across many parts of the world (Dje *et al*., 2000; Manzelli *et al*., 2007; Ali *et al*., 2008; Danquah., 2019; Ng'uni *et al*., 2011, 2012; Burow *et al*., 2012; Mofokeng *et al*., 2014; Motlhaodi *et al*., 2014, 2017).

Ethiopia is the center of origin for sorghum (Doggett, 1965), the country is the highest contributor of sorghum genetic resources for the world and a potential source of novel genes (Haussmann *et al*., 2002, Reddy *et al*., 2009). For Ethiopian sorghum, genetic diversity studies based on microsatellite markers have been reported (Bisrat Ghebru *et al*., 2002; Amelework Beyene, 2012; Asfaw Adugna *et al*., 2013; Asfaw Adugna, 2014; Tesfaye Disasa, 2016; Alemu Tirfessa *et al*., 2020, etc.).

Understanding the Ethiopian sorghum germplasm is an important step for the manipulation and conservation of the crop. Plant improvement and breeding programs depend on the genetic diversity of the plants for efficient selection (Shapter *et al*., 2018). For these reasons, it is vital to assess the genetic diversity of sorghum towards developing a new breed with good features. The aim of this study was, therefore, to assess the genetic diversity and population structure of Sorghum landraces from Ethiopia using SSR markers.

1.2. Statement of the problem

Genetic diversity in crop species can be assessed based on morphological, biochemical, and molecular characteristics (Muthusamy *et al*., 2008). As it can be influenced by environmental factors, genetic analysis based on morphological traits is less reliable. Molecular markers have a clear advantage over morphological and biochemical markers as they are numerous and independent of environmental conditions (Raman *et al*., 2014).

Various efforts have been done to uncover the genetic diversity and population structure of Ethiopian sorghum landraces using various marker systems such as (RAPD by Amsalu Ayana *et al*., 2000, ISSR by Hailekiros tadese and Tiliye Feyissa, 2013, SSR by Asfaw Adugna, 2014, etc.). One of the most convenient and popular molecular methods for identifying and studying genetic polymorphism is using SSR markers. Microsatellites are a marker of choice for genetic analysis because of their Co-dominantly inherited, highly informative, and widely distributed along the genomes.

Genetic diversity assessments are extremely critical for understanding the level and distribution of species diversity and for designing appropriate conservation strategies (Endashaw Bekele, 1985). The examination of the genetic diversity of sorghum is indispensable for genetic improvement and conservation.

Ethiopia, as a sorghum diversity hotspot, is a major contributor of sorghum genetic resources to the world (Reddy *et al*., 2008) and a potential source of novel genes (Haussmann *et al*., 2002; Reddy *et al*., 2009); however, published works on the microsatellite diversity of Ethiopian sorghum landraces are inadequate in terms of area coverage and the used seed materials. Therefore, Ethiopian sorghum landraces should have to be characterized further to broaden the knowledge and the base of genetic potential for breeders. Moreover, there are more than 9432 sorghum landraces conserved in Ethiopian Biodiversity Institute (EBI, 2021). Despite the huge sorghum landrace collections in the Ethiopian gene bank (9432) that might hold a great amount of genetic diversity, Limited studies have been conducted to uncover the genetic structure of Ethiopian sorghum landrace and one can reasonably decide to contribute towards the study on the landraces.

Hence, this study was carried out to assess the genetic diversity and population structure of sorghum landraces in Ethiopia using SSR markers to generate valuable information for sorghum improvement through breeding and its wise management. Therefore, the present study was designed with the following general and specific objectives.

1.3. Objectives

1.3.1. General objective

- To assess the genetic diversity and population structure of Ethiopian *S. bicolor* landraces using SSR markers.

1.3.2. Specific objectives

- To investigate the genetic diversity of *S. bicolor* landraces using SSR markers.
- To determine the population structure of the *S. bicolor* landraces in Ethiopia based on SSR markers.

1.4. Significance of the Study

The information from the present study:

- Will provide information on the genetic diversity of *S. bicolor* landraces which may assist in the identification and selection of genetic materials for improvement through breeding and germplasm conservation
- Can be used as a reference point for other researchers to do related works.

2. LITERATURE REVIEW

2.1. Origin, domestication, and distribution of sorghum

Sorghum is an important grain and its primary center of origin and diversity is northeastern Africa (Harlan, 1992; Dillon *et al*., 2007). Ethiopia is considered as the center of origin of sorghum (Doggett, 1988, Dogget and Rao, 1995). Archaeological evidence also suggests that the domestication of sorghum was begun in 5000 BC (Krzyzaniak, 1978) and was introduced to Egypt from Ethiopia around 3000 BC (Doggett, 1965). After its domestication, sorghum spread to various countries around the world, including India and China (Wang *et al*., 2013). *S. bicolor* is recently distributed globally, but Africa has the highest diversity of cultivated sorghums (Kimber *et al.*, 2013).

In Ethiopia, sorghum is the most important crop particularly in the lowland areas due to its drought tolerance and is widely distributed in the country (Yilma Kebede, 1991). The crop is cultivated in the altitude range of 400 m.a.s.l to 2500 m.a.s.l in all regions of Ethiopia, mostly at lower altitudes (EIAR, 2014). Therefore, the country is one of the main contributors of sorghum germplasm to the world collections at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) (Reddy *et al*., 2002).

2.2. Taxonomic classification and races

Sorghum is a member of the *Poaceae* family, the *Panicoid* subfamily, tribe *Andropogoneae*, the genus *Sorghum* and encompasses many species (De Wet, 1978; Clayton and Renvoize, 1986) (Table 1). It was believed that the genus *Sorghum* contained 20 to 30 species, of which 25 are currently recognized and taxonomically classified into five subgenera or sections: *Eusorghum, Chaetosorghum, Heterosorghum, Parasorghum, and Stiposorghum* based upon the node, panicle, and spikelet morphology (Garber, 1950). The section Eusorghum includes three species: *S. halepense* (L.) Pers. from India, *S. propinquum* (Kunth) Hitchc from Southeast Asia, and *S. bicolor* (L.) Moench from Africa (De Wet, 1978).

Table 1. Taxonomic classification of *S. bicolor* (Source: CABI)

Taxonomic Hierarchy	Taxonomic Name
Kingdom:	Plantae (plants)
Division:	Tracheophyta (vascular plants)
Subdivision:	Spermatophytina (seed plants)
Class:	Magnoliopsida
Order:	Poales
Family:	*Poaceae* (*Gramineae*)
Subfamily:	*Panicoid*
Tribe:	*Andropogoneae*
Subtribe:	Sorghinae
Genus:	*Sorghum*
Species:	*S. bicolor*

Sorghum is traditionally divided into five basic races based on inflorescence and grain morphology, which are bicolor, caudatum, durra, guinea, and kafir (Harlan and De Wet, 1972; De Wet, 1978) (Figure 1). The most important taxon among these races is *S. bicolor* (Menz *et al*., 2002). Ethiopia is the home for the four basic races (bicolor, caudatum, durra, and guinea) except Kaffir (Yilma Kebede, 1991).

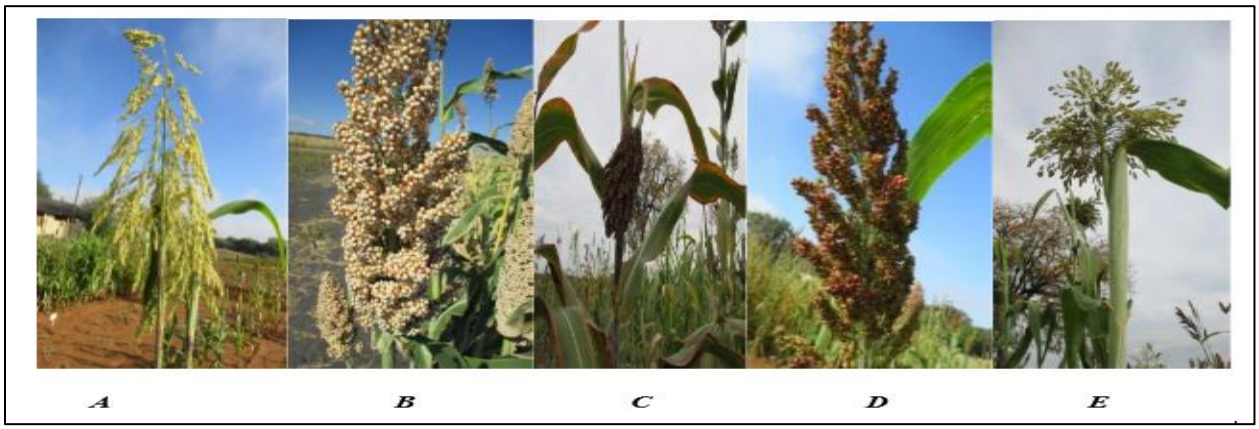


Figure 1. Sorghum basic races by their basic spikelet types. Where; Guinea (A), Caudatum (B), Durra (C), Kafir (D), and Bicolor (E) based on Harlan and De Wet (1972). (Photo by: Motlhaodi *et al.*, 2017).

2.3. Agroecology of sorghum

Sorghum grows in a variety of agro-ecologies, particularly in drought-prone areas where other crops fail to survive and food insecurity is very common (Asfaw Adugna, 2007). It is a warm-season crop, and needs a high temperature for adequate germination and growth. As reported by Mundia *et al*. (2019), an average temperature between 24 ℃ and 27 ℃ is ideal for best grain yields after germination. Sorghum requires 400 to 800 mm of annual rainfall (Ng'uni *et al*., 2011). A positive attribute of sorghum plants is that it is adapted to a variety of soil types such as deep, well-drained, and fertile soil for optimal growth (Abera Deresa *et al*., 2005), however, it grows poorly on sandy soils. For the reason that it is more alkaline salt tolerant than other grain crops, it can successfully grow with a pH between 5.5 and 8. Tolerance and adaptability to drought, high temperatures, and salinity make it become one of the best crops that can tolerate climatic change (ICRISAT, 2015).

2.4. Sorghum production, constraints, and utilization

The production areas of sorghum were divided into three adaptation zones, namely lowlands (less than1600 m.a.s.l), intermediate (1600–1900 m.a.s.l), and highland (greater than1900 m.a.s.l) (Brhane Gebrekidan and Belaynesh Gebrehiwot, 1981). Global sorghum production is expected around 57,893,378 tons yearly (FAOSTAT, 2019). In Africa and Asia, sorghum is produced as a principal food for millions of people (Gebisa Ejeta and Grenier, 2005).

Africa has become the main producer of sorghum with around 49.4% (28,619,588 tons) of total production annually (FAOSTAT, 2019), which is followed by America, Asia, and Europe with a total production of 32.6% (18,875,728 tons), 13.6 % (7,862,817 tons) and, 2.4 % (1,369,774 tons) in the year 2019, respectively. Ethiopia is the third top global producer of sorghum with 5,265,580 tons next to Nigeria (6,665,000 tons) and the United States of America (8,673,480 tons) (FAOSTAT, 2021) (Table 2). It is grown in 13 of the 18 most important agro-ecological zones of the country (Gebeyehu Geremew *et al*., 2004). Oromia, Amhara, and Tigray are the top three sorghum producer regions in Ethiopia (Table 3).

Table 2. Top 10 sorghum producer countries in the world (Source: FAOSTAT, 2021)

Countries	Production (million tons)	Area harvested (ha)
USA	8,673,480	1,891,930
Nigeria	6,665,000	5,397,024
Ethiopia	5,265,580	1,828,182
Mexico	4,352,947	1,324,783
Sudan	3,714,000	6,827,520
China	3,600,000	750,000
India	3,475,410	4,093,290
Brazil	2,672,245	818,301
Niger	1,896,638	3,747,664
Burkina Faso	1,871,791	1,890,172

Table 3. Sorghum producer regions in Ethiopia (CSA, 2020)

Region	Area (ha)	Production (quintals)
Oromia	714,492.91	21,626,963.10
Amhara	641,613.53	18,146,464.95
Tigray	232,636.49	7,008,560.77
SNNP	105,255.96	2,849,141.51
Somali	56,365.24	942,073.45
Benishangul Gumuz	52,005.79	1,540,894.15
Dire Dawa	11,328.24	227,764.61
Harari	8,030.19	176,684.64
Gambella	3,689.05	82,773.67
Afar	2,765.09	54,479.74

*SNNP = Southern Nation Nationality and Peoples.

Sorghum production is influenced by biotic and abiotic factors that can result in economic losses around the world. Some biotic constraints are grain mold, charcoal rot, downy mildew (*Peronosclerspora sorghi*), anthracnose, rust, insect shoot fly, leaf blight, insect stem borers, quelea birds, *Striga hermonthica* spp. and several other viral diseases cause significant economic losses (Wortmann *et al.,* 2006; Thakur *et al*., 2011). *Striga* is one of the most significant obstacles to sorghum production in Africa, causing 10% to 100% yield losses in many countries (Srinivasa *et al*., 2014). In Ethiopia and Sudan, it can cause 65% to 100% losses of the sorghum yield (Gebisa Ejeta *et al*., 2002). Drought, salinity, high temperature, flooding, heat, low soil fertility, oxidative stress, heavy metal toxicity, and other abiotic stresses can all have an impact on sorghum production (Bray *et al*., 2000). Yet, the crop is produced for numerous kinds of utilizations in different parts of the world.

The production of Sorghum grain is usually utilized for food, and the stalks, are useful for building materials and fodders, and its grain or sweet stem is generally utilized as animal feed in developed countries, and it is used to make sugar, syrup, and molasses (Doggett and Prasada, 1995; Duncan, 1996). Sorghum is an important source of energy, protein, vitamins, and minerals in the semi-arid tropics (Duodu *et al*., 2003). It is a good source of energy consisting of nearly 70% of carbohydrates and 12% of proteins of grain's dry weight (Ng'uni *et al*., 2012).

In Ethiopia, sorghum is a food crop that is a staple local diet, which is produced not only for its grains but also for its use as a source of animal feed, fuel-wood, and as a construction material. After teff, sorghum is the preferred crop for making Ethiopian cultural and indigenous food called injera, local beverages (*tela* and *areke*) (Adamu Zegeye, 1997). The stalk is used as fuelwood, construction material, and feed to animals. It has also received a lot of attention as a result of its newer use as a 'biofuel feedstock' (Zegada-Lizarazu and Mont, 2012).

2.5. Genetic diversity of sorghum

Genetic diversity is defined as the extent to which heritable materials differ within a group of plants and it is important in understanding the genetic variability that is available for use in breeding programs. It is also essential for the development of improved cultivars, as well as the survival of the species. Comprehending, accessing, and utilizing the diversity available in domestic and wild relatives are essential in expanding the genetic base of varieties to maintain improvement

(Singh, 2005). Crop genetic diversity is important for crop species analysis, which is an important part of crop improvement programs as it provides information about genetic diversity and helps understand the available biological variability and its potential use in breeding programs (Albert, 2014). Additionally, this genetic diversity is important to develop appropriate sampling techniques for the collection and preservation of germplasm (Ramanatha and Hodgkin, 2002).

In various parts of the world, vast collections of sorghum germplasm (approximately 168, 500 accessions) were kept (Billot *et al*., 2013). The northeastern quadrant of Africa, which stretches from Ethiopia through Sudan to Chad, is the center of diversity for cultivated and wild sorghum (Mann *et al*., 1983). Consequently, Sorghum has a high genetic diversity in East Africa, especially in Sudan and Ethiopia (Brhane Gebrekidan, 1982). Ethiopia is the highest contributor of sorghum genetic resources for the world and a potential source of novel genes (Haussmann *et al*., 2002). Genetic diversity in sorghum landraces and core collections based on race, the latitude of origin, seed quality, and DNA markers have shown that sorghum has a significant polymorphism (Wu *et al*., 2004). The genetic variation can be measured with different genetic marker systems.

2.6. Marker systems used in crop genetic diversity study

Biological markers can be of many different types but are mainly divided into morphological, biochemical, and molecular markers (Govindaraj *et al.*, 2015). Each type of marker system has advantages and disadvantages. The type of marker is chosen based on species, availability, and a cost-benefit analysis (Kassa Semagn *et al*., 2006).

2.6.1. Morphological markers

Morphological markers were the first classical genetic markers based on phenotypic traits that are visible to human eyes and used to assess variation with highly distinguishable traits such as flower color, seed shape, growth habits, pigmentation, and the like. Due to genotype-by-environment interactions, morphological markers have complex inheritance, late expression, lower polymorphism, and vulnerability to environmental influences (Smith and Smith, 1992). However, this does not imply that any of the biochemical, molecular, or both markers replace morphological markers; rather, they are used in tandem (Tadele Temesgen, 2014).

2.6.2. Biochemical markers

To overcome the limitations of morphological markers, protein, and DNA-based markers were developed. Protein markers (isozyme) are commonly referred to as biochemical markers. They are based on protein polymorphisms obtained by electrophoretic separation (Tadele Temesgen, 2014). Protein markers have several advantages, including co-dominant inheritance, little environmental influence, and a low assay cost. Their major drawbacks include lower polymorphism, lower genome coverage, technical difficulties, the need for fresh tissue, and the requirement for specific developmental stages for certain enzymes (Tadele Temesgen, 2014).

2.6.3. Molecular markers

According to Tabbasam (2006), molecular markers are DNA sequences that can be found at certain points in the entire genome that are passed down from generation to generation as per the laws of inheritance.

DNA markers proved to be more powerful and reliable tools for detecting variation within crop genotypes than biochemical markers because they are relatively simple to identify, abundant in the entire genome, independent of environmental conditions, and detectable in every stage of plant development (Gepts, 1993; Gupta *et al*., 2000). A desirable DNA marker would be co-dominant, evenly distributed across the genome, and highly reproducible (Mondini *et al*., 2009). To characterize and estimate the genetic diversity of sorghum, various types of molecular markers, such as (RFLP, RAPD, AFLP, ISSR, SSR, SNP, etc.) have been developed.

The DNA-based marker systems are generally classified as non-polymerase chain reaction (non-PCR) (hybridization-based) markers and PCR based markers. Non-PCR hybridization-based marker techniques were used relatively early in plant research, which combined the use of restriction endonucleases. The hybridization-based method includes RFLP (Kumar *et al*., 2009).

PCR based markers use thermocycler for amplification of the fragments of DNA sequence. PCR is a molecular biological technique that amplifies a single or a few copies of DNA fragment by several orders of magnitude and produces thousands to million copies of specific DNA sequence, which can be used to study the variation within the DNA sequence without the involvement of a living organism. PCR-based markers require less DNA per assay and have higher throughput than

RFLP markers (Molosiwa, 2012). PCR based molecular markers include RAPD, AFLP, ISSR, Microsatellites, SNPs, etc. (Kumar *et al*., 2009).

2.6.3.1. Microsatellites

Microsatellites or SSRs, are polymorphic loci found in DNA made up of repeating units ranging in length from one to six base pairs (Bidichandani *et al*., 1998). Tandem repeating motifs of 1-6 nucleotides are common in genomes of most taxa (Beckmann and Weber, 1992). These repetitions are widely distributed along the genomes of plants and animals, resulting in a high genetic variation due to differences in the number of repeating units of a locus (Jonah *et al*., 2011).

Microsatellites have several advantages over other types of molecular markers; for example, they are evenly distributed throughout the genome and co-dominant (Farooq and Azam, 2002), highly repeatable, inexpensive, simple to run, and require a small amount of medium quality DNA (Ahmed, 2012). Due to the features listed above, microsatellites have become genetic markers of choice in many plants stems (Tharachand *et al*., 2012). These markers have several disadvantages: costly, laborious, and time-consuming to develop (Kumar *et al*., 2009).

3. MATERIALS AND METHODS

3.1. Plant materials

In this study, 77 sorghum landraces were used for diversity analyses, among them, 48 were obtained from the Ethiopian Biodiversity Institute (EBI) and 29 landraces were newly collected (coded = SB) from different parts of the country (Appendix 1). They were collected from five sorghum-growing regions of the country including 22 from Oromia [East Hararge (8), West Hararge (7), and East Wellega(8)], 25 from Amhara [North Wollo (8), South Wollo (7), and North Shewa(10)], 16 from Tigray [South Tigray (10) and Central Tigray (6)], 6 from Benishangul-Gumuz [Metekel (6)], and 8 from Southern Nations and Nationalities and Peoples (SNNP) [Benchi-Maji (8)] (Figure 2). Seeds of the 77 sorghum landraces were grown in a greenhouse and fresh leaves were collected from two-weeks-old plants for genomic DNA extraction.

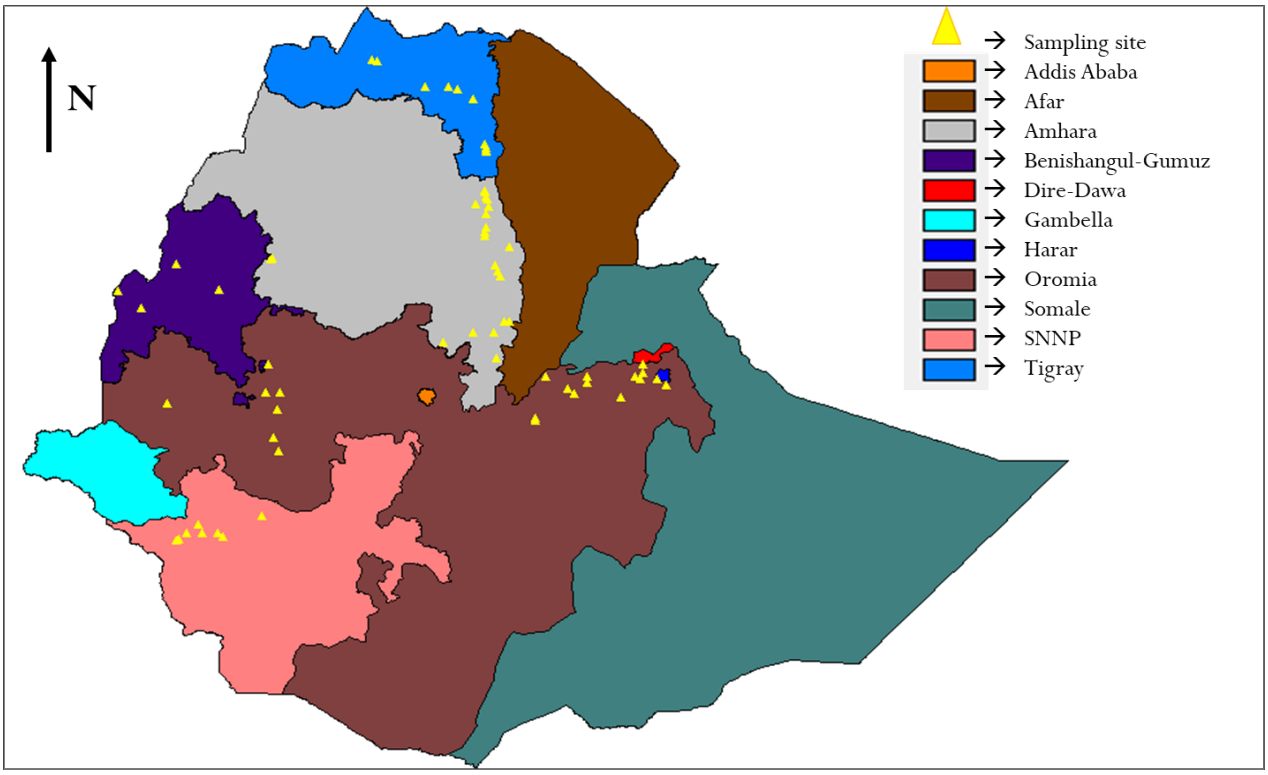


Figure 2. Map of Ethiopia illustrating the collection region and sites of sorghum landraces

3.2. DNA extraction

Genomic DNA extraction and downstream activities were carried out at the Plant genetic research laboratory of the Institute of Biotechnology, Addis Ababa University. Approximately equal amounts of 2-week-old single leaf samples were collected in zip-lock plastic bags containing silica gel (Appendix 4). About 100 mg of each sample were Pulverised by Geno-grinder (mixer-miller) machine at 30 RPM for 4 min. Genomic DNA extraction was carried out using modified Cetyl trimethyl ammonium bromide (CTAB) protocol (Govers, 2005) (Appendix 5).

The DNA quality and concentration were checked by electrophoresis and Nanodrop (NanoDrop ™ 2000/2000c) spectrophotometer (Thermo Scientific ™ USA). A 1% agarose gel (100 ml, 1x TAE, and 1g agarose, stained with 2 µl Ethidium bromide) was prepared and 4 µl of each genomic DNA sample mixed with a 2 µl of loading dye was loaded and electrophoresed at a constant voltage of 85 V for 45 minutes. The gel picture was taken under UV transilluminator by Biodoc Analyses 2.0 with a digital canon camera. Then, the final concentration of each genomic DNA sample was adjusted (diluted) to 100 ng μL^{-1} using nuclease-free sterile (double distilled) water for use in PCR amplification and stored at -20 °C until use.

3.3. Genotyping

The genotyping assay involves simple sequence repeat (microsatellite) markers. The primers' polymorphism and reproducibility were checked on six randomly selected sorghum genotypes. In this study, 11 polymorphic microsatellite markers were used to assess the genetic structure of 77 sorghum landraces. The microsatellite markers were retrieved from previous reports of related works (Ng'uni *et al*., 2012; Tesfaye Disasa, 2016; Alemu Tirfessa *et al*., 2020) (Table 4).

Table 4. Primer's sequence, repeat motif, chromosome number, annealing temperatures, and the range of molecular weight detected using 11 microsatellite markers of sorghum in Ethiopia

Marker	Forward (F) and Reverse (R) primer (5′ to 3′)	Repeat Motif	Chr. No.	Reference	[a]AT (°C)	[b]Fragment Size (Base pair)	
						Min	Max
SbAGB02	F: CTCTGATATGTCGTTGTGCT R: ATAGAGAGGATAGCTTATAG-CTCA	(AG)35	7	x	44	95	126
Xcup74	F: GTCGCCATTGTGATGAAGAG R: CAGTAGTCCAGCAAAACGGC	(TG)9	2	x	56	174	185
Xtxp136	F: GCGAATAGCATCTTACAACA R: ACTGATCATTGGCAGGAC	(GCA)5	5	z	56	238	242
Xcup67	F: GGTCAGTGCTTACACAGATTCC R: GGGGATTGCAGGTGTCATAG	(TA)6	10	x	56	264	294
Xcup14	F: TACATCACAGCAGGGACAGG R: CTGGAAAGCCGAGCAGTATG	(AG)10	3	w	60	209	237
Xgap084	F: CGCTCTCGGGATGAATGA R: TAACGGACCACTAACAAATGATT	(AG)14	2	z	50	176	209
Xtxp010	F: ATACTATCAAGAGGGGAGC R: AGTACTAGCCACACGTCAC	(CT)14	9	z	60	139	148
Xtxp021	F: GAGCTGCCATAGATTTGGTCG R: ACCTCGTCCCACCTTTGTTG	(AG)18	4	z	52	175	190
Xtxp141	F: TGTATGGCCTAGCTTATCT R: CAACAAGCCAACCTAAA	(GA)23	10	y	50	141	182
Xtxp265	F: GTCTACAGGCGTGCAAATAAAA R: TTACCATGCTACCCCTAAAAGTGG	(GAA)19	6	z	56	183	220
Xtxp273	F: GTACCCATTTAAATTGTTTGCAGTAG R: CAGAGGAGGAGGAAGAGAAGG	(TTG)20	8	z	56	198	231

Where; Min = minimum, Max = maximum, Chr. No. = Chromosome Number, AT=Annealing temperature. [a], [b] = resulted in this study. Sources: (Folkertsma *et al.*, 2005, Ng'uni *et al.*, 2012; Tesfaye Disasa, 2016; Alemu Tirfessa, 2020). W is Folkertsma *et al.*, 2005, x is Ng'uni *et al.* (2012), y is Tesfaye Disasa (2016) and z is Alemu Tirfessa (2020) in the reference column.

The PCR was conducted using Biometra 2003 Thermocycler machine. PCR Amplification was carried out in a 20 µl final reaction volume containing, 12.6 µl H_20, 2 µl 10x PCR buffer ($MgCl_2$ free), 1.6 µl $MgCl_2$ (50 mM), 0.5 µl dNTPs (10x), 0.5 µl Forward primer, 0.5 µl Reverse primer, 0.3 µl Taq Polymerase (5 u/µl) and 2 µl template DNA (100 ng μL^{-1}). The PCR steps were programmed as presented in Table 5.

Table 5. The PCR cycles and annealing temperature used to amplify the 11 SSR markers in 77 landraces of sorghum

PCR steps	Temperature in ^{o}C	Primer 1-3	Primer 4-9	Primer 10 and 11
Initial denaturation	95 ^{o}C	5 min	5 min	5 min
Cycle denaturation	95 ^{o}C	2 min	1 min	1 min
Annealing temperature	Vary with the marker used	2 min	1 min	1 min
Cycle extension	72 ^{o}C	2 min	2 min	1 min
Final extension	72 ^{o}C	10 min	10 min	10 min
Number of cycles		40x	35x	44x

The PCR products were stored at 4 ^{o}C until loaded on the gel for electrophoresis. The products were fractionated in 3% agarose gel using 1x TAE buffer at 100 V for 3 hours. The gel was stained in 2 µl ethidium bromide and an 8 µl amplification product of each sample with 2 µl loading dye was loaded on the gel. The gel was, visualized under UV light and subsequently photographed using a BIO-RAD Gel Doc TM EZ Imaging System. A 100 bp DNA ladder was used to estimate the size of the amplified fragments.

3.4. Data scoring and statistical analysis

The sizes of all PCR-amplified microsatellite regions were estimated using Image-lab 6.1 software (Bio-Rad Laboratories, Inc.) with the 100 bp standard ladder. The scored marker data were used to analyze genetic diversity and population structure. The standard indices of genetic diversity were calculated using various statistical software packages. GenAlEx 6.5 was used to calculate the mean effective number of alleles (Ne), observed number of alleles (Na), observed heterozygosity (Ho), expected heterozygosity (He), fixation index (F), Shannon information index (I), and gene flow (Nm) for each SSR marker and population (Peakall and Smouse, 2012). The number of private alleles (NPA), number of locally common alleles, and percentage of polymorphic loci (PPL) of the population and regions were also calculated using GenAlEx 6.5. PowerMarker was used to calculate the gene diversity (h), major allelic frequency (MAF), and the polymorphic information content (PIC) (Liu and Muse, 2005). According to Vaiman *et al*. (1994), the PIC value less than 0.25, between 0.25 and 0.5, and higher than 0.5 is considered to be slightly informative, moderately informative, and highly informative, respectively. Arlequin ver. 3.5.2.2 was used for the analysis of Molecular Variance (AMOVA) within and among populations (Excoffier and Lischer, 2010). The principal coordinate analysis (PCoA) was carried out using GenAlEx 6.5. The same software was used to estimate pairwise population genetic distances and gene flow (Nm). Based on Nei's genetic distance (Nei, 1983), the unweighted Neighbour-Joining clustering of individuals, populations, and regions was performed using PowerMarker v 3.25 (Liu and Muse, 2005). The resulting trees were visualized by using MEGA-X (Kumar *et al*., 2018). Population structure and mixing patterns were determined with the STRUCTURE software ver. 2.3.4 with a Bayesian model-based clustering algorithm (Pritchard *et al.*, 2000). To estimate the actual number of population clusters (K), a burn time of 100,000 was used on each run, and data was collected on 250,000 replicas from Markov Chain Monte Carlo (MCMC) for K = 1 to K = 10 using 20 iterations for each K. The optimal value of K was determined according to the simulation method of Evanno *et al.* (2005) using the web-based STRUCTURE HARVESTER ver. 0.6.92 (Earl & vonHoldt, 2012). A bar chart for the optimal K was determined with the beta version of Clumpak (Kopelman *et al*., 2015).

4. RESULTS

4.1. Microsatellite markers based polymorphism

The 11 microsatellite markers resulted in a total of 49 alleles (Table 6). The number of alleles per locus ranged from three (Xtxp136 and Xcup14) to six (SbAGB02 and Xtxp265) with an overall mean of 4.5. All microsatellite markers were polymorphic, for example, as presented on the gel image of a microsatellite locus (Figure 3).

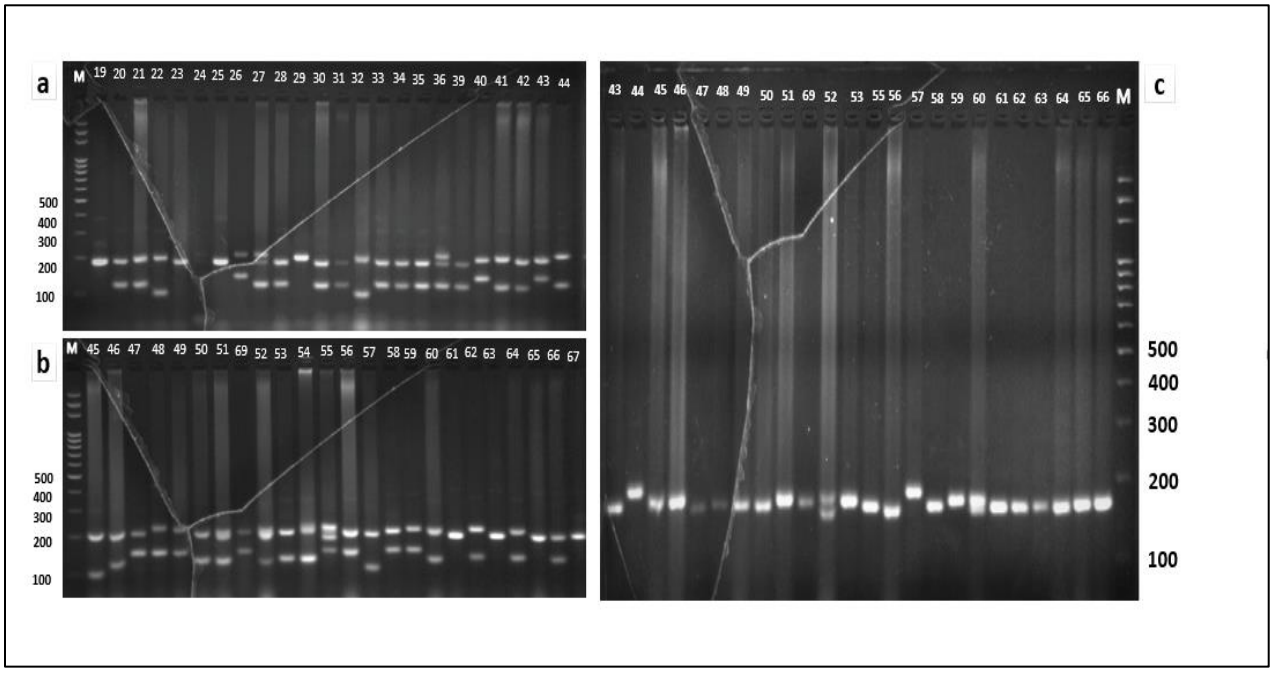


Figure 3. Generated electrophoretic bands of SSR markers of sorghum landraces collected from different parts of Ethiopia. Where; (A: Xtxp265 with samples 19 - 44, B: Xtxp265 with samples 45 - 67 and C: Xtxp021 with samples 43 - 66), M = marker (100 bp DNA ladder), and the numbers in lanes indicates the landraces.

Across the entire population, major allele frequencies ranged from 0.38 for locus SbAGB02 to 0.86 for Xtxp136 with an overall mean of 0.59. The number of effective alleles ranged from 1.1 (Xcup67 and Xcup14) to 3.19 for the locus Xtxp265 with an overall mean of 1.84. The locus gene diversity ranged from 0.25 for Xtxp136 to 0.72 for the locus SbAGB02 with an overall mean of 0.53 (Table 6).

The polymorphic information content of the loci ranged from 0.23 (Xtxp136) to 0.67 (SbAGB02 and Xtxp265) with an overall mean of 0.48. Out of 11 loci, six of the loci (SbAGB02, Xgap084, Xtxp021, Xtxp141, Xtxp265, and Xtxp273) were found to be highly informative (PIC ≥ 0.5). The microsatellite loci (Xcup74, Xcup67, Xcup14 and Xtxp010 were detected to be moderately informative (0.25 ≤ PIC < 0.5) while one locus (Xtxp136) was found to be less informative (PIC < 0.25) (Table 6).

Table 6. Summary of genetic diversity indices for 11 microsatellite markers across seventy-seven sorghum landraces

Locus	MAF	Na	Ne	H	I	Ho	He	Fst	Nm	PIC
SbAGB02	0.38	6	2.26	0.72	0.91	0.05	0.54	0.12	1.8	0.67
Xcup74	0.53	4	1.58	0.57	0.49	0	0.32	0.26	0.71	0.49
Xtxp136	0.86	3	1.24	0.25	0.22	0	0.14	0.2	1.02	0.23
Xcup67	0.78	4	1.1	0.36	0.17	0.03	0.08	0.05	5.3	0.33
Xcup14	0.79	3	1.1	0.34	0.13	0	0.07	0.1	2.17	0.3
Xgap084	0.47	5	2.16	0.62	0.8	0.04	0.49	0.19	1.06	0.55
Xtxp010	0.81	4	1.48	0.34	0.4	0	0.23	0.2	1.03	0.32
Xtxp021	0.48	5	2.21	0.61	0.83	0.07	0.49	0.19	1.05	0.53
Xtxp141	0.48	5	2.07	0.66	0.77	0.02	0.42	0.35	0.47	0.6
Xtxp265	0.42	6	3.19	0.71	1.27	0.85	0.68	0.06	3.8	0.67
Xtxp273	0.52	4	1.87	0.65	0.73	0	0.45	0.16	1.3	0.6
Mean	0.59	4.5	1.84	0.53	0.61	0.1	0.36	0.17	1.79	0.48

Key: MAF-Major allelic frequency, Na-Number of alleles, Ne- Number of effective alleles, I-Shannon's Information Index, Ho= observed heterozygosity, He- Expected heterozygosity, h- Gene diversity, Nm = Gene flow estimate from Nm = [(1 / Fst) - 1] / 4, Fst = Inbreeding coefficient within subpopulations relative to total (genetic differentiation among subpopulations), PIC- polymorphic information content.

4.2. Genetic diversity indices among populations and regions

A summary of the different genetic diversity estimates overall loci across regions and populations are presented in Table 7 and Table 8, respectively. The numbers of alleles per locus (Na) within

the regions ranged from 2.18 (Tigray) to 2.73 (Amhara), with an average of 2.51 alleles per locus (Table 7). The same diversity index across populations ranged from 1.45 (South Tigray and South Wollo) to 2.64 (Benchi-Maji and North Shewa), with an overall mean of 2.02 alleles per locus (Table 8).

The number of effective alleles (Ne) within regions and populations ranged from 1.37 (Tigray) to 2.24 (SNNP) (Table 7), and 1.18 (South Tigray) to 2.24 (Benchi-Maji) (Table 8), respectively. The Shannon diversity index (I) ranged from 0.36 (Tigray) to 0.76 (SNNP) for the regions (Table 7), and 0.18 (South Tigray) to 0.76 (Benchi-Maji) for the populations (Table 8).

The percentage of polymorphic loci (PPL) for the regions and the populations ranged from 64% (Tigray) to 91% (Amhara) (Table 7) and 37% (NWL, STG, SWL) to 91% (NSH) (Table 8), respectively. Except for Oromia regions where the number of private alleles (NPA = 0.00), all others have the same number of private alleles (NPA =0.09) (Table 7). From ten populations Central Tigray, Benchi-Maji, Metekel, and North shewa showed an equal number of private alleles (NPA = 0.09) with a mean value of 0.04 (Table 8). Likewise, the number of locally common alleles found in regions and populations ranged from 0.18 (Tigray) to 0.36 (Amhara and SNNP) for regions (Table 7), while it becomes 0.09 (South Tigray) to 0.82 (for Benji-Maji and Metekel) for population (Table 8).

Single private alleles were observed for Amhara and Tigray, SNNP, and Benishangul-Gumuz on locus Xtxp265, Xtxp010, and SbAGB02, respectively. Similarly, single private alleles were also observed for the population of Central Tigray on locus Xtxp265, Benchi-Maji on locus Xgap084, Metekel on locus SbAGB02, and North Shewa on locus Xtxp265.

Table 7. Allelic patterns and diversity indices across the regions averaged over the 11 SSR loci

Region	N	Na	Ne	I	Ho	He	F	PPL	NPA	NLCA
Amhara	25	2.73	1.9	0.67	0.05	0.39	0.91	91%	0.09	0.36
Tigray	16	2.18	1.37	0.36	0.09	0.2	0.62	64%	0.09	0.18
Oromia	22	2.64	1.78	0.62	0.11	0.36	0.76	82%	0	0.27
SNNP	8	2.64	2.24	0.76	0.1	0.45	0.82	82%	0.09	0.36
Benishangul-Gumuz	6	2.36	1.93	0.64	0.12	0.38	0.75	73%	0.09	0.27
Mean	15.5	2.51	1.84	0.61	0.1	0.36	0.78	78%	0.07	0.29

Key: N=Number of samples, Na = Observed number of alleles, Ne = Number of effective alleles, NPA = Number of Private Alleles (i.e., the number of alleles unique to a single population), I = Shannon's information statistic, Ho= observed heterozygosity, He= expected heterozygosity, F= Fixation index, PPL= Percentage of polymorphic loci, and NLCA = Number of Locally Common Alleles found in populations. N.B., SNNP = Southern Nations, Nationalities, and People's Region.

Table 8. Allelic patterns and diversity indices across populations averaged over the 11 SSR loci

Pop	N	Na	Ne	I	Ho	He	F	PPL	NPA	NLCA
NWL	10	1.55	1.3	0.26	0.07	0.16	0.65	37%	0	0.18
STG	10	1.45	1.18	0.18	0.08	0.1	0.6	37%	0	0.09
CTG	6	2.09	1.52	0.45	0.09	0.26	0.6	64%	0.09	0.55
EH	7	2.09	1.76	0.5	0.13	0.3	0.53	64%	0	0.55
WH	7	1.91	1.61	0.44	0.12	0.27	0.57	64%	0	0.36
EWL	8	2	1.67	0.47	0.09	0.28	0.81	64%	0	0.27
BM	6	2.64	2.24	0.76	0.1	0.45	0.82	82%	0.09	0.82
MET	6	2.36	1.93	0.64	0.12	0.38	0.75	73%	0.09	0.82
SWL	7	1.45	1.25	0.23	0.04	0.15	0.76	37%	0	0.36
NSH	10	2.64	2.11	0.72	0.06	0.42	0.89	91%	0.09	0.82
Mean	7.7	2.02	1.66	0.46	0.09	0.28	0.71	60.91%	0.04	0.48

Key: N=Number of samples, Na = Observed number of alleles, Ne = Number of effective alleles, NPA = Number of Private Alleles (i.e., the number of alleles unique to a single population), I = Shannon's information statistic, Ho= observed heterozygosity, He= expected heterozygosity, F= Fixation index, PPL= Percentage of polymorphic loci, and NLCA = Number of Locally Common Alleles found in populations. N.B., NW=North Wollo, ST=South Tigray, CT=Central Tigray, EH= East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SW=South Wollo, and NSH=North Shewa.

4.3. Genetic Distance, differentiation, and gene flow between regions and populations

The existing genetic distance between populations was estimated by Pairwise Nei's genetic distance and gene flow (Nm) values between the studied regions and population. The highest measure of genetic distance (0.17) with the lowest gene flow (Nm=1.23) was observed between Tigray and SNNP regions. The lowest measure of genetic distance (0.05) with the highest gene flow (Nm=19.26) was observed between Amhara and SNNP, which was followed by the genetic distance of Oromia and Amhara (0.06), regions with a gene flow (Nm= 6.32) (Table 9). Moreover, the population-wise genetic distance was the highest when comparing South Tigray with Metekel (0.59) having a gene flow of (4.37), and the lowest when South Tigray paired with North Wollo (0.02) with a gene flow of (27.77). The largest gene flow for the population was observed between North Shewa and Benchi-Maji (50.76) and the lowest gene flow was observed between possible pairs of South Tigray and South Wollo, and South Tigray and Metekel (Table 10).

Table 9. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among the five regions of Ethiopia

Pairwise	Amhara	Tigray	Oromia	SNNP	Benishangul-Gumuz
Amhara	-	2.37	6.32	19.26	2.8
Tigray	0.08	-	1.62	1.23	0.63
Oromia	0.06	0.08	-	6.84	1.67
SNNP	0.05	0.17	0.1	-	4.85
Benishangul-Gumuz	0.12	0.3	0.16	0.13	-

Key: SNNP = Southern Nations, Nationalities, and People's Region.

Table 10. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among the ten populations in Ethiopia

pairwise	NWL	STG	CTG	EH	WH	EWL	BM	MET	SWL	NSH
NWL	-	27.77	8.68	12.06	30.91	8.08	17.25	10.44	9.16	4.27
STG	0.02	-	4.15	6.38	9.50	5.11	4.85	4.37	3.19	3.46
CTG	0.05	0.10	-	6.18	46.67	5.77	6.33	6.58	4.46	8.61
EH	0.11	0.19	0.13	-	16.16	35.38	5.54	7.39	4.59	4.45
WH	0.23	0.24	0.13	0.17	-	6.18	19.19	24.84	12.75	5.94
EWL	0.13	0.15	0.23	0.16	0.15	-	9.14	4.17	13.44	3.66
BM	0.20	0.27	0.22	0.16	0.13	0.11	-	6.38	6.38	50.76
MET	0.36	0.59	0.4	0.34	0.33	0.3	0.20	-	4.46	7.39
SWL	0.50	0.57	0.56	0.32	0.44	0.45	0.23	0.33	-	5.11
NSH	0.12	0.21	0.14	0.10	0.11	0.07	0.03	0.19	0.36	-

Key: NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SWL=South Wollo, NSH=North Shewa.

The pair-wise genetic differentiation between the regions ranged from 0.01 to 0.28. The highest population differentiation was observed among Tigray and Benishangul-Gumuz (Fst=0.28), followed by Tigray and SNNP (Fst=0.17). The lowest population differentiation (Fst=0.01) was observed among SNNP and Amhara, which was followed by the population differentiation of Oromia and Amhara (Fst=0.04) and Oromia and SNNP (Fst=0.04), regions, respectively (Table 11). Furthermore, the population-wise genetic differentiation ranged from 0 to 0.08. The highest population differentiation was observed among North Shewa and South Tigray (Fst=0.16). However, most possible pairs of populations showed low to moderate population differentiation (Table 12).

The genetic differentiation was low to moderate between possible pairs of regions except for the combination of Tigray with both SNNP and Benishangul-Gumuz (Table 11). All possible pairs of populations showed low genetic differentiation (Table 12).

Table 11. Estimate of genetic differentiation measured by Fst (below the diagonal) among five regions

Pairwise	Amhara	Tigray	Oromia	SNNP	Benishangul-Gumuz
Amhara	0.00				
Tigray	0.10	0.00			
Oromia	0.04	0.13	0.00		
SNNP	0.01	0.17	0.04	0.00	
Benishangul-Gumuz	0.08	0.28	0.13	0.05	0.00

Key: SNNP = Southern Nations, Nationalities, and People's Region.

Table 12.Estimate of genetic differentiation measured by Fst (below the diagonal) among ten populations

pairwise	NWL	STG	CTG	EH	WH	EWL	BM	MET	SWL	NSH
NWL	0.00									
STG	0.01	0.00								
CTG	0.03	0.03	0.00							
EH	0.02	0.04	0.04	0.00						
WH	0.01	0.03	0.01	0.02	0.00					
EWL	0.03	0.05	0.04	0.01	0.03	0.00				
BM	0.01	0.05	0.04	0.04	0.01	0.03	0.00			
MET	0.02	0.05	0.04	0.03	0.01	0.05	0.04	0.00		
SWL	0.03	0.07	0.05	0.05	0.02	0.02	0.04	0.05	0.00	
NSH	0.06	0.08	0.03	0.05	0.04	0.06	0.01	0.04	0.06	0.00

Key: NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SWL=South Wollo, NSH=North Shewa.

4.4. Analysis of molecular variance

Analysis of molecular variance (AMOVA) was computed with the populations and regions by using collection regions and zones as a grouping criterion. The AMOVA analysis significantly partitioned the total genetic variation between regions and populations that showed 20.78% of the genetic variation resulted from the difference among populations, while variation within the population has contributed 79.22% of the total genetic variability observed. The genetic variation among regions and within regions accounted for 17.79% and 82.21%, respectively. The overall genetic differentiation coefficient among the populations and regions was Fst = 0.21, 0.18 with gene flow of 0.95and 1.16, respectively (Table 13).

Table 13.Analysis of molecular variance (AMOVA) showing the partitioning of genetic variation within and among populations, and regions of Ethiopian sorghum landraces

CO	SV	DF	SS	VC	PV (%)	FI	P-value
population	Among population	9	22.95	0.13	20.78	F_{ST}: 0.21	0.00000***
	within population	144	73.09	0.51	79.22		
	Nm	0.95					
	Total	153	96.04	0.64	100%		
Regions	Among regions	4	15.69	0.12	17.79	F_{ST}:0.18	0.00000***
	within regions	149	80.35	0.54	82.21		
	Nm	1.16					
	Total	153	96.04	0.66	100%		

Key: Fst = the coefficient of genetic differentiation, where AP = the variance among populations, calculated as= P – 1 and WP = the variance within populations calculated as=*2N – P*, Total =2N – 1, Gene flow (Nm) calculated as [(1/Fst) −1]/4; among population and regions. CO=components; SV=source of variation; DF=degrees of freedom; SS=sum of squares; VC=variance component; PV= percentage of variation; FI=fixation index.

4.5. Cluster analysis

The Neighbor-Joining- based clustering of the 77 individuals belonging to five regions and ten populations identified three major clusters (C1, C2, and C3). The first (C1), second (C2), and the third (C3) clusters consisted of 13%, 61%, and 26%, from different populations respectively (Figure 4).

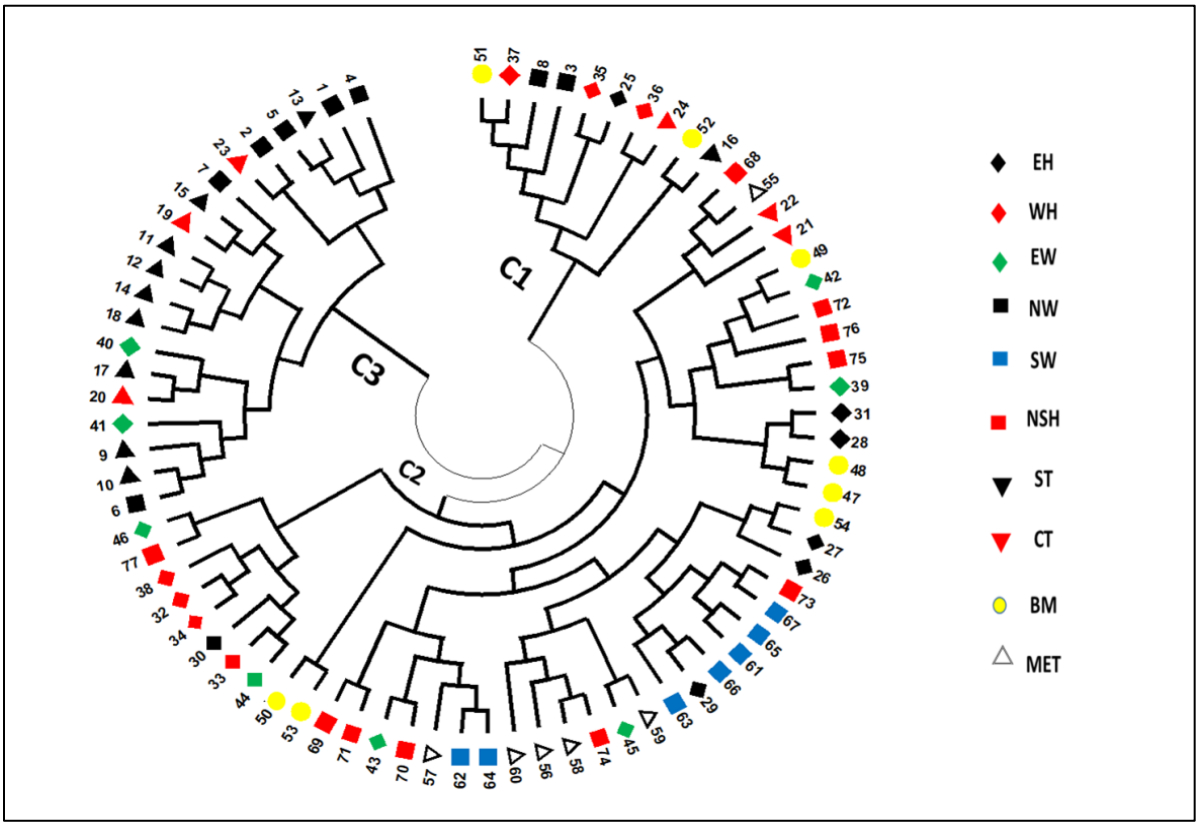


Figure 4. Neighbor-Joining tree showing relatedness among the 77 sorghum landraces of regions and populations. Key: C1, C2, and C3 are Cluster 1, 2, and 3, respectively, NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SWL=South Wollo, NSH=North Shewa, SNNP = Southern Nations, Nationalities, and People's Region (Amhara= NWL, SWL, NSH), (Tigray=STG, CTG), (Oromia= EH, WH, EWL), (SNNP=BM), (Benishangul Gumuz= MET).

Table 14. Number of genotypes contributed by populations to the clusters

Region	Population	Size	Member of		
			C1	C2	C3
Amhara	North Wollo	8	2	0	6
	South Wollo	7	0	7	0
	North Shewa	10	0	10	0
Tigray	South Tigray	10	1	0	9
	Central Tigray	6	1	2	3
Oromia	East Hararge	7	1	6	0
	West Hararge	7	3	4	0
	East Wellega	8	0	6	2
SNNP	Benchi-Maji	8	2	6	0
Benishangul-Gumuz	Metekel	6	0	6	0
Total		77	10	47	20

Key: C1, C2, and C3 are Cluster 1, 2, and 3, respectively. SNNP= Southern Nations, Nationalities, and People's Region.

The Neighbor-Joining- based clustering grouped both the regions and the populations into three major clusters (I, II, and III) (Figure 5). The first cluster (I) is composed of one region Amhara, the second cluster (II) is composed of Oromia and Tigray regions while Benishangul Gumuz and SNNP (Southern Nations, Nationalities regions) constituted the third cluster (III). In Population-wise clustering, only the South Wello population made the first cluster (I) while Metekel, North Shewa, and Benchi-Maji constituted the third cluster (III) and the remaining population constructed the second cluster (II) (Figure 5).

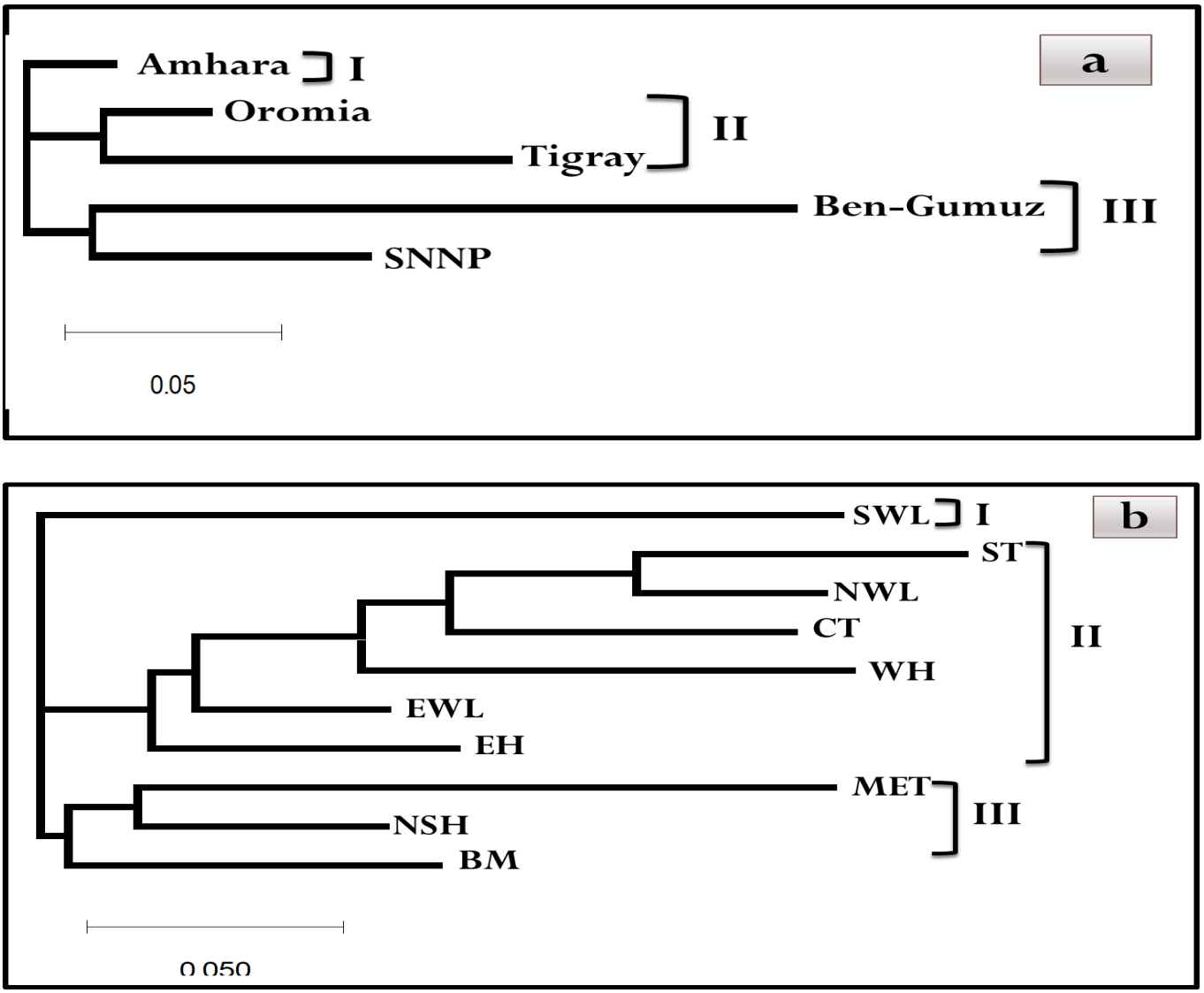


Figure 5. The Neighbor-Joining- based clustering showing genetic relationships among regions (a) and populations (b). Key: NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SWL=South Wollo, NSH=North Shewa, B-Gumuz=Benishangul-Gumuz, SNNP = Southern Nations, Nationalities, and People's Region.

4.6. Principal coordinates analysis

Principal coordinate analysis (PCoA) of 77 sorghum landraces showed almost a uniform distribution around the center of the three-dimensional coordinate plane and resulted in a scatter plot with poor population clustering. The first three principal coordinate axes explained 12.96%,

11.70%, and 9.88% of the variation for the region and 12.96%, 11.70%, and 9.88% of the variation for populations, respectively (Figure 6).

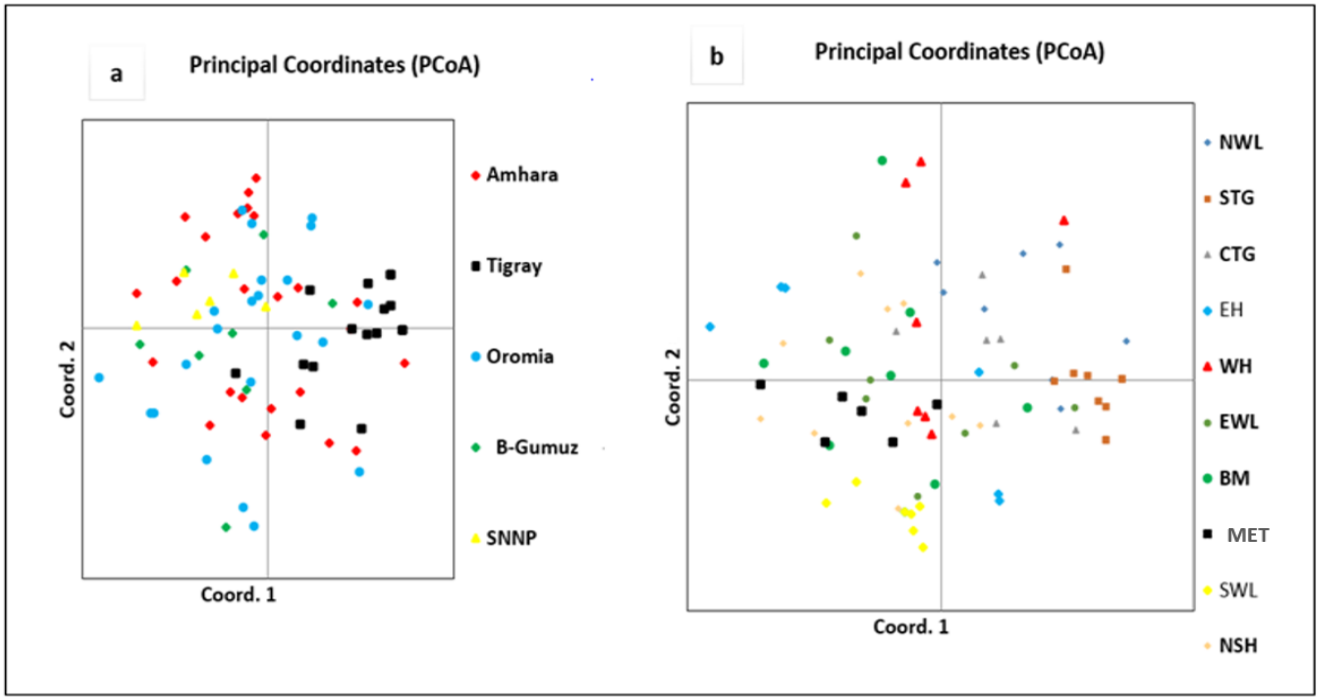


Figure 6. Principal coordinate analysis (PCoA) of 77 Ethiopian sorghum landraces based on regions (a) and populations (b) using 11 SSR markers. Key: NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM= Benchi-Maji, MET= Metekel, SWL= SouthWollo, NSH=North Shewa, B-Gumuz=Benishangul-Gumuz, SNNP = Southern Nations, Nationalities, and People's Region. Samples coded with the same symbol and colors belong to the same population.

4.7. Population structure

The output of the Structure harvester revealed that the delta K (ΔK) values reached a sharp peak at K = 2 (Figure 7, Appendix 11) indicating the presence of two major subgroups (Evanno *et al*., 2005). The Clumpak result (bar plot) detected a greater degree of genetic admixture between the studied populations and regions (Figure 8). It also indicated the absence of clear- geographical origin-based structuring in the sorghum populations.

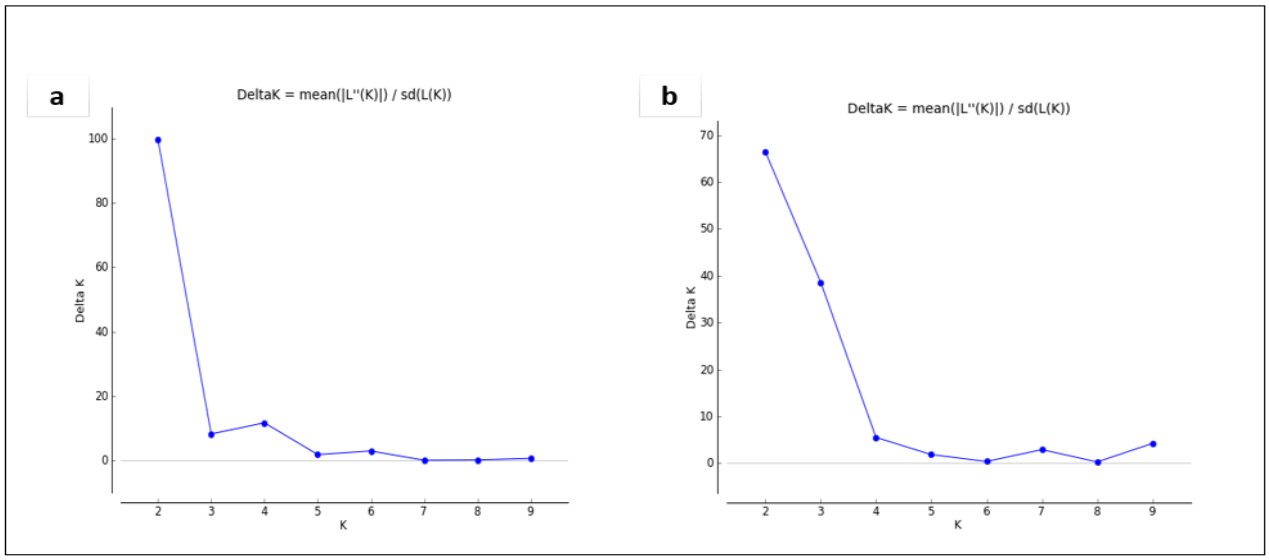


Figure 7. Population structure of 77 Ethiopian sorghum landraces using the maximum peak of K values based on Evanno *et al*. (2005) prediction among the studied regions (a) and populations (b).

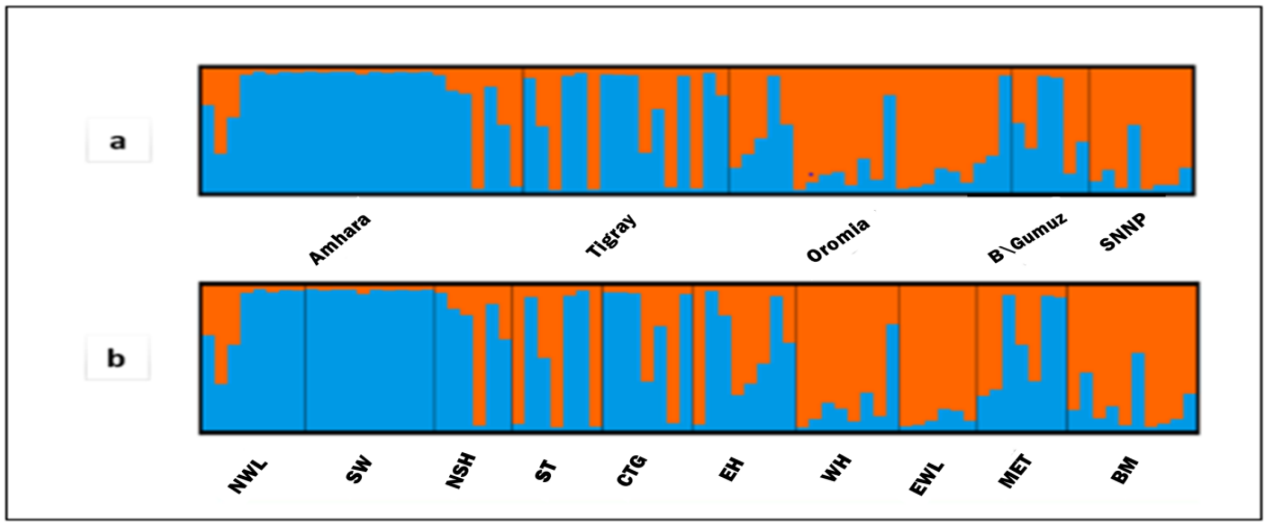


Figure 8. The population structure bar graphs of the studied regions (a), populations (b) of the 77 sorghum landraces of Ethiopia. The different colures (blue and orange) represent genetic groups or sub-populations designated by Structure Harvester: the x-axis represents individual samples and the y-axis represents the proportion of ancestry to each cluster. Key: NWL=North Wollo, STG=South Tigray, CTG=Central Tigray, EH=East Hararge, WH=West Hararge, EWL=East Wellega, BM=Benchi-Maji, MET=Metekel, SWL=South Wollo, NSH=North Shewa, B-Gumuz=Benishangul-Gumuz, SNNP = Southern Nations, Nationalities, and People's Region.

5. DISCUSSION

5.1. Polymorphism level of the microsatellite markers

In this study, eleven microsatellite markers were used to determine the genetic structure of 77 sorghum landraces of Ethiopia. The observed SSR fragment sizes were within the range of previous reports (Folkertsma *et al.*, 2005; Ng'uni *et al.*, 2012; Tesfaye Disasa *et al.*, 2016; Alemu Tirfessa *et al.*, 2020).

The average number of alleles observed in this study (Na = 4.5) is lower than the report of Anas and Yoshida (2004), Tesfamichael Abraha (2014), and Alemu Tirfessa *et al.* (2020) who reported 4.8 alleles in 22 inbred lines using nine SSR, 4.8 alleles in 98 landraces using 29 SSR markers and 6.7 alleles per locus in 200 sorghum landraces, respectively. But, the average number of alleles observed in the present study is higher than that of Danquah *et al.* (2019) who reported a mean of 4.2 by using 22 markers. The difference observed between our finding and formerly mentioned reports could be the difference in the size of the population, size of landraces, and/or the number of SSR markers used (Hague and Routman, 2016).

In the present study, except for one, all the used microsatellite loci showed moderate (PIC=0.23) to high (PIC=0.67) PIC as classified by Vaiman *et al*. (1994), indicating that they are highly informative and useful genetic tools to uncover the genetic structure of sorghum. It also indicates the high-resolution potential of the used microsatellite markers to detect differences between sorghum landraces (Salih *et al.*, 2016). The mean polymorphic information content (PIC) value of the current study (PIC=0.48) was fairly similar to the mean PIC reported in previous genetic diversity of sorghum by Muui *et al*. (2016) who reported 0.49 but lower than the report of Cuevas and Prom (2013) (PIC = 0.78), Tesfaye Disasa*et al*. (2016) (PIC=0.69), and Gamachu Olani *et al*. (2021) (PIC=0.76). However, it was considerably higher than that of Danquah *et al.* (2019) who reported 0.44. The variation could be due to the different nature of the landraces used for genotyping, the difference in the size of the population, the number of markers, and the nature of the SSR markers (Hague and Routman, 2016). The observed moderate to high PIC values suggests the potential of the genetic marker to detect differences between sorghum lines Vaiman *et al*. (1994).

The average gene diversity (h) observed in the current study (h=0.53) is exactly similar to the gene diversity (h=0.53) reported by Ng'uni *et al*. (2011) in sorghum using SSR markers. The present average gene diversity (h=0.53) is considerably higher than Danquah *et al.* (2019) who reported a mean gene diversity of 0.48 but lower than the report of Thudi and Frakrudin (2011) (h=0.80). The high allelic diversity observed in the present study could be an indicator of the wide range of genetic diversity existed in Ethiopian landraces (Harlan, 1992; Dillon *et al*., 2007). This will provide breeders the opportunity to select genotypes for breeding and conservation purposes.

5.2. The genetic diversity of populations and regions

The high mean number of alleles per locus, effective number of alleles, Shannon diversity index, expected heterozygosity, and percent of polymorphic loci observed across the 11 regions and populations revealed the existence of high genetic diversity within Ethiopian sorghum populations. Similarly, Tesfamichael Abraha *et al.* (2014) reported an average of 3.07 alleles per locus, expected heterozygosity of 0.49, and percentage of polymorphic loci of 86.32%, which are considerably higher than the current study. The difference could be due to the larger number of genotypes (98) and SSR markers (29) used.

The high mean percentage of polymorphic loci (PPL) scored 78% for the regions and 60.91% for populations indicates the presence of considerable genetic variation among regions and populations of Ethiopian sorghum landraces, suggesting the need to consider the study populations in sorghum breeding and conservation programs. The observed single private alleles for Amhara, Tigray, SNNP, and Benishangul-Gumuz regions, and populations of Central Tigray, Metekel, and North Shewa, suggests the ability of these landraces to have a certain level of adaptations to different environmental conditions.

The larger value of Shannon's diversity index, expected heterozygosity, and percent of polymorphic loci observed in Amhara, Oromia, SNNP, and Benishangul-Gumuz regions could indicate that these regions are the hot spot for sorghum diversity to be exploited for breeding and conservation purposes. Overall, population-based molecular diversity analysis showed that the population of Benchi-Maji and North Shewa had the highest diversity, indicating the presence of relatively high genetic diversity within sorghum populations of the regions.

5.3. Genetic distance, differentiation, and gene flow

Genetic diversity is influenced by gene flow, genetic distance, and differentiation. The observed low to moderate genetic differentiation between possible pairs of regions and the lower genetic differentiation between all possible pairs of populations is probably indicative of high seed exchange. Wright (1965) found an inverse relationship between gene flow and population differentiation, FST, where high differentiation within the population results in less gene flow between the populations and vice versa, and suggested that FST values range from 0 to 1, and it can be classified as little (0.0–0.05), moderate (0.05–0.15), high (0.15–0.25) and very high (>0.25).

According to Slatkin (1985), the gene flow (Nm) values can be divided into three categories: high (Nm> 1), medium, (0.25-0.99), and low (0.000-0.249). The low to moderate value of genetic distance and differentiation between the possible pairs of regions and populations in the current study showed the exchange of seed materials between these regions. The observed high gene flow between SNNP and Amhara regions followed by the gene flow value between Oromia and SNNP, and Oromia and Amhara regions could likely be due to the 1984 resettlement of farmers with their seeds from northern highlands (Wollo and Gojjam) to southern lowlands of the country through the redundant problem of famine in the northern highlands of Ethiopia (Sivini, 1986), and the sharing of the longest common border of Oromia with both SNNP and Amhara might have contributed to the high gene flow through seed-exchange and pollen movement across populations, respectively. The observed lower gene flow along with larger genetic distance and differentiation between the region of Tigray with both Benishangul-Gumuz and SNNP could be due to the long geographical distance among these regions.

The observed highest gene flow with the lowest genetic distance between the populations of North Shewa and Benchi-Maji, is probably due to the high seed exchange between these populations that confirms the highest gene flow value observed between Amhara and SNNP regions because of resettlement of farmers from northern to the southern part of Ethiopia (Sivini, 1986). The lowest gene flow observed between possible pairs of South Tigray with both South Wollo and Metekel with the highest genetic distance and differentiation is probably due to lower seed exchange between South Tigray and South Wollo, and the long geographical distance between South Tigray and Metekel.

5.4. Genetic relationship

Neighbor-joining cluster analysis of the 77 landraces revealed an admixed clustering pattern confirming the presence of moderate to high gene flow among the regions and populations, indicating that the genetic background of sorghum populations may not always correlate with their geographical areas of sampling. None of the major clusters was composed of exclusively landraces from a particular region or population, confirming the existence of considerable intermixing of the landraces.

The Neighbor-Joining- based clustering grouped both the regions and the populations into three major clusters. The two major clusters (II and III) of populations were constituted of landraces originated from different geographical origins, confirming the existence of high gene flow among populations. Asfaw Adugna (2012) reported that the agro-ecological zonation, which divided Ethiopia into 32 agro-ecologies, of which sorghum grows in 12 of them, was more informative for sorghum germplasm collection than using geographical regions of origin. Moreover, several other studies have reported weak differentiation among landraces according to geographic region (Amsalu Ayana *et al*., 2000; Uptmoor *et al*., 2003; Asfaw Adugna, 2012; Motlhaodi *et al*. 2014).

The absence of a sharp, genetically distinct clustering pattern was also evidenced by the results of structure and principal coordinate analysis (PCoA). PCoA detected a high genetic intermixing among *S. bicolor* for both regions and populations. This intermixing of genotypes originated from different populations is probably due to the high genetic materials exchange rate among the various sorghum-growing regions and populations through marketing or relocation of farmers.

5.5. Population structure

The observed lower contribution of variation existed among population and regions (21% and 18%), respectively as revealed by AMOVA, indicating the presence of gene flow resulted due to the existence of seed exchange between farmers in the studied regions and populations which is confirmed by high gene flow (Nm=0.95 and 1.16) among population and region, respectively. Similar to the present study, a lower proportion of among-population genetic variations of 0.38%, 2.75%, and 12% were reported by Tesfamichael Abraha (2014), Muui *et al*. (2016), and Alemu Tirfessa *et al*. (2020), in Eritrea, Kenya, and Ethiopia, respectively. Conversely, the observed high

within populations (79%) and within regions (82%) genetic variations could be due to the presence of mutation and recombination. Similar higher proportions of within population genetic variations of 99.62%, 91.61%, and 89.7% were reported in the previous studies of sorghum landraces in Ethiopia, Kenya, and Burkina Faso by Alemu Tirfessa *et al*. (2020), Muui *et al*. (2016), Barro-Kondombo *et al*. (2010), respectively.

STRUCTURE analysis also evidenced the presence of weak sub-division (close relationship) among the samples from different regions and collection zones. It detected two sub-groups (K =2) with a greater degree of genetic admixture as a result of gene flow that led to weak population differentiation. The population structure of sorghum for both regions and populations has not belonged to a specific genetic group that showed some genetic admixture which is complementary to the Neighbor-Joining- based clustering of both regions and populations showing relatedness among the 77 sorghum landraces. Indicating that each individual population shares genetic background. Asfaw Adugna, (2014) also reported two sub-groups (K=2) in Ethiopian cultivated *S. bicolor* landraces using SSR markers.

6. CONCLUSION AND RECOMMENDATIONS

6.1. CONCLUSION

In general, the assessment of genetic diversity is an important step in any crop improvement package. Knowledge of a crop's genetic diversity usually aids breeders in selecting desirable parents for breeding programs. It also has a great implication in plant genetic material conservation. The present study targeted to uncover the genetic structure of 77 Ethiopian sorghum landraces using 11 microsatellite markers. The used microsatellite markers were highly informative, indicating that they are useful genetic tools for diversity analysis. Among these markers, ten of them were modestly to highly informative for diversity analysis. Ethiopian sorghum landraces showed high regional and population diversity. The observed genetic distance and differentiation between possible pairs of regions and populations were low to moderate which is likely due to the presence of high gene flow through germplasm exchanges in the form of seeds and pollen grains movements. Therefore, this high genetic distance and differentiation between these populations and regions are very important in breeding to develop new varieties. Clustering, PCoA, and Structure analysis did not sharply distinctly cluster the populations following their sampling area. AMOVA revealed 20.78 % and 17.79 % of the variation was accounted for among pupations and among regions, respectively. Therefore, the present study has generated useful information on the genetic structure of sorghum landraces of Ethiopia to be used in the sorghum breeding program and conservation of the sorghum gene pool, in the future.

6.2. RECOMMENDATIONS

Based on this study, the following key recommendations were made:

- These highly informative markers in this study can also use for further diversity analysis
- High Genetic diversity in Ethiopian sorghum landraces, could be a hot spot for conservation
- Individuals from highly distant populations (South Tigray & Metekel, and South Tigray & South Wollo), and regions (Tigray and Benishangul Gumuz) should be considered in the breeding program of sorghum

Future line of work

- Some populations and regions were represented by small numbers of landraces due to various justifiable reasons. Therefore, a greater number of sorghum accessions from the national wide collection can provide better information
- The present study involved a microsatellite-based analysis of genetic diversity. Thus, High-density markers such as SNPs and further sequence-based genetic diversity analysis shall be used to provide additional information.

7. REFERENCES

Abera Deresa, Tesfaye Tesso and Asfaw Adugna. (2005). Sorghum Production and Research Experiences.*African Journal of Biotechnology*, *8*(5), pp.18-29.

Adamu Zegeye. (1997). Acceptability of injera with stewed chicken. *Food Quality and Preference*, *8(*4), pp. 293-295.

Ahmed L.A-M. (2012). DNA Based Techniques for Studying Genetic Diversity, Genetic Diversity in Microorganisms, Prof. Mahmut Caliskan (Ed.), *In Tech open*, pp. 95-122

Albert, A. (2014). Analysis of Genetic Diversity of Some Groundnut (*Arachis hypogaea L*.) Genotypes Using Principal Component Analysis. M.Sc. Thesis. Ahmadu Bello University. Zaria, Nigeria.

Alemu Tirfessa, Tesfaye Tesso, Asfaw Adugna, Hussein Mohammed, & Kiambi, D. (2020). Genetic diversity among Ethiopian sorghum [*Sorghum bicolor* (L.) Moench] gene bank accessions as revealed by SSR markers. *African Journal of Biotechnology*, *19*(2), pp. 84-91.

Ali, M. L., Rajewski, J. F., Baenziger, P. S., Gill, K. S., Eskridge, K. M., & Dweikat, I. (2008). Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers.*Molecular Breeding*, *21*(4), pp. 497-509.

Allard, R. W. (1999). John Wiley & Sons. *Principles of plant breeding, 2nd ed.*, pp.71.

Amelework Beyene (2012). Genetic diversity analysis of lowland sorghum [*Sorghum bicolor* (L.) Moench] landraces under moisture stress conditions and breeding for drought tolerance in North-Eastern Ethiopia (Doctoral dissertation).

Amsalu Ayana, Endashaw Bekele& Bryngelsson, T. (2000). Genetic variation of Ethiopian and Eritrean sorghum [*Sorghum bicolor* (L.) Moench] germplasm assessed by random amplified polymorphic DNA (RAPD). *Genetic Resources and Crop Evolution*, *47*(5), pp. 471-482.

Anas, & Yoshida, T. (2004). Sorghum diversity evaluated by simple sequence repeat (SSR) markers and phenotypic performance. *Plant production science*, 7(3), pp. 301-308.

Asfaw Adugna. (2007). The role of introduced sorghum and millets in Ethiopian agriculture. *An Open Access Journal,3*(1), pp. 45-52.

Asfaw Adugna. (2012). Population genetics and ecological studies in wild sorghum [*Sorghum bicolor* (L.) Moench] in Ethiopia: Implications for germplasm conservation (Doctoral dissertation, Addis Ababa University).

Asfaw Adugna., Snow, A. A., Sweeney, P. M., Endashaw Bekele, & Mutegi, E. (2013). Population genetic structure of in situ wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. *Genetic resources and crop evolution*, *60*(4), pp. 1313-1328.

Asfaw Adugna. (2014). Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. *Springer Plus*, *3*(1), pp. 1-14.

Ashutosh K. (2013). Studies on Genetic Diversity of Groundnut (*Arachis hypogaea L*.) Using Morphological and Molecular Marker. M.Sc Thesis. Plant Breeding and Genetics, and Department of Agricultural Biotechnology, College of Agriculture, OUAT, Bhubaneswar. Bhubaneswar, India.

Barro-Kondombo, C., Sagnard, F., Chantereau, J., Deu, M., vom Brocke, K., Durand, P., Goze, E., & Zongo, J. (2010). Genetic structure among sorghum landraces as revealed by morphological variation and microsatellite markers in three agro-climatic regions of Burkina Faso. *Theoretical and Applied Genetics*, *120*, pp. 1511–1523.

Beckmann, J.S. & Weber, J.L. (1992). Survey of human and rat microsatellites.*Genomics*, *12*, pp. 627–631.

Bidichandani, S., Ashizawa, T., & Patel, P.I. (1998). The GAA triplet-repeat expansion in *Friedreich ataxia* interferes with transcription and may be associated with an unusual DNA structure. *American Journal of Human Genetics, 62*, pp. 111–121.

Billot, C., Ramu, P., Bouchet, S., Chantereau, J., Deu, M., Gardes, L., & Hash, C. T. (2013). Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *for Public Library of Science*, *8*, pp. 76-79.

Bio-Rad Laboratories, Inc. https://www.bio-rad.com/en-sg/product/image-lab software ID=KRE6P5E8Z. (Assessed on date 11/1/2022).

Bisrat Ghebru, Schmidt, R., & Bennetzen, J. (2002). Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. *Theoretical and Applied Genetics*, *105*(2-3), pp. 229-236.

Bray, E. A. (2000). Response to abiotic stress. *Biochemistry and molecular biology of plants*, 1158-1203.

Brhane Gebrekidan & Belaynesh Gebrehiwot. (1981, October). Sorghum injera preparations and quality parameters.In *Proceedings of the International Symposium on Sorghum Grain Quality, 21-31 October*, (pp. 55-65).

Brhane Gebrekidan. (1982). Utilization of accession in sorghum improvement. In Sorghum in the eighties.*Proceedings of the international symposium on sorghum, ICRISAT, Patancheru*.

Burow, G., Franks, C. D., Xin, Z., & Burke, J. J. (2012). Genetic diversity in a collection of Chinese sorghum landraces assessed by microsatellites.

Centre for Agriculture and Bioscience International (CABI). Invasive Species Compendium. Sorghum bicolor (sorghum) (cabi.org) (assessed on 11/1/2022)

Clayton, W. D., & Renvoize, S. A. (1986). Genera Graminum Grasses of the World. *Kew Bulletin Additional Series,13*, pp. 338–345.

Craufurd, P., Mahalakshmi, V., Bidinger, F., Mukuru, S., Chantereau, J., Omanga, P., Qi, A., Roberts, E., Ellis, R. & Summerfield, R. (1999). Adaptation of sorghum: characterization of genotypic flowering responses to temperature and photoperiod. *Theoretical and Applied Genetics*, *99*(5), pp. 900-911.

CSA (2020). Ethiopian statistical abstract for the year 2016/17. Report on area and production of major crops.

Cuevas, H.E. and Prom, L.K. (2013). Assessment of molecular diversity and population structure of the Ethiopian sorghum [*Sorghum bicolor* (L.) Moench] germplasm collection

maintained by the USDA–ARS National Plant Germplasm System using SSR markers. *Genetic Resources and Crop Evolution*, *60*(6), pp. 1817-1830.

Danquah, A., Galyuon, I. K., Otwe, E. P. and Asante, D. K. (2019). Genetic diversity in some Ghanaian and Malian sorghum [*Sorghum bicolor* (L) Moench] accessions using SSR markers. *African journal of biotechnology,18* (27), pp. 591-602.

De Wet, J. (1978). Special paper: systematics and evolution of sorghum sect. Sorghum (Gramineae). *American journal of botany,65* (4), pp. 477-484.

Dillon, S. L., Lawrence, P. K., Henry, R. J., & Price, H. J. (2007). Sorghum resolved as a distinct genus based on combined ITS1, ndhF, and Adh1 analyses. *Plant Systematics and Evolution*, *268*(1), pp. 29-43.

Djè, Y., Heuertz, M., Lefebvre, C., & Vekemans, X. (2000). Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. *Theoretical and Applied Genetics*, *100*(6), pp. 918-925.

Doggett, H. (1965). The development of the cultivated sorghums.*Genetic Resources and Crop Evolution*, pp. 50-69.

Doggett, H. (1988). Sorghum.2nd ed. John Wiley & Sons, New York.

Doggett, H. & Prasada, R. (1995). Sorghum. p. 173-180. In: Smartt, J. & N.W. Simmonds (eds.) Evolution of Crop Plants 2nd edition. Longman Group Limited, UK.

Duncan, R. R. (1996). Breeding and improvement of forage sorghums for the tropics. *Advances in agronomy*, *57*, pp. 161-185.

Duodu, K., Taylor, J., Belton, P. & Hamaker, B. (2003). Factors affecting sorghum protein digestibility.*Journal of cereal science*, *38*, pp. 117–131.

Earl, D.A. & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resource,4*(2), pp. 359-361.

Endashaw Bekele. (1985). The biology of cereal landrace populations: I. Problems of gene conservation, plant-breeding selection schemes, and sample size requirement. *Hereditas*, *103*(1), pp. 119-134.

EBI (Ethiopian Biodiversity Institute) https://www.ebi.gov.et/biodiversity/conservation/genetic-material-holdings/ (2021) (assessed on 7/1/2022).

Ethiopian Institute of Agricultural Research (EIAR), (2014). Ethiopian strategy for sorghum. 2014-2024, pp. 1-10.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology, 14*(8), pp. 2611-2620.

Excoffier, L. & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, *10*(3), pp. 564-567.

FAO. (2011, March). WHO food standards programme codex committee on contaminants in foods. In *Fifth Session [displayed 10 February 2014]. Available at ftp://ftp. fao. org/codex/meetings/CCCF/cccf5/cf05_INF. pdf*.

FAOSTAT. (2019). Food and Agriculture Organization of the United Nations (FAO). *FAOSTAT database.*

FAOSTAT. (2021). Food and Agriculture Organization of the United Nations Database of agricultural production.

Farooq, S. & Azam, F. (2002). Molecular markers in plant breeding – 1: concepts and characterization. *Pakistan journal of biological sciences*, *5* (10), pp. 1135-1140.

Folkertsma, R. T., Rattunde, H. F. W., Chandra, S., Raju, G. S., & Hash, C. T. (2005). The pattern of genetic diversity of Guinea-race Sorghum bicolor (L.) Moench landraces as revealed with SSR markers. *Theoretical and Applied Genetics*, *111*(3), pp. 399-409.

Gamachu Olani, Tileye Feyissa, & Tesfaye Disasa. (2021). Genetic diversity analysis of sorghum [*Sorghum bicolor* (L.) Moench] races in Ethiopia using simple sequence repeats (SSR) markers. *African Journal of Biotechnology*, *20*(11), pp. 431-439.

Garber, E. D. (1950). Cytotaxonomic studies in the genus Sorghum. *University of California Publications in Botany*, *23*, pp. 283-362.

Gebeyehu Geremew, Asfaw Adugna, Taye Tadesse, Tesfaye Tesso, Ketema Belete, & Michael, H. S. (2004). Developments of varieties and hybrids for dryland areas of Ethiopia.*Uganda Journal of Agricultural Science*, *9*, pp. 594-805.

Gebisa Ejeta, & Grenier, C. (2005). Sorghum and its weedy hybrids. *Crop fertility and volunteerism. CRC Press, Boca Raton, FL*, pp. 123-135.

Gebisa Ejeta., Babiker, A. G. T. & Butler, L. (2002). New approaches to the control of *Striga*, a training workshop on *Striga* resistance. *Melkassa, May*, pp. 14-17.

Gepts, P. (1993). The use of molecular and biochemical markers in crop evolution studies.*Evolutionary biology*, pp. 51-94.

Govers, (2005). DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol", International Scholarly Research Notices, 6:62-71.

Govindaraj, M., Vetriventhan, M. & Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international*, *2015*.

Gupta, P. K. & Varshney, R. K. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat.*Euphytica*, *113*(3), pp. 163-185.

Hague, M. T., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, *116*(1), pp. 92-98.

Hailekiros Tadesse and Tileye Feyissa (2013). Analysis of genetic diversity of Sorghum bicolor ssp. bicolor (L.) Moench using ISSR markers. *Asian Journal of Plant Sciences*, *12*(2), pp. 61-70.

Hariprasanna, K. & Rakshit, S. (2016). Economic importance of sorghum.In The sorghum genome.*Springer, Cham*, pp. 1-25.

Harlan, J. R. and De Wet, J. M. (1972). A simplified classification of cultivated sorghum.*Crop Science,12* (2), pp. 172-176.

Harlan, J. R. (1992). Indigenous African agriculture.*Crops & Man*, (cropsandman), 175-191.

Haussmann, B., Mahalakshmi, V., Reddy, B., Seetharama, N., Hash, C., & Geiger, H. (2002). QTL mapping of stay-green in two sorghum recombinant inbred populations. *Theoretical and Applied Genetics*, *106*(1), pp. 133-142.

ICRISAT. (2015). **[***Sorghum bicolor* (L.) Moench**]**. International Crops Research Institute for the Semi-Arid Tropics.

Jonah, P.M., Bello, L.L., Lucky, O., Midau, A., Moruppa, S.M. & Moruppa, S.M., 2011. Review: The importance of molecular markers in plant breeding programs. Global Journal of Science Frontier Research, 11(5), pp.4-12.

Kassa Semagn, Bjørnstad, Å. & Ndjiondjop, M. N. (2006). An overview of molecular marker methods for plants.*African journal of biotechnology*, *5*(25). pp. 2540-2568.

Kidanemaryam Wagaw. (2019). Review on mechanisms of drought tolerance in sorghum **[***Sorghum bicolor* (L.) Moench**]** basis and breeding methods. *Academic Research Journal of Agricultural Science and Research*, *7*(2), pp. 87-99.

Kimber, C. T., Dahlberg, J. A. & Kresovich, S. (2013). The gene pool of *Sorghum bicolor* and its improvement.*Genomics Information*, pp. 23-41.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology*, *15*(5), pp. 1179-1191.

Krzyzaniak, L. (1978). New light on early food production in the Central Sudan.*The Journal of African History*, *19*(2), pp. 159-172.

Kumar, P., Gupta, V.K., Misra, A.K., Modi, D.R., & Pandey, B.K. (2009). Potential of Molecular Markers in Plant Biotechnology. *Plant Omics*, *2*(4), pp. 141-162.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution, 35*, pp. 1547-1549.

Liu, K. & Muse, S.V. (2005). Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics, 21*(9), pp. 2128-2129.

Mann, J.A., Kimber, C.T. & Miller, F.R. (1983). The origin and early cultivation of sorghums in Africa. *Texas Agricultural Experimental Station.21, pp.* 14-54.

Manzelli, M., Pileri, L., Lacerenza, N., Benedettelli, S., & Vecchio, V. (2007). Genetic diversity assessment in Somali **[***Sorghum bicolor* (L.) Moench**]** accessions using microsatellite markers. *Biodiversity and conservation, 16*(6), pp. 1715-1730.

Mehmood, S., Bashir, A., Ahmad, A., Akram, Z., Jabeen, N., & Gulfraz, M. (2008). Molecular characterization of regional *Sorghum bicolor* varieties from Pakistan.*Pakistan Journal of Biotechnology, 40*(5), pp. 2015-2021.

Menz, M. A., Klein, R. R., Mullet, J. E., Obert, J. A., Unruh, N. C. & Klein, P. E. (2002). A high-density genetic map of **[***Sorghum bicolor* (L.) Moench**]** based on 2926 AFLP, RFLP, and SSR markers. *Plant Molecular Biology,48*(5), pp. 483-499.

Missihoun, A.A., Adoukonou-Sagbadja, H., Sedah, P., Agbangla, C., Ahanhanzo, C. & Dagba, R.A., (2015). Genetic diversity of **[***Sorghum bicolor* (L.) Moench**]** landraces from Northwestern Benin as revealed by microsatellite markers. *African journal of biotechnology, 14*(16), pp. 1342-1353.

Mofokeng, A., Shimelis, H., Tongoona, P., & Laing, M. (2014). Genetic diversity analysis of South African sorghum genotypes using SSR markers.*South African Journal of Plant and Soil, 31*(3), pp. 145-152.

Mohammed, H. E. A. (2017). Genetic diversity estimation using SSR markers and some yield components in seven forage sorghum **[***sorghum bicolor* (l.) Moench**]**. *Pakistan Journal of Biotechnology,14* (4), pp. 739-744.

Molosiwa O. O. (2012). Genetic diversity and population structure analysis of Bambara groundnuts *(Vignasub terranea (L.) Verdc.)* Landraces using morphoagronomic characters and SSR markers.Ph.D. Thesis.University of Nottingham. Nottingham, UK.

Mondini, L., Noorani, A. & Pagnotta, M. A. (2009). Assessing plant genetic diversity by molecular tools. *Diversity, 1*, pp. 19–35.

Motlhaodi, T., Mulatu Geleta, Bryngelsson, T., Fatih, M., Chite, S., and Ortiz, R. (2014). Genetic diversity in-situ' conserved sorghum accessions of Botswana as estimated by microsatellite markers. *Australian Journal of Crop Science*, *8*(1), pp. 35-43.

Motlhaodi, T., Mulatu Geleta, Chite, S., Fatih, M., Ortiz, R., and Bryngelsson, T. (2017). Genetic diversity in sorghum [*Sorghum bicolor* (L.) Moench] germplasm from Southern Africa as revealed by microsatellite markers and agro-morphological traits.*Genetic Resources and Crop Evolution,64* (3), pp. 599-610.

Mundia, C.W., Secchi, S., Akamani, K. & Wang, G. (2019). A regional comparison of factors affecting global sorghum production: The case of North America, Asia, and Africa's Sahel. *Sustainability*, *11*(7), pp. 2135.

Muthusamy, S., Kanagarajan, S. & Ponnusamy, S. (2008). Efficiency of RAPD and ISSR markers system in accessing genetic variation of rice bean (Vigna umbellata) landraces. *Electronic Journal of Biotechnology*, *11*(3), pp. 32-41.

Muui, C. W., Muasya, R. M., Kirubi, D. T., Runo, S. M. & Karugu, A. (2016). Genetic variability of sorghum landraces from lower Eastern Kenya based on simple sequence repeats (SSRs) markers. *African journal of biotechnology,15* (8), pp. 264-271.

Nei, M., Tajima, F., & Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. *Journal of molecular evolution*, *19*(2), pp. 153-170.

Nemera Geleta., & Labuschagne, M. T. (2005). Qualitative traits variation in sorghum **[***Sorghum bicolor* (L.) Moench**]** germplasm from, eastern highlands of Ethiopia.*Biodiversity & Conservation*, *14*(13), pp. 3055-3064.

Ng'uni, D., Mulatu Geleta. & Bryngelsson, T. (2011). Genetic diversity in sorghum **[***Sorghum bicolor* (L.) Moench**]** accessions of Zambia as revealed by simple sequence repeats (SSR). *Hereditas, 148* (2), pp. 52-62.

Ng'uni, D., Mulatu Geleta, Hofvander, P., Fatih, M. & Bryngelsson, T. (2012). Comparative genetic diversity and nutritional quality variation among some important Southern African sorghum accessions [*Sorghum bicolor* (L.) Moench]. *Australian Journal of Crop Science, 6* (1), pp. 56.

Peakall, R. & Smouse, R., (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics, 28*(19), pp. 2537-2539.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155* (2), pp. 945-959.

Ramanatha., R. & Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture, 68,* pp. 1-19.

Raman, R., Cowley, R., Raman, H. & Luckett, D. (2014). Analyses using SSR and DArT molecular markers reveal that Ethiopian accessions of white lupin (*Lupinus albus L*.) represent a unique gene pool. *Open journal of genetics, 4,* pp. 87-98.

Reddy, V. G., Rao, N. K., Reddy, B. V. S., & Rao, K. E. P. (2002). Geographic distribution of basic and intermediate races in the world collection of sorghum germplasm. *International Sorghum and Millets Newsletter, 43,* pp. 15-17.

Reddy, B. V., Kumar, A. A., Reddy, P. S., & Elangovan, M. (2008). Sorghum germplasm: diversity and utilization.

Reddy, B. V., Ramesh, S., Reddy, P. S., & Kumar, A. A. (2009). 3 Genetic Enhancement for Drought Tolerance in Sorghum. *Plant Breeding Reviews, 31,* pp. 189.

Ruiz-Chután, J. A., Salava, J., Janovská, D., Žiarovská, J., Kalousová, M. & Fernández, E. (2019). Assessment of genetic diversity in *Sorghum bicolor* using RAPD markers. *Genetika,51* (3), pp. 789-803.

Salih, S. A., Herslman, L., Labuschange, M. T. & Mohammed, A. H. (2016). Assessment of genetic diversity of sorghum [*Sorghum bicolor* (L.) Moench] germplasm in East and Central Africa.*World Journal of Microbiology and Biotechnology, 1* (03), pp. 113-120.

Shapter, F. M, Crowther, A., Fox G., Godwin, I. D. & Watson-Fox, L. (2018). The domestication, spread, and use of sorghum as a crop. In: achieving suitable cultivation of sorghum. (Rooney, B.). *Science Publishing*, 31-64.

Singh, S.P. (2005). Common Bean (Phaseolus vulgaris L.). Genetic Resources, Chromosome Engineering, and Crop Improvement Serie, Vol.1, Taylor and Francis, Boca Raton London, New York, Singapore. (edt book). In: Ram J. Singh (ed.), Vol.1, pp.13-57.

Sivini, G. (1986). Famine and the resettlement program in Ethiopia. *Africa*, pp. 211-242.

Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, *39*(1), pp.53-65.

Poehlman, J., and Sleper, D. (1979). Breeding field crops. Amsterdam: *Springer*. (No. Ed. 5).

Smith, J. S. C., & Smith, O. S. (1992). Fingerprinting crop varieties. In *Advances in agronomy.* Academic Press, Vol. 47, pp. 85-140.

Srinivasa, P., Reddy, B. V., Nagaraj, N., & Upadhyaya, H. D. (2014). Sorghum production for diversified uses, pp. 01-27.

Tabbasam, N., & Zafar, Y. (2006). DNA-based genotyping of sorghum hybrids. *Pakistan Journal of Botany (Pakistan)*. *38* (5), pp. 1599- 1604.

Tadele Temesgen. (2014). Genetic Diversity of Chickpea (*Cicerarietinum L*.) Cultivars from Ethiopia by Using Inter Simple Sequence Repeat Markers. M.Sc. Thesis. Haramaya University. Haramaya, Ethiopia.

Tesfamichael Abraha., Githiri, S. M., Kasili, R. W., Skilton, R. A., Solomon, M., & Nyende, A. B. (2014). Genetic diversity analysis of Eritrean sorghum **[***Sorghum bicolor* (L.) Moench**]** germplasm using SSR markers.*Molecular Plant Breeding, 5*.

Tesfaye Disasa. (2016). Genetic Diversity Analysis and QTL Mapping of Selected Traits in Sorghum **[***Sorghum bicolor* (L.) Moench**]**. Addis Ababa University,

Thakur, J.S., Prinja, S, Garg, C.C., Mendis, S. & Menabde, N. (2011). Social and economic implications of non-communicable diseases in India.*Indian Journal of Community Medicine*, *36*(1), pp. S13-S22.

Tharachand, C., Immanuel, S.C. & Mythili, M.N. (2012). Molecular Markers in Characterization of Medicinal Plants: An overview.*Research in plant biology, 2*(2), pp. 01-12.

Thudi, M. & Fakrudin, B. (2011). Identification of unique alleles and assessment of genetic diversity of rabi sorghum accessions using simple sequence repeat markers. *Journal of Plant Biochemistry and Biotechnology, 20* (1), pp. 74-83.

Uptmoor, R., Wenzel, W., Friedt, W., Donaldson, G., Ayisi, K. & Ordon, F. (2003). Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs, and SSRs. *Theoretical and Applied Genetics*, *106*(7), pp. 1316-1325.

Vaiman, D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lépingle, A. & Guérin, G. (1994). A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mammalian Genome, 5*(5), pp. 288-297.

Vara, P. P.V. & Staggenborg, A. S. (2009). Growth and production of sorghum and millets. In: Soils, Plant Growth and Crop Production. (Willy H. Verheye). *Encyclopedia of Life Support Systems*, UK, pp. 1-27.

Wang, Y.H., Upadhyaya, H.D., Burrell, A.M., Sahraeian, S.M., Klein, R.R. & Klein, P.E. (2013). Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant *Sorghum bicolor*. *3*(5), pp. 783–793.

Wortmann, C. S., Martha Mamo, Girma Abebe., Mburu, C., Kayuki, K. C., Letayo, E. & Xerinda, S. (2006). Atlas of sorghum **[***Sorghum bicolor* (L.) Moench**]**: production in five countries of Eastern Africa.

Wright, S. (1965). The interpretation of population structure by F- statistics with special regard to systems of mating.*Evolution,* 19 (3), pp. 395-420.

Wu, J.L., Wu, C.J., Lei, C.L., Baraoidan, M., Bordeos, A. & Madamba, M.R.S. (2004). Sorghum diversity evaluated by simple sequence repeat (SSR) markers and phenotypic performance. *Plant Production Science,7*, pp. 301–308.

Yaqoob, M., Hussain, N. & Rashid, A. (2015). Genetic variability and heritability analysis for yield and morphological traits in sorghum **[***Sorghum bicolor* (L.) Moench**]** genotypes. *Journal of Agricultural Research*, *53*(3), pp. 331.

Yilma Kebede. (1991). The role of Ethiopian sorghum germplasm resources in the national breeding programme. In: Plant Genetic Resources of Ethiopia (eds JM Engels, JG Hawkes, and M Woreda), Cambridge University Press, *Cambridge*, pp. 315-322.

Zegada-Lizarazu, W. & Monti, A. (2012). Are we ready to cultivate sweet sorghum as a bioenergy feedstock? A review of on-field management practices. *Biomass and Bioenerg*, *40*, pp. 1–12.

Zhu, M., Chen, J., Yuyama, N., Luo, L., Xiao, X., Lv, Y., Liu, Y. & Cai, H. (2020). Genetic Diversity and Population Structure of Broomcorn Sorghum Investigated with Simple Sequence Repeat Markers. *Tropical Plant Biology*, *13* (1), pp. 62-72.

8. APPENDICES

Appendix 1. List of sorghum landraces used in this study.

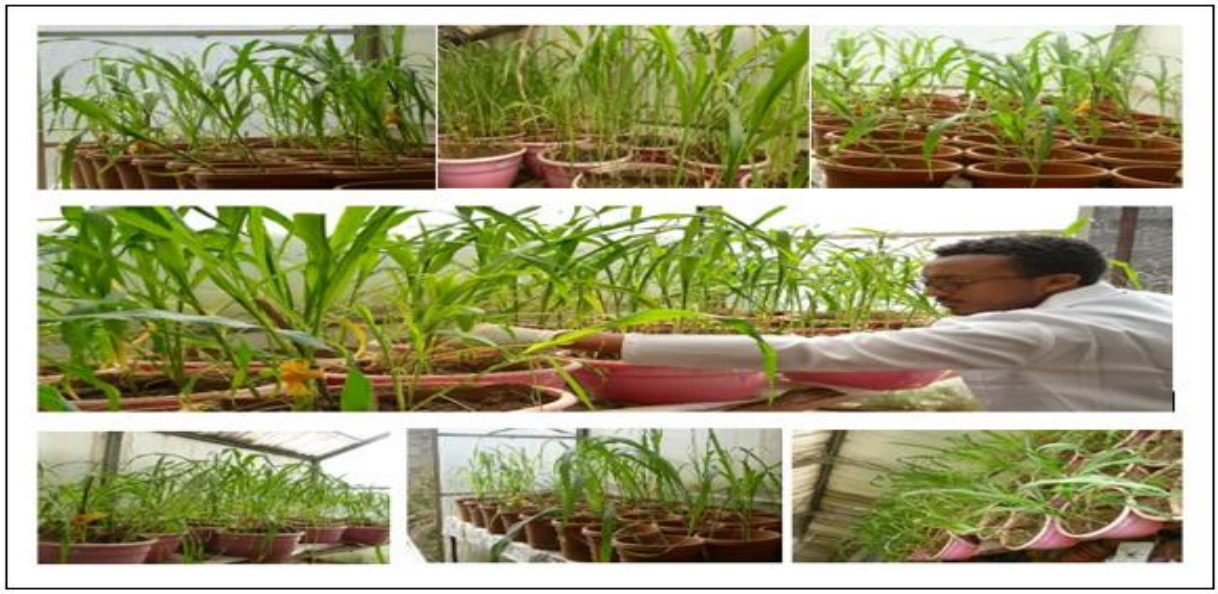
S/No	Accessions	Region	Zone	Altitude	Longitude	Latitude
1	SB42	Amhara	North Wello	1495.3	E 39 º 37. 21	N 12 º 4.38
2	SB51	Amhara	North Wello	1616.5	E 39 º 39·5	N 11 º 58.24
3	SB59	Amhara	North Wello	1456.7	E 39 º 37.17	N 12 º 6.42
4	SB55	Amhara	North Wello	1667.2	E 39 º 38.0	N 12 º 0,30
5	SB70	Amhara	North Wello	1894.7	E 39 º 37.50	N 11 º 44.29
6	SB75	Amhara	North Wello	1762.9	E 39 º 38.1	N 11 º 32.56
7	SB96	Amhara	North Wello	1483.5	E 39 º 40.28	N 11 º 52.7
8	SB106	Amhara	North Wello	1892.8	E 39 º 29.40	N 11 º 53,33
9	SB2	Tigray	South Tigray	1650.8	E 39 º 38.44	N 12 º 42.25
10	SB3	Tigray	South Tigray	1675.7	E 39 º 38.41	N 12 º 42.24
11	SB4	Tigray	South Tigray	1675.7	E 39 º 38.41	N 12 º 42.24
12	SB7	Tigray	South Tigray	1679.5	E 39 º 38.42	N 12 º 42.27
13	SB13	Tigray	South Tigray	1748.9	E 39 º 38.26	N 12 º 45.23
14	SB14	Tigray	South Tigray	1738.9	E 39 º 38.25	N 12 º 45.36
15	SB18	Tigray	South Tigray	1729.5	E 39 º 38.33	N 12 º 45.11
16	SB19	Tigray	South Tigray	1729.4	E 39 º 38.33	N 12 º 45.10
17	SB29	Tigray	South Tigray	1844.8	E 39 º 37.10	N 12 º 48.17
18	SB30	Tigray	South Tigray	1843.4	E 39 º 37.4	N 12 º 48.13
19	235462	Tigray	Central Tigray	1500.00	E 39 º 06.00	N 13 º 41.00
20	26144	Tigray	Central Tigray	2096.00	E 38 º 45.48	N 1 º 40.65
21	234086	Tigray	Central Tigray	1810.00	E 38 º00.00	N 14 º 05.00
22	235477	Tigray	Central Tigray	2260.00	E 39 º 14.00	N 13 º 39.00
23	237274	Tigray	Central Tigray	1450.00	E 39 º 27.00	N 13 º 30.00
24	238405	Tigray	Central Tigray	2170.00	38 º 05.00	N 14 º 04.00
25	9125	Oromia	East Harerge	1745.00	E 41 º 34.22	N 08 º 57.49
26	29507	Oromia	East Harerge	1350.00	E 42 º 13.42	N 09 º 08.49

27	69273	Oromia	East Harerge	1970.00	E 41 º47.00	N 09 º 16.00
28	241238	Oromia	East Harerge	2030.00	E 41 º53.25	N 09 º 27.77
29	241263	Oromia	East Harerge	1850.00	E42 º 05.43	N09º14.40
30	241268	Oromia	East Harerge	1970.00	E41 º 52.90	N09º19.87
31	241270	Oromia	East Harerge	2100.00	E41 º 51.08	N09º1450
32	69233	Oromia	West Harerge	2090.00	E40 º 48.00	N09º05.00
33	210941	Oromia	West Harerge	2310.00	E40 º 30.00	N09º17.00
34	69265	Oromia	West Harerge	1965.00	E41 º 05.00	N09º17.00
35	70442	Oromia	West Harerge	1972.00	E40 º 52.00	N08 º61.00
36	241197	Oromia	West Harerge	1720.00	E40 º 19.78	N08º37.79
37	15888	Oromia	West Harerge	1985.00	E41 º 05.17	N09º10.60
38	241196	Oromia	West Harerge	1720.00	E40 º 19.79	N08º36.56
39	16134	Oromia	East Wollega	1867.00	E36 º 35.34	N80º55.18
40	16147	Oromia	East Wollega	2133.00	E36 º 40.21	N08º46.14
41	69118	Oromia	East Wollega	1440.00	36 º 31.00	N09º28.00
42	69123	Oromia	East Wollega	1270.00	E36 º 31.00	N09º28.00
43	69523	Oromia	East Wollega	1990.00	E36 º 41.00	N09º02.00
44	69524	Oromia	East Wollega	2070.00	E36 º 29.00	N09º02.00
45	69525	Oromia	East Wollega	2260.00	E36 º 39.00	N08º47.00
46	69529	Oromia	East Wollega	1480.00	E35 º 04.00	N08º52.00
47	27909	SNNP	Bench Maji	1372.00	E35 º 12.15	N06º46.60
48	27911	SNNP	Bench Maji	1399.00	E35 º 13.33	N06º48.06
49	27917	SNNP	Bench Maji	1214.00	E35 º 31.13	N06º61.46
50	69546	SNNP	Bench Maji	1780.00	E36 º 25.00	N07º10.00
51	27921	SNNP	Bench Maji	228.00	E 35 º33.56	N06º53.45
52	27922	SNNP	Bench Maji	2088.00	E35 º 51.37	N06º50.58
53	27924	SNNP	Bench Maji	2071.00	E35 º 47.22	N06º54.04
54	27925	SNNP	Bench Maji	2074.00	E35 º 20.20	N06º53.59
55	23639	Benishangul-Gumuz	Metekel	559.00	E35 º 11.48	N10º59.49
56	69147	Benishangul-Gumuz	Metekel	1740.00	E35 º 48.04	N10º35.49

57	229833	Benishangul-Gumuz	Metekel	1350.00	E 36 º 34.00	N 11 º 05-00
58	229832	Benishangul-Gumuz	Metekel	1300.00	E 36 º 33.00	N 11 º 04-00
59	29368	Benishangul-Gumuz	Metekel	735.00	E 34 º 21.37	N 10 º35-09
60	29371	Benishangul-Gumuz	Metekel	1358.00	E 34 º 42.13	N 10 º 19.40
61	SB85	Amhara	South Wello	1638.8	E 39 º 37. 2	N 11 º 24.46
62	SB86	Amhara	South Wello	1619.4	E39 º 36. 46	N 11 º 25.12
63	SB88	Amhara	South Wello	1578.5	E 39 º 37.9	N 11 º 27.50
64	SB89	Amhara	South Wello	1626.2	E 39 º 36.46	N 11 º 25.12
65	SB90	Amhara	South Wello	1620	E39 º 36 .45	N 11 º 25.13
66	SB111	Amhara	South Wello	1714.3	E39 º 46 .22	N 10 º 58.18
67	SB112	Amhara	South Wello	1714.3	E 39 º 46 22	N 10 º 58.18
68	30196	Amhara	North Shewa	1847.00	E 39 º47.09	N 09 º 33.44
69	30197	Amhara	North Shewa	2610.00	E 39 º01.08	N 09 º 47.28
70	17538	Amhara	North Shewa	2603.00	E 39 º 27.14	N 09 º 56.34
71	17509	Amhara	North Shewa	2457.00	E 09 º 56.28	N 39 º 44.31
72	17508	Amhara	North Shewa	1250.00	E 10 º 07.02	N 39 º 53.34
73	SB124	Amhara	North Shewa	1490.7	E39 º 48 .38	N 10 º 53.8
74	SB118	Amhara	North Shewa	1469.5	E39 º 48 .39	N 10 º 53.0
75	SB125	Amhara	North Shewa	1156.7	E 39 º 581.3	N 10 º 7.13
76	SB128	Amhara	North Shewa	1190	E 39 º 58.5	N 10 º 75.6
77	SB131	Amhara	North Shewa	1469.9	E 39 º 50.2	N 10 º 48.13

Appendix 2. Some of the sorghum landraces (Seeds) used for planting in this study. SB85 (a), 229833 (b), 69524 (c), SB128 (d)

Appendix 3. Sorghum seeds while growing in the greenhouse.



Appendix 4. Leaf samples dried in plastic bags filled with silica gels.

Appendix 5. Plant DNA extraction by using CTAB immersed leaf samples procedure and chemical used.

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 a 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) to 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. Steps 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) to 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 inverting 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 the 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 by using a 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. The samples were centrifuged at 13000 rpm for 12 minutes.

26. 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)

27. Repeat steps 21 to 24 with 3 M Na-AC-solution (4°C, half the volume) then proceed with Step 28.

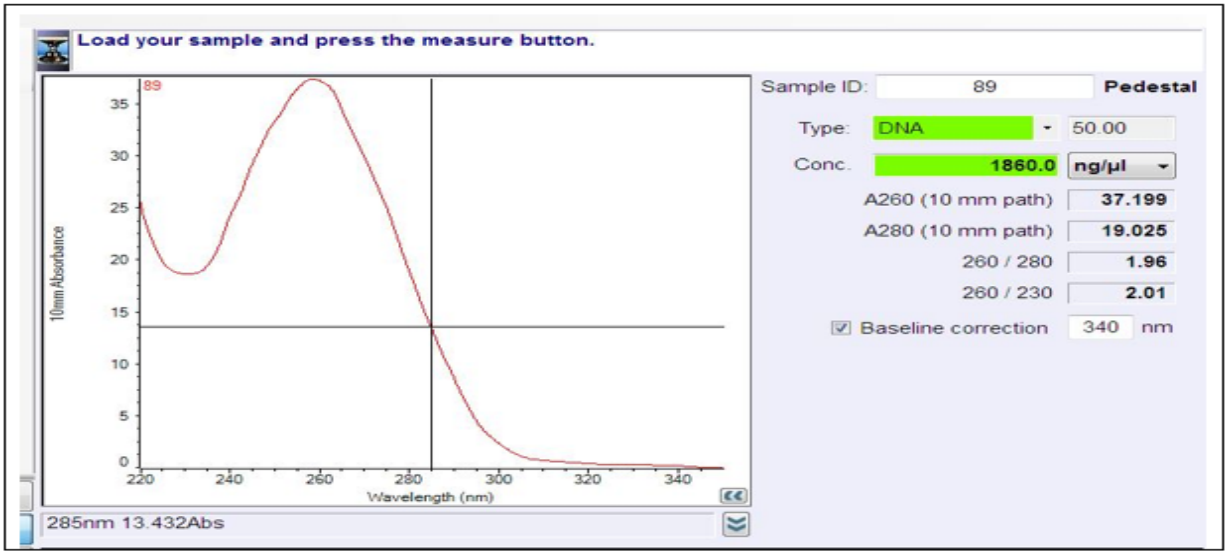
28. Centrifuge 10 min. at 15000 rpm. Aspirate liquid and dry pellet at room temperature.

Dissolve the pellet in 100 µl TE (1x, p.a. grade) and store at -20^{o}c for further use.

Appendix 6. Materials and chemicals used for DNA extraction.

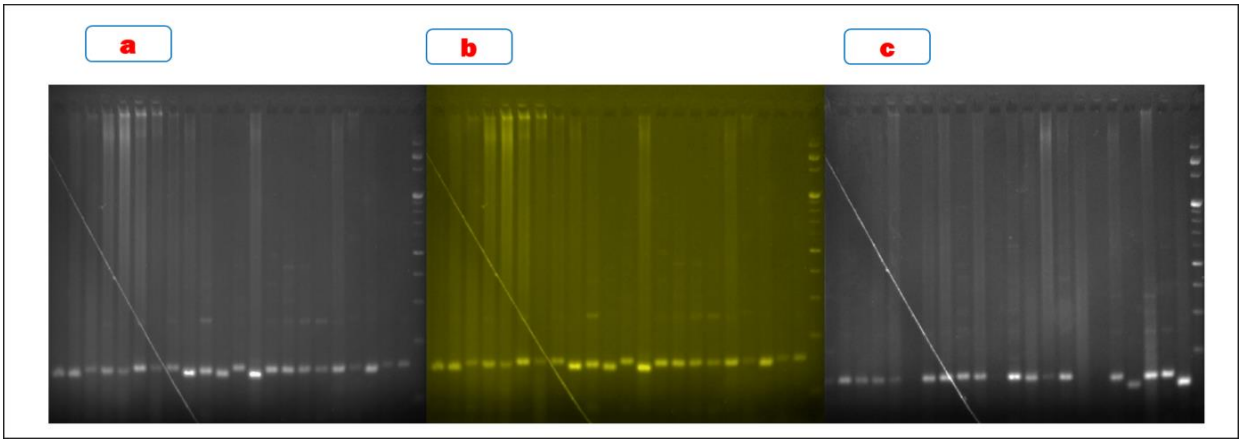
Materials	Chemicals
Beads	CTAB - 3 x 700 µl per sample
Eppenfdorf cup (|Tube)	ß-Mercapto-Ethanol
Water bath	Chloroform
Centrifuge	Isopropanol - 100 %, 4°C
Mixer-mill	Ethanol - 100 % and 70% in sterile ddH2O
Tips	TE - 1x, p.a. grade
Cut 1000 µl tips - Sterile	NH4Ac - 7.5 M Solution, sterile
	NaAc - 3 M Solution, sterile

Appendix 7. The result of genomic DNA obtained via quantification and qualification of Nanodrop.

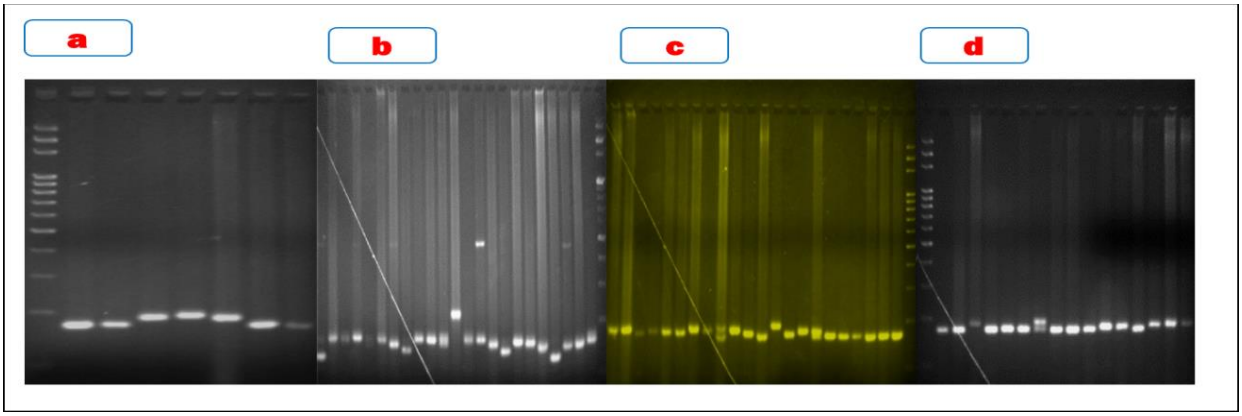


Appendix 8. Electrophoretic bands of SSR markers.

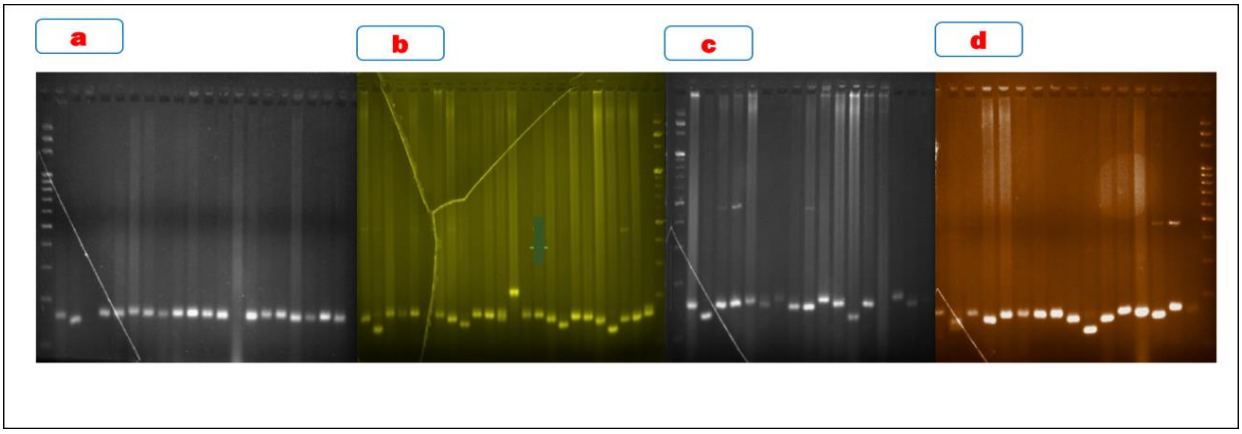
Primer Xtxp 010 (A. Sample from 1- 30, B. 31-61, and c. 62-81)



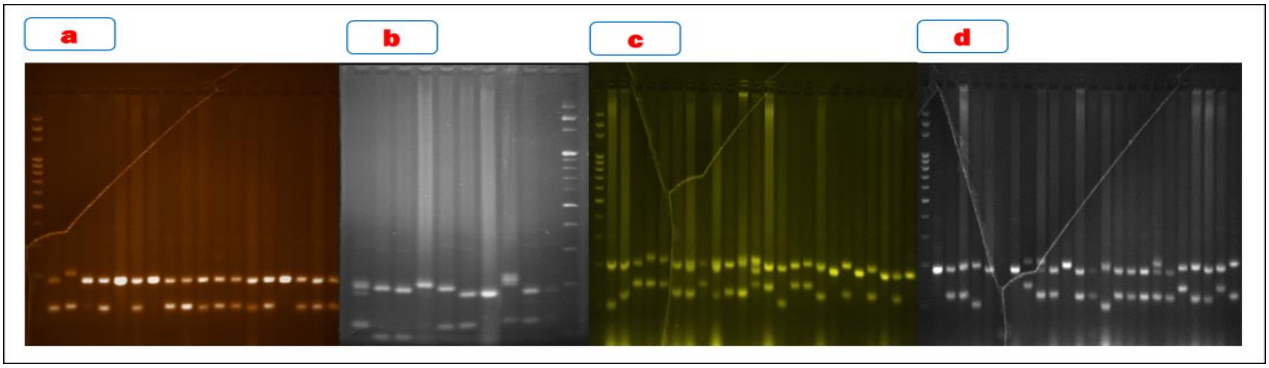
Primer Xtxp 021 (A. Sample from 1- 30, B. 31-61, and c. 62-81)



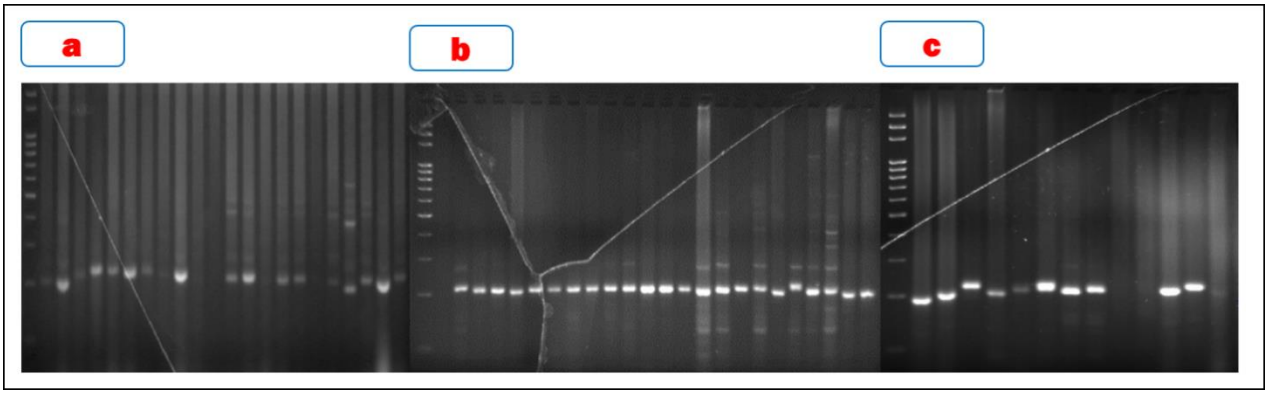
Primer Xtxp 141 (A. Sample from 1- 30, B. 31-61, and c. 62-81)



Primer Xtxp 265 (A. Sample from 1- 30, B. 31-61, and c. 62-81)



Primer Xtxp 273 (A. Sample from 1- 30, B. 31-61, and c. 62-81).



Appendix 9. Representative pictures to give a further illustration of all practical works beginning from sample collection to the end of the laboratory session

Appendix 10. The image that shows how ARLEQUIN calculates for: variance among populations (AP), variance within populations (WP), Total (T), sum of squares (SSD).

Source of variation	Degrees of freedom	Sum of squares (SSD)	Expected mean squares
Among Populations	$P - 1$	SSD(AP)	$n\sigma_a^2 + \sigma_b^2$
Within Populations	$2N - P$	SSD(WP)	σ_b^2
Total:	$2N - 1$	SSD(T)	σ_T^2

Appendix 11. Identification of the true number of clusters (K) based on delta k (ΔK) of the structure harvester (a=region and b=population).

a

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	20	-1093.210000	0.304181	–	–	–
2	20	-964.785000	0.607215	128.425000	60.545000	99.709265
3	20	-896.905000	2.516781	67.880000	20.610000	8.189033
4	20	-849.635000	2.170926	47.270000	25.305000	11.656315
5	20	-827.670000	5.560869	21.965000	9.880000	1.776701
6	20	-795.825000	4.827539	31.845000	14.040000	2.908314
7	20	-778.020000	7.311750	17.805000	0.045000	0.006154
8	20	-760.260000	13.863105	17.760000	1.355000	0.097741
9	20	-743.855000	9.340488	16.405000	5.440000	0.582411
10	20	-732.890000	9.096494	10.965000	–	–

b

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	20	-1095.820000	0.475284	—	—	—
2	20	-966.385000	0.885720	129.435000	58.805000	66.392310
3	20	-895.755000	0.926496	70.630000	35.775000	38.613229
4	20	-860.900000	2.415738	34.855000	13.370000	5.534541
5	20	-839.415000	4.991495	21.485000	9.305000	1.864171
6	20	-808.625000	4.494192	30.790000	1.825000	0.406080
7	20	-779.660000	4.029679	28.965000	11.795000	2.927032
8	20	-762.490000	8.348457	17.170000	2.520000	0.301852
9	20	-742.800000	3.669935	19.690000	15.605000	4.252119
10	20	-738.715000	5.868495	4.085000	—	—