ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Genetic Diversity and Population Structure Analysis of Released and Landrace Sorghum (*Sorghum bicolor* (L.) Moench) Genotypes of Northern and Eastern Ethiopia as Revealed by SSR Markers

M.Sc. Thesis

Abebaw Misganaw Ambaw

April, 2021

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

Genetic Diversity and Population Structure Analysis of Released and Landrace Sorghum (*Sorghum bicolor* (L) Moench) Genotypes of Northern and Eastern Ethiopia as Revealed by SSR Markers

A Thesis Submitted to the Institute of Biotechnology, Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biotechnology

By

Abebaw Misganaw Ambaw

April, 2021

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

INSTITUTE OF BIOTECHNOLOGY

THESIS APPROVAL SHEET

We certify that Abebaw Misganaw Ambaw's M.Sc. thesis entitled "**Genetic Diversity and Population Structure Analysis of Released and Landrace Sorghum (*Sorghum bicolor* (L.) Moench) Genotypes of Northern and Eastern Ethiopia as Revealed by SSR Markers"** has been conducted under our direct supervision. Therefore, we kindly request the Institute of Biotechnology of Addis Ababa University to final approval and acceptance of the thesis.

Examinee:	Signature	Date
Abebaw Misganaw Ambaw	----------	----------
Examiners:	Signature	Date
Teklehaimanot Haileselassie (PhD, Associ. Prof.), IoB, AAU	----------	----------
Demesachew Guadie (PhD) IoB, AAU	----------	----------
Main Advisor:		
Tileye Feyissa (Ph.D., Assoc. Prof.), IoB, AAU	----------	----------
Co-Advisors:		
Tilahun Mekonnen (MSc., PhD candidate), IoB, AAU	----------	---------
Tesfaye Disasa (PhD), EIAR, NABRC	----------	----------
Obssi Desalegne (MSc.), EIAR, NABRC	----------	----------

Dr. Tesfaye Sisay Signature ---------- Date----------

Director, Institute of Biotechnology

Addis Ababa University

DEDICATION

I am very glad to dedicate this thesis research to my late Mother, Tege Dagnew Bogale, for raising me with love and affection. Her selflessness will always be remembered. Also, I dedicate this thesis to my life partner Melesech Teshale Gabi and my kid Amanda Abebaw.

ACKNOWLEDGMENTS

First and foremost, I am very glad to express my warmest gratitude to all individuals and institutions that contributed to my study directly or indirectly. I am so grateful for all of them. However, I find it very vital to mention some of them. I would like to thank the Institute of Biotechnology of Addis Ababa University for giving me admission and the National Agricultural Biotechnology Research Center of Ethiopian Institute of Agricultural Research (EIAR) for the complete financial support to carry out my thesis research and for giving me a leave of absence during the entire study period of my Master's program.

I am grateful to my supervisors Dr: Tileye Feyissa and for his careful supervision with beneficial comments from the beginning to the end of my thesis preparation. I am also grateful to pass my special gratitude to my second co-supervisors Mr. Tilahun Mekonnen, Dr. Tesfaye Disasa (who provided me the SSR primers) and Obssi Desalegne for their support, helpful suggestions, and honorable support in conducting the research work and thesis preparation. Specially, I am very grateful to Mr. Tilahun Mekonnen, for his unforgettable help during my data analysis, providing support for statistical software, elaborating about the scientific principles of genetic diversity and he is very kind, committed and I was always knocking on his door and come with novel idea and suggestions.

I also extend my thanks to all members of the Institute of Biotechnology group of Addis Ababa University and the National Agricultural Biotechnology Research Center for their help and moral support during the study and write-up of the thesis.

Finally, I am very much indebted to my family for nursing me with their warmness from the early days to this day. Finally, yet highly, my friends for their moral and absolute round help that contributed directly or indirectly towards the completion of my whole academic successes.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATION

AMOVA	Analysis of Molecular Variance
FIS	Inbreeding Coefficient
FIT	Overall Inbreeding Coefficient
FST	Fixation Index
GD	Genetic Distance
GI	Genetic Identity
He	Heterozygosity
I	Shannon's Index
MAF	Major Allele Frequency
NA	Number of Allele
NABRC	National Agricultural Biotechnology Research Center
Ne	Number of Effective Alleles
Nm	Gene Flow
PAGE	Polyacrylamide Gel Electrophoresis
PCoA	Principal Component Analysis
PIC	Polymorphic Information Content
SSR	Simple Sequence Repeats

Genetic Diversity and Population Structure Analysis of Released and Landrace Sorghum (*Sorghum bicolor* (L.) Moench) Genotypes of Northern and Eastern Ethiopia as Revealed by SSR Markers

Abebaw Misganaw Ambaw

Institute of Biotechnology, Addis Ababa University.

Abstract

Sorghum (*Sorghum bicolor* (L.) Moench) is the most stable and important food security crop in Ethiopia accounting for nearly 40% of human calorie intake. Knowledge of the natural genetic composition of the crop provides the option to further exploit its genetic potential through breeding. However, there are limited reports on the genetic variability of Ethiopian sorghum using a medium-throughput marker system. Hence, the current study was designed to evaluate the genetic variability of released and landrace Ethiopian sorghum genotypes using polymorphic microsatellite markers. A 92 sorghum accessions collected from five Ethiopian ecological zones were targeted using 12 SSR markers. The study resulted in 77 alleles across the entire loci and populations. All the used microsatellite loci were highly polymorphic with PIC ranging from 0.66 to 0.82 and an overall mean of 0.76. The analysis confirmed the presence of high gene diversity ranging from 0.71 to 0.84 with overall mean of 0.79. There was a higher genetic differentiation (F_{ST}=0.21) showing the presence of moderate gene flow. The analyzed molecular variances indicated the existence of large genetic differentiation (F_{ST}=0.21) where 90% of the total variation was accounted for within populations genetic variability, leaving only 10% for the among populations variation. The PCoA, clustering, and population structure did not cluster the studied populations into a separate groups according to their geographical areas of sampling due to the presence of considerable gene flow (Nm= 2.13). In conclusion, based on the overall evaluated loci the highest intra-population diversity was observed among populations of North Gondar (Het= 0.75) and South Tigray (Het= 0.74), and hence these areas can be considered as hot spots for the identification of genotypes for breeding program. Therefore, the present study generated valuable information for sorghum breeding programs, and for conservation measures.

Keywords: Cluster, Gene flow, Genetic differentiation, Genetic distance, Landraces, PIC, Released

1. INTRODUCTION

Sorghum (*Sorghum bicolor* (L.*)* Moench 2n=2X=20) is the most substantial and stable cereal crop with more than 60 Mt of global total production every year (Mundia *et al*., 2019). It is the second most basic tropical and sub-tropical cereal crop with a total annual production of 20 Mt and over 100 million African people are depending on it (Borrell *et al*., 2001). In Africa, the major sorghum-producing countries such as Nigeria, Burkina Faso, Sudan, and Ethiopia are approximately producing 72% of the overall sorghum production (FAO, 2015). Sorghum supplies and guarantees for 40% of human calorie intake in the majority of regions of Sudan and Ethiopia (Kresovich *et al*., 2005).

Ethiopia contributed a tremendous amount of sorghum germplasm to the world sorghum collection representing the country as the center and origin of different sorghum genotypes (Upadhyaya *et al*., 2017). Basically, in the lowland areas, sorghum is the most vital cereal crop for the Ethiopian national economy (Abraham Reda, 2014). It has been reported that sorghum was domesticated in Ethiopia and nearby countries around 4000-3000 BC (Dillon *et al*., 2007). Logical proof and evidence are signifying that sorghum was first trained and cultivated in western Ethiopia (Doggett, 1988). This is additionally confirmed by the presence of different number of wild sorghum genotypes in Ethiopia (Berhane Gebrekidan, 1982). Ethiopia is the central point for a variety of sorghum genotypes and is a potential region to discover novel genotypes that could be comprehensively adaptable to different biotic and abiotic stresses (Reddy *et al*., 2009). For example, the two well recognized sorghum lines (B35 and E36-1) are local to Ethiopia and used as a source for drought tolerance genes (stay green) (Reddy *et al*., 2009), and Singh and Axtell (1973) announced two Ethiopian sorghum genotypes (IS11167 and IS11758) with high lysine content.

Likewise, sorghum varieties that originated from Ethiopia and that are superior to others have been released in Eritrea, Zambia, Burundi, Burkina Faso, and Tanzania (Reddy *et al*., 2006), and are highly contributing to the national economy in these countries.

In Ethiopian agriculture, 1,679, 277.06 hectare areas are allocated for sorghum production (CSA, 2020). Sorghum is adaptive and growing in a wide range of agro-ecologies and resulting in the development of plenty of sorghum landraces. It is produced in various regions of the country and plays an important role to meet and sustain the food demands of million populations (Roa *et al*., 2002). Nationally, sorghum is also progressively gaining importance for bioethanol energy generation (Mathur *et al*., 2017).

Furthermore, it is widely recognized that the low water requirement nature of the crop provided extra opportunity to be the most suitable and ideal cereal crop under a low input crop-growing system (Gnansounou *et al.*, 2005). It is also suitable for intensive mechanized agriculture due to seed-based planting and cultivation methods (Gnansounou *et al.*, 2005). The incidence of environmental stresses such as stem borer, molds, anthracnose, and Striga, including drought has highly constrained sorghum production (Zigale Semahegn, 2019). Subsequently, to avoid those production constraints, the effort of plant breeders and marker technology is always indispensable to understand the sorghum genetic diversity. Consequently, the rapid genetic characterization of sorghum genotypes for traits of interest is becoming the most important step to develop superior sorghum genotypes adaptive to a wide range of agro-ecologies (Huang *et al.*, 2015). Today, the collective efforts of sorghum breeders with the application of both phenotypic and molecular markers are ongoing to meet sorghum breeding objectives. However, phenotypic assessment-based sorghum yield improvement is slow, expensive, and inefficient to achieve the national sorghum breeding objectives. As a result, to improve the productivity of sorghums in the semi-arid regions of the

world, marker technologies and conventional breeding approaches need to be integrated (Cuevas and Prom, 2013).

Understanding the genetic association among sorghum relatives using modern breeding tools along with marker technology would be elemental to assist and speed up conventional breeding programs for better exploitation of sorghum genetic resources. Besides, the magnitude and importance of genetic variability assessment for crop yield improvement have been demonstrated and applied for sorghum yield improvement (Ejeta Gabissa, 2007). Consequence, various molecular diversity studies have been conducted and indicated the potential input of sorghum genetic variability analysis for economically desirable traits improvement. The presence of huge genetic diversity in sorghum has been confirmed using various marker systems including ISSR (Yangg *et al*., 1997), RFLP (Jordann *et al*., 1999), RAPD (Uptmoor *et al*., 2003), AFLP (Ritter *et al*., 2007), SSR (Cuevas and Prom, 2013), EST-SSR (Ramue *et al*., 2013), (DArT) (Macea *et al*. 2008) and SNP (Morriss *et al*., 2013).

However, the so far reported genetic structure analysis of Ethiopian sorghum relatives is insufficient to characterize the huge Ethiopian sorghum genetic resource. Consequently, there has been a limited utilization of the available genetic resource to improve agronomically important traits. Hence, there is a need to characterize, classify and utilize these novel genetic resources in the large scale sorghum breeding to boost sorghum production and productivity to ensure food security and to meet the food demand in Ethiopia. Therefore, this research was conducted to analyse the genetic structure of released and landrace Ethiopian sorghum using informative SSR markers to generate valuable information for future sorghum breeding and yield improvement program.

1.1 General Objective

- To investigate the genetic structure of released and landrace Ethiopian sorghum genotypes using SSR markers.

1.2 Specific objectives

- To assess the genetic diversity of released and landrace sorghum genotypes of Ethiopia using microsatellite markers.
- To determine the population structure of released and landraces of sorghum genotypes.

2. LITERATURE REVIEW

2.1 Origin and Distribution of Sorghum

The use of wild sorghum as food grain was started in the Sahara regions (Venkateswaran *et al*., 2019). Likewise, the fourth millennium BC has a historical indication that eastern Sudan is the first place where the earliest sorghums originated and were domesticated for cultivation (Winchell *et al*., 2017). It was notorious and controversial to know and point out the exact regions for sorghum domestication and cultivation (Venkateswarann *et al*., 2019), however, according to the archaeological report eastern Sudan was predicted as the potential place for sorghum domestication (Fuller, 2018). On the other side, according to some recent reports and implications the event of domestication may probably be from more than a single geographical place, for example, the gnus *guinea-margaritiferum* was originated from multiple geographical sites (Macee *et al*., 2013). Based on the archaeological evidence, the commonly distributed and cultivated sorghum races such as *S. bicolor* originated from its wild relative *Sorghum bicolor* (L.) Moench subsp. *verticilliflorum* (Doggett, 1988). According to Doggett (1988), three major hypothesis were projected to confirm and clarify the exact sorghum origin and domestication site. The first hypothesis declared that West Africa is the possible place of sorghum origin and domestication. The second hypothesis is that sorghum has originated and domesticated in eastern Sahara and the final proposed hypothesis stats that sorghum. Race durra has originated and domesticated in India.

Sorghum was physically distributed into various agro-ecologies from its first ancestor that originated in Africa into the globe mostly by trade routes. Also, human migration was another way for sorghum distribution and circulation, which is originated from east Africa across southern and eastern Africa (Mann *et al*., 1983). Consequently, it was re-introduced to the Middle East towards India through trade

routes (Mann *et al*., 1983). Afterward, sorghum becomes popular and playing a significant economical role in the Indian agriculture system (Kleih *et al*., 2000) and then India got recognized as the second center for sorghum diversity (Appa *et al*., 1996).

Furthermore, the introduction of sorghum into different regions of China from India continued through sea and territorial trade routes. River Indochina was the one responsible for sorghum race distribution between China and India (Venkateswaran *et al*., 2019). Genetically, the eastern African sorghum relatives are related to Amber cane whereas the introduced Indian *Sorghum bicolor* is the progenitor of the Kaoliangs (Doggett, 1988). There is also an archeological confirmation that explains that Kaoliangs might have originated from the local wild diploid sorghum relatives (Harlan, 1995). Based on the evidence of archaeological studies report Yellow River Valley was the anticipated geographical area where the most primitive sorghum races were originated and cultivated (Venkateswaran *et al*., 2019). In the 19th century, the sorghum race Chinese Amber was introduced from China into the USA with the help of slave traders (Martin, 1936). In the 1900s, Americans were the most responsible to introduce Sorghum races into Queensland, Australia (Venkateswaran *et al*., 2014). Finally, sorghum became the foremost important summer crop in Australia and responsible for 5% of the global sorghum export (Venkateswaran *et al*., 2019).

2.2 Taxonomic Classification and Race

Taxonomically, sorghum belongs to the kingdom, class, order, family and genus, of *Plantae*, *Liliopsida*, *Cyperales*, *Poaceae,* and *Sorghum* respectively (Hariprasanna, 2015). According to Snowden's taxonomy, which was based on a phenotypic trait, sorghum race has been classified into two sub groups such as *Eu-sorghum* and *Para-sorghum* (Snowden, 1955). Nevertheless, the domesticated sorghum relatives are now grouped into five sub-genera according to their

morphological features: *Eu-sorghum*, *Stiposorghum*, *Hetero-sorghum*, *Chaeto-sorghum*, *Para-sorghum* (Lazaridess *et al*., 1990) (Figure 2.1). Even though the *Sorghum bicolour (L.)* races were domesticated and cultivated in Eastern Africa, now a day 17 *Sorghum* sub species are totally local and indigenous to South Australia and 13 of them are endemic. In this diverse genus, knowing the definite number of sorghum sub species is notorious and not clearly documented yet. Based on Diillon *et al*., (2001) report, the 25 *Sorghum* sub species are geographically distributed mostly in the three continents such as (Australia, Asia, and much of in Africa). For instance the species *Kuntze*, that indigenous and local to Mexico are growing only in limited agro ecological locations (Spangler, 2003) with only limited taxonomic information. Kew's Angiosperm DNAC-values data base, currently lists and provided taxonomic information for a 32 Sorghum sub species (Leitcch *et al*., 2018). These differing taxonomic classifications are referring different morphological parameter that is why sorghum taxonomy is a complex process and controversial area for scientific and taxonomic study (Figure 2.1). Now a day, there are 24 accepted sorghum species ((USDA, 2020)

According to Venkateswarann *et al*. (2014) report, taxonomic classification was by referring the agro-morphological features and is less important and accurate as it causes taxa and genus overlapping. However, the recent studies which is based on the molecular parameters, such as DNA sequencing data, was able to classify the species with clear and precise grouping systems (Venkateswarann *et al*., 2014). The molecular based diversity analysis confirmed the occurrence of large genetic variation describe Ethiopia as a centre for sorghum origin (Vavilov, 1951).

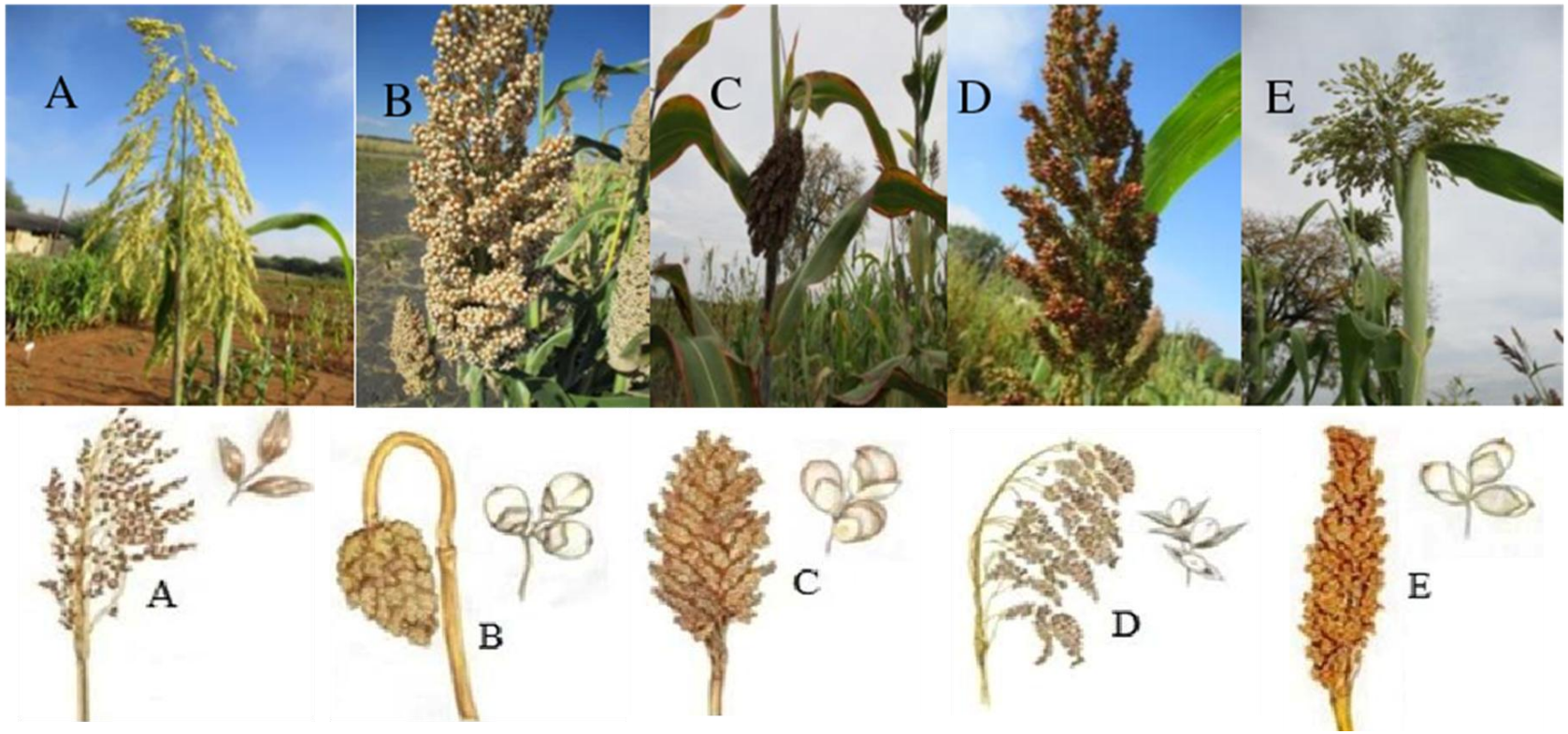


Figure 2-1 Sorghum morphological Species diversity showing the five sorghum races: (Harla, 1972)

Hence, all reported proof indicated that the east part of Africa, particularly Ethiopia, the centre for many sorghum species. Therefore, the greatest genetic variability of different sorghum is local to Ethiopia and the surrounding eastern African countries. Ethiopia is known as the center of origin and diversity of sorghum (Doggett, 1988).Therefore, this gives an opportunity to look for useful genes from our diverse sorghum landraces to address production environmental constraints of the crop.

2.3 Sorghum production and use

Sorghum is ranked 5th based on its annual production around the world with over 60 Mt (Mundia *et al*., 2019). In Africa, sorghum is the second most elemental cereal crop with a total of 61% production area and 41% production to secure the food demands of over 100 million African people (Mundia *et al*., 2019). Nigeria and Burkina Faso together with Sudan and Ethiopia are the fore most sorghum-producing countries of the continent, whose production is approximately 72% of the total sorghum production in Africa (FAO, 2015). More specifically, in most regions of Sudan and Ethiopia, sorghum sourced about 40% of human calorie intake (Kresovich *et al.*, 2005).

In Ethiopia, sorghum is the second most important cereal crop, mostly in the low land areas where rainfall is irregular and unpredictable (Abraham Reda, 2014). From the total cultivated area, sorghum production covers about 1,679, 277.06 hectare areas for sorghum production (CSA, 2020). In Ethiopia, sorghum is used for daily human consumption like flatbread (Kita), “injera”, and also to brew alcoholic drinks (CSA, 2017). The straw is also used for animal feed. The crop is commonly adaptive to a wide range of environments ensuing in the evolution of ample landraces cultivated in various production regions of the country (Roa *et al*., 2002). This crop is growing in a wide range of agro-ecologies and playing a significant role as a

stable food grain for millions of food-insecure populations. In addition to its use as grain, sorghum is also gradually becoming useful for bioethanol production (Mathur *et al*., 2017).

Sorghum is the most appropriate and ideal cereal crop under low input agriculture due to the low water requirement nature of the crop (Gnansounou *et al.*, 2005). The crop is cultivated during the warm season because it is highly adapted to drought, high temperature, and low soil fertility that make the crop major food grain in Arid and semi-Arid environments. Sorghum uses a C4 photosynthesis mechanism that helps for the survival of the crop under harsh environmental conditions by improving its carbon assimilation during high temperatures. Besides, it can reduce water loss by widening its root system and by having waxy bloom on the leave. Seed-based cultivation also makes it suitable for intensive mechanized agriculture (Gnansounou *et al.*, 2005). Sorghum production is highly constrained by both biotic stresses such as stem borer, molds, anthracnose, and Striga, and abiotic stresses such as drought (Zigale Semahegn, 2019).

2.4 Sorghum Genetic Diversity

To develop superior sorghum genotypes, genetic characterization is the most ideal and important step for genetic improvement with particular traits of interest (Huang *et al.*, 2015). Even though, sorghum improvement based on phenotypic assessment is slow and inefficient (Menkir *et al*., 1997), sorghum breeders with the development and the joint application of phenotypic and molecular markers are underway to meet the national sorghum breeding objectives.

These days, new marker technologies and approaches are emerging to be integrated with the conventional breeding approaches to improve the productivity of sorghums and to meet the food demand for millions of lives in the world. Also, to improve food security, better utilization of the

sorghum genetic resources using modern breeding approaches is fundamental to support conventional breeding programs by assessing the genetic relationship among sorghum genotypes.

Various genetic diversity and genetic distance studies have been conducted based on phenotypic, biochemical, and genetic markers. In the past, the utilization of biochemical markers such as isozymes and functional proteins was comprehensively used to assess the plant genetic resources. The outcome indicated the significance of sorghum genetic diversity analysis for crop improvement. Consequently, today different types of genetic markers are being applied to characterize and evaluate the genetic diversity of sorghum. For instance, Randomly Amplified Polymorphic DNA (RAPDs) (Uptmoor *et al*., 2003); Restriction Fragment Length Polymorphism (RFLP) (Jordan *et al*., 1998); Inter Simple Sequence Repeats (ISSRs) (Yang *et al*., 1996); Amplified Fragment Length Polymorphism (AFLP) (Ritter *et al*., 2007); Simple Sequence Repeat (SSR) markers (Cuevas and Prom, 2013); Expressed Sequence Tage Simple Sequence Repeat (EST-SSR) markers (Ramu *et al*., 2013); (DArT) (Mace *et al*. 2008) and Single Nucleotide Polymorphism (SNP) markers (Morris *et al*., 2013). Those studies reported the existence of moderate to high sorghum genetic diversity. Accordingly, genetic characterization of sorghum is a good demonstration of the importance of genetic diversity assessment for crop improvement (Ejeta Gabissa, 2007).

However, for genetic diversity analysis, the choice of molecular markers depends on cost, time, marker heritability, and quantity of DNA required, technical labour, and degrees of polymorphism, number of loci, precision, resolution, environmental stability, and statistical power of the tests (Garcia *et al*., 2004). Therefore, for successful genetic diversity analysis, there

is no single and recommended method. However, it is frequent and popular to use DNA-based genetic markers, which are more reliable and robust methods for crop genetic diversity analysis (Amsalu Ayana *et al*., 2001).

Simple Sequence Repeats (SSR) also named microsatellites; which are PCR-based markers with short tandem repeats (STRs) and short nucleotide motifs (2-6 bp/ntd). These repeat motifs can be di-, tri- and tetra, e.g. (AT)n, (TAT)n, and (CATA)n, which are commonly dispersed right through the genomes of the crop plants. Their repeat varies across individuals and leads to polymorphic genetic information in the plant's genome. The flanking DNA sequences surrounding the microsatellite regions are typically conserved, and specific genetic primers which are specific for the microsatellite regions are considered to be used for PCR amplification and genotyping. The prior sequence information is required for primer design.

SSR marker requires DNA in nanogram, moderate DNA quality, is co-dominant, produces higher polymorphism, highly reproducible, 1-3 loci can be analyzed, and are easy to use (Nadeem *et al*., 2018). They are quite many in eukaryote genomes and present ample genetic information and are hyper-variable, locus-specific, and multi-allelic, are reasonable cost-wise, and are applicable in genetic studies, and can be analyzed by fast and simple PCR (Pereira *et al*., 2013).

Thus, SSR marker is popular to analyze and quantify genetic diversity, and finding genetic relationships among collections originated from a diverse geographical location for effective utilization of genetic resources (Xu *et al*., 2016). Practically SSR markers are being applied for characterization and fingerprinting of the natural genotypes (Menz *et al.*, 2004), evaluation of genetic distances and identity among populations (Asfaw Adugna, 2014), for agronomically

important QTL detection, linkage analysis and to discover agronomically important traits from wild sorghum relatives through linkage map approaches (Klein *et al.,* 2001).

Therefore, of these DNA markers, SSR markers are the foremost important and popular markers of choice under the current breeding programs. As a result, the conducted genetic diversity studies on Ethiopian sorghum relatives using SSR markers indicated the presence of considerable sorghum diversity mostly within the populations collected and cultivated from diverse agro-ecologies. Classification of Ethiopian sorghum relatives based on adaptation zone is more reliable than the center of origin and confirmed that Ethiopia is both centers of diversity and domestication for the sorghum race (Doggett and Rao, 1995).

Also, a total of 80 sorghum relatives collected from Ethiopia and Eritrea has been genetically analyzed using RAPD markers and showed ample genetic diversity within ecologies than origins indicating that the existence of weak sorghum genetic differentiation based on regional bases, confirming the existence of out crossing between cultivated sorghum and weedy relatives, plus seed exchange by trade routes (Amsalu Ayana *et al*., 2000).

The previous study on a cluster and population structure analysis using Ethiopian sorghum materials established at the USDA using SSR markers showed the presence of high sorghum genetic variability in Ethiopia and confirmed the event of gene flow among highly related sorghum relatives of the sub-population (Cuevas and Prom, 2013). Moreover, there has also been reported by other genetic studies about the presence of wider sorghum (bi-color) genetic diversity in Eastern Africa regions, for instance, the SSR genetic studies conducted for the sorghum materials collected from Eastern Kenya (Muui *et al*., 2016), Eritrea (Tesfamichael Abraha *et al*.,

2014), and Zambia (N G'Uni *et al*., 2011) provided additional insight for sorghum domestications and origins. Likewise, SSR-based analysis conducted by Tesfaye Disasa *et al*. (2016) reported the existence of high genetic diversity within the Ethiopian sweet sorghum materials and suggesting that Ethiopia might be the center of diversity for sweet sorghum relatives too.

The foremost important factors that have contributed to the higher genetic variability among sorghum populations are the existence of diverse agro-ecologies, gene flow among wild and cultivated relatives, and patterns of out breeding among sorghum landraces (Yao *et al.*, 2004). The event of national and local seed exchange and the growing of varietal sorghum mixtures, the agricultural system may also contribute to these variations (Frew Mekbib, 2008). Besides, collecting leaf samples from different plants during DNA extraction may be responsible for the variation among plants since landraces are highly diverse (Cuevas and Prom, 2013).

Therefore, according to Missihoun *et al*. (2015) implication, conducting genetic diversity analysis based on the botanical and morphological future of the crop plant could support the better conservation and utilization of the crop genetic resources.

3. MATERIALS AND METHODS

3.1 Plant Materials

The study involved a total of 92 (48 landraces and 44 released) sorghum genotypes (Appendix. 1). The landraces were collected from five Administrative zones of Ethiopia (South Tigray, North Wello, North Gondar, East Hararge, and West Hararge) (Fig. 3.1). Also, some released sorghum genotypes were obtained from the federal and regional agricultural research centers including Mehoni, Serinka, and Gondar Agricultural research centres. Their local and breeding names are presented in Appendix 1.

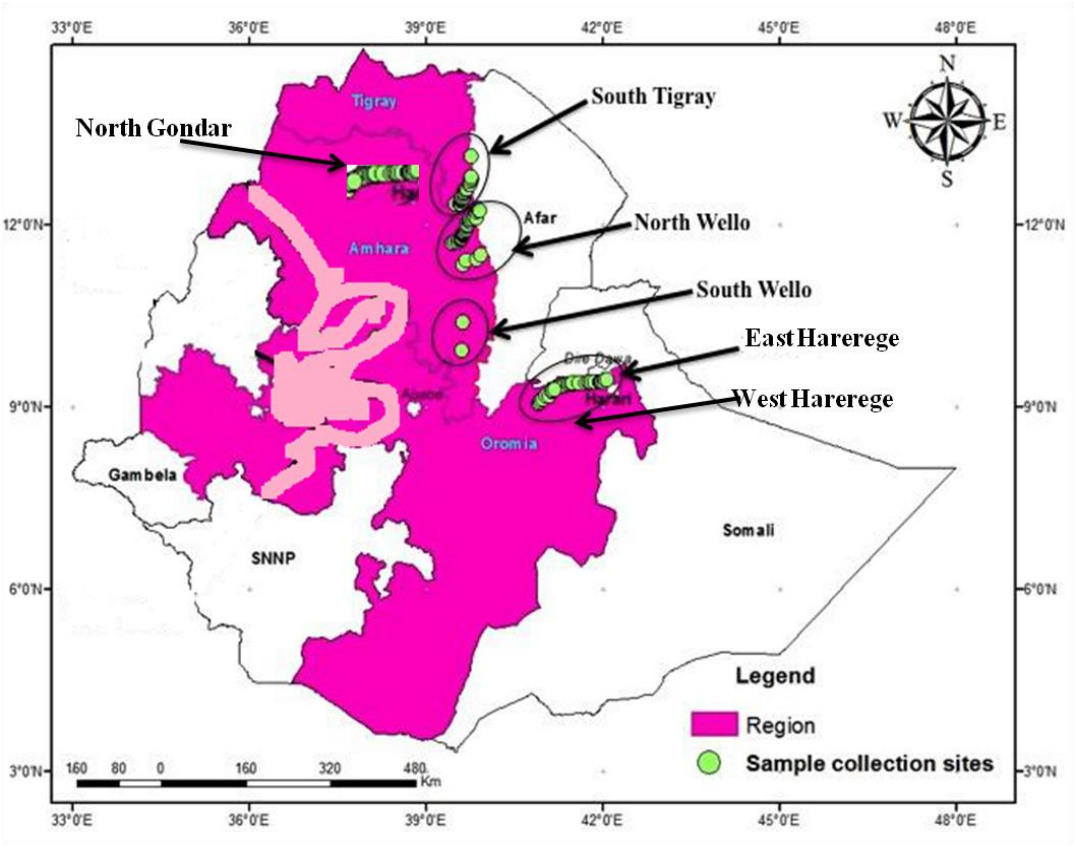


Figure 3-1 Map of Ethiopia showing sorghum collection sites and regions

3.2 DNA Extraction

Seeds of each genotype were grown in a plastic pot of 15 cm x 8cm size filled with mix of compost, soil and sand at 1:1:1 (v/v/v) ratio, respectively at greenhouse condition of National Agricultural Biotechnology Research Center (NABRC), Holeta, Ethiopia following (Mace *et al*., 2012) (Appendix 2. A). For genomic DNA extraction, fresh young leaves were collected from two to three-week-old seedlings. Between each sample, used scissors were cleaned using 70% ethanol to avoid samples cross-contamination and surface contaminates. Collected leaves were dried using silica gel at room temperature. About 50 ng of dried leaf samples of each sample was grinded in the presence of metal beads using Geno Grinder (MM-200, Retsch). The genomic DNA extraction was carried out following the modified CTAB procedure as described by Mace *et al*. (2012). The extracted genomic DNA quality and concentration were checked using a Nano-drop Spectrophotometer (ND-8000, 8-sample Spectrophotometer, Thermo Scientific). The presence of high molecular weight DNA was confirmed by running a mix of 5 µL of DNA and 2 µL of 1X loading dye containing gel red on 0.8% agarose gel at 100 V for 40 minutes. The gel was visualized under UV light and subsequently photographed. A 1 kb ladder was used to estimate the molecular size of the extracted genomic DNA (Appendix 3. B). Portion of the extracted genomic DNA was diluted to 50 ng/µl working solution using 1xTE buffer and stored at -20 °C until used for PCR.

3.3 Polymerase Chain Reaction and Electrophoresis

A total of 12 SSR pairs of primers were used in this study (Table 3.1). The SSR regions of the sorghum genotype were amplified in total volume of 13 µl reaction mixtures composed of 1 µL of each primer pair, 2 µl of template DNA of 50 ng/µl, 3 µl of nuclease-free H_2O, 6 µl of GoTaq® Master Mixes). The PCR program followed an initial denaturation of 4 min at 95 °C,

followed by 35 cycles of denaturation for 30 s at 94 °C, primer annealing for 45 s at 49-61 °C, primer extension for 30 s at 72 °C and final extension for 10 min at 72 °C, and holding at 4 °C (Billot *et al*., 2013).

For SSR-based genotyping, polyacrylamide gel electrophoresis was used (Wang *et al.,* 2009). The accuracy of the PCR product was quantified using 1.2 % agarose gel (Appendix 2. D, E), and Polyacrylamide gel was used to count alleles and to see the patterns of SSR in the genome of the entire genotypes (Cleaver CS500 volt) (Appendix 2. F). A 100 bp DNA ladder was used to estimate the amplification fragment size. Finally, the gel was visualized under UV light and subsequently photographed using a BioDoc-It TM Imaging System.

Table 3.1. Primer sequences, annealing Temperature and their range of fragment size detected at the 12 microsatellite loci

S/No	Marker-Name	Motive	Forward Primer (5´ to 3´)	Reverse Primer (5´ to 3´)	Ta	Size (bp)	Chrom*
1	Xtxp -298	(TG)9	GTTCTTGGCCCTACTGAAT	TCACCTGTAACCCTGTCTTC	55	108-112	SBI08
2	Xtxp -034	(GT)10	TAGTCCATACACCTTTCA	TCTCTCACACACATTCTTC	49	170-180	SBI08
3	Xtxp -284	(AG)16	TGCCACCACTCTGGAAAAGGCT	CTGAGGACTGCCCCAAATGTAG	55	183-203	SBI06
4	Xtxp-312	(TC)12	CCCATAATACTTGACCTTC	ACTTACTCCCTCTGTCCC	50	189-197	SBI06
5	Xtxp -023	(AC)9	GCTTCTATACTCCCCTCCAC	TTTATGGTAGGATGCTCTGC	55	112-134	SBI01
6	Xtxp- 270	(TG)9	ATCAGGTACAGCAGGTAGG	ATGCATCATGGCTGGT	50	165-173	SBI01
7	Xtxp -211	(GT)9	TTGAGAGCGGCGAGGTAA	AAAAGCCCAAGTCTCAGTGCTA	61	104-110	SBI07
8	Xcup02	(GCA)6	GACGCAGCTTTGCTCCTATC	GTCCAACCAACCCACGTATC	54	192-204	SBI09
9	Xgap206	(AC)13/(AG)20	ATTCATCATCCTCATCCTCGTAG	AAAAACCAACCCGACCCACTC	55	106-146	SBI09
10	Xcup63	(GGATGC)4	GTAAAGGGCAAGGCAACAAG	GCCCTACAAAATCTGCAAGC	54	133-145	SBI02
11	Xgap84	(AG)14	CGCTCTCGGGATGAATGA	TAACGGACCACTAACAAATGAT	55	183-217	SBI02
12	Xtxp114	(AGG)8	CGTCTTCTACCGCGTCCT	CATAATCCCACTCAACAATCC	50	211-217	SBI03

Source: Billot *et al*. (2013). Note: Ta= Annealing temperature.

3.4 Data scoring and analysis

The clear and visible amplified PCR products generated by the 12 microsatellite loci across the genome of the 92 sorghum genotypes were scored and the resulted data matrix were subjected to diverse genetic diversity analysis software programs. The natural genetic variation and population structure of the sorghum materials were evaluated using an admixture model assuming that each genotype has ancestry from genetically diverse sources and population clustering was computed using STRUCTURE 2.3.3 (Pritchard *et al*., 2003). To confirm the anticipated population cluster and genetic structure a 500 burn-in period and 50,000 Markov Chain Monte Carlo (MCMC) iterations were used based on the 12 SSR markers which were scattered across the sorghum genome. The number of sub-populations for the studied materials was computed using K-value ranging from 1 to 10 using 20 iterations for each K. From the STRUCTURE run output, the most likelihood K value was recognized following the simulation method of Evanno *et al.* (2005) by using an online structure harvester (Earl and von Holdt, 2012). A phylogenetic tree analysis based on dissimilarity index was performed using Darwin 6.0 software (Perrier *et al*., 2003). Analysis of Molecular Variance (AMOVA) among and between populations, estimated allele frequencies and estimated heterozygosity over population for each locus, number of alleles (N), number of effective alleles (Ne), expected heterozygosity (He), and observed heterozygosity (Ho), pairwise genetic distance (GD) and genetic identity (GI) were analyzed using GenAlEx 6.5 (Peakall and Smouse, 2012). The number of migrants per generation was considered to estimate gene flows (Nm) as (1-FST)/4FST. Furthermore, major allele frequency (MAF), heterozygosity, polymorphic information content (PIC), and the total number of alleles were calculated using a PowerMarker 3.25 (Liu and Muse, 2005).

4. RESULTS

4.1 Marker Polymorphism

In the present study, a total of 77 alleles were detected across the genome of the 92 sorghum genotypes as revealed by the 12 microsatellite loci (Table 4.1). The generated number of alleles ranged from five (xtxp-284, xtxp-270, and xtxp-211) to eight (xtxp-312) with a mean number of 6.42 alleles per locus across the genome of the entire populations. The major allele frequency (MAF) ranged from 0.18 (xtxp-298) to 0.45 for the two microsatellite loci (xcup02 and xgap84) with overall mean of 0.30.

The largest Nei's gene diversity (GD = 0.84) and polymorphic information content (PIC = 0.82) were observed for three loci: (Xcup63, Xtxp-298 and Xtxp-312) While, the lowest gene diversity (0.71) PIC (0.66) were recorded for Xcup02 (Figure 4.1). The highest effective number of alleles (Ne=1.22) and Shannon information index (I = 0.52) were recorded for Xtxp-023 and Xgap206, respectively (Table. 4.1). The heterozygosity of the studied materials ranged from 0.15 (xgap206) to 0.48 (xcup02) with overall mean value of 0.29 (Table 4.1). In terms of the overall polymorphic information content, all the microsatellite loci used were found to be highly informative (PIC > 0.5) (Table 4.1). Likewise, the highest fixation index (Fst = 0.14) and the lowest value for gene flow (Nm =1.60) were observed for Xtxp-211 (Table 4.1). Moreover, three of the microsatellite loci (Xtxp-211, Xtxp-312 and Xtxp114) significantly deviated from the HW-equilibrium over the entire populations (Table 4.1).

Table 4.1. Genetic diversity summary statistics of the 12 SSR markers across the sorghum genotypes

Marker	MAF	NA	Ne	GD	Het	PIC	I	Fst	Nm	P-value	P_{HWE} [a]
Xcup02	0.45	6.00	1.05	0.71	0.49	0.66	0.09	0.06	3.64	0.035	0.001***
Xgap206	0.28	7.00	1.57	0.81	0.15	0.78	0.52	0.04	6.55	0.012	0.045*
Xcup63	0.20	7.00	1.01	0.84	0.21	0.82	0.03	0.06	3.65	0.031	0.025*
Xgap84	0.45	7.00	1.09	0.74	0.11	0.72	0.12	0.07	3.31	0.035	0.021*
Xtxp114	0.22	7.00	1.06	0.82	0.28	0.80	0.15	0.06	3.79	0.052	0.088ns
Xtxp-298	0.18	7.00	1.12	0.84	0.17	0.82	0.17	0.03	7.87	0.005	0.029*
Xtxp-034	0.33	7.00	1.08	0.78	0.20	0.75	0.14	0.07	3.55	0.021	0.021*
Xtxp-284	0.32	5.00	1.06	0.73	0.63	0.68	0.10	0.07	3.16	0.048	0.025*
Xtxp-312	0.26	8.00	1.10	0.84	0.19	0.82	0.17	0.06	3.67	0.034	0.113ns
Xtxp-023	0.30	6.00	1.22	0.79	0.44	0.76	0.31	0.06	3.70	0.029	0.002**
Xtxp-270	0.27	5.00	1.07	0.77	0.27	0.73	0.22	0.05	5.10	0.021	0.025**
Xtxp-211	0.29	5.00	1.14	0.78	0.40	0.75	0.18	0.14	1.60	0.085	0.072ns
Mean	**0.30**	**6.42**	**1.131**	**0.79**	**0.29**	**0.76**	**0.18**	**0.07**	**4.13**	**0.034**	**0.314**

MAF=Major Allele frequency; NA = Total number of Alleles; Ne= Number of effective alleles; GD= Gene Diversity; Het= Heterozygosity; PIC=polymorphic information content, I =Shannon´s index, Fst = Fixation index, Nm =Gene flow , *p* = Differentiation statistics probabilities, P_{HWE} [a] = *p*-value for deviation from Hardy Weinberg Equilibrium, ns = not significant, ***** = $p < 0.05$, ****** = $p < 0.01$ and ******* = $p < 0.001$.

4.2 Genetic variation within and among populations

Among the five tested populations the highest number of alleles (Na = 6.25), number of effective alleles (Ne =4.33), Shannon information index (I =1.54), heterozygosity (He =0.74), number of private alleles (No. PA =0.42) and Number of local common alleles (>=50%) (NLCA=0.75) were recorded for the population from South Tigray (Table 4.2). The populations from East Hararege (EH) and West Hararge (WH) showed similar lower number of alleles, effective number of alleles and other detected population diversity values. Analysis of percentage of polymorphic loci (PPL) revealed that all of the loci were 100% polymorphic in each population studied.

Table 4.2. Estimated allele frequencies and heterozygosity over population

Population	Na	Na Freq. >= 5%	Ne	I	No. PA	NLCA	OHet	EHet
EH	4.25	4.25	3.31	1.26	0.00	0.25	0.68	0.73
WH	4.25	4.25	3.65	1.33	0.00	0.33	0.71	0.79
ST	6.25	5.00	4.33	1.54	0.42	0.75	0.74	0.75
NW	5.08	4.08	3.53	1.33	0.08	0.58	0.68	0.71
NG	5.92	4.58	4.17	1.52	0.25	0.50	0.74	0.76
Mean	5.15	4.44	3.8	1.4	0.15	0.49	0.71	0.75

EH= East Hararege, WH= West Hararge, ST = South Tigray, NW= North Wollo, NG= North Gondar, OHet=Observed Heterozygosity, EHet = Expected Heterozygosity, Na= number of alleles; Ne= number of effective alleles; I=Shannon information index; No.PA= number of private alleles and NLCA = No. Local Common Alleles (>=50%))

4.3 Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance (AMOVA) and variance components were computed to partition the total natural genetic variation existing within and among the studied sorghum population. Hence, based on the administrative zones, evaluating the genetic diversity of the 92 sorghum genotype using 12 microsatellite loci revealed that 10% of the total genetic variation was accounted by the variation among populations (Table 4.3). The analysis also confirmed that within populations genetic differences explained 90% of the total genetic variation. The study confirmed the presence of higher genetic differentiation (FST: 0.21) among the populations with higher values of overall genetic differentiation coefficient (PhiRT = 0.105, p = 0.001) and gene flow (Nm: 2.13) (Table 4.3). On the other side, based on released and landrace classification, significant overall genetic differentiation coefficient among the populations (PhiRT = 0.06, p = 0.001) and gene flow (Nm: 3.89) were observed in which 6% and 94% of the total genetic variation were explained by among and within populations variations, respectively (Table 4.3).

Table 4.3. Summary of AMOVA as revealed by 12 microsatellite loci

Components	Source of V	d.f	Sum of Square	Mean Square	Est.Var	Percentage Variation (%)	F-stat	P-value	
Administrative Zone	Among Pops	4	7.909	1.977	0.028	10	0.22	0.001	PhiRT = 0.06
	Within Pops	87	147.500	1.603	1.603	90	0.34	0.001	
	Total	91	248.027		1.632	100	0.29	0.001	
Gene flow (Nm)	2.13								
Released and Landraces	Among Pops	1	24.357	24.357	0.39	6	0.21	0.001	PhiRT =0.105
	Within Pops	90	555.068	6.16	6.16	94	0.31	0.001	
	Total	91	579.435		6.56	100	0.29	0.001	
Gene flow (Nm)	3.89								

Note: F-stat =F-statistics across all evaluated genomic loci: PhiRT = 0.105 and between released and landraces, overall genetic differentiation coefficient among the populations (PhiRT = 0.06, *p*-value= 0.001).

4.4 Genetic distance between the populations

The pairwise population genetic distance and gene flow for both landrace and released sorghum genotypes are presented in Table 4.4. Generally, lower genetic distances were observed between pairs of populations. The pairwise Nei's standard genetic distance between populations varied from 0.009 (between WH and NG) to 0.013 (between WH and NW) (Table 4.4). A higher gene flow (Nm=11.32) was observed between the sorghum populations collected from North Gondar and South Tigray. Conversely, the lowest (1.80) gene flow was observed between the populations collected from East Hararege and North Wollo (Table 4.4). Similarly, the genetic differentiation (Fst) between pairs of population was the lowest (0.01, p=0.05) between the populations West Hararge and East Hararge, and the highest (0.22, p=0.02) between East Hararge and North Wollo (Table 4.5).

Table 4.4. Pairwise Population comparison on Nei Genetic Distance (below the diagonal) and Gene flow (above the diagonal)

	EH	WH	ST	NW	NG
EH	------	6.178	5.815	1.796	4.36
WH	0.005	------	2.34	3.05	6.75
ST	0.003	0.004	------	5.62	11.32
NW	0.010	0.013	0.006	------	4.91
NG	0.006	0.009	0.003	0.010	------

Note: EH= East Hararege, WH= West Hararge, ST= South Tigray, NW= North Wollo, NG= North Gondar.

Table 4.5. Pairwise Population fixation index (Fst) (below the diagonal) and *p*-values (above the diagonal)

	EH	WH	ST	NW	NG
EH	--	0.09	0.04	0.02	0.001
WH	0.04	--	0.05	0.03	0.01
ST	0.04	0.01	--	0.01	0.02
NW	0.22	0.08	0.04	--	0.03
NG	0.05	0.04	0.02	0.05	--

Note: EH= East Hararege, WH=West Hararge, ST = South Tigray, NW = North Wollo and NG= North Gondar.

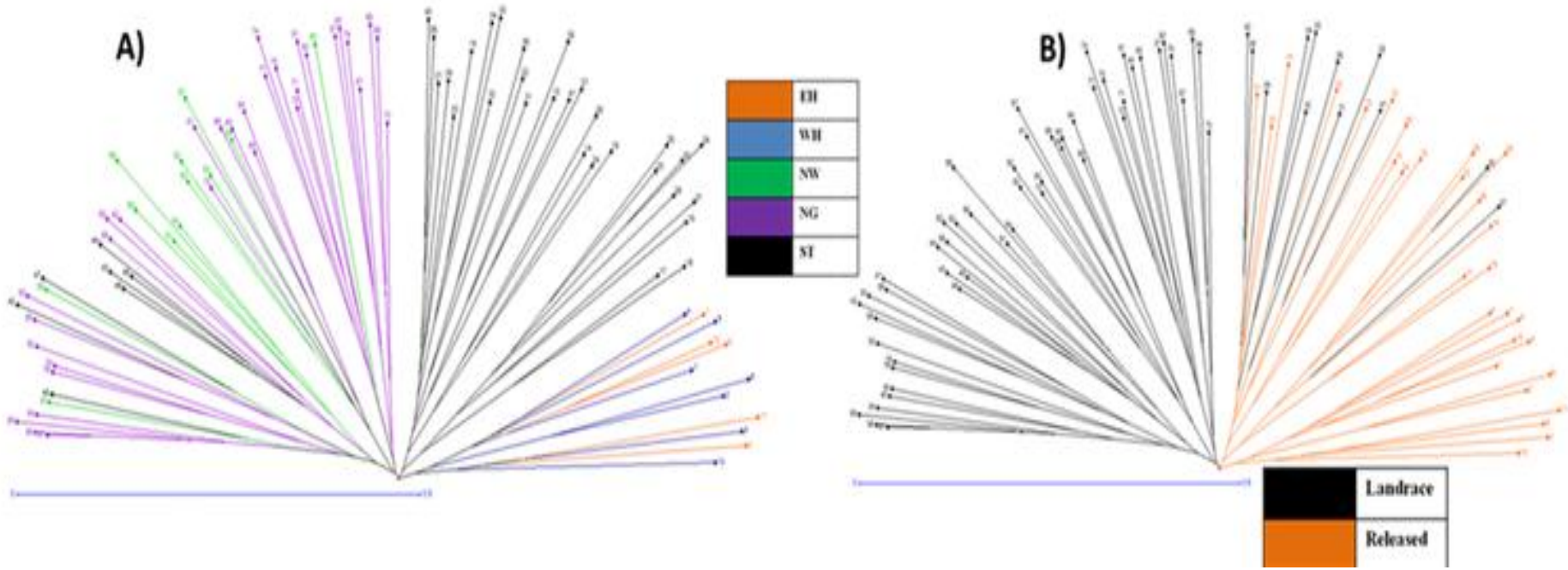
4.5 Population structure and cluster analysis

Cluster analysis based on Unweighted Neighbor-joining method clustered the study materials into three major groups (Figs. 4.1A, 4.2A). It was observed that individuals from North Wello and North Gondar formed the same cluster, the second cluster comprised of populations of East Hararge and West Hararge and the third cluster contained only the populations of South Tigray (Figs. 4.1A, 4.2A). Clustering based on breeding materials resulted two groups where most of the released (red) genotypes were clustered together with few mix from landraces, and the second cluster composed of the landraces (black) (Figs. 4.1B, 4.3B).

The conducted principal component analysis (PCoA) results indicated that 23% of the total variation among the population was explained by the three-axis coordinates (Fig. 4.3). However, PCoA did not cluster the individuals sharply according to their geographical areas of sampling.

There was some sort of overlapping between the populations of East Hararge and West Hararge. The populations of North Gondar showed admixture with populations of North Wollo and South Tigray. Clustering based on breeding material grouped the samples nearly into two separate clusters as released and landraces with some sort of admixture between the two (Fig. 4.3).

Similarly, The Bayesian model-based population structure analysis using the STRUCTURE output with harvester online tool predicted k=3 to be the most likely number of clusters labeled as sub-population one (SP1), sub-population two (SP2) and sub-population 3 (SP3) (Fig 4.1 D).


A)
EH
WH
NW
NG
ST
B)
Landrace
Released

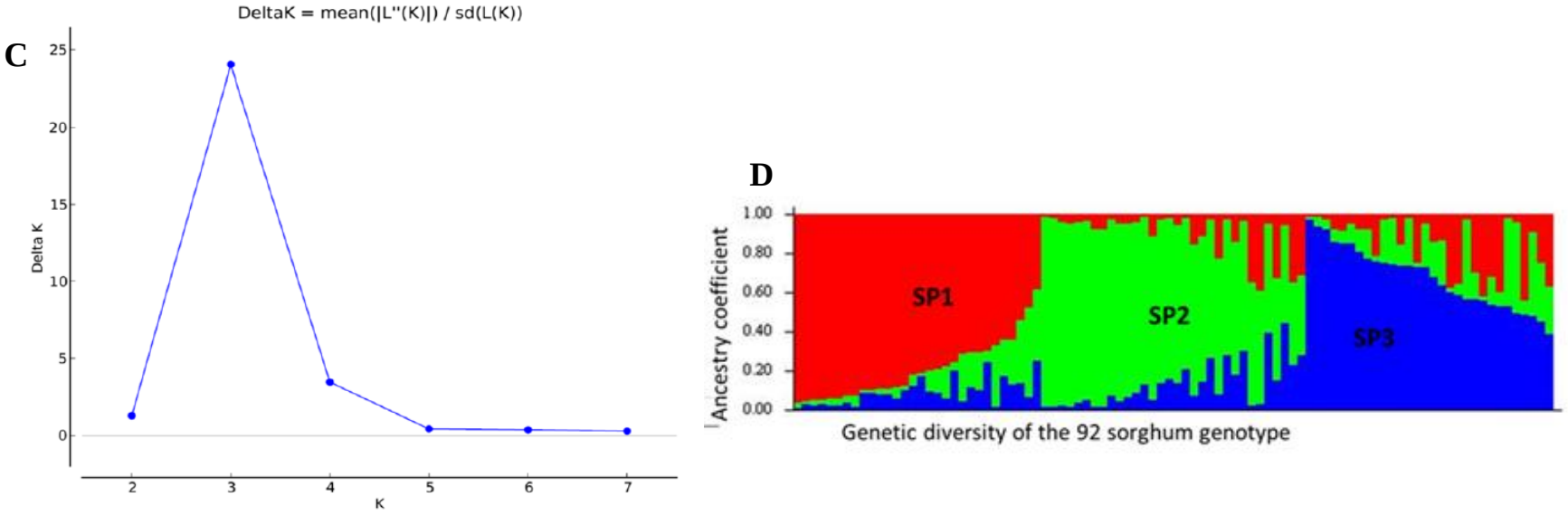
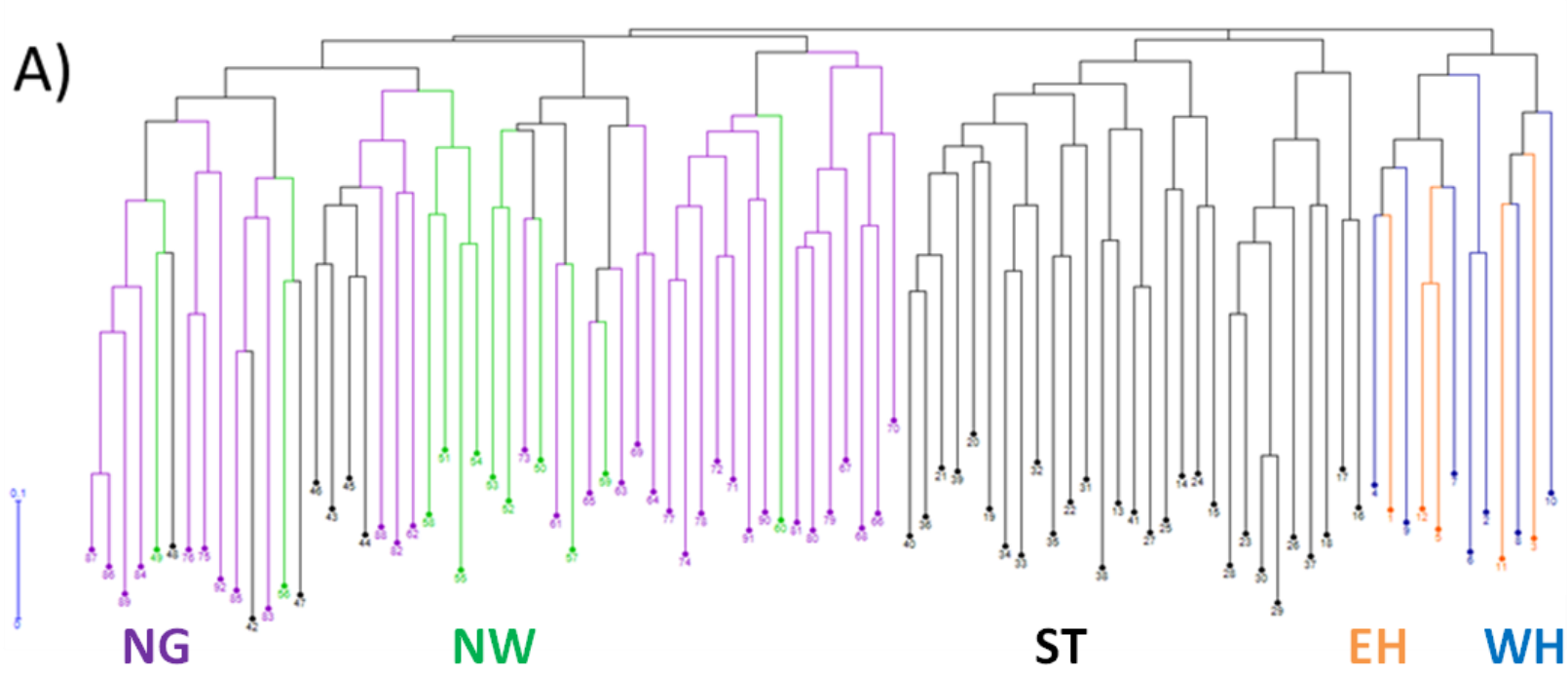


Figure 4-1. Population structure and cluster analysis. A) UPGMA based clustering of sorghum material :East Hararge (Orange) and West Hararge (blue) clustered together, and North Wollo (green) and North Gondar (violet) clustered together, and southern Tigray (black) clustered alone, B) Clustering based on breeding materials: released (red) and landrace (black), C) Maximum likelihood of delta value, and D) Population structure.


A)
NG
NW
ST
EH
WH

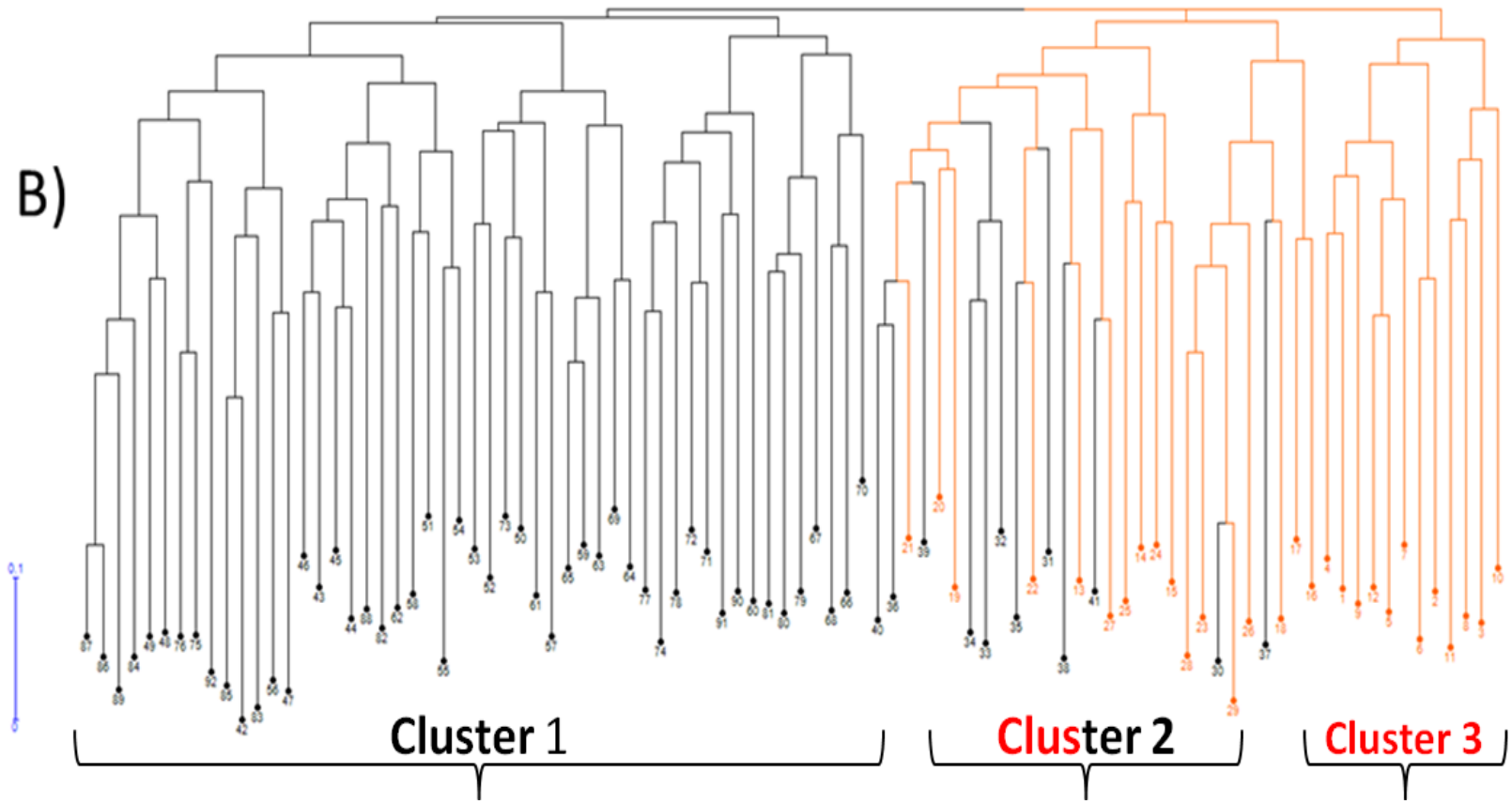


Figure 4-2. Dendrogram for 92 Ethiopian sorghum genotypes based on Unweighted Neighbor-joining. A) Clustering based on the Administration zones : East Hararge (red) and West Hararge (blue) clustered together, and North Wollo (green) and North Gondar (violet) clustered together, and Southern Tigray (black) clustered alone B) Based breeding materials: released (red) and landrace (black)

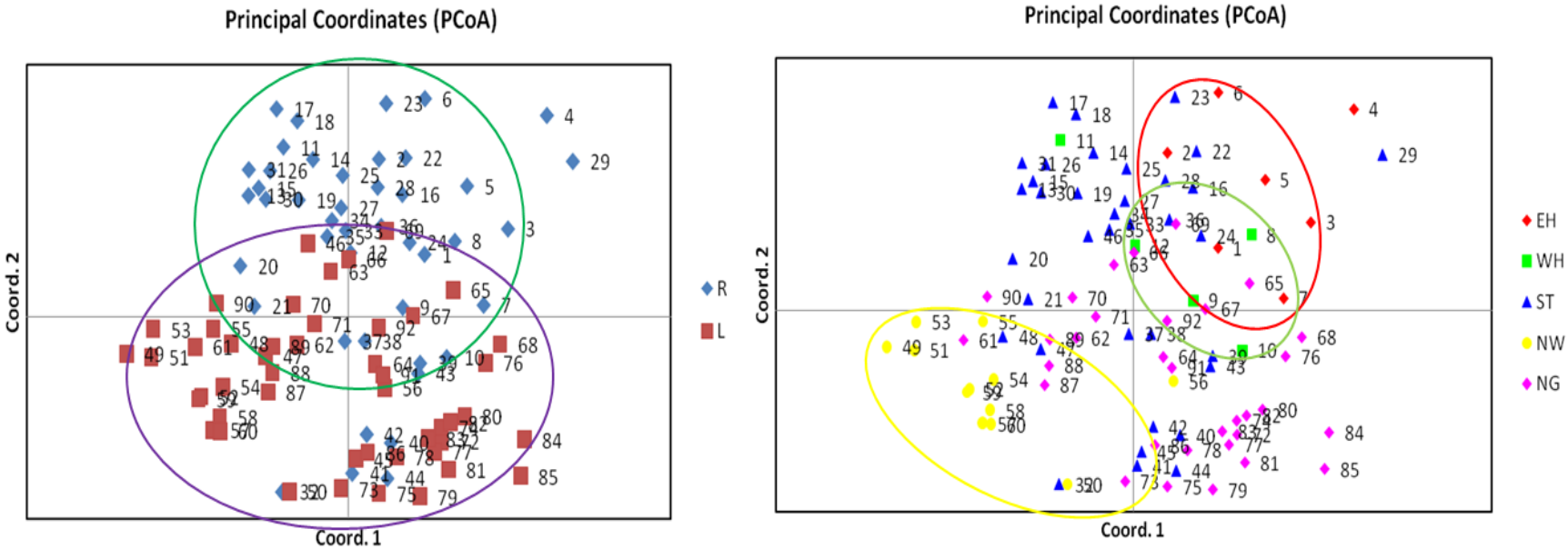


Figure 4-3. Principal component analysis (PCoA) of the sorghum population of Released and Landraces (Left) and across administration zones (Right): R= Released, L= Landraces, EH= East Hararege, WH= West Hararge, ST= South Tigray, NW= North Wollo, NG=North Gondar.

5. DISCUSSION

Genetic diversity analysis is the most crucial steps in the current plant breeding programs. It provides a deeper understanding of crop genetic variability to select appropriate parental lines for crossing, and thus, to speed up plant breeding objectives. Thus, the present study targeted to scan the genetic structure of both released and landrace Ethiopian sorghum genotypes using microsatellite markers. A total of 92 (44 released and 48 landraces) sorghum genotypes were used to assess the genetic variation and population structure using 12 microsatellite loci with a final goal of providing breeding information, wise use and conservation measures.

5.1 SSR allelic diversity

In the current study, high genetic diversity was detected among the studied Ethiopian sorghum populations. Moreover, a total number of 77 alleles with an average number of alleles (6.42) per locus were recorded across all sorghum relatives using 12 SSR markers and this result revealed a higher average number of alleles compared to Zambia, 4.4 (Ng'Uni *et al.*, 2011), Eastern Kenya, 5.03 (Muui *et al*., 2016), Eritrea, 4.8 (Tesfamichael Abraha *et al*., 2014), but lower than Eastern Benin, 7.8 (Missihoun *et al*., 2015) and Egypt, 7.7 (El-awady *et al*., 2008) which is a similar study to diversity analysis of sorghum with SSR markers.

The relatively higher mean number of alleles observed in the present study indicates the extent of higher marker polymorphism in the studied population. This result is contrasting with the previous study that reported lower mean number of alleles which suggests the presence of lower marker polymorphism across the studied materials. This may be due to the inclusion of a small number of study plant materials, the studied materials may be growing and/or collected from the same geographical area of the same agro-ecologies. Hence, our study reports may represent this

high genetic diversity is may be due to habitual seed exchange among farmers in the Ethiopian agricultural production system (McGuire, 2000). This seed exchange between farmers may increase the natural allele variations and allele frequencies in sorghum landraces and even rare alleles that may avoid the event of genetic drift, and the introduced rare alleles could be considered as an additional source of agronomically important traits. Sorghum is grown in Ethiopia in a range of agro-ecologies. Consequently, these broader agro-ecologies where all studied sorghum populations collected may accelerate natural variation among populations due to the diverse - agro-climate.

In the present study, all the microsatellite loci used were polymorphic and highly informative (PIC = 0.76), and thus they are useful genetic tools to unfold the genetic structure of sorghum. This result was in agreement with a previous study by Ceuvas and Prom (2013) who reported a mean PIC value of 0.78 in sorghum genotypes using SSR markers. In contrast, a lower PIC value of 0.46 and 0.33 were reported by Missihoun *et al.* (2015) and Muui *et al.* (2016), respectively. Although sorghum is a self-pollinating plant species, the incidence of cross-pollination is believed to be 8% in a given crop environment (Clarissa *et al*., 2013). Therefore, the observed high allele frequencies in the current study may be due to cross-pollination with wild relatives (House, 1985). In addition to these biological factors, the selected markers for this study were highly polymorphic, retrieved from previous related studies Billot *et al*. (2013), and this may be an attribute for the observed heterozygosity in the current study.

5.2 Population genetic diversity

Generally, older and larger populations show higher genetic diversity as compared with newly established and smaller ones because of several generations' accumulated genetic variation

which increases their wide adaptation and hence reduce their probability of local extinction. In the present study the higher values for Shannon's information index (1.4) and expected heterozygosity (0.71) confirmed the presence of higher genetic variation within populations. This could be due to the broad genetic base of the crop, and/ the inclusions of landraces collected from divers agro-ecological areas of the country (Tesfaye Disasa *et al*., 2016). Populations from South Tigray were identified to be more divers than others as exhibited by higher number of observed alleles (6.25), effective number of alleles (4.33), heterozygosity and Shannon diversity index (1.54) followed by populations from North Gondar, and hence the places representing these populations might be considered as sorghum genetic diversity hot spots. In contrast, population of East Hararege and West Hararege exhibited least genetic diversity likely due to the presence of various population bottlenecks leading to fast genetic erosion from these area or due to presence of higher artificial selection pressure for higher yield and other desired economical traits. Moreover, the relatively larger value for private alleles in South Tigray population indicates presence of certain level of independent evolution of its gene pool that allowed it to maintain the private alleles at a population level.

Based on the overall evaluated locus, the highest intra-population genetic diversity was observed among populations collected from North Gondar and South Tigray (Het = 0.75 and 0.74 respectively). Our observations were equivalent with other similar studies (Adugna Asfaw, 2014; Tesfaye Disasa *et al*., 2016) and the observed highest intra-population genetic diversity may be due to the event of frequent seed exchange because of the geographical closeness of the two zones (Gemechu Olani, 2017).

5.3 Genetic differentiation of the populations

The AMOVA confirmed the existence of lower genetic diversity among populations (10%) than within-population (90%). The low across loci Fst value (0.07**)** and between all pairs of populations (highest Fst= 0.22) indicates the presence of low to moderate genetic differentiation (Wright, 1943). This is supported by the presence of higher gene flow (Nm > 1). Population differentiation could be high if the gene flow which decreases population sub-structuring became low. The observed high gene flow (Nm =2.13) could be due to long distance pollen-grain movement among populations, and/or germplasm (seed) exchange among study areas where the different populations were sampled, Similar low to moderate population genetic differentiation was reported by previous studies (Adugna Asfaw *et al*., 2013; Adugna Asfaw, 2014; Labeyrie *et al*., 2014). The observed pair-wise population differentiation in the current study was lowest between the populations of West Hararge and East Hararge likely due to the large possibility for gene flow to occur between them. Conversely, the pair-wise population differentiation was observed among sorghum populations of East Hararge and North Wollo. This result was in agreement with the previous study (Tesfaye Disasa *et al*., 2016).

Pair-wise population genetic distance analysis confirmed the existence of relatively low genetic distances among the various sorghum populations likely due to the presence of higher gene flow among the populations. Pair-wise genetic Nei's standard genetic distance was relatively higher between sorghum populations of North Wollo and West, due to the existed geographical distance between the two zones. Conversely, the lowest genetic distance was observed between North

Gondar and West Hararge (0.009) possibly due to the presence of higher gene flow through seed exchange or other means between these zones. This result is comparable and consistent with the previous study (Adugna Asfaw, 2014; Tesfaye Disasa *et al*., 2016).

5.4 Clustering and population structure Analysis

Neighbour joining cluster analysis clustered the population into three major groups but not sharply according to their geographical origins. Although each cluster contains most of the genotypes collected from the same geographical area, the clustering pattern is not strong enough to support the principle of "isolation by distance". The present result is consistent with the previous genetic diversity studies on sorghum(Amsalu Ayana *et al*., 2000; Wang *et al*., 2013; Kitavi *et al*., 2014; Motlhaodi *et al*., 2014; Motlhaodi, 2016). Grouping based on biological status however clustered the sorghum materials almost into two distinct clusters as released and landrace with few admixture, This result is consistent and similar to the reports of Billot *et al*. (2013). However, some reports are not indicating sorghum clustering being either based geographical or racial (Motlhaodi *et al*., 2014). For instance, the phylogenetic and cluster analysis of sorghum collected from Ethiopia and Eritrea failed to form clusters according to their ecological and adaptation zones as analyzed by RAPD (Amsalu Ayana *et al*., 2000). Similarly, for the sorghum collection of Botswana, Namibia, Zambia, and Zimbabwe a cluster analysis failed to form a group based on their racial classification and ecological zones (Motlhaodi, 2016).

The clustering analysis conducted confirmed the occurrence of duplication of improved sorghum lines whose copies were found in the landrace and vice versa. In the case of landraces, the

movement of seeds among farmers of different regions got a local name and considered as different sorghum genotypes (Ng'uni *et al*., 2011). Whereas in the case of improved /released sorghum lines, different research institutes may release the same sorghum genotype under different names in different seasons or genotype selections based on a few genetic loci (Tesfaye Disasa *et al*., 2016). Hence, clustering analysis revealed the genetic relationship among the divers study materials, and generated sufficient information to enable the selection of genetically diverse and superior parental lines for crossing and other downstream breeding (Kristensen *et al*., 2018). Likewise, structure analysis confirmed the presence of closes relationship among samples collected from the different sorghum growing zones of Ethiopia, and it detected three sup-populations (K=3) with high potential of genetic admixture.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

In conclusion, understanding the genetic diversity and population structures of sorghum breeding materials using powerful molecular markers are the foremost important step in the current plant breeding program. The highly polymorphic and informative markers used in the present study had detected a larger number of alleles, where some of them were private alleles that could be linked to important agronomic traits. The study uncovered the genetic structure of sorghum for selection of germplasm to be used in further breeding program and to improve several economically important traits in the crop. Among the sorghum populations derived from different administrative zones, those from South Tigray exhibited the largest values for all the genetic diversity parameters. This is indicating that this location is a hot spot for source of useful alleles to be used in further sorghum breeding program. This baseline information is useful for sorghum improvement, sustainable use and conservation measures.

6.2 Recommendations

For a successful sorghum breeding, the integrated application of molecular studies using a high density markers and phenotypic components covering the major sorghum growing areas will enable and speed up the use of genetic diversity analysis for large scale breeding purpose.

- In our study we only consider molecular marker based diversity analysis. Thus, we suggest integrating the morphological analysis with molecular markers in sorghum germplasm characterizations would provide reliable genotypic information.
- The current study only considered five sorghum growing administrative zones and hence considering other sorghum growing areas would increase the resolution and accuracy of the diversity analysis for breeders to exploit the genetic potential of the crop to improving its production and productivity through breeding.
- Also, we used low density SSR marker and hence we suggest using of high-density SSR markers system would provide reliable genotypic information on the genetic structure of Ethiopian sorghum genotype which could be supportive for large scale sorghum breeding.
- For the future, marker technologies and approaches need to be integrated with the conventional breeding approaches to improve the productivity of sorghums and to sustain millions of lives in food-insecure semi-arid regions of the world, particularly in Ethiopia

7. REFERENCES

Abreham Reda (2014). Achieving food security in Ethiopia by promoting productivity of future world food tef: A review. *Adv. Plants Agric. Res*. **2**: 10-45.

Adugna Assefaw (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**:1-14.

Adugna Assefaw, Snow, A.A., Sweeney, P.M., Endashaw Bekele and Mutegi, E. (2013). Population genetic structure of *in situ* wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. *Genet. Resour. Crop Evol*. **60**:1313-1328.

Amsalu Ayana, Bryngelsson, T. and Endashaw Bekele (2000). Genetic variation of Ethiopian and Eritrean sorghum (*Sorghum bicolor (L.) Moench*) germplasm assessed by random amplified polymorphic DNA (RAPD). *Genet. Resour. Crop Evol.* **47:** 471–482.

Amsalu Ayana, Bryngelsson, T. and Endashaw Bekele (2001). Geographic and altitudinal allozyme variation in sorghum (*Sorghum bicolor* (L.) Moench) landraces from Ethiopia and Eritrea. *Hereditas* **135**:1-12.

Appa, R., Prasada Rao, K., Mengesha, M. and Reddy, V. (1996). Geographical distribution, diversity and gap analysis of east African sorghum collection conserved at the ICRISAT Gene bank. *Genet. Resour. Crop Evol.* **43**: 559–567.

Billot, C., Ramu, P., Bouchet, S., Chantereau, J. *et al*. (2013). Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLoS ONE* **8**: 591-714.

Borrell, A., Hammer, G., and Van Oosterom, E. (2001). Stay-green: A consequence of the balance between supply and demand for nitrogen during grain filling. *Ann. Appl. Biol.* **138**: 91-95.

Clarissa, T., Kimber, Jeff, A., Dahlberg, D. and Stephen, K. (2013). The Gene pool of *Sorghum bicolor* and its improvements. **In**: *Genomics of the Saccharinae, Plant Genetics and Genomics: Crops and Models*, pp. 23-39. (A. H. Paterson ed.). Texas, USA.

CSA (2020). *Report on area and production of major crops: Private peasant holdings of Meher season* (Vol. I). Addis Ababa, Ethiopia.

Cubry, P., Musoli, P., Legnate, H., Pot, D., De Bellis, F., Poncet, V., Anthony, F., Dufour, M., and Leroy, T. (2008). Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome*. **51**: 50 – 63.

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. *Genet Resour Crop Evol.* **60**: 1817–1830.

Dillon, S. L., Lawrence, P. K. and Henry, R. J. (2001). The use of ribosomal ITS to determine phylogenetic relationships within Sorghum. *Plant Syst. Evol.* **230**: 97–110.

Dillon, S.L., Shapter, F.M., Henry, R.J., Cordeiro, G., Izquierdo, L. and Lee, L.S. (2007). Domestication to Crop Improvement: Genetic Resources for Sorghum and Saccharum (*Andropogoneae*). *Ann. Bot.* **100**: 975–989.

Doggett, H. and Prasada Rao, K.E. (1995). Sorghum. **In**: *Evolution of Crop Plants, 2nd ed*, pp 140–159, (Smartt, J. and Simmonds, N.W., eds). Cambridge University Press, Cambridge.

Doggett, H., (1988). Sorghum. Harlow, Essex, England New York: Longman Scientific & Technical. *Wiley*, *18*, p.512.

Earl, D.A. and vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**: 359–361.

El-awady, M., Youssef, S. and Selim, E. E. M. (2008). Genetic diversity among *Sorghum bicolor* genotypes using simple sequence repeats (SSRs) markers. *Arab J. Biotechnol.* **11**: 181–192.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**: 2611–2620.

FAO of United Nations. (2015). www.fao.org.

Frew Mekbib (2008). Genetic erosion of sorghum (*Sorghum bicolor* (L.) Moench) in the centre of diversity, Ethiopia. *Genet. Resour. Crop Evol.* **55**: 351–364.

Fuller, D. Q. and Stevens, C. J. (2018). Sorghum domestication and diversification: a current archaeobotanical perspective. **In**: *Plants and People in the African Past*, pp.427–452, (Mercuri, A. M., D'andrea, A. C., Fornaciari, R. and Höhn, A. Eds.). Berlin, Germany.

Gabissa Ejeta (2007). Breeding for Striga resistance in sorghum: exploitation of intricate host–parasite biology. *J. Crop Sci.* **47**: S-216.

Garcia, A.A.F., Benchimol, L.L., Barbosa, A.M.M., Geraldi, I.O., Souza Jr., C.L. and Souza, A.P. (2004). Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genet. Mol. Biol.* **27**: 579–588.

Gebrekidan Brhane (1981). Utilization of germplasm in sorghum improvement. **In**: *Proceedings of the International Symposium on Sorghum, (House, L.R., Mughogho, L.K. and Peacock, J.M., eds.)* (pp. 2-7).

Gemechu Olani (2017). *Genetic Diversity Analysis of Sorghum (Sorghum bicolor (L.) Moench) Races in Ethiopia Using SSR Markers.* Msc thesis. Addis Ababa University. Addis Ababa.

Gnansounou, E., Dauriat, A. and Wyman, C.E. (2005). Refining sweet sorghum to ethanol and sugar: economic trade-offs in the context of North China. *Bioresource Technol.* **96**: 985-1002.

Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S.K. and Prasad, M. (2014). Population structure and association mapping of yield contributing agronomic traits in foxtail millet. *Plant Cell Rep.* **33**: 881–893.

Hao, C., Wang, L., Ge, H., Dong, Y. and Zhang, X. (2011). Genetic Diversity and Linkage Disequilibrium in Chinese Bread Wheat (*Triticum aestivum L.)* Revealed by SSR Markers. *PLoS ONE* **6**: 172-179.

Hariprasanna, K., and Patil, J.V. (2015). Sorghum: Origin, classification, biology and improvement. **In**: *Sorghum molecular breeding*, pp. 3–20, (Madhusudhana, R., P. Rajendrakumar, P. and Patil, J. V., Eds.). New Delhi, India.

Harlan, J. R. (1992). Indigenous African agriculture. **In**: *Crops*, pp.175-191, (Harlan, J. R., Ed.). Wisconsin, USA.

Harlan, J. R. and de Wet, J. M. J. (1972). A simplified classification of cultivated Sorghum. *Crop Sci*. **12**:148-172.

Harland, S. (1951). Vavilov Cultivated Plants. *Nature* **170**: 48–49.

House, L. R. (1985). *A Guide to Sorghum Breeding, Second Edition.* International Crops Research Institute for the Semi-Arid Tropics, Andhra Pradesh 502, India.

Huang, B.E., Verbyla, K.L, Verbyla, A.P., Raghavan, C., Singh, V.K, Gaur, P., Leung, H., Varshney, R.K. and Cavanagh, C.R. (2015). MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* **128**: 999-1017.

Jordan, D.R., Tao, Y.Z., Godwin, I.D., Henzell, R.G., Cooper, M. and McIntyre, C.L. (1998). Loss of genetic diversity associated with selection for resistance to sorghum midge in Australian sorghum. *Euphytica* **102**: 1–7.

Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Mol. Ecol.* **17**: 4015–4026.

Kitavi, M.N., Kiambi, D.K., Haussman, B., Semagn, K., Muluvi, G., Kairichi, M. and Machuka, J. (2014). Assessment of the genetic diversity and pattern of relationship of West African sorghum accessions using microsatellite markers. *Afr. J. Biotechnol*. **13**:1503-1514.

Kleih, U., Ravi, S. B., Rao, B. D. and Yoganand, B. (2000). *Industrial utilization of sorghum in India*. ICRISAT, India.

Klein, R.R., P.E. Klein, A.K. Chhabra, J. Dong, S. Pammi, K.L. Childs, J.E. Mullet, Rooney, W.L. and Schertz, K.F. (2001). Molecular mapping of the *rf1* gene for pollen fertility restoration in sorghum (*Sorghum bicolor* L.). *Theor. Appl. Genet.* **102**: 1206-1212

Kresovich, S., Barbazuk, B., Bedell, J. A., Borrell, A., Buell, C.R. and Burke, J. (2005). Toward sequencing the sorgum genome. *Plant physiol.***138**: 1898-1902.

Kristensen, P.S., Jahoor, A., Andersen, J.R., Cericola, F., Orabi, J., Janss, L.L. and Jensen, J. (2018). Genome-Wide Association Studies and Comparison of Models and Cross-Validation Strategies for Genomic Prediction of Quality Traits in Advanced Winter Wheat Breeding Lines. *Front. Plant Sci.* **9**: 34-69.

Labeyrie, V., Deu, M., Barnaud, A., Calatayud, C., Buiron, M., Wambugu, P., Manel, S., Glaszmann, J.C. and Leclerc, C. (2014). Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in eastern Kenya. *PLoS One*. **9**: 21-78.

Lazarides, M., Hacker, J. B. and Andrew, M. H. (1991). Taxonomy, cytology and ecology of indigenous Australian sorghums (*Sorghum Moench: Andropogoneae: Poaceae*). *Aust. Syst. Bot.* **4**: 591–635.

Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O. and Bennett, M. D. (2019). Angiosperm DNA C-values database (release 9.0, Apr 2019). Available at: https://cvalues.science.kew.org/ (Accessed 23rd March 2021).

Liu, K. and Muse, S.V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, ***21***: 2128-2129.

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J. and Bian, L., *et al*. (2013). Whole genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**:1–9.

Mace, E. S., Xia, L., Jordan, D.R., Halloran, K., Parh, D.K., Huttner, E., Wenzl, P. and Kilian, A. (2008). DArT markers: diversity analyses and mapping in Sorghum bicolor. *BMC Genomics* **9**: 12-26.

Mace, E.S., Buhariwalla, K. K., Buhariwalla, H.K. and Crouch, J.H. (2012). A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol. Biol. Rep.* **21**: 459-460.

Mann, J. A., Kimber, C. T. and Miller, F. R. (1983). *The Origin and Early Cultivation of Sorghum in Africa*. Texas, USA.

Martin, J. H. (1936). Sorghum improvement. **In**: *The Year book of Agriculture,* pp. 523–560 (United States Department of Agriculture, Ed.). Washington DC, USA.

Mathur, S., Umakanth, A.V., Tonapi, V.A., Sharma, R. and Sharma, M.K. (2017). Sweet sorghum as biofuel feedstock: recent advances and available resources. *Biotechnol. Biofuels* **10**: 134-146.

McGuire, S. (2000). Farmers' management of sorghum diversity in Eastern Ethiopia and use of genetic diversity of needs. **In**: *Encouraging Diversity: The Conservation and Development of Plant Genetic Resources,* pp. 43–48, (C.J.M. Almekinders, W. de Boef, eds.). London, Intermediate Technology.

Menz, M.A., Klein, R.R., Unruh, N.C., Rooney, W.L., Klein, P.E. and Mullet, J.E. (2004). Genetic diversity of public inbreeds of sorghum determined by mapped AFLP and SSR markers. *Crop Sci.* **44**:1236-1244.

Missihoun, A. A., Adoukonou-sagbadja, H. and Sedah, P. (2015). Genetic diversity of *Sorghum bicolor* (L) Moench landraces from Northwestern Benin as revealed by microsatellite markers. *Afri. J. Biotechnol.* **14**: 1342–1353.

Morris, G.P., Ramu, P., Deshpande, S.P., Hash, C.T., Shah, T., *et al*. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *PNAS.* **110**: 453–458.

Motlhaodi, T. M. (2016). *Genetic Diversity and Nutritional Content of Sorghum (Sorghum bicolor (L) Moench) Accessions from Southern Africa*. Ph.D. thesis, Swedish University of Agricultural Sciences, pp. 24-27.

Motlhaodi, T., Geleta, M., Bryngelsson, T., Fatih, M., Chite, S. and Ortiz, R. (2014). Genetic diversity in *ex-situ* conserved sorghum accessions of Botswana as estimated by microsatellite markers. *Aust. J. Crop Sci*. **8**:35-43.

Mundia, C.W., Secchi, S., Akamani, K. and Wang, G. (2019). A regional comparison of factors affecting global sorghum production: The case of North America, Asia and Africa's Sahel. *Sustainability* ***11***:21-35.

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

Muui, C. W., Muasya, R. M., Kirubi, D. T., Runo, S. M. and Karugu, A. (2016). Genetic variability of sorghum landraces from lower Eastern Kenya based on simple sequence repeats (SSRs) markers. *Afri. J. Biotechnol.* **15**: 264–271.

Nadeem, M.A., Nawaz, M.A., Shahid, M.Q., Doğan, Y., *et al*. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **32**: 261–285.

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

Peakall, R.O.D. and Smouse, P.E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**:288-295.

Pereira, G.S., Nunes, E.S., Laperuta, L.D.C., Braga, M.F., Penha, H.A., Diniz, A.L., Munhoz, C.F., Gazaffi, R., Garcia, A.A.F. and Vieira, M.L.C.(2013). Molecular polymorphism and linkage analysis in sweet passion fruit, an outcrossing species: Molecular map in sweet passion fruit. *Ann. Appl. Biol.* **162**: 347–361.

Perrier, X., Flori, A. and Bonnot, F.(2003). Data analysis methods. **In**: *Genetic diversity of cultivated tropical plants*, pp 43 – 76, (Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C., Eds). Enfield, Science Publishers. Montpellier.

Pritchard, J.K., Wen, W. and Falush, D. (2003). Documentation for the structure software, version 2. Department of Human Genetics, University of Chicago, Chicago.

Ramu, P., Billot, C., Rami, J.-F., Senthilvel, S., Upadhyaya, H.D., Ananda Reddy, L. and Hash, C.T. (2013). Assessment of genetic diversity in the sorghum reference set using EST-SSR markers. *Theor. Appl. Genet.* **126**: 2051–2064.

Reddy, B.V., Ramesh, S., Reddy, P.S. and Kumar, A.A. (2009). Genetic Enhancement for Drought Tolerance in Sorghum. *Plant Breed. Rev.* **31**:189-201.

Ritter, K.B., McIntyre, C.L., Godwin, I.D., Jordan, D.R. and Chapman, S.C. (2007). An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor ssp. bicolor (L.)* Moench, using AFLP markers. *Euphytica* **157**: 161–176.

Roa, N.G., Murty, U.R. and Rana, B.S. (2002). Sorghum. **In**: *Evolution and Adaptation of Cereal Crops*, pp. 213–238, (Chapra, V.L., Prakash, S. eds). Science Publishers, Inc., Enfield, USA.

Singh R. and Axtell, J.D. (1973). High lysine mutant gene (hl) that improves protein quality and biological value of grain sorghum. *Crop Sci.* **13**: 456-535.

Snowden, J. D. (1955). The cultivated races of sorghum. **In**: *Andropogoneae evolution and generic limits in Sorghum (Poaceae) using ndhF sequences*, pp. 267–281, (Spangler, R., Zaitchik, B., Russo, E., and Kellogg, E., eds.). London, Adlard.

Spangler, R. E. (2003). Taxonomy of *Sarga, Sorghum* and *Vacoparis* (*Poaceae: Andropogoneae*). *Aust. Syst. Bot*. **16**: 279–299.

Tesfamichael, Abraha, Githiri, S. M., Kasili, R. W., Skilton, R. A., Solomon, M. and Nyende, A. B. (2014). Genetic diversity analysis of Eritrean sorghum (*Sorghum bicolor* (L.) Moench) germplasm using SSR markers. *J. Plant Mol.* **5**: 1–12.

Tesfaye Disasa, Tileye Feyissa, and Belayneh Admassu (2016). Characterization of Ethiopian sweet sorghum accessions for 0 Brix , morphological and grain yield traits. *Sugar Tech.* **19**: 72-82.

Tesfaye Disasa, Tileye Feyissa, Belayneh Ademassu, Paliwal, R., De Villiers, S.M. and Odeny, D.A. (2016). Molecular evaluation of Ethiopian sweet sorghum germplasm and their contribution to regional breeding programs. *Aust J Crop Sci.* **10**: 520–527.

Upadhyaya, H.D., Narsimha Reddy, K., Vetriventhan, M., Ahmed, M., Murali Krishna, G., Thimma Reddy, M. and Singh, S.K. (2017). Sorghum germplasm from West and Central Africa maintained in the ICRISAT genebank: Status, gaps, and diversity. *Crop J*. **5**: 518–532.

Uptmoor, R., Wenzel, W., Friedt, W., Donaldson, G., Ayisi, K. and Ordon, F. (2003). Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. *Theor. Appl. Genet.* **106**: 1316–1325.

USDA and Agricultural Research Service, National Plant Germplasm System (2020). Germplasm Resources Information Network (GRIN-Taxonomy) (Beltsville, Maryland: National Germplasm Resources Laboratory). Available at: https://npgsweb.ars-grin.gov/gringlobal/taxonomydetail.aspx?id=489172 (Accessed 25th March 2021).

Venkateswaran, K., Elangovan, M. and Sivaraj, N. (2019). Origin, Domestication and Diffusion of Sorghum bicolor. **In**: *Breeding Sorghum for Diverse End Uses, pp.15-31,*

(C. Aruna, K.B.R.S. Visarada, B. V. Bhat and V. A. Tonapi, eds.). Cambridge, United Kingdom.

Venkateswaran, K., Muraya, M., Dwivedi, S. L. and Upadhyaya, H. D. (2014). Wild sorghums-Their potential use in crop improvement. **In**: *Genetics, genomics and breeding of sorghum*, pp.78–11, (Y. Wang, H. D. Upadhyaya and K. Chittaranjan, Eds.). Florida, USA.

Wang, M.L., Zhu, C., Barkley, N.A., Chen, Z., Erpelding, J.E., Murray SC, Tuinstra, M.R., Tesso, T., Pederson, G.A. and Yu, J. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* **120**:13-23.

Wang, M.L., Zhu, C., Barkley, N.A., Chen, Z., Erpelding, J.E., Murray SC, Tuinstra, M.R., Tesso, T., Pederson, G.A. and Yu, J. (2009). Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor. Appl. Genet.* **120**:13-23.

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

Winchell, F., Stevens, C. J., Murphy, C., Champion, L. and Fuller, D. Q. (2017). Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan Spikelet Morphology from Ceramic Impressions of the Butana Group. *Curr. Anthropol.* **58**: 673–683.

Xu, C., Gao, J., Du, Z., Li, D., Wang, Z., Li, Y. and Pang, X. (2016). Identifying the genetic diversity, genetic structure and a core collection of *Ziziphus jujuba Mill*. var. jujuba accessions using microsatellite markers. *Sci. Rep.* **6**: 31-50.

Yang, W., Oliveira, A.C., Godwin, I., Schertz, K. and Bennetzen, J.L. (1996). Comparison of DNA Marker Technologies in Characterizing Plant Genome Diversity: Variability in Chinese Sorghums. *Crop Sci.* **36**: 1669–1676.

Yao, D., Myriam, H., Mohammed, T., Claude, L. and Xavier, V. (2004). *In situ* estimation of outcrossing rate in sorghum landraces using micro-satellite markers. *Euphytica* **138**: 205–212.

Zigale Semahegn (2019). The Genetic Improvement of Sorghum in Ethiopia: Review. *JBAH*. **9**: 2224-3208.

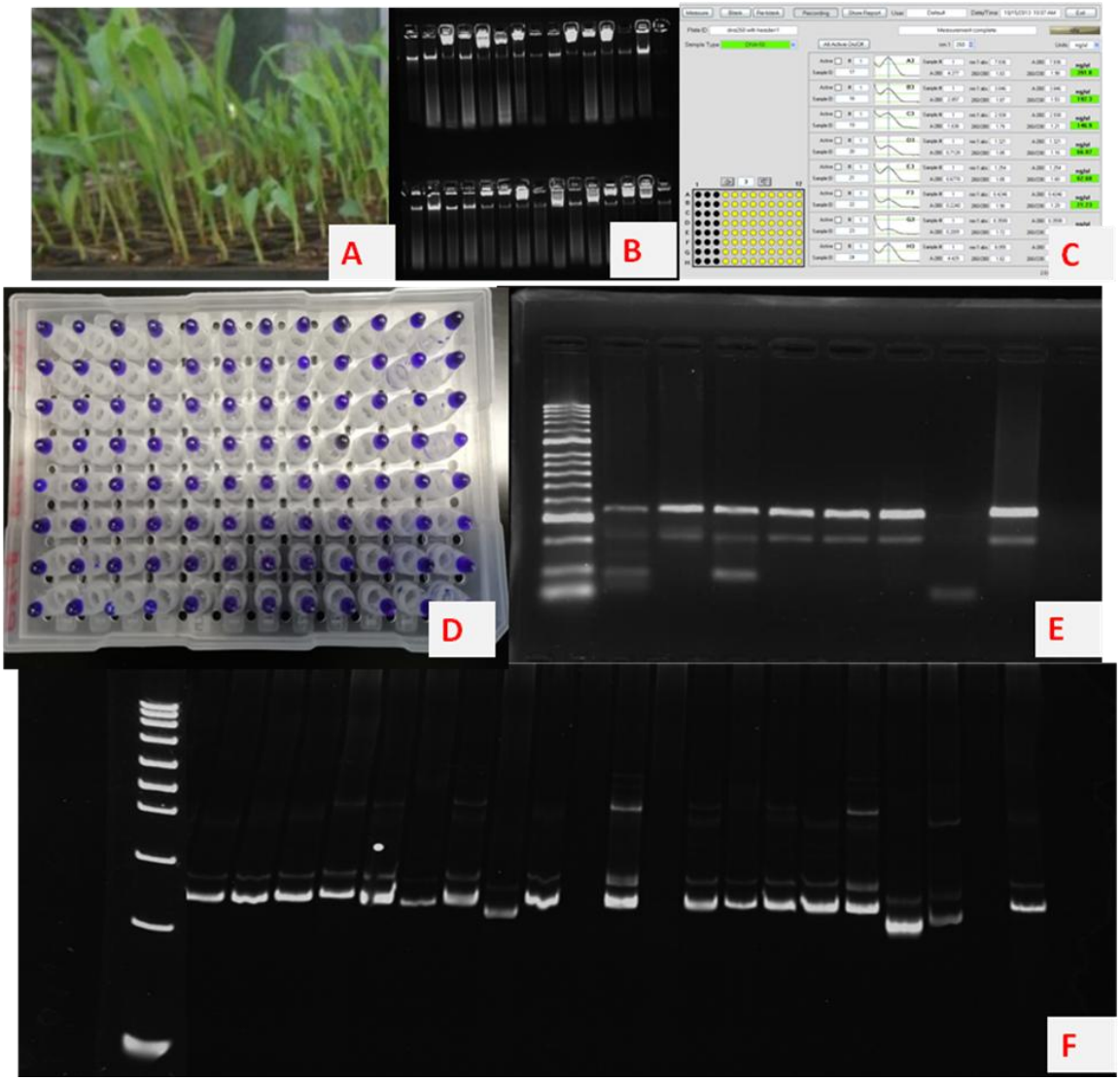
8. APPENDICES

Appendix 1. List of landraces and released sorghum genotypes used in this study

S/No.	Region	Zone	Name of Var.	Source
1	Harere	mirab hareregia	Abedelota	Farmer
2	Harere	misrak hareregia	Aferia	Farmer
3	Harere	mirab hareregia	Ageda	Farmer
4	Harere	misrak hareregia	Amedeye	Farmer
5	Harere	mirab hareregia	Chefera	Farmer
6	Harere	misrak hareregia	Chekuria	Farmer
7	Harere	misrak hareregia	Cherechere	Farmer
8	Harere	misrak hareregia	Nano	Farmer
9	Harere	misrak hareregia	Sherudia	Farmer
10	Harere	misrak hareregia	Wayelo	Farmer
11	Harere	mirab hareregia	Wogeria	Farmer
12	Harere	mirab hareregia	Worabi	Farmer
13	Tigray	S/Awi	Masia	Mehoni ARC
14	Tigray	S/Awi	ESH-1	Mehoni ARC
15	Tigray	S/Awi	Dekeba	Mehoni ARC
16	Tigray	S/Awi	ESH-3	Mehoni ARC
17	Tigray	S/Awi	Cheria	Mehoni ARC
18	Tigray	S/Awi	Gobia	Mehoni ARC
19	Tigray	S/Awi	Teshale	Mehoni ARC
20	Tigray	S/Awi	Horemate	Mehoni ARC
21	Tigray	S/Awi	Birhane	Mehoni ARC
22	Tigray	S/Awi	Red swize	Mehoni ARC
23	Tigray	S/Awi	Gerania-1	Mehoni ARC
24	Tigray	S/Awi	Mesaye	Mehoni ARC
25	Tigray	S/Awi	Meko	Mehoni ARC
26	Tigray	S/Awi	Melkam	Mehoni ARC
27	Tigray	S/Awi	Abeshere	Mehoni ARC
28	Tigray	S/Awi	Aberia	Mehoni ARC
29	Tigray	S/Awi	Raya	Mehoni ARC
30	Tigray	S/Awi	Mesekire	Mehoni ARC
31	Tigray	S/Awi	cherekit/mera	Mehoni ARC
32	Tigray	S/Awi	Red sorgum	Mehoni ARC
33	Tigray	S/Awi	Jeru	Mehoni ARC
34	Tigray	S/Awi	jeregutia/America	Mehoni ARC
35	Tigray	S/Awi	Horedem	Mehoni ARC
36	Tigray	S/Awi	boro/nechi mashela	Mehoni ARC
37	Tigray	S/Awi	Merawi	Mehoni ARC
38	Tigray	S/Awi	Ajibia	Mehoni ARC
39	Tigray	S/Awi	Abaiera	Mehoni ARC
40	Tigray	S/Awi	Misinego	Mehoni ARC
41	Tigray	S/Awi	Cherkite	Mehoni ARC
42	Tigray	S/Awi	Dengalite	Mehoni ARC
43	Tigray	S/Awi	Gano	Mehoni ARC
44	Tigray	S/Awi	Wedehakere	farmer
45	Tigray	S/Awi	hagose areya	Farmer
46	Tigray	S/Awi	Kodem	farmer

S/No.	Region	Zone	Name of Var.	Source
47	Tigray	S/Awi	Woteria	Farmer
48	Tigray	S/Awi	Baro	farmer
49	Amhara	N/wello	Wedehakere	Farmer
50	Amhara	N/wello	Zenegada	Farmer
51	Amhara	N/wello	Jamiye	farmer
52	Amhara	N/wello	Jegurtia	farmer
53	Amhara	N/wello	Degalite	farmer
54	Amhara	N/wello	gobye 2016	Serinka ARC
55	Amhara	N/wello	Mesekere 2016	Serinka ARC
56	Amhara	N/wello	chere 2016	Serinka ARC
57	Amhara	N/wello	Melkam-2016	Serinka ARC
58	Amhara	N/wello	Dekeba2016	Serinka ARC
59	Amhara	N/wello	masia- 2016	Serinka ARC
60	Amhara	N/wello	teshale 2016	Serinka ARC
61	Amhara	N/Gondar	Tekemeche	Farmer
62	Amhara	N/Gondar	Wenecho	Farmer
63	Amhara	N/Gondar	denebia zenegada	Farmer
64	Amhara	N/Gondar	Belesia	Farmer
65	Amhara	N/Gondar	yeker demozia	Farmer
66	Amhara	N/Gondar	Bulia	Farmer
67	Amhara	N/Gondar	abatu zenegada	GARC
68	Amhara	N/Gondar	Girana	GARC
69	Amhara	N/Gondar	Abeshere	GARC
70	Amhara	N/Gondar	Mogede	GARC
71	Amhara	N/Gondar	wdi-areba	GARC
72	Amhara	N/Gondar	Melkam	GARC
73	Amhara	N/Gondar	Jolia	GARC
74	Amhara	N/Gondar	Teshale	GARC
75	Amhara	N/Gondar	Gobia	GARC
76	Amhara	N/Gondar	Mesekere	GARC
77	Amhara	N/Gondar	Dekeba	GARC
78	Amhara	N/Gondar	Gamebala	GARC
79	Amhara	N/Gondar	Masha	GARC
80	Amhara	N/Gondar	Bajo	GARC
81	Amhara	N/Gondar	76T-23	GARC
82	Amhara	N/Gondar	Lalo	GARC
83	Amhara	N/Gondar	Meko	GARC
84	Amhara	N/Gondar	Berehane	GARC
85	Amhara	N/Gondar	IS-9302	GARC
86	Amhara	N/Gondar	Beremashe	GARC
87	Amhara	N/Gondar	Dageme	GARC
88	Amhara	N/Gondar	Chemeda	GARC
89	Amhara	N/Gondar	Yeju	GARC
90	Amhara	N/Gondar	gamebia 1107	GARC
91	Amhara	N/Gondar	Charia	GARC
92	Amhara	N/Gondar	Gerema	GARC

Appendix 2. A simple experimental procedure of the study (CTAB method). (A: Sorghum seeds planted in pots for gDNA extraction after two weeks, B: The extracted gDNA on 0.8% agarose gel, C) The quantity and quality of the extracted gDNA, D) A 2µl gDNA of the 92 sorghum samples and master mix was transferred into PCR plates sequentially for PCR amplification, E) A 1.2% agarose gel showing the presence of PCR product for marker *xtxp114*) with 100bp DNA ladder, and F) A polyacrylamide gel showing the fragment generated from the genome of the studied plant materials for genotyping together with the 100bp DNA ladder.



Appendix 3. Estimated Allele Frequencies and Heterozygosity over population for each locus

Components	Parameters	Xcup02	Xgap206	Xcup63	Xgap84	Xtxp114	Xtxp _298	Xtxp _034	Xtxp _284	Xtxp_ 312	Xtxp _023	Xtxp_ 270	Xtxp _211
Ecological Zone	**Na**	0.800	2.000	0.800	0.800	1.200	1.200	1.200	1.200	1.600	2.000	1.200	1.600
	Ne	1.049	1.573	1.012	1.087	1.061	1.116	1.083	1.057	1.103	1.223	1.065	1.143
	I	0.085	0.522	0.031	0.121	0.114	0.173	0.140	0.102	0.172	0.312	0.109	0.219
	OHe	0.043	0.346	0.012	0.068	0.055	0.095	0.072	0.050	0.088	0.175	0.055	0.118
	EHe	0.044	0.363	0.012	0.069	0.058	0.096	0.075	0.051	0.094	0.183	0.059	0.123
Released and Landraces	**Na**	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
	Ne	1.045	1.620	1.022	1.152	1.044	1.188	1.059	1.059	1.068	1.262	1.044	1.143
	I	0.106	0.566	0.060	0.242	0.102	0.291	0.123	0.123	0.136	0.357	0.102	0.246
	He	0.043	0.379	0.022	0.127	0.042	0.157	0.055	0.055	0.062	0.205	0.042	0.125
	uHe	0.044	0.383	0.022	0.129	0.042	0.159	0.055	0.055	0.063	0.207	0.042	0.127

(Abber. Na = No. of Different Alleles, Ne= number of effective alleles, I =Shannon´s index, OHe=Observed Heterozygosity, EHe= Expected Heterozygosity)

Appendix 3. Comparison of the allele frequencies of Ethiopian sorghum genotypes among ecological zones

For each population																Across total population		
Locus	**EH**			**WH**			**ST**			**NW**			**NG**			**<5%**	**5-50%**	**>50%**
	Ra	Ca	Aa	Ra	Ca	Aa	Ra	Ca	Aa	Ra	Ca	Aa	Ra	Ca	Aa	Ra	Ca	Aa
Xcup02	5	4	-	3	6	-	3	6	-	5	4	-	4	5	0	20	25	-
Xgap206	2	6	-	1	7	-	2	5	-	2	6	-	2	6	0	9	21	-
Xcup63	1	5	-	2	4	-	1	5	-	3	3	-	1	5	0	9	21	-
Xgap84	3	4	-	4	3	-	3	4	-	2	5	-	2	5	0	14	21	-
Xtxp114	2	5	-	3	4	-	1	6	-	3	4	-	3	4	0	12	23	-
Xtxp _298	2	5	-	2	5	-	1	6	-	0	7	-	2	5	0	7	28	-
Xtxp _034	5	5	-	4	6	-	1	9	-	2	8	-	1	9	0	13	37	-
Xtxp _284	1	4	-	1	4	-	0	5	-	1	4	-	1	4	0	4	21	-
Xtxp_ 312	0	5	-	2	3	-	2	3	-	3	2	-	2	3	0	9	16	-
Xtxp _023	1	3	-	1	3	-	1	3	-	0	4	-	1	3	0	4	16	-
Xtxp_ 270	2	4	-	2	4	-	2	4	-	1	5	-	1	5	0	9	21	-
Xtxp _211	5	3	-	4	4	-	2	6	-	5	5	-	3	5	0	19	21	-

(Abbreviations, EH, East Hararege, WH, West Hararge, ST, South Tigray, NW, North wollo, NG, North Gondar, Ra - Rare alleles that were present in <5% of the genotypes; Ca - Common alleles that were present in 5-50% of the genotypes; Aa- Abundant alleles that were present in more than 50% of the genotypes)

Appendix 4. Comparison of the allele frequencies between landraces and released Ethiopian sorghum genotypes

Locus	Released			Landraces		
	Ra	Ca	Aa	Ra	Ca	Aa
Xcup02	3	6	-	4	5	-
Xgap206	2	4	-	2	4	-
Xcup63	1	5	-	1	5	-
Xgap84	3	4	-	2	5	-
Xtxp114	1	6	-	2	5	-
Xtxp _298	0	7	-	0	7	-
Xtxp _034	1	9	-	1	9	-
Xtxp _284	0	5	-	1	4	-
Xtxp_ 312	1	4	-	2	3	-
Xtxp _023	1	3	-	0	4	-
Xtxp_ 270	2	4	-	1	5	-
Xtxp _211	3	5	-	3	5	-

(Abbreviations, Ra - Rare alleles that were present in <5% of the genotypes; Ca - Common alleles that were present in 5-50% of the genotypes; Aa- Abundant alleles that were present in more than 50% of the genotype