

Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from
Major Growing Areas of Ethiopia Using Simple Sequence Repeat (SSR) Markers



By

BETHELIHEM NEMERA

A Thesis Submitted to the department of Applied Biology

School of Applied Natural Sciences

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in
Biotechnology

Office of Graduate Studies

Adama Science and Technology University

June, 2021

Adama, Ethiopia

Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from
Major Growing Areas of Ethiopia Using Simple Sequence Repeat (SSR) Markers

By

BETHELIHEM NEMERA

Advisor: MULUGETA KEBEDE (PhD)

Co-advisor: TILEYE FEYISSA (PhD)

A Thesis Submitted to the department of Applied Biology

School of Applied Natural Sciences

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in
Biotechnology

Office of Graduate Studies

Adama Science and Technology University

June, 2021

Adama, Ethiopia

Declaration

I hereby declare that this MSc thesis entitled “Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from Major Growing Areas of Ethiopia Using Simple Sequence Repeat (SSR) Markers” is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation.

Name of the student

Signature

Date

Recommendation

I/we, the advisor(s) of this thesis, hereby certify that I/we have read the revised version of the thesis entitled “Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from Major Growing Areas of Ethiopia Using Simple Sequence Repeat (SSR) Markers” prepared under my/our guidance by Bethelihem Nemera submitted in partial fulfillment of the requirements for the degree of Master’s of Science in Biotechnology.

Therefore, I/we recommend the submission of revised version of the thesis to the department following the applicable procedures.

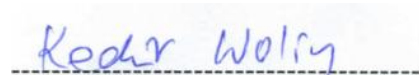
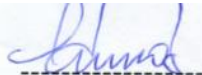
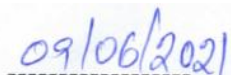
_____	_____	_____
Major Advisor	Signature	Date
_____	_____	_____
Co-advisor	Signature	Date

Approval Page of M.Sc. Thesis

we, the advisors of the thesis entitled “Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from Major Growing Areas of Ethiopia Using Simple Sequence Repeat (SSR) Markers” and developed by Bethelihem Nemera, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

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

We, the undersigned, members of the Board of Examiners of the thesis by Bethelihem Nemera have read and evaluated the thesis entitled “Genetic Diversity of Sorghum (*Sorghum bicolor* (L.) Moench) Germplasm from Major Growing Areas of Ethiopia 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

Acknowledgement

First of all, I would like to thank the Almighty GOD for His works of mercy in my life, may his holy name be worshipped forever and ever. I would like to thank Adama Science and Technology University for their financial support and for giving me the sponsorship opportunity to continue my post graduate study. I would also like to thank the National Agricultural Biotechnology Research Center (NABRC) for providing the molecular laboratory facilities to conduct this research.

I would like to thank my research main advisor, Dr. Mulugeta Kebede, for his skilful guidance, unfailing patience throughout the course of my studies as well as research continuation up till the accomplishment of this thesis. I also acknowledge and express my sincere thanks to my co-advisor, Dr. Tileye Feyissa for accepting me as a student to work under the Swedish International Development Agency (SIDA) through “Sorghum: screening and association mapping of drought-tolerant sorghum genotypes in Ethiopia research project. I am very thankful to him for his constructive suggestions, guidance and reviewing my draft thesis.

My special appreciation goes to Dr. Tesfaye Disasa for the chemical support during the laboratory work. I am also deeply grateful to Dr. Geleta Dugassa for his constructive guidance and endless help. I am also very much grateful to Emebet Regasa, Girma Gashu and Wubishet Mamo for their thoughtfulness, encouragement, advice and support during laboratory work, data analysis and for their genuine cooperation in every aspect. So many people have made meaningful contribution to achieve my goal. I am thankful to all friends who I have met along the way even if your names do not appear here.

Last but not least, I would like to pass my heartfelt thanks to my family without their support, encouragement and unconditional love I wouldn't have been here.

TABLE OF CONTENTS

Contents	Pages
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF APPENDIX	ix
ABBREVIATIONS	x
ABSTRACT	xi
1. INTRODUCTION	1
1.1 Statement of the problem	3
1.2 Objectives.....	4
1.2.1 General objective.....	4
1.2.2 Specific objectives.....	4
1.3 Significance of the study	4
2. LITERATURE REVIEW	5
2.1 Sorghum origin and distribution	5
2.2 Taxonomy and botanical description of sorghum	5
2.3 Sorghum growth requirements	6
2.4 Sorghum production areas in Ethiopia	7
2.5 Studies of sorghum genetic diversity	9
2.6 Sorghum genome.....	9
2.7 Molecular markers in genetic diversity studies.....	10
2.7.1 Non-polymerase chain reaction method	11
2.7.2 Polymerase chain reaction based molecular markers	11
3. MATERIALS AND METHODS.....	14
3.1 Plant materials	14
3.2 DNA extraction	14
3.3 PCR and gel electrophoresis	14
3.4 Data scoring and analysis	16
4. RESULTS	18
4.1 Genetic diversity within and among populations	20

4.2 Genetic differentiation, distance and gene flow between populations.....	21
4.3 Analysis of molecular variance	22
4.4 Cluster analysis and population structure.....	23
5. DISCUSSION	28
6. CONCLUSION AND RECOMMENDATIONS	31
7. REFERENCES	33
8. APPENDICES	42

LIST OF TABLES

Tables	Pages
Table 1. Sorghum producing areas by region (CSA, 2013)	8
Table 2. Characteristics of the 11 SSR markers used in the study.....	15
Table 3. The genetic diversity indices summary of the 11 SSR loci across five population of sorghum in Ethiopia.....	19
Table 4. Summary of allelic patterns and diversity indices across populations averaged over the 11 SSR loci.....	21
Table 5. Estimate of population genetic differentiation measured by F_{ST} (below the diagonal) among different geographic regions.....	21
Table 6. Pairwise Nei's genetic distance (below diagonal) and gene flow (N_m) values (above diagonal) among five sorghum populations from Ethiopia.....	22
Table 7. Analysis of molecular variance (AMOVA) of 80 sorghum germplasm genotyped using 11 SSR markers.....	23

LIST OF FIGURES

Figures	Pages
Figure 1. Electrophoretic bands of SSR markers generated from 13 sorghum germplasm	20
Figure 2. Unweighted Neighbor-joining tree for 80 sorghum germplasm.....	24
Figure 3. Unweighted pair-group method with arithmetic mean (UPGMA) dendrogram showing genetic relationships among the five populations of sorghum.....	25
Figure 4. Principal coordinate analysis of 80 sorghum germplasm using 11 SSR markers.....	26
Figure 5. Population structure of 80 sorghum germplasm representing five populations in Ethiopia.....	27

LIST OF APPENDICES

Appendices	Pages
Appendix 1. List of sorghum germplasm used in this study.....	42
Appendix 2. Plant DNA Extraction Protocol for DArT.....	45
Appendix 3. Electrophoretic bands of SSR markers Xtxp217, Xtxp096, Gpsb067 and Xtxp057 respectively.....	47

ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of Molecular Variance
DArT	Diversity Array Technology
ISSR	Inter Simple Sequence Repeat
MAS	Marker Assisted Selection
PCoA	Principal Coordinate Analysis
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SSR	Simple Sequence Repeat
TAE	Tris-Acetate-EDTA
UPGMA	Unweighted Pair Group Method with Arithmetic Average

ABSTRACT

Ethiopia is the center of origin of sorghum having wide variation of the crop. Evaluations and understanding of the genetic diversity of the species is fundamental to design appropriate conservation and management strategies. Therefore, the present study was designed to assess the genetic diversity and population structure of sorghum germplasm collected from major growing regions of Ethiopia. Eighty sorghum germplasm representing five populations were analyzed with 11 Simple Sequence Repeat (SSR) markers. A wide range of diversity indices including polymorphic information content (PIC), number of alleles, effective number of alleles, number of private alleles, Shannon's diversity index, expected heterozygosity and percentage of polymorphic loci (PPL) were computed to determine genetic variation within populations. Analysis of molecular variance (AMOVA) was executed to see genetic variation within and among populations. The unweighted Neighbour-joining-based cluster analysis, Principal components of analysis (PCoA) and structure analysis were done to elucidate clustering of populations based on their coefficient of similarity. All the microsatellite markers used in the present study were polymorphic and highly informative. The computed polymorphic information contents (PIC) ranges from 0.50 to 0.86 with a mean value of 0.75. A high within-populations genetic diversity was confirmed with gene diversity values ranging from 0.51 to 0.77 with an overall mean of 0.67. Moreover, AMOVA revealed that 93.26% of the total genetic variation was accounted for within populations, leaving only 6.74% for among populations. The unweighted Neighbour-joining-based cluster, PCoA and structure analysis failed to reveal a clear grouping of germplasm according to their geographical origins, confirming the high gene flow ($N_m=6.65$) among populations. In conclusion, the SSR markers used in the present study were highly informative and revealed the genetic structures of sorghum populations in Ethiopia. Among the studied populations, those from Oromia and Amhara displayed gene diversity greater than the grand mean value of 0.67 suggesting that these locations should be targeted to generate baseline information for breeding and conservation of Sorghum. The result of this study suggested that future germplasm collection and utilization strategies should take into consideration the magnitude and pattern of genetic diversity established at the genotypic level.

Key words: Ethiopia, genetic diversity, population structure, sorghum bicolor, SSR markers

1. INTRODUCTION

Sorghum (*Sorghum bicolor* (L.) Moench) belongs to the family Poaceae, tribe Andropogoneae, genus *Sorghum* (De Wet, 1978). *S.bicolor* includes all cultivated sorghums as well as a group of semi-wild and wild plants regarded as weeds (Mutegi *et al.*, 2011). Both the cultivated and wild varieties of *Sorghum bicolor* have diploid chromosomes number ($2n=20$). Sorghum is a tropical cereal and C4 plant, monocotyledon with a genome size of ~730 Mbp, and well known for tolerating drought (Gnansounou *et al.*, 2005).

Sorghum originated in northeastern Africa, an area currently occupied by Ethiopia, Eritrea and Sudan and it was domesticated 3000 –5000 years ago and then spread to different regions all around the world (De Wet and Harlan, 1971). In addition to this, Dogget (1965) suggested that the practice of cereal cultivation was introduced from Ethiopia to Egypt about 3000 B.C. As reported by Vavilov (1951), Ethiopia is the centre of origin of sorghum due to the wide variation of the crop. All the basic races of sorghum, except Kaffir, are believed to be found in Ethiopia though the Durras makes the largest proportion of the race composition (Stemler *et al.*, 1975).

Sorghum is the fifth in the order of cereal crops global production principally after wheat, maize, rice and barley (Cuevas *et al.*, 2014). According to FAOSTAT (2017), United States was the world's largest producer of sorghum followed by Mexico, Nigeria, Sudan and India. Africa is the world's regional leader in the total production of sorghum. Ethiopia is the third-largest producer of sorghum in Africa behind Nigeria and Sudan with a contribution of about 12% of annual production (Wani *et al.*, 2011). In Ethiopian agriculture, sorghum holds the third-largest share of total cereal production after teff [*Eragrostis tef* (Zucc.) Trotter.] and maize (*Zea mays* L.). Teff, maize, sorghum and wheat (*Triticum aestivum* L.) are accounting for about 24.0, 16.8, 14.6, and 13% of the total cereal crop cultivated area, respectively (CSA, 2014). The crop is cultivated in all regions of Ethiopia between 400 m and 2500 m altitude, mostly at lower altitudes along the country's Western, South Western, North Eastern, Northern and Eastern peripheries (EIAR, 2014). It is a very important crop for food and feed for needy people and as ingredients in various alcoholic beverages as malt in semi-arid areas in the world (Tariq *et al.*, 2014). Sorghum can also be widely used as fodder, biofuel, fiber and as well as biodegradable packaging (Elangovan *et al.*, 2012). The new direction has now given major attention to its use as renewable bioethanol production. In this regard, sweet sorghum has been used as smart biofuel crops because of its

sugar-rich stalk (Disasa *et al.*, 2016) and currently, 12% of grain sorghum production in the US is used to make ethanol.

The national average sorghum productivity in Ethiopia is 2.3 tons per hectare (CSA, 2014) which is far below the global average is 3.2 tons per hectare (FAO, 2005). Several productions constraints were identified as a hindrance for sorghum production and productivity enhancement. The major constraints include drought, Striga, insect pest (stalk borer, midge, and shoot fly), diseases (anthracnose and smut), soil fertility decline, inadequate adoption of existing improved varieties, lack of high yielding and good quality sorghum varieties, and post-harvest management practices.

The success in any plant breeding program depends on the knowledge of genetic variability for the efficient selection process. Therefore, the study of genetic diversity can be considered as one of the most important conditions for developing and improving new and high yielding varieties (Yaqoob *et al.*, 2015). Identification of these genes involves either the application of conventional breeding and/or modern molecular approaches. The evaluation of genetic diversity based on the conventional morphological method is time-consuming and highly dependent on environmental conditions. The application of molecular markers is crucial and more effective and efficient for the selection of traits of interest in breeding programs as well as to assess genetic diversity since these markers are not influenced by the environment (Smith and Smith, 1992). Molecular markers play a major role in the determination of genetic variability and relationships. These included restriction fragment length polymorphism (RFLPs) (Botstein *et al.*, 1980), random amplified polymorphic DNA (RAPDs) (Williams *et al.*, 1991), amplified fragment length polymorphism (AFLPs) (Vos *et al.*, 1995), expressed sequence tags (EST) (Kantety *et al.*, 2002) and simple sequence repeats (SSRs) (Tautz, 1989) markers have shown their great potential in detecting the genetic diversity and relationships of organisms due to their allele specificity and co-dominant nature. In addition, SSR marker is the preferred marker system for many sorghum genomics and molecular breeding applications due to its technical simplicity, high throughput level and potential for automation (Missihoun *et al.*, 2015). In order to develop a new breed with good traits, it is necessary to assess the genetic diversity. Therefore, the aim of this study was to assess the genetic diversity of sorghum collected from major growing areas in Ethiopia using SSR markers and identify highly diverse genotypes to broaden the genetic base of sorghum improvement in Ethiopia.

1.1 Statement of the problem

Genetic diversity in the crop species is one of the precious gifts of nature to us, and it arises due to geographical isolation or genetic boundary to gene flow. The methods to study the genetic diversity of sorghum extend from analysis of morphological characteristics to biochemical and molecular traits (Muthusamy *et al.*, 2008). Knowledge at the molecular level can potentially bridge to identify and organize the available genetic resources associated with adaptive or agronomically important traits (Mengesha *et al.*, 2011). Evaluations of genetic diversity are extremely fundamental to understand the level and distribution of diversity in the species and for the design of appropriate conservation strategies (Bekele, 1985).

For genetic improvement and conservation of sorghum, genetic diversity study is important. Though some studies were conducted at the molecular level to understand the genetic diversity of sorghum, the proportion of the studies, compared with the number of germplasm in the gene bank and the amount of diversity which the country has, is small. Therefore, this study was conducted to assess the genetic diversity of sorghum germplasm collected from different geographical origins for its improvement based on SSR molecular marker.

1.2 Objectives

1.2.1 General objective

- The main objective of this study is to assess the genetic diversity of sorghum (*Sorghum bicolor* (L.) Moench) germplasm collected from major growing regions of Ethiopia using SSR markers.

1.2.2 Specific objectives

- To determine the genetic diversity of sorghum germplasm using SSR markers.
- To evaluate the genetic variability within and among populations of sorghum.
- To identify the population structure of the sorghum germplasm collected from the major sorghum growing regions of Ethiopia.

1.3 Significance of the study

The result of this study:

- Will provide information on the general genetic variability of sorghum (*Sorghum bicolor* (L.) Moench) germplasm which may assist in the identification and selection of the genetic materials for conservation and improvement for different regions of Ethiopia.
- Enhance the knowledge of sorghum diversity using the molecular marker.
- Can be used as a reference for other researchers to do similar works.

2. LITERATURE REVIEW

2.1 Sorghum origin and distribution

Sorghum originated in north-eastern Africa, and it was domesticated 3000–5000 years ago and then spread to different regions all around the world such as India and China (De Wet and Harlan, 1971). Anthropological evidence suggests that hunter-gatherers consumed sorghum as early as 8000 BC (Smith and Frederiksen, 2000). The primary centers of origin and diversity for sorghum are believed to be in sub-Saharan Africa extending from the extreme East to West Africa (Tarekegne, 2014). However, the complex diversity of agro-ecologies and the presence of wide genetic variation describe Ethiopia as a centre of origin of sorghum (Vavilov, 1951). Therefore, the greatest genetic diversity for both cultivated and wild forms of sorghum is found in Ethiopia and the surrounding eastern African countries. It is also the 2nd staple cereal after tef, *Eragrostis tef*, and in total national production, it ranks 3rd after maize and tef (CSA, 2014).

2.2 Taxonomy and botanical description of sorghum

Sorghum is a member of the family Poaceae, the genus *Sorghum*, order Poales, family Gramineae and the panicoid subfamily, tribe Andropogoneae, includes many species and classified under the genus *Sorghum* (De Wet, 1978). It was believed that there were 20 to 30 species of genus *Sorghum* that are recognized until now. Currently, 25 species of sorghum are recognized, and these are classified into five sections: *Stiposorghum*, *Parasorghum*, *Eusorghum*, *Heterosorghum* and *Chaetosorghum*. Under the section *Eu-sorghum*, three species are recognized: *S. halepense* (L.) Pers. occurring in India, *S. propinquum* (Kunth) Hitchc found in Southeast Asia and *S. bicolor* (L.) Moench, which originated in Africa (De Wet, 1978) all classified under the genus *Sorghum*. De Wet's recognized *S. bicolor* (L.) Moench representing all annual cultivated, wild and weedy sorghums along with two rhizomatous taxa, *S. halepense* and *S. propinquum*. All types of the *S. bicolor* (the primitives) propagated with seeds except the rhizomatous taxa's which reproduces through both seed and rhizome production. *Sorghum bicolor* was further classified into three subspecies: *S. bicolor* subsp *bicolor*, *S. bicolor* subsp *drummondii* and *S. bicolor* subsp *verticilliflorum* (Harlan and De Wet, 1972).

The cultivated sorghums are classified as *S. bicolor* subsp *bicolor* and represented by agronomic types such as grain sorghum, sweet sorghum, sudangrass and broomcorn. All the cultivated sorghum taxa of the world have been classified by inflorescence type, grain and glumes into five races: Durra, Bicolor, Caudatum, Kaffir and Guinea and there are very high levels of diversity that exist among and within the races (Harlan and De Wet, 1972). *Sorghum bicolor* is the most important taxon among these races (Menz *et al.*, 2002). From the five basic races, four races (Bicolor, Caudatum, Durra and Guinea) except Kaffir are found in Ethiopia (Teshome *et al.*, 1997). Durra is dominant in Ethiopia and the main crop of the eastern highland region and mid-elevation terrace of the north, while Caudatum race is grown primarily in hot, dry valleys and lowland savannas in the south and west of Ethiopia. The intermediate race Durra-Bicolor predominates in the southwestern highland region, where cooler temperature and rain are higher than in the eastern and northern regions. In contrast, Bicolor and Guinea races represent a very small part of Ethiopian sorghum diversity and both are mostly found in the Rift valley region.

2.3 Sorghum growth requirements

Sorghum is a warm short cycle annual, adapted to withstand higher average temperatures than most other cereal crops (Hall, 2000). Mundia *et al.* (2019) reported that after germination, an average temperature between 24°C and 27°C is ideal for best yields. Low temperatures can be limiting to sorghum growth and most plants will die when exposed to below freezing temperatures. In addition to this Du (2003) reported that sorghum is often grown in regions that get between 350–700 mm of precipitation annually. As a predominantly rain-fed crop, its yield depends largely on its drought resistance. The ideal soil moisture during germination ranges between 25% and 50% of field capacity and the sorghum plant can survive flooding events as it is more tolerant of wet soils. Sorghum is often grown in shallow to medium-deep or light to medium-textured soils and or medium to deep soils of high-water capacity after the rainy season. As reviewed by Mundia *et al.* (2019), since sorghum is not only a food crop, varieties are chosen to meet specific uses. Its yield is partly determined by the varieties of cultivars and the cultural practices of the region.

2.4 Sorghum production areas in Ethiopia

Sorghum is one of the major staple crops grown in the poorest and most food-insecure regions of Ethiopia. The crop is typically produced under adverse conditions such as low input use and marginal lands. It is well adapted to a wide range of precipitation and temperature levels and is produced between 400 m and 2500 m altitude (Fetene *et al.*, 2011). Mainly it is grown at lower altitudes along the country's western, southwestern, north eastern, northern and eastern peripheries. As represented in Table 1 Oromiya, Amhara and Tigray are the three major producers of sorghum covering 86% of the total area and 89% of the total production in the last 9 years (CSA, 2013). Sorghum is produced mainly in four traditional agro-ecologies of dry lowland, humid lowland, intermediate and high land altitudes vernacularly named as Kolla, Erteb Kolla, Woinadega and Dega respectively. More than 65% of the sorghum growing areas fall in dry lowland agro-ecology, and it is characterized by erratic rain fall, low soil fertility and a fragile eco-system.

Table 1. Sorghum producing areas given by regions and zones or special districts (CSA, 2013)

Region	Total area coverage(ha)	Total production(Quintals)	Zones/special districts/woredas
Amhara	733,017.68	14,320,673.73	North Gondar, South Gondar, North Wello, South Wello, North Shewa, East Gojam, Waghemra, Oromiya zone, Argoba
Benishangul Gumuz	60,122.37	1,108,206.34	Metekel, Assosa, Kemashi, Mao Komo, pawe
Dira Dawa	9,389.52	138,116.32	Dire Dawa
Gambella	3,324.79	54,973.60	Agnuwak, Mezhenger, Itang
Harari	7,701.01	148,510.94	Harari
Oromia	743,379.32	16,267,374.41	West Welega, East Welega, Illubabor, Jimma, West Shewa, North Shewa, East Shewa, Arsi, West Hararghe, East Hararghe, Bale, South West Shewa, Kelem, Horo guduru, West Arsi
Southern Nations,Nationalities and Peoples Region (SNNP)	122,731.01	1,967,158.93	Gurage, Hadiya, Kembata-Tembaro, Sidama, Wolayita, South Omo, Sheka, Keffa, Gamo Gofa, Benchi-maji, Yem, Konso, Derashe, Dawro, Basketo, Konta, Silite, Alaba, Burji
Somali	28,368.93	757,042.84	Shinele, Jijiga
Tigray	215,142.84	4,736,781.77	North West Tigray, Central Tigray, East Tigray, South, Western
Total	1,923,178.35	34,498,838.88	

2.5 Studies of sorghum genetic diversity

Genetic diversity is defined as the heritable variation within and among populations analyzed by a specific method or a combination of methods. Several approaches ranging from morphological to DNA sequence level analysis can be used to assess the genetic diversity of the crops. In sorghum, several studies have been carried out using morphological, biochemical, molecular and sequence-based diversity analysis.

Despite the fact that, morphological method has its limitations, it is an important step in complementing with molecular diversity study. Most of these studies focused on cultivated grain sorghum collected from a particular region of the country (Noor *et al.*, 2012). Globally, limited studies using morphological based genetic diversity have also been carried out using sweet sorghum and wild sorghum germplasms (Adugna *et al.*, 2013). A complement to morphological markers, little protein (allozymes) studies have been applied to the geographic and altitudinal variation among landraces collected from Ethiopia and Eritrea (Ayana *et al.*, 2001). Similarly, Dje *et al.* (1999) assessed the population genetic structure of sorghum landraces from north-western Morocco using allozyme and microsatellite markers. Some other studies measured genetic diversity on selected sorghum genotypes at the molecular level using various types of DNA markers such as RFLP (Menz *et al.*, 2004), RAPD (Ayana *et al.*, 2000), AFLP (Geleta *et al.*, 2006), ISSR (Tadesse and Feyissa, 2013), SSR (Tirfessa *et al.*, 2020), expressed sequence tags (EST) (Ramu *et al.*, 2009). Those studies differ in their power of genetic resolution, quality of information content, extents of polymorphism, degree of environmental stability, number of loci and practicality. Generally, molecular markers provide a better estimate of genetic diversity among sorghum genotypes and increase the resolution of measurement of genetic variation.

2.6 Sorghum genome

Sorghum is the first C4 plant that sequenced in 2009 by Andrew Paterson and his group involving 20 laboratories across the USA, Germany, China, Switzerland, and India. It has a comparatively small genome (~730 Mb), making it an attractive model for functional genomics study of Saccharinae and other C4 grasses (Paterson *et al.*, 2009). According to Hariprasanna and Patil (2015), due to its small genome size combined with its high drought tolerance nature and high level of inbreeding ranging from 94 to 99.4%, sorghum is considered an attractive model crop for

functional genomics and association studies of Saccharinae and other C4 grasses. In addition, Paterson *et al.* (2009) reported that nearly 61% of the sorghum genome is characterized by repetitive sequence and contains 27,640 protein-coding genes, about 24 % of the genes are grass-specific whereas 7% of the genes are sorghum-specific. The availability of the genome sequence has tremendous importance in the sorghum improvement program particularly in the application of marker-assisted breeding. Moreover, the genome sequence information will create an excellent foundation for the improvement of drought tolerance traits of other related cereal crops.

2.7 Molecular markers in genetic diversity studies

Molecular markers showed powerful and reliable tools for variation within crop genotypes compared with the biochemical markers (Gepts, 1993). A molecular marker is a sequence of DNA, which are located with a known position on the chromosome (Kumar, 1999). An ideal molecular marker should be co-dominant, evenly distributed throughout the genome, highly reproducible and having the ability to detect a higher level of polymorphism (Mondini *et al.*, 2009).

Molecular markers differ from morphological traits in that they usually occur in greater numbers, they can be distinguished without relying on the complete development of the plant and their expression is not altered by the environment (Assar *et al.*, 2009). Molecular markers have proved valuable in crop breeding, especially in studies on genetic diversity, gene mapping for understanding and determination of genetic variability and relationships. According to the approach employed in the classification, these markers are grouped as methodical principle, the aim of use, or the type of polymorphism (Gupta *et al.*, 1999). Those DNA markers are divided into three classes: (a) Hybridization-based DNA markers (e.g., RFLP); (b) PCR-based DNA markers (e.g., RAPD, AFLP, SSR, and ISSR) and (c) and DNA chip sequencing-based DNA markers (SNP). Different kinds of markers have been used efficiently to characterize and estimate the genetic diversity of sorghum (Shehzad *et al.*, 2009). For example, restriction fragment length polymorphism (RFLPs) (Menz *et al.*, 2004), amplified fragment length polymorphism (AFLPs) (Geleta *et al.*, 2006) random amplified polymorphic DNA (RAPDs) (Iqbal *et al.*, 2010), and simple sequence repeats (SSRs) (Adugna, 2014).

2.7.1 Non-polymerase chain reaction method

Molecular markers based on non-PCR techniques were employed relatively early in the field of plant studies and combined the use of restriction endonucleases and the hybridization method (Southern, 1975).

Restriction Fragment Length Polymorphism (RFLP) is the first molecular marker technique developed and used in MAS for plant breeding (Botstein *et al.*, 1980). It is a technique that exploits variations in homologous DNA sequences. It refers to the difference between samples of homologous DNA molecules that come from differing locations of restriction enzyme sites. RFLPs correspond to DNA fragments, usually within the range of 2-10 kb, that resulted from the digestion of genomic DNA with restriction enzymes. The RFLPs markers are relatively highly polymorphic, co dominantly inherited and highly reproducible. The technique centers around the digestion of genomic DNA with restriction enzymes (Madhumati, 2014). Differences in fragment lengths resolved by gel electrophoresis and Southern blotting, Specific banding patterns are then visualized by hybridization with labelled probe (Mondini *et al.*, 2009).

2.7.2 Polymerase chain reaction based molecular markers

The Polymerase chain reaction (PCR) is a molecular biology technique for enzymatically replicating (amplifying) small quantities of DNA and useful in studying DNA sequence variation without using a living organism. It is used to amplify a short (usually up to 10 kb), well-defined part of a DNA strand (between two specific priming sites in the genome) from a single gene or just a part of a gene. PCR based markers require less DNA per assay than RFLP and are higher throughput (Molosiwa, 2012).

2.7.2.1 Random Amplified Polymorphic DNA (RAPD)

RAPDs were the first PCR-based molecular markers to be employed in genetic variation analyses (Williams *et al.*, 1991). In the last decade, the RAPD technique based on PCR has been one of the most commonly used molecular techniques to develop DNA markers. The principle of RAPD is that a single, short oligonucleotide primer, which binds to many different loci, is used to amplify random sequences from a complex DNA template. This means that the amplified fragment

generated by PCR depends on the length and size of both the primer and the target genome (Kumari and Thakur, 2014).

2.7.2.2 Amplified Fragment Length Polymorphism (AFLP)

AFLP was developed to overcome some of the shortcomings of reproducibility of RAPDs as the technique combines the digestion of DNA with some specific restriction endonucleases with a PCR-based technique (Vos *et al.*, 1995). AFLP analysis first uses the restriction digestion of genomic DNA with a combination of rare cutting (EcoRI or PstI) and frequent cutting (MseI or TaqI) restriction enzymes (Vos *et al.*, 1995), followed by ligation of double-stranded adaptors to the sticky ends of the restriction fragment to generate template DNA. The nucleotide sequence of the adaptors and the adjacent restriction sites serve as primer binding sites during selective PCR amplification of some of these fragments. The amplified fragments are separated by electrophoresis on polyacrylamide gels and visualized either through autoradiography or fluorescence methods (Hartl and Seefelder, 1998).

2.7.2.3 Simple Sequence Repeat

Simple sequence repeats (SSR) are also called microsatellites, short tandem repeats (STR) and variable number tandem repeats (VNTR). SSRs are polymorphic loci present in DNA that consist of repeating units of one to six base pairs units arranged in repeats of mono-, di-, tri-, tetra and penta-nucleotides (A, T, AT, GA, AGG, AAAG, etc.). These repeats are widely distributed throughout the plants and animal genomes that display a high level of genetic variation based on differences in the number of tandemly repeating units of a locus (Jonah *et al.*, 2011). Because the DNA sequences flanking microsatellite regions are usually conserved, primers specific for these regions are designed for use in the PCR reaction. The technique requires nucleotide information for primer design, labour-intensive marker development process and high start-up costs for automated detections. They are usually characterized by their hyper-variability, reproducibility, co-dominant nature, locus-specificity, and random genome-wide distribution in most cases for genetic studies (Jiang, 2013). In addition, they have been widely applied for the identification and fingerprinting of genotypes (Menz *et al.*, 2002). Yet among DNA markers, SSR markers remain the markers of choice for practical breeding applications.

2.7.2.4 Inter Simple Sequence Repeat

Inter simple sequence repeat (ISSR) marker technique is a PCR based method, which involves amplification of DNA present at an amplifiable distance between two identical microsatellite repeats oriented in opposite direction. The technique uses microsatellites, usually, 16-25 bp long as primers in a single primer PCR reaction targeting multiple genomic loci to amplify inter simple repeat sequences of different sizes. The microsatellite repeats used as primers for ISSRs can be di, tri, tetra or penta-nucleotides (Madhumati, 2014). ISSRs can have problems in reproducibility, and their dominant inheritance and homology of co-migrating amplification products result in similar problems to RAPDs (Omondi *et al.*, 2016).

3. MATERIALS AND METHODS

3.1 Plant materials

Eighty sorghum germplasm used for diversity analyses in this study were previously collected from major sorghum growing regions of Ethiopia by Ethiopian Biodiversity Institute (EBI). A total of 80 germplasm, 28 from Oromia, 28 from Amhara, 14 from Tigray, 5 from Dire dawa and 5 from Southern Nations, Nationalities and Peoples Regions (SNNP) were used in this study (Appendix 1). The study was carried out at the Melkassa Agricultural Research Center (MARC) and National Agricultural Biotechnology Research Center (NABRC). Fifteen seeds from each germplasm were planted at Melkassa Agricultural Research Center (MARC) and from each germplasm, a pool of five healthy young leaves were collected from two-week-old plants and put in a plastic bag containing silica gel to dry and make it ready for genomic DNA extraction.

3.2 DNA extraction

The genomic DNA was extracted from young leaves using the plant DNA extraction protocol described in Diversity Array Technology (DArT) with some modifications at National Agricultural Biotechnology Research Center (NABRC) (Appendix 2). Extracted DNA quality was checked by loading 3 µl of DNA mixed with 2 µl of 6x loading dye with gel red on a 1% agarose gel and separated at 100 V for 40 min. The genomic DNA was quantified using a Nano-Drop spectrophotometer (ND-8000, 8-sample Spectrophotometer). The DNA was stored at -20°C until used for the analyses.

3.3 PCR and gel electrophoresis

Fifteen SSR primer-pairs, which were previously developed by Billot *et al.* (2012), were used for initial screening for amplification, polymorphism and specificity to target loci. According to the manufacturer's instruction, the primers were dissolved using nuclease-free water to a final concentration of 100 mol/µl. Out of the 15 screened primer pairs, 11 of them successfully amplified the genomic DNA and showed polymorphism across tested sorghum germplasm (Table 2).

DNA amplification was performed in 12.5 µl reaction volume containing 6.25 µl, One Taq® 2X Master Mix with Standard Buffer, 2 µl 50ng/µl template DNA, 2 µl forward and reverse primers, 0.25 µl DMSO (Dimethyl sulfoxide) and 2 µl nuclease-free water. Touch-down PCR amplification was carried out using BIO-RAD T100 Thermal cycler. The PCR program for the first six primer

pairs (Table 2) consisted of 15 min at 94°C for the initial denaturation followed by twelve cycles of 94°C for 30 s, 60°C for 45 s (ramp of 1°C per cycle) and 72°C for 1 min, then by 31 cycles of 94°C for 30 s, 38°C for 45 s and 72°C for 1 min, a final extension of 20 min at 72°C and then put on hold at 4°C, at infinity (∞). The PCR conditions for the remaining primer pairs 7 to 11 (Table 2) involved 15 min at 94°C for the initial denaturation followed by twelve cycles of 94°C for 30 s, 60°C for 45 s (ramp of 1°C per cycle) and 72°C for 1 min, then by 31 cycles of 94°C for 30 s, 51°C for 45 s and 72°C for 1 min, a final extension of 20 min at 72°C before holding at 4°C. The PCR amplified products were separated in 3% agarose gel electrophoresis by loading 5 μ l of each of the PCR product mixed with 2 μ l of gel red using 1x TAE buffer at 100V for three hours. The product was visualized under gel documentation system (BioDoc-It™ imaging system) and subsequently photographed. A 100 bp size ladder was used to estimate the size of the amplified products.

Table 2. Characteristics of the 11 SSR markers used in the study

No	Marker	Chromosome	DNA Motif	Forward primer	Reverse primer	Range of fragment size(bp)
1	Xtxp021	4	(AG)18	GAGCTGCCATAGAT TTGGTGC	ACCTCGTCCCAC CTTTGTTG	157-214
2	Msbcir246	5	(CA)7+(GA)5	TTTTGTTGCACTTTT GAGC	GATGATAGCGAC CACAAATC	88-108
3	Xtxp141	7	(GA)23	TGTATGGCCTAGCT TATCT	CAACAAGCCAAC CTAAA	139-189
4	Gpsb067	8	(GT)10	TAGTCCATACACCT TTCA	TCTCTCACACAC ATTCTTC	163-180
5	Xtxp057	9	(GT)21	GGAAC TTTTGACGG GTAGTGC	CGATCGTGATGT CCCAATC	209-237
6	Xtxp096	2	(GA)24	GCTGATGTCATGTT CCCTCAC	CATTCGTGGACT CTGTCCG	161-227
7	Xtxp217	10	(GA)23	GGCCTCGACTACGG AGTT	TCGGCATATTGA TTTGGTTT	144-219
8	Xcup53	1	(TTTA)5	GCAGGAGTATAGGC AGAGGC	CGACATGACAAG CTCAAACG	187-196
9	Xcup02	6	(GCA)6	GACGCAGCTTTGCT CCTATC	GTCCAACCAACC CACGTATC	188-197
10	Xtxp321	8	(GT)4+(AT)6+(CT)21	TAACCCAAGCCTGA GCATAAGA	CCCATTACACACA TGAGACGAG	180-200
11	Msbcir276	3	(AC)9	CCCCAATCTAACTA TTTGGT	GAGGCTGAGATG CTCTGT	221-238

3.4 Data scoring and analysis

All the PCR amplified products were scored by comparing with known fragments size ladders that were electrophoresed along with each germplasm using PyElph 1.4 software package (Pavel and Vasile, 2012). Genetic diversity and population structure analyses were carried out based on the scored bands. Different statistical software packages were employed to compute the standard indices of genetic diversity. Locus based diversity indices including major allele frequency (MAF), the number of alleles (Na), gene diversity (GD) and Polymorphic information contents (PIC) were analyzed using Power marker version 3.25 software (Liu and Muse, 2005). Effective number of alleles (Ne), Shannon's Information index (I), Population differentiation (Fst) and Gene flow (Nm) were determined using POPGENE version 1.32 (Yeh *et al.*, 1999). Population's genetic pattern; number of alleles (Na), effective number of alleles (Ne), number of private alleles (NPA), Shannon's Information index (I), expected heterozygosity (He), Percentage of polymorphic loci (PPL) and estimate of the deviation from Hardy-Weinberg equilibrium (HWE) were computed using GenAlEx version 6.5 software (Peakall and Smouse, 2012). The same software was used to compute pairwise population genetic distances and gene flow, and to perform the genetic differentiation test (Fst) over 999 bootstrap replications.

Analysis of molecular variance (AMOVA) and estimate of the variance components were conducted using Arlequin version 3.5.2.2 (Excoffier and Lischer, 2010). Gene flow (Nm) among populations was estimated using the formula,

$$Nm \text{ (Haploid)} = [(1 / F_{st}) - 1] / 2$$

where F_{st} = the variance among populations/total genetic variations.

A genetic dissimilarity matrix was analyzed using neighbour-joining (NJ) method and Nei's standard genetic distance (DST, corrected) (Nei, 1972) based Unweighted Pair Group Method with Arithmetic Mean (UPGMA) trees were generated using DARwin version 6.0 (Perrier and Jacquemoud-Collet, 2006) and Power Marker version 3.25 (Liu and Muse, 2005), respectively.

Principal Coordinate Analysis (PCoA) was performed using GenAlex version 6.5 (Peakall and Smouse, 2012) software package. Population structure and admixture patterns were determined using the admixture model based on the Bayesian algorithm implemented in STRUCTURE software version 2.3.3 (Pritchard *et al.*, 2000). The admixture model with correlated allele frequencies was used, assuming that the genome of each individual resulted from the mixture of K ancestral populations. To estimate the true number of population cluster (K), a burn-in period of 100,000 was used in each run, and data were collected over 250,000 Markov Chain Monte Carlo (MCMC) replications for K = 1 to K = 10 using 20 iterations for each K. The optimum K value was predicted following the simulation method of Evanno *et al.* (2005), using the web-based STRUCTURE HARVESTER version 0.6.92 (Earl, 2012). A bar plot for the optimum K was determined using Clumpak beta version (Kopelman *et al.*, 2015).

4. RESULTS

The eleven SSR markers generated a total of 70 alleles, across all 80 sorghum germplasm with an average of 6.36 alleles per marker. The analysis showed that the major allele frequency ranged from 0.20 (Xtxp217) to 0.54 (Xcup53) with a mean frequency of 0.31 per locus (Table 3). The lowest major allele frequency was 0.20 and the highest number of allele was 9. The effective number of alleles (7.80), Shannon's Information Index (2.13), gene diversity (0.87) and expected heterozygosity (0.77) were obtained for marker Xtxp217 (Table 3). Conversely, the highest major allele frequency (0.54) and the lowest number of alleles (3), effective number of alleles (2.39), Shannon's Information Index (0.96), gene diversity (0.58), and expected heterozygosity (0.56) was obtained for marker Xcup53 (Table 3). Although the highest genetic differentiation (0.21) and the lowest gene flow (0.91) was obtained for marker Msbcir276, the lowest genetic differentiation (0.06) and the highest gene flow (3.70) was obtained for marker Xcup02 (Table 3).

The PIC value for the SSR loci ranged from 0.50 for marker xcup53 to 0.86 for marker Xtxp217 with a mean value of 0.75 (Figure 1, Appendix 3). With regard to the test for the Hardy Weinberg equilibrium, all the 11 SSR markers showed a highly significant ($p < 0.0001$) deviation from HWE (Table 3).

Table 3. The genetic diversity indices summary of the 11 SSR loci across five populations of Sorghum in Ethiopia

Marker	MAF	Na	Ne	I	GD	He	Fst	Nm	PIC	P_{HWE}
gps067	0.40	5.0	3.95	1.50	0.75	0.68	0.10	2.15	0.71	0.000***
msbcir246	0.21	6.0	5.77	1.77	0.83	0.69	0.15	1.36	0.80	0.000***
xcup02	0.43	4.0	2.80	1.15	0.64	0.60	0.06	3.70	0.56	0.000***
xcup53	0.54	3.0	2.39	0.96	0.58	0.56	0.08	2.87	0.50	0.000***
xtxp141	0.33	7.0	4.63	1.71	0.78	0.58	0.20	0.98	0.75	0.000***
xtxp021	0.24	6.0	5.34	1.72	0.81	0.69	0.14	1.55	0.78	0.000***
xtxp057	0.21	9.0	7.22	2.07	0.86	0.74	0.11	2.09	0.85	0.000***
xtxp096	0.23	8.0	6.41	1.96	0.84	0.71	0.14	1.53	0.82	0.000***
xtxp217	0.20	9.0	7.80	2.13	0.87	0.77	0.11	2.04	0.86	0.000***
xtxp321	0.24	7.0	5.50	1.80	0.81	0.74	0.10	2.18	0.80	0.000***
msbcir276	0.35	6.0	4.70	1.67	0.79	0.59	0.21	0.91	0.76	0.000***
Mean	0.31	6.36	5.14	1.68	0.78	0.67	0.13	1.66	0.75	

Where MAF = Major allele frequency, Na= Number of alleles, Ne = Effective number of alleles, I = Shannon's Information Index, GD= Gene diversity, He = Expected heterozygosity, Fst = Inbreeding coefficient within subpopulations relative to total (genetic differentiation among subpopulations), Nm = gene flow estimated from $F_{st} = 0.25 (1-F_{st}) / F_{st}$, PIC = Polymorphic information content, P_{HWE} = P-value for deviation from Hardy Weinberg equilibrium, ns = not significant, * = $P < 0.0001$ and hence highly significant.

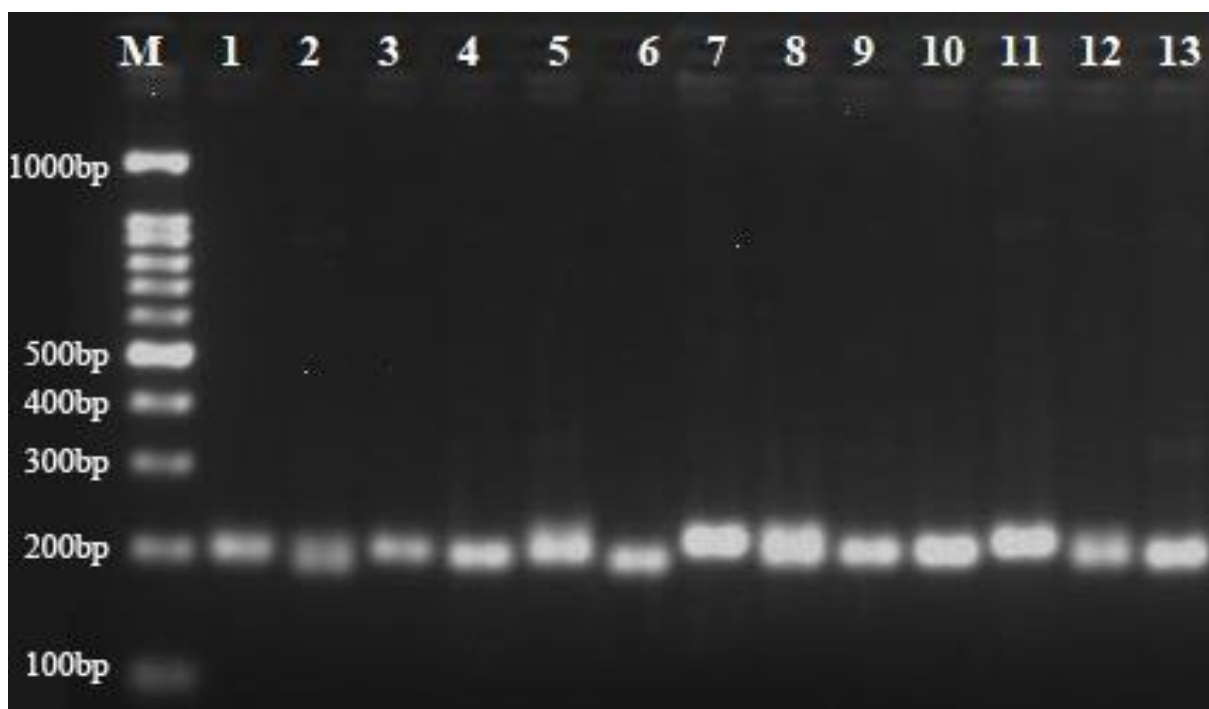


Figure 1. Electrophoretic bands of SSR markers (Xtxp217) generated from 13 sorghum germplasm collected from different geographical areas in Ethiopia. Where; M is a 100bp DNA ladder and the numbers in lanes indicate the germplasm.

4.1 Genetic diversity within and among populations

A summary of the different genetic diversity indices over the eleven markers for the five populations was presented in Table 4. The overall genetic diversity estimates within the populations had several alleles ranged from 2.64 to 6.18 with a mean value of 4.47. The Oromia population showed the highest number of alleles (6.18), effective number of alleles (4.89), Shannon's Information Index (1.63), and expected heterozygosity (0.77). On the contrary, the SNNP population showed the lowest number of alleles (2.64), effective number of alleles (2.32), Shannon's Information Index (0.84) and expected heterozygosity (0.51). Except for the population of Amhara where the number of private alleles is 0.09, there was no private allele unique to a single population. The percentage of polymorphic loci per population ranged from 90.91% (SNNP) to 100% (Oromia, Amhara, Tigray and Dire Dawa) with a mean of 98.18% (Table 4).

Table 4. Summary of allelic patterns and diversity indices across populations averaged over the 11 SSR loci

Population	N	Na	Ne	NPA	I	He	NPL	PPL
Oromia	28.00	6.18	4.89	0.00	1.63	0.77	11	100.00%
Amhara	28.00	6.00	4.71	0.09	1.58	0.75	11	100.00%
Tigray	14.00	4.18	3.17	0.00	1.23	0.67	11	100.00%
SNNP	5.00	2.64	2.32	0.00	0.84	0.51	10	90.91%
DireDawa	5.00	3.36	2.97	0.00	1.12	0.64	11	100.00%
Mean	16.00	4.47	3.61	0.02	1.28	0.67		98.18%

Where N= Number of samples, Na= Number of alleles, Ne = Effective number of alleles, NPA = Number of Private Alleles, I = Shannon's Information Index, He = Expected heterozygosity, NPL = Number of polymorphic loci, PPL= Percentage of polymorphic loci.

4.2 Genetic differentiation, distance and gene flow between populations

The pair-wise genetic differentiation between the populations ranged from 0.02 to 0.16 (Table 5). The highest population differentiation was observed among Tigray and SNNP ($F_{st}=0.16$), followed by Dire Dawa and SNNP ($F_{st}=0.15$). The lowest population differentiation ($F_{st}=0.02$) was observed among Oromia and Amhara (Table 5). The Pairwise Nei's genetic distance and gene flow (N_m) of each population from the other populations ranged from 0.11 to 0.74 and 0.94 to 16.68, respectively (Table 6). The highest measure of genetic distance (0.74) with the lowest gene flow (0.94) was observed between the populations of Tigray and SNNP (Table 6). The second-highest genetic distance (0.62) was observed between the populations of Amhara and SNNP. The lowest measure of genetic distance (0.11) with the highest gene flow (16.68) was observed between the populations of Oromia and Amhara (Table 6). The second-lowest gene flow (1.27) was observed between the populations of Dire Dawa and SNNP.

Table 5. Estimate of population genetic differentiation measured by Fsts (below the diagonal) among different geographic regions

	Oromia	Amhara	Tigray	SNNP	DireDawa
Oromia	0.00				
Amhara	0.02	0.00			
Tigray	0.05	0.04	0.00		
SNNP	0.12	0.13	0.16	0.00	
DireDawa	0.05	0.07	0.09	0.15	0.00

Table 6. Pairwise Nei's genetic distance (below diagonal) and gene flow (Nm) values (above diagonal) among five sorghum populations from Ethiopia

	Oromia	Amhara	Tigray	SNNP	DireDawa
Oromia	0.00	16.68	3.34	1.46	5.86
Amhara	0.11	0.00	3.47	1.34	3.04
Tigray	0.30	0.28	0.00	0.94	2.30
SNNP	0.59	0.62	0.74	0.00	1.27
DireDawa	0.27	0.42	0.44	0.58	0.00

4.3 Analysis of molecular variance

Analysis of molecular variance (AMOVA) was computed among and within populations. The variability among populations and within populations accounted for 6.74% and 93.26% of the total genetic variations, respectively. The overall fixation index value, used as a measure of population differentiation was moderate ($F_{st} = 0.07$) with high gene flow (≥ 6.65) (Table 7).

Table 7. Analysis of molecular variance (AMOVA) of 80 sorghum germplasm genotyped using 11 SSR markers

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	Fixation Index
Among populations	4	50.41	0.30	6.74	FST: 0.07
Within populations	75	635.84	4.10	93.26	
Total	79	686.25	4.40	100	
Nm (haploid) populations	6.65				

4.4 Cluster analysis and population structure

The neighbour-joining-cluster analysis divided the 80 individuals into four main clusters (C1, C2, C3 and C4) (Figure 2). The first cluster (C1) consisting of 22.5% from different populations (7.5% from Amhara, 3.75% from Dire Dawa, 5% from Oromia, 1.25% from SNNP and 5% from Tigray). The second cluster (C2) consisting of 13.75% from different populations (5% from Amhara, 1.25% from Dire Dawa, 6.25% from Oromia and 1.25% from SNNP). The third cluster (C3) consisting of 12.5% from different populations (2.5% from Amhara, 5% from Oromia and 5% from Tigray). The fourth cluster (C4) consisting of 51.25 % from different populations (20% from Amhara, 1.25% from Dire Dawa, 18.75% from Oromia, 3.75% from SNNP and 7.5% from Tigray). Unweighted Pair Group Method with Arithmetic mean (UPGMA) clustered the five populations into two major clusters (I and II) (Figure 3).

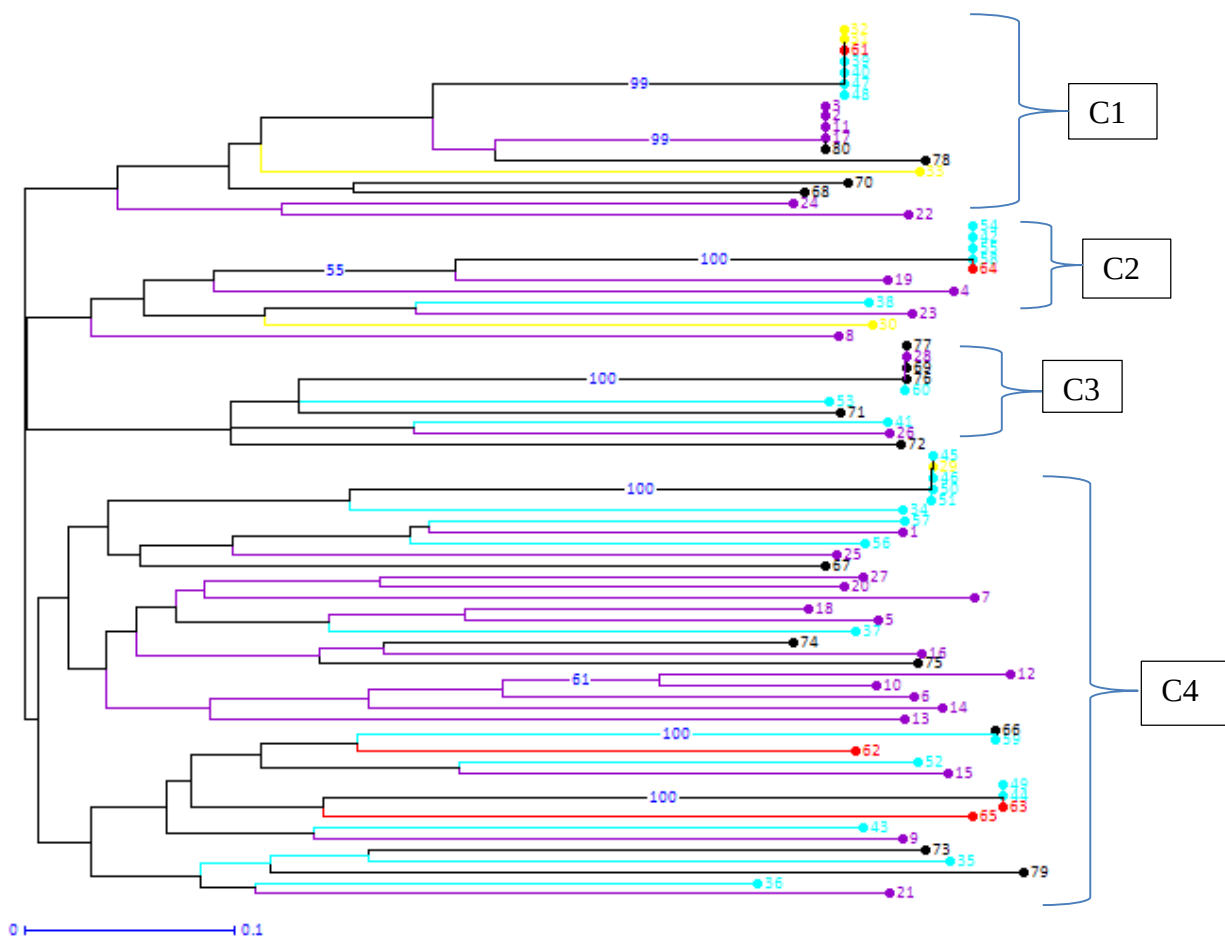


Figure 2. Unweighted Neighbor-joining tree for 80 sorghum germplasm. Numbers at the roots of the branches represent percentages of bootstrap values, and values less than 50% were not indicated. Each color represents a population: purple = Amhara; Yellow = Dire Dawa; Aqua = Oromia; Red = SNNP and Black = Tigray.

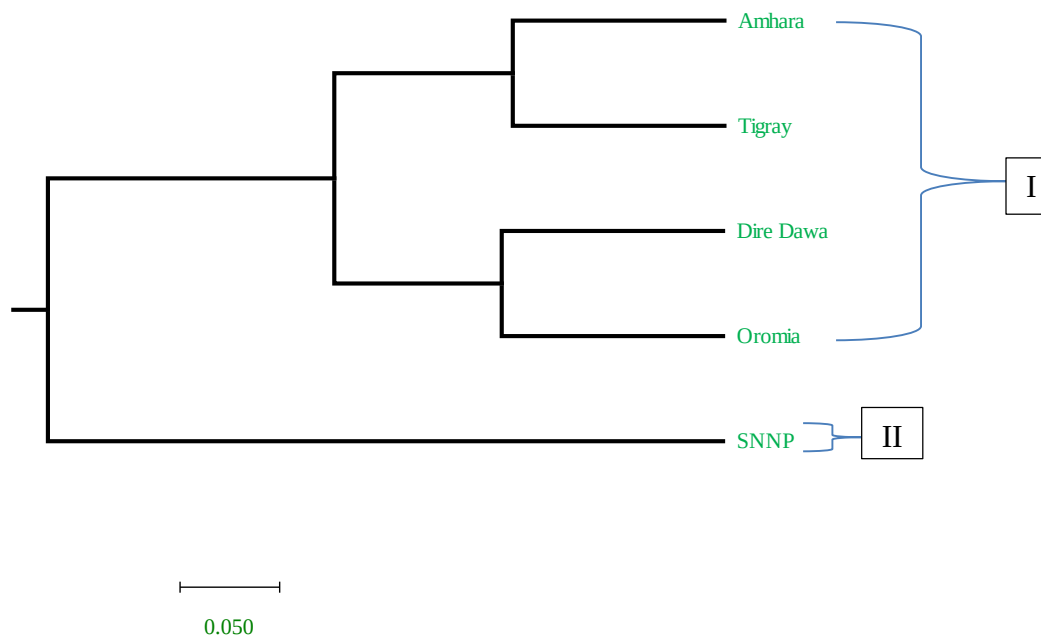


Figure 3. Unweighted pair-group method with arithmetic mean (UPGMA) dendrogram showing genetic relationships among the five populations of sorghum

The genetic relatedness among sorghum germplasm was further investigated using principal coordinate analysis (PCoA) (Figure 4). The three principal coordinates accounted for 29.69% of the total variations. The first, second, and third principal coordinates axes showed 12.13%, 9.07% and 8.49%, respectively.

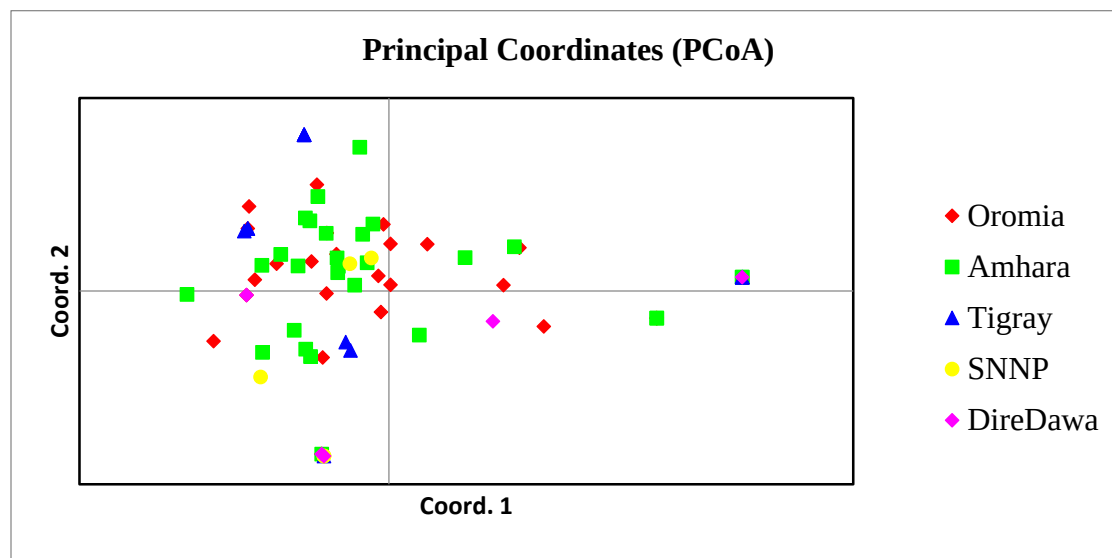


Figure 4. Principal coordinate analysis of 80 sorghum germplasm using 11 SSR markers. Samples coded with the same symbol and colors belong to the same population. The percentages of variation explained 29.69% of the total variations and the first 3 axes (1, 2 and 3) are 12.13%, 9.07% and 8.49%, respectively.

The population structure analysis revealed that maximum delta K becomes a sharp peak at K=2 (Figure 5A), suggesting two clusters. Based on this value, Clumpak result (bar plot) detected a genetic admixture and hence there was no clear geographic origin-based structuring of populations (Figure 5B).

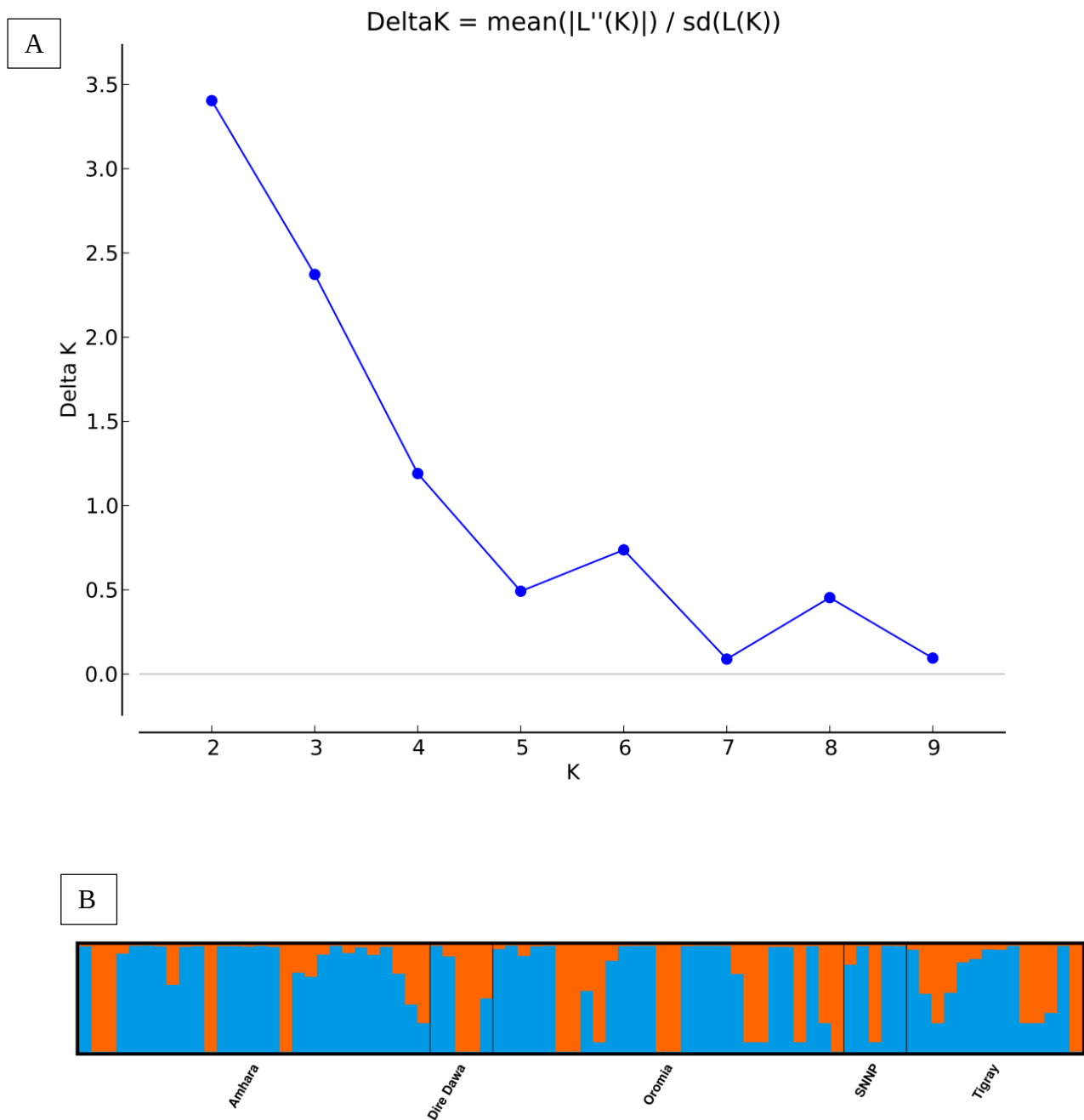


Figure 5. Population structure of the 80 sorghum germplasm representing five populations in Ethiopia. (A) Best delta K value estimated using Evano *et al.* (2005); method and (B) Bayesian model-based estimated population structure for $K = 2$ according to geographical locations. The different (blue and orange) colores represent genetic groups or sub-populations designated by Structure Harvester.

5. DISCUSSION

In the framework of this study, 11 SSR markers amplified a total of 70 alleles. The number of alleles per locus ranged from three for Xcup53 to nine for Xtxp057 and Xtxp217 with a mean of 6.36 alleles per locus. The mean number of alleles observed in the current study is similar to the report of Mofokeng *et al.* (2014), who observed a total of 306 alleles with a mean of 6.4 alleles per locus with a range of 2 to 15 alleles using 30 SSR markers in 103 sorghum genotypes but less than Cuevas and Prom (2013) population structure and diversity study for 137 Ethiopian germplasm conserved at USDA-ARS National Plant Germplasm System, that is 14 per locus. The difference could be due to the size of the population, the number of germplasm and the types of SSR markers used. The mean gene diversity of the markers observed in this study (0.78) was fairly similar to that previously reported by Thudi and Frakrudin (2011) (0.8) among rabi sorghum genotypes. Ng'Uni *et al.* (2011) reported the mean gene diversity value of 0.53 which is lower than the value obtained in the current study.

The high level of gene diversity of SSR markers observed in this study was probably associated with the genetic diversity in the sorghum germplasm that represented different geographic origins. The polymorphic information content (PIC) values are usually calculated to assess the level of polymorphism of a marker. As described by Botstein *et al.* (1980), a marker with PIC values below 0.25 is slightly informative, a marker with PIC between 0.25 and 0.5 is considered as reasonably informative, and those greater than 0.5 is said to be highly informative. In the present study, all the eleven SSR loci used had PIC values greater than 0.5. The PIC of loci ranged from 0.5 to 0.86 with a mean of 0.75. These results were fairly similar to the mean PIC ranges obtained in other sorghum genetic diversity study by Cuevas and Prom (2013) who reported a mean PIC value of 0.78. However, the observed PIC value is higher than that of Muui *et al.* (2016), who reported 0.49. High PIC values indicate the high discrimination ability of the selected SSR markers for the studied sorghum genotypes.

As indicated by diversity indices across populations by Shannon's information index (1.28) and expected heterozygosity (0.67), relatively all of the populations in Ethiopia showed higher genetic diversity. Among the studied populations, those from Oromia and Amhara displayed gene diversity greater than the grand mean value of 0.67 suggesting that these populations should be targeted to generate baseline information for breeding and conservation. The study also indicated that populations from the Amhara region, where the number of private alleles was 0.09, suggesting a certain level of independent evolution of their gene pools that allowed maintenance of private alleles at a population level.

Analysis of Molecular Variance (AMOVA) revealed very low genetic differentiation among populations (6.74%) and a high proportion (93.26%) of within-population of the total genetic variations. Similar results were reported by previous studies, Tirfessa *et al.* (2020), Muui *et al.* (2016) and Ng'Uni *et al.* (2011) also observed large genetic variation (99.62%, 91.61% and 82%) within populations across Ethiopia, Kenya and Zambia sorghum genotypes, respectively. Basically, the fixation index ranges from 0 (indicating no differentiation between the overall population and its subpopulations) to a theoretical maximum of 1; and it can be considered as small (0 to 0.05), moderate (0.05 to 0.15), large (0.15 to 0.25) or very high (>0.25) (Wright, 1965). The study also revealed the presence of moderate genetic differentiation among populations of sorghum ($F_{ST} = 0.07$). This moderate genetic differentiation might be because of the high variability within individuals and high gene flow through a continuous exchange of genes through sharing common markets among several of the adjacent areas where different populations were collected. The high variability within individuals observed in this study could be due to high gene flow from other populations mainly through the long-distance transfer of pollen and migration of individuals (seed). The overall observed gene flow (N_m) or gene migration value observed in this study was 6.65, which showed the approximate number of individual's migration from one population to the other. Based on Slatkin (1985) and Waples (1987), N_m values grouped into three categories: 0.000 - 0.249 low, 0.25 - 0.99 intermediate and $N_m > 1.00$ high. However, seed exchange among neighbouring sorghum growing regions and higher gene flow ($N_m = 6.65$) could also contribute to the observed low variation among populations. This indicates geographical locations of the populations have non-significant effect on the genetic variation of sorghum germplasm under study.

The Unweighted Neighbour-joining-based cluster and PCoA analysis failed to reveal a clear grouping of germplasm according to their geographical origins possibly confirming the existence of high gene flow among populations. The clustering pattern is weak enough to support the concept of “separation by distance”. This is probably due to gene flow during seed exchange by the farmers and the geographical regions in Ethiopia are not completely different in terms of climate and other soil variables; rather they share similar agro-ecologies. In addition to this, Adugna (2012) suggested that the latest agro-ecological zonation, which classified Ethiopia into 32 agro-ecologies out of which sorghum grows in 12 of them, were more informative to follow for sorghum germplasm collection than geographical regions of origin and the traditional classification using the four agro-ecologies (dry lowland, wet lowland, intermediate, and highland). Similar to this study, both hierarchical clustering and PCoA failed to reveal a clear grouping of Ethiopian sorghum accessions using SSR based on either altitude or region of origin (Tirfessa *et al.*, 2020). In addition, Motlhaodi *et al.* (2014) reported that in sorghum there was no observed geographical origin or racial clustering pattern. In contrast, Missihoun *et al.* (2015) reported that the 61 samples from Benin structured according to their botanical race and morph-physiological characteristics of sorghum grains using 20 SSR. Population genetic structure analysis also revealed the presence of weak sub-groups (K=2) of the five populations of sorghum with a high potential of admixture. All populations possessed genetic background (alleles) from both clusters confirming the presence of gene flow among the populations.

6. CONCLUSION AND RECOMMENDATIONS

Sorghum is one of the most important multipurpose crops grown globally. It is one of the major staple crops grown in the poorest and most food-insecure regions of Ethiopia. Knowledge of the genetic diversity of crop plants is the foundation for the sustainable development of new varieties. So there is a need to characterize the diverse genetic resources using different statistical tools and utilize them in the breeding programme. In the present study, 80 sorghum germplasm collected from five sorghum growing regions of Ethiopia were analyzed for allelic diversity by 11 SSR markers. All the markers were highly polymorphic and informative to describe the genetic diversity and population structure of the genotype.

The study also revealed the existence of high genetic diversity within sorghum populations of Ethiopia, where 93.26% of the total genetic diversity resides within populations. All of the study populations showed high gene diversities ranging from 0.51 to 0.77. Among the studied populations, those from Oromia and Amhara displayed gene diversity greater than the grand mean value of 0.67 suggesting that these locations should be targeted to generate baseline information for breeding and conservation. The Unweighted Neighbour-joining based cluster analysis showed the existence of weak genetic differentiation and all the sorghum population shared genetic background originated from two sub-populations confirming the presence of high genetic exchange or gene flow among populations. This information is significantly crucial for the development of pure sorghum breeding lines. The result of this study suggested that future germplasm collection and utilization strategies should take into consideration the magnitude and pattern of genetic diversity established at the genotypic level.

Based on the study, the following recommendations were made:

- Because of the limited financial resources for sample collection, the present study was delimited to five regions. Therefore, similar studies should be carried out on other populations from the remaining parts of the country to generate a nationwide picture of the diversity.
- The study involved microsatellite-based analysis of the genetic diversity. Thus, further sequence-based identification of genetic diversity analysis would provide additional information.

7. REFERENCES

- Adugna, A., 2012. *Population genetics and ecological studies in wild sorghum [Sorghum bicolor (L.) Moench] in Ethiopia: Implications for germplasm conservation* (Doctoral dissertation, Addis Ababa University).
- Adugna, A., 2014. Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. *SpringerPlus*, 3(1), pp.1-14.
- Adugna, A., Snow, A.A., Sweeney, P.M., Bekele, E. and 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.
- Assar, A.H.A., Uptmoor, R., Abdelmula, A.A., Wagner, C., Salih, M., Ali, A.M., Ordon, F. and Friedt, W., 2009. Assessment of sorghum genetic resources for genetic diversity and drought tolerance using molecular markers and agro-morphological traits. *Institute of Crop Science and Plant Breeding I, Heinrich-Buff-Ring*, pp.26-32.
- Ayana, A., Bryngelsson, T. and Bekele, E., 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.
- Ayana, A., Bryngelsson, T. and Bekele, E., 2001. Geographic and altitudinal allozyme variation in sorghum (*Sorghum bicolor* (L.) Moench) landraces from Ethiopia and Eritrea. *Hereditas*, 135(1), pp.1-12.
- Bekele, E., 1985. The biology of cereal land race populations: I. Problems of gene conservation, plant-breeding selection schemes and sample size requirement. *Hereditas*, 103(1), pp.119-134.
- Billot, C., Rivallan, R., Sall, M.N., Fonckea, D., Deu, M., Glaszmann, J.C., Noyer, J.L., Rami, J.F., Risterucci, A.M., Wincker, P. and Ramu, P., 2012. A reference microsatellite kit to assess for genetic diversity of *Sorghum bicolor* (Poaceae). *American Journal of Botany*, 99(6), pp.245-250.

- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), p.314.
- CSA (Central Statistics Agency). Report on area and production of crops of central statistics agency (2013) Central Statistics Agency, Addis Ababa, Ethiopia.
- CSA (Central Statistics Agency). Report on Area and Production of Major Crops (2014) (Private Peasant Holdings, Meher Season).
- 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.
- Cuevas, H.E., Prom, L.K., Erpelding, J.E. and Brotons, V., 2014. Assessments of genetic diversity and anthracnose disease response among Zimbabwe sorghum germplasm. *Plant Breeding*, 133(2), pp.234-242.
- De Wet, J.M.J. and Harlan, J.R., 1971. The origin and domestication of sorghum bicolor. *Economic Botany*, 25(2), pp.128-135.
- De Wet, J.M.J., 1978. Special paper: systematics and evolution of sorghum sect. Sorghum (Gramineae). *American Journal of Botany*, 65(4), pp.477-484.
- Disasa, T., Feyissa, T. and Admassu, B., 2017. Characterization of Ethiopian sweet sorghum accessions for 0 Brix, morphological and grain yield traits. *Sugar Technology*, 19(1), pp.72-82.
- Djè, Y., Forcioli, D., Ater, M., Lefèbvre, C. and Vekemans, X., 1999. Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. *Theoretical and Applied Genetics*, 99(1-2), pp.157-163.
- Doggett, H., 1965. *Striga hermonthica* on sorghum in East Africa. *The Journal of Agricultural Science*, 65(2), pp.183-194.

- Du Plessis, J., 2003. Maize Production, Department of Agriculture, Pretoria, South Africa.
- Earl, D.A., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), pp.359-361.
- Elangovan, M., Jain, S.K. and Patel, N.V., 2012. Characterization of sorghum germplasm collected from gujarat. *Indian Journal of Plant Genetic Resources*, 26(1), pp.42-46.
- Ethiopian Institute of Agricultural Research (EIAR), (2014). Ethiopian strategy for sorghum 2014-2024, Pp 1-10.
- Evanno, G., Regnaut, S. and 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. and 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 (2005). Food and Agriculture Organization of the United Nations.
- FAOSTAT (2017), “Production – Crops –Area harvested/yield/production quantity – sorghum – 2014”, *FAO Statistics* (online database), Food and Agriculture Organization, www.fao.org/faostat/en (accessed 22 September 2017).
- Fetene, M., Okori, P., Gudu, S., Mneney, E.E. and Tesfaye, K., 2011. Delivering new sorghum and finger millet innovations for food security and improving livelihoods in Eastern Africa.
- Geleta, N., Labuschagne, M.T. and Viljoen, C.D., 2006. Genetic diversity analysis in sorghum germplasm as estimated by AFLP, SSR and morpho-agronomical markers. *Biodiversity and Conservation*, 15(10), pp.3251-3265.

- Gepts, P., 1993. The use of molecular and biochemical markers in crop evolution studies. *Evolutionary Biology*, pp.51-94.
- 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 Technology*, 96(9), pp.985-1002.
- Gupta, P.K., Varshney, R.K., Sharma, P.C. and Ramesh, B., 1999. Molecular Markers and Their Applications in Wheat Breeding. *Plant Breeding*, 118, pp.369-390.
- Hall, A.E., 2000. *Crop Responses to Environment*. CRC press.
- Hariprasanna, K. and Patil, J.V., 2015. Sorghum: origin, classification, biology and improvement. In *Sorghum Molecular Breeding* (pp. 3-20). Springer, New Delhi.
- Harlan, J.R. and de Wet, J.M., 1972. A simplified classification of cultivated sorghum 1. *Crop Science*, 12(2), pp.172-176.
- Hartl, L. and Seefelder, S., 1998. Diversity of selected hop cultivars detected by fluorescent AFLPs. *Theoretical and Applied Genetics*, 96(1), pp.112-116.
- Iqbal, A., Sadia, B., Khan, A.I., Awan, F.S., Kainth, R.A. and Sadaqat, H.A., 2010. Biodiversity in the sorghum (*Sorghum bicolor* L. Moench) germplasm of Pakistan. *Genetics and Molecular Research*, 9(2), pp.756-764.
- Jiang, G.L., 2013. Molecular markers and marker-assisted breeding in plants. *Plant Breeding from Laboratories to Fields*, pp.45-83.
- Jonah, P.M., Bello, L.L., Lucky, O., Midau, A., Moruppa, S.M. and Moruppa, S.M., 2011. Review: The importance of molecular markers in plant breeding programmes. *Global Journal of Science Frontier Research*, 11(5), pp.4-12.
- Kantety, R.V., La Rota, M., Matthews, D.E. and Sorrells, M.E., 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology*, 48(5), pp.501-510.

- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. and Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), pp.1179-1191.
- Kumar, L.S., 1999. DNA markers in plant improvement: an overview. *Biotechnology Advances*, 17(2-3), pp.143-182.
- Kumari, N. and Thakur, S.K., 2014. Randomly amplified polymorphic DNA-a brief review. *American Journal of Animal and Veterinary Sciences*, 9(1), pp.6-13.
- Liu, K. and Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21(9), pp.2128-2129.
- Madhumati, B., 2014. Potential and application of molecular markers techniques for plant genome analysis. *International Journal of Pure and Applied Bioscience*, 2(1), pp.169-188.
- Mengesha, B., Alemaw, G. and Tesfaye, B., 2011. Genetic divergence in Ethiopian coriander accessions and its implication in breeding of desired plant types. *African Crop Science Journal*, 19(1).
- Menz, M.A., Klein, R.R., Mullet, J.E., Obert, J.A., Unruh, N.C. and 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.
- Menz, M.A., Klein, R.R., Unruh, N.C., Rooney, W.L., Klein, P.E. and Mullet, J.E., 2004. Genetic diversity of public inbreds of sorghum determined by mapped AFLP and SSR markers. *Crop Science*, 44(4), pp.1236-1244.
- Missihoun, A.A., Adoukonou-Sagbadja, H., Sedah, P., Agbangla, C., Ahanhanzo, C. and 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. and Laing, M., 2014. A genetic diversity analysis of South African sorghum genotypes using SSR markers. *South African Journal of Plant and Soil*, 31(3), pp.145-152.
- Molosiwa, O.O., 2012. *Genetic diversity and population structure analysis of bambara groundnuts (Vigna subterranea (L.) Verdc.) landraces using morpho-agronomic characters and SSR markers* (Doctoral dissertation, University of Nottingham).
- Mondini, L., Noorani, A. and Pagnotta, M.A., 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, 1(1), pp.19-35.
- 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. *Australian Journal of Crop Science*, 8(1), pp.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(7), p.2135.
- Mutegi, E., Sagnard, F., Semagn, K., Deu, M., Muraya, M., Kanyenji, B., De Villiers, S., Kiambi, D., Herselman, L. and Labuschagne, M., 2011. Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. *Theoretical and Applied Genetics*, 122(5), pp.989-1004.
- Muthusamy, S., Kanagarajan, S. and 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. and 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.

- Ng'uni, D., Geleta, M. and 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.
- Noor, M., Shah, I.A., Ali, F., Shah, S.M.A. and Mehmood, N., 2012. Characterization of sorghum germplasm for various morphological and fodder yield parameters. *African Journal of Biotechnology*, 11(56), pp.11952-11959.
- Omondi, E.O., Debener, T., Linde, M., Abukutsa-Onyango, M., Dinssa, F.F. and Winkelmann, T., 2016. Molecular markers for genetic diversity studies in African leafy vegetables. *Advances in Bioscience and Biotechnology* 7 (2016), Nr. 3, 7(3), pp.188-197.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A. and Schmutz, J., 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 457(7229), pp.551-556.
- Pavel, A.B. and Vasile, C.I., 2012. PyElph-a software tool for gel images analysis and phylogenetics. *BMC Bioinformatics*, 13(1), pp.1-6.
- Peakall, R. and Smouse, P.E., 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28(19), pp.2537-2539.
- Perrier, X. and Jacquemoud-Collet, J.P., 2006. DARwin—dissimilarity analysis and representation for windows. *Version*, 5, p.158.
- Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945-959.
- Ramu, P., Kassahun, B., Senthilvel, S., Kumar, C.A., Jayashree, B., Folkertsma, R.T., Reddy, L.A., Kuruvinashetti, M.S., Haussmann, B.I.G. and Hash, C.T., 2009. Exploiting rice–sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. *Theoretical and Applied Genetics*, 119(7), pp.1193-1204.
- Shehzad, T., Okuizumi, H., Kawase, M. and Okuno, K., 2009. Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its

- evaluation by morphological traits. *Genetic Resources and Crop Evolution*, 56(6), pp.809-827.
- Slatkin, M., 1985. Rare alleles as indicators of gene flow. *Evolution*, 39(1), pp.53-65.
- Smith, C.W. and Frederiksen, R.A. eds., 2000. *Sorghum: Origin, History, Technology, and Production* (Vol. 2). John Wiley and Sons.
- Smith, J.S.C. and Smith, O.S., 1992. Fingerprinting crop varieties. *Advances in Agronomy*, 47, pp.85-140.
- Southern, E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98(3), pp.503-517.
- Stemler, A.B., Harlan, J.R. and de Wet, J.M., 1975. Evolutionary history of cultivated sorghums (*Sorghum bicolor* [Linn.] Moench) of Ethiopia. *Bulletin of the Torrey Botanical Club*, pp.325-333.
- Tadesse, H. and Feyissa, T., 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.
- Tarekegne, Z.G., 2014. Physiological and Agronomic Performance Evaluation of Stay Green (SG) *Sorghum* (*Sorghum bicolor* (L.) Moench) Varieties at Shewa Robit, Amhara Regional State, Ethiopia.
- Tariq, A.S., Akram, Z., Shabbir, G., Khan, K.S., Mahmood, T. and Iqbal, M.S., 2014. Heterosis and combining ability evaluation for quality traits in forage sorghum (*Sorghum bicolor* L.). *SABRAO Journal of Breeding & Genetics*, 46(2).
- Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids Research*, 17(16), pp.6463-6471.
- Teshome, A., Baum, B.R., Fahrig, L., Torrance, J.K., Arnason, T.J. and Lambert, J.D., 1997. *Sorghum* [*Sorghum bicolor* (L.) Moench] landrace variation and classification in north Shewa and south Welo, Ethiopia. *Euphytica*, 97(3), pp.255-263.

- Thudi, M. and 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.
- Tirfessa, A., Tesso, T., Adugna, A., Mohammed, H. and 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
- Vavilov, N.I., 1951. The Origin. *Variation, Immunity and Breeding of Cultivated Plants*, pp.1-364.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T.V.D., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic acids Research*, 23(21), pp.4407-4414.
- Wani, S.P.R. and Albrizio, N.R., 2011. Vaija (2011) Sorghum crop yield response to water. *FAO, Rome, Italy*.
- Waples, R.S., 1987. A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution*, 41(2), pp.385-400.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V., 1991. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids Research*, 18(22), pp.6531-6535.
- Wright, S., 1965. The interpretation of population structure by F- statistics with special regard to systems of mating. *Evolution*, pp.395-420.
- Yaqoob, M., Hussain, N. and 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).
- Yeh, F.C., Yang, R.C. and Boyle, T., 1999. POPGENE version 1.32: Microsoft Windows-based freeware for population genetic analysis, quick user guide. *Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada*, pp.1-29.

8. APPENDICES

Appendix 1. List of sorghum germplasm used in this study

S/No	Germplasm	Region	Zone	Altitude	Longitude	Latitude
1	226057	Amhara	South Wello	2450	E38°48.00	N10°37.00
2	69310	Amhara	North Gondar	1380	E37°52.00	N13°20.00
3	239222	Amhara	South Wello	1580	E39°46.40	N10°58.17
4	SB60	Amhara	North Wello	1488	E39°37.19	N12°6.42
5	239225	Amhara	South Wello	1580	E39°46.40	N10°58.17
6	69291	Amhara	North Shewa	1486	E39°51.00	N09°57.00
7	239151	Amhara	South Wello	1560	E39°46.20	N10°58.33
8	212541	Amhara	North Shewa	1490	E38°54.00	N09°54.00
9	239206	Amhara	South Wello	1580	E39°46.40	N10°58.17
10	212542	Amhara	North Shewa	1580	E38°54.00	N09°56.00
11	212546	Amhara	North Shewa	1540	E38°57.00	N10°03.00
12	226056	Amhara	South Wello	2540	E38°53.00	N10°42.00
13	SB78	Amhara	South Wello	1832	E39°37. 32	N11°32.50
14	226062	Amhara	South Wello	2535	E38°48.00	N10°46.00
15	SB107	Amhara	North Wello	1899	E29°42.00	N11°53.32
16	239249	Amhara	North Wello	2320	E38°55.73	N11°43.65
17	239217	Amhara	South Wello	1580	E39°46.40	N10°58.17
18	237045	Amhara	North Shewa	2020	E39°20.00	N09°30.00
19	239193	Amhara	South Wello	1580	E39°46.40	N10°58.17
20	SB130	Amhara	South Wello	1271	E39°58.24	N10°9.55
21	70933	Amhara	North Shewa	2880	E39°44.00	N09°35.00
22	75284	Amhara	North Shewa	1763	E39°56.00	N10°25.00
23	210951	Amhara	North Wello	1970	E39°35.00	N11°48.00
24	SB107	Amhara	North wello	1899	E39°29.42	N11°53.32
25	210911	Amhara	North Gondar	1960	E37°22.00	N12°27.00
26	17538	Amhara	North Shewa	2603	E39°27.14	N09°56.34
27	239162	Amhara	South Wello	1560	E39°46.20	N10°58.33

Appendix 1. (Continued...)

S/No	Germplasm	Region	Zone	Altitude	Longitude	Latitude
28	210972	Amhara	North Wello	1870	E39°15.00	N11°31.00
29	241255	Dire Dawa	Dire Dawa	1520	E42°10.10	N09°16.01
30	239124	Dire Dawa	Dire Dawa	1640	E41°44.79	N09°29.64
31	239114	Dire Dawa	Dire Dawa	1220	E41°46.38	N09°35.74
32	239134	Dire Dawa	Dire Dawa	1430	E42°06.46	N09°35.83
33	240759	Dire Dawa	Dire Dawa	1900	E42°37.75	N09°29.92
34	69034	Oromia	Arsi	2275	E39°44.00	N08°23.00
35	241207	Oromia	West Harerge	1765	E40°58.34	N09°06.78
36	ESH1	Oromia	East Shewa	2275	E39°44.00	N08°23.00
37	236730	Oromia	West Shewa	2100	E37°28.00	N08°59.00
38	69529	Oromia	West Wollega	1480	E35°04.00	N08°52.00
39	241237	Oromia	East Harerge	2020	E41°54.77	N09°27.24
40	23655	Oromia	West Wollega	1263	E35°18.08	N07°13.15
41	223575	Oromia	East Shewa	1860	E38°34.00	N07°13.00
42	210955	Oromia	West Harerge	1750	E40°51.00	N09°05.00
43	223505	Oromia	East Wollega	1830	E36°38.00	N09°02.00
44	241189	Oromia	West Harerge	2000	E40°54.31	N09°03.16
45	223576	Oromia	East Shewa	1860	E38°34.00	N07°13.00
46	B35	Oromia	East Shewa	1860	E38°34.00	N07°13.00
47	27897	Oromia	West Wollega	1263	E35°18.08	N07°13.15
48	Argity	Oromia	East Shewa	2275	E39°44.00	N08°23.00
49	69344	Oromia	North Shewa	2540	E38°14.00	N10°01.00
50	9142	Oromia	East Harerge	1626	E41°16.55	N09°23.58
51	241208	Oromia	West Harerge	2130	E41°01.29	N09°08.83
52	241279	Oromia	East Harerge	1560	E41°34.62	N09°54.85
53	236733	Oromia	West Shewa	2270	E37°54.00	N08°48.00
54	69540	Oromia	East Wollega	1900	E36°21.00	N08°21.00
55	69227	Oromia	West Harerge	1750	E40°51.00	N09°05.00

Appendix 1. (Continued...)

S/No	Germplasm	Region	Zone	Altitude	Longitude	Latitude
56	223581	Oromia	East Shewa	1930	E38°45.00	N07°21.00
57	241185	Oromia	West Harerge	1600	E40°50.45	N09°06.30
58	241289	Oromia	West Harerge	1430	E40°39.34	N09°11.36
59	241239	Oromia	East Harerge	1560	E41°47.30	N09°26.35
60	236734	Oromia	West Shewa	2100	E37°54.00	N08°48.00
61	69525	Oromia	East Wollega	2260	E36°39.00	N08°47.00
62	27909	SNNP	Bench Maji	1372	E35°12.15	N06°46.60
63	69500	SNNP	Bench Maji	1300	E37°24.00	N05°25.00
64	69546	SNNP	Bench Maji	1780	E36°25.00	N07°10.00
65	69493	SNNP	North Omo	1150	E37°24.00	N05°25.00
66	223573	SNNP	Sidama	1950	E37°48.00	N07°18.00
67	SB29	Tigray	South Tigray	1844	E39°37.10	N12°48.17
68	SB14	Tigray	South Tigray	1738	E39°38.25	N12°45.36
69	SB7	Tigray	South Tigray	1679	E39°38.42	N12°42.27
70	19628	Tigray	Central Tigray	1952	E38°57.8	N14°19.10
71	SB17	Tigray	South Tigray	1738	E39°38.25	N12°45.36
72	SB22	Tigray	South Tigray	1738	E39°38.25	N12°45.36
73	238390	Tigray	Central Tigray	2000	E39°35.00	N13°04.00
74	19626	Tigray	Central Tigray	1946	E38°57.10	N14°19.28
75	19629	Tigray	West Tigray	1276	E38°04.59	N14°07.25
76	238392	Tigray	Central Tigray	2000	E38°04.00	N14°02.00
77	SB15	Tigray	South Tigray	1751	E39°38.26	N12°45.22
78	238406	Tigray	East Tigray	1920	E39°09.00	N13°09.00
79	19624	Tigray	Central Tigray	1670	E38°58.40	N14°19.16
80	238407	Tigray	East Tigray	1870	E39°09.00	N13°09.00

Appendix 2. Plant DNA Extraction Protocol for DArT

Extraction buffer stock

- ❖ 0.35M sorbitol
- ❖ 0.1 M Tris HCL pH. 8.0
- ❖ 5m M EDTA pH. 8.0

To make 500ml

- ❖ 31.9g sorbitol
- ❖ 50ml 1M Tris HCL pH. 8.0
- ❖ 5ml 0.5 M EDTA pH 8.0
- ❖ Fill up to 500ml MiliQ H2O

Lysis buffer stock to make 500ml

- ❖ 0.2M Tris HCL pH 8.0
- ❖ 0.05M EDTA pH. 8.0
- ❖ 2M NaCl
- ❖ 2% CTAB

To make 500 ml:

- ❖ 100ML IM Tris HCL pH. 8.0
- ❖ 50ml 0.5M EDTA pH. 8.0
- ❖ 200ml 5 M NaCl
- ❖ 10 g CTAB
- ❖ Fill up to 500 ml with MilliQ H2O

Sarcosyl stock 5% (w/v)

- ❖ 0.5% (w/v) sodium disulfide
- ❖ 2% (w/v) PVP- 40 (K29-32) sigma
- ❖ Dissolved in required volume of extraction buffer stock, add same volume of lysis buffer stock and 0.4 volume of extraction (=lysis) buffer stock of sarcosyl stock.

PROTOCOLS

1. Homogenize leaf tissue in liquid nitrogen to fine powder
2. Weigh 300 mg of homogenized leaf in a 2 ml eppendorf tube
3. Add 750 μ l EB and 100 μ l SDS (10%)
4. Incubate at 65°C for 20min (in water bath)
5. Add 250 μ l KAc and mix
6. Put the tubes in ice for at least 30 min (possible to take pause)
7. Centrifuge at 14000 rpm for 15min
8. Take supernatant to new eppendorf tube
9. Add 1 volume (same volume as supernatant) of cold isopropanol (freeze at -20°C)
10. Centrifuge at 14000 rpm for 10min and discarded (pour out) the supernatant
11. Dissolve the pellet in 250 μ l TE and after the pellet has dissolved entirely (if not can be put in water bath at 65°C to facilitate dissolution) and add 250 μ l CTAB
12. Incubate at 65°C for 15min (water bath)
13. Extraction with 1 volume of chloroform
 - ✓ Add 1 volume (same volume of chloroform)
 - ✓ Centrifuge at 14000 rpm for 15min
 - ✓ Collect the supernatant in new eppendorf tube (pool same samples in one tube)
14. Repeat the chloroform extraction
15. Precipitate with one volume of cold isopropanol
16. Centrifuge at 14000 rpm for 15min
17. Discard (pour out) the supernatant
18. Dilute the pellet in 100 μ l TE overnight
19. Add 2.5 μ l RNase (1mg/ml) in 100 μ l of TE- DNA
20. Incubate the samples at 37°C for 30min

Appendix 3. Electrophoretic bands of SSR markers Xtxp217, Xtxp096, Gpsb067 and Xtxp057 respectively

