GENETIC DIVERSITY OF ETHIOPIAN COFFEE (*Coffea arabica* L.) COLLECTIONUSING SSR MARKER

MSc. THESIS

GUDETA DIDA HORDOFA

MARCH, 2020

JIMMA, ETHIOPIA

Genetic Diversity of Ethiopian Coffee (*Coffea arabica* L.) CollectionUsing SSR Marker

By

Gudeta Dida Hordofa

MSc. Thesis

Submitted to School of graduate studies College of Agriculture and VeterinaryMedicine department of horticulture and plant sciences in partial fulfillment ofthe requirements for degree of master of sciences in plant Biotechnology

Major advisor: - Kassahun Bantte (Prof.)

Co- advisor: - Tesfaye Disasa (PhD)

March, 2020

Jimma, Ethiopia

SCHOOL OF G TUDIES

JIMMA Y

COLLEGE OF AGRICULTU ERNARY MEDICINE

MSC THESIS APPROVAL SHEET

We, the undersigned, member of the Board of Examiner of the final open defense by Gudeta Dida Hordofa have read and evaluated his thesis entitled "**Genetic Diversity of Ethiopian Coffee (*Coffea arabica* L.) Collection Using SSR Marker**" and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Plant Biotechnology.

_____ _____ _____

Name of the Chairperson Signature Date

_____ _____ _____

Name of Major Advisor Signature Date

_____ _____ _____

Name of Internal ExamineSignatureDate

_____ _____ _____

Name of External ExaminerSignatureDate

DEDICATION

I dedicate this thesis to my mother Dessi Demisse and my brother Daniel Dida for their unflinching support during my study.

STATEMENT OF THE AUTHOR

I declare that the thesis hereby submitted by me for the Degree of Master of science (MSc.) in Plant Biotechnology to the School of Graduate Studies of Jimma University is my own independent work and has not previously been submitted by me or anybody else at another university. The materials obtained from other sources have been duly acknowledged in the thesis.

Brief quotations from this thesis are allowable without special permission provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Head of the major department or the Dean of the School of Graduate Studies when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Name: Gudeta Dida Hordofa

Signature: _____

Place: Jimma University

Submission Date: _____

BIOGRAPHICAL SKETCH

The author was born from his father Dida Hordofa and his mother Dessi Demisse. He attended elementary school at Berity Elementary and Junior School whereas secondary school in Gebra Gurracha Secondary and Preparatory School. Then he joined Jimma University and graduated with BSc. degree in Biology in 2014. After graduation, the author was recruited as a researcher by Ethiopian Institute of Agricultural Research and was assigned at National Agricultural Biotechnology research center in Microbial Biotechnology Research Program as a Junior Researcher in April 2015. Then he joined Jimma University College of Agriculture and Veterinary Medicine in October, 2017 to pursue his MSc. degree in Plant Biotechnology.

ACKNOWLEDGMENT

First of all, I thank God for bringing me this far in my academic pursuit. My sincere gratitude goes to Ethiopian Institute of Agricultural Research for granting my study leave and financial support. Thanks to Jimma University for admitting me to study MSc. program in Plant Biotechnology. I acknowledge National Agricultural Biotechnology research center administrative staff for availing and providing me the molecular lab facilities and all consumable services. I highly acknowledge the special assistance from the Ethiopian Biodiversity Institute for the supply of *Coffea arabica* from their germplasm collections site, which constituted the plant materials used for this research work.

My special appreciation goes to my advisor professor Kassahun Bantte (PhD) for his valuable comments and suggestions during proposal and thesis manuscript preparation. Furthermore, I am very grateful to him for filling my knowledge gaps that were found in the study and encouragement that I have received throughout the work. His willingness to share his rich experience on molecular study has greatly contributed to the success of this study.

I would like to express my earnest gratitude to my Co-advisor Tesfaye Disasa (PhD) (National Agricultural Biotechnology Research center, Center director) for his encouragement, advice throughout the work and providing me all the PCR components, chemicals and all consumable materials necessary for this study. I am deeply grateful for his unreserved professional guidance and supervision in the course of my study. I also thank for his supervision during my laboratory work, advises and devotion whenever I needed them and during data analysis. In general, Dr. Tesfaye, I do not have a word to express my gratitude; you have been around me all the time.

My special thanks go to Tilahun Negesse (PhD candidate), Addis Ababa University (Institute of Biotechnology) for providing me guidance and assistance in laboratory work and helping me in molecular data analysis. I also wish to express my deep sense of gratitude to my friends, Fedessa Kebede, Doni Hinsene, Demilew and Mogos for their genuine cooperation in every aspect.

I am also thankful to staff members of Molecular Biotechnology Laboratory for their generous assistance. Special thanks go to Dejene, Yitaseb and Lidiya. I am sincerely grateful to Diriba Guta, Adenech Teshome and Lellise Legesse for their assistance in statistical data analysis.

Special thanks also goes to staff members of Jimma University, College of Agriculture and Veterinary Medicine who gave me valuable input and support. I would like to express my special thanks to Plant Biotechnology lecturers.

I would like to express my deepest thanks and regards to my dear family Daniel Dida, Tolera Dida, Fire Dida, Maritu Dida, Dessi Demisse, Dida Hordofa and other close relatives for their continuous love and daily prayers for my happiness and better life. Thank you very much for your love and encouragement.

I am highly indebted to my classmates Yonas Getachew, Birhanu Babiye and Tsegaye Abebe for your encouragement and diverse forms of assistance during all these years. I must indicate that I have really enjoyed your accompany and may we all continue to hold on to this good spirit of diligence. God bless us all. I am indebted to everyone who might have made an impact in my life in one way or the other particularly regarding my academic pursuit, especially my friends, co-workers, family members, and all who have good thoughts about me.

Finally, there is special individual that I need not to forget, Kelkidan Tesfu, for her valuable support such as, chemicals, consumable laboratory materials and good counsel. I would not have been able to achieve this great feat in life without her support. I say may the Almighty God richly bless you.

ABBREVIATIONS AND ACRONYMS

AFLP Amplified fragment length polymorphism

AMOVA Analysis of molecular variance

DNA Deoxyribonucleic acid

EBI Ethiopian Biodiversity Institute

EIAR Ethiopian Institute of Agricultural Research

NABRC National Agricultural Biotechnology Research Center

NJ Neighbor Joining

PCoA Principal Coordinate Analysis

PCR Polymerase chain reaction

PIC Polymorphic information content

RAPD Randomly amplified polymorphic DNA

RAMP Randomly amplified microsatellite polymorphism

RFLP Restriction fragment length polymorphism

SNP Single nucleotide Polymorphism

SSR Simple sequence repeat

SNNP Southern Nation Nationalities and Peoples

UPGMA Unweight Pair Group Methods with Arithmetic average

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (*continued*)

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF APENDIX

Page

ABSTRACT

Many studies reported wide genetic variability of Arabica coffee accessions collected from Ethiopia compared with commercial cultivars demonstrating the great potential of these accessions for future breeding purposes. However,little work has been done on the molecular genetic structure and diversity of Ethiopian Arabica coffee populations. Hence, this study was initiated with the objective of studying the genetic diversity of Coffea arabica populationscollected from different regions of Ethiopia using SSR markers. A total of 20 SSR markers were used to genotype 86 accessions and produced a total of 112 alleles, ranging from 3 to 11 with an average of 5.6 alleles per locus.All the loci across the entire populations were found to be highly polymorphic and informative with PIC values ranging from 0.45 to 0.75 with a mean of 0.6 confirming the good discriminatory power of the SSR loci used.Average observed heterozygosity andallelic richness across all populations ranged from (0.22-0.27 and 3.52-4.26), with a mean of 2.43 and 3.97, respectively. AMOVA showed 63% of the variation to be within populations, 33% among individuals within populations and 4% among population based on geographical origin.The smallerFst (0.037) observed indicates the presence of lower population geneticdifferentiation as a result of higher gene flow (Nm =2.45) between the C. arabica populations. The lowest mean genetic distance observed between C. arabica populations was 0.21. The populations from Bale and Hararge were highly distant from other populations.The unweighted pair group methods with arithmetic meanbased cluster analysis and principal coordinate analysis poorly grouped the individuals into distinct clusters confirming the presence of population admixture due to the long distance movement of C. arabica seeds and high gene flow among populations of adjacent geographical regions. The observed higher genetic variability in all populations indicates that the country has huge coffee genetic diversity which can serve coffee improvement.Comparatively, the populations from Omo, Ilubabor and Benchi Maji were more diverse than other populations. Thus, special attention on these populations may be useful in future Coffea arabica breeding program, germplasm conservation.

Key words:*Coffea arabica,*Genetic diversity,Gene flow, Heterozygosity, SSR marker.

1. INTRODUCTION

Coffee belongs to the genus *Coffea* in the Rubiaceae family and comprises approximately 124 species, but only 10 species are cultivated (Davis *et al*., 2011). Among the hundreds species in the genus, *Coffea arabica* and *Coffea canephora (*Robusta coffee) are the two most important commercial species with approximately 60% and 40% of global coffee production, respectively (ICO, 2019). *Coffea arabica* produces a highquality beverage, with pleasant aroma and flavor (Tran *et al*., 2016). *Coffea arabica* is an allotetraploid (2n = 4x = 44) that originated from two different diploid (2n = 2x = 22) ancestors *Coffea canephora and Coffea eugenioides* (Lashermes *et al*., 1999). Due to the nature of its origin, reproductive biology (i.e. autogamy), evolution and the narrow gene pool from which it spread around the world, Arabica coffee has very low genetic diversity (Vega *et al.,* 2008). However, the indigenous cultivars of Arabica coffee in Ethiopia have wide genetic variability in natural Arabica coffee populations (Mesfin & Bayeta, 1987; Kassahun *et al*., 2014).It is self-compatible and mostly reproduces by self-fertilization (Alemayehu *et al*., 2010).

Geographically, most of the coffee species are originated from tropical African countries: Ethiopia for the tetraploid *C. arabica*. During the early centuries, the Arabica coffee species was disseminated to other parts of the world where it is being produced in mass nowadays. However, the Arabica coffee plants in major producing areas such as Latin and Central America, and Asian countries are believed to have a narrow genetic bases attributed to the few seeds/plants used for dissemination, the successive genetic reduction due to human impacts and reproduction nature (Pestana *et al*., 2015).

Coffee is one of the most valuable commodity crops in the world trade. More than 90% of its production occurs in developing countries providing an income for millions of smallholder farmers around the world that are dependent on coffee for their subsistence (Tran *et al*., 2016). It contributes largely to the economy of more than 50 countries in Asia, Latin America and Africa (Davis *et al*., 2011). Beside its tremendous contribution to the foreign exchange, it serves as a means of livelihood for millions of people and plays a vital role in their socio economic values (Stieger *et al*., 2002). The share of coffee in total export earnings has a

positive and significant reflection on economic growth for developing and least developed countries (Al-Abdulkader *et al*., 2018).

Coffee production in Ethiopiahas long tradition which dates back to dozens of centuries. Ethiopia is endowed with a good production environment for growing coffee with a combination of appropriate altitude, rain fall, and temperature. The country possesses a diverse genetic base for Arabica coffee with considerable heterogeneity(Habtamu*et al.*, 2018) and is the center of origin for *Coffea arabica*(FAO, 1968). The total area coverage of coffee land in the country is 1.2 million hectares. Of which 900,000 hectares of land is estimated to be productive. According to some studies, about 92-95% of coffee is producedby 4.7 million small scale farmers and 5-8% large scale plantations (Aigaforum, 2018). Theannual coffee production in the country is 500,000-700,000 tones andthe average national productivity is 7 quintal per hectare. Ethiopia accounts for around 3% of the global Coffee market. Around 40% of foreign income comes from coffee. In 2017/18, Ethiopia's export was estimated at 3.98 million bags of coffee (Aigaforum, 2018).

Coffee breeding programs invested intense efforts to release cultivars with high productivity, with climate change and simultaneous change in biotic and abiotic pressures stresses tolerance and high biochemical quality of the beans (Tran *et al*., 2016). However, several factors are limiting the genetic gains in breeding programs (Vieira *et al*., 2010). Commercial coffee plants are originated from a limited number of cultivars; consequently, only narrow genetic base is available to support breeding programs.Susceptibility of commercial coffee cultivars to pests and diseases,climate change and emerging pest and outbreak of diseases are important challenges to global coffee production (Bunn *et al*., 2015). In addition, the reproductive behavior of *C. arabica* also contributes to the narrow genetic diversity available in this species. As expected, several studies based on molecular markers demonstrated the low genetic variability available among commercial *C. arabica* varieties (Setotaw *et al*., 2013; Pestana *et al*., 2015; da Silva *et al*., 2019).

Many studies reported wide phenotypic diversity of Arabica coffee accessions collected from Ethiopia regarding leaf size, height, biotic and abiotic stresses tolerance and yield (Tran *et al.*,

2016) and the presence of higher genetic variability compared with cultivars, demonstrating the potential of these accessions for breeding purposes (Aerts *et al.*, 2013; Sant'Ana *et al.*, 2018).The accessions also showed a great variability for chlorogenic acids, lipids, sucrose and diterpenes contents of coffee beans and the discovery of decaffeinated coffee varieties (Silvarola *et al.,* 2004)underlines the great potential existing in Ethiopian coffee accessions for Arabica coffee future breeding programs.There are few studies on molecular genetic diversity of Ethiopia coffee (Alemayehu*et al.*, 2010; Kassahun *et al.*, 2014)whichwere restricted to germplasm from some geographical regions of the country, especially in South-Western part of the country. But each coffee producing region has its own unique germplasm which should be conserved independently and can be used for future breeding efforts.

Presently, Ethiopian coffee genetic resources are under greatest threat, mainly due to deforestation of the natural habitat for timber and crop production, replacement of farmers' variety by a few high yielding varieties, establishment and expansion of modern plantations and illegal and legal settlements, increasing incidence of drought, the spread and severity of devastating fungal pathogens (Kassahun, 2006; da Silva *et al*., 2019). In eastern and south eastern parts of the country, farmers are obliged to switch their attention more towards khat production. As a result, there is a huge land use shift in favor of khat production and consequently affecting coffee genetic resources especially in Yirgachefe (known for its premium price fetching coffee),Harar and Bale.

To minimize the loss of coffee genetic resources, Ethiopian Biodiversity Institute has established field gene banks (ex-situ conservation) in different parts of the country to conserve coffee genetic resources.However, even if more than 4000 *Coffea arabica* accessions are conserved in EBI gene banks,the diversity of the accessionswere not studiedusing modern molecular genetictools,Rising temperatures, diseases and pests, lack of care and extreme weather are putting pressure on this field gene banks.In addition, *C. arabica'*s natural habitat in the highland forests of Ethiopia (the world's only in-situ collection) is rapidly disappearing, thatseverely threaten the future of world's coffee production (Krishnan, 2017).

To design any conservation strategy, analyzing the genetic diversity using different marker systems are vital for sustainable use, the efficient utilization of plant germplasm for improvement purpose and conservation strategy. Therefore, the knowledge about population structure and genetic relationships of the Ethiopian coffee accessions is important for conservation and efficient utilization of these genotypes in Arabica coffee future breeding program.

A number of DNA based techniques were used in different coffee genetic studies. These include RAPD (Diniz *et al*., 2005), AFLP (Steiger *et al*., 2002), ISSR (Kassahun *et al.,* 2014) and microsatellite (SSR) markers (Lashermes *et al.,* 1995; Alemayehu *et al.,* 2010; Geleta *et al*., 2012; Motta *et al*., 2014; Sousa *et al*., 2017**;**da Silva *et al*., 2019). Among the several types of molecular markers, microsatellites (SSR) are most commonly used in genetic diversity study because of several advantages, including high degree of polymorphism, repeatability, reproducibility, codominance, technical simplicity, speed and multi allelism (Vieira *et al*., 2010). Hence, this study wasinitiated with the following objective.

Objective:

- To study the level of genetic diversity among and within populations of *Coffea arabica* accessionsthereby to generate information for conservation, breeding program, sustainable use and germplasm management in the country

2. LITERATURE REVIEW

2.1 Botany, Origin and Distribution of Arabica Coffee

2.1.1 Botany

A tropical woody genus, *Coffea* belongs to the Rubiaceae family. *Coffea arabica* is an allotetraploid (2n = 4x = 44) that originated from two different diploid (2n = 2x = 22) wild ancestors, *C, canephora* and *C, eugenioides* (Lashermes *et al*., 1999). Due to the nature of its origin, reproductive biology, evolution and the narrow gene pool from which Arabica coffee spread around the world, it has very low genetic diversity (Lashermes *et al.*, 1999; Anthony *et al*., 2002). It is self-compatible and most of the time reproduces by self-fertilization (Fazuoli *et al*., 2000; Alemayehu *et al*., 2010).

The Arabica coffee tree is a small tree with the potential in the wild to reach 9 to 12 meters in height, growing at an altitude of 900 to 2,000 meters above sea level. From seed germination to first fruit production, the coffee plant takes about three years, whenit reaches full maturity. The fruit of Arabica coffee is known as a cherry and the seed inside is known as the bean. The fruit is comprised of the epicarp (skin), mesocarp (pulp),endocarp (parchment), integument (silver skin), endosperm (bean) and embryo. The tree has an open branching system with a main vertical (orthotropic) stem from which ariseprimary plagiotropic branches from “head of series” buds. The four to six serialbuds generate either flowers or orthotropic suckers. The leaves are opposite, dark green,shiny, and waxed. The flower consists of white, five-lobed corolla, a calyx, five stamens,and the pistil. The ovary at the base of the corolla consists of two ovules, which whenfertilized become two coffee beans (Meyer, 1968; Wintgens, 2009).

2.1.2 Origin and distribution of arabica coffee

There are two main species of commercially important coffee, Arabica coffee (*Coffea arabica*) and Robusta coffee (*C. canephora*) and two minor commercial species, Liberica coffee (*Coffea liberica*) and Excelsa coffee (*Coffea excelsa*) (Mangal, 2007). The Arabica coffee is originated in the highland forests of south-western Ethiopia whereas that of

Robusta coffee in the wild forest of humid hot lowland areas of Middle East, Central, and West Africa. In south-western part of Ethiopia about 400,000 ha of an ancient forest where coffee occurs as understory shrubs still remain (Ervine, 1969). Moreover, there is also high genetic diversity of coffee in the region that is used as source of plant stock for the selection of disease resistance, drought resistance, high yields and top quality in terms of aroma and flavor (Melkamu, 2015). Arabica coffee is the earliest known and most widely distributed coffee species. The distribution of Arabica coffee from its center of origin to Yemen (secondary origin) is probably through Persian invaders in 575 and Arab traders in 13th century. From Yemen, it spread to Middle Eastern countries and India in the 14th century from where European traders subsequently took coffee seeds to other parts of Asia, Africa, eventually to South America (Ervine, 1969).

2.2 Taxonomy and Ecology of Arabica Coffee

2.2.1 Taxonomy

Coffee belongs to the genus *Coffea* in the Rubiaceae family, and is mostly grown in the tropical and subtropical regions of the world. Of the hundred known species in the genus *Coffea*, Arabica coffee (*Coffea arabica* L.) and Robusta coffee (*Coffea canephora* P.) are the two most important commercial species. In terms of quality profile rating, Arabica coffee stands out and contributes more than 60 percent of the world coffee production (Lashermes *et al*., 1997; Anthony *et al.*, 2002; Stieger *et al.*, 2002; ICO, 2019).

The first botanical description of a coffee tree under the name *Jasminum arabicanum* was made in 1713 by A. de Jussieu, at Amsterdam botanic garden. However, Linnaeus (1737) classified it as a separate genus *Coffea* with the then only one known species Arabica coffee. However, many more species of *Coffea* were discovered during exploration of the tropical forests of Africa since the second half of the nineteenth century. Efforts of several botanists to describe the species in the genus had led to confusion such that numerous epithets were later discovered to be synonyms of the same species. However, the respective work of Lebrun (1941) in Central Africa (Zaire) and Chevalier (1947) in Africa and Madagascar on coffee is remarkable. Chevalier (1947) grouped the species within the

genus *Coffea* into four sections: Argocoffea, Paracoffea, Mascarocoffea, and Eucoffea. However, Leroy's (1967) excluded Argocoffea from the genus *Coffea* and classified as a sub-genus of Psilanthus, because the seed does not resemble coffee beans. Classifications within Eucoffea and Mascarocoffea now mostly fit correctly in the genus *Coffea*. Within the Eucoffea, there are five subsections (Chevalier, 1947) based on some diverse criteria: tree height (Nanocoffea), leaf thickness (Pachycoffea), fruit color (Erylhrocoffea, Melanocoffea) and geographical distribution (Mozambicoffea) (Lashermes, 1997).

2.2.2 Ecology of arabica coffee

In Ethiopia, coffee grows at various altitudes, ranging from 550-2,750m above sea level. However, the bulk of Arabica coffee is produced in the altitudes ranging from 1,300 - 1,800 m. Annual rainfall in the coffee-growing regions of the country varies from 1,500-2,500mm. Where precipitation is less, as in the eastern part of the country (about 1,000 mm), coffee is supplemented with irrigation. Arabica coffee grows best in the cool, shady environment of Ethiopian highland forests (Aerts *et al*., 2013; Habtamu *et al*., 2018).

The ideal temperatures for coffee production are considered to be 15-25ºC. These temperatures prevail in most of the country's coffee growing areas. Arabica coffee requires fertile, friable, loamy soils, with a depth of at least 1.5m and relatively high water holding capacity. The fertility of coffee soils is naturally maintained through organic recycling of litters falls from coffee, and shade trees. Relatively acidic soil is suitable for Arabica coffee pH (5-6.8). The soil of most coffee growing regions in Ethiopia satisfies these characteristics of coffee soil (Melkamu, 2015).

2.3 Economic Importance of Coffee

Coffee is the most important agricultural commodity in the world. More than 90% of its production occurs in developing countries providing an income for millions of smallholder farmers around the world that are dependent on coffee for their subsistence (Tran *et al*., 2016). Worldwide, an estimated 125 million people are dependent on coffee for their livelihoods. More than 121 countries including Ethiopia export and/or re-export coffee to more than 165 countries worldwide. More than 50 developing countries, 25 of them in

Africa, export coffee in different parts of the word(NCA, 2017).In many coffee producing countries, beside its tremendous contribution to the foreign exchange, it serves as a means of livelihood for millions of people and plays a vital role in their socio economic values (Stieger *et al*., 2002). The share of coffee in total export earnings has a positive and significant reflection on economic growth (Tigist, 2015) and on the Gross Domestic Product (GDP) for most of the producing countries, particularly developing and least developed countries (Al-Abdulkader *et al*., 2018).

The agriculture-based Ethiopian economy is also highly dependent on coffee asforeign exchange earnings. Moreover the sector provides income for approximately eight million smallholder households which are participating in the various activities in the value chain of coffee (Melkamu, 2015). Coffee is grown in both highland and lowland conditions, nurtured with care by the farmers, giving the beans a range of unique flavors and textures. Like any commodity trade, the coffee trade has been characterized by boom and bust cycles mainly due to an imbalance of supply and demand. In the early 20^{th} century, attempts to stabilize coffee prices rested on efforts of individual countries, especially in Brazil. In the following decades, the price of coffee has alternately soared and dived, with the market hitting the lowest at 40 cents per pound in New York, while farmers' production costs amounted to about 70 cents a pound. This has led to poverty and food insecurity in countries where the majority of coffee producers are subsistence farmers (Thurston, 2013).

Coffee production is generally characterized by considerable instability, with a large crop one year followed by a smaller crop the next. In the world coffee market, as is the case for many commodities, price volatility is a major concern for all stakeholders. In exporting countries, price volatility leads to instability in producer incomes and uncertainty of export earnings and tax revenues. In importing countries, price volatility affects profit margins for roasters, traders, and stockholders. All these factors make the coffee crop less attractive throughout the supply chain, especially to growers, who will seek other, more remunerative crops to replace coffee. Despite these challenges, world coffee production has grown steadily since the 1960s, although it will be difficult to maintain this trend due

to the continued rise in production costs, problems related to climate change, and the higher incidence of pests and diseases (Krishnan, 2017).

Coffee does not only have an economic benefit, but also has its own social values**.** Coffee plays a vital role in both cultural and social life of Ethiopian community. Among coffee producing countries in the world, Ethiopia is the first country in consumption of coffee. From the average total annual production about 50% is consumed in the country. Preparation and drinking of coffee is a unique culture in Ethiopia; coffee ceremony. Coffee is not drunk alone. It is a social activity to be shared with others. Sharing coffee with others means you are at peace with them and cultivates community and friendship. Coffee is typically made by roasting and brewing on a small charcoal burner. Cups (cinis) are usually laid out in a square on a tray dressed with fresh grass and served with a snack such as fresh popcorn (Melkamu, 2015).

2.4. Genetic Markers for Genetic Diversity Analysis

A genetic marker is a gene or DNA sequence with a known location on a chromosome that can detect variation in either a protein or DNA sequence. The analysis of genetic diversity within and among populations routinely involves the use of different genetic markers. Nowadays, genetic markers are used in both basic plant research and plant breeding programs to characterize plant germplasm, for gene isolation, marker assisted introgression of favorable alleles, production of improved varieties and to obtain information about the genetic variation of populations (Henry, 2001; Adebabay, 2015). Genetic markers can be categorized in to three broad classes; morphological markers; those based on visual assessable traits, biochemical markers; those based on gene product and molecular markers those based on a DNA assay. Each of the various marker systems utilized for the characterization has its own specific merits and demerits (Henry, 2001).

Different genetic markers have been used to characterize*coffea arabica* germplasm. These markers include; morphological (Mesfin and Bayeta, 1987; Carvalho, 1988), Biochemical (Silvarolla *et al*., 2000) and PCR-based markers such as RAPD (Diniz*et al*., 2005; Gichimu *et al*., 2012), AFLP (Steiger *et al*., 2002), ISSR (Kassahun *et al*.,2014) and

microsatellite (SSRs) markers (Lashermes *et al.*, 1995; Cubry *et al*., 2008; Alemayehu *et al.*, 2010; Hussein *et al*., 2017; Omingo *et al.*, 2017 and da Silva*et al*., 2019).

2.4.1. Morphological markers

Morphological markers are visually distinguishable qualities like seed structure, flower color, growth habit and other important agronomic traits. Morphological markers are easy to use, with no requirement for specific instruments. They do not require any specialized biochemical and molecular technique. Since ancient times, humans have successfully used various morphological markers to investigate the variation for utilization in plant breeding.Walyaro (1983) successfully investigated the diversity of eleven Arabica coffee genotypes using different morphological characteristics. Gichimu and Omondi (2010) also determined the morphological diversity among five newly developed and two existing commercial cultivars of Arabica coffee in Kenya. The study demonstrated low morphological variation hence, low genetic variation among the varieties tested. Montagnon and Bouharmont (1996) classified wild and cultivated coffee genotypes from Ethiopia according to their geographic origin using 18 agro morphological traits. The FAO Coffee collection team has observed phenotypic variation in branching habit, young leaf color, fruit color, persistence of sepals, leaf and fruit size on Ethiopian coffee. Many authors observed growth habit variation such as compact and spreading type genotypes from national coffee collections (Dessalegni, 2017).

2.4.2. Biochemical markers

Biochemical markers, or Isozymes, are multi-molecular forms of enzymes which are coded by various genes, but have the same functions. They are allelic variations of enzymes and thus gene and genotypic frequencies can be estimated with biochemical markers. Biochemical markers have been successfully applied in the detection of genetic diversity, population structure, gene flow and population sub division. They are codominant, easy to use and cost effective. However, they are less in number; they detect less polymorphism and they are affected by various extraction methodologies, plant tissues and different plant growth stages (Mondini *et al*., 2009).

The study of diversity based on biochemical markers is depending on the separation of proteins into specific banding patterns. It is a fast method when compared to another diversity study method and requires only small amounts of biological material. However, only a limited number of enzymes are available and thus, the resolution of diversity is limited. Allozymes, being allelic variants of enzymes, provide an estimate of gene and genotypic frequencies within and between populations (Mondini *et al.,* 2009).

2.4.3. Molecular markers

Molecular markers are based on naturally occurring polymorphisms in DNA sequences due to base pair deletions, insertions, and substitutions. Molecular markers are superior to both morphological and biochemical markers because they are highly polymorphic relatively simple to detect, abundant throughout the genome even in a highly inbred cultivars and completely independent of environmental conditions and can be detected at any stage of plant development (Bhandari *et al.,* 2017). However, major disadvantage is the need for technically more complex equipment.

The rapid development of molecular techniques over the last few decades, now offers a good technical approaches for plant genotyping or genome analysis. Which technique is most appropriate depends upon (i) the extent of genetic polymorphism required; (ii) the analytical or statistical approaches available for the technique's application, and (iii) the pragmatics of time and costs of materials (Parker *et al*.,1998).The discovery of the polymerase chain reaction (PCR) was a landmark in molecular marker evolution and has proved to be a unique process for thedevelopment and utilization of a battery of new very sensitive and quick approaches,such as AFLP or microsatellites (SSR) (Paglia and Morgante, 1998). PCR is an *in vitro* technique that allows the amplification of a specific DNA region that lies between two regions of known DNA sequence. Orozco-Castillo *et al.* (1994) demonstrates the power of the polymerase chain reaction technology for the generation of genetic markers for long-lived perennial tree and bush crops.

2.4.3.1. Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) was developed first and was initially used for human genome mapping (Botstein *et al*., 1980). Later, RFLP markers are one of the most important tools for plant genome mapping (Jiang, 2013) and they are classified as hybridization-based markers. RFLP involves the extraction of genomic DNA followed by its digestion with specific restriction enzymes, which cut the DNA into fragments. RFLP results when there is variation in restriction enzyme cleavage sites, arising due to base substitutions, insertions, deletions or translocations in the genomic DNA (Gupta *et al*., 2002).

The major strength of RFLP markers are their high reproducibility, high genomic wide abundance, codominant inheritance, and good transferability between laboratories which provides locus specific markers that allow synteny (conserved order of genes between related organisms) studies (Kesawat and Das, 2009; Jiang, 2013). Still, there are several limitations for RFLP analysis: it requires the presence of high quantity and quality of template DNA. RFLP markers can be applied in diversity, phylogenetic and fingerprinting studies ranging from individuals within populations or species to closely related species. It is widely used in gene mapping studies because of their high genomic abundance, the ample availability of different restriction enzymes and random distribution throughout the genome (Kesawat and Das, 2009).

2.4.3.2. Random Amplified Polymorphic DNA (RAPD)

Random Amplified Polymorphic DNA (RAPD) is based on the amplification of genomic DNA with single primer of arbitrary nucleotide sequence, usually 10 bp long (Williams *et al*., 1990). RAPD Primers detect polymorphisms in the absence of specific nucleotide sequence information and the polymorphism detected is used as genetic markers to construct genetic maps. RAPD polymorphisms arise when genomic regions vary for the presence or absence of complementary primer annealing sites due to insertion or deletion between two priming sites, which results in different lengths of the amplification products. The main advantages of the RAPD technology include: (i) suitability for work on

anonymous genomes, (ii) applicability to problems where only limited quantities of DNA are available and (iii) less expense. However, there are various limitations and considerations in RAPD analysis, which include specificity of the marker in genome scanning (Hadrys *et al*., 1992), reproducibility of amplification products (Williams *et al*., 1990), and unclear and non-reproducible fragments (Pan *et al*., 1997).

Yet, in coffee, RAPDs have been used to evaluate genetic diversity among coffee varieties (Diniz*et al*., 2005). RAPD markers generated by arbitrary decamers have been successfully employed to detect the level of genetic polymorphisms between different coffee species and between *Coffea arabica* genotypes. The information gained from RAPD profiles were used to construct dendrogram and these were consistent with the known history and evolution of *Coffea arabica*. Material originating from Ethiopia and the arabica sub-groups *C. arabica* variety typica and *C. arabica* variety bourbon were clearly distinguished by RAPD. RAPD analysis therefore reflects morphological differences between the sub-groups and the geographical origin of the coffee materials (Diniz*et al*., 2005).

2.4.3.3. Amplified Fragment Length Polymorphism (AFLP)

AFLP is a PCR-based molecular marker for the rapid screening of genetic diversity studies and intra specific variation. It is a potent fingerprinting technique for genomic DNAs of any origin or complexity and rapidly generates a number of highly replicable markers that allow high resolution genotyping. The strength of AFLPs includes its high genomic abundance, highreproducibility, highly polymorphic, generation of many informative bands per reaction; small amount of template DNA is needed, and the fact that no sequence information for primer construction is required (Saal and Wricke, 2002). Possible reasons for AFLP polymorphisms are: (i) sequence variations in a restriction site, (ii) insertions or deletions within an amplified fragment, and (iii) differences in the nucleotide sequence immediately adjoining the restriction site.

AFLPs have been used for the analysis of genetic diversity, DNA fingerprinting, the construction of linkage maps (Cho *et al*., 1998) and to locate traits of interest (Hartl*et al*., 1999). Pearl *et al.* (2004) used AFLPs to construct a genetic linkage map on a pseudo-F2 population of Arabica coffee *(Coffea arabica* L.*)* derived from a cross between the cultivars Mokka hybrid and Catimor. Sixty three of these populations were selected on the basis of plant height distribution to construct a linkage map. A total of 456 dominant markers and eight codominant markers were generated from 288 AFLP primer combinations (Dessalegni, 2017).

2.4.3.4. Microsatellite or Simple Sequence Repeat (SSR)

SSRs are very short motifs (about 1-6bp) usually characterized by a high degree of repetition and occur at many thousand loci in the nuclear genome (Singh *et al*., 2010). They are ubiquitous and highly polymorphic, owing to the mutation affecting the number of repeat units. The hyper variability of SSRs among related organisms makes them an informative and excellent choice of markers for a wide range of applications in population and evolutionary biology (Chen *et al*., 2009), which include estimate genetic diversity, study population structure and gene flow and develop gene mapping. The popularity of microsatellites stems from a unique combination of several important advantages: the relatively abundance with uniform genome coverage, the enormous extent of allelic diversity, the hyper variability, the codominant inheritance, the ease of detection by PCR using pair of flanking primers, and requirement for only a small amount of template DNA (Chen *et al*., 2009). These characteristics make this marker a favorite for genetic mapping studies, marker assisted selection (MAS), genetic diversity surveys, QTL analysis and germplasm maintenance. Despite these advantages, the number of SSR markers for the genus *Coffea* is limited mainly by the difficulties in developing specific primers (Ferrao*et al.,* 2015).

Genetic structure of Arabica coffee and diversity of wild and cultivated accessions of *Coffea arabica* were assessed using Simple Sequence Repeat (SSR) markers (Silvestrini *et al*., 2007). In addition, accessions of *C. eugenioides, C. racemosa* and *C. canephora* were also sampled. By cluster analysis based on Jaccard's coefficient, all species were

distinguished and cultivated. *C. arabica* accessions were distinguished from spontaneous and sub-spontaneous ones. The Brazilian cultivars were distinguished from Yemen-cultivated accessions; however, both groups exhibited a very low genetic diversity. Their result similar with the previous initial remark, that the cultivated germplasm of *C. arabica* has a narrow genetic base. The SSR and EST-SSR markers were successfully used for genetic diversity evaluation of valuable accessions of a Brazilian coffee breeding program. The SSR markers were more efficient in this evaluation, especially in differentiating *C. arabica* related accessions. Nevertheless, the combined use of gSSR and EST-SSR markers was recommended by Missio*et al*. (2011), because they may provide complementary information. Their investigation provided a selection protocol of a more informative combination of gSSR and EST -SSR markers for further studies.

Generally, the use of molecular markers of the SSR types in the study of diversity was efficient in carrying out the molecular characterization of coffee genotypes between and within *C. arabica and C. canephora*. Motta *et al.* (2014) reported that microsatellites markers were efficient in estimating the genetic similarity and could be used to increase the efficiency in classifying the materials and selecting the candidates for parental crosses. Therefore, for this study, SSR markers selected because the technique is simple, fast, cost effective and easily accessible.

2.4.3.5. Single Nucleotide Polymorphism (SNP) marker

Single Nucleotide polymorphism (SNP) is a substitution of a single nucleotide that occurs at a specific position in the genome sequence where each variation is present at a level of more than 1% in the population. SNPs are the most abundant form of DNA variation and markers developed based on SNPs could reach higher density than any other marker types (Li *et al*., 2009). With the advent of SNP markers, the possibility of simultaneous analysis of a set of loci becomes more real. A SNP is created when a single nucleotide base in a DNA sequence is replaced with a different nucleotide base. The SNP markers are based on the most fundamental alterations of the DNA molecule and mutations in the bases of unique chain of nitrogenous bases (Adenine, Cytosine, Guanine and Thymine). SNP are extremely abundant in genomes of an indivduals. (Li *et al*., 2009). High-throughput DNA

sequencing technology has now been widely applied to develop massive genotyping arrays, which allowed fast and efficient identification of SNP markers for large numbers of individuals (Ganal*et al*., 2014).

SNP markers are the most common marker type used in breeding programs. Nowadays, SNP markers had a wide application for different purposes in different crops, such as identification of plant varieties and cultivars, genetic diversity analysis, QTL analysis, construction of high density genetic map and genome wide association analysis (Delourme*et al*., 2008; Han *et al*., 2016).These markers can be associated with the genes that control the main traits of agronomic interest and have low cost per data point. These facts, combined with the development of genomic selection algorithms, have increased the accuracy of the selection methods used in breeding programs for several plant species. However, in order for a chip to be developed for a novel species, it is necessary that the SNPs be known a priori (Crossa*et al*., 2010).

The overall genetic diversity and genetic structure of *Coffea arabica* was studied with SNP markers thatare distributed throughout the *C. arabica* genome (Sousa *et al*., 2017b). The use of these SNP markers allowed the accurate discrimination of all the genotypes analyzed in the study, and even detected genetic mixtures among full-sib individuals which are highly inbred lines.

2.5. Statistical Measures of Genetic Diversity

Many statistical measures of genetic diversity are available in the literature depending on the data set. The commonly used measures of genetic diversity are; genetic distance, number of alleles per locus, allelic richness,rate of polymorphism, observed and expected heterozygosity, effective number of allele (N*e*), average number of alleles per locus and fixation indices.

(1) Genetic distance is a mean quantitative measure of genetic difference calculated between individuals, populations or species at DNA sequence level or allele frequency level. Genetic distance and/or similarity between two genotypes, populations, or

individuals may be calculated by various statistical measures depending on the data set. The commonly used measures of genetic distance are (i) Nei's and Li's coefficient (1979) and (ii) Jaccard's coefficient (1908). Genetic distance determined by the above measures can be estimated as follows:

$(i)\ GD = 1 - [2N_{11}/(2N_{11} + N_{10} + N_{01})]$

$(ii)\ GD = 1 - [N_{11}/(N_{11} + N_{10} + N_{01})]$

Where N_{11} is the number of bands/alleles present in both individuals; N_{10} is the number of bands/alleles present only in the individual i; N_{01} is the number of bands/alleles present only in the individual j; and N represents the total number of bands/alleles. There are two main ways of analyzing the resulting distance, principal coordinate analysis (PCoA) and dendrogram (or clustering, tree diagram). PCoA is used to produce a 2 or 3 dimensional scatter plot of the samples that reflect the genetic distances among samples. A dendrogram (tree diagram) group samples together in clusters that are more genetically similar to each other (Govindraj*et al*., 2015).

(2) The rate of polymorphism refers to number of polymorphic loci (n_i) divided by the total number of loci (n_{total}). It expressed as;

$P = n_{pi} / n_{total.}$

A gene is described as polymorphic if the frequency of one of its alleles is less than or equal to 0.95 or 0.99 (Govindraj *et al*., 2015).

(3) Average number of alleles per locus (Allelic diversity) – this is the sum of all detected alleles in all loci divided by the total number of loci. It can be expressed as;

Na = ni /nl where, ni is the total number of alleles over all loci; nl is the number of loci.

(3) Effective number of alleles (N*e*). This is the number of alleles that can be present in a population. It expressed as;

$Ne = \frac{1}{1-h} = \frac{1}{\sum Pi^2}$ Where,

P_i= frequency of the *i*th allele in a locus, h= $1 - \sum Pi^2$ Heterozygosity in a locus.

(4) Heterozygosity. There are two types of Heterozygosity: - observed (Ho) and expected Hetrozygosity (He). He is the probability that, at a single locus, any two alleles, chosen at random from the population are different to each other whereas Ho is the portion of genes

that are heterozygous in a given population. The value of He is an estimate of the overall extent of genetic variability in the population. Typically values for He and Ho range from 0 (no Heterozygosity) to 1 (all are heterozygous). It is necessary to compare the He and Ho to make inferences about the evolutionary history of the population (Lynch and Walsh, 1998). If Ho and He are similar (they do not differ significantly), mating in the populations is random. If Ho smaller than He, the population is inbreeding; if Ho is greater than He, the population has a mating system avoiding inbreeding (Govindaraj *et al*., 2015). He &Ho can be calculated as follows;

$He = 1 - \sum_i^n \mathrm{p}i^2$Where,

pi is the frequency of the i*th* allele and Ho can be calculated for each locus as the total number of heterozygotes divided by sample size.

(5) The effective population size (Ne), is another important measurement which can indicate the rate of genetic drift, the rate of genetic diversity loss, and increase of inbreeding within a population. It provides a measure of genetic variability of a set of individuals in a given situation. It is a basic parameter that largely determines allelic retention, preservation, and conservation over generations when studying genetic diversity of landraces (Vencovsky and Crossa, 2003). The static calculations of Ne depend on the genetic parameters used and reference generation.

(7) Fixation indices (F-statistics), widely applied in population genetics to measure the amount of allelic fixation by genetic drift (Govindraj*et al*., 2015). Fixation indexes can be calculated as follows:

Fit = 1 − (H*oa*/H*tot*),

Fis = 1 − (H*oa*/H*s*),

Fst = 1 − (H*s*/H*tot*), where,

H*oa*' is the average Ho within each population, H*s*' is the average He of subpopulations assuming random mating within each population, and H*tot*' is the He of the total population assuming random mating within subpopulations and no divergence of allele frequencies among subpopulations.

2.6. Genetic Diversity in Arabica Coffee

Genetic variability, which is due to the genetic differences among individuals within a population, is the core of plant breeding because proper management of diversity can produce permanent gains in the performance of plants and can resist against seasonal fluctuations and climate changes (Sharma, 1998). These genetic variations can be enumerated at three levels: species, populations and individual levels. Since Ethiopia is the only centers of origin and diversifications of *Coffea arabica,* there is a high genetic diversity, which is mainly attributed to its diverse ecological features such as suitable altitude, ample rainfall, optimum temperature, fertile soils etc. and the presence of different indigenous methods used in Coffee production system in the country (Habtamu *et al.,* 2018).

More genetically diverse strains of Arabica Coffee exist in Ethiopia than anywhere else worldwide, which has lead many botanists, breeders, and scientists to consent that Ethiopia is the center for origin, diversification, and dissemination of the Arabica coffee (Mekuria *et al.,* 2004). Several phenotypic and molecular studies revealed that the populations of *Coffea arabica* from the south western part of Ethiopia have high genetic variability, which is suitable for *in situ*conservation of the species. Sylvian (1958) and Meyer (1968) observed a high diversity of several phenotypic characters among Ethiopian Coffee populations collected from different geographical area of the country. Montagnon and Bouharmont (1996) also found higher phenotypic diversity among the populations of *Coffea arabica* collected from Ethiopia as compared to cultivated populations of Arabica coffee species that collected from different parts of the world. According to the study of genetic variation among forty nine *Coffea arabica* accessions from Limmu Ethiopia, confirmed the presence of trait diversity within coffee accessions (Olika *et al*., 2011). The study of genetic variation among 100 *Coffea arabica* accessions from Hararge, Ethiopia were also confirmed the presence of trait diversity within 14 characters suggesting that the presence of high variability among the accessions (Mesfin & Bayeta, 2008).

Higher level of genetic variability with molecular markers was observed among spontaneous and sub spontaneous accessions of this species collected from Ethiopia

(Anthony *et al.*, 2002). The existence of two subgroups of partial genetic differentiation within germplasm of *C. arabica* into accessions collected from West (Kaffa, Ilubabor and Wolegga) and East (Sidamo and Hararge) of Great Rift Valley was established by an analysis with molecular markers (RAPD) (Lashermes *et al.*, 1996) and also by a multivariate analysis of phenotypic characters (Montagnon and Bouharmont, 1996). In this perspective it would appear that the coffee cultivated in Yemen from where almost all cultivated *Coffea arabica* derive, had its origin in Ethiopia (Dessalegni, 2017).

3. MATERIALS AND METHODS

The experiment was conducted atthe National Agricultural Biotechnology research center (NABRC) which is located at Holeta, 29 km west of Addis Ababa, 09^{o}04 Nlatitude and 38^{o}30 E. Longitude.

3.1 Plant Materials

In this study, a total of 86 *Coffea arabica* accessions were used. The accessions were obtained from Ethiopian Biodiversity Institute (EBI). They were collected fromdifferent agro-ecologies varying in altitude, soil type, rainfall, temperature and representthe major coffee growing regions of Ethiopia and were preserved at different EBI field gene banks. The list of accessions used is given in Table 1.

Table 1: List of *Coffea arabica* accessions used in the study

S/N	Code	Ac. No.	Region	Zone/Region	Latitude	Longitude	Altitude
1	HG01	244460	Oromia	Hararge	N09º04.95	E41º21.68	1700
2	HG02	244335	Oromia	Hararge	N09º11.12	E41º30.01	1960
3	HG03	244438	Oromia	Hararge	N09º05.95	E41º21.24	1780
4	HG04	244355	Oromia	Hararge	N09º17.37	E41º31.47	1880
5	HG05	244430	Oromia	Hararge	N09º04.99	E41º22.53	1720
6	HG06	244434	Oromia	Hararge	N09º04.92	E41º22.52	1700
7	HG07	244386	Oromia	Hararge	N09º02.02	E41º19.45	1700
8	HG08	244341	Oromia	Hararge	N09º11.36	E41º29.90	1960
9	HR09	244216	Harari	Harar	N09º23.20	E42º14.73	1970
10	HR10	244230	Harari	Harar	N09º15.34	E42º08.99	1840
11	HR11	244227	Harari	Harar	N09º22.53	E42º14.99	2980
12	HR12	244224	Harari	Harar	N09º23.34	E42º14.63	2010
13	HR13	244232	Harari	Harar	N09º15.27	E42º08.93	1830
14	HR14	244229	Harari	Harar	N09º15.43	E42º08.99	1830

Table 1. continued

15	HR15	244233	Harari	Harar	N09º15.23	E42º08.94	1830
16	HR16	244239	Harari	Harar	N09º15.85	E42º08.60	1910
17	IB17	30744	Oromia	Ilubabor	N09º27.18	E81º45.43	1424
18	IB 18	30731	Oromia	Ilubabor	N09º26.97	E81º57.86	1461
19	IB 19	30745	Oromia	Ilubabor	N09º27.16	E81º45.53	1423
20	IB 20	30655	Oromia	Ilubabor	N08º11.13	E35º43.03	1706
21	IB 21	30638	Oromia	Ilubabor	N08º30.10	E35º42.39	1320
22	IB 22	30655	Oromia	Ilubabor	N08º11.13	E35º43.03	1706
23	IB 23	212085	Oromia	Ilubabor	N08º34.12	E35º47.18	1500
24	IB 24	212095	Oromia	Ilubabor	N08º18.16	E35º34.36	1530
25	IB25	22873	Oromia	Ilubabor	N07º11.13	E61º45.53	1950
26	YC26	24909	SNNP	Gedeo/Yirgachefe	N06º06.19	E38º12.04	2036
27	YC27	24909	SNNP	Gedeo/Yirgachefe	N06º06.19	E38º12.04	2036
28	YC28	24902	SNNP	Gedeo/Yirgachefe	N06º05.59	E38º12.11	2100
29	YC29	211947	SNNP	Gedeo/Yirgachefe	N06º08.76	E38º12.39	1880
30	YC30	24910	SNNP	Gedeo/Yirgachefe	N06º09.55	E38º12.54	2030
31	YC31	24907	SNNP	Gedeo/Yirgachefe	N06º79.03	E38º22.00	2222
32	YC32	24912	SNNP	Gedeo/Yirgachefe	N06º04.14	E38º09.21	1920
33	YC33	24887	SNNP	Gedeo/Yirgachefe	N06º02.11	E38º08.56	1718
34	YC34	24900	SNNP	Gedeo/Yirgachefe	N06º06.08	E38º12.09	2076
35	WG35	8714	Oromia	Wolegga	N09º13.25	E36º30.15	1900
36	WG36	8721	Oromia	Wolegga	N09º08.58	E36º24.24	1882
37	WG37	8730	Oromia	Wolegga	N09º12.10	E36º22.06	1579
38	WG38	8744	Oromia	Wolegga	N09º11.11	E36º27.00	1664
39	WG39	8750	Oromia	Wolegga	N09º11.37	E36º30.28	1706
40	WG40	212403	Oromia	Wolegga	NA	NA	1500
41	WG41	8753	Oromia	Wolegga	N09º09.17	E36º31.02	1888

Table 1. continued

42	WG42	8760	Oromia	Wolegga	N10º30.37	E35º40.02	1745
43	WG43	8754	Oromia	Wolegga	N09º08.35	E36º31.23	1868
44	KF44	211971	SNNP	Kefficho/Shekicho	NA	NA	1700
45	KF45	211973	SNNP	Kefficho/Shekicho	NA	NA	1750
46	KF46	211978	SNNP	Kefficho/Shekicho	NA	NA	1650
47	KF47	211975	SNNP	Kefficho/Shekicho	NA	NA	1720
48	KF48	211927	SNNP	Kefficho/Shekicho	NA	NA	1640
49	KF49	219392	SNNP	Kefficho/Shekicho	NA	NA	1650
50	KF50	212250	SNNP	Kefficho/Shekicho	NA	NA	1940
51	KF51	219394	SNNP	Kefficho/Shekicho	NA	NA	1840
52	KF52	211926	SNNP	Kefficho/Shekicho	NA	NA	1700
53	BM53	212033	SNNP	Bench maji	N06º59.66	E035º34.11	1520
54	BM54	212030	SNNP	Bench maji	N07º02.03	E035º32.77	1440
55	BM55	212032	SNNP	Bench maji	N07º02.91	E035º29.74	1320
56	BM56	212032	SNNP	Bench maji	N07º02.88	E035º2974	1650
57	BM57	212039	SNNP	Bench maji	N07º02.91	E035º29.75	1350
58	BM58	212036	SNNP	Bench maji	N07º04.13	E035º37.74	1400
59	BM59	212039	SNNP	Bench maji	N07º02.88	E035º29.74	1600
60	BM60	212048	SNNP	Bench maji	N06º59.69	E035º34.9	1650
61	OM61	8654	SNNP	Omo	N05º58.39	E36º34.27	1643
62	OM62	8662	SNNP	Omo	N05º47.30	E36º32.51	1500
63	OM63	8678	SNNP	Omo	N05º49.6	E037º49.11	1360
64	OM64	8693	SNNP	Omo	N06º26.44	E36º32.7	1500
65	OM65	8711	SNNP	Omo	N06º18.9	E36º37.26	1860
66	OM66	8812	SNNP	Omo	N07º11.53	E37º14.24	1465
67	OM67	8789	SNNP	Omo	N07º09.41	E37º03.55	1642
68	OM68	8800	SNNP	Omo	N07º08.41	E37º10.43	1426
69	OM69	8820	SNNP	Omo	N06º58.09	E37º19.43	1484

Table 1. continued

70	JM70	218931	Oromia	Jimma	N07º40.43	E036º48.76	1740
71	JM71	218940	Oromia	Jimma	N07º40.46	E036º48.79	1640
72	JM72	211967	Oromia	Jimma	N07º40.58	E036º48.75	1750
73	JM73	218930	Oromia	Jimma	N07º41.87	E036º48.13	1950
74	JM74	212330	Oromia	Jimma	N07º31.37	E036º53.44	1900
75	JM75	218921	Oromia	Jimma	N07º31.28	E036º53.63	1610
76	JM76	211932	Oromia	Jimma	N07º46.80	E036º48.84	1800
77	JM77	211966	Oromia	Jimma	N07º44.89	E036º46.87	1740
78	JM78	212335	Oromia	Jimma	N07º41.89	E036º49.94	1760
79	BL79	23376	Oromia	Bale	N06º24.00	E39º35.00	1520
80	BL80	23384	Oromia	Bale	N06º20.00	E39º37.00	1540
81	BL81	23368	Oromia	Bale	N06º24.00	E39º35.00	1540
82	BL82	23427	Oromia	Bale	N06º24.00	E39º35.00	1560
83	BL83	23440	Oromia	Bale	N06 º24.00	E39º35.00	1560
84	BL84	23447	Oromia	Bale	N07º08.00	E40º48.00	1510
85	BL85	23453	Oromia	Bale	N07º08.00	E40º48.00	1510
86	BL86	23484	Oromia	Bale	N07º08.00	E40º48.00	1670

SNNP= South Nation Nationalities and Peoples, NA= absent

3.2 DNA Extraction

Young leaf samples that are free from disease were collected from the growing tips of the coffee branches and preserved and dried using silica gel (Blulux laboratory (p) Ltd. -121005) and transported to National Agricultural Biotechnology Research Center for DNA extraction. Genomic DNA was extracted from dried leaf samples according to Khanuja*et al.* (1999) protocol with small modifications. About 0.5g of dried leaf tissue was placed in 2ml eppendrof tube and ground using a geno-grinder (Retsch MM200, Germany). One ml of freshly prepared extraction buffer was added to the ground material and mixed inversely. The

mixed solution was incubated at 60°C in a shaking water bath (100 rpm) for 2hours followed by addition of 1ml chloroform: isoamylalcohol (24:1).The solution was mixed by inverting the tubes for about 15 minute and then centrifuged at 8000 rpm for 10 min at 25-30°C. Carefully, the upper clear aqueous layer was transferred to another 2ml eppendrof tube and properly mixed with 0.75 ml of 5M NaCl. About 0.6 volume of isopropanol was added and placed at room temperature for1hour. After 1hour, the mixture was centrifuged at 10,000 rpm for 10minutes at 25-30°C and the supernatant was discarded and the pellet washed with 80% ethanol. The pellet was dried in a vacuum for 15 minutes and dissolved in 0.5 ml of high salt TE buffer (1M NaCl, 10mM Tris-Cl (pH 8.0) and 1mM EDTA). After the pellet wasdissolved in TE buffer, 2ml of RNase was added and incubated at 37°C for 30minutesin order to degrade RNA. Equal volume of chloroform: isoamyl alcohol (24:1) was added, centrifuged and the aqueous layer was transferred to fresh 1.5ml microfuge tube and 2 volumes of cold ethanol was added to precipitate the DNA. The mixture was centrifuged at 10,000 rpm for 10minutes at 25-30°C and the pellet was washed with 80%ethanol. Finally, the pellet was dried in a vacuum and dissolved in 100µlof nucleasefree water.

3.3 Determination of Concentration and Quality of Extracted of Genomic DNA

The genomic DNA concentration was estimated at 260 and 280 nm using the nano drop (Nano Drop®ND-800). Similarly, the quality of DNA was checked using 0.8% agarose gel by loading 5µl genomic DNA after mixing with 2µL 6X loading dye with Gel red. The band of DNA was visualized using gel documentation system (Bio Doc-It™ imaging system) under UV light. High molecular weight genomic DNA was stored at -20°C until use for genotyping. The quality and concentration of the genomic DNA of some of the accessions were indicated as samples in appendix 1.

3.4 Marker Selection

Initially 24 SSR markers previously described as being polymorphic in *C. arabica*were selected from published literature (Rovelli *et al*., 2000; Combes *et al*., 2000; Moncada and McCouch, 2004 and Cubry *et al*., 2008)**.**The annealing temperatures of the primers were optimized by performing gradient PCR. Finally, 20 polymorphic SSR markers were selected

by excluding four markers that failed to amplify properly. Details of the SSR markers used for this study are listed in table 2.

Table 2: List of selected SSR markers with their marker name, forward/reverse primer sequences and repeat motifs

S/N	Locus	5'.3' Forward/5'.3' Reverse primer	Repeat motifs	Annealing T^o	Reference
1	AJ308782	F:AAAGGAAAATTGTTGGCTCTGA R: TCCACATACATTTCCCAGCA	(GT)15	58.1	Rovelli *et al* (2000)
2	AJ308779	F: TCCCCGATCTTTTTCTTTCC/ R:GGGAGTGTTTTTGTGTTGCTT	(TG)17	57.8	Rovelli *et al (*2000)
3	AJ308753	F:CTTGTTTGAGTCTGTCGCTG/ R:TCCAGAAGTCTTGGGTT	(CA)15	56.1	Rovelli *et al (*2000)
4	AJ250251	F:ATTCTCTCCCCCTCTCTGC/ R: TGTGTGCGCGTTTTCTTG	(CA)8	58.1	Combes *et al* (2000)
5	CFGA92	F:TGAGGGCAAAGGAGTAAGAAAG/ R:TCAAACTTCAACAATCAAATACCC	(AG)10	60.1	Moncada and McCouch (2004)
6	AJ250253	F:GGAGACGCAGGTGGTAGAAG/ R:TTTCCCTCCCAATGTCTGTA	(GA)5(GT)8TT(GT)4TT(GT)7(GA)11(TC)2(CT)3GT	59.2	Combes *et al* (2000)
7	CFGA91	F:CTTCTCCAGCTTTAGGTTCACTTTG/ R:TTTTGAATACTGGCTCGTGAACTT	(AG)17	55	Moncada and McCouch (2004)
8	CFGA69	F:TGGTGGAGTGGCTTTGATTGATG/ R:GCAACTTATGAGCCTAATCC	AG)14(GT)11	55.8	Moncada and McCouch (2004)
9	CFGA502	F:AAGCCACCCAGAAAACAGCACATC/ R:ATTTGCTTCTCATGTTCCCTTTCA	(AG)27	62	Moncada and McCouch(2004)
10	AJ250255	F:CCCTCCCTGCCAGAAGAAGC/	(GT)5CT(GT)2/(GT)12	57.8	Combes *et al* (2000)

		R:AACCACCGTCCTTTTCCTCG			
11	CFGA465	F:ACATCCCCTTGCCATTTCTTC/	(AG)18	55.7	Moncada and
		R:ACCCTTTACTCATTATTTACTCTC			McCouch (2004)
12	AJ250257	F:GACCATTACATTTCACACAC/	(CTCACA)4/(CA)9	55.7	Combes *et al(*2000)
		R:GCATTTTGTGCACACTGTA			
13	AJ308774	F:GCCACAAGTTTCGTGCTTTT/	(CT)5 (CA)7	55	Cubry *et al* (2008)
		R:GGGTGTCGGTGTAGGTGTATG			
14	AJ250258	F:AACTCTCCATTCCCGCATTC/	(CA)3/(CA)3/(CA)18	56.9	Combes *et al*(2000)
		R:CTGGGTTTTCTGTGTTCTCG			
15	CFGA100	F:TTGACTCTTTTCTCTCCCAA/	(AG)15	56.1	Moncada and
		R:ATTTAGCAGGCTTGGCATTTTT			McCouch (2004)
16	AJ308809	F:AGCAAGTGGAGCAGAAGAAG/	(CA)15(CG)4CA	60.1	Cubry *et al* (2008)
		R:CGGTGAATAAGTCGCAGT			
17	AJ250260	F:TGATGGACAGGAGTTGATGG/	(CT)9(CA)8/(CT)4/(CA)5	58.1	Combes *et al*(2000)
		R:TGCCAATCTACCTACCCCTT			
18	AJ308790	F:TTTTCTGGGTTTTCTGTGTTCTC	(GT)21	57.8	Rovelli *et al* (2008)
		R:TAACTCTCCATTCCCGCATT			
19	AJ308755	F: CCCTCCCTCTTTCTCCTCTC	(CA)20	56.9	Cubry *et al* (2008)
		R:TCTGGGTTTTCTGTGTTCTCG			
20	AJ308837	F:CTCGCTTTCACGCTCTCTCT	(GT)16(GA)11	61.1	Rovelli *et al* (2000)
		R:CGGTATGTTCCTCGTTCCTC			

F: = Forward primer, R: = Reverse primer

3.5 PCR Amplification

PCR for SSR markers was conducted in 12.5µl reaction volume containing:-One Taq® 2X Master Mix with Standard Buffer, 10µM of forward primers, 10µM of reverse primers, Nuclease free water and 20ng/µl genomic DNA template. DNA free 10.5µl reaction volume was used as a negative control.

The PCR Conducted in Bio-Rad T100™ thermal cycler. Thermal cycling conditions for PCR were conducted as follow:

Steps	Temperature	Time	Cycles
Initial DNA denaturation	94°C	3 minute	1
DNA denaturation	94°C	1 minute	30
Primer annealing	as per as Ta	1 minute	30
Primer extension	72°C	2 minute	30
Final extension	72°C	10 minutes	1
Hold	4°C	-	-

3.6 Gel Electrophoresis

The amplified PCR products were separatedon 3.5% (w/w)agarose gel by loading 10µl of each of the PCR product mixed with 2µl of loading dye and gel red. Based on the resolution with DNA Ladder on the adjacent lane, the amplified products were loaded on high resolution agarose gel in 1X TAE (40mMTris, 40mM Acetate, 1mM EDTA, pH 8) buffer and gels were run at 85Voltage, 100mA for 3.5 hours. Finally, the amplified bands were documented under UV light in gel documentation.

3.7 Data Scoring and Analysis

The amplified SSR fragment sizes on gel were estimated using PyElph 1.4 software package. Genetic diversity analysis was carried out on the basis of 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 allele (Na), gene diversity; Polymorphic information contents (PIC) were computed using

Power marker v3.25 software (Liu and Muse, 2005). Effective number of alleles, Shannon's Information index (I) and Gene flow (Nm) were determined using POPGENE version 1.31 (Yeh and Yang, 1999). Allelic frequency, observed Heterozygosity (Ho), expected Heterozygosity (He), fixation index (F) and estimate of the deviation from Hardy-Weinberg equilibrium (HWE) over the entire populations were computed with GenAlEx version 6.501 software (White and peakall, 2015). The same software package was used to compute population differentiation test: Wrights fixation indices and pairwise population differentiation (Fst), analysis of molecular variance (AMOVA) and estimate of the variance components. Moreover, rarified allelic richness (Ar) and private rarified allelic richness (Arp) were computed using HP-Rare 1.1 software (Kalinowski, 2005). To examine the genetic relationship between the different populations, a genetic dissimilarity matrix was calculated using Jaccard's formula and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram were generated using DARwin var. 6.0.14 (Perrier and Jacquemoud-Collet, 2006). Principal coordinate analysis (PCoA) was performed using the GenALEX to produce Eigen values and Eigen vectors which reveal both total variance and the loci that are important to the variance. Biplots were generated to reveal associations among accessions.

The frequency of each allele was calculated using the formula,

$$\frac{\boldsymbol{ni}}{\boldsymbol{nt}}$$

Where $\mathbf{n_i}$ is number ofgenotypesexhibited$\mathbf{i^{th}}$ alleleand $\mathbf{n_t}$is the sum ofnumber of detected allelesamong genotypes

The PIC of SSR markers wascalculated according to Botstein et al. (1980)using the following formula,

$$1 - \sum_{i=1}^{n} pi^2 \; - \sum_{i=1}^{n-1} \; \sum_{j=i-1}^{n} 2pi^2 \; pj^2$$

Where, p_i^2 is the frequency of i^{th} allele

Expected hetrozygosity (He) calculated as;

$\mathrm{He} = 1 - \sum_i^n \mathrm{p}i^2$Where,

pi is the frequency of the i[th] allele and Ho can be calculated for each locus as the total number of heterozygotes divided by sample size.

Effective number of alleles (N*e*). This is the number of alleles that can be present in a population. It expressed as;

$$\mathrm{Ne} = \frac{1}{1-h} = \frac{1}{\sum P\ i^2}$$Where,

P_i= frequency of the *i*th allele in a locus, $h= 1 - \sum \mathrm{Pi}^2$,Heterozygosity in a locus.

4. RESULTS AND DISCUSSION

4.1. Polymorphism and Allelic Diversity among SSR Markers

A total of 112 SSR alleles were produced across 20 loci, with estimated product size range of 100 to 385 base pairs (Table 3), out of which 5 (4.46%) were rare (frequency < 0.01) and 11 (9.82%) were scarce (frequency between 0.01 and 0.05). The frequency of 16 (14.28%) was between 0.05 and 0.1 whereas 80 alleles (71.43%) had a frequency of 0.1 or higher (Appendix 1). The mean number of alleles was 5.6. The detected number of alleles range from 3 (AJ250253 and CFGA69) to 11 (CFGA465), and most of the markers produced 5 and 6 alleles.This result is higher than aprevious study by Anthony *et al.* (2002) who reported an average number of 4.7 alleles per SSRs using six SSRs in Arabica coffee collections containing four Typica, five Bourbon and 10 sub spontaneous derived accessions.Moncada & McCouch (2004) reported an average of 2.5 alleles per SSR in 11 wild Arabica coffee genotypes and 12 cultivated Arabica coffee, respectively, with the number of alleles ranging from one to eight using 34 SSRs. Maluf *et al. (*2005) also reported an average number of 2.87 alleles in 28 cultivated Arabica lines using 23 SSR Markers. Bigirimana *et al.* (2013)reported an average of 3.2 alleles in 14 genotypes using six SSR markers. In the present study the higher average number of allele (5.6) may be due to the difference in genotypes, sample sizes, enrichment of Ethiopian *C. arabica* genotypes and the number and types of markers included in the study.

In other study, Alemayehu *et al.* (2010) reported a mean of 6.5 number of alleles per locus that ranged from 2 to 14 in 133 Coffee genotypes (78 accessions from Ethiopia and 55 cultivars) and 5.9 mean number of alleles ranged from 2 to 12 for Ethiopian Arabica coffee with 32 SSR markers that higher than the present study. This is might be due to the few markers (20) used in the present study to evaluate 86 genotypes and Alemayehu *et al.* (2010) used more diverse genotypes from differentsources compared to the genotypes included inthe present study. It is therefore expected that with increase in number of markers a higher number of alleles could be obtained. In other study, da Silva *et al.* (2019) reported 6.9 mean number of allele per locus in 36 *Coffea* genotypes, including 33 *C. arabica* genotypes: Twenty-five from the Ethiopian collection, four cultivars and four lines developed by the

breeding programs using 30 SSR markers that are higher than the present study. Hussein *et al.* (2017) also reported higher mean number of allele per locus (10.7) among seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (*Coffea arabica* L.*)* germplasm using 15 SSR markers.Aerts *et al.* (2013) in a study based on populations from two *Coffea arabica* production systems of Ethiopia (forest coffee and semi-forest coffee) and detected 159 alleles ranged from 2 to 19 alleles across 703 *Coffea arabica* wild accessions with 24 SSR markers.Generally, the author suggested the number of alleles depends on markers, platform used for resolution of amplified products and the sources of genotypes used.

The use of molecular markers for efficient selection of genotypes with desirable traits and enhancing the efficiencyof breeding by allowing effective simultaneous selection of various desirable traits is a well-establishedapproach (Edwards, 1992). Hence, the large number of alleles detected in the present study suggests the suitability of microsatellites in general and in this study particular for genetic diversity study. From the detected alleles, smaller number of allele was rare and scarce. Some of the SSR markers showed tremendous discriminative performance during the estimation of the genetic diversity among *Coffea arabica* genotypes and resulted in the identification of private alleles. Moreover, the private alleles observed at several loci (Example: AJ250260, Aj250251, CFGA91 and Aj308779) in different accessions (HR12, KF52, BM57, OM65 and BL82) could offer a good opportunity to evaluate *Coffea arabica* genetic materials for the association of particular alleles with traits of interest and for conservation. Such alleles are useful in comparing diversity between species or populations and also for measuring genetic distinctiveness of individuals in a population (Kalinowski, 2004).

Table 3: In-formativenessand levels of different diversity indices of the SSR loci across *Coffea arabica* populations

Locus	EPS	MAF	GD	PIC	Na	Nm	I	Ar	Arp	He	Ho	Ne	F	P_{HWE}[a]
Aj308782	130-150	0.25	0.79	0.67	5	0.82	1.08	3.56	0.00	0.61	0.00	2.65	1.00	0.001**
Aj308779	130-180	0.27	0.79	0.66	6	2.78	1.40	4.61	0.01	0.72	0.99	3.86	-0.37	0.008**
Aj308753	340-310	0.17	0.87	0.75	8	1.25	1.45	4.91	0.00	0.72	0.00	4.10	1.00	0.001**
Aj250251	300-370	0.57	0.61	0.48	5	1.38	0.94	3.26	0.00	0.52	0.00	2.41	1.00	0.001***
CFGA92	345-385	0.28	0.78	0.65	6	2.05	1.30	4.22	0.78	0.69	0.01	3.35	0.98	0.001***
AJ250253	320-350	0.42	0.63	0.48	3	1.51	0.87	2.69	0.00	0.55	0.00	2.24	1.00	0.001**
CFGA91	270-300	0.40	0.69	0.55	7	2.24	1.10	3.49	0.19	0.62	0.06	2.90	0.91	0.001**
CFGA69	190-220	0.50	0.62	0.45	3	2.10	0.89	2.79	0.00	0.55	0.00	2.27	1.00	0.005**
CFGA502	285-320	0.52	0.64	0.50	4	2.10	0.98	3.18	0.00	0.57	0.00	2.47	1.00	0.001**
AJ250255	200-230	0.42	0.71	0.57	4	1.39	1.09	3.56	0.00	0.60	0.00	2.76	1.00	0.001***
CFGA465	180-280	0.33	0.83	0.73	11	3.36	1.71	6.54	0.03	0.77	0.88	4.62	-0.14	0.170^{ns}
AJ250257	130-170	0.38	0.76	0.62	5	6.98	1.44	4.81	0.00	0.73	0.98	3.82	-0.34	0.008**
AJ308774	160-190	0.42	0.70	0.55	5	8.33	1.25	4.03	0.00	0.68	0.97	3.17	-0.42	0.088^{ns}
AJ250258	100-120	0.40	0.65	0.40	4	1.18	0.88	2.78	0.00	0.54	0.00	2.32	1.00	0.001**

Table 3: continued

CFGA100	300-325	0.27	0.81	0.69	6	1.85	1.34	4.24	0.00	0.71	0.00	3.67	1.00	0.003**
AJ308809	140-170	0.27	0.83	0.71	7	1.98	1.41	4.46	0.00	0.73	0.00	3.86	1.00	0.001***
AJ250260	180-220	0.33	0.77	0.63	6	1.03	1.13	3.74	0.88	0.62	0.01	2.82	0.98	0.013*
AJ308790	140-160	0.41	0.70	0.56	5	1.61	1.04	3.18	0.00	0.61	0.01	2.69	0.98	0.001**
AJ308755	110-150	0.21	0.84	0.72	7	3.24	1.64	5.77	0.00	0.78	0.92	4.55	-0.18	0.062^{ns}
AJ308837	150-175	0.35	0.74	0.61	5	1.76	1.16	3.65	0.008	0.65	0.03	3.06	0.96	0.001**
Mean	-	0.36	0.72	0.60	5.6	2.45	1.21	3.97	0.095	0.65	0.24	3.18	0.67	-

EPS= Estimated product size in base pair, MAF=major allele frequency, Na= number of allele, PIC= Polymorphism information content, Ne = Number of effective alleles, Ho = observed heterozygosity, He = expected heterozygosity, F = fixation index, I = Shannon's diversity index, ns = not significant, * P<0.05, ** P<0.01, *** P<0.001= significant, $P_{HWE}{}^{a}$ = P-value for deviation from Hardy Weinberg equilibrium, Nm = Gene flow estimated from Fst= 0.25(1- Fst)/Fst, GD= gene diversity, Arp= private allelic richness, Ar= allelic richness

The polymorphic information content value, which is a measure of the in-formativeness of the marker, was calculated for each of the 20 simple sequence repeats to estimate the extent of genetic diversity among 86 *Coffea arabica* genotypes. It variedfrom 0.45 to 0.75 with a mean value of 0.6.About 16 (~80%) of the markers werehighly informative (i.e. PIC > 0.50). The rest of four markers were moderately informative withPIC values less than 0.5. Al-Murish *et al.* (2013) reported 0.43 mean PIC value that is smaller than the present result with 16 SSR markers in seventy *Coffea arabica* genotypes. Similarly, Moncada and McCouch (2004) also reported smaller average PIC value of 0.35 among 11 *Coffea arabica* genotypes. In the present study, the higher mean PIC value (0.6) could be attributed to differences in genotypes as well as SSR markers used.

In other studies, da Silva *et al.* (2019) reported 0.72 mean PIC values per locus in 36 *Coffea arabica* genotypes using 30 SSR markers that is higher than the present study. Similarly,Hussein *et al.* (2017) reported 0.84mean PIC value among seventeen genotypes involving 16 commercial cultivars and oneaccession of Yemeni coffee (*Coffea arabica* L.) germplasm using 15 SSR markers that is higher than the present study.SuchPIC value that observed in present study and others provides an estimate of the discriminatory powerof a locus (Marulanda *et al*., 2014), and the allelic diversity suggest greatpotential of the markers for use in future genetic studies.

The values of major allele frequency of the markers ranged from 0.16 to 0.57 with a mean of 0.36. The lowest and highest number of major allele frequency was observed in AJ30875 and AJ250251 markers, respectively. The gene diversity of the marker ranges from 0.61 (Aj250251) to 0.87 (Aj308753) with a mean of 0.72 indicating the markers used are polymorphic in this study. The mean number of alleles (5.6) and the mean gene diversity (0.72) per locus revealed by the current study indicate diversity among the populations of *Coffea arabica.* The effective number of alleles ranged from 2.24 to 4.62 with a mean of 3.18. The disparity in the number of alleles and effective number of alleles is accounted for the differences in heterozygosity. Variation in heterozygosity is the corollary differences in allele frequencies in a locus. On other hand, the difference between the number of alleles in each

locus andnumber of effective alleles showed the existences of exclusive/specificalleles in the genotypes.

The average observed (0.24) and expected heterozygosity (0.65) under Hardy Weinberg equilibrium values across the loci are presented in the Table 3. About 17 (85%) of the markers showed significant ($p < 0.05$) deviation from HWE. Only three markers (CFGA465 p = 0.17), (AJ308774 p = 0.088) and (AJ308755 p = 0.062) showed non-significant deviation from Hardy- Weinberg equilibrium (HWE). The loci studied displayed differences between Ho and He in which some of them showed excess heterozygositythat led to a significant departure from Hardy-Weinberg equilibrium (HWE) across populations. Such excess heterozygosity is expected in sexually reproducing organisms that can maintain their heterozygosity through sexual recombination, or other factors such as natural selection pressure and excess gene flow (Fekadu *et al*., 2018). In other ways, a high number of alleles and large gene diversity is ordinarily expected from large sample size and large proportion of di-nucleotide repeats employed that consequently give high value of expected heterozygosity (Liu *et al*., 2005).

On the other hand, the lowest fixation index (-0.42) were observed for AJ308774 marker (Table 3). The fixation index of the markers (F) ranged from 0 to 1 with a mean of 0.67 (Table 3). Five markers had negative Fixation index (F) values indicating an excess of heterozygotes. For loci Aj308779 (72%), CFGA465 (77.4%), AJ250257 (73.2%), AJ308774 (68%) and AJ308755 (77.8%) genotypes expected to be heterozygous under the specific locus under random mating conditions. However, 98.8%, 88.3%, 97.8%, 96.5% and 91.9% of the genotypes were heterozygotes for Aj308779, CFGA465, AJ250257, AJ308774 and AJ308755, respectively. The F value revealed that, five of the twenty markers showed excess of heterozygotes (negative F value). It might be due to mutation at specific loci.

4.2. Genetic Diversity Within and Among Populations

4.2.1. Analysis of genetic diversity parameters

Summary of the different genetic diversity indices over the twenty markers for the ten populations are presented in Table 4.

Table 4: Summary of genetic diversity indices of Arabica coffeepopulations classified by area of collection

Populations	Genetic parameter								
	N	Na	Ne	Ar	Arp	I	He	Ho	%p
Hararge	8.00	3.550	2.92	3.52	0.00	1.11	0.63	0.24	100
Harar	8.00	3.750	2.90	3.71	0.05	1.12	0.62	0.26	100
Ilubabor	9.00	4.350	3.45	4.21	0.00	1.28	0.68	0.23	100
Yirgachefe	9.00	4.300	3.09	4.15	0.00	1.21	0.63	0.24	100
Wolegga	9.00	3.950	3.13	3.85	0.01	1.19	0.65	0.22	100
Kefficho	9.00	4.350	3.18	4.16	0.05	1.24	0.66	0.24	100
Benchi Maji	8.00	4.150	3.27	4.09	0.04	1.25	0.67	0.24	100
Omo	9.00	4.400	3.60	4.26	0.04	1.31	0.69	0.24	100
Jimma	9.00	4.250	3.34	4.14	0.01	1.23	0.65	0.238	100
Bale	8.00	3.700	2.91	3.65	0.04	1.11	0.62	0.25	100
Mean	-	4.08	3.18	3.97	0.024	1.21	0.65	0.24	100

N = number of individual within each population, Na = Number of different alleles, Ne = number of effective alleles, Ar = allelic richness, Arp = private allelic richness, I = Shannon diversity index, He = expected heterozygosity, Ho = observed heterozygosity, %p = percentage of polymorphic loci

The comparative analysis showed that there is no much difference among the ten populations of *Coffea arabica* with regard to genetic diversity indices including Number of different alleles, effective number of alleles, private allelic richness, Shannon's diversity index, observed heterozygosity, expected heterozygosity and percentage of polymorphic loci. Comparatively, the populations from Omo and Ilubabor scored higher measures of diversityindices parameters. The population from Omo scored the highest, number of different alleles (4.4), number of effective allele (3.60), allelic richness (4.26), Shannon diversity index (1.31), expected heterozygosity (0.69) and the population from Ilubaborscored the higher number of different alleles (4.35), number of effective allele (3.45), allelic richness (4.21), Shannon diversity index (1.28) and expected heterozygosity (0.68). Next to population from Omo and Ilubabor, the population from Bench Maji also scored the higher number of genetic diversity indices parameters. On other hand, the populations from Bale, Hararge and Harar scored the least number of genetic diversity indices parameters in order of their magnitude.

Population from Bale scored the least number of different alleles (3.70), effective number of allele (2.91), allelic richness (3.65), Shannon diversity index (1.11), expected heterozygosity (0.62) and followed by population from Harar and Hararge.

The Populations from Omo, Ilubabor and Benchi Maji are genetically more diverse than the other populations as estimated by parameters such as number of different alleles, number of effective allele, allelic richness, Shannon diversity index and heterozygosity.Hence, the areas representing these populations could be considered as genetic diversity hot spots and potential*in-situ* conservation sites for *C. arabica*germplasm. Similarly, depending on the diversity parameters Shimekit*et al.* (2014) reported higher genetic diversity in population that were collected from Debub Omo, among eighty seven*Coffea arabica*germplasms collected from southern Ethiopia using five ISSR markers. In other hands, among the ten populations studied,the populations from Bale, Harar and Hararge have the least genetic diversity indices which might suggest current rapid genetic erosion in the area.

In addition to this, in terms of allelic richness the populations from Omo, Ilubabor, Benchi Maji and Kefficho are the top four in that order, and hence they are more interesting in terms of genetic and evolutionary studies because allelic richness is more informative in this regard as it is sensitive to the presence of private alleles (Leberg, 1991) which is prominent and population bottlenecks when compared to other parameters such as heterozygosity. Moreover, these populations except Ilubabor, Hararge and Yirgachefe, bears relatively small proportion of private alleles which may indicate certain level of independent evolution of their gene pools that allowed maintenance of private alleles at a population level (Slatkin, 1985; Fekadu *et al*., 2018).

Another fundamental diversity indices parameter is heterozygosity which measures the level of genetic variability within populations, because genetic diversity can be measured as the amount of actual or potential heterozygosity. There are two types of heterozygosity, observed heterozygosity (Ho) and expected heterozygosity (He). The Ho is the portion of genes that are heterozygous in a population and He is estimated fraction of all individuals that would be heterozygous for any randomly chosen locus. The mean observed heterozygosity (0.24)

observed in present study, is lower than the half mean of expected heterozygosity (0.65) observed (Ho< He) indicated that the population is inbreeding (Govindraj *et al*., 2014). Comparatively, the highest number of observed heterozygosity (0.26) was observed in population from Harar and the least number of observed heterozygosity (0.22) was observed in population from Wolegga with a mean of 0.24 across the populations. Three populations: Wolegga, Ilubabor and Jimma in the order of their magnitude, scored slightly less than the mean observed heterozygosity (Ho), four populations: Hararge, Kefficho, Bench Maji and Omo scored similar Ho value (0.24) and population fromYirgachefe had scored a mean Ho value whereas the population from Bale scored higher observed heterozygosity next to population from Harar.

The mean value of observed heterozygosity (Ho) observed in the present study is higher than that of Hussein *et al.* (2017) who reported 0.212 mean observed heterozygosity value among seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (*Coffea arabica* L.) using 15 polymorphic SSR markers.Combes *et al.* (2000) also reported 0.043 average Ho value in*Coffea arabica* using eleven SSR primers that is lower than the present study. In other studies,Al-murish *et al.* (2013) reported 0.32 mean observed heterozygosity (Ho) that higher than the present study.Overall the lower observed heterozygosity level in this study and other studies might be due to the autogamous nature of *C. arabica*which contribute to lowheterozygosity levels whereas the higher average observed heterozygosity degree observed in this study and others seems to indicate the populations may have received enough external genes, which is likely the reason of the excess heterozygosity (Bodia *et al.*, 2012).

The polymorphism detected in the present studywas similar with the result reported by Hussein *et al.* (2017) who observed 100% polymorphism of markers among seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (*Coffea arabica* L.) using 15 polymorphic SSR markers.Thishigher degree of polymorphism might be due to the difference in discrimination power of the genetic marker and the source of genotypes used in the present study. Hence, SSR markers are more powerful and very

informative for geneticdiversity analysis than any other markers (Vieira *et al*., 2010; Pestiana *et al*., 2015).

In contrast to the present study,Tiago *et al.* (2017) reported 74.07% of polymorphism in thirty- four *Coffea arabica* cultivars in Brazil using 31 microsatellite markers which is smaller than the present study. This is due to the fact that Tiago *et al.* (2017) used*Coffea arabica* cultivars that have high genetic similarity. In several studies, a small number of polymorphic loci were observed for *C. arabica* (Combes *et al*., 2000; Anthony *et al*., 2002; Diniz *et al*., 2005; Vieira *et al*., 2010; Pestana *et al*., 2015). These low polymorphic loci might be due to the low genetic variability of species and narrow genetic base of origin.

Population may differ with respect to all aspects of diversity and show variation in number of alleles, allele distribution and frequency (Rao and Hodgkin, 2002). Variation in population may be attributed to the breeding system of the species and ecological factors such as latitude, altitude, temperature, moisture availability and other factors. Inter specific diversity can be valuable source as intra specific diversity for crop improvement (Benson *et al*., 2013). Higher genetic diversity is expected in larger and older populations when compared with small and newly established ones because of higher levels of accumulated and maintained genetic variation which is important in increasing fitness and therefore reduces the risk of local land race extinction (Futuyma, 2008).The mean diversity indices parameters (observed heterozygosity (0.24), expected heterozygosity (0.65), Shannon's diversity index (1.21) and allelic richness (3.97) and number of effective allele (3.79) obtained in the present study showed a high level of genetic variationwithin populations of Ethiopian coffee. This might be due to a relativelydifferent genetic basis of the populations included in this study that resultedfrom different germplasm resources accessible to farmers,or due to addition in population size, both naturalas well as human factors such as, sexual reproduction of the *C. arabica* and free movement of germplasm between different geographic areas.

4.2.2. Population genetic differentiation and gene flow

Analysis of molecular variance (AMOVA) partitioned the total molecular variance within and among the populations based on their area of collections. There were highly significant

(P<0.001) molecular variances among populations, among individuals within populations and within individuals. The highest proportion, 63% of the variation wasattributed to genetic variability among individuals within populations, while 33% wasdue to variation within individuals in the population. In contrast, a smaller portion of the total variation (4%) was among populations suggestingthat all populations are genetically similar (Table 5).

In *Coffea arabica* populations, the variation among population was expected than the variation within population as the plant is predominantly autogamous. However, the pollination biology study was carried out in non-native habitats of Arabica coffee and hence might not be representative for Arabica coffee native habitat where pollinators are actively engaged in facilitating active pollen exchange among individual coffee plant in Ethiopia (Shimekit *et al.,* 2014). Krishnan (2014) reported the higher proportion of variation within populations (96%) whereas the variation among populations was very low (4%) by using Ethiopian and Panamanian *C. arabica* populations that agree with the present study. In several studies, Esayas *et al.* (2003), Kassahun (2006), Solomon (2007) and Shimekit *et al.* (2014) observed high variation within population ascompared to among population variation, which is in agreement with the results of this study.

The higher variation within populations' genetic diversity might be accounted to two contrary reasons. *Coffea arabica* is affected by multiple evolutionary forces which operate within historical and biological context of the plant species. This includes the mating types, gene flow, mode of reproduction and natural selection. For this reason, it could be speculated from this result that *Coffea arabica* might have mixed mating system (partial out-crossing by pollen and seed, and partial selfing) for which some extent of gene flow is in high variation within population genetic diversity (Loveless and Hamrik, 1984; Esayas *et al*., 2003; Kassahun, 2006).In addition, the high genetic diversity observed within populations of the *Coffea arabica* might be due to preferential adaptive gene complexes adapted to environmental changes being evolved during long evolutionary period in a given region. In this case, *Coffea arabica* population uses selfing as mechanisms to prevent influx of the gene from another portion of the populations that might reduce diversity through disrupting adaptive genes (Loveless and Hamrik, 1984; Shimekit *et al.,* 2013).

Table 5: Analysis of molecular variance (AMOVA) across the full data set of 86 *Coffea arabica* accessions according to area of collection

Source	DF	SS	MS	Estimated Variance	Percentage variation	p-value	F-Statistics
Among Populations	9	149.878	16.653	0.276	4%	0.001	Fst=0.037
(Among individuals							
Within populations)	76	905.237	11.911	4.717	63%	0.001	Fis =0.656
Within Individuals	86	213.048	2.477	2.477	33%	0.001	Fit =0.668
Total	171	1268.16		7.470	100%	-	

DF = degree of freedom, SS=sum of squares and MS=mean squares

The overall observed gene flow (Nm) or gene migration value observed in this study on Ethiopian Arabica coffee was2.45,which showed the approximate number of individual's migration from one population to the other. Based on Slatkin (1985) and Waples (1987), Nm values grouped into three categories: Nm > 1.00 high, 0.25-0.99 intermediate and 0.000 - 0.249 low. Therefore, the high Nm value observed in this study indicates high gene flow between populations which will agree with the AMOVA result showing low variation between populations. Gene flow between population could occurred through either seeds or seedlings exchange. The gene flow could be enhanced via birds, insects, wild and domestic animals by facilitating exchange of pollens and seeds. Moreover, coffee farmers could also contribute to gene flow by exchanging seeds and seedlings of enhanced landraces among nearby districts with the objective of improving productivity of coffee. This leads to maximize the diversity of local genotypes and increase the distribution of alleles among different populations regardless of their geographical distance.

To explain the properties of subdivided populations; the magnitude between and within population differentiation were quantified using F-statistics (Fst, Fis and Fit) also known as Wright fixation indices (Wright, 1951). They are used to measure the amount of allelic fixation by genetic drift and mutation and they are related to heterozygosity and genetic drift. Since inbreeding increases the frequency of homozygotes, as a consequence, it decreases the

frequencies of heterozygotes and genetic diversity (Govindraj *et al*., 2014). Fixation index (Fst) is a measure of population differentiation due to genetic structure. According to Wright (1951), the threshold to determine the level of Fst value ranges from 0 to 0.05 considered as low, 0.05 to 0.15 moderate, 0.15 to 0.25 large and those greater than 0.25 mean very large genetic differentiations among populations.

In the current study the overall F- statistics used as a measure of population differentiation was low (Fst= 0.037 which is < 0.05),which indicates that *C. arabica* has very low genetic differentiationamong populations, which accounted onlyfor 3.7% of the total genetic variation (Table 5). This low genetic differentiation among population may be due to high gene flow that resulted from the movementof seeds from one location to another location through long miles movement of the germplasm exchange through marketing and human migration pattern. The presence of private alleles in some genotypes contributed tothe significant differentiation obtained.The spread/exchange of single genes, genotypes and even thewhole population in different regions resulted by the processes of exchange of gametes,individuals and population on geographic scale which is considered as gene flow inconjugation with other evolutionary forces is sufficient to prevent differentiation betweenpopulations (McDermott and McDonald, 1993).

The pair-wise genetic differentiation among populations within the source of origin ranged from 0.041 to 0.108 (Table 6). The lowest population differentiation was observed between Benchi Maji and Wolegga, Omo and Wolegga and Jimma and Wolegga whereas the highest population differentiation observed between Bale and Hararge, Kefficho and Hararge, Jimma and Bale populations in order of their magnitude. The low value of Fst implies that there is high frequency of identical alleles among *Coffea arabica* accessions. This might be due to the exchange of genetic materials like seeds and seedlings among the two populations. The lowest Fst value observed between a pair of populations can be explained by high level of gene flow, which leads to genetic similarity of populations. The larger pair wise Fst value observed between populations might be due to low genetic material exchange and geographical area.
Hence, the present study showed that *C. arabica* has very little populationgenetic differentiation.The low population genetic differentiation is supported by high gene flow

owing to step-wise pollen movement across *Coffea arabica* populations, germplasm exchange in the form of seedlings and seeds through sharing common markets among several of the adjacent geographical areas where different populations were collected. This could be explained by the extensive exchange of seedlings as planting materials among farmers (gene flow), common origin of the populations, the interest of the farmers to agronomically superior individuals in which only a limited number of individuals contribute seeds to the next generation, which gradually leads to recent or old population bottlenecks and hence, facilitate genetic drift. This study also showed the minimal effects of regions or geographic origins on genetic variation among *Coffea arabica* populations.

Table 6: Pair wise population matrix Fst values for ten populations of *C. arabica*population

	HG	HR	IB	YC	WG	KF	BM	OM	JM	BL
HG	0.000									
HR	0.051	0.000								
IB	0.085	0.055	0.000							
YC	0.082	0.066	0.045	0.000						
WG	0.087	0.088	0.060	0.063	0.000					
KF	0.105	0.100	0.065	0.056	0.047	0.000				
BM	0.075	0.085	0.059	0.053	0.041	0.050	0.000			
OM	0.088	0.079	0.052	0.062	0.041	0.047	0.045	0.000		
JM	0.097	0.085	0.057	0.051	0.041	0.049	0.051	0.055	0.000	
BL	0.108	0.098	0.097	0.104	0.090	0.089	0.099	0.078	0.105	0.000

HG= Hararge, HR= Harar, IB= Ilubabor, YC= Yirgachefe, WG= Wolegga, KF= Kefficho/Shekicho, BM= Bench Maji, OM= Omo, JM= Jimma, BL= Bale, all pair wise Fst values are significant at P = 0.05

4.3. Genetic Distance amongPopulations

Genetic distance is a measure of the genetic divergence between species or between populations within a species, whether the distance measures time from common ancestor or degree of differentiation. Populations with many similar alleles have small genetic distances. This indicates that they are closelyrelated and have a recent common ancestor. The Genetic distance between populations reflects more or less the level of similarity and the degree of

relatedness between populations. The magnitude of genetic distance between *Coffea arabica* populations was estimated based on Nei's (1978). The pair wise Nei's unbiased genetic distance value ranged from 0.048 to 0.426 whereas the mean Nei's unbiased genetic distance of the populations ranged from 0.16 to 0.35 with an overall mean of 0.21 (Table 6). About 70% of the populations (Ilubabor, Yirgachefe, Wolegga, Kefficho, Benchi Maji, Omo and Jimma)scored below overall mean of Nei's unbiased geneticdistance (0.21) whereas 30% of the populations (Bale, Hararge and Harar) scored above overall mean of Nei's unbiased geneticdistance.

The Population from Hararge showed the highest (0.426) pairwise Nei's unbiased geneticdistance with Kefficho population. The population from Bale also, separated from Jimma and Hararge populations at higher pair wise Nei's unbiased genetic distance of 0.406 and 0.4 respectively. The population from Bale separated from all populations at highest mean Nei's unbiased genetic distance (0.35). This population was the most genetically distinct population with the highest mean Nei's unbiased genetic distance. The population from Hararge also scored the higher mean Nei's unbiased genetic distance (0.29) next to Bale population. The limitation of genetic material exchange by human or natural factors may be considered as the main factor for the larger genetic distance obtained. This can be partly explained by the fact that these *Coffea arabica* populations were collected from a relatively pocket location and are separated from the other populations with a relatively longer geographic distance that probably restricted recent germplasm exchange. Hence, these populations may serve as potential sources of new genetic variation of important traits that can be used in breeding programs.

On other hand, the smallest pair wise Nei's unbiased genetic distance was observed between Jimma and Wolegga and between Benchi Maji and Wolegga populations in order of their magnitude. Thus, the population from Wolegga showed the smallest pair wise Nei's unbiased genetic distance with Jimma and Benchi Maji populations with Nei's unbiasedgenetic distance of 0.048 and 0.051, respectively. Similarly, the population from Wolegga and Benchi Maji scored the lowest mean Nei's unbiased genetic distance (0.16) (Table 6).

Hue (2005) reported lower genetic distance (0.205) that is in agreement with the present study (0.21) in eighty four coffee (*C. arabica*) varieties of local coffee grown in Northern New South Wales, Australia with AFLP and SSR data profiles. Mulatu *et al.* (2012) reported the pair wise genetic distance ranged from 0.001 to 0.392 from twenty-six populations representing eight Arabica coffee varieties collected from Nicaragua by using 12 SSR markers that is smaller than the present study (0.048 to 0.426). This might be due the nature and number of molecular characterization of the used SSR loci, and the selected group of genotypes included in the study.

The result of present study indicates the availability of low pair-wise genetic distance and low overall genetic distance among all populations of *Coffea arabica*. This might be due to the existence of relatively related genetic makeup between geographically adjacent populations and basis that the population may be have beenoriginated in the same geographical region. Another factors affecting genetic distance for a given population is selection. Selection may be posed by natural abiotic and biotic factors or may be practiced by humans (artificial selection). Natural selection usually leaves more heterogeneous population with broad genetic variation while artificial selection results in more uniform population with improved economically important traits and low genetic base (Jolley *et al*., 2004).

Table 7: Pair wise population matrix of Nei's unbiased genetic distance of *Coffea arabica* populations from different source of origin

	HG	HR	IB	YC	WG	KF	BM	OM	JM	BL
HG	0.000									
HR	0.074	0.000								
IB	0.287	0.116	0.000							
YC	0.247	0.170	0.073	0.000						
WG	0.295	0.292	0.164	0.161	0.000					
KF	0.426	0.396	0.204	0.120	0.087	0.000				
BM	0.225	0.272	0.155	0.104	0.051	0.102	0.000			
OM	0.333	0.245	0.125	0.150	0.059	0.085	0.075	0.000		
JM	0.336	0.259	0.123	0.095	0.048	0.093	0.097	0.118	0.000	
BL	0.4	0.325	0.377	0.383	0.315	0.315	0.383	0.254	0.406	0.000
Mean	0.29	0.24	0.17	0.17	0.16	0.2	0.16	0.16	0.175	0.35
Overall mean = 0.21										

HG = Hararge, HR= Harar, IB= Ilubabor, YC= Yirgachefe, WG= Wolegga, KF= Kefficho/Shekicho, BM= Bench Maji, OM= Omo, JM= Jimma and BL= Bale

4.4 Cluster Analysis and Relationships among Accessions

The genetic relationships reflect the interactions among genotypes with regard to their long-term evolutionary history, mutations, recombination, genetic drift, reproductive system, gene flow and natural selection (Rao and Hodgkin, 2002; Govindraj*et al*., 2015). Thus, an understanding of the level and relationships of Ethiopian Coffee is necessary for the conservation and efficient use of the germplasm available for breeding purposes. Cluster analysis of 86 *Coffea arabica* genotypes revealed three major clusters, consisting of 2.3%, 58% and 39.5 % of the total populations in clusters I, II and III respectively (Figure 1). Similarly, Hussein *et al.* (2017) grouped Yemeni coffee into three major clusters that contain seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (*Coffea arabica* L.) germplasm collected from different Governorates in Yemen, using 16

SSR markers. In other study, Shimekit *et al.* (2014) reported two major clusters of 87 *Coffea arabica* that collected from southern Ethiopia using five ISSR markers.

The first cluster (Cluster I) contain only two accessions (YC30 andWG40) from Yirgachefe and Wolegga populations. These two accessions grouped together in cluster I regardless of their geographical origin.Cluster II and III, classified in to different hierarchical sub-groups. The Second cluster included most of the genotypes (50) from populations of all of the ten populations (Hararge= 100%, Harar = 75%, Ilubabor= 55.5 %, Yirgachefe = 66.65, Wolegga= 55.5 %, Kefficho =33.3 %, Benchi Maji= 50 %, Omo=55.5 %, Jimma 22.2 % and Bale= 75 %). It is dominated by accessions from Hararge, Harar and Bale. Cluster III included 34 genotypes from nine populations (Harar= 25%, Ilubabor= 44.5%, Yirgachefe = 22.2%, Wolegga= 33.3%, Kefficho =66.7%, Benchi Maji= 50%, Omo=44.5%, Jimma = 78.8% and Bale= 25 %) only the Accessions from Hararge were not included in cluster III. This cluster is dominated by accessions collected from Jimma and Kefficho.

Accessions from different populations were clustered together, which may imply the existence of gene flow between and within populations. Hararge accessions were grouped with geographically distant accessions from Omo in cluster II. Accessions from Bale were also grouped with Ilubabor accessions. In all of the clusters many accessions are grouped with geographically distant populations. This indicates accessions in one cluster might be evolved from different lines of ancestry. In addition, the independent events of evolutionary forces such as genetic drift, mutation, migration, natural and artificial selection and germplasm exchange might be separated them into related but different gene pools (Slatkin, 1985). These also indicated that genotypes from different seed types might have similar genetic background for microsatellite markers.

Generally, the unweighted pair group method with arithmetic mean cluster analysis of 86 *Coffea arabica* individual plants that grouped in to ten populations revealed a weak clustering pattern confirming low genetic differentiation among the populations and suggesting that the genetic background of *Coffea arabica* populations do not always correlate with their geographical origin. The accessions in each populations revealed that they were distributed

and inter-mixed with accessions of another populations. Hence, the inter-mixed UPGMA dendrogram of 86 *Coffea arabica* individual plants indicates high level of genetic variation among individual genotypes investigated and the clustering pattern is weak to support the concept of "isolation by distance" (Figure 1).

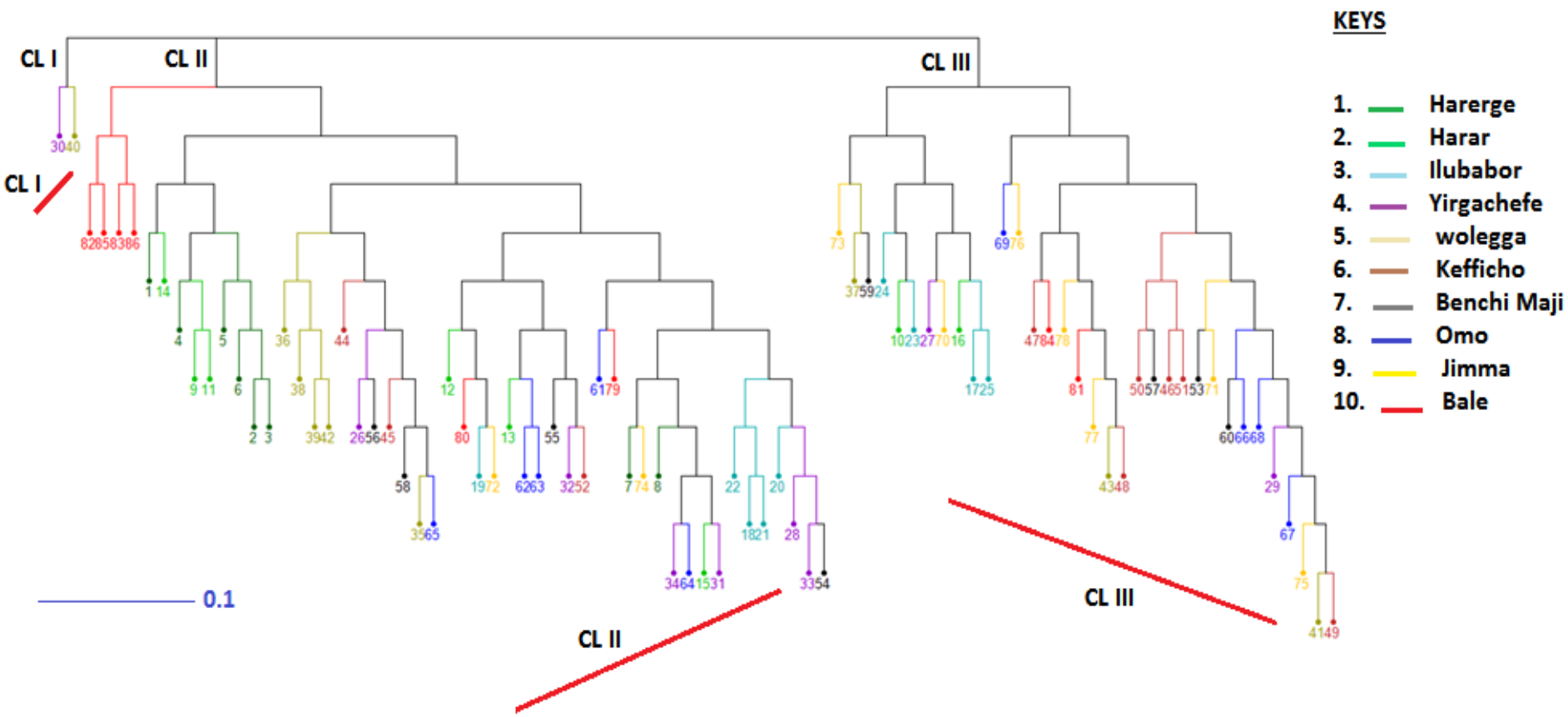


Figure 1: UPGMA dendrogram for 86 *Coffea arabica* accessionsbased on the Jaccard's coefficient as revealed using 20microsatellite markers**.**

4.5. Principal Coordinates Analysis and Relationships among Accessions

Principal coordinate analysis (PCoA) showed that the first two coordinates accounted for about 21.22 % of the genetic variation present in SSR molecular data derived from the study. The first and second principal coordinates explained about 14.27% and 6.95%of the gross variation respectively, (Table 8).

Table 8: Percentage of variation explained by the first two principal components using 20 SSR markers across 86 *Coffea arabica* accessions

Axis	1	**2**
Individual %	14.27	6.95
Cumulative %	14.27	21.22

The PCoA analysis in the two dimensional plot displayed in Figure 2 showed that accessions from different collection sites often grouped together. There was no separate group formed by a single population. This, in turn, agrees with the results of the UPGMA dendrogram in thatthere was no unique clustering among accessions from the same population. In some cases, accessions of the same population such as Hararge and Ilubaborformed sub cluster in the major groups. Even though, some of the accessions forming a sub cluster in their specific group, there was no separate group formed by a singlepopulation. The overallgrouping pattern of PCoA corresponds with the clustering of UPGMA dendrogram (figure 1)which explains about conformity of the results obtained from the cluster analysis.

The presence of seed exchange and high gene flow between and within populations orcollection sites may be the probable explanation behind the mixed clustering of accessionsfrom different populations.In the present study a principal coordinate analysis (PCoA) demonstrated that all populations of Ethiopian *C. arabica* were related, and inter mixed with no clear cut defined genetic clusters. Principal coordinates analysis also revealed the absence of structure in accordance with the geographical origin of the country. The

overlap between accessions from different populations indicates high rate of inter-mixing between Ethiopian *C. arabica*.

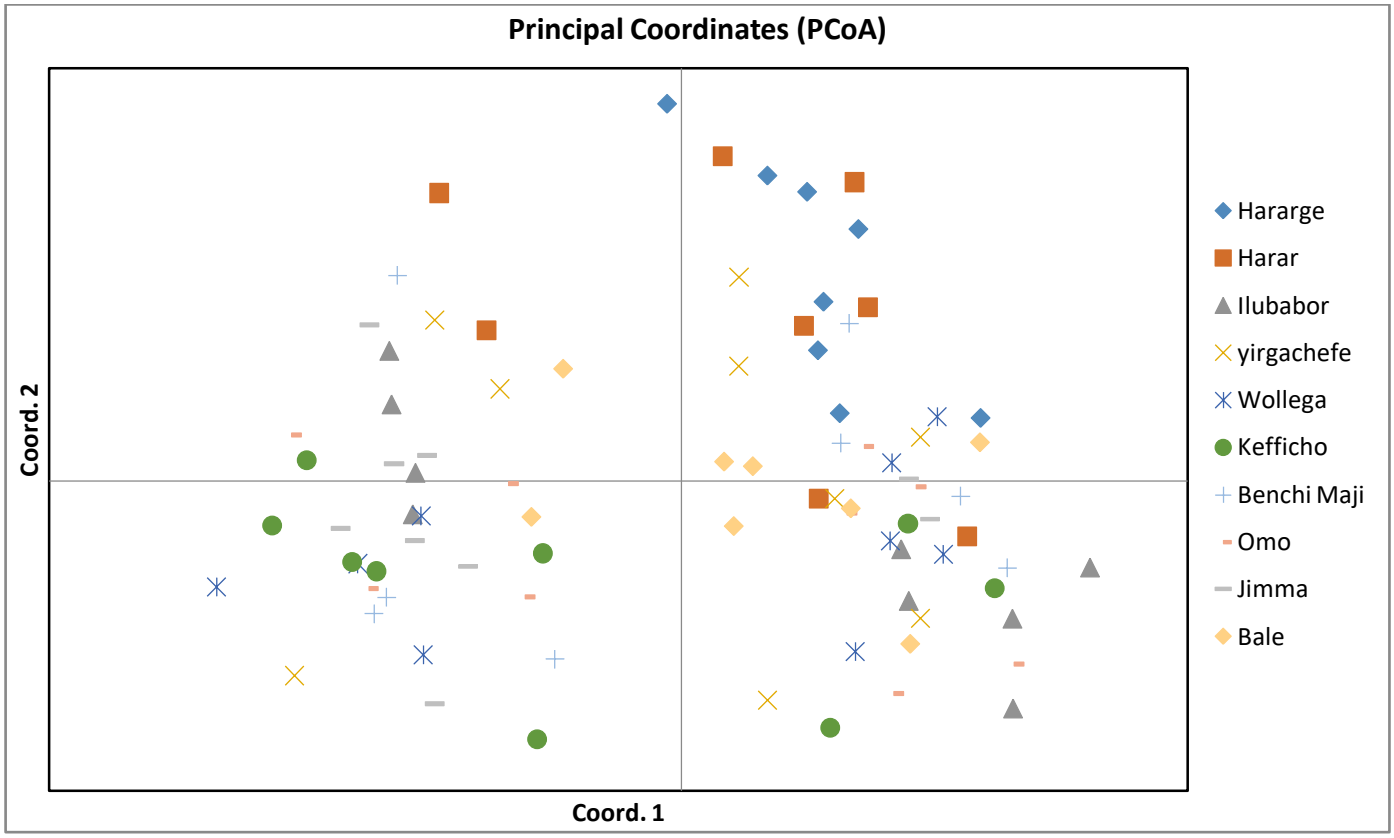


Figure 2: Principal coordinates analysis (PCoA) bi-plot showing the clustering pattern of 86 *Coffea arabica* accessionsbased on 20 SSR. Accessions coded with the same symbol and color belongs to the same population.

5. SUMMARY AND CONCLUSIONS

Coffea arabica is the most important commercial species in the world. It contribute to foreign exchange and serves as a means of livelihood for millions of people worldwide. In many countries including Ethiopia, coffee is one of the primary economic bases which provideemployment and income for millions of people involved in its production, processing andmarketing. Hence, it is an indispensably important commodity crop to the world and Ethiopians alike. In Ethiopia around 40% of foreign income comes from coffee and hence there are needs and plan to improve coffee productivity. Even though Ethiopia is the land of origin and diversity for *Coffea arabica*; manipulating the genetic potential of the crop at molecular level is very low because of the limited information available regarding the molecular diversity and characterization state of the germplasm. More than 4000 collections are available in Ethiopian Biodiversity Institute; however, there is littlemolecular characterization for this germplasm. Hence, the objective of this study was to study the genetic diversity of Ethiopian coffee populations.

In this study, 86 *Coffea arabica* accessionsthat from ten populations based on geographical origin were used to assess the extent of genetic diversity, using polymorphic 20 SSR markers. Atotal of 112 alleles were detected with estimated product size range of 100 to 385 base pairs.The numbers of alleles ranged from three to 11 per markerswith a mean of 5.6 per locus. All the markers were polymorphic with major allele frequency rangedfrom 0.16 to 0.57 with a mean of 0.36 and the polymorphic information contents (PIC) value ranged from 0.45 to 0.75 with a mean of 0.6 per locus. Sixteen SSR markers were highly withPIC value of greater than0.50.

The analysis of diversity parameter indices based on number of alleles, effective number of alleles, private allelic richness, Shannon's diversity index, observed heterozygosity, expected heterozygosity and percentage of polymorphic loci clearly indicate that there is no much difference among the ten populations of *Coffea arabica* and the mean observed different diversity parameter indices indicates the existence of high genetic diversity within populations of *C. arabica.* The high percentage of polymorphism for populations investigated indicates

the presence of high genetic polymorphism and high genetic variability in Ethiopian Coffee. Comparatively, the analysis showed that, the population from Omo, Ilubabor, Benchi Maji and Kefficho scored higher diversity parameter indices that indicate the populations were genetically more diverse than the other populations.Therefore, these areas were considered as hot spot or diversificationlocations, a potential in-situ conservation sites for Ethiopian Coffee and more interesting in terms of genetic and evolutionary studies.On the contrary, Bale, Harar and Hararge populations have the least genetic diversity relatively, which might suggest current rapid genetic erosion from the area.

The analyses of molecular variance (AMOVA) indicate high variation within populations and low variation among populations. The overall observed gene flow (Nm) observed was very high (Nm= 2.45) and the measure of population genetic differentiation was low (Fst= 0.037) which revealed that *C. arabica* has very low genetic differentiationamong populations which indicates high gene flow between populations that implies high frequency of identical alleles among *Coffea arabica* accessions which will agree with the AMOVA result showing low variation between populations. The pair-wise genetic differentiation among populations ranged from 0.041 to 0.108. The lowest pair wise population differentiation was observed between Benchi Maji and Wolegga, Omo and Wolegga, Jimma and Wolegga whereas the highest population differentiation observed between Bale and Hararge, Kefficho and Hararge, Jimma and Bale.

The pair wise genetic distance value ranged from 0.048 to 0.426 whereas the mean genetic distance among populations ranged from 0.16 to 0.35 with an overall mean of 0.21. This result indicates the low genetic distance among the studied populations of Ethiopian coffee. The highest pairwise geneticdistance (0.426) was observed between Hararge and Kefficho populations, Bale and Hararge populations (0.406) and Bale and Jimma populations (0.4).The population from Bale separated from all populations at highest mean genetic distance (0.35). This population was the most genetically distinct population. The population from Hararge also scored the higher mean genetic distance (0.29) next to Bale population. The limitation of genetic material exchange by human or natural factors may be considered as the main factor for the larger distance obtained. On other hand, the smallest pair wise genetic distance was

observed between (Jimma and Wolegga) and (Benchi Maji and Wolegga) populations in order of their magnitude. Similarly, the population from Wolegga and Benchi Maji scored the lowest mean genetic distance (0.16) compared to other populations.

The UPGMA cluster analysis of 86 *Coffea arabica* genotypes revealed three major clusters, consisting of 2.3%, 58% and 39.5 % of the total populations in clusters I, II and III respectively. The cluster analysis revealed a weak clustering pattern confirming low genetic differentiation among the populations and the genetic background of *Coffea arabica* populations does not always correlate with their geographical origin. The genotypes in each population were distributed and inter-mixed with genotypes of other populations. Hence, the inter-mixed UPGMA dendrogram of 86 *Coffea arabica* individual plants indicates high level of genetic variation among individual genotypes investigated.

The principal co-ordinate analysis (PCoA) of 86 accessions of Ethiopian Coffee showed that the first two coordinates accounted forabout 21.22 % of the genetic variation present in SSR molecular data derived from the study.PCoA demonstrated that all populations of Ethiopian *C. arabica* were related, and inter mixed. PCoA also revealed the absence of structure in accordance with the geographical origin of the country. The overlap between species from different populations was observed that indicates high rate of inter-mixing between Ethiopian *C. arabica* that agree with the UPGMA dendrogram.

In conclusions,

- The SSR markers discriminated all the genotypes in all populations, detected high number of allele and moreover, most of the SSR markers used were highly informative
- High genetic variability was observed within Ethiopian coffee populations
- Among populations studied the populations from Omo, Ilubabor, Benchi Maji and Kefficho were relatively more genetically diverse than the others and hence, they can be considered as hot spots for conservation and sources of desirable alleles for breeding values.

- Overall there is a genetic diversity within Ethiopian coffee populations which could be used to further broaden the genetic base and enlarge number of available Ethiopian coffee germplasm.
- Low population genetic differentiation, low population genetic distance and high rate of gene flow were observed among Ethiopian Coffee(*Coffea arabica* L.).

Future line of works,

- The number of markers used in this study was limited, and it will be good to use more number of markers that covering more of *Coffea arabica* genome in the future.
- Although SSR markers are multi-allelic and codominant, it is good to use high resolution markers such as SNP markers due to its genome wide abundance and amenability for high to ultra-high-throughput detection platforms in future works.
- The present study was conducted on very limited number of genotypes, due to the limitation of the resources, further studies that include germplasm from the remaining area would be important to reveal additional potential sites for conservation and breeding needs.

6. REFERENCES

Adebabay, T., 2015.Molecular Characterization of Introduced Sugarcane (*Saccharumofficinarum*L.) Genotypes in Ethiopia Using Inter Simple Sequence Repeat (ISSR) Markers. MSc thesis, Addis Ababa University.129p.

Aerts, R., Gazahegni, B., Gijbels, P., Kittessa, H., Van Glabeke, S., Vandepitte, K., Muys, B., Roldan-Ruiz, I. and Honnay, O., 2013. Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rain forests. *Evolutionary Applications*, **6(2):**243-252

Aigaforum, 2018.Economic Benefit of Ethiopian Coffee. Retrieved 19 Feb. 2018, from (http://www.aigaforum.com/article2018/Economic-Benefit-of-Ethiopian-Coffee.pdf).

Al-Abdulkader, A.M., Al-Namazi, A.A., Al Turki, T.A., Al-Khuraish, M.M. and Al-Dakhil, A.I., 2018. Optimizing coffee cultivation and its impact on economic growth and export earnings of the producing countries: The case of Saudi Arabia. *Saudi journal of biological sciences*, **25(4):**776-782

Alemayehu, T., Crouzillat, D, Petiard,V., Brouhan, P., 2010. Genetic diversity of Arabica coffee (*Coffea arabica L.)* collections. *EJAST*,**1(1):**63-79

Al-Murish, T.M., Elshafei, A.A., Al-Doss, A.A. and Barakat, M.N., 2013. Genetic diversity of coffee (*Coffea arabica L.)* in Yemen via SRAP, TRAP and SSR markers. *Journal of Food, Agriculture & Environment*, **11(2):**411-416.

Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G. and Lashermes, P., 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theory of Applied Genetics*, **104**:894–900.

Benson, O.N., Beatrice, A.W., Samuel, G., Otto,G.D., Augustino, O., 2013. Genetic diversity in cultivated sesame and related wild species in East Africa. *J Crop Sci Biotech*,**16**: 9-15

Bhandari, H.R., Bhanu, A.N., Srivastava K., Singh M.N., Shreya, 2017. Assessment of genetic diversity in crop plants. An overview: *Advance Plants Agricultural Research*,**7(3):**00255

Bigirimana, J., Njoroge, K., Muthomi, J.W., Gahakwa, D., Phiri, N.A., Gichuru, E.K. and Walyaro, D.J., 2013. Genetic diversity among disease resistant coffee varieties and cultivars in Rwanda based on RAPD and SSR markers. *Journal of Renewable Agriculture*, **1(6):**106-112.

Bodia, A., M. A., Elhoumaizi, K. N. Ndir, A. Hasnoui, M. Nachtigall and P. Wehling, 2012. Genetic diversity analysis of date palm (*Phoenix dactylifera L.)* cultivars from Figuig oasis (Morocco) using SSR markers. *IJSAT*, **2(3):**96–104.

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. *Am. J. Hum. Genet*.**32:**314-331

Bunn, C., Läderach, P., Ovalle-Rivera, O., Kirschke, D., 2015. A bitter cup: climate change profile of global production of Arabica and Robusta coffee. *Clim. Change*, **129**: 89-101,

Carvalho, A., 1988. Principles and practice of coffee plant breeding for productivity and quality factors: *Coffea arabica*. *Coffee agronomy*, **4**:129-165

Chen, P. H., Pan, Y. B., Chen, R. K., Xu, L. P. and Chen, Y. Q., 2009. SSR marker-based analysis of genetic relatedness among sugarcane cultivars (*Saccharum spp*. hybrids) from breeding programs in China and other countries. *Sugar Technol*. **11:** 347-354

Chevalier, A., 1947. Les caféiers du globe. Fasc. 3: Systematique des cafeierset faux-caféiers maladies et insectesnuisibles. Paul Lechevalier. 356p.

Cho, Y. G., McCouch, S. R., Kuiper, M., Kang, M. R., Pot, J., Groenen, J. T. M. and Eun, M.Y., 1998. Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). *Theor, Appl. Genet*. **97:** 370-380

Combes, M.C., S. Anderzejewski, F. Anthony, B. Bertrand, P. Rovelli, G. Graziosi and P. Lashermes, 2000. Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology*,**9**: 1178-1180.

Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueno, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J. and Arief, V., 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*,**186**:713–724.

Cubry, P., P. Musali, H. Legnate, D. Pot, F.D. Bellis, V. Poncet, F. Anthony, M. Dufour and T. Leroy, 2008. Diversity in Coffee assessed with SSR markers: Structure of the genus *Coffea* and perspectives for breeding. *Genome*,**51**:50-63.

da Silva, B.S.R., Sant'Ana, G.C., Chaves, C.L., Androcioli, L.G., Ferreira, R.V., Sera, G.H., Charmetant, P., Leroy, T., Pot, D., Domingues, D.S. and Pereira, L.F.P., 2019. Population structure and genetic relationships between Ethiopian and Brazilian *Coffea arabica* genotypes revealed by SSR markers. *Genetica*, **147(2):** 205-216.

Davis, A.P., Tosh, J., Ruch, N., Fay, M.F., 2011. Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot J Linn Soc***. 167(4):**357–377

Delourme, R., Piel, N., Horvais, R., Pouilly, N., Domin, C., Vallee, P., 2008. Molecular and phenotypic characterization of near isogenic lines at QTL for quantitative resistance

to *Leptosphaemaculans* in oilseed rape (*Brassica napus* L.). *Theor, Appl. Genet.* **117**: 1055–1067.

Dessalegni, A., 2017. Review on Genetic Diversity of Coffee (*Coffea arabica L.)* in Ethiopia. *International Journal of forestry and Horticulture*, **3(2):**10.

Diniz, L.E.C., Ruas, C.D.F., Carvalho, V.D.P., Torres, F.M., Ruas, E.A., Santos, M.D.O., Sera, T. and Ruas, P.M., 2005. Genetic diversity among forty coffee varieties assessed by RAPD markers associated with restriction digestion. *Brazilian Archives of Biology and Technology*, **48(4):**511-521.

Edwards, M.D., Helentjaris, T., Wright, S. and Stuber, C.W., 1992. Molecular-marker-facilitated investigations of quantitative trait loci in maize. *Theoretical and Applied Genetics*, **83(6-7):**765-774.

Ervine, F.R., 1969. West African Agriculture: West African Crops, 3rd Edition, Vol. 2. 272p.

Esayas, A., Bryngelsson T., Endashew, B., Salomon, B., 2003. Genetic diversity of forest arabica coffee (*Coffea arabica L.)* in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis. *Hereditas,***138**:36-46

FAO, 1968. Coffee Mission to Ethiopian 1964-1965. Food and Agriculture Organization, Rome, Italy. 200p.

Fazuoli, L.C., Maluf, M.P., GuerreiroFilho, O., Medina Filho, H. and Silvarolla, M.B., 2000. Breeding and biotechnology of coffee. pp. 27-45. In: Sera T., Soccol C.R., Pandey A., Roussos S. (eds).Coffee biotechnology and quality, Springer, Dordrecht.

Fekadu, G., Kassahun, T., Kifle D. and Mulatu, G., 2018. Genetic diversity and population structure analyses of *Plectranthusedulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC genetics*, **19(1):**92.

Ferrao, L.F.V., Caixeta, E.T., Pena, G., Zambolim, E.M., Cruz, C.D., Zambolim, L., Ferrao, M.A.G. and Sakiyama, N.S., 2015. New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. *Molecular breeding*, **35(1):**31

Futuyma, D.J., 2008. Sympatric speciation: norm or exception? Speciation and radiation. pp. 136-148. In: Futuyma, D.J. (eds) the evolutionary biology of herbivorous insects. University of California Press.

Ganal, M.W., Wieseke, R., Luerssen, H., Durstewitz, G., Graner, E.M., Plieske, J. and Polley, A., 2014.High-throughput SNP profile of genetic resources in crop plants using genotyping arrays, PP 113-130. In: R. Tuberosa, A. Graner, and E. Frison (eds). Genomics of Plant Genetic Resources. Dordrecht: Springer.

Gichimu, B.M. and Omondi, C.O., 2010. Morphological characterization of five newly developed lines of Arabica coffee as compared to commercial cultivars in Kenya. *Int. J. plant breed. Genet*,**4(4):**238-246.

Gichimu, B.M., Kathurima, C.W., Kenji, G.M., Muhoho, S.M., Boulanger, R. and Gichuru, E.K., 2012. Genetic diversity among commercial coffee varieties, advanced selections and museum collections in Kenya using molecular markers.*International Journal of Biodiversity and Conservation*,**4(2):** 39-46.

Govindaraj, M., M.Vetriventhan and M. Srinivasan, 2015. Importance of genetic diversity Assessment in crop plants and its recent advances: An overview of its analytical perspectives. A review article. *Genetic Research International*, **2015:**1-14.

Gupta, P.K., Varshney, R.K., Prasad, M., 2002. Molecular Markers: Principles and Methodology. pp. 9-54. In: Jain S.M., Brar D.S., Ahloowalia B.S. (eds). Molecular Techniques in Crop Improvement. Springer, Dordrecht.

Habtamu, G., Gizachew, A., Meseret, D., Ashenafi A., 2018. Arabica Coffee *(Coffea arabica L.)* Hybrid Genotypes evaluation for Growth Characteristics and Yield Performance under Southern Ethiopian Growing Condition. *Acad. Res. J. Agri. Sci. Res.***6(2):** 89-96.

Hadrys, H., Balick, M. and Schierwater B., 1992. Applications of Random Amplified Polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol*. **1:** 55-63.

Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D.A., Yang, Z., Zhao, L., Zhou, G., Wang, Z., Huang, L. and Zhang, Z., 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist*, ***209*(2)**:871-884.

Hartl, L., Mohler, V., Zeller, F. J., Hsam, S. L. K. and Schweizer, G., 1999. Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticumaestivum*L.*).Genome,* **42:** 322-329.

Henry, R. J., 2001. The DNA Fingerprinting of Plants: Plant Genotyping CAB1 Publisher Wallingford, UK. 512p.

Hue, T.M.H., 2005. Genetic variation in cultivated coffee (*Coffea arabica L.)* accessions in northern New South Wales, Australia. Master thesis. Southern Cross University. 117p.

Hussein, M.A.A., Al-Azab, A.A.A., Habib, S.S., El Sherif, F.M. and El-Garhy, H.A., 2017. Genetic Diversity, Structure and DNA Fingerprint for Developing Molecular Ids of Yemeni Coffee (*Coffea arabica L.)* germplasm assessed by SSR markers. *Egypt. J. Plant Breed*. **21(4):**713– 736

International Coffee Organization (ICO) (2019) http://www.ico.org/price s/po-production pdf Accessed 19 February, 2019.

Jiang, G. L., 2013. Molecular markers and marker-assisted breeding in plants. pp. 46-83. In**:** Jiang, G. L. (Eds).PlantBreeding from Laboratories to Fields, Black Well Publisher, London, UK.

Jolley, K. A., D. J. Wilson, P. Kriz, G. McVean and M. C. J. Maiden, 2004. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. *Molecular Biology and Evolution,***22(3):**562-569

Kalinowski S. T., 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol. Ecol. Notes,* **5**:187–189.

Kalinowski, S.T., 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation genetics, 5***(4):**539-543.

Kassahun, T., 2006. Genetic Diversity of wild Arabica Coffee populations in Ethiopia as a contribution to conservation and use planning. Ecology and Development Series 44. University of Bonn, Germany.

Kassahun, T., Govers, K., Endashew, B. and Borsch, T., 2014. ISSR fingerprinting of *Coffea arabica* throughout Ethiopia reveals high variability in wild populations and distinguishes them from landraces. *Plant systematics and evolution,* **300(5):** 881-897

Kesawat, M. S. and Das, B. K., 2009. Molecular markers: It's application in crop improvement. *J. Crop Sci. Biotech*. **12:** 169-181.

Khanuja, S.P., Shasany, A.K., Darokar, M.P. and Kumar, S., 1999. Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Molecular Biology Reporter*, **17(1):**74-74.

Krishnan, S., 2017. Sustainable Coffee Production. Oxford Research Encyclopedia of Environmental Science. Retrieved 19 Sep. 2019, from,https://oxfordre.com/environmentalscience/view/10.1093/acrefore/9780199389414.001.0001/acrefore-9780199389414-e-224.

Lashermes, P., Comes, M.C., Trouslot, P., Anthnony, F., Hamon, S. and Charrier, A., 1999. Molecular characterization and origin of the *Coffea arabica L*. genome. *Molecular Genomics and Genetics,***261**:259-266.

Lashermes, P., Combes, M.C., Cros, J., Trouslot, P., Anthony, F. and Charrier, A., 1995. Origin and genetic diversity of *Coffea arabica L*. based on DNA molecular markers. pp. 528-535. In *16th Conference of ASIC, Kyoto (Japan)*.

Lashermes, P., Combes, M.C., Trouslot, P. and Charrier, A., 1997. Phylogenetic relationships of coffee-tree species (*Coffer L.)* as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics*, **94(6-7):**947-955.

Lashermes, P., Trouslot, P., Anthony, F., Combes, M.C. and Charrier, A., 1996. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica*, **87(1):** 59-64.

Leberg, P.L., 1991. Influence of fragmentation and bottlenecks on genetic divergence of wild turkey populations. *Conservation Biology*, **5(4):**522-530.

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

Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*,**15**: 65–95.

Lynch, M. and Walsh, B., 1998. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates, 980p.

Maluf, M. P., Silvestrini, M., Ruggiero, L.M., Guerreiro, Filho, O., Colombo, C., 2005. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Sci. Agric*.**62(4):**366–373

Mangal, S.K., 2007. Coffee: Planting, Production and Processing. Gene-Tech Books, New Delhi, 266p.

Marulanda, M.L., Lopez, A.M., Isaza, L. and Lopez, P., 2014. Microsatellite isolation and characterization for Colletotrichumspp, causal agent of anthracnose in Andean blackberry. *Genet.Mol. Res*. **13**:7673-7685.

McDermott, J.M. and McDonald, B.A., 1993. Gene flow in plant patho systems. *Annual review of phytopathology*, **31(1):** 353-373.

Mekuria, T., D. Neuhoff and U. Kopke, 2004. The Status of Coffee Production and the Potential for Organic Conversion in Ethiopia. Berlin: Conference on International Agricultural Research for Development, 5-7 October 2004. Deutscher, Tropentag.

Melkamu, A., 2015. Ethiopian Highlands: Home for Arabica Coffee (*Coffea arabica* L.*)*. pp.58-65. In: Enyew, A., Mekete D., Jan N. (Eds). Tropical lakes in a changing environment: water, land, biology, climate and humans. Bahir Dar, Ethiopia, Bahir Dar University.

Mesfin, A. and Bayeta, B., 2008. Phenotypic Diversity in the Hararge Coffee (*Coffea arabica* L.) Germplasm for Quantitative Traits. *East African Journal of Science*,**2(1):**13-18.

Mesfin, A. and Bayeta, B., 1987. Field evaluation of resistance to stress conditions in crosses and parents of *Coffea arabica L. EJAST,**14**:4-9*

Meyer, F.G., Fernie, L.M., Narasimhaswamy, R.L., Monaco, L.C., Greathead, D.J., 1968. Food and Agriculture organization. Coffee mission to Ethiopia 1964–1965. FAO, Rome, 200p.

Missio, R.F., Caixeta, E.T., Zambolim, E.M., Pena, G.F.,Zambolim, L., Dias, L.A.S. and Sakiyama, N.S., 2011. Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. *Genet Mol Res*.**10(4):**2.

Moncada, P., and Couch, S. M., 2004. Simple sequence repeats diversity in diploid and tetraploid *Coffea* species. *Genome*,**47**:501-509.

Mondini, L., Noorani, A. and Pagnotta, M., 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, **1(1**):19-35.

Montagnon, C. and Bouharmont, P., 1996. Multivariate analysis of phenotypic diversity of *Coffea arabica* .*Genetic. Resour. Crop. Evol*.**43**: 221-227.

Motta, L.B., Soares, T.C.B., Ferrao, M.A.G., Caixeta, E.T., Lorenzoni, R.M. and Souza Neto, J.D.D., 2014. Molecular characterization of arabica and Conilon coffee plants genotypes by SSR and ISSR markers. *Brazilian Archives of Biology and Technology*,**57(5):**728-735.

Mulatu, G., Herrera I, Monzón A, Bryngelsson T., 2012. Genetic diversity of Arabica coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. *Sci. World J*.**2012**:1–11

NCA (National Coffee Association), 2017. Coffee around the world. Retrieved 24 Sep, 2018 from (http://www.ncausa.org/About-Coffee/Coffee-Around-the-World).

Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*,**89**: 583-590.

Olika, K., Sentayehu, A., Taye, K. and Wayessa, G., 2011. Variability of quantitative traits in limmu coffee (*Coffea arabica* L.) in Ethiopia. *International journal of agricultural research*, **6(6):** 482-493.

Omingo, D.O., Omondi, C.O., Cheserek, J., Runo, S. and Okun, D., 2017. Diversity analysis of selected coffee genotypes using microsatellites and random amplified polymorphic DNA in Kenya. *International Journal of Biotechnology and Food Science*,**5(1):** 1-9

Orozco-Castillo, C., Chalmers, K.J., Waugh, R. and Powell, W., 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and applied genetics*, **87(8):**934-940.

Paglia, G. and Morgante, M., 1998. PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. *Mol. Breed.* **4:**173-177.

Pan, Y. B., Burner, D. M., Ehrlich, K. C., Grisham, M. P. and Wei, Q., 1997. Analysis of primer-derived non-specific amplification products in RAPD-PCR. *Bio techniques,***22:** 1071-1077.

Parker, P. G., Snow, A. A, Schug, M. D., Booton, G. C. and Fuerst, P. A., 1998. What molecules can tell us about populations: choosing and using a molecular marker? *Ecol.* **79:** 361-382.

Pearl, H.M., Nagai, C., Moore, P.H., Steiger, D.L., Osgood, R.V. and Ming, R., 2004. Construction of a genetic map for Arabica coffee. *Theoretical and Applied Genetics,* **108 (5):**829-835.

Perrier X. and Jacquemoud-Collet J. P., 2006. DARwin software. http://darwin.cirad.fr/.

Pestana, K.N., Capucho, A.S., Caixeta, E.T., de Almeida, D.P., Zambolim, E.M., Cruz, C.D., Zambolim, L., Pereira, A.A., de Oliveira, A.C.B. and Sakiyama, N.S., 2015. Inheritance study and linkage mapping of resistance loci to Hemileiavastatrix in Híbrido de Timor UFV 443-03. *Tree genetics & genomes*, **11(4):** 72.

Rao, V.R. and Hodgkin, T., 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant cell, tissue and organ culture*, **68(1):**1-19.

Rovelli, P., Mettulio, R., Anthony, F., Anzueto, F., Lashermes, P. and Graziosi, G., 2000. Microsatellites in *Coffea arabica L*. pp.123-133. In: Sera T., Soccol C.R., Pandey A., Roussos S. (eds).Coffee biotechnology and quality, Springer, Dordrecht.

Saal, B. and Wricke, G., 2002. Clustering of amplified fragment length polymorphism markers in a linkage map of rye. *Plant Breeding*, **121(2):**117-123.

Sant'Ana ,G.C., Pereira L.F., Pot D., Ivamoto, S.T, Domingues, D.S., Ferreira, R.V., Pagiatto, N.F., da Silva, B.S.R., Nogueira, L.M., Kitzberger, C.S., Scholz, M.B., de Oliveira, F.F., Sera, G.H,, Padilha, L.L,, Labouisse, J.P,, Guyot, R., Charmetant, P., Leroy, T., 2018. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica L*. *Sci Rep*.**8**:465.

Setotaw, T.A., Caixeta, E.T., Pereira, A.A., Oliveira, A.C.B, de CRUZ, C.D., Zambolim, 2013. Coefficient of parentage in *Coffea arabica L*. cultivars grown in Brazil. *Crop Sci*.**53(4):**1237-1247.

Sharma, J.R., 1998. Statistical and biometrical techniques in plant breeding. New Age International (P) limited, publishers. New Delhi, 432p.

Shimekit, T., Firew, M. and Kassahun, T., 2014. Genetic diversity of coffee (*Coffea arabica L.)* landraces from southern Ethiopia as revealed by inter simple sequence repeat marker. *Global Advanced Research Journal of Agricultural Science*, **3(1):**024-034.

Silvarolla, M.B., Mazzafera, P. and Lima, M.M.A.D., 2000. Caffeine content of Ethiopian *Coffea arabica* beans. *Genetics and Molecular Biology*, **23(1):**213-215.

Silvarolla, M.B., Mazzafera, P. and Fazuoli, L.C., 2004. Plant biochemistry: a naturally decaffeinated Arabica coffee. *Nature*, ***429*(6994**):826.

Silvestrini, M., Junqueira, M.G., Favarin, A.C., Guerreiro-Filho, O., Maluf, M.P., Silvarolla, M.B. and Colombo, C.A., 2007. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica L*. accessions using microsatellites markers. *Genetic Resources and Crop Evolution*, **54(6):**1367-1379.

Singh, R. K., Mishra, S. K., Singh, S. P., Mishra, N. and Sharma, M. L., 2010. Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. *Aust. J. Crop Sci*. **4:**116-125.

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

Solomon, B., 2007. Genetic Diversity Analysis of the Wild *Coffea arabica* L. Populations from Harenna Forest, Bale Mountains of Ethiopia, Using Inter Simple Sequence Repeats (ISSR) Marker. MSc thesis. Addis Ababa University. Addis Ababa, Ethiopia. 89p.

Sousa, T.V., Caixeta, E.T., Alkimim, E.R., de Oliveira A.C.B., Pereira, A.A.,Zambolim, L., Sakiyama, N.S., 2017a. Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica*, **213**:75

Sousa, T.V., Caixeta, E.T., Alkimim, E.R., de Oliveira, A.C.B., Pereira, A.A., Sakiyama, N.S., de Resende Junior, M.F.R. and Zambolim, L., 2017b. Population structure and genetic diversity of coffee progenies derived from Catuai and Hibrido de Timor revealed by genome-wide SNP marker. *Tree genetics & genomes*, ***13*(6):**124.

Steiger, D., Nagai, C., Moore, P., Morden, C., Osgood, R. and Ming, R., 2002. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theoretical and applied genetics*, **105(2-3):**209-215.

Sylvain, P.G., 1958. Ethiopian Coffee: Its significance to world coffee problems. *Botany*,**12**: 111-139.

Thurston, R.W., Morris, J. and Steiman, S., 2013. Coffee*:* A comprehensive guide to the bean, the beverage, and the industry. Lanham, MD: Rowman & Littlefield, 275p.

Tigist, Y., 2015. Impact of agricultural exports on economic growth in Ethiopia: the case of coffee, oilseed and pulses. Master's thesis, Egerton University. 82p.

Tran, H.T., Lee. L.S., Furtado, A., Smyth, H., Henry, R.J., 2016. Advances in genomics for the improvement of quality in coffee. *J Sci Food Agric*.**96**:3300–3312

Vega, F.E., Ebert, A.W. and Ming, R., 2007. Coffee germplasm resources, genomics and breeding. *Plant breeding reviews,* **30:**415-447.

Vencovsky, R. and Crossa, J., 2003. Measurements of representativeness used in genetic resources conservation and plant breeding. *Crop Science*, **43**:1912-1921.

Vieira, E.S.N., vonPinho, E.V.R, Carvalho, M.G., Esselink, G.D., Vosman, B., 2010. Development of microsatellite markers for the identification of Brazilian *Coffea arabica* varieties. *Genet Mol. Biol.***33**:507-514.

Walyaro, D.J.A., 1983. Considerations in breeding for improved yield and quality in Arabica coffee.Doctoral dissertation, Wagengen University,Neitherlands. 129p.

Waples, R.S., 1987. A multi species approach to the analysis of gene flow in marine shore fishes. *Evolution,***41(2):**385-400.

White head and Peakall, 2015. GENALEX 6.502: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes,* **6**: 288-295.

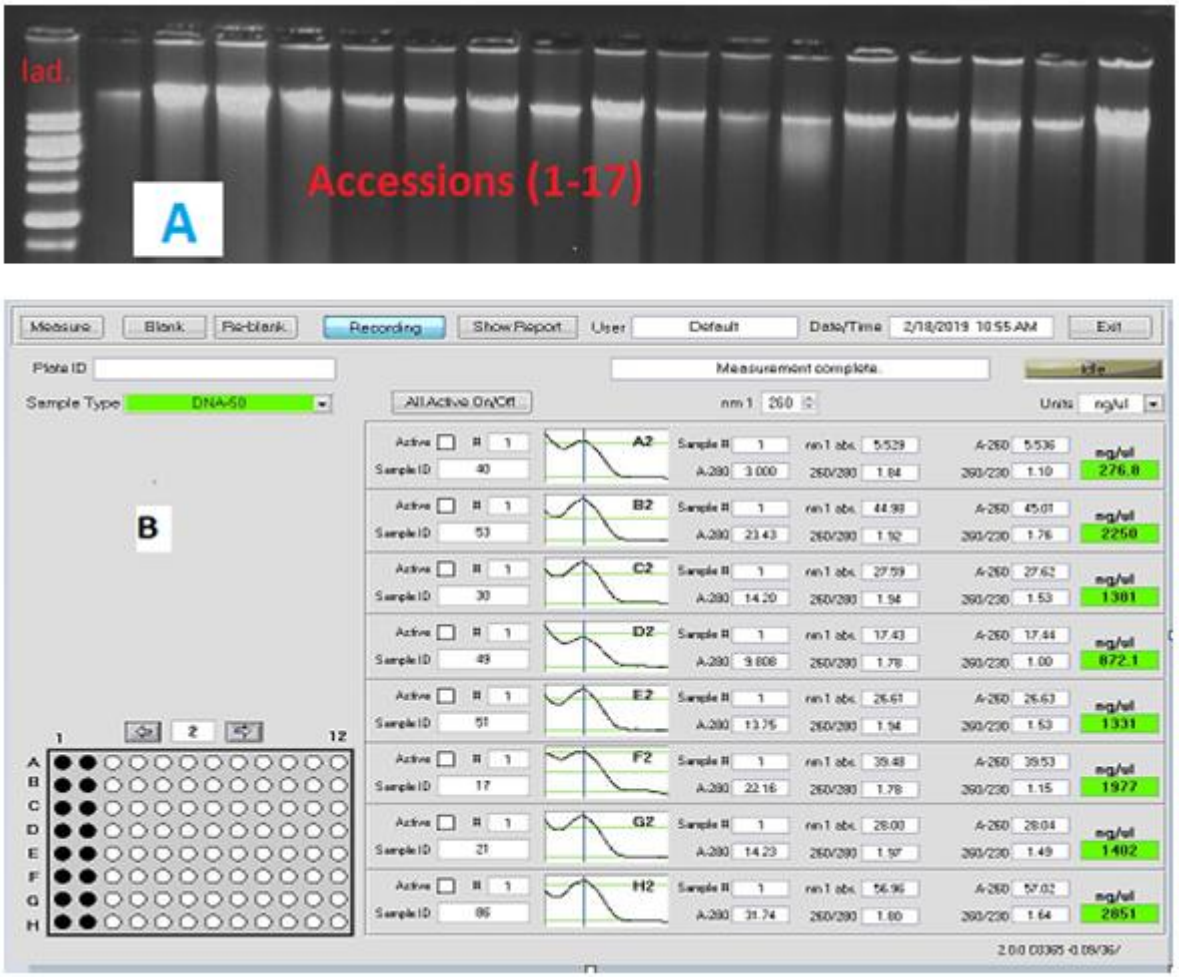
Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research*, **18(22):**6531-6535.

Wintgens, J.N., 2009. Coffee: growing, processing, sustainable production. A guidebook for growers, processors, traders and researchers. Wiley-VchVerlagGmbh& Co. Kgaa. 996p.

Wright, S., 1951. The genetic structure of populations. *Annual Eugenics*, **15**: 323-354.

Yeh, F. C. and Yang, R., 1999. POPGENE VERSION 1.31: Microsoft Window-based Freeware for Population Genetic Analysis. Department of Renewable Resources University of Alberta, Canada. pp. 1-29

7. APPENDIX



Appendix 1: Quality and concentration test of genomic DNA: A) on 0.8% agarose gel electrophoresis and B) withNano Drop Spectrophometry.

Appendix 2: Summary of the number of alleles with their respective frequencies

Locus	Number of alleles with frequency				
	Rare(<0.01)	Scarce (0.01 - 0.05)	0.05 - 0.1	0.1 or higher	Total
Aj308782	0	0	0	5	5
Aj308779	0	1	1	4	6
Aj308753	0	0	1	7	8
Aj250251	0	1	1	3	5
CFGA92	1	0	1	4	6
AJ250253	0	0	0	3	3
CFGA91	2	2	0	3	7
CFGA69	0	0	0	3	3
CFGA502	0	0	0	4	4
AJ250255	0	0	0	4	4
CFGA465	0	4	5	2	11
AJ250257	0	0	0	5	5
AJ308774	0	0	2	3	5
AJ250258		1	0	3	4
CFGA100	0	0	1	5	6
AJ308809	0	1	1	5	7
AJ250260	1	0	1	4	6
AJ308790	1	0	0	4	5
AJ308755	0	0	2	5	7
AJ308837	0	1	0	4	5
percentage	4.46	9.80	14.28	71.43	112