

**ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES**

**GENETIC DIVERSITY OF WILD POPULATIONS OF ARABICA
COFFEE (*Coffea arabica* L.) IN YAYU FOREST OF ETHIOPIA
USING INTER-SIMPLE SEQUENCE REPEAT (ISSR) MARKER**

**By
TAMIRU OLJIRA**

*A Thesis
Presented to the school of Graduate Studies of the Addis Ababa University
in partial fulfillment of the requirement for the Degree of Masters in
Biology (Applied Genetics)*



Addis Ababa
July 2006

Dedication

I dedicate this thesis to my parents: my mother Desaye Reggaa and my father
Oljira Raga

I am also thankful to my colleagues particularly, Abel Gizaw, Edossa Fikru and others for every help they gave me. I thank Abel for giving me some comments and providing me NTSYSpc program and his personal Lap Top computer as well as in preparing slides.

I thank the CoCE project for covering the whole cost of the study and accepting me as M. Sc student to be involved in the project.

I thank the Administration of the Department of Biology, Addis Ababa University, particularly Dr. Dawit Abate, Mrs. Mulu-Alem, and Banchu and all other staffs for their kind assistance.

Haromaya University is highly appreciated for giving me chance to join my postgraduate study and for paying me salary during my study time. I am very much thankful to the staff of the Department of Biology particularly Mr. Meseret Chimdessa and Mr. Olana kumssa for every support they gave me during my study time.

I would like to thank all people that helped me in one way or the other in this thesis work.

Thank You All!!!

Finally I praise and glorify the only Almighty God whose help made all possible!! Thank God.

Table of Contents

ACKNOWLEDGEMENTS.....	I
TABLE OF CONTENTS.....	III
LIST OF TABLES.....	V
LIST OF FIGURES.....	VI
ABSTRACT.....	VIII
1. GENERAL INTRODUCTION.....	1
1.1. GENERAL OVERVIEW OF COFFEE.....	1
1.2. GENETIC DIVERSITY AND ITS USE.....	2
1.3. LOSS OF GENETIC DIVERSITY.....	3
1.4. METHODS FOR CONSERVATION OF GENETIC RESOURCES.....	4
1.5. DETECTING GENETIC DIVERSITY IN GENETIC RESOURCES.....	6
2. LITERATURE REVIEWS.....	8
2.1. TAXONOMY, REPRODUCTIVE BIOLOGY AND MORPHOLOGY OF C. ARABICA.....	8
2.1.1. <i>Taxonomy</i>	8
2.1.2. <i>Morphology and reproductive biology</i>	10
2.2. COFFEE IN ETHIOPIA.....	13
2.2.1. <i>Origin and distribution of C.arabica</i>	13
2.2.2. <i>Coffee production in Ethiopia</i>	15
2.3. GENETIC DIVERSITY OF C.ARABICA.....	16
3. MOLECULAR MARKERS.....	22
3.1. NON-PCR BASED MARKERS.....	24
3.2. PCR-BASED DNA MARKERS.....	26
3.3. INTER-SIMPLE SEQUENCE REPEATS (ISSRs) MARKERS.....	27
4. BACKGROUND AND OBJECTIVES.....	2
4.1. GENERAL OBJECTIVE.....	2
4.2. SPECIFIC OBJECTIVES.....	3
5. MATERIALS AND METHODS.....	4

5.1. DESCRIPTION OF STUDY SITE.....	4
5.2. SAMPLING.....	5
5.3. DNA EXTRACTION.....	7
5.4. INTER-SIMPLE SEQUENCE REPEAT (ISSR-PCR).....	8
5.4.1. <i>Primers used</i>	8
5.4.2. <i>PCR amplification</i>	9
5.4.3. <i>Gel electrophoresis</i>	9
6.1. LOCUS DIVERSITY.....	13
6.2. POPULATION LEVEL POLYMORPHISM.....	13
6.3. CLUSTER ANALYSIS.....	16
6.4. PARTITIONING OF GENETIC DIVERSITY INTO WITHIN AND AMONG POPULATIONS.....	18
6.5. AMONG-POPULATION GENETIC DIFFERENTIATION.....	23
6.6. GENETIC DISTANCE/SIMILARITY BETWEEN POPULATIONS.....	23
7. DISCUSSION.....	25
7.1. THE USE OF ISSR MARKERS IN GENETIC DIVERSITY.....	25
7.2. GENETIC VARIATION.....	25
7.3. DISTRIBUTION OF THE VARIABILITY.....	27
8. CONCLUSION AND RECOMMENDATIONS.....	30
8.1. CONCLUSION.....	30
8.2. RECOMMENDATIONS: IMPLICATION FOR CONSERVATION AND BREEDING.....	31
8.2.1. <i>Implication for conservation</i>	31
8.2. 2. <i>Implication for breeding</i>	32
8. 3. FUTURE RESEARCH DIRECTIONS.....	32
REFERENCES.....	33
DECLARATION.....	44

List of Tables

TABLE 1. COMPARISON OF THE MAJOR TECHNIQUES FOR IDENTIFICATION OF PLANT GENETIC DIVERSITY. PROTEINS (ISOZYMES) ARE ALSO INCLUDED FOR COMPARISON.....	25
TABLE 2. LIST OF <i>C. ARABICA</i> POPULATIONS INCLUDED IN THIS STUDY WITH SAMPLE SIZE (ORIGINAL) AND SITE LOCATION.....	7
TABLE 3. LIST OF PRIMERS USED IN THE ISSR ANALYSIS ALONG WITH THEIR NUCLEOTIDE SEQUENCES, NUMBER OF POLYMORPHIC BANDS AND ESTIMATED MOLECULAR SIZE RANGE.....	14
TABLE 4. FREQUENCY (%) OF THE POLYMORPHIC ISSR MARKERS WITHIN THE FOUR POPULATIONS.....	16
TABLE 5. SUMMARY OF POPULATION LEVEL AND THE WHOLE DATA GENETIC DIVERSITY MEASURES ASSUMING HARDY-WEINBERG DISEQUILIBRIUM (FIS=0.95, ESAYAS, 2003).....	17
TABLE 6. SUMMARY OF NEI'S (1973) GENE DIVERSITY STATISTICS CALCULATED IN SUBDIVIDED POPULATIONS (NEI, 1987), FOR EACH POLYMORPHIC LOCUS.....	20
TABLE 7. SUMMARY OF THE MEAN VALUES OF NEI (1973) GENE DIVERSITY STATISTICS IN SUBDIVIDED POPULATIONS OVER ALL THE LOCI (BOTH POLYMORPHIC AND MONOMORPHIC) FOR ALL THE SAMPLES.....	21
TABLE 8. SUMMARY OF SHANNON'S INFORMATION INDEX FOR EACH POLYMORPHIC LOCUS OVERALL THE SAMPLES	21
TABLE 9. PARTITIONING OF THE GENETIC VARIATION INTO WITHIN AND AMONG POPULATIONS BASED ON SHANNON'S INFORMATION INDEX CALCULATED OVER ALL THE LOCI FOR THE WHOLE DATA (78 PLANTS).....	22
TABLE 10. PARTITIONING OF GENETIC VARIATION INTO WITHIN AND AMONG POPULATIONS BY ANALYSIS OF MOLECULAR VARIANCE (AMOVA).....	22
TABLE 11. SUMMARY OF NEI'S UNBIASED MEASURES OF GENETIC DISTANCE/IDENTITY MATRIX BETWEEN POPULATIONS (NEI, 1978).....	24

List of Figures

- FIGURE 1. *PICTURES SHOWING SOME MORPHOLOGICAL FEATURES OF C. ARABICA AND CROSS-SECTION OF ITS FRUITS.*
PICTURE A SHOWS COFFEE TREE WITH BRONZE-TIPPED LEAVES AND B SHOWS THE BRANCHING PATTERN, THE RIPE AND UNRIPE BERRIES AND THE MAIN STEM. PICTURE C SHOWS THE CROSS-SECTION OF THE FRUIT (ESAYAS, 2006)
..... 11
- FIGURE 2. *SCHEMATIC REPRESENTATION OF THE MAIN STEPS IN THE HISTORY OF COFFEE CULTIVATION AND THE GENETIC DIVERSITY OF ARABICA COFFEE POPULATION OF SOUTHWEST ETHIOPIA, COMPARED TO CULTIVATED POPULATION ABROAD (ANTHONY ET AL., 2002).....* 19
- FIGURE 3. *DNA FRAGMENTS FLANKED BY SSRs (INDIVIDUAL 1 AND 2) ARE AMPLIFIED USING A SSR PRIMER (GREEN ARROW). LENGTH POLYMORPHISMS THAT ARE CAUSED BY INSERTIONS OR DELETIONS (RED MARK) IN DIFFERENT INDIVIDUALS LEAD TO PCR FRAGMENTS (BLUE LINES) VARYING IN SIZE.....* 30
- FIGURE 4. *MAP OF ETHIOPIA SHOWING THE RELATIVE POSITION OF THE STUDY SITE (YAYU). IT IS MARKED AS RED ELLIPSE ON THE ROAD FROM ADDIS ABABA TO METU.....*5
- FIGURE 5. *SAMPLED PLOTS OF YAYU FOREST FOR THE PRESENT ANALYSIS. THE FOUR PLOTS IN THE PRESENT STUDY ARE IV-1, IV-2, IV-4, AND IV-5 AND REPRESENT YAYU-1, YAYU-2, YAYU-4, AND YAYU-5 RESPECTIVELY. SITE IV-3 IS EXCLUDED FROM THE PRESENT STUDY SINCE IT IS NEARER TO AND IN BETWEEN SITES IV-1 AND IV-2. THE MAP IS DRAWN USING GPS COORDINATES TAKEN FROM THE PLOT.....*6
- FIGURE 6. *ETHIDIUM BROMIDE-STAINED GELS OF ISSR BANDING PATTERNS GENERATED FOR 36 INDIVIDUAL COFFEE TREES FROM TWO PLOTS IN YAYU (GEBA DOGI). A IS BASED ON A DI-NUCLEOTIDE 810-H AND B ON A TETRA-NUCLEOTIDE PRIMER CoIS001. OUTSIDE LANES (M) SHOW AN EXTENDED 100-BP LADDER. THE ARROWS ON THE RIGHT SIDE OF BOTH GEL PICTURES INDICATE FRAGMENT SIZES (ARROWS CORRESPOND TO 1000 BP, 1500 BP AND 2000 BP FROM TOP TO BOTTOM).....* 15
- FIGURE 7. *DENDROGRAM GENERATED BY UPGMA BASED ON JACCARD SIMILARITY COEFFICIENT USING 81 SAMPLES, THREE OF THEM FROM SORE FOREST FOR COMPARISON. THE C-SERIES LABELLING INDICATES THE NUMBER OF SAMPLES IN COCE PROJECT FOR MOLECULAR ANALYSIS AND THE ROMAN NUMERALS FOLLOWED BY ARABIC NUMERALS INDICATE YAYU (SITE-4 OF COCE), SAMPLE PLOT, AND SAMPLE NUMBER WITHIN A PLOT.....* 18
- FIGURE 8. *DENDROGRAM GENERATED BASED ON NEI UNBIASED GENETIC DISTANCE (NEI, 1987) AND UPGMA FOR THE FOUR POPULATIONS.....* 24

ABSTRACT

Arabica coffee (Coffea arabica L.), which contributes over 70 percent of the world's coffee productions, is characterized by a low genetic diversity, attributed to its allopolyploidy origin, predominately self-pollinating nature and its recent evolution. A total of 81 wild coffee trees from four populations of Yayu (Geba-Dogi) forest and Sore forest (only three samples for comparison) were used to assess the within and among populations genetic diversity using inter-simple sequence repeat (ISSR) marker. Two populations (Yayu-1 and Yayu-2) are from disturbed region of the forest and two populations (Yayu-4 and Yayu-5) are from undisturbed region of the forest. The samples range in age from about 2 years old to very old trees. Larger genetic diversity estimate was observed for Yayu-2 ($P=27.27\%$, $H=0.079$, $H'j=0.175$) and lower for Yayu-4 ($P=10.91\%$, $H=0.31$, $H'j=0.07$). Individuals in disturbed regions are clearly distinguished from those in undisturbed regions on dendrogram generated based on Jaccard similarity coefficient and unweighted pair-group method using arithmetic averages (UPGMA). Smaller genetic distance exists between Yayu-1 and Yayu-2 (0.011) and larger genetic distance exists between Yayu-1 and Yayu-4 (0.109). Larger proportion of the total genetic diversity exists within population (above 55%), which is the pattern for populations with mixed-mating. The results of the present study suggest that coffee populations in the semi-disturbed regions of the forest have higher genetic variability and need to be given priority for both in situ and ex situ conservation as well as for breeding purposes. Those from undisturbed populations should also be conserved as they may have some unique genes.

Key words: *Coffea arabica*, ISSR, Yayu, intraregional diversity, Ethiopia

1. General Introduction

1.1. Background

1.1.1. General overview of Coffee

Coffee provides one of the most widely consumed beverages in the world, and is a very important source of foreign exchange income for many countries. It is currently being produced in over 60 countries, and it is among the five most important raw goods of the world. For many developing countries it is the most important source of foreign currency, (up to 80% of all foreign income currency), only being surpassed by crude oil (OXFAM, 2002). According to the FAO, in 1999 the value of world coffee imports amounted to US\$ 13.4 billion, but this value decreased significantly to USD 10 billion due to the current coffee crisis. Coffee consumption was approx. US\$ 50 billion per year. Commercial coffee production relies on two species of coffee, *Coffea arabica* and *Coffea canephora* (Coste, 1992). Arabica coffee accounts for over 70 percent/about two-thirds of the world market and production. Arabica beans are generally considered to produce better quality, and more flavourful results (Raina *et al.*, 1998). At least 14 countries are dependent on coffee for 10% of their export earnings and three countries namely Burundi, Ethiopia and Uganda depend on coffee for more than 50% of their export earnings. In Ethiopia coffee production is exclusively from *C. arabica* L. and Arabica coffee is the most important plant in the country's economy as foreign exchange earner, amounting to more than 67 percent of foreign exchange (OXFAM, 2002). Ethiopia is currently the seventh largest coffee producer worldwide and ranked ninth in coffee export. In Ethiopia coffee grows on approx. 400,000 ha. The average annual production amount to about 200,000 t. Coffee is by far Ethiopia's most

important export crop. Based on the current coffee crisis, Ethiopia's exports from coffee decreased from over US \$ 250 million in 1999/2000 to less than US \$ 150 million in 2000/2001(OXFAM, 2002; Tadesse, 2003). Ninety-six percent (96%) of the Ethiopian coffee is produced in traditional coffee production systems of which a significant share is threatened through the coffee crisis.

1.2. Genetic Diversity and its Maintenance

Biological diversity refers to the number, variety, and variability among species of plant, animal, and microorganism and the ecological systems in which they live (UNCED, 1992; IPGRI, 1993). Biological diversity can be defined at three levels: Species diversity, ecosystem diversity, and genetic diversity. Species diversity refers to the variety and abundance of different species and ecosystem diversity is the variety of habitats, such as grasslands or wetlands, occurring within a region. Finally, genetic diversity refers to the variation of genes within species, i.e., the heritable variation within and between populations of organisms. The generation of new genetic variation occurs continuously in individuals through chromosomal and gene mutations, as well as recombination but in sexually reproducing organisms. Genetic variation is also influenced by selection. Modern crops with stored genetic diversity are the results of thousands of years of evolutionary processes (Endashaw, 1983, 1985, 1986). Plant genetic resources (PGR) include the reproductive or vegetative propagated material of (i) cultivars in current use and newly developed varieties, (ii) traditional cultivars and landraces (iii) wild relatives of cultivated species and (iv) elite breeding materials, aneuploids and mutants (Singh, 1996; Karp *et al.*, 1997; IPGRI, 2005).

The existence of wide genetic diversity in a crop plant is also important to individual farmers, communities and to the agriculture of the world. Genetic diversity provides farmers and plant

breeders with options to develop, through selection and breeding, new and more productive crops, that are resistant to virulent pests and diseases and adapted to changing environments (Singh, 1996; Rao, 2004). Diverse genetic resources allow humans to select and breed plants and animals with desired characteristics, thus increasing agricultural productivity. Genetic diversity helps to overcome stresses imposed by pests, parasites, pathogens and environment on crop varieties that range from a reduced productivity to extinction in evolutionary perspective (Endashaw, 1985). Narrow genetic base of a species thus can lead to loss or extinction of that species.

The genetic diversity of crop varieties is particularly high at the centers of diversity (Singh, 1996). Most centers of diversity are found where crops were first domesticated, primarily in today's developing countries (Singh, 1996; Karp *et al.*, 1997). Ethiopia is one of the Vavilonian centers of diversity (Vavilov, 1951 cited in Singh, 1996; Demel, 1999). *C. arabica* is of the most important crops originated and diversified from Ethiopia. It is largely represented in the country by genetically diverse forms of landraces and wild types that are uniquely adapted. It occurs naturally in forests of Ethiopian highlands of SW Ethiopia between 1,400-1,900m (Purseglove, 1968; Demel, 1999; Tadesse, 2003).

Although *C. arabica* L. is cultivated nowadays in over 60 countries of the world, it has a very narrow genetic base which is represented by two baseline cultivars, *Bourbon* and *Typica* (Lashermes *et al.*, 1993; Lashremes *et al.*, 1996; Anthony *et al.*, 2001). However, the genetic diversity in Ethiopia is larger as compared to the other countries (Demel, 1999; Tadesse, 2003). Many different types of landraces and forest/spontaneous forms of the crop represent this diversity.

1.3. Loss of Genetic diversity

Genetic resources are disappearing at unprecedented rates (IPGRI, 1993, 2005). The reasons for this loss are many and include deforestation, developmental activities such as hydroelectric projects, road construction, urbanization and changes in agricultural practices, and finally modern agriculture and introduction of new and uniform varieties (Tadesse, 2003; Rao, 2004). More than 15 million hectares of tropical forest are lost each year (Rao, 2004). Over-grazing and changes in land-use pattern are taking heavy toll on diversity available in the wild species. Wild populations of *C. arabica* L. are one of such highly endangered species due to increasing settlement and land-use pressure on the montane rain forests of SW Ethiopia (Tadesse *et al.*, 2001). Currently only about 2,000 km² forests that is extremely fragmented remain (Tadesse *et al.* 2001). Large-scale deforestation of the montane rain forest, therefore, leads to loss of the wild Arabica coffee gene pool. Despite all these factors of threats on the forest coffee ecosystem, a number of wild coffee populations, which can be used as sources of germplasm, exist in rain forest of SW and eastern part of the Great Rift Valley system (Tadesse *et al.*, 2002). Many factors are contributing to the erosion of the gene pool from Ethiopia (Tadesse, 2003). Population increment in the areas leads to land-use change, which in turn leads to deforestation of the natural ecosystem of wild coffee. To avoid high rate of erosion of genetic variability collection, characterization and conservation of germplasm is highly important (Demel, 1999; Tadesse *et al.*, 2002).

1.4. Methods for conservation of Genetic Resources

Conserving and using plant genetic diversity is vital to meeting the world's future development needs (IPGRI, 1993, 2005). There are two main approaches for conservation of plant genetic resources, namely *in situ* and *ex situ* (Endashaw, 1986; Karp *et al.*, 1997; Karp, 2002; Rao, 2004). *In-situ* conservation involves maintaining genetic resources in the natural habitats where they occur, allowing evolution to continue and increasing genetic diversity to be conserved (Endashaw, 1986). *In situ* conservation is considered to be the method of choice for conserving forest species and wild crop relatives and there is an increasing interest in the use of *in-situ* conservation for crops themselves as on-farm conservation (Karp *et al.*, 1997; Karp, 2002). The conservation of wild species in genetic reserves, termed gene management zones (GMZs), involves the target species, location, designation, management and monitoring of genetic diversity in a particular natural location (Tan and Tan, 2002).

Ex-situ conservation on the other hand, involves conservation outside the native habitat and is generally used to safeguard populations in danger of destruction, replacement or deterioration (Endashaw, 1986; Rao, 2004). Approaches to *ex situ* conservation include methods like seed storage, field genebanks and botanical gardens. DNA and pollen storage also contribute indirectly to *ex situ* conservation of PGR. Among the various *ex situ* conservation methods, seed storage is the most convenient for long-term conservation of plant genetic resources. This involves desiccation of seeds to low moisture contents and storage at low temperatures. However, there are a large number of important tropical and sub-tropical tree species, which produce recalcitrant seeds that quickly lose viability and do not survive desiccation; hence conventional seed storage strategies are not possible. *C. arabica* is one of such crops. Such plants are conserved either *in*

situ or in field genebanks. Field genebanks are also important to conserve crop species that are sterile or do not easily produce seeds, or their seed is highly heterogenous and clonal propagation is preferred to conserve elite genotypes. Examples of crop species in this category are banana, sweet potato, sugarcane, cassava, yam, potato, and taro. Although field genebanks provide easy access to conserved material for use, they run the risk of destruction by natural calamities, pests and diseases (Endashaw, 1986).

There are a number of alternative methods to avoid the risks posed by using field genebanks (Rao, 2004). *In vitro* methods such as tissue culture techniques that help to establish safety duplicates of the living collections for conservation. *In vitro* conservation also offers advantages such as maintaining the material in a pathogen-tested state, so as to facilitate safer distribution and to protect the cultures from environmental disturbances.

Effective conservation of genetic resources requires the use of the combination of the methods (Endashaw, 1986; Karp *et al.*, 1997). There are many field genebanks in the world and in Ethiopia (FAO, 1968), which have collections of *C. arabica* from Ethiopia and other parts of the world (Tadesse, 2003). In Ethiopia there are two field gene banks in SW at Jimma and Chochie (Tadesse, 2003). There is a need to supplement these collections by maintaining some genetic resources of *C. arabica* in their native environment, *in-situ*.

1.5. Detecting Genetic Diversity in Genetic Resources

Genetic resources conservation and use activities, be it *in situ* or *ex situ*, require characterization of the diversity in both the gene pools and the gene banks (Karp *et al.*, 1997; Tan and Tan, 2002). Therefore, genetic diversity measurements are needed to designate the appropriate genetic reserve of target species with high variation. Diversity measurements are also needed to monitor any change in the pattern of diversity in the *in situ* GMZs.

Study of genetic diversity is the process by which variation among individuals or groups of individuals or populations is analysed by a specific method or a combination of methods (Haig, 1998; Karp, 2002; Rao, 2004). Diverse data sets, often involving numerical measurements, are used in genetic diversity studies of plants. The genetic diversity analysis may be made at different levels: individual genotypes such as inbred lines or pure lines or clones, populations, germplasm accessions, and species (Karp *et al.*, 1997). To conserve and use genetic variation it should first be assessed, i.e., the extent and distribution of the variability needs to be determined.

There are many methods by which genetic diversity could be assessed (Murphy *et al.*, 1996; Karp *et al.*, 1997; Parker *et al.*, 1998; Karp, 2002). Detection of genetic variation can be done with the help of a genetic marker. Genetic marker is a measurable character that can detect variation in either a protein or DNA sequence. A difference, whether phenotypic or genotypic, may act as a genetic marker if it identifies characteristics of an individual's genotype and/or phenotype, and if its inheritance can be followed through different generations. A genetic trait may not have necessarily observable consequences on an individual's performance. There are three major types

of genetic markers: morphological traits, protein/biochemical markers, and DNA-based/molecular markers.

Traditionally, diversity is assessed by measuring variation in phenotypic traits such as flower colour, growth habit or quantitative agronomic traits like yield potential, stress tolerance, etc., which are of direct interest to users (Karp *et al*, 1997; Amsalu, 2001;Karp, 2002). However, since these sets of characters lack adequate coverage of the genome, are strongly influenced by environmental factors, and are apparently controlled by several genes their use in genetic diversity studies is limited. In the 1960s, biochemical methods based on seed protein and enzyme electrophoresis were introduced, which proved particularly useful in analysis of genetic diversity as they reveal differences between seed storage proteins or enzymes encoded by different alleles at one (allozymes) or more gene loci (isozymes) (Murphy *et al*, 1996; Parker *et al*, 1998). Like morphological markers protein markers are also limited by being influenced by the environment and changes in different developmental stages (Murphy *et al.*, 1996; Rao, 2004). DNA-based techniques introduced over the past two decades have potential to identify polymorphisms represented by differences in DNA sequences (Dowling *et al.*, 1996; Karp *et al.*, 1997; Parker *et al.*, 1998; Wolfe and Liston, 1998; Karp, 2002). These methods are being used as complementary strategies to traditional approaches for assessment of genetic diversity, the major advantage being that they analyse the variation at the DNA level itself, excluding all environmental influences.

Whichever method is used the aim is to find out the level and distribution of genetic variability. The analysis of genetic relationships among samples or populations and estimates of population genetics parameters, particularly diversity and its partitioning at different levels are important analyses in genetic resources conservation (Karp *et al.*, 1997). Measures of genetic

similarity/distance (Nei, 1987) and cluster analysis show the relationship between samples or populations. Measurement of genetic diversity can be done by the proportion of polymorphic markers (P) and Nei gene diversity (H)(Nei, 1987) and also by Shannon-Weaver diversity index. The genetic structure (within and among populations) can be analyzed as G-statistics (Nei, 1973), F-statistics of Sewal Wright (Wright, 1951 cited in Nei, 1987).

~~Molecular techniques proved useful in a number of ways to improve the conservation and management of PGR (Karp et al., 1997). In particular, genetic diversity data provides information on gaps in terms of coverage in gene pools as well as redundancies, i.e., material with similar characteristics that wastes resources through increased cost of management (Rao, 2004).~~

2. Literature Reviews

2.1. Taxonomy, Reproductive biology and Morphology of *C. arabica*

2.1.1. Taxonomy

Taxonomically, *C. arabica* belongs to genus *Coffea* L., which in turn belongs to a large family of flowering plants, Rubiaceae (Wellman, 1961; Purseglove, 1968; Charrier and Berthaud, 1985). The classification was made by Carlos Linnaeus in 1737 using the then known single species *C. arabica* putting it as a separate genus *Coffea* L. However, since the second half of the nineteenth century many new species of *Coffea* have been discovered by exploration of tropical forests of Africa, where most of the species are found (Charrier and Berthaud, 1985).

The taxonomy of the genus is still unsettled and many new species are being discovered. Many new species of the genus have been discovered in recent years in the forests of eastern Africa (Charrier and Berthud, 1985) and Madagascar. The number of species described under the genus

varies from author to author, ranging from 60 to 100 (Bridson and Verdcourt, 1988; Coste, 1992) all of which are found in the tropics of Africa, Madagascar and Mascarnes. Although the taxonomy is not settled, the genus is classified into sections and subgenera by different authors based on different criteria. For example, Chevalier, (1947, cited in Charrier and Berthaud, 1985) classified the genus into four sections namely *Argocoffea*, *Paracoffea*, *Mascarocoffea* and *Eucoffea*. He also further classified the sections into subsections. Sections *Eucoffea* (more correctly *Coffea*) and *Mascarocoffea* include most of the presently known coffee species. Leroy (1980, cited in Tadesse, 2003) classified the genus into three subgenera as subgenus *Coffea*, subgenus *Baracoffea* and a monotypic subgenus *Psilanthopsis* based on growth habit (monopodial vs. sympodial) and type of inflorescence, axillary vs. terminal. The first two sections, *Argocoffea* and *Paracoffea*, are excluded from the genus by Leroy (1967, cited in Charrier and Berthaud, 1985). Recently Bridson (1994, cited in Tadesse, 2003) recommends that *Psilanthopsis* be placed in subgenus *Coffea*.

All the caffeine-containing coffee trees belong to subgenus *Eucoffea* that probably originated from central and east African centers and differentiated into different species such as *C. arabica* from Ethiopia, *C. eugenioides* from Kenya and *C. canephora* as well as *C. liberica* from West Africa. Some other species like *C. congensis* and *C. excelsa* come from the Central African Republic (Berthou, 1983; Lashermes *et al.*, 1997).

Within *C. arabica* there are several phenotypically distinct populations and lines of cultivars or varieties (Var. *bourboun* Choussy and *typica* Cramer) and mutations from these varieties (Coste, 1992). From Ethiopia many landraces or farmer's cultivars, about 130 in number, have recently been recorded from the different coffee growing regions (Demel, 1999; Tadesse *et al.*, 2001).

2.1.2. Morphology and reproductive biology

Arabica coffee (*C. arabica*) is a shrub or small tree, and if untended, it may reach a size of 4 to 5 meters (Purseglove, 1968; Van der Graff, 1981, Coste, 1992; Demel, 1999). The plant has a dimorphic habit of branching in which vertical (orthotropic) branches form horizontal (plagiotropic) branches, which bear the flowers and the fruits in clusters (Figure 1). Flowers are with short corolla, long style and exerted stamen and are typical of the genus *Coffea*. Inflorescences develop from serial buds mainly on horizontal branches. Each inflorescence normally carries one to five flowers. The flowers have a short pedicel and a rudimentary calyx. The petals are fused and form corolla with five lobes. The pistil consists of an inferior ovary and a long style with two stigmatic lobes. The ovary is bilocular each with one anatropous ovule. Flower initiation occurs after sufficient rainfall following a dry period. The total period of flowering is normally not more than three days with the majority of flowers opening on the first and the second day. Pollen shedding starts very soon after opening of the flowers early in the morning and the stigma is then receptive. Flowers wither in one or two days after pollination. It takes six to eight months from flowering to fruit ripening.





Figure 1. Pictures showing some morphological features of *C. arabica* and cross-section of its fruits. Picture A shows coffee tree with bronze-tipped leaves and B shows the branching pattern, the ripe and unripe berries and the main stem. Picture C shows the cross-section of the fruit (Esayas, 2006)

The coffee fruit usually contains two seeds. Ripe fruits have a thick fleshy mesocarp (pulp) and a hard endocarp (parchment). In addition, each seed is enveloped in a silver skin (testa), which is a remnant of the integument (perisperm). The tough endocarp is to protect the seed from digesting enzyme activities in the gut of frugivores such as birds and mammals. The fleshy, sugar containing mesocarp and the vivid coloration due to anthocyanins of the exocarp act as a reward and attract the dispersing animals, respectively (Urbaneja *et al.*, 1996, cited in Esayas, 2006).

The coffee bean consists of an endosperm and small embryo embedded at the basal end of the seed. There is no seed dormancy in coffee, and seed viability is normally lost within three to six months after harvesting. The coffee seed is recalcitrant, but it is possible to preserve the viability for up to two and half years when coffee seeds at moisture content of 41% is stored at 15° C (Purseglove, 1968).

The floral morphology of *C. arabica* would permit natural cross-pollination. Nevertheless, *C. arabica* is largely autogamous (self-fertile), and fruit set after self-pollination is 60% or higher (Carvalho *et al.*, 1969). Most studies on the degree of natural cross-pollination were carried out on cultivars of *C. arabica*, which underwent many cycles of selection. Van der Vossen (1974) found 7 to 15 percent of natural cross-pollination in *C. arabica*. The rate of out-crossing is expected to be higher in its centre of origin than in the cultivars growing elsewhere, since natural pollinating agents that evolved with the species are found in the natural habitats. About 40-60% self-fertility was observed in wild coffee plants in SW Ethiopia (Meyer, 1965) indicating the possibility of higher out-crossing rate. Most diploid species of *Coffea* have proved to be highly self-incompatible, allogamous (Berthaud and Charrier, 1985).

The genome of *C. arabica* is tetraploid ($2n=4x=44$). Recent investigations established that *C. arabica* has arisen by natural hybridization between two closely related diploid species followed by unreduced gamete formation (Charrier and Berthud, 1985; Anthony *et al.*, 2002). The suggested putative ancestral parents are *C. canephora* and *C. eugenioides* (Lashermes *et al.*, 1999) or *C. liberica* or *C. congensis*, or species very closely related or identical to *C. eugenioides* and to *C. canephora* (or *C. congensis*) are the most likely maternal and paternal progenitors of *C. arabica*, respectively (Lashermes *et al.*, 1995; Raina *et al.*, 1998). Recently, sequencing of regions of chloroplast genome by Kassahun *et al.* (inpress) suggests the probability of *C. eugenioides* or its ancestor to be the maternal parent of *C. arabica*.

2. 2. Coffee in Ethiopia

2.2.1. Origin and distribution of C.arabica

Ethiopia has diverse climatic conditions varying from hot and dry deserts in the lowland areas, up 116 meters below sea level (mbsl), to humid alpine habitats in highlands, which rise to over 4000 metres above sea levels (masl). There exists high diversity of species of plants and animals because of the climatic diversity (Demel, 1999; Paulos and Demel, 2000). The montane rain forest is one among the very important homes for endemic and indigenous plants and animals (Tadesse, 2003). Ethiopia is also the center of origin of several cultivated crops such as tef (*Eragrostis tef*), niger seed (*Guizotia abyssinica*), and *C. arabica* (Demel, 1999). These crops are largely represented in the country by landraces and wild types.

All coffee species are believed to have been originated from Africa and the cultivated ones, *C. arabica*, *C. canephora* and *C. liberica*, were also domesticated in Africa for the first time (Wellman, 1961; Purseglove, 1968; Demel, 1999). Ethiopia is widely believed to have been the center of origin and proliferation of *C. arabica* (Wellman, 1961; Haarer, 1962; Meyer, 1965; Purseglove, 1968; Friis, 1979). It occurs naturally in forests of Ethiopian highlands of SW Ethiopia between altitudes of 1400 and 1900 masl (Purseglove, 1968). It is also native to the Boma plateau of Sudan and Mount Marsabit in N. Kenya (Friis, 1979; Anthony *et al.*, 2001).

Today the species is cultivated in most of the tropical countries of the world and it is distributed to these regions via Arabia, the present Yemen Republic (Wellman, 1961; Friis, 1979). Arabica coffee is being distributed in South and Central America, Caribbean Islands in the pacific, India, Indonesia, Viet Nam, New Guinea, the Philippines, high land areas of Africa, Kenya, Tanzania,

Zimbabwe, Rwanda, Burundi, Zaire, Angola and Cameroon (Coste, 1992). It is not clearly known when Arabica coffee was taken from Ethiopia to Arabia. Some authors consider the first introduction to be at about A.D. 575 (Wellman, 1961), but Haarer (1962) states the absence of reputable evidence of coffee in Arabia during the 13th century. However, it is clear that it has been cultivated in Yemen for at least five centuries, since the 15th century when the discovery of brewing coffee was made (Purselove, 1968).

The Dutch introduced variety *typica* now called *C.arabica var. arabica* into Java in 1696 and subsequently in 1699 and spread to South East Asia about 1700(Wellman, 1961; Friis, 1979; Demel, 1999). A single plant was taken from Java to the Amsterdam Botanic Garden in 1706 whose vigorous-progeny was taken to Paris in 1713. From Amsterdam the progenies were taken to Surinam in 1718, then to Cayenne in 1722 and then to Brazil in 1727. Progeny from Paris plant were sent to Martinique about 1720 from where it was taken to Jamaica in ten years later, where it gave rise to the cultivar 'Blue Mountain'. From these two sites, Amsterdam and Paris, the variety spread to Caribbean and Central and South America. Originated from a single tree, *typica* coffee, has a low genetic diversity. Thus the Latin America has a very low genetic variability (Lashermes, 1996). Another variety commonly called 'Bourbon', which occurs sub spontaneously in Ethiopia, appears to have been taken by the French to Bourbon Islands (now R'eunion) about 1718 from unknown source (Friis, 1979). Progenies from this introduction were taken to the New world and other parts of the world including Africa.

2.2.2. Coffee production in Ethiopia

Arabica coffee (*C. arabica*) is the only coffee species that exists naturally in Ethiopia and it is geographically isolated from the rest of coffee species (Tadesse *et al.*, 2002). There are wild and cultivated types of *C. arabica* in Ethiopia grown generally under four different production systems: forest, semi-forest, garden and plantation (Demel, 1999; Workafes and Kassu, 1999; Tadesse *et al.*, 2002). In the forest production system, coffee is harvested directly from wild coffee plants growing in afro-montane rain forests of west and southwest of Ethiopia by subsistence farmers. Wild populations of *C. arabica* are under the forest production system and grow naturally in the undergrowth of the montane rain forests in southwest and Eastern part of Ethiopia (Tadesse *et al.* 2001). These forests are part of a natural forest ecosystem and represent 9% of the total land covered by coffee and the production system contributes about 5-6% of the total coffee production in the country (Tadesse *et al.* 2002). There is no any kind of management of the forest and the coffee other than harvesting the berries.

In the semi-forest coffee production system, management of forest such as thinning out of trees and shrubs competing with coffee in the lower storey as well as large trees in the canopy layer and also planting of coffee seedlings either from the natural forest or raised in nurseries are possible. This production system represents 24% of the total land covered by coffee and contributes about 20% of the total coffee production in the country. These managements can affect the genetic diversity structure of the forest coffee gene pool. Both the forest and semi-forest coffee production systems are part of the forest coffee ecosystem. The two production systems, forest and semi-forest, together account for about 33% of land covered by coffee and 25-26% of the total coffee production in Ethiopia. Plantation and garden coffee production

systems together contribute to the remaining percentage of production and are mainly found in the south and eastern parts of the country. They are produced on open lands and higher level of managements is carried out.

The forest ecosystems of the SW and the south parts of Ethiopia which harbor the wild coffee are disappearing at an alarming rate due to deforestation and land use change (Paulos and Demel, 2000; Tadesse *et al.*, 2001). Deforestation is estimated at 10,000 ha/year in the coffee growing areas in the southwest Ethiopia (Tadesse *et al.*, 2002). There are only few fragmented natural forest areas remaining with wild coffee in Ethiopia. The following areas are important wild coffee areas in Ethiopia (Demel, 1999; Paulos and Demel, 2000; Tadesse, 2003): (1) Geba-Dogi/Yayu (2) Boginda-Yeba, (3) Berhane-Kontir, (4) Amora Gedel, (5) Dawo Tobi, (6) Mankera, (7) Maji, and (8) Bale Mountains.

2.3. Genetic diversity of *C.arabica*

The world coffee industry is based on a very narrow genetic base of *C. arabica* as it is originated from few base-line individuals. These are represented by the two base line varieties, typica and Bourbon, and some cultivars and mutations derived from them (Anthony *et al.*, 2001). Currently, three cultivars are grown worldwide: cv. Mundo Novo, a Bourbon and Typica hybrid, cv. Caturra, a Bourbon mutant with short internode length, and cv. Catuaia Caturra and Mundo Novo hybrid. These cultivars are high yielding and produce a high quality beverage. However, breeding programs are limited by the very narrow genetic base, especially in improving pest and disease resistance (Van der Graff, 1981). The narrow genetic base can lead to production and economic loss to a country. For example, Sri Lanka was forced to totally abandon coffee production in 1869 because of production loss by coffee leaf rust disease.

A genetic alternative to breeding limitations exists in the spontaneous and subsponaneous genotypes collected in the Ethiopian highlands (Mesfin, 1991; Lashermes *et al.*, 1996). Many biologists who visited Ethiopia in the 1960s and before observed the presence of high phenotypic variability among the wild coffee populations and landraces (Sylvain, 1955; Meyer, 1965; van der Graff, 1981). The wide genetic diversity in Ethiopia warrants the availability of large gene pools having desired characteristics for improvement of the crop and disease and pest resistance (Demel, 1999; Paulos and Demel, 2000). The diversity is represented by different forms of landraces and wild types that are uniquely adapted to different agro-ecologies and resistant to coffee diseases and pests. There are many varieties that are resistant to coffee berry disease (CBD) and other diseases and pests (Van der Graff, 1981). Although there was greater production loss by CBD in the years from 1971-78, coffee production did not stop due to the presence of resistant varieties (Mesfin and Bayetta, 1984; Mesfin, 1991).

The existence of wide genetic diversity of *C. arabica* in Ethiopia, as compared to cultivars, was confirmed by both phenotypic/morphological and recently molecular markers. However, most of these studies were carried out on materials collected and stored in field genebanks out of Ethiopia (Carvalho *et al.*, 1969; Orozco-Castillo *et al.*, 1994; Lashermes *et al.*, 1996; Montagnon and Bouharmont, 1996; Anthony *et al.*, 2001, 2002). Wild and cultivated coffee from Ethiopia was classified according to their geographic origin (i.e. southwestern and south/southeastern Ethiopia) using agro-morphological observations (Montagnon and Bouharmont, 1996). The two base line varieties for the cultivated materials are also morphologically distinct from each other. Bourbon-derived cultivars are characterized by a more compact and upright growth habit, higher yield and better cup quality than the typica-derived cultivars (Carvalho *et al.*, 1969).

Some recent investigations on materials in field genebanks, using molecular markers, showed that accessions from SW Ethiopia have greater genetic variability than cultivars (Orozco-Castillo *et al.*, 1994; Lashermes *et al.*, 1996; Anthony *et al.*, 2001, 2002). Polymorphism among the subsponaneous accessions was much higher than among the cultivated accessions (Anthony *et al.*, 2001). The polymorphism was very low within the genetic bases of cultivars, confirming the historical documentation on their dissemination. The results enabled a discussion of the genetic diversity reductions that successively occurred during the dissemination of *C. arabica* from its primary center of diversify, SW Ethiopia (Anthony *et al.*, 2002). The schematic representation of the Ethiopian and the cultivars is shown in Figure 2. Wider genetic variability exists in Ethiopian materials as compared to the world wide cultivated cultivars.

In general it is confirmed that *C. arabica*, as compared to other coffee species, is characterized by a low genetic diversity. This is attributed to the recent evolution of the species by allopolyploidy and its predominantly self pollinating nature (Lashermes *et al.* 1996). Restriction fragment polymorphism (RFLP) analyses (Lashermes *et al.* 1996) and Random amplified fragment polymorphism (RAPD) analysis (Lashermes *et al.*, 1993; Anthony *et al.*, 2001) indicate a rather low levels of polymorphism among *C. arabica* accessions. However, genetic diversity, as expressed by the number of markers detected, and polymorphism appeared

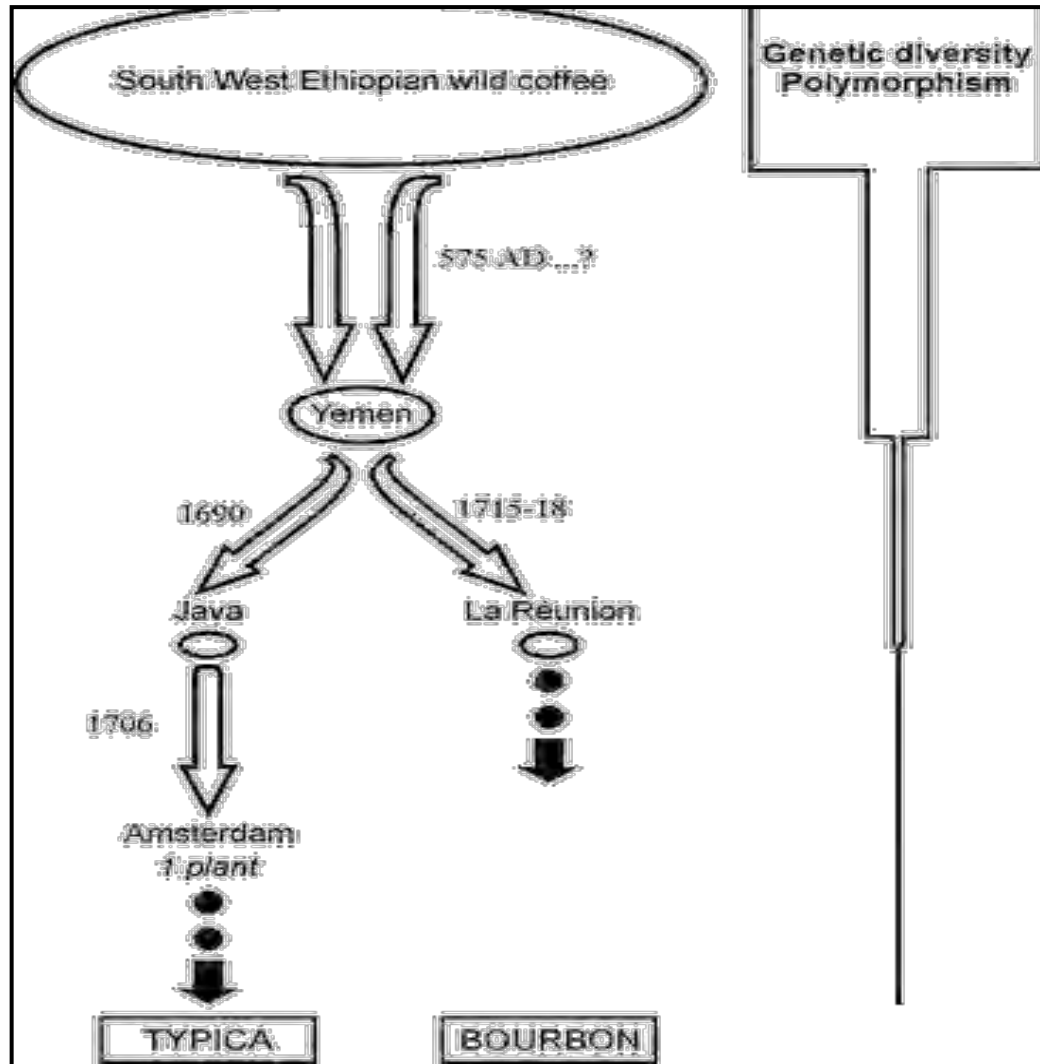


Figure 2. [Schematic representation of the main steps in the history of coffee cultivation and the genetic diversity of Arabica coffee population of southwest Ethiopia, compared to cultivated population abroad \(Anthony *et al.*, 2002\)](#)

[to be higher among accessions derived from collections of sub-spontaneous \(forest coffee ecosystem\) from Ethiopia as compared to Bourbon and Typica-derived accessions and Yemen cultivars \(Anthony *et. al.*, 2001\).](#) But these analyses did not detect within collection polymorphism. High level of polymorphism among [subspontaneous materials of Ethiopia](#) was also confirmed by AFLP and SSR analyses (Anthony *et al.*, 2002).

AFLP markers were used to assess polymorphism between and within Typica- and Bourbon-derived accessions and Yemen cultivars (Anthony *et al.*, 2002). Their relationship with wild material was defined analyzing accessions derived from spontaneous trees collected in the center of origin of the species. Microsatellite loci were used to confirm the structure of the genetic diversity detected by the AFLP markers and to compare the polymorphism within the genetic groups. The results indicate that wild materials have larger polymorphism followed by Yemen-derived cultivars and Typica- and Bourbon-derived cultivars have the lowest genetic diversity as seen in Figure 2.

Recent molecular studies of the genetic variability among wild coffee in Ethiopia indicated the existence of moderate level of polymorphism among and within populations of wild coffee (Esayas, 2006; Kassahun *et al.*, in press). Inter-simple sequence repeat (ISSR) marker separated wild materials from their respective landraces with the wild materials having larger variability (Kassahun *et al.*, in prep.). These two studies using molecular markers such as RAPD (Essays *et al.*, 2003), ISSR (Esayas, *et al.*, 2005; Kassahun *et al.*, in prep), ISTR (Esayas *et al.*, 2005) and SSRs (Esayas, 2006) are the first genetic diversity studies of wild coffee in Ethiopia. These studies found moderate amount of genetic diversity within and between wild *C. arabica* populations from different geographic regions (interregional analysis) in the country. The results also indicated high levels of intraregional diversity and coffee plants of a single plot are usually related closest, indicating hierarchical-geographical patterns. In general, intraregional (e.g. within Yayu) genetic diversity was found to be high, compared to interregional differences.

Although the above studies showed higher diversity within regions than between regions, only limited numbers of samples per region was used (less than 15 per region). To know the amount of diversity and its distribution between different plots of a region, the use of large number of samples from different plots is required. Such detailed analysis, using efficient marker systems, at a regional stage is a crucial step for *in situ* conservation and breeding purposes and also to manage the genetic diversity. Therefore, the present study aims at conducting the analysis among wild *C. arabica* in one of the CoCE regions: Yayu(Geba-Dogi) forest.

3. ~~DNA-based~~ ~~MM~~molecular markers

With the development of new molecular biology techniques such as restriction digestion, polymerase chain reaction (PCR), southern blots and hybridization and others, to manipulate DNA many DNA-based molecular markers have been and are being developed (Karp *et al.*, 1997; Parker *et al.*, 1998; Wolfe and Liston, 1998; Karp, 2002). DNA polymorphisms can be detected in nuclear and organellar DNA, which is found in mitochondria and chloroplasts (Parker *et al.*, 1998). Molecular markers concern the DNA molecule itself and as such, are considered to be objective measures of variation. The diverse arrays of molecular markers have become available for high-resolution (even to the level of detecting single nucleotide mutations) genetic studies of population-level processes. DNA-based markers have several advantages over protein-based markers (Dowling *et al.*, 1996; parker *et al.*, 1998). For example, one can find DNA in nearly all cells of all organisms and can recover it from both dead and living tissues, tissues can be easily stored under field conditions, and in most cases very small amount of DNA (in nanograms) can be used (Parker *et al.*, 1998). Furthermore, it covers the whole genome including both coding and

noncoding regions. In addition, the methods are, for the most part, general to any type of DNA and genotype rather than phenotype is assayed (Dowling *et al.*, 1996).

Many different types of DNA molecular markers with different properties exist, each with its own advantages and disadvantages. However, the following features characterize a good molecular marker (Karp *et al.*, 1997):

1. **Polymorphic**- a molecular marker has to be variable among individuals. The degree of polymorphism detected depends on the technology used to measure it. It also depends on stores of genetic variability maintained by several factors.
2. **Reproducible**- a molecular marker needs to be reproducible in any laboratory experiment, whether within experimental events in the same laboratory or between different laboratories performing identical experiments.
3. **Codominant**. Depending on the type of application, the selected technology must be able to distinguish between homozygotes and heterozygotes, i.e., it has to be codominantly inherited. A heterozygous individual shows simultaneously the combined genotype of the two homozygous parents. Some markers such as RAPD, ISSR and AFLP are dominantly inherited and do not distinguish between homozygotes and hererozygotes (Table 1).
4. **Wide genome coverage and even distribution throughout the genome**. The more distributed and dense genome coverage is, the better the assessment of polymorphism.
5. **High discriminating power**- a good genetic marker has to be able to detect differences between closely related individuals.
6. **Not subject to environmental influences**. Unlike the case of protein markers a DNA marker should not be influenced by environment and the developmental stage of the individuals under investigation.

7. **Neutral.** The allele present at the marker locus is independent of, and has no effect on, the selection pressure exerted on the individual. This is usually an assumption, because no data are usually available to confirm or deny this property.
8. **Inexpensive.** Easy, fast and cheap in detecting across numerous individuals. If possible, the equipment should be of multipurpose use in the experiment.

In general, the most appropriate molecular marker is the one that show high-resolution i.e., the one that produces a large number of alleles at a single locus and/or many loci with two or more common alleles are better, with a minimum amount of effort and expense. However, there is no single molecular marker that fulfils the above requirements. Thus many authors suggest the use of more than one type of molecular markers in a single experiment (Karp *et al.*, 1997).

The choice of a method depends on the level of genetic variation required for a particular study and the amount of effort and expense required (Parker *et al.*, 1998). Each molecular marker has its advantages and drawbacks. Therefore, application of molecular marker techniques to diversity questions must take into account whether or not the data derived from a technique provide the right type of information for answering the question being addressed (Karp *et al.*, 1997; Karp, 2002). This in turn depends on the taxonomic levels of the material being studied (different species, subspecies, populations, cultivars and individuals). The closer the relationship of the materials to be studied, the more necessary it may be to consider highly discriminatory techniques. Besides, the choice of appropriate molecular markers also depends on the accessibility and cost effectiveness of the marker techniques. Depending on whether PCR is involved or not molecular markers can be divided into two groups: PCR-based and non-PCR based techniques.

3.1. Non-PCR based markers

Plant DNA polymorphism assays include sequencing of a known region of a genome; non-PCR-based DNA and PCR-based DNA markers (Dowling *et al.*, 1996; Karp *et al.*, 1997). The two main non-PCR based technologies are Restriction Fragment Length Polymorphism (RFLP) and Variable number of Tandem Repeats (VNTRs). In restriction fragment length polymorphism (RFLP) analysis, the DNA is digested with restriction enzymes and the resultant fragments are separated by gel electrophoresis (Dowling *et al.*, 1996; Parker *et al.*, 1998). The restricted DNA fragments are then transferred to a filter by Southern Blotting (Karp *et al.*, 1997). Specific

Table 1. [Comparison of the major techniques for identification of plant genetic diversity. Proteins \(Isozymes\) are also included for comparison](#)

Markers	Number of loci	Codominant	Polymorphism	Locus specificity	Technicality	Cost
AFLP	Unlimited	No	High	No	Medium	Medium
Isozymes	<90	Yes	Low	Yes	Low	Low
ISSR	Unlimited	No	High	Yes	Low/Medium	Low/Medium
RAPD	Unlimited	No	Medium	No	Low	Low
RFLP	Unlimited	Yes	Medium	Yes	High	Medium
Sequencing	Unlimited	Yes	High	Yes/No	High	High
SNP	Unlimited	Yes	Very High	Yes	High	High
SSR	Unlimited	Yes	Very High	Yes	Low^a	Low^a

^a [When microsatellites have already been identified and primers designed.](#)

[Taken and modified from a module prepared by IPGRI and Cornell University \(2003\).](#)

probe/enzyme combinations will give highly reproducible restriction fragment patterns for a given individual, but variation between individuals can arise when mutations alter the sequence in the restriction sites, thus preventing the enzyme from cutting, or the DNA sequence in the fragment lengths between them, by the creation of new restriction sites, or the insertion or deletion of base pairs resulting in an alteration of the fragment length between the sites (Karp *et al.*, 1997). RFLPs are highly reproducible and inherited in a codominant manner (Table .1). The main disadvantages of the RFLP analysis include finding probes that can detect polymorphisms at the cultivar or population level in some species, RFLPs are time-consuming and they are not easy to automate. RFLPs require high quantities of good quality DNA, e.g. 10 µg per digestion (Devarumath *et al.*, 2002).

Variable number of tandem repeats (VNTR) are hypervariable, tandemly repeated regions of DNA interspersing the genome of higher organisms (Tautz, 1989; Parker *et al.*, 1998). There are two classes of VNTRs: microsatellites, or simple sequence repeats (SSRs), where the basic repeat

unit is around 2-8 base pairs in length, and minisatellites for longer repeat units of around 16-100 base pairs (Wolfe and Liston, 1998). The length of the basic units differs from author to author. Hybridization to restricted DNA with micro- or minisatellite probes gives multilocus patterns, which can resolve variation at the levels of populations and individuals (Dowling *et al.*, 1996). The variation results from changes in the number of copies of the basic repeat and is often referred to as Variable Number of Tandem Repeats (VNTRs). Both of them have high rate of mutation, 1/100 per gamete per generation (Dowling *et al.*, 1996).

3.2. PCR-based DNA markers

With the development of the polymerase chain reaction (PCR), many PCR-based DNA molecular techniques have been, and still are being developed for plant genome analysis (Weber and May, 1989; Dowling *et al.*, 1996; Karp *et al.*, 1997; Parker *et al.*, 1998; Wolfe and Liston, 1998; Karp, 2002; Rao, 2004). The techniques could be categorized into two main groups: arbitrary (or semi-arbitrary) primed techniques and site-targeted PCR techniques (Karp *et al.*, 1997). Techniques in the first category use single primers which are designed arbitrarily/or semi-arbitrarily, i.e., without the knowledge of flanking sequence of the region to be amplified. These techniques include Random Amplified Polymorphic DNA (RAPD) (Welsh and McClelland, 1990; Williams *et al.*, 1990), AFLP (amplified fragment length polymorphism) (Vos *et al.*, 1995), ISSR (inter-simple sequence repeat) (Zietkiewicz *et al.*, 1994). These are also called multi-locus approaches.

Techniques in the second category depend on primers that target a single known site, such as a gene (Karp *et al.*, 1997; Wolfe and Liston, 1998). These are alternative approaches to multi-locus profiling. PCR-DNA sequencing, sequence-tagged microsatellite (STMs), thermal gradient gel electrophoresis (TGGE), denaturing gradient gel electrophoresis (DGGE), cleaved amplified

polymorphic sites (CAPS), single-strand conformational polymorphisms (SSCP), are important examples of these category (Dowling *et al.*, 1996; Karp *et al.*, 1997; Wolfe and Liston, 1998). Since, they give information on a single locus, they are particularly important when information is required on gene frequency or genealogical information for genetic diversity management and when information on heterozygosity is required (Karp, 2002).

3. 3. Inter-simple sequence repeats (ISSRs) Markers

Inter-simple sequence repeat (ISSR) is a ~~an alternative~~ technique to study polymorphism based on the presence of microsatellites throughout genomes (Zietkiewicz *et al.* 1994; Ramesh *et al.*, 1995; Nagaoka and Ogiwara, 1997; Wolfe and Liston, 1998; Devarumath *et al.*, 2002). ISSRs are regions found between Simple Sequence Repeats (SSRs) or microsatellite repeats. Microsatellites are present in both coding and non-coding regions and are usually characterized by a high degree of length polymorphism (Tautz, 1989). The technique is based on PCR amplification of DNA segments present at an amplifiable distance between two identical microsatellite repeat regions that are oriented in opposite direction by using a single primer (Wolfe and Liston, 1998) (Figure-3). ISSR-PCR bands are very reproducible, as compared to RAPDs—, abundant and polymorphic in plant genomes.

The primers are usually 16–25 base pair long and are developed from within the SSRs themselves with or without a 5'-or- 3' anchor sequence [e.g., 5'-AA (CG)₆-3', 5'-(AGC)₄-GC-3']. The anchors are useful to guarantee locus-specificity to the inter-repeat region only. The microsatellite repeat motifs used as primers can be di-nucleotide, tri-nucleotide, tetra-nucleotide or penta-nucleotide. Because of the known abundance of repeat sequences spread all over the

genome, this single primer PCR reaction targets multiple genomic loci to amplify mainly the inter-SSR sequences of different sizes (Figure 32).

The high reproducibility of ISSR markers is possibly due to the use of longer primers as compared to RAPD, which permits the subsequent use of a high annealing temperature (45–60° C) leading to higher stringency. The amplified products are usually 200-2000 bp long and amenable to detection by both agarose and polyacrylamide gel electrophoresis and can be detected with ethidium bromide, silver staining and/or radioisotopes.

The sources of variation in ISSR markers could be: (1) mutations at the priming site (SSR), which could prevent amplification of a fragment as in RAPD markers and thus give a presence/absence polymorphism; (2) an insertion/deletion event within the SSR region or the amplified region would result in the absence of a product or length polymorphism depending on the amplifiability of the resulting fragment size (Zietkiewicz *et al.* 1994; Wolfe and Liston, 1998).

The ISSR marker technique has several advantages and disadvantages. Some of the advantages, in addition to reproducibility as mentioned above, include relative technical simplicity and quick typing because it does not require prior sequence information, identification of significant level of variation, variation within unique regions of the genome may be found at several loci simultaneously, microsatellite sequence specificity and its very useful for DNA profiling. As a result of these advantages and their universality and easiness of development (no need to sequence data), ISSR markers are more and more requested (Abbot, 2001; Bernet and Branchard, 2004; Sudupak, 2004). ISSR markers were

shown to be efficiently used to study genetic differentiation of the *Coffea* species and to identify the parentage of *Coffea* interspecific hybrids

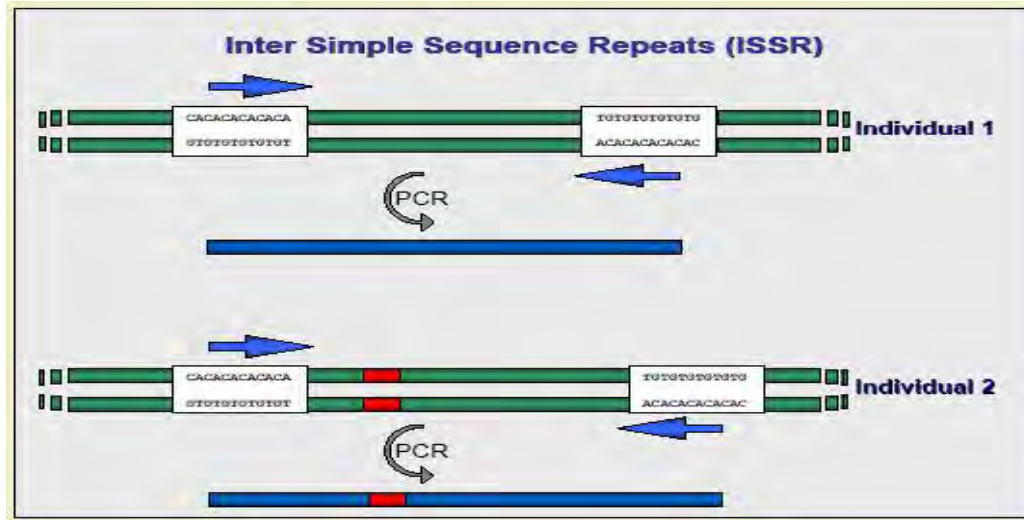


Figure 3. DNA fragments flanked by SSRs (individual 1 and 2) are amplified using a SSR primer (Green arrow). Length polymorphisms that are caused by insertions or deletions (red mark) in different individuals lead to PCR fragments (blue lines) varying in size

(Ruas *et al.*, 2003) and to study their genetic diversity (Esayas *et al.*, 2005; Kassahun *et al.*, in press).

The technique is not without limitations. For instance, there is the possibility, as in RAPD, that fragments with the same mobility may originate from non-homologous regions, which can contribute to some distortion in the estimates of genetic similarities (Sanchez *et al.*, 1996). In addition since the marker is dominant heterozygote cannot be distinguished from homozygote. The use of polyacrylamide gel electrophoresis and detection with silver staining or radioisotopes are also disadvantages.

23

~~Therefore, the present study used one of the molecular techniques known as inter-simple sequence repeat (ISSR), discussed in other sections, to assess the genetic variability and structure of four populations of forest *C. arabica L.* in~~

~~Yayu (Geba_Dogi) forest. Yayu is one of the few natural forests remaining with wild *C.arabica L.* populations in the SW Ethiopia proposed for *in situ* coffee gene reserve (Demel et al., 1998; Gole, 2003). These proposals were made on the basis of some preliminary surveys of the area (Demel et al., 1998) and on the basis of vegetation ecology studies (Gole, 2003). Hitherto, there are no detailed studies carried out on the genetic diversity and structure of the coffee populations in the forest using molecular markers. The present study was aimed to cover this gap to some extent.~~

~~*2. Objective of the Study*~~

~~*2.1. General objectives*~~

~~To design effective conservation strategy and thereby to identify sites with high diversity estimate, prior characterization of the existing wild coffee populations is important. The identification of these sites has paramount importance for germplasm collection mission and appropriate *in-situ* site selection.~~

~~In general, the objective of this study was to characterize the genetic variation that exists within and among the coffee populations in Yayu forest and based on the results suggest areas of priority for conservation (both *in-situ* and *ex-situ*) and use of the genetic resources.~~

~~The knowledge of the genetic diversity and relationship available in *Coffea* collections is important for planning breeding strategies, identification of interspecific hybrids, and for germplasm conservation (Ruas et al., 2003).~~

~~2.2. Specific Objectives~~

~~The specific objectives of this study was:~~

- ~~1. To investigate the genetic diversity within and among populations of *C. Arabica L.* in Yayu forest and how the diversity is distributed.~~
- ~~2. To generate information for *ex-situ* conservation collection and target wild *C. arabica* populations for *in-situ* conservation.~~
- ~~3. To generate information for further improvement of *C. arabica L.*~~

- ~~1. to identify populations/plots (four in number) for characterization of genetic variability.~~
- ~~2. to collect leaf samples from about twenty-five individuals for DNA extraction.~~
- ~~3. to isolate DNA from each sample and analyze with simple sequence repeat (SSR) primers using Polymerase chain reaction(PCR) for further running on gel.~~
- ~~4. to score the ISSR fragments and analyze population genetic diversity.~~
- ~~5. to compare individual level and population level genetic diversity from the results.~~
- ~~6. to suggest conservationists on areas of priority for management of coffee gene pool in the forest.~~

~~1.2. Literature Reviews~~

~~232323~~

~~1.232. 1. Ethiopia as a Centre of Origin and Diversification of *C. Arabica* 23.L.~~

~~Ethiopia is the center of origin of several cultivated crops such as sorghum, tef, noog, chickpea, and *C. arabica* L. largely represented in the country by landraces and wild types that are uniquely~~

adapted to, genetically diverse forms of these crops. Ethiopia is widely believed to have been the center of origin and proliferation of *C. arabica* L. (Wellman, 1961; Meyer, 1965; Purseglove, 1968; Friis, 1979). It occurs naturally in forests of Ethiopian highlands of SW Ethiopia between 4,500-6,000fts (Purseglove, 1968) and also native to the Boma plateau of Sudan and Mount Marsabit in N. Kenya (Friis, 1979; Anthony et al., 2001). Today the species is cultivated in most of the tropical countries of the world and it is distributed to these regions via Arabia (Currently Yemen). It is not clearly known when Arabica coffee is taken from Ethiopia to Arabia. Wellman (1961) considers the first introduction was about A.D. 575, but Haarer (1962) cited in Purseglove (1968) states the absence of reputable evidence of coffee in Arabia during the 13th century. However, it is clear that it has been cultivated in Yemen for at least five centuries, since the 15th century when the discovery of brewing coffee was made (Purseglove, (1968).

The Dutch introduced variety *typica* now called *C.arabica var. arabica* into Java in 1690 and subsequently in 1699 and spread to South East Asia about 1700(Wellman, 1961). A single plant was taken from Java to the Amsterdam Botanic Garden in 1706 whose vigorous progeny was taken to Paris in 1713. From Amsterdam the progenies were taken to Surinam in 1718, then to Cayenne in 1722(by the French) and then to Brazil in 1727. Progeny from Paris plant were sent to Martinique about 1720 from where it was taken to Jamaica in ten years later, where it gave rise to the cultivar 'Blue Mountain'. From these two sites (Amsterdam and Paris) the variety spread to Caribbean and Central and South America. Originated from a single tree, 'typica' coffee, which still predominate the Latin America, has a very low genetic variability.

The variety was taken also to India and Ceylon (towards the end of the 17th century), Philippines in 1740, and to Hawaii in 1825. The French also took it to their African territories, such as Uganda in 1900.

Another variety commonly called 'Bourbon', which occurs sub-spontaneously in Ethiopia, appears to have been taken by the French to Bourbon Islands (now Reunion) about 1718 from unknown source. Progenies from this introduction were taken to the New world and other parts of the world including Africa.

23

Ethiopia has got wide genetic diversity of Arabica coffee and this diversity warrants the availability of large gene pools having desired characteristics improvement of the crop and disease and pest resistance (Paulos & Demeil, 1999).

Arabica coffee (*C. arabica*) is the only coffee species occurring in Ethiopia and it is geographically isolated from the rest Coffee species and is naturally restricted to two isolated mountain forests on the western and eastern sides of the Great Rift Valley in the southern Ethiopia (Tadesse Gole et al., 2002). There are wild and cultivated types of *C. arabica* in Ethiopia grown generally under four different production systems; Forest, semi-forest, garden and plantation Workafes & Kassu, 1999; . Wild populations of *C. arabica* are under the forest production system and grow naturally in the undergrowth of the montane rain forests in southwest and Eastern part of Ethiopia at altitudes between 1,400 and 1,900 m. (Gole et al. 2001). This forest is part of a natural forest ecosystem and it represents 9% of total land covered by coffee and the production system contributes about 5-6% of the total coffee production in the country.

out.

Despite the importance of the Ethiopian forests and coffee production systems for the conservation and use of coffee genetic resources and other forest biodiversity, they are disappearing at an alarming rate due deforestation and land use change. Deforestation is estimated at 10,000 ha/year in the coffee growing areas in the southwest Ethiopia (Tadesse et al., 2002).

Taxonomy, Reproductive biology and Morphology of *C. arabica*

Taxonomy

Taxonomically, *C. arabica* L. belongs to genus *Coffea* L., which in turn belongs to a large family of flowering plants, *Rubiaceae* (Wellman, 1961; Charrier and Berthaud, 1985). A. de Jussieu made the first botanical description of a coffee tree in 1713 under the name of *Jasminum arabicanum laurifolium* by studying a single plant originating from the botanic garden of Amsterdam. However, in 1737 C. Linnaeus classified it as a separate genus *Coffea* L. with the then only one species known, *C. arabica* L. However, since the second half of the nineteenth century many new species of *Coffea* have been discovered by exploration of tropical forests of Africa, where most of the species are found.

The taxonomy of the genus is very much confused as many new species are being discovered. The number of species described under the genus varies from author to author, ranging from 60 to 100 (Bridson and Verdecourt, 1988; Coste, 1992). Although the taxonomy is not settled, the genus is classified into sections and subgenera by different authors based on different criteria. For example, Chevalier, (1947) cited in (Charrier and Berthaud, 1985) classified the

genus in to four sections namely *Argo Coffea*, *Paracoffera*, *Mascarocoffea* and *Eucoffea*. He also further classified the sections into subsections. Sections *Eucoffea* (more correctly *Coffea*) and *Mascarocoffea* include most of the presently known coffee species. Leroy (1980) classified the genus into two subgenera as subgenus *Coffea* and subgenus *Baracoffea* based on growth habit (monopodial vs. sympodial) and type of inflorescence (axillary vs. terminal). However, the first two sections are excluded from the genus by Leroy (1967) cited in (Charrier and Berthaud, 1985).

All the caffeine-containing coffee trees belong to subgenus *Eucoffea* that probably originated from central and east Africa centers and differentiated in to different species such as *C. arabica* L. from Ethiopia, *C. eugenioides* from Kenya and *C. canephora* as well as *C. liberica* from West Africa. Some other species like *C. congensis* and *C. excelsa* come from the Central African Republic (Berthou, 1983).

Morphology and reproductive biology of *C. arabica*

Coffea arabica is a shrub or small tree, and if untended, it may reach a size of 4 to 5 meters. The plant has a dimorphic habit of branching in which vertical (orthotropic) branches form horizontal (plagiotropic) branches, which bear the flowers and the fruits in clusters. Flowers of *C. arabica* with short corolla, long style and exerted stamen are typical of the genus *Coffea*. Inflorescences develop from serial buds mainly on horizontal branches. Each inflorescence normally carries one to five flowers. The flowers have a short pedicel and a rudimentary calyx. The petals are fused and form corolla with five lobes. The pistil consists of an inferior ovary and a long style with two stigmatic lobes. The ovary is bilocular each with one anatropous ovule. Flower initiation occurs after sufficient rainfall following a dry period. The total period of flowering is normally not more than three days with the majority of flowers opening on the first and the second day. Pollen shedding starts very soon after opening of the flowers early in the morning and the stigma is then receptive. Flowers wither in one or two days after pollination. It takes six to eight months from flowering to fruit ripening.

The coffee fruit usually contains two seeds. Ripe fruits have a thick fleshy mesocarp (pulp) and a hard endocarp (parchment). In addition, each seed is enveloped in a silver skin (testa), which is a remnant of the integument (perisperm). The tough endocarp is to protect the seed from digesting enzyme activities in the gut of frugivores such as birds and mammals. The fleshy, sugar-containing mesocarp and the vivid coloration due to anthocyanins of the exocarp act as a reward and to attract the dispersing animals, respectively (Aga, 2006).

The coffee bean consists of an endosperm and small embryo embedded at the basal end of the seed. There is no seed dormancy in coffee, and seed viability is normally lost within three to six months after harvesting. The coffee seed is recalcitrant, but it is possible to preserve the viability for up to two and half years when coffee seeds at moisture content of 41% is stored at 15°C (Van der Vossen, 1979).

The floral morphology of *C. Arabica* would permit natural cross-pollination, but nevertheless, *C. arabica* is largely autogamous, and fruit set after self-pollination is 60% or higher (Carvalho *et al.*, 1969). Most studies on the degree of natural cross-pollination were carried out on cultivars of *C. arabica*, which underwent many cycles of selection. Using the recessive marker genes Cera (Yellow endosperm) and Purpurascens (purple leaves) Van der Vossen (1974) in Kenya found 7 to 15 percent of natural cross-pollination in *C. arabica*. Most diploid species of *Coffea* have proved to be highly self-incompatible, and incompatible, are

The genome of *C. arabica* is tetraploid ($2n=4x=44$). Recent investigations established that *C. arabica* L. is formed by natural hybridization between two closely related diploid species, *C. canephora* and *C. eugenioides* (Lashermes et al., 1999). *C. arabica* is mainly autogamous (inbreeding), although the floral morphology would permit natural cross-pollination. It has short corolla, long style and exerted stamens (fig.). Studies made on wild populations of *C. arabica* in Ethiopia showed variation in the degree of self-fertility, the rate of cross-pollination ranging from 40-60 percent (Meyer, 1965, cited in Charrier & Berthud, 1985). Mixed mating is known to cause more genetic variability within populations and is a potential for differentiation, depending on the level of selfing and may vary in time, in the genetic structure of among populations (Loveless & Hamrick, 1984).

23

Genetic diversity of C. arabica L.

Genetic diversity refers to the variation of genes within species, that is, the heritable variation within and between populations of organisms. . In the end, all variation resides in the sequence of the four base pairs that compose the DNA molecule and, as such, constitute the genetic code. Other kinds of genetic diversity can also be identified at all levels of organisation in the nucleus, including the amount of DNA per cell, chromosome number and DNA structure.

The generation of new Genetic variation is affected by life history and evolutionary forces such as selection, mutation, recombination and genetic drift (Karp et al., 1997; Tan and tan, 2002). occurs continuously in individuals through chromosomal and gene mutations, which, in organisms with sexual reproduction, are propagated by recombination. Genetic variation is also influenced by selection. The consequences of these phenomena are changes in gene and allele frequencies that account for the evolution of populations. Similar situations can occur through artificial selection such as breeding.

~~Plant genetic resources comprise the present genetic variation that is potentially useful for the future of humankind. These resources include traditional varieties, landraces, commercial cultivars, hybrids, and other plant materials developed through breeding; wild relatives of crop species; and others that could be used in the future for either agriculture or environmental benefits. Hence, plant genetic resources should be conserved. Conserved, with the ultimate reason being to eventually use them as a source of potentially useful genetic variation.~~

~~Methods of Detecting Genetic Diversity~~

~~Study of genetic diversity is the process by which variation among individuals or groups of individuals or populations is analyzed by a specific method or a combination of methods (Rao, 2004). Diverse data sets, often involving numerical measurements, are used in genetic diversity studies of plants. The genetic diversity analysis may be made at different levels: individual genotypes such as inbred lines or pure lines or clones, populations, germplasm accessions, and species.~~

~~This section deals with the different methods of genetic diversity study, starting with the general overview of the methods followed by their use in study of *C. arabica* L.~~

~~Methods of measuring genetic diversity (variation)~~

~~*To conserve and use genetic variation, it should first be assessed, that is, the extent and its distribution need to be determined. Variation can be evaluated on the phenotypic and genotypic levels.* Assessment of phenotypic variation focuses on morphological traits—those characteristics that define the shape and appearance of a set of individuals. Some of these traits can be considered as ‘genetic’ if~~

~~their presence in related individuals is heritable and not dependent on the environment, meaning that they are associated with a particular DNA sequence. Assessment of genotypic variation is at the level of the DNA molecule responsible for transmitting genetic information. The DNA molecule is composed of nucleotides, which are organized in a double-helix configuration in increasing levels of complexity up to the chromosomal units.~~

~~The different methods for detection of genetic diversity may include morphological characteristics, molecular markers and even karyotypic variations. Each approach has its own advantages and disadvantages and each data set provide different types of information and differ in scopes of studying objectives analysis costs. The choice of analytical method(s) depends on objective(s) of the study, the level of resolution required, the resources and technological infrastructure available, and the operational and time constraints. Thus, careful planning is needed for choosing a suitable method before carrying out research on genetic diversity. The sections below deal with some of the relevant methods.~~

~~*Genetic Marker*~~

~~A genetic marker is a measurable character that can detect variation in either a protein or DNA sequence. A difference, whether phenotypic or genotypic, may act as a genetic marker if it identifies characteristics of an individual's genotype and/or phenotype, and if its inheritance can be followed through different generations.~~

~~A genetic trait may not have necessarily observable consequences on an individual's performance. Sometimes, however, this trait may be linked to, or correlated with, other traits that are more difficult to measure and do affect the individual's performance. In such cases, these unobservable genetic traits may be used as genetic markers for the linked traits~~

because they indirectly indicate the presence of the characteristics of interest. The two measures can be correlated, using an analysis of inheritance and studying the distribution of the characteristics in both parents and offspring.

Types of genetic Markers

There are three major types of genetic markers. Morphological traits, protein (biochemical) markers, and DNA (molecular) markers. The sections that follow describe each type briefly.

Morphological characters

Traditionally, diversity within and between populations was determined by assessing differences in morphology. These measures have the advantage of being readily available, do not require sophisticated equipment and are the most direct measure of phenotype, thus they are available for immediate use, an important attribute. However, morphological determinations need to be taken by an expert in the species, they are subject to changes due to environmental factors and may vary at different developmental stages and their number is limited.

Genetically based morphological characteristics were among the earliest genetic markers used for assessment of variation and are still of great importance. They can be continuous or discrete and may be determined by many or few genes with some input from the environment. Thus the total phenotypic variation should be partitioned into its genotype, environmental, and genotype-by-environment components. These characters are inexpensive and simple to score when the characters are discrete and huge work is required

~~when they are continuous. For example, assessment of morphological characters in perennial plants such as coffee, often require a lengthy and expensive evaluation during the whole vegetative growth (Aga, 2005).~~

~~The sharing of physical features is also often accepted as an indication of relatedness. There are several sets of physical character assessment for different crops at different developmental stages such as seed, juvenile, adult vegetative, flower and fruit. However, since these sets of characters lack adequate coverage of the genome, are strongly influenced by environmental factors, and are apparently controlled by several genes (Wang & Tanksley, 1989) their use in genetic diversity studies is limited.~~

~~*Protein-based (Biochemical) molecular markers*~~

~~To overcome the limitations of morphological traits, other markers have been developed at both the protein level (phenotype) and the DNA level (genotype). Protein markers (seed storage proteins and isozymes) are generated through electrophoresis, taking advantage of the migrational properties of proteins and enzymes, and revealed by histochemical stains specific to the enzymes being assayed. Detecting polymorphisms—detectable differences at a given marker occurring among individuals—in protein markers is a technique that shares some of the advantages of using morphological ones. However, protein markers are also limited by being influenced by the environment and changes in different developmental stages (Rao, 2004).~~

~~*Molecular markers*~~

Recent developments in molecular biology such as protein electrophoresis and DNA-based molecular markers have been replacing or complementing traditional morphological and agronomic characterization, since they are virtually unlimited, cover the whole genome, are not influenced by the environment, and less time consuming. Protein markers are usually named 'biochemical markers' but, more and more, they are mistakenly considered as a common class under the so-called 'molecular markers'. Each molecular marker has its advantages and drawbacks. Therefore, application of molecular marker techniques to diversity questions must take into account whether or not the data derived from a technique provide the right type of information for answering the question being addressed (Karp et al., 1997). This in turn depends on the taxonomic levels of the material being studied (different species, subspecies, populations, cultivars and individuals). The closer the relationship of the materials to be studied, the more necessary it may be to consider highly discriminatory techniques. Besides, the choice of appropriate molecular markers also depends on the accessibility and cost effectiveness of the marker techniques.

Both protein electrophoresis and DNA-based molecular marker techniques have certain properties in common; all involve careful extraction of molecules (proteins or DNA) and electrophoretic separation of the molecules on a gel so that polymorphisms can be detected. The type of genetic variation that is sampled by these methods includes allelic variation that is in gene products (allozymes) and length variation in specific DNA fragments (Parker et al., 1998). DNA sequencing can be used for greater resolution of DNA sequence divergence.

Protein-based (Biochemical) molecular markers

~~To overcome the limitations of morphological traits, other markers have been developed at both the protein level (phenotype) and the DNA level (genotype). Protein markers are usually named 'biochemical markers' but, more and more, they are mistakenly considered as a common class under the so-called 'molecular markers'. Protein markers (seed storage proteins and isozymes) are generated through electrophoresis, taking advantage of the migrational properties of proteins and enzymes, and revealed by histochemical stains specific to the enzymes being assayed. Detecting polymorphisms—detectable differences at a given marker occurring among individuals—in protein markers is a technique that shares some of the advantages of using morphological ones. However, protein markers are also limited by being influenced by the environment and changes in different developmental stages. Even so, isozymes are a robust complement to the simple morphometric analysis of variation.~~

~~Both enzymes and non-enzyme proteins have been used as genetic markers (Murphy et al, 1996).. Non-enzyme proteins (in particular storage proteins) are usually analysed by one or two dimensional polyacrylamide gel electrophoresis (PAGE) where normally several bands or spots relating to different molecular forms of the protein are observed. In the case of enzymes, staining based on their specific activity is usually exploited for their detection analysis, therefore non-denaturing starch or polyacrylamide gels are used. Differences or polymorphisms due to different molecular forms of the enzyme but with conserved activity are detected by differential migration within the gels.~~

~~Enzymes include isozymes and allozymes. Two general forms of protein data can be gathered simultaneously using electrophoretic methods. One is derived from isozymes,~~

~~which are all functionally similar forms of enzymes, including all polymers of subunits produced by different gene loci or by different alleles at the same locus. A The other data set consists of allozymes , which are the variants of polypeptides representing different allelic alternatives of the same gene locus (Murphy et al, 1996). Both methods involve proteins that can be separated on the basis of net charge and size. The differences in movements may be due either to changes at the DNA level, which causes amino-acid substitutions and changes in charge of the protein, or post-translational modifications such as glycosylation, which lead to changes in molecular weight. After electrophoresis, enzyme-specific stains are applied to the gel so that the positions of different allozymes can be visualized as coloured bands and interpretation of the electromrph bands requires thorough understanding of the genetic basis of allelic variation for each enzyme (Murphy et al, 1996; Parker et al, 1998). Homozygotes at a given locus typically yield one band, while heterozygotes typically yield two, three, or five bands, depending on the quaternary structure of the enzymes.~~

~~Protein based markers have the advantages that they are relatively inexpensive and technically simple once established for a species and also need no sophisticated equipment. Comparable data from previous studies and a wealth of standard statistical procedures make allozymes appealing for studies of both fine and broad-scale genetic variation. They are usually co-dominant making them appropriate for heterozygosity studies. However, the main drawbacks to their use are the limited number of enzyme systems available, the use of specific detection methods for each enzyme, and only genomic regions coding for expressed proteins can be analysed resulting in low level of polymorphism. Although many statistical models in population genetics assume that phenotypic differences among allozymes are~~

minimal and selectively neutral, some allozymes may differ in metabolic function (Parker et al., 1998).

DNA-based molecular markers

With the development of new molecular biology techniques such as restriction digestion, polymerase chain reaction (PCR), southern blots and hybridization and others, to manipulate DNA many DNA-based molecular markers have been and are being developed (Karp et al., 1997; Palmer et al., 1998; Wolfe & Liston, 198). DNA polymorphisms can be detected in nuclear and organellar DNA, which is found in mitochondria and chloroplasts (Parker et al., 1998). Molecular markers concern the DNA molecule itself and as such, are considered to be objective measures of variation. The diverse arrays of molecular markers have become available for high resolution (even to the level of detecting single nucleotide mutations) genetic studies of population-level processes. DNA-based markers have several advantages over protein-based markers. For example, one can find DNA in nearly all cells of all organisms and can recover it from both dead and living tissues, tissues can be easily stored under field conditions, and in most cases very small amount of DNA (in nanograms) can be used. Furthermore, it covers the whole genome including both coding and noncoding regions.

DNA polymorphisms can be detected in nuclear and organellar DNA, which is found in

~~mitochondria and chloroplasts (Parker et al., 1998). Molecular markers concern the DNA molecule itself and, as such, are considered to be objective measures of variation. They are not subject to environmental influences; tests can be carried out at any time during plant development, and, best of all, they have the potential of existing in unlimited numbers, covering the entire genome.~~

~~Many different types of DNA molecular markers with different properties exist, each with its own advantages and disadvantages. However, the following features characterize a good molecular marker (Karp et al., 1997b):~~

- ~~• Polymorphic, that is, it is variable among individuals. The degree of polymorphism detected depends on the technology used to measure it.~~
- ~~• Reproducible in any laboratory experiment, whether within experimental events in the same laboratory or between different laboratories performing identical experiments.~~
- ~~• Codominant. Depending on the type of application, the selected technology must be able to detect the marker's different forms, distinguishing between homozygotes and heterozygotes (codominant inheritance). A heterozygous individual shows simultaneously the combined genotype of the two homozygous parents.~~
- ~~• Evenly distributed throughout the genome. The more distributed and dense genome coverage is, the better the assessment of polymorphism.~~

- ~~Discriminating, that is, able to detect differences between closely related individuals.~~
- ~~Not subject to environmental influences. The inference of a marker's genotype should be independent of the environment in which the individual lives or its developmental stage.~~
- ~~Neutral. The allele present at the marker locus is independent of, and has no effect on, the selection pressure exerted on the individual. This is usually an assumption, because no data are usually available to confirm or deny this property.~~
- ~~Inexpensive. Easy, fast and cheap in detecting across numerous individuals. If possible, the equipment should be of multipurpose use in the experiment.~~

~~In general, the most appropriate molecular marker is the one that show high-resolution i.e., the one that produces a large number of alleles at a single locus and/or many loci with two or more common alleles are better, with a minimum amount of effort and expense.~~

~~W~~

~~The choice of a method depends on the level of genetic variation required for a particular study and the amount of effort and expense (Palmer et al., 1998). Each molecular marker has its advantages and drawbacks. Therefore, application of molecular marker techniques~~

~~to diversity questions must take into account whether or not the data derived from a technique provide the right type of information for answering the question being addressed (Karp et al., 1997; Karp, 2002). This in turn depends on the taxonomic levels of the material being studied (different species, subspecies, populations, cultivars and individuals). The closer the relationship of the materials to be studied, the more necessary it may be to consider highly discriminatory techniques. Besides, the choice of appropriate molecular markers also depends on the accessibility and cost effectiveness of the marker techniques.~~

~~The most appropriate molecular marker is the one that show high-resolution i.e., the one that produces a large number of alleles at a single locus and/or many loci with two or more common alleles are better, with a minimum amount of effort and expense.~~

~~Plant DNA polymorphism assays include sequencing of a known region of a genome; non-PCR-based DNA markers such as restriction fragment length polymorphisms (RFLP) and PCR-based DNA markers. In coffee, DNA-based molecular marker technology has already been implemented in germplasm characterization and management, detecting genetically divergent breeding subpopulations (for example to predict hybrid vigour), establishing gene introgression from related species and molecular marker-assisted selection (Lashermes et al., 1996a).~~

~~PCR-based DNA markers~~

~~With the development of the polymerase chain reaction (PCR), many PCR-based DNA molecular techniques have been, and still are being developed for plant genome analysis. The techniques could be categorized into two main groups: arbitrary (or semi-arbitrary) primed techniques and site-targeted PCR techniques (Karp et al., 1997). Techniques in the first category use single primers which are designed arbitrarily/or semi-arbitrarily, i.e., without the knowledge of flanking sequence of the region to be amplified. These techniques include RAPD (random amplified polymorphic DNA) (Welsh & McClelland, 1990; Williams et al., 1990), AFLP (amplified fragment length polymorphism) (Vos et al., 1995), ISSR (inter-simple sequence repeat) (Zietkiewicz et al., 1994). These are also called multi-locus approaches,~~

~~Techniques in the second category depend on primers that target a single known site, such as a gene.~~

~~These are alternative approaches to multi-locus profiling. PCR-DNA sequencing sequence-tagged microsatellite (STMs), thermal gradient gel electrophoresis (TGGE), denaturing gradient gel electrophoresis (DGGE), cleaved amplified polymorphic sites (CAPS), single-strand conformational polymorphisms (SSCP), are important examples of these category.~~

~~Since, they give information on a single locus, they are particularly important when information is required on gene~~

~~frequency or genealogical information for genetic diversity management and when~~

~~information on~~

~~heterozygosity is required.~~

~~SSR (simple sequence repeat) or microsatellite (Morgante & Olivieri, 1993) and retrotransposons based markers such as inverse sequence tagged repeat (ISTR) (Rohde, 1996), retrotransposon-based insertion polymorphism (RBIP) (Flavell et al. 1998), inter-retrotransposon amplified polymorphism (IRAP) and retrotransposon-microsatellite amplified polymorphism (REMAP) (Kalendar et al. 1999).~~

Random amplified polymorphic DNA (RAPD)

~~This technique employs single primers with 10 arbitrary nucleotide sequences and at least 50% GC content. PCR products are separated on agarose gels and detected by staining with ethidium bromide. To obtain an amplification product with only one primer, there must be two identical or at least highly similar target sequences in close vicinity to each other on different strands in an opposite orientation. RAPD polymorphisms can theoretically result from several types of events: (1) insertion of a large piece of DNA between the primer binding sites may exceed the capacity of PCR, resulting in fragment loss; (2) insertion or deletion of a small piece of DNA will lead to a change in size of the amplified fragment; (3) the deletion of one of the two primer annealing sites results in either the loss of a fragment or an increase in size; (4) a nucleotide substitution within one or both primer target sites may affect the annealing process, which can lead to a presence versus absence of polymorphism or to a change in fragment size (Weising et al., 2005).~~

~~The greatest advantage of the RAPD approach is its technical simplicity, paired with the independence of any prior DNA sequence information. One obvious disadvantage that RAPD share with other multilocus markers is their dominant nature, which limits their use in population genetics and mapping studies (Lynch & Milligan, 1994).~~

~~Besides, RAPD is sensitive to slight changes in reaction conditions, which interfere with the reproducibility of banding patterns between separate experiments, PCR instrumentation, and laboratories (Ellsworth et al., 1993; Muralidharan & Wakeland, 1993; Penner et al., 1993). This high sensitivity is at least in part a consequence of the non-stringent PCR conditions, which are needed to allow for mismatch priming (Weising et al., 2005). An obvious measure to enhance reproducibility is to carry out replicate experiments, and exclude inconsistent bands from the analysis. Because reproducibility mainly depends on appropriate optimization of PCR components, it is advisable to determine optimal RAPD conditions empirically by performing a set of pilot experiments. Given that the outcome of RAPD experiments is influenced by many interacting variables, complete optimization can only be achieved if each component is tested independently and across a wide concentration range.~~

Amplified fragment length polymorphism (AFLP)

~~Amplified fragment length polymorphisms are DNA fragments obtained from endonuclease restriction, followed by ligation of oligonucleotide adapters to the fragments and selective amplification by the polymerase chain reaction (PCR). The PCR primers consist of a core sequence (part of the adapter), a restriction enzyme specific sequence and 1–3 selective nucleotides. Typically two successive PCRs are performed on the restricted template, using specifically designed primers that allow only a subset of the restriction fragments to be amplified. In the standard procedure described by Vos et al. (1995) one of the selective primers is radioactively labeled, the amplification products are separated on highly resolving sequencing gels, and banding patterns are visualized by autoradiography. The other alternative method is using non-labeled primers and visualize the bands with silver staining technique. Polymorphisms between two or more genotypes may arise from sequence variation in one or both restriction sites flanking a particular fragment, insertion or deletions within an amplified fragment, and differences in the nucleotide sequences immediately adjacent to the restriction sites. The advantages of the AFLP technology include no need of prior DNA sequence information, and the possibility of applying high stringency during PCR, which ensures high reproducibility of the method. Although it is a very powerful approach, it has a number of limitations such as dominance of markers, clustering of some markers in distinct genomic regions, limited levels of polymorphism in some cultivated species and the requirement of good quality DNA to ensure complete restriction (Weising et al., 2005).~~

Microsatellites (simple sequence repeats) (SSR)

The simple sequence repeats (SSRs) or microsatellite-based markers are also very attractive for plant genetic studies as they can be efficiently analysed by rapid and simple polymerase chain reaction (PCR) assays. Microsatellites are tandemly repeated motifs of 1-6 bases found in all prokaryotic and eukaryotic genomes analysed to date. They are present in both coding and non-coding regions and are usually characterized by a high degree of length polymorphism. Microsatellites are surprisingly common in the vicinity of genes, and tri-nucleotide repeats preferably occur in exons (Morgante et al., 2002). Slippage of DNA polymerase during DNA replication and failure to repair mismatches is considered as a mechanism for creation and hypervariability of microsatellites (Levinson & Gutman, 1987). Microsatellites were widely employed in many fields soon after their first description (Litt & Luty, 1989; Tautz, 1989; Weber & May, 1989). The large number of alleles and high level of variability among closely related organisms made PCR amplified microsatellites the marker system of choice for a wide variety of applications.

The popularity of microsatellite markers stems from a combination of several important advantages, namely their codominant inheritance, high abundance, enormous extent of allelic diversity, and the ease of assessing size variation by PCR with pairs of flanking primers (Weising et al., 2005). The major drawback of microsatellites is the necessity of sequence information for primer design that they need to be isolated de novo from most species being examined for the first time. However, there is some cross species transferability of SSRs among closely related species that can help to defray the initial development costs (Peakall et al. 1998).

Inter-simple sequence repeats (ISSRs)

~~Inter-simple sequence repeat (ISSR) is an alternative technique to study polymorphism based on the presence of microsatellites throughout genomes (Zietkiewicz et al. 1994). Inter-simple sequence repeats (ISSRs) are regions found between microsatellite repeats. The technique is based on PCR amplification of DNA segments present at an amplifiable distance between two identical microsatellite repeat regions that are oriented in opposite direction by using a single primer (Wolfe & Liston, 1998). ISSR PCR gives multilocus patterns which are very reproducible, abundant and polymorphic in plant genomes. The primers are usually 16–25 base pair long and are developed from within the SSRs themselves with or without a 5' or 3' anchor sequence [e.g., 5'-AA (CG)₆-3', 5'-(AGC)₄-GC-3']. The anchors are useful to guarantee locus-specificity to the inter-repeat region only. The microsatellite repeat motifs used as primers can be di-nucleotide, tri-nucleotide, tetra-nucleotide or penta-nucleotide. Because of the known abundance of repeat sequences spread all over the genome, this single primer PCR reaction targets multiple genomic loci to amplify mainly the inter-SSR sequences of different sizes.~~

~~Inter-simple sequence repeats (ISSRs) have high reproducibility possibly due to the use of longer primers as compared to RAPD, which permits the subsequent use of a high annealing temperature (45–60° C) leading to higher stringency. The amplified products are~~

~~usually 200-2000 bp long and amenable to detection by both agarose and polyacrylamide gel electrophoresis and can be detected with ethidium bromide, silver staining and/or radioisotopes.~~

~~The sources of variation in ISSR markers could be: (1) mutations at the priming site (SSR), which could prevent amplification of a fragment as in RAPD markers and thus give a presence/absence polymorphism; (2) an insertion/deletion event within the SSR region or the amplified region would result in the absence of a product or length polymorphism depending on the amplifiability of the resulting fragment size.~~

~~The ISSR marker technique has several advantages and disadvantages. Some of the advantages (in addition to reproducibility as mentioned above) include relative technical simplicity and quick typing because it does not require prior sequence information, identification of significant level of variation, variation within unique regions of the genome may be found at several loci simultaneously, microsatellite sequence specificity and its very useful for DNA profiling. As a result of these advantages and their universality and easiness of development (no needs to sequence data), ISSR markers are more and more requested (Bornet and Branchard 2004).~~

~~The technique is not without limitations. For instance, there is the possibility as in RAPD that fragments with the same mobility may originate from non-homologous regions, which can contribute to some distortion in the estimates of genetic similarities (Sanchez et al., 1996). In addition since the marker is dominant heterozygote cannot be distinguished from homozygote. The use of polyacrylamide gel electrophoresis and detection with silver staining or radioisotopes are also disadvantages.~~

Specific Genetic diversity studies of C. arabica Arabica

~~The world coffee industry is based on a very narrow genetic base of the most important species *C. arabica* L, as it is originated from few base-line individuals. These are represented by the two base-line varieties, Typica and Bourbon, and some cultivars and mutations derived from them (Anthony et al., 2001). Currently, three cultivars are grown worldwide: cv. Mundo Novo, a Bourbon and Typica hybrid, cv. Caturra, a Bourbon mutant with short internode length, and cv. Catuai, a Caturra and Mundo Novo hybrid. These cultivars are high yielding and produce a high quality beverage. However, breeding programs are limited by the very narrow genetic base, especially in improving pest and disease resistance (Van der Graff, 1981).~~

~~A genetic alternative to breeding limitations exists in the spontaneous and subspontaneous genotypes collected in the centre of origin of the species, the Ethiopian highlands (Anthony et al., 1999). Botanists and geneticists who visited Ethiopia in the 1960s and before observed the presence of high phenotypic variability among the wild coffee populations and landraces (Sylvain, 1955; Meyer, 1965; van der Graff, 1981; Tadesse, 2003). Ethiopia has got wide genetic diversity of Arabica coffee and this diversity warrants the availability of large gene pools~~

having desired characteristics for improvement of the crop and disease and pest resistance (Paulos & Demel, 1999).

In spite of its great economic importance of *C. arabica* L., relatively few studies have been carried out on its genetics, particularly on populations in Ethiopia. However, some recent investigations on materials in field genebanks, using molecular markers, showed that accessions from SW Ethiopia have greater genetic variability than cultivars (Orozco-Castillo et al., 1994; Lashermes et al., 1996; Anthony et al., 2001, 2002). Polymorphism among the subsponaneous accessions was much higher than among the cultivated accessions. The polymorphism was very low within the genetic bases, confirming the historical documentation on their dissemination. The results enabled a discussion of the genetic diversity reductions that successively occurred during the dissemination of *C. arabica* from its primary centre of diversity, SW Ethiopia (Anthony et al., 2002). The schematic representation of the Ethiopian and the world's coffee is shown in figure 2. Wider genetic variability exists in Ethiopian materials as compared to the world wide cultivated cultivars.

Figure 2. Schematic representation of the main steps in the history of coffee cultivation and the genetic diversity of Arabica coffee population of southwest Ethiopia, compared to cultivated population abroad (Anthony et al., 2002).

~~All categories of genetic diversity studies have been applied to characterize coffee genetic diversity. However, in light of the great economic importance of coffee, so far only a limited number of genetic studies have been carried out, both on wild populations and cultivars.~~, 2002

In light of the great economic importance of coffee, so far only a limited number of genetic studies have been carried out, both on wild populations and cultivars. It is confirmed, in general, that *C. arabica*, as compared to other coffee species, is characterized by a low genetic diversity. This is attributed to the recent evolution of the species by allopolyploidy and its predominantly self-pollinating nature (Lashermes *et al.*, 2000). However, genetic diversity, as expressed by the number of markers detected, and polymorphism appeared to be higher among accessions derived from collections of sub-spontaneous (forest coffee ecosystem) from Ethiopia as compared to Bourbon and Typica-derived accessions and Yemen cultivars (Anthony *et al.*, 2001).

~~All categories of genetic diversity studies have been applied to characterize coffee genetic diversity. Most of these studies were conducted on either cultivated materials or spontaneous/subspontaneous materials that are found in the field genebanks (Carvalho *et al.* 1969; Orozco-Castillo *et al.* 1994; Lashermes *et al.* 1996; Anthony *et al.* 2001). Most of these studies found a relatively low genetic diversity in the materials. Agro-morphological characters and, more recently, DNA-based genetic markers have enabled the characterisation of the two genetic bases of the cultivated material (Typica and Bourbon). Bourbon-derived cultivars are characterized by a more compact and upright growth habit, higher yield and better cup quality than the Typica-derived cultivars (Carvalho *et al.* 1969). Wild and cultivated coffee from Ethiopia was classified according to their geographic origin (i.e. southwestern and~~

south/southeastern Ethiopia) using agro-morphological observations (Montagnon and Bouharmont 1996).

Some attempts to characterize *C. arabica* germplasm using a number of enzyme systems revealed no polymorphism (Aga et al, 2006).

Recently a number of DNA-based molecular markers have been used to characterize coffee and to distinguish cultivated and wild types. Random amplified polymorphic DNA (RAPD) markers have shown a distinction between cultivated and wild material (Lashermes et al. 1996; Anthony; Anthony et al. 2001). The application of AFLP markers on coffee has proven to be effective in detecting introgression fragments in lines and derivatives of a spontaneous interspecific hybrid (*C. arabica* X *C. canephora*) (Lashermes et al. 2000).

Eleven microsatellite loci were recently identified in coffee, of which six were polymorphic within *C. arabica* (Combes et al., 2000). AFLP markers were used to assess polymorphism between and within Typica and Bourbon-derived accessions and Yemen cultivars (The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers, (Antony et al., 2002). Their relationship with wild material was defined analysing accessions derived from spontaneous trees collected in the center of origin of the species *C. arabica*. Microsatellite loci were used to confirm the structure of the genetic diversity detected by the AFLP markers and to compare the polymorphism within the genetic groups.

Inter-simple sequence repeat (ISSR) markers were also shown to be efficiently used for genetic differentiation of the *Coffea* species and to identify the parentage of *Coffea* interspecific hybrids (Ruas et al, 2003) and to study their genetic diversity (Aga et al, 2006; Tesfaye et al, submitted in

~~press). These two studies using molecular markers such as RAPD (Essays et al., 2003), ISSR (Essayas, et al., 2005; Kassahun et al., submitted), ISTR (Esayas et al., 2005) and SSRs (Esayas et al., submitted) are the first genetic diversity studies of wild coffee in Ethiopia. These studies found moderate amount of genetic diversity within and between wild *C. Arabica* populations from different geographic regions (Interregional analysis) in the country.~~

~~wild, (wild,) is also done recently Kassahu et al., Tesfaye et al. submitted. differentiation between wild plants and landraces and cultivars. The results also show high levels of intraregional diversity and coffee plants of a single plot are usually related closest, indicating hierarchical geographical patterns. Di and tetra nucleotide SSR primers were used for amplification. Six dinucleotide primers are completely coded and analyzed and results show extremely different levels of genetic diversity within regions, is was found to be~~

~~Although the above studies showed higher diversity within regions than between regions, only limited numbers of samples per region was used (used (less than 15). To know the amount of diversity and its distribution within regions, the use of large number of samples from different plots is required. The remaining potential genetic diversity of forest coffee in Ethiopia, Such detailed analysis, using efficient marker systems, at a regional stage which is still requires detailed study using efficient marker systems before it is lost forever. Therefore, more detailed genetic diversity analyses of wild *C. arabica* both at interregional and intraregional levels using ISSR markers is still important or (hereafter referred to as Yayu)~~

24. Background and Objectives ~~statment~~statement

The present study used one of the molecular techniques, ~~known as~~ inter-simple sequence repeat (ISSR) markers, ~~discussed in the previous sections,~~ to assess the genetic ~~variability~~ diversity and its distribution among populations of four ~~populations~~ plots of forest *C. arabica* ~~L.~~ in Yayu (~~Geba_Dogi~~) forest. Yayu is one of the few natural forests remaining with wild *C.arabica* ~~L.~~ populations in the SW Ethiopia proposed for *in situ* coffee gene reserve (Paulos and Demel, ~~1999~~2000; Taddesse, 2003). These proposals were made on the basis of some preliminary surveys of the area (Demel *et al.*, 1998; Paulos and Demel, ~~1999~~2000) and on the basis of vegetation ecology studies (Taddesse, 2003) at first. -

~~Interregional analysis (CoCE regions) of *C. arabica* populations (wild, land races, and commercial cultivars) using ISSR markers is also done recently (Tsfaye et al.). Di and tetra nucleotide SSR primers were used for amplification. Six dinucleotide primers are completely coded and analyzed and results show extremely different levels of genetic diversity within regions. In general interaregional genetic diversity analysis high level of within region genetic diversity was observed (Kassahun *et al.*, submitted in press). Yayu showed the higher genetic diversity from all other regions investigated indicating that the forest is area of priority for conservation of wild coffee gene pools in Ethiopia. To conserve and use this remaining potential genetic diversity of forest coffee in Ethiopia, further detailed analysis at the region/intraregional level is needed with the help of compared to interregional differences. Tetr-~~

nucleotide primers yield higher variability but seem to provide less resolution at the interregional level.

The remaining potential genetic diversity of forest coffee in Ethiopia, which is a crucial step for *in situ* conservation and breeding purposes, still requires detailed study using efficient marker systems and larger sample sizes before it is lost forever. Therefore, more detailed genetic diversity analyses of wild *C. arabica* both at interregional and intraregional levels using ISSR markers is still important. Therefore, this study aims at conducting the detailed analysis using ISSR markers among wild *C. arabica* individuals from different plots of one of a CoCE sites: Yayu (Geba-dogi) forest under the objectives stated in the following sections.

3. Objective of the Study

3.4.1. General Objective

To design effective conservation strategy and thereby to identify sites with high diversity estimate, prior characterization of the existing wild coffee populations is important. The identification of these sites has a paramount importance for germplasm collection mission and appropriate *in situ* site selection.

Thus, the general objective of this study was to characterize the genetic variation that exists within and among the coffee populations in Yayu forest and based on the results suggest areas of priority for conservation (both *in situ* and *ex situ*) and for use of the genetic resources in breeding.

3.4.2. Specific Objectives

The specific objectives of this study were:

1. To investigate the genetic diversity within- and among-populations of *C.arabica* in Yayu forest and see how the diversity is distributed among the populations.
2. To generate information for *ex situ* conservation collection and target wild *C. arabica* populations for *in situ* conservation.
3. 3.—To generate information for further improvement of *C.arabica* using parents from the wild populations.

5. Materials and Methods

5.1. Description of study site

Yayu (Geba-Dogi) forest is located in the southwestern of the northwestern highlands, in the Illubabor Zone of Oromia State (Figure 4) at 550 km from the capital of Ethiopia, Addis Ababa and at about 60 Km from Metu. The Institute of Biodiversity Conservation and Research (IBCR) protects the forest area covering a total of 10,200 ha as Geba-Dogi forest coffee conservation project since 2003 and it is one of the five research sites for CoCE project. The soil of the study area is generally characterized by red or brownish ferrisols derived from volcanic parent material (Demel, 1999; Taddesse, 2003). Yayu forest is one of the eight important wild coffee areas in Ethiopia proposed for conservation. The forest is within the Afromontane rain forest system of the SW Ethiopia and is characterized by hot and humid climatic condition. The mean annual temperature is about 20⁰ C with mean minimum of 12.7⁰ C to mean maximum of 26.1⁰ C. The mean annual rainfall is 2100 mm per year ranging from about 1400 mm to 3000 mm per year and show uni-modal pattern, with lowest rain fall being in January and February and gradually increasing to the peak between May and October, and then decreasing in November and December. Two major rivers, Geba and Dogi, dissect the forest area. Dogi river drains into the Geba, which flows southwest ending in Baro river, which in turn flows into Nile crossing and draining most of the forested areas of the southwest part of the country.

The district of Yayu has the highest percentage forest cover (55.8%) compared to other districts in Ethiopia (Taddesse *et al.*, 2002). This is by far above the percent forest cover for the southwest parts of Ethiopia (18%) and that of the country as a whole (2.7%). Although most forest areas in the district are demarcated as National forest priority areas, the local people are

heavily dependent on the forest mainly for coffee and honey bees. The vegetation of the area is of the afro montane rain forest of the SW Ethiopian highlands, which is characterized by a mixture of *Podocarpus* and broad-leaved angiosperm species (Taddesse, 2003).

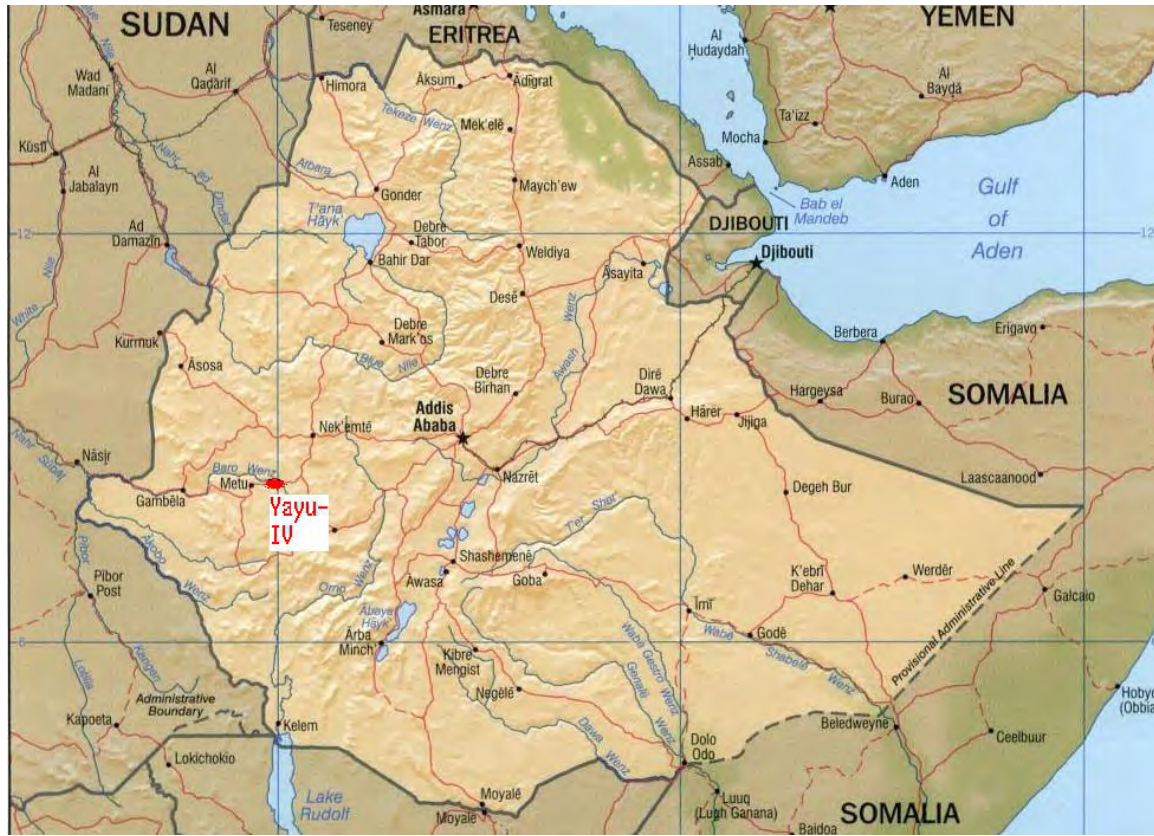


Figure 4. Map of Ethiopia showing the relative position of the study site (Yayu). It is marked as red ellipse on the road from Addis Ababa to Mettu.

5.2. Sampling

Four study sites (plots) that are found between 8 ° 22 ' 379 "N and between 8 ° 23 ' 994 "N latitude and between 35°47 ' 718 "E and 35 ° 48 ' 829 "E longitudes are selected in the forest (Table 2). The altitudinal variation of these sites ranges from 1388 masl to 1499 masl. These sites and some other additional sites were selected by CoCE project (Conservation and use of the wild populations of *Coffea arabica* in the montane rainforests of Ethiopia) for multidisciplinary

investigation. Yayu-3 is excluded from this study because it is very closer to two other sites, Yayu-1 and Yayu-2. Two of the selected plots, Yayu-1 and Yayu-2, are in the semi-forest (semi-managed forest) production systems and two, Yayu-4 and Yayu-5, are in the undisturbed (unmanaged) region of the forest and under forest coffee production system.

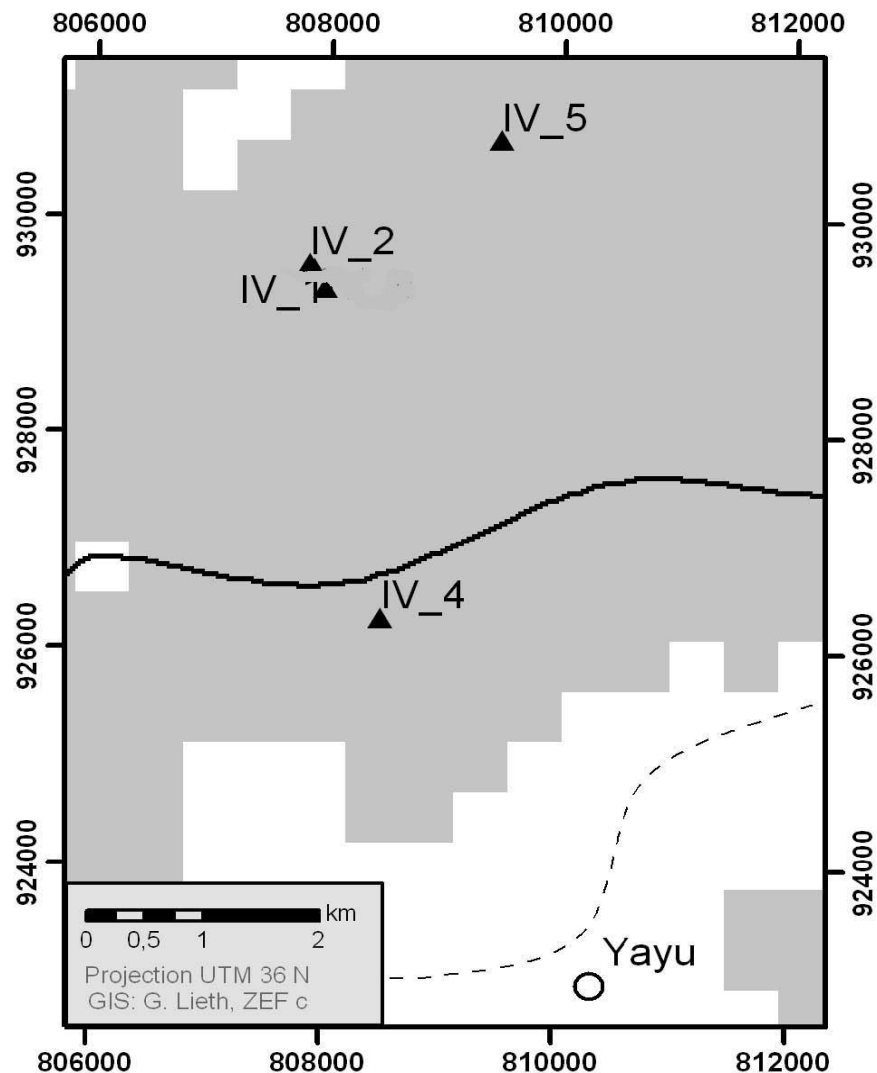


Figure 5. Sampled plots of Yayu forest for the present analysis. The four plots in the present study are IV-1, IV-2, IV-4, and IV-5 and represent Yayu-1, Yayu-2, Yayu-4, and Yayu-5 respectively. Site IV-3 is excluded from the present study since it is nearer to and in between sites IV-1 and IV-2. The map is drawn using GPS coordinates taken from the plot

From each site, originally, 25 coffee trees, ranging from two years age to very old trees, were randomly sampled from an area of approximately 50m x by 50m. However, not all of these samples were used in the analysis because of either failure to amplify or get DNA during extraction. In addition, three samples of *C. arabica* L. from Sore forest were included for comparison, only in the dendrogram. In all cases, four to six young leaves that are not exposed to fungi and other epiphytes were picked and put in plastic bag containing silica-gel for rapid drying and were transported to the laboratory at the Science Faculty of Addis Ababa University.

Table 2. List of *C. arabica* L. populations included in this study with sample size (original) and site location

<u>Population/site</u>	<u>Sample size</u>	<u>Site information</u>
<u>Yayu -1 (dist)</u>	<u>25</u>	<u>8 ° 23' 994" N/ 35° 47' 718" E/1499 masl</u>
<u>Yayu -2 (dist)</u>	<u>25</u>	<u>8° 23' 711" N / 35° 47' 728" E/1490 masl</u>
<u>Yayu -4 (undist)</u>	<u>25</u>	<u>8° 23' 484" N / 35° 47' 882" E/1468 masl</u>
<u>Yayu-5 (undist)</u>	<u>25</u>	<u>8° 22' 379" N / 35° 48' 029" E/1388 masl</u>
<u>Sore (dist)</u>	<u>3</u>	<u>8° 19' 14.8" N/ 35° 42' 20.8" E/1330 masl</u>

Dis – *disturbed forest*

Undis-*undisturbed forest*

N.B: Samples from Sore are only included in the dendrogram for comparison. They are not included in any other analysis.

5.3. DNA extraction

Total genomic DNA was extracted from silica-gel-dried young leaves of coffee by a modified version of CTAB method (2% cetyltrimethylammoniumbromide, 1% polyvinylpyrrolidone, 100 mM Tris (pH 8), 20 mM EDTA, 1.4 M NaCl, 0.2% beta-Mercapto-ethanol) (Borsch *et al.*, 2003).

The isolation procedure used triple CTAB extractions to yield optimal quantities of high-quality DNA from tissues with considerable amounts of secondary compounds such as alkaloids that occur in coffee. About 100 mg of dry tissue were ground in sand (quartz) and incubated at 65°C

for 30 min with 700 μ L of CTAB. After centrifuging and transferring the supernatant into a clean tube, the same tissue was re-incubated twice with CTAB solution. All three preparations were kept separate. The CTAB solutions were then extracted with chloroform twice, and the DNA was subsequently precipitated with ethanol. After separately re-suspending the pellets from all extraction steps in TE, two cleaning steps were carried out: the first by adding one-half volume 7.5 M ammonium acetate and precipitating with 100% ethanol, and the second by adding one-half volume 3 M sodium acetate and precipitating with ethanol. Genomic DNA from the second and third extractions was usually clean enough to be directly used for polymerase chain reaction (PCR) amplification and sometimes diluted to 1:5.

Only one of the three extractions was used for ISSR-PCR analyses. Selection was made based on the quality and quantity of the genomic DNA by running Esaygaasrose gel electrophoresis at concentration of 0.98% (*NEEO*, Germany). In most of the cases, the third extraction is relatively pure and selected. The original DNA was also diluted by 1:5 in most cases and by 1:10 in some cases before using for ISSR-PCR. EsayasIf PCR amplification was inhibited, further cleaning steps were taken by QIAquick PCR-purification kit (Qiagen GmbH., Hilden, Germany).

5.4. Inter-Simple Sequence Repeat (ISSR-PCR)

5.4.1. Primers used

Inter-simple sequence repeats (ISSRs) analyses in this study are used to investigate the genetic diversity of wild *C. arabica* in Yayu forest. Six simple-sequence primers, five di-nucleotide and one tetra-nucleotide, that showed polymorphisms were selected based on previous publications of the interregional ISSR analysis (Govers *et al.*, in press; Kassahun *et al.*, in press). The primers

and their 5'-3' sequences are given in Table 3. All of the di-nucleotide primers are 3'-anchored and the primers range in size from 16-18 base pairs.

5.4.2. PCR amplification

Each DNA amplification was performed in a total volume of 26µl containing 2.6µl of 10x *Taq* buffer (50mM KCl, 10mM Tris-HCl, PH= 8.4, 2.5mM MgCl₂, 0.1µ g/µl of gelatin)(PeQLab), 250µM of each of dATP, dGTP, dTTP and dCTP, (PeQLab), 0.23µM of primer (PeqLab), 1U/reaction of *Taq* DNA-polymerase (MWG-Biotech AG), and 1µl of genomic DNA.

The amplification was done in Biometra T₃ PCR machine (Biometra, Personal) under the following cycling conditions. The lid temperature was adjusted to 105 ° C and after initial preheating for 4 minutes at 94°C 40 cycles of denaturation (at 94 °C for 15 seconds), annealing (for 1 minute, at 45°C for primers 810-H, 812-H, 814-H, and CoIS001 and at 48°C for primers 818 and 844-H), polymerisation (at 72 °C for 1:30 minutes) followed by a 7 min. final extension step: at 72 °C.

5.4.3. Gel electrophoresis

To separate fragments, 8µl of the PCR amplification product pre-mixed with 2µl of 6x loading dye was subjected to 1.67 percent Esagyas rose gel electrophoresis for about two hours under a power supply of 100 volts. It was run in 1x TBE (containing 10.8g of Tris, 5.5 g of boric acid 0.05 M Na-EDTA) buffer and a 100 base pair DNA ladder (*ROTH*) was used as a molecular weight marker to estimate fragment size. The gel was visualized under UV light at 365nm (BioDocAnalyze, Biometra-035-300) after staining with ethidium bromide (1.75µg/ml) for

about 30 min. and de-staining by distilled water for about 30minutes. Gel picture was taken by digital camera, connected to PC with BioDocAnalyze program and saved for later data scoring.

5.5. Data Scoring and Analysis

Out of the 100 samples collected from the four populations only 78 gave good amplification products for the five di-nucleotide primers and these are used in the analysis below. Since bands for the tetra-nucleotide primer are very close to each other (Figure 6), it is difficult to score and use in analysis. Each amplified DNA fragment was considered as an independent character (locus), and scored as present (1) or absent (0). Each amplified product was named by the code of the primer followed by Arabic number starting from the larger fragment to the smaller ones. Since ISSR markers are dominant, a locus was considered to be polymorphic if the presence and absence of the bands were observed in various individuals, and monomorphic if the bands were present in all individuals. No distinction was made between fragments of the same molecular size that varied in intensity although such differences in intensity of bands were observed.

The synthesised band profile was subjected to pair-wise comparison to compute Jaccard similarity (Jaccard, 1908) and Nei standard genetic distance (Nei, 1972) matrices for each pair of individuals. Cluster analyses of the similarity and genetic distance matrices were made using unweighted pair-group method using arithmetic averages (UPGMA) (Sneath and Sokal, 1973) algorithm using Free tree program (Pavlieek *et al.*, 1999), version: 0.9.1.50. Genetic distance between each pair of populations was calculated by Nei's unbiased genetic distance (Nei, 1978) using popgene software version 1.32 (Yeh *et al.*, 1999) and cluster analysis of the genetic distance/similarity matrix was done by UPGMA algorithms using NTSYSpc program version ~~2.11T~~(2.11T (Rolf, 1997).

Genetic diversity and its distribution were calculated, based on variants frequencies, using Nei's gene diversity statistics (Nei 1987) and Shannon's information index. Gene diversity was calculated for each locus and each population according to Nei (1987) using Popgene software version 1.32(Yeh *et al.*, 1999). The same program was used to calculate the percent polymorphic loci for each population and for the whole data. Analysis of gene diversity in the subdivided population was also done by Nei (1973) formula using the same program. Comparison was made between results obtained over all the loci and those obtained only over polymorphic loci.

Shannon diversity index was calculated for each locus for each population as $H_j = -\sum p_i \log_2 p_i$, where p_i is the frequency of the presence or absence of ISSR band in that population. The average diversity for each locus across the four populations was calculated as $H_{pop} = \sum H_j/n$, where n is the number of populations, while the mean observed Shannon diversity of each population across all loci was calculated as $H_{loci} = \sum H_j/L$, where L is the number of loci studied. Similarly, the overall Shannon diversity within the species for each locus was calculated as $H_{sp} = -\sum p_s \log_2 p_s$ where P_s is the frequency of the presence or the absence of ISSR band across the genotypes within the whole data. The overall diversity of the species/whole population (H'_{sp}) was then partitioned into proportion of within and among population genetic diversity for each locus as H'_{POP}/H'_{sp} and $1-(H'_{POP}/H'_{sp})$, respectively.

The extent of population differentiation was calculated as $G_{ST} = (H_T - H_S) / H_T$ from Nei's formula (Nei 1973), $G'_{ST} = 1 - (H'_{pop}/H'_{sp})$ from Shannon index, and F_{ST} from analysis of molecular variance (AMOVA) (Weir and Cockerham, 1984; Excoffier *et al.*, 1992). H_T is the measure of total genetic variation for the whole population and H_S and D_{ST} ~~is—the~~ are the measure of the

amount of genetic variation within- and among-populations, respectively. G_{ST} and G'_{ST} were calculated for each locus and their mean values were calculated over all the loci and over only polymorphic loci for comparison. Esayas Analysis of molecular variance (AMOVA) was done by Arlequin program version 3.01 (Excoffier, 2006).

Although the data available is not sufficient to make comparison of the plants according to their age structure, attempt was made to do so. The analysis was done by categorizing the plants into two groups: Young and old plants. Plants that are approximately less than ten years of age are considered to be young plants and those above ten years of age are considered to be old plants. A total of 31 plants, 17 and 14 plants, respectively for young and old plants for which data was available were included in this analysis. Only comparisons of their genetic diversity were made.

6. Results

6.1. Locus diversity

Only 78 plants were used in this analysis from the four populations of Yayu, the rest 22 were excluded due to incomplete amplification over all the primers or at least 50% of the primers. The five di-nucleotide primers produced a total of fifty-five scoreable and reproducible bands of which 21 bands (38.18%) were polymorphic across the 78 coffee tree samples investigated. A fragment was considered as polymorphic if it exists in some of the individuals and absent from other and monomorphic if it exists in all individuals Seven to 13 bands were obtained from each primer with mean of 11 bands per primer (Table 3). One to eight polymorphic fragments were observed per primer (4.2 on average). The fragments ranged from 200 to 2,500bp in size (Table 3). The tetra-primer is excluded from the analysis. However, if polyacrylamide gel was used instead of agarose gel the bands would be well separated and the pattern would give better information. The samples of electrophoretic patterns obtained with the ISSR markers are illustrated in Figure 6, one (A) for di-nucleotide 810-H and the other (B) for tetra-nucleotide primers, CoIS001.

6.2. Population level polymorphism

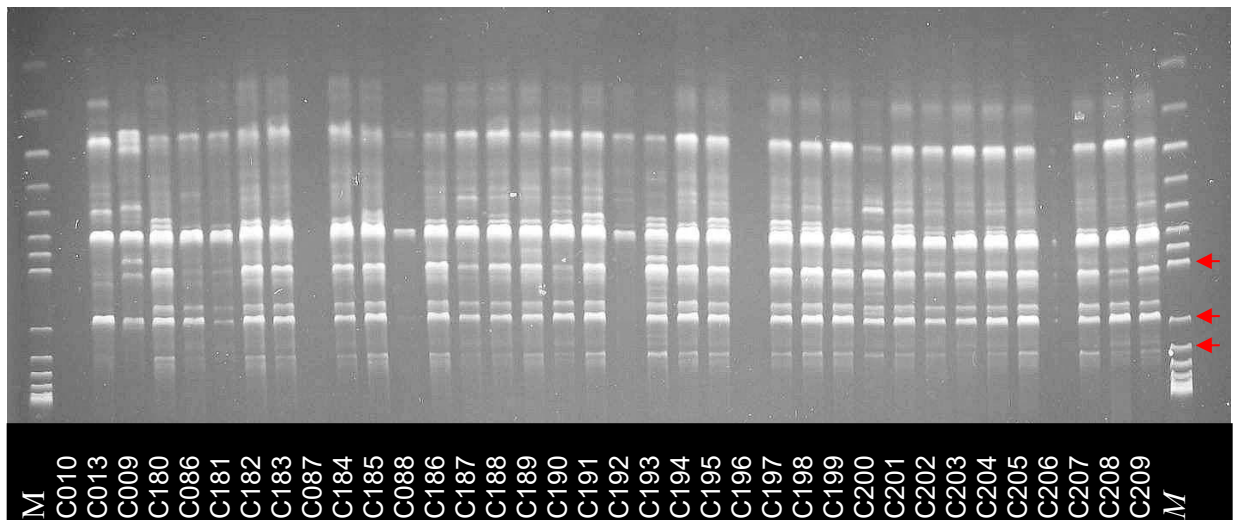
The frequency, in percentage, of the polymorphic ISSR markers within the four populations is summarized in Table 4. Of the twenty-one polymorphic fragments scored, 3, 1, 6, and 4 fragments are absent from Yayu-1, Yayu-2, Yayu-4, and Yayu-5 respectively. Thirteen markers are common to all the populations although they differ in frequency. Three markers are specific to Yayu-1 and Yayu-2 and one marker is specific to Yayu-4 and Yayu-5. One marker (812-6) is

Table 3. List of primers used in the ISSR analysis along with their nucleotide sequences, number of polymorphic bands and estimated molecular size range

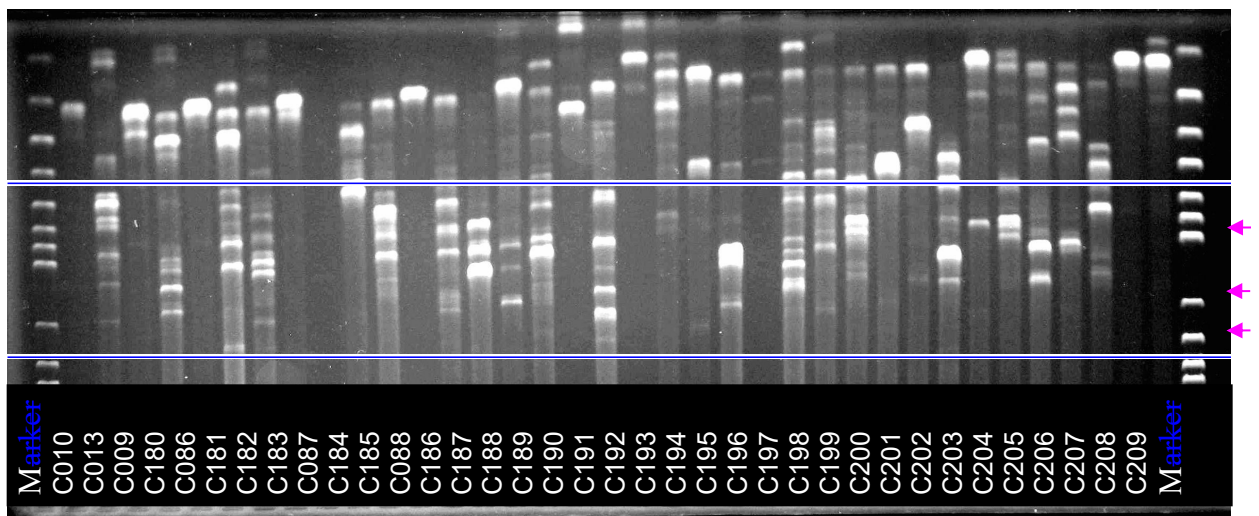
<u>Primer</u>	<u>Sequences</u> <u>5' to 3'</u>	<u>No. of</u> <u>scoreable and</u> <u>reproducible</u> <u>bands</u>	<u>No. of</u> <u>polymorphic</u> <u>bands</u>	<u>%</u> <u>Polymorphic</u> <u>bands</u>	<u>Molecular</u> <u>size range in</u> <u>bp</u>
<u>810-H</u>	<u>(GA) 8T</u>	<u>13</u>	<u>7</u>	<u>53.85</u>	<u>200-2,500</u>
<u>812-H</u>	<u>(GA) 8A</u>	<u>13</u>	<u>1</u>	<u>7.69</u>	<u>200-1,500</u>
<u>814-H</u>	<u>(CT) 8A</u>	<u>7</u>	<u>2</u>	<u>28.57</u>	<u>300-2,000</u>
<u>818-H</u>	<u>(CA) 8G</u>	<u>10</u>	<u>3</u>	<u>30</u>	<u>300-1,000</u>
<u>844-H</u>	<u>(CT) 8RC</u>	<u>12</u>	<u>8</u>	<u>66.67</u>	<u>300-2,000</u>
<u>CoIS0</u>	<u>(CCTA) 4</u>	<u>-</u>	<u>-</u>	<u>-</u>	<u>400-2,500</u>
<u>01</u>					
<u>Total</u>		<u>55 for the</u> <u>dinucleotides</u>	<u>21</u>	<u>38.18</u>	

specific to Yayu-2. Two markers are absent from Yayu-4 and are common to the rest of the populations and one marker is absent from Yayu-1 and is common to all the rest populations. Six to fifteen polymorphic fragments, 11.5 on average, were detected within the populations (Table 5). Yayu-1 and Yayu-2 presented higher number of polymorphic fragments (15) whereas Yayu-4 and Yayu-5 presented few numbers, six and ten fragments, respectively.

The population level genetic diversity statistics are summarized in Table 5. In general the mean gene diversity over all the loci of Yayu-1 and Yayu-2 are nearly equal and 0.072 and 0.079, respectively. When all loci are considered Yayu-2 is the most diverse (P=27.27%, H= 0.079, H_j= 0.175) and Yayu-4 is the least diverse (P=10.91%, H=0.031, H_j = 0.070) populations. When only the polymorphic loci are considered, however, Yayu-1 is the most diverse and Yayu-4 is the least diverse populations.



A



B

Figure 6. *Ethidium bromide-stained gels of ISSR banding patterns generated for 36 individual coffee trees from two plots in Yayu (Geba Dogi). A is based on a di-nucleotide 810-H and B on a tetra-nucleotide primer CoIS001. Outside lanes (M) show an extended 100-bp ladder. The arrows on the right side of both gel pictures indicate fragment sizes (arrows correspond to 1000 bp, 1500 bp and 2000 bp from top to bottom).*

As can be seen from Table 5, the mean gene diversities for the populations both over all loci and over polymorphic loci only show consistency except for the case of Yayu-1 and Yayu-2 in which the value over polymorphic loci for Yayu-1 is greater than that of Yayu-2. The mean Shannon diversity is larger for Yayu-2 and smaller for Yayu-4 both over all loci and on polymorphic loci.

Table 4. Frequency (%) of the polymorphic ISSR markers within the four populations

<u>Marker</u>	<u>Yavu-1</u>	<u>Yavu-2</u>	<u>Yavu-4</u>	<u>Yavu-5</u>
<u>810-2</u>	<u>6</u>	<u>52</u>	<u>100</u>	<u>100</u>
<u>810-5</u>	<u>94</u>	<u>100</u>	<u>100</u>	<u>100</u>
<u>810-6</u>	<u>88</u>	<u>100</u>	<u>100</u>	<u>100</u>
<u>810-9</u>	<u>41</u>	<u>35</u>	<u>28</u>	<u>83</u>
<u>810-10</u>	<u>76</u>	<u>100</u>	<u>100</u>	<u>89</u>
<u>810-11</u>	<u>94</u>	<u>100</u>	<u>100</u>	<u>94</u>
<u>810-12</u>	<u>0</u>	<u>52</u>	<u>100</u>	<u>6</u>
<u>812-6</u>	<u>0</u>	<u>4</u>	<u>0</u>	<u>0</u>
<u>814-4</u>	<u>73</u>	<u>82</u>	<u>0</u>	<u>65</u>
<u>814-5</u>	<u>54</u>	<u>45</u>	<u>0</u>	<u>0</u>
<u>818-4</u>	<u>88</u>	<u>78</u>	<u>0</u>	<u>0</u>
<u>818-7</u>	<u>75</u>	<u>78</u>	<u>0</u>	<u>0</u>
<u>818-8</u>	<u>75</u>	<u>91</u>	<u>38</u>	<u>100</u>
<u>844-3</u>	<u>94</u>	<u>89</u>	<u>100</u>	<u>100</u>
<u>844-5</u>	<u>88</u>	<u>86</u>	<u>100</u>	<u>94</u>
<u>844-6</u>	<u>81</u>	<u>90</u>	<u>100</u>	<u>100</u>
<u>844-7</u>	<u>100</u>	<u>100</u>	<u>84</u>	<u>100</u>
<u>844-8</u>	<u>100</u>	<u>95</u>	<u>84</u>	<u>89</u>
<u>844-9</u>	<u>100</u>	<u>91</u>	<u>0</u>	<u>6</u>
<u>844-11</u>	<u>0</u>	<u>0</u>	<u>11</u>	<u>28</u>
<u>844-12</u>	<u>6</u>	<u>10</u>	<u>5</u>	<u>50</u>

The results of the age-wise comparisons of the plants indicate that the younger plants are more diverse (P=27.27%, H= 0.097) than the older ones (P= 20%, H= 0.073) (Table 5). From analysis of gene diversity in these categories, most of the total gene diversity (HT= 0.088) is found within groups (H_s= 0.085) and only 3% exists among the groups (D_{ST}= 0.003). The unbiased Nei's genetic distance between the young and the old plants is only 0.003.

6.3. Cluster Analysis

Dendrogram was obtained using 81 individuals including three samples from Sore forest based on Jaccard similarity coefficient and UPGMA algorithms. The dendrogram has two main clusters

Table 5. Summary of population level and the whole data genetic diversity measures assuming Hardy-Weinberg disequilibrium (fis=0.95, Esayas, 2003)

<u>Populations</u>	<u>Polymorphic loci</u>		<u>Gene Diversity</u>		<u>Shannon's index</u>	
	<u>Number</u>	<u>Percent</u>	<u>Overall Loci</u>	<u>Over polymorphic loci</u>	<u>Overall the loci</u>	<u>Over Polymorphic loci</u>
<u>Yayu-1 (18)</u>	<u>15</u>	<u>27.27</u>	<u>0.072</u>	<u>0.279</u>	<u>0.162</u>	<u>0.425</u>
<u>Yayu-2 (23)</u>	<u>15</u>	<u>27.27</u>	<u>0.079</u>	<u>0.272</u>	<u>0.175</u>	<u>0.459</u>
<u>Yayu-4 (18)</u>	<u>6</u>	<u>10.91</u>	<u>0.031</u>	<u>0.081</u>	<u>0.070</u>	<u>0.183</u>
<u>Yayu-5 (19)</u>	<u>10</u>	<u>18.18</u>	<u>0.045</u>	<u>0.160</u>	<u>0.103</u>	<u>0.271</u>
<u>Mean</u>	<u>11.5</u>	<u>20.91</u>	<u>0.056</u>	<u>0.198</u>	<u>0.128</u>	<u>0.334</u>
<u>Entire data</u>	<u>21</u>	<u>38.18</u>	<u>0.100</u>	<u>0.306</u>	<u>0.223</u>	<u>0.583</u>
<u>Young (17)</u>	<u>15</u>	<u>27.27</u>	<u>0.097</u>	=	=	=
<u>Old (14)</u>	<u>11</u>	<u>20</u>	<u>0.073</u>	=	=	=

N.B. *The numbers in the bracket are samples used in the analysis*

in which individuals from disturbed and undisturbed regions are separated (Figure 7). The first main group consists of mainly individuals from the undisturbed regions of the forest (Yayu-4 and Yayu-5), and this main cluster further separated Yayu-4 from Yayu-5 with their own sub-clusters. The second main cluster is dominated by individuals from the disturbed regions of the forest (Yayu-1 and Yayu-2). There are four individual trees from Yayu-4 and two from Yayu-5 within the second main cluster. Two of the three samples from Sore are also within this cluster. This second main cluster does not separate well individuals from the two populations although there are some sub-clusters dominated by individuals from either population. The grouping of individuals from Sore with populations from disturbed habitats seems logical because Sore forest is also disturbed and gene flow is expected.

6.43. Partitioning of genetic diversity into within and ~~among~~ *Between* Populations

Genetic diversity was partitioned into within and among populations using three types of analyses: Nei's gene diversity in subdivided populations (Nei, 1973), Shannon's diversity index and analysis of molecular variance (AMOVA). The summary of genetic diversity in subdivided populations for the polymorphic loci is given in Table 6. Larger total gene diversity ($H_T > 0.25$) was observed for ten of the twenty-one polymorphic loci, of which five presented a high coefficient of genetic differentiation ($G_{ST} > 0.5$). Three markers showed a relatively high level of gene diversity within populations ($H_S > 0.25$); of which one marker (810-10) has the highest H_S ($H_S = 0.399$). The results indicate higher proportion of the genetic diversity within populations (mean $H_S = 0.056$ and 0.148 over all the loci and over polymorphic loci, respectively) as compared to the mean among population diversity ($D_{ST} = 0.044$ and 0.114 , respectively over all loci and over polymorphic loci). The mean values for the parameters ~~all~~ over-all loci are summarized in Table 7. The mean population differentiation calculated both over all loci and over polymorphic loci are nearly equal (Table 6 and Table 7).

Summary of genetic diversity in subdivided populations using Shannon diversity index is given in Table 8 and Table 9 over the ~~polymorphic~~ polymorphic loci and over all the loci, respectively. These values also show that larger proportion of the variation exists within population rather than among populations. The values for the means of H'_{pop} , H'_{sp} and G'_{st} were 0.334 , 0.574 and 0.435 respectively over the polymorphic loci only. Thirteen of the 21 polymorphic loci have H'_{pop} values of greater than 0.30 , which is also true for the Nei (1973) gene diversity. From

comparisons of the mean values for the within populations diversity (H'_{pop}/H'_{sp}) and the proportion of among populations diversity $[1-(H'_{pop}/H'_{sp})]$ larger genetic diversity exists within-populations. These mean values are 0.573 and 0.427 for ~~only~~ polymorphic loci ~~only~~ respectively. Only five of the 21 polymorphic loci show larger proportion of among-populations variation and the rest 16 markers showed larger within-populations variation.

Table 6. Summary of Nei's (1973) gene diversity statistics calculated in subdivided populations (Nei, 1987), for each polymorphic locus

Marker	Sample size	H_T	H_s	D_{ST}	G_{ST}
810-2	76	0.457	0.150	0.307	0.673
810-5	76	0.029	0.028	0.001	0.044
810-6	76	0.056	0.051	0.005	0.090
810-9	76	0.497	0.399	0.100	0.196
810-10	76	0.159	0.140	0.020	0.116
810-11	76	0.057	0.055	0.002	0.030
810-12	76	0.477	0.150	0.327	0.688
812-6	74	0.022	0.021	0.001	0.033
814-4	72	0.496	0.292	0.204	0.413
814-5	70	0.369	0.245	0.124	0.336
818-4	72	0.484	0.144	0.340	0.702
818-7	72	0.471	0.183	0.288	0.612
818-8	72	0.369	0.251	0.118	0.320
844-3	52	0.085	0.080	0.005	0.055
844-5	74	0.155	0.147	0.008	0.047
844-6	74	0.134	0.122	0.012	0.090
844-7	74	0.079	0.069	0.010	0.129
844-8	74	0.153	0.145	0.008	0.051
844-9	74	0.500	0.072	0.428	0.856
844-11	74	0.167	0.143	0.024	0.143
844-12	74	0.285	0.219	0.066	0.234
Mean	73	0.262	0.148	0.114	0.435

Total gene diversity (H_T), gene diversity within subpopulations (H_s), gene diversity among populations (D_{ST}), and the coefficient of gene differentiation among the populations (G_{ST}).

The mean values of the proportion of variation within (H'_{pop}/H'_{sp}) and among ($1-(H'_{pop}/H'_{sp})$) populations calculated both over all loci and only on polymorphic loci are nearly equal (Tables 8 and 9).

Table 7. Summary of the mean values of Nei (1973) gene diversity statistics in subdivided populations over all the loci (both polymorphic and monomorphic) for all the samples

<u>Parameter</u>	<u>H_s</u>	<u>D_{ST}</u>	<u>H_T</u>	<u>G_{ST}</u>
<u>Mean</u>	<u>0.056</u>	<u>0.044</u>	<u>0.100</u>	<u>0.436</u>

Table 8. Summary of Shannon's information index for each polymorphic locus overall the samples

<u>Marker</u>	<u>H'_{pop}</u>	<u>H'_{sp}</u>	<u>H'_{pop}/H'_{sp}</u>	<u>$1-(H'_{pop}/H'_{sp})$</u>
<u>810-2</u>	<u>0.326</u>	<u>0.939</u>	<u>0.347</u>	<u>0.653</u>
<u>810-5</u>	<u>0.077</u>	<u>0.101</u>	<u>0.766</u>	<u>0.234</u>
<u>810-6</u>	<u>0.126</u>	<u>0.176</u>	<u>0.717</u>	<u>0.283</u>

<u>810-9</u>	<u>0.849</u>	<u>0.996</u>	<u>0.853</u>	<u>0.147</u>
<u>810-10</u>	<u>0.317</u>	<u>0.399</u>	<u>0.795</u>	<u>0.205</u>
<u>810-11</u>	<u>0.155</u>	<u>0.176</u>	<u>0.882</u>	<u>0.118</u>
<u>810-12</u>	<u>0.326</u>	<u>0.975</u>	<u>0.334</u>	<u>0.666</u>
<u>812-6</u>	<u>0.067</u>	<u>0.103</u>	<u>0.646</u>	<u>0.354</u>
<u>814-4</u>	<u>0.613</u>	<u>0.991</u>	<u>0.618</u>	<u>0.382</u>
<u>814-5</u>	<u>0.493</u>	<u>0.800</u>	<u>0.616</u>	<u>0.384</u>
<u>818-4</u>	<u>0.324</u>	<u>0.991</u>	<u>0.327</u>	<u>0.673</u>
<u>818-7</u>	<u>0.390</u>	<u>0.980</u>	<u>0.398</u>	<u>0.602</u>
<u>818-8</u>	<u>0.545</u>	<u>0.764</u>	<u>0.713</u>	<u>0.287</u>
<u>844-3</u>	<u>0.207</u>	<u>0.318</u>	<u>0.649</u>	<u>0.351</u>
<u>844-5</u>	<u>0.361</u>	<u>0.406</u>	<u>0.888</u>	<u>0.112</u>
<u>844-6</u>	<u>0.285</u>	<u>0.357</u>	<u>0.800</u>	<u>0.200</u>
<u>844-7</u>	<u>0.157</u>	<u>0.245</u>	<u>0.643</u>	<u>0.357</u>
<u>844-8</u>	<u>0.355</u>	<u>0.406</u>	<u>0.874</u>	<u>0.126</u>
<u>844-9</u>	<u>0.195</u>	<u>0.999</u>	<u>0.195</u>	<u>0.805</u>
<u>844-11</u>	<u>0.334</u>	<u>0.452</u>	<u>0.741</u>	<u>0.259</u>
<u>844-12</u>	<u>0.522</u>	<u>0.671</u>	<u>0.233</u>	<u>0.221</u>
<u>Mean</u>	<u>0.334</u>	<u>0.583</u>	<u>0.573</u>	<u>0.427</u>

H'pop=genetic variation for the populations; H'sp=genetic variation for the whole population; H'pop/H'sp=proportion of within populations genetic variation; 1-(H'pop/H'sp)= proportion of genetic variation among populations.

Table 9. Partitioning of the genetic variation into within and among populations based on Shannon's information index calculated over all the loci for the whole data (78 plants).

<u>Parameter</u>	<u>Mean</u>
<u>H'pop</u>	<u>0.128</u>
<u>H'sp</u>	<u>0.223</u>
<u>H'pop/H'sp</u>	<u>0.574</u>
<u>1-(H'pop/H'sp)(G'ST)</u>	<u>0.426</u>

Hpop=mean genetic variation for the populations; Hsp=mean genetic variation for the entire population; Hpop/Hsp= proportion of genetic variation within populations; 1-(Hpop/Hsp)=proportion of genetic variation between populations.

Analysis of molecular variance (AMOVA) was done for the whole population (entire data) over all loci by considering all the populations as one geographic region. The measure was done by

computing distances between “haplotypes”, each individual’s data pattern as one “haplotype” and computing variance components for each level (Excoffier *et al.*, 1992). These results also support the larger within populations variation as compared to the among populations variation (Table 10). The percentage values for these estimates are 54.06 and 45.94, respectively.

~~A locus-by-locus AMOVA was also calculated for each polymorphic locus (data not shown) and the result is consistent with the calculation for the whole data together.~~

Table 10. Partitioning of genetic variation into within and among populations by Analysis of molecular variance (AMOVA)

<u>Source of</u>	<u>d.f</u>	<u>Sum of</u>	<u>Variance</u>	<u>Percentage of</u>
------------------	------------	---------------	-----------------	----------------------

<u>variation</u>		<u>squares</u>	<u>components</u>	<u>variation</u>
<u>Among populations</u>	<u>3</u>	<u>26.07</u>	<u>0.427</u>	<u>45.94</u>
<u>Within populations</u>	<u>73</u>	<u>36.71</u>	<u>0.503</u>	<u>54.06</u>
<u>Total</u>	<u>76</u>	<u>62.78</u>	<u>0.930</u>	

6.54. Among-Between-population genetic Differentiation

The amount of genetic differentiation can be seen from dendrogram produced using all individuals over the whole ~~data~~ data (Figure 7). Moderate to high genetic differentiation was seen, i.e., the dendrogram separated at least individuals from the two ~~regions~~ habitat conditions: from disturbed and undisturbed. Also individuals from the two populations of the ~~undisturbed~~ habits are separated from each other and form their own sub-clusters. The extent of genetic differentiation was also estimated using Shannon's diversity index as G'_{st} for each locus, Nei (1973) gene diversity estimate as G_{ST} for each locus and analysis of molecular variance (AMOVA) as F_{st} using the entire data set as well as only polymorphic loci.

G'_{ST} and G_{ST} measure the proportion of variation among populations relative to the total species diversity estimated as H_{SP} and H_T , respectively. The mean G_{ST} value was 0.436 over all the loci (Table 7) and 0.435 over the polymorphic loci only (Table 6). Eight of the polymorphic loci have a G_{ST} value of greater than 0.300 indicating greater differentiation among the populations (Table 6). The mean G'_{ST} values over polymorphic loci and over all the loci are 0.426 (Table 8) and

0.427 (Table 9), respectively. Ten (48%) of the polymorphic markers (48%) showed a G'_{ST} value of greater than 0.300, which is nearer to the values obtained from Nei (1973) statistics. The mean F_{ST} values over all the loci and over polymorphic loci are 0.459 and 0.469 (data not indicated), respectively.

6.65. Genetic Distance/Similarity between populations

From the analysis of Nei unbiased genetic distance/similarity among the four populations (Nei, 1978) Yayu-1 and Yayu-4 are the most genetically separated groups with (genetic distance of =0.109) as compared to the rest groups and Yayu-1 and Yayu-2 are the least genetically differentiated groups, with a genetic distance of 0.011) (Table 11 and Figure 8). ~~Yayu-1 and Yayu-2 are the least differentiated (0.011) populations and Yayu-1 and Yayu-4 are the most differentiated groups (0.109). Yayu-4 and Yayu-5 are also less differentiated with a genetic distance of 0.043.~~ The genetic differentiation among the four populations is summarized by a dendrogram generated after UPGMA based on the ~~original~~ unbiased Nei genetic distance/similarity matrix (Nei 198772) (Figure.8-7).

Table 11. Summary of ~~Nei's Unbiased measures of genetic Identity and Genetic distance/identity matrix between populations (Nei, 1978). Nei's genetic identity (above diagonal) and genetic distance (below diagonal)~~

Population	Yayu-1	Yayu-2	Yayu-4	Yayu-5
Yayu-1	****	0.989	0.891	0.921
Yayu-2	0.011	****	0.93	0.94
Yayu-4	0.109	0.070	****	0.957

<u>Yayu-5</u>	<u>0.079</u>	<u>0.060</u>	<u>0.043</u>	<u>****</u>
---------------	--------------	--------------	--------------	-------------

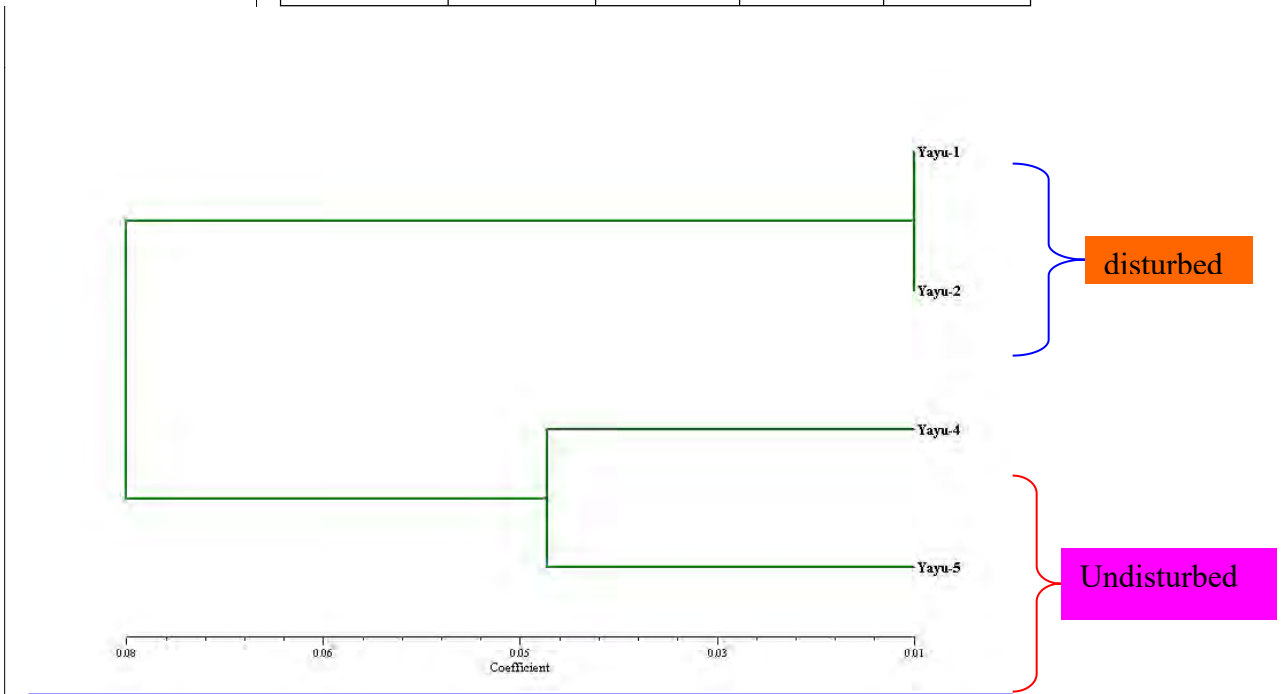


Figure 8. Dendrogram generated based on Nei unbiased genetic distance (Nei, 1978, 1987) and UPGMA for the four populations.

7. Discussion

7.12. The Use of ISSR markers in genetic Diversity

Molecular markers are being used widely for genetic variability studies and cultivar identification in a large number of species. ISSR is one of such molecular markers used in the analysis of variability at both intra- and inter-specific levels (Wolfe and Liston, 1998). Reliability and repeatability are essential features of a molecular technique to be used in fingerprinting. Techniques that use polymerase chain reaction (PCR) require high-level quality DNA to get clear and repeatable bands on the gel (Wolf and Liston, 1998). Many different steps are taken in different studies to get quality DNA. In the present study the use of triple extraction, a modified version of the CTAB method, for each sample and further cleaning with QiaGen kit enabled to

get pure DNA that can be used for ISSR analysis. More than 99% of the scored bands were reproducible. The high reproducibility of the ISSR fragments in this study and other studies in Coffee (e.g. Ruas, *et al.*, 2003; Esayas *et al.*, 2005; Kassahun *et al.*, in prepin press) as compared to RAPD analyses in most studies may be due to the use of longer primers, which allow more stringent (higher) annealing temperatures, than those of RAPD.

In the present study 38.18% of the bands scored from five di-nucleotide primers were polymorphic. ~~But the sample sites of Esayas et al., (2005) was not strictly similar with the sample sites of the present study.~~ This percent polymorphism is larger as compared to results obtained by Esayas *et al.* (2005) using 11 primers, partly due to differences in the sites used by the two studies.

7.2. Genetic Variation

The present study reveals ~~low to~~ moderate to high levels of genetic diversity as estimated by percent polymorphism (%P), Nei's gene diversity (H) (Nei, 1978), and Shannon diversity index (H'_{je}) over all the loci and only when polymorphic loci are considered. The genetic variability was high for Yayu-2 (P=27.27%, H=0.079, H'_{je} = 0.162) and lower for Yayu-4 (P=10.91%, H=0.031, H'_{je} =0.070). These results are comparable with what Esayas *et al.* (2005) found using the 11 primers for populations found in Illubabor ~~Zone. Esayas in Zone.~~ Again this comparison should be roughly, since they did not discriminate between disturbed and undisturbed regions in their analysis. Higher and lower ISSR variability was obtained for populations from disturbed and undisturbed regions of the forest. The total genetic diversity for the whole population or the species was (P=38.18%, H=0.100, H'_{spI} =0.223) and is comparable with what Kassahun *et al.*, in press found, using 10 primers and greater than what they found using only di-nucleotide primes

~~($P=20.5\%$, $H=0.07$, $H'spI=0.30$). The use of larger sample size in the present study could be why~~
~~A relatively larger or comparable genetic variability observed in the present study could be due to~~
~~the use of larger sample size is found in the present study.~~

The higher level of variability in the two populations from disturbed (managed) regions, ~~(yayu-1~~
~~and Yayu-2,)~~ of the forest ~~than from undisturbed regions~~ could be for two reasons. These two
populations are owned by the local people and are managed to some extent by thinning under
story small trees that compete with coffee and some times larger trees in the canopy ~~could also be~~
~~cleared~~ (cleared (Taddesse, 2003). Such managements could contribute to reduce genetic erosion
at early age (juvenile stage) of the coffee and result in higher genetic diversity. The second type
of management in these areas is the enrichment of the coffee trees by using seedlings from other
parts of the forest. ~~and may be also from the local landraces.~~ This seedling flow into the regions
could increase the genetic diversity of the two populations in the ~~disturbed regions~~. Generally, the
~~relatively low total genetic variability is expected for of~~ *C. arabica* ~~is~~ is small as it is a
predominantly self-pollinated and a recently evolved species (Lashermes *et al.*, 1996), by a rare
event of allopolyploidization. ~~This is further evidenced with the absence of variation of variability~~
~~chloroplast in chloroplast genome of~~ *C. arabica* ~~collected from wider geographic regions~~
(TefayeKassahun *et al.*, ~~Submit in~~ press). Although total variability was small because of large
number of monomorphic loci, about 48% of the polymorphic loci have higher total genetic
~~diversity~~ (diversity ($H_T > 0.25$) (Table .6) and this is greater than the result obtained by Anthony *et*
al. (2001) using 29 polymorphic RAPD markers.

7.3. Distribution of the Variability

Plant population structure in space and time is produced by the different evolutionary forces, which operate within historical and biological context of each species. Any factor that affects reproduction and dispersal is important in determining the genetic structure of a species (Loveless and Hamrick, 1984). Breeding system of a species was found to be the most important factor to determine genetic structure (Nybom and Bartish, 2000). Inbreeding species have lower within-population and larger among-population variability and out-crossing species have the reverse kind of genetic diversity distribution. Species with mixed-mating have a somewhat intermediate distribution of genetic diversity. Mixed-mating is known to cause more genetic variability within populations and is a potential for differentiation. The among-populations genetic structure depends on the level of selfing and may vary in time (Loveless and Hamrick, 1984).

Arabica coffee is known to be a predominantly self-pollinated species and is expected to have the pattern of inbreeding populations. Previous studies such as Esayas (2006) implied concordant data with some amount of gene flow between local populations and also between geographically separated populations of wild coffee in Ethiopian. The results of the present study showed that most of the diversity was distributed within populations rather than among populations. All the three estimates, G_{ST} (0.436), G'_{ST} (0.426) and F_{ST} (0.460), indicate larger within population than among-populations variability. The results indicate that about 54-57% of the variation exists within-populations. ~~-between-~~ Such results are typical of species with mixed-mating system and for which some extent of gene flow is expected (Loveless and Hamrick, 1984; Nybom and Bartish, 2000). The low out-crossing rate estimated for *C. arabica* (7-15 %) by Carvalho (1974)

was not confirmed in the study areas and the rate could be higher than what is expected from other studies conducted out of Ethiopia. The relatively low coefficient of population differentiation among populations could also be due to gene flow between the closely located areas by human and wild animals such as baboon, apes and others (Kassahun *et al.*, in press).

The individual level cluster analysis indicated that individuals from undisturbed regions of the forest (Yayu-4 and Yayu-5) formed their own clusters (Figure 7). Within this cluster they have their own sub-clusters that separate individuals from the two regions. The failure of individuals from the two disturbed plots to cluster on the basis of their respective populations could be due to the presence of substantial gene flow between local populations mainly in the form of seedling, seed and so on by human and other animals. Population-level genetic distance/similarity measures between every possible pairs of populations - and the resulting clustering also clearly separate populations from disturbed and an undisturbed plots of the forest (Table 11 and Figure 8). The smaller and the larger genetic distance values are between Yayu-1 and Yayu-2 (0.011) and between Yayu-2 and yayu-4 (0.109), respectively. In general, a relatively, larger genetic distance exists between pairs of populations from disturbed and undisturbed regions rather than between pairs from either disturbed or undisturbed regions indicating less genetic differentiation between plots under similar habitat condition.

The special genetic structuring of the plots in the undisturbed regions, which is indicated by the relatively larger genetic distance between them could be an indication that in the absence of management, populations will have their own genetic structure. ~~They are patches of have their~~ differentiated genetically genetically from those patches each other and also from populations in the disturbed - and less genetic differentiation observed between patches from the disturbed

~~regions-~~regions. This ~~could be indicates~~ partly due to the absence of human interference in the two populations and also because they are evolving independently. The fact that they have some unique markers could be an indication of this situation.

~~Yyure~~87

8. Conclusion and Recommendations

8.1. Conclusion

In spite of the smaller number of primers, ~~(only five dinucleotide primers,)~~ that is not as polymorphic as the tetra-nucleotide primers used in the present study, the genetic diversity

obtained for some of the populations was large ~~or~~ /moderate. This clearly indicates that one can use ISSR in the study of intraregional genetic structure of a population. Populations are clearly differentiated on the basis of management levels carried out in them. Thus, management difference could create genetic diversity ~~difference of forest coffee difference among the forest regions~~ under different management systems.

Yayu forest is a highly valuable forest ecosystem for biodiversity conservation. Coffee is found in many ~~regions~~ area of the forest. The abundance of coffee in the forest was found to be higher than that of any other ~~coffee~~ forest in Ethiopia. The genetic diversity of the wild coffee populations in this forest is found to be larger and deserve conservation for future utilization. ~~Low to m~~ Moderate levels ~~genetic~~ of genetic diversity estimates were obtained for populations in this study. The genetic structure obtained was typical of populations with mixed-mating rather than populations with purely inbreeding pattern although *C. arabica* is ~~kn~~ shown to be predominantly self-pollinating species. This deviation could be because of gene flow among the local populations by many agents such as human in the form of seedlings, some animals such as ~~as~~ Baboons ~~as~~ Baboons, apes, birds etc. in the form of seeds, and increased out-crossing rate among the plots studied by honey bees.

8.2. Recommendations: Implication for conservation and Breeding

8.2.1. Implication for conservation

Molecular tools do contribute to conservation efforts. Understanding the relative levels of within- and among -populations differentiation can help focus efforts on specific populations for conservation (Haig, 1998). In *in situ* conservation approach genetic diversity measurements are needed to designate the appropriate genetic reserve of target species with high variation (Tan and

Tan, 2002). The present study indicated higher ISSR diversity in populations from Yayu-1 and Yayu-2, i.e., in populations from disturbed regions. These populations and probably other populations in the managed areas of the forest need to be given priority for *in situ* conservation of wild *C. arabica* L. gene pool. Populations from undisturbed regions of the forest showed lower diversity indices, nevertheless, the strong genetic structuring observed between them and between populations in disturbed regions is an indication of the presence of unique individuals with unique genes in the populations. Thus, these populations also should deserve conservation attention. *Ex situ* conservation efforts for field gene bank collections should also give attention to these populations with higher variability.

From the fact that higher ISSR variability is observed in managed areas, it seems that management such as thinning of understory forests and enrichment of coffee with seedlings from other areas, would increase the genetic diversity of coffee in the forest. Thus, it is recommended that such managements be carried out in the forest for *in situ* conservation of wild coffee. However, care should be taken not to harm the forest ecosystem that would lead to loss of other organisms species and degradation of the natural forest ecosystem. An alternative approach to this one could be to establish coffee field gene bank in a nearby area and make collection of coffee in these field gene banks where management is possible. Such field gene bank would provide an environment that is nearly similar to the forest ecosystem and also provide give management possibilities options that would increase the genetic diversity of coffee. Seedlings could be collected from the undisturbed regions for replanting in the field gene banks.

8.2. 2. Implication for breeding

The present result would also help coffee breeders to focus on the populations with higher genetic variability to select parental coffee plants for crossing to maximize the level of variation present in segregating populations. ~~Esayasin-Here also~~ populations from managed areas need to be given priority. Again populations in the undisturbed areas should not be ignored for breeding since they showed some markers of their own and since such markers could code for agronomically important traits. ~~in this respect also.~~ Coffee productivity was found to be higher in hybrids, for example, 30 to 60% heterosis in hybrid over the better parent has been observed in Ethiopia for *C. arabica* L. (Mesfin, 1990).

8.8. 33. Future research directions

The genetic diversity of the forest coffee should be assessed in time series analysis to confirm the present genetic structure and to monitor any change in the structure. As one type of molecular marker is not sufficient for characterization of genetic diversity, the use of co-dominant DNA molecular markers such as simple sequence repeats and complementing it with some morphological studies is recommended. Demographic studies are also recommended to monitor any population number change and its impact on the magnitude of genetic variability of *C. arabica* populations in the forest. Since it also seems that the wild populations of *C. arabica* in Ethiopia have larger out-crossing rate than expected for world wide cultivars, it is recommended that the pollination biology of the species be included in future research.

9. References

- Abbot, P. (2001). Individual and population variation in invertebrates revealed by Inter-simple Sequence Repeats (ISSRs). *Journal of Insect Science*, 1.8.
- Anthony, F., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J. and Charrier, A. (2001). Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers *Euphytica*, **118**: 53-65.
- 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. *Theor. Appl. Genet.* **104**:894-900.
- Amsalu Ayana, A. (2001). Genetic diversity in sorghum (*Sorghum Bicolor*(L.) Moench) germplasm from Ethiopia and Eritrea. *Doctoral Thesis*, Addis Ababa University, Addis Ababa, Ethiopia, Pp.229.
- Berthou, F. (1983). Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. *Theor. Appl. Genet.* **65**: 77-84.
- Bornet, B., and Branchard, M. (2004). Use of ISSR fingerprints to detect microsatellites and genetic diversity in several related *Brassica* taxa and *Arabidopsis thaliana*. *Hereditas* **140**: 245-248.
- Bridson, D. M. and Verdcourt, B. (1988). *Flora of Tropical East Africa- Rubiaceae*. (Polhill, R. M., ed.), Part 2, Balkema, Brookfield, Rotterdam.
- Borsch, T., Hilu, K. W., Quandt, D., Wilde, V., Neinhuis, C., and Barthlott, W. (2003). Noncoding plastid trnT-trnF sequences reveal a well-resolved phylogeny of basal angiosperms. *J. evol. biol.* **16**: 558–576.

Carvalho, A., Ferwerda, F.P., Frahm-Leliveld, J.A., Medina, D.M., Mendes, A.J.T. and Monaco, L.C. (1969). Coffee (*Coffea arabica* L. and *C. canephora* Pierre ex Froehner). In: *Outlines of perennial crop breeding in the tropics* (Ferwerda, F. P. and Wit, F., eds.), Veenman and Zonen, Wageningen, pp. 189–241. Chang, R.Y., O'Donoghue, L.S. and Bureau, T.E. 2001. Inter-MITE.

Charrier, A., Berthaud, J. (1985). Botanical classification of Coffee. In: *Coffee: botany, biochemistry and production of beans and beverage*. (Cliffor M.N. and Wilson, K.C., eds.), Croom Helm, London, pp.13-47.

Coste, R. (1992). *Coffee: The plant and the product*. CTA Wageningen, Netherlands.

Demel Teketay (1999). *History, botany and ecological requirements of coffee*. Ethiopian Agricultural Research Organization, Addis Ababa, Ethiopia.

Demel Teketay, Paulos Dubale and Ababu Anage (1998). Report on the first field trip report of the team of experts organized to prepare a FCCP proposal. Mimeographed, Addis Ababa.

Devarumath, R. M., Nandy, S., Rani, V., Marimuthu, S., Muraleedharan, N. and Raina, S. N. (2002). RAPD, ISSR and RFLP fingerprints as useful markers to evaluate genetic integrity of micropropagated plants of three diploid and triploid elite tea clones representing *Camellia sinensis*(China type) and *C. assamica ssp. assamica*(Assam-India type). *Plant Cell Rep.***21**:166–173.

Dowling, T.E., Moritz, C., Palmer, J. D., and Reisberg, L.H. (1996). Nucleic acids III: Analysis of fragments and restriction sites. . In: *Molecular Systematics* (Hillis, D. M., Moritiz, C. and Mable, B.K., eds), SINAUERDITED, Sunderland, Massachusetts, U.S.A.

??

- Endashaw [Bekele \(1983\). Some measures of gene diversity analysis on landrace populations of Ethiopian barley. *Hereditas*, **98**:127-143.](#)
- Endashaw [Bekele \(1985\). The biology of cereal landrace population. *Hereditas* **103**: 119-134.](#)
- Endashaw [Bekele \(1986\). The interfaces of ecology, genetics, and biosystematics in conservation. *Symb. Bot. Ups. XXVI*, **2**: 101-113.](#)
- Esayas [Aga, \(2006\). Molecular Genetic Diversity Study of Forest Coffee Tree \[*Coffea arabica* L.\] Populations in Ethiopia: Implications for Conservation and Breeding. Doctoral thesis, Swedish University of Agricultural Sciences.](#)
- [Esayas Aga, and Bryngelsson, T. \(2005\). Inverse sequence-tagged repeat \(ISTR\) analysis of genetic variability in forest coffee \(*Coffea arabica* L.\) from Ethiopia. *Genetic Resources and Crop Evolution* \(in press\).](#)
- Esayas [Aga, a, E, Bryngelsson, T., Endashaw Bekele, and Salamon, B. \(2003\). Genetic diversity of forest coffee \(*Coffea arabica* L.\) in Ethiopia as revealed by random amplified polymorphic DNA \(RAPD\) analysis. *Hereditas*, **138**: 36-46.](#)
- [Esayas Aga, Endashaw Bekele, and Bryngelsson, T. \(2005\). Inter-simple sequence repeat \(ISSR\) variation in forest coffee tree \(*Coffea arabica* L.\) populations from Ethiopia. *Genetica*, **124**: 213-221.](#)
- [Excoffier, L., Smouse, P., and Quattro, J. \(1992\). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**:479-491.](#)
- [Excoffier, L. \(2006\). An Integrated Software Package for Population Genetics Data Analysis. Computational and Molecular Population Genetics Lab. \(CMPG\), Institute of Zoology, University of Berne, Bern, Switzerland.](#)

- Food and Agricultural Organization (1968). FAO Coffee mission to Ethiopia 1964-65. Report, FAO. ed, Rome (Italy), 200p.
- Friis, I.B. (1979). The wild populations of *Coffea arabica* L. and cultivated coffee. In: *Taxonomic aspects of African economic Botany* (ed., Kunkel,G.), *Proceedings of the IX plenary meeting of AETFAT, Las Palmas de Gran Canaria, 18-23 march, 1978.*
- Govers, K., Kassahu Tesfaye, and Borsch, T. (in prep): An ISSR marker system for measuring genetic diversity in *Coffea Arabica* L. *Plant Biology*.
- Haig,S.M.(1998). *Molecular contributions to conservation. Ecology* **79**: 413-425.
- Haarer, A. E. (1962). *Modern coffee production*, 2nd ed. London.
- IPGRI (1993). *Diversity and Development. The strategy of the International plant Genetic resources Insittute. IPGRI, Rome, Italy.*
- IPGRI (2005). *Crop Genetic Resources: An Economic Appraisal/EIB-2, Economic Research Service/USDA, IPGRI, Rome, Italy.*
- Jaccard, P. (1908). Nouvelles rescherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, **44**: 223-270.
- Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G. and Hodgkin, T. (1997). Molecular tools in plant genetic conservation: a guide to the techniques. International Plant Genetic Resources Institute Technical Bulletin, No. 2, pp. 5–47.
- Karp, A. (2002). The New Genetic Era: Will it Help us in Managing Genetic Diversity? In: *Managing Plant Genetic Diversity* (eds J.M.M. Engels, V. Ramanatha Rao, A.H.D. Brown and M.T. Jackson), CAB International / IPGRI.

Kassahun Tesfaye, Borsch, T., Govers, K. and Endashaw Bekele (in press). Characterisation of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. *Genome*.

Kassahun Tesfaye, Govers, K., Endashaw Bekele and Borsch, T. (in press). Genetic Diversity of *Coffea arabica* throughout its native range in Ethiopia based on ISSR fingerprint data. *Plant Genet. Resou. and Crop Evolution*.

Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M.C., Dussert, S., Graziosi, G., Trouslot, P. and Anthony, F. (2000). Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *Theor. Appl. Genet.* 100, 139–146.

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,” *16th Int. Sci. Colloq. On Coffee*, pp. 528–536, ASIC, Paris.

Lashermes, P., Combes, P. M. C., Robert, J., Trouslot, P., D’Hont, P. A., Anthony, F. and Charrier, A. (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet.* 26: 259–266.

Lashermes, P., Combes, M. C., Trouslot, P. and Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theor. Appl. Genet.* 94: 947-955.

Lashermes, P., Cros, J., Marmey, P. and Charrier, A. (1993). Use of random amplified DNA markers to analyse genetic variability and relationships of *coffea* species. *Genetic Resource and Crop Evolution*, 40: 91-99.

[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**:59-64.](#)

[Levinson, G. and Gutman, G.A. \(1987\). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**: 203–221.](#)

[Litt, M. and Luty, J.A. \(1989\). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401.](#)

[Loveless, M. D. and Hamrick, J. L. \(1984\). Ecological determinants of genetic structure in plant populations. *Ann.Rev. Ecol. Syst.*, **15**: 65-95.](#)

[Lynch, M. and Milligan, B.G. \(1994\). Analysis of population structure with RAPD markers. *Mol. Ecol.* **3**, 91–99.](#)

Mesfin Ameha and Bayetta Bellachew (1984). Resistance of the F1 to coffee berry disease in six parental diallel crosses in coffee Pp. 167-177. **In:** *Proc. 1st Reg. workshop on coffee berry disease*. 19-23 July, 1982 Addis Ababa.

Mesfin [Ameha \(1990\) . Heterosis and Arabica coffee breeding in Ethiopia. *Plant Breedin. Abstr.* **60**: 594–598.](#)

Mesfin Ameha (1991) . Significance of Ethiopian coffee genetic resources to coffee improvement. **In:** *Plant genetic resources of Ethiopia*, (Engels, J.M.M, Hawkes, J.G. and Melaku Worede, eds). Cambridge University Press, Cambridge, UK, Pp. 354-359.

[Meyer, F.G. \(1965\). Notes on wild *Coffea arabica* from Southwestern Ethiopia, with some historical considerations. *Econ. Bot.* **19**:136-151.](#)

- Morgante, M. and Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant J.* **3**: 75–182.
- Montagnon, C. and Bouharmont, P. (1996). Multivariate analysis of phenotypic diversity of *Coffea arabica*. *Genetic Resour. Crop Evol.* **43**: 221–227.
- Murphy, R. W., Sites, J. W., Donald, Jr., Buth, G., and Haufler, C.H. (1996). Proteins: Isozyme electrophoresis. **In**: *Molecular Systematics* (Hillis, D. M., Moritz, C. and Mable, B.K., eds), SINAUERDITED, Sunderland, Massachusetts, U.S.A.
- Nagaoka, T. and Ogihara, Y. (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet.* **94**: 597-602.
- Nei, M. (1972). Genetic distance between populations. *Amer. Natur.* **106**: 283-292.
- Nei, M., (1973). Analysis of gene diversity in subdivided Populations. *Proc. Natl. Acad. Sci. USA* **70**: 3321–3323.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.
- Nei, M. (1987). *Molecular Evolutionary Genetics*, Columbia University Press, New York, NY, USA.
- Nybom, H. and Bartish, I. (2000). *Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants*. Urban and Fischer Verlag, **3/2**: 93-114.
- Orozaco-Castillo, C., Chalmers, K.J., Waugh, R. and Powell, W. (1994). Detection of genetic diversity and selective gene introgration in coffee using RAPD markers. *Theor. Appl. Genet.* **87**: 934–940.

OXFAM (2002). *Bitter coffee: how the poor are paying for the slump in coffee prices*. Oxfam International, London.

Parker, P. G., Snow, A.A., Schug, M. D., Booaton, G. C. and Fuerst, P. A.(1998). *What molecules can tell us about populations: Choosing and using a molecular marker*. *Ecology* **79**: 361-382.

Paulos Dubale and Demel Teketay (2000). The need for forest coffee germplasm conservation in Ethiopia and its significance in the control of coffee diseases. In: *Proceedings of the Coffee Berry Disease workshop*, Ethiopian Agricultural Organization (EARO) Addis Ababa, Ethiopia, pp 125-135.

Pavlieek, A., Pavlieek, T. and Flegr, J. (1999). Free Tree-prerelease version. *Folia Biologica*, (Praha).

Purseglove, J. W. (1968). *Tropical crops: Dicotyledone 2*. Longmans, Green and Co.Ltd., London.

Raina, S.N., Mukai, Y. and Yamanoto, M. (1998). In situ hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae). *Theor. Appl. Genet.* **97**:1204-1209.

Ramesh, V., Kantety, Jeffrey L. Bennetzen and Zehr, B. E. (1995). Use of inter-simple sequence repeat (ISSR) amplification to detect multiple polymorphism in popcorn germplasm. Plant Genome III Conference, Town and Country Conference Center, San Diego, CA, January.

Rao, N. K. (2004). Plant genetic resources: Advancing conservation and use through biotechnology. *African J. of Biotech.* **3**:136-145.

Rolf, F. J. (1997). NTSYSpc v 2.11T Numerical Taxonomy and Multivariate Analysis System. Exter software. New York.

Ruas, P. M., Ruas, C. F., Rampim, L., Carvalho, V. P., Ruas, E. A. and Sera, T. (2003). Genetic relationship in Coffea species and parentage determination of interspecific hybrids using ISSR (Inter- Simple Sequence Repeat) markers. *Genet. and Molec. Biol.* **26**: 319-327.

[Sanchez, de la Hoz, M.P., Dávila, J.A., Loarce, Y. and Ferrer, E. \(1996\). Simple sequence repeat primers used in polymerase chain reaction amplifications to study genetic diversity in barley. *Genome* **39**: 112–117.](#)

Singh, P. (1996). *Essentials of plant breeding*, 1st ed, Kalayani Publishers; Ludhiana. New Delhi. Noida(U.P). Hyderabad. Madras. Culcutta. Cuttack

[Sneath, P. H. A. and Sokal, R. R. \(1973\). *Numerical Taxonomy*. Freeman. San Francisco. Pp. 573.](#)

[Sudupak, M. A. \(2004\). Inter and intra-species inter simple sequence repeat \(ISSR\) variations in the genus *Citrus*. *Euphytica* **135**: 229-238.](#)

[Sylvain, P.G. \(1955\). Some observation on *Coffea arabica* L. in Ethiopia. *Turrialba* **5**:37-52.](#)

[Tadesse Woldemariam Gole \(2003\). *Vegetation of the Yayu forest in sw Ethiopia: impacts of human use and implications for in situ conservation of wild *Coffea arabica* L.* Doctoral thesis, Center for Development Research\(ZEF\), Univeristy of Bonn, Bonn.](#)

[Tadesse Woldemariam Gole, Denich, M., Demel Teketay, and Borsch, T. \(2001\). Diversity of traditional coffee production systems in Ethiopia and their contribution to the conservation of coffee genetic diversity. Conference on International Agricultural Research for Development. Deutscher Tropentag, Bonn, 9-11 October.](#)

~~[Tadesse W/m Gole Tadesse, and Demel Teketay \(2001\)a. The forest coffee ecosystems: ongoing crisis, problems and opportunities for coffee gene conservation and sustainable](#)~~

utilization. Pp. 131-142. **In:** *Imperative problems associated with forestry in Ethiopia* (Ed. BSE) Biological society of Ethiopia, Addis Ababa.

Tadesse ~~W/m Gole~~Tadesse, Denich, M., Demel ~~Teketay~~ and ~~Teketay~~ and Vlek, P.L.G.(2002+)^b. Human impacts on *Coffea arabica* genetic pool in Ethiopia and the need for its *in situ* conservation. **In:** *Managing plant genetic diversity* (Engels, J., Ramanatha Rao, V., Brown, A. H. D. and Jackson, M., eds). CAB International / IPGRI.

~~Tadesse W/m Gole~~Tadesse, Denich, M., Demel Teketay, and Borsch, T. (2001) c. Diversity of traditional coffee production systems in Ethiopia and their contribution to the conservation of coffee genetic diversity. Conference on International Agricultural Research for Development. Deutscher Tropentag, Bonn, 9-11 October.

Tan, A. and Tan, A. S. (2002). *In situ* conservation of wild species related to crop plants: the Case of Turkey. **In:** *Managing Plant Genetic Diversity* (Engels, J. M. M., Rao, V.R., Brown, A. H. D., and Jackson, M. T. eds). IPGRI

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17, 6463-6471.

~~Kassahun Tesfaye, Borsch, T., Govers, K. and Endashaw Bekele (submitted) a:~~ Characterisation of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. *Genome*.

~~Kassahun Tesfaye, Govers, K., Endashaw Bekele and Borsch, T. (submitted) b:~~ Genetic Diversity of *Coffea arabica* throughout its native range in Ethiopia based on ISSR fingerprint data. *Plant Genet. Resour. and Crop Evolution*.

United Nations Conference on the Environment and Development (UNCED) (1992). Convention on Biodiversity. UNCED, Rio de Janero.

- Van der Graaff, N.A. (1981). Selection of Arabica coffee types resistant to coffee berry disease in Ethiopia. *Doctoral thesis*, University of Wageningen, Netherlands, pp.110.
- Van der Vossen, H.A.H. (1974). *Plant breeding. Coffee Research Foundation, Kenya, Annual Report 1973-74, pp. 40-51.*
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Plot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407-4414.
- ??Weber, J.L. and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388-396.
- Weir, B.S. and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**:1358-1370.
- Wellman, F.L. (1961). *Coffee: botany, cultivation and utilization*. New York.
- Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**: 7213-7218.
- 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 Res.* **18**: 6531-6535.
- Workafes Woldetsadik and Kassu Kebede (1999). Coffee production systems in Ethiopia. *Proc. Of the workshop on Control of coffee Berry Disease (CBD) in Ethiopia*, Addis Ababa.
- Wolfe, A.D. and Liston, A. (1998). Contributions of PCR-based methods to plant systematics and evolutionary biology. In: *Molecular Systematics of Plants* (eds, Soltis, D. E, Soltis, P.S. and Doyle, J.J.), PP.43-86. Kluwer Academic Publishers, Boston.

[Yeh, F.C., Yang, R.C. and Boyle, T. \(1999\). Popgene program. Version 1.31.](#)

[Zietkiewicz, E., Rafalski, A. and Labuda, D. \(1994\). Genome fingerprinting by simple sequence repeat \(SSR\)-anchored polymerase chain reaction amplification. *Genomics* 20: 176–183.](#)

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in any other university and that all sources of material used for the thesis has been duly acknowledged.

Tamiru Oljira _____

July 2006

Addis Ababa, Ethiopia

