Molecular Genetic Diversity Study of Forest Coffee Tree [*Coffea arabica* L.] Populations in Ethiopia: Implications for Conservation and Breeding

Esayas Aga
Faculty of Landscape Planning, Horticulture and Agricultural Science
Department of Crop Science
Alnarp

Doctoral thesis
Swedish University of Agricultural Sciences
Alnarp 2005

Acta Universitatis Agriculturae Sueciae
2005: 79

ISSN 1652-6880
ISBN 91-576-6978-3

Tryck: Repro, Alnarp 2005

Abstract

Aga, E. 2005. Molecular genetic diversity study of forest coffee tree [*Coffea arabica* L.] populations in Ethiopia: Implications for conservation and breeding. Doctoral dissertation ISSN 1652-6880, ISBN 91-576-6978-3

Coffee provides one of the most widely drunk beverages in the world, and is a very important source of foreign exchange income for many countries. *Coffea arabica*, which contributes over 70 percent of the world's coffee productions, is characterized by a low genetic diversity, attributed to its allopolyploidy origin, reproductive biology and evolution. *C. arabica* has originated in the southwest rain forests of Ethiopia, where it is grown under four different systems, namely forest coffee, small holders coffee, semi plantation coffee and plantation coffee. Genetic diversity of the forest coffee (*C. arabica*) gene pool in Ethiopia is being lost at an alarming rate because of habitat destruction (deforestation), competition from other cash crops and replacement by invariable disease resistant coffee cultivars. This study focused on molecular genetic diversity study of forest coffee populations in Ethiopia using PCR based DNA markers such as random amplified polymorphic DNA (RAPD), inverse sequence-tagged repeat (ISTR), inter-simple sequence repeats (ISSR) and simple sequence repeat (SSR) or microsatellites. The objectives of the study are to estimate the extent and distribution of molecular genetic diversity of forest coffee and to design conservation strategies for it's sustainable use in future coffee breeding. In this study, considerable samples of forest coffee collected from four coffee growing regions (provinces) of Ethiopia were analysed. The results indicate that moderate genetic diversity exists within and among few forest coffee populations, which need due attention from a conservation and breeding point of view. The cluster analysis revealed that most of the samples from the same region (province) were grouped together which could be attributed to presence of substantial gene flow between adjacent populations in each region in the form of young coffee plants through transplantation by man. In addition wild animals such as monkeys also play a significant role in coffee trees gene flow between adjacent populations. The overall variation of the forest coffee is found to reside in few populations from each region. Therefore, considering few populations from each region for either *in situ* or *ex situ* conservation may preserve most of the variation within the species. For instance, Welega-2, Ilubabor-2, Jima-2 and Bench Maji-2 populations should be given higher priority. In addition, some populations or genotypes have displayed unique amplification profiles particularly for RAPD and ISTR markers. Whether these unique bands are linked to any of the important agronomic traits and serve in marker assisted selections in future coffee breeding requires further investigations.

Keywords: Breeding, Conservation, Ethiopia, forest coffee, genetic diversity, molecular markers

Author's address: Esayas Aga, SLU Department of Crop Science, Sundsvägen 14, P.O. Box 44, SE-230 53 Alnarp, Sweden. E-mail: esayas.aga@vv.slu.se

Dedicated to the memory of my late parents: ***Aga Guda*** *&* ***Shume Gamada***

"The more you know, the more you realize how the little you know"

Contents

Appendix

Papers I-IV

This thesis is based on the following papers, which will be referred to by their Roman numerals:

I. Aga, E., Bryngelsson, T., Bekele, E. & 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*.

II. Aga, E. & 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*).

III. Aga, E., Bekele, E. & Bryngelsson T. (2005). Inter-simple sequence repeat (ISSR) variation in forest coffee tree (*Coffea arabica* L.) populations from Ethiopia. *Genetica 124, 213-221*.

IV. Aga, E. & Bryngelsson, T. Genetic variability of forest coffee tree (*Coffea arabica* L.) populations from Ethiopia estimated by microsatellite markers (*submitted*).

Introduction

Some geographical features of Ethiopia

Geographically, the country is located on the horn of Africa and extends from 3° to 15° N latitude, and 33° to 48° E longitude. The heartland, which covers nearly two-thirds of the country, is a high plateau with a mean altitude of 2000 to 2500 meters above sea level. Mt. Ras Dashen, at an altitude of 4620 meter, is the fourth highest mountain in Africa. The Great Rift Valley bisects the country in a northeasterly direction from Lake Rudolf on the southwest to the Danakil Depression in the northeast near the Red Sea. Danakil Depression is the lowest sink in Africa, which lies at 125 meters below sea level. The high plateau of Ethiopia is largely treeless, except for limited areas of forest in canyons and river valleys. Already more than 90% of the forest cover of Ethiopia has vanished due to timber cutting for construction purposes and increased land requirement for cultivation of food crops. At present a belt of mountain rain forests are limited mostly to the southwest and southeast regions. In altitude, the rain forests range from approximately 1000 to 2000 meters above sea level. Geographically, the southwestern Ethiopian rain forest is a disjunction from the broad evergreen forest belt of the Congo basin by the valley of the White Nile (Meyer, 1965).

The Coffee tree: a botanical overview

Coffee belongs to the genus *Coffea* in the family Rubiaceae, and is mostly grown in tropical and subtropical regions (Berthaud & Charrier, 1988; Bridson & Vercourt, 1988). About 100 species of the genus *Coffea* have been identified so far (Bridson & Verdcourt, 1988), but commercial production relies only on two species, *Coffea arabica* L and *Coffea canephora* Pierre ex Froehner, which represent about 70% and 30% of the total coffee market, respectively. Coffee provides one of the most widely drunk beverages in the world, and is a very important source of foreign exchange for many countries. *C. arabica* is the only allopolyploid (2n = 4x = 44) coffee species and self-fertile at approximately 90%. The other *Coffea* species are diploid (2n = 2x = 22) and self-sterile except for C. *heterocalyx* and C. sp. Moloundou, which are diploid but self-fertile (Lashermes *et al*., 1999; Coulibaly *et al*., 2002).

Allopolyploidy origin of *C. arabica*

Coffea arabica, the only tetraploid species in the genus *Coffea*, is indigenous to the highlands of southwestern Ethiopia and southeastern Sudan. *C. arabica* follows one of the typical patterns of distribution of polyploids with peripheral expansion outside the

range of distribution of the other diploid species of the genus. An allopolyploid is an organism containing separate sets of non-homologous chromosomes due to hybridization between different species. *C. arabica* could have arisen from natural hybridization between two ancestral diploid coffee species followed by unreduced gamete formation. *C. eugenioides* and *C. canephora* (or *C. liberica* or *C. congenesis*) have often been assumed to be the ancestral parents of *C. arabica* (Narasimhaswamy, 1962; Cramer, 1957). However, meiotic pairing of chromosomes of *C. eugenioides* and *C. canephora* in interspecific hybrids was observed to be better than in dihaploid plants of *C. arabica*. Species very closely related or identical to *C. eugenioides* and to *C. canephora* (or *C. congensis*) are indeed the most likely maternal and paternal progenitors of *C. arabica*, respectively (Lashermes *et al*., 1995). Raina *et al*. (1998), arrived at similar conclusions in a cytogenetic study of *C. arabica* using genomic and fluorescent *in situ* hybridisation techniques.

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*. Such floral morphology 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 have proved to be highly self-incompatible, and are allogamous (out crossing). 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 (Urbaneja *et al*., 1996).

Figure 1. Typical coffee tree with clusters of fruits. Note the vertical and lateral branches (http://www.coffeeresearch.org).

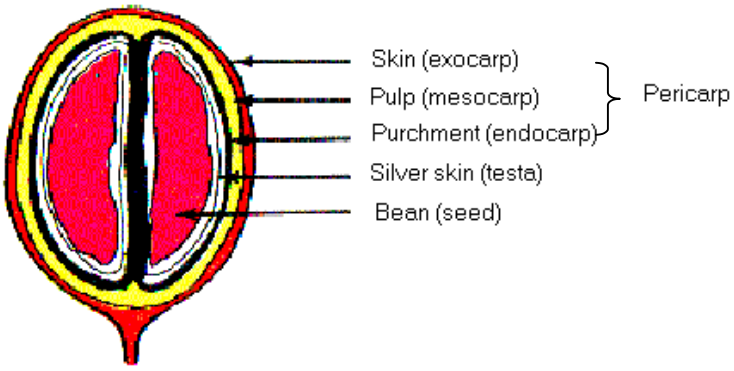

Figure 2. A cross section of a coffee fruit (http://www.coffeeresearch.org).

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).

Center of origin and distribution of *C. arabica*

There is every ground for the belief that Ethiopia is the original home of *C. arabica*, which is confirmed by the fact that within small areas four or more genetically different types of coffee trees can be found, and *Coffea aethiopica* would therefore have been a more correct name than *Coffea arabica* (Strenge, 1956). The first belief that this plant was of Arabic origin was due to the fact that the first knowledge of the beverage and the tree was obtained from Arabia, hence the scientific name given by Linnaeus (Sylvain, 1958). Yemen was almost the sole source of coffee germplasm over most of the recorded history of *C. arabica*. Yet, there is no reason to suggest that *C. arabica* is indigenous to Yemen. Yemen, as part of old Arabia Felix, is largely a desert area with high mountains in the center of the country where coffee is grown on terraces and watered from wells (Sylvain, 1955). The *C. arabica* plant is a rain forest species with a tolerance to grow in a wide range of climatic and ecologic conditions. The date when coffee was introduced into Yemen from Ethiopia is unknown. Yet no other country of Africa seems to be associated with the history of the *arabica* coffee plant, although the closest relatives of *C. arabica* are definitely tropical African (Meyer *et al*., 1968).

All botanists, who have explored the forests on the south-western highlands of Ethiopia, agree in there observation that this is the center of diversity of *C. arabica*, but that it is very difficult to find truly wild populations (Sylvain, 1955; Strenge, 1956, Meyer 1965). On the other hand, Thomas (1942) found wild populations of *C. arabica* in secondary forest on the Boma plateau in southeastern Sudan. Berthaud & Charrier (1988) also reported the presence of *C. arabica* populations on Mount Imantong in Sudan and Mount Marsabit in Kenya. It is not clear whether these *C. arabica* trees are truly wild or in earlier times brought from Ethiopia by man. Self-compatibility would have provided opportunity for quick occupation of the new region away from the original range of distribution. Besides plantation of young coffee plants in coffee tree deprived forests by man, the expansion of the coffee plant in Ethiopia seems to be accomplished with the help of animals, among which the baboons and monkeys are important distribution agents.

Coffea arabica has been cultivated in Yemen for at least five centuries but spread to South East Asia about 1700. In the early 18th century, progenies of a single plant from Indonesia, cultivated in Amsterdam and Paris, were spread to Latin America. Other introductions followed in the late 18th century from Yemen to Brazil, via Bourbon Island (now Réunion) (Haarer, 1956). These base populations gave rise to many cultivars and were described as two distinct varieties commonly called *typica* and *bourbon*, respectively (Carvalho *et al*., 1969).

The significance of Ethiopian Coffee (*C. arabica*)

Coffee accounts for over 60% of foreign currency income in Ethiopia. Its production is estimated to be between 250,000 and 300,000 tones per a year, out of which over a half is consumed locally. Generally, breeding programs of *C. arabica* are limited by very narrow genetic bases, especially in countries outside of its origin in improving pest and disease resistance (Van der Vossen, 1985). A genetic alternative to breeding limitations exists in the spontaneous and subspontaneous genotypes collected in the center of origin of the species, the northeast African highlands (Anthony *et al*., 2001). Since 1928, several teams and individual researchers have visited Ethiopia to collect coffee seed material and study the available variability. Two extensive collecting trips were later organized in Ethiopia, in 1964-65 by the FAO and in 1966 by ORSTOM (now IRD) with an emphasis on spontaneous and sub-spontaneous material (Meyer *et al*., 1968). This material was established in field gene banks in several African and Latin American countries. Many characteristics of agronomic interest were observed in Ethiopian coffee, such as the incomplete resistance to orange leaf rust *Hemileia vastatrix*, resistance to the nematode *Meloidogyne incognita* and to coffee berry disease caused by *Colletotrichum kahawae* (Van der Vossen, 2001). Ethiopian coffee is also a potential source of caffeine free *C. arabica* genotype, which is an alternative to transgenic caffeine free coffee plants proposed by Moisyadi *et al*. (1999). In spite of these facts, relatively few investigations have been carried out on forest coffee populations in Ethiopia. Moreover, the use of these genetic resources has been hampered by the limited knowledge of the structure of genetic diversity of this species in its native habitat (Anthony *et al*., 2001).

Cultivation of coffee (*C. arabica*) in Ethiopia

Coffea arabica is the only representative of the genus *Coffea* in Ethiopia. Coffee is generally grown on deep reddish-brown clay soils, which are derived from volcanic outflows and are slightly acidic with pH values from 4.5 to 6 (Van der Graaf, 1981). Practically all coffee appears to be under some form of cultivation.

In Ethiopia, coffee is grown under four different systems: 1) Forest coffee which is some times referred to as "wild" coffee accounts for 60% of coffee production in which self sown seedlings have been transplanted irregularly in the forest. Planting is done with spontaneous seedlings, which are pulled out and planted in a hole made with stick. Seedlings are also bought from traders who carried them bare-rooted on horse or donkey back. 2) Small holder coffee accounts for 37% of coffee production, and is plots of varying sizes around dwellings. 3) Semi plantation coffee is coffee plants in which seedlings raised in nurseries are planted more or less regularly in thinned forest. 4) Plantation coffee is established on previously cleared land in which seedlings are raised in nurseries and regularly planted together with shade plants (Van der Graaff, 1981). *C. arabica* grows in many parts of Ethiopia. The main cultivation is limited to the southern and southwestern regions. The crop is often picked only once and green, ripe, and overripe cherries are stripped off. The berries are sun dried, often on the ground. Improved practices such as drying on concrete and wet processing are being gradually adopted.

The rain forest habitat of forest coffee (*C. arabica*) in Ethiopia

The coffee rain forest in Ethiopia lie between latitude 6° and 9° N, and longitude 34° and 40° E. Already nine–tenth of the forest cover of Ethiopia has vanished due to timber cutting for construction purposes and increased land requirement for cultivation of food crops. At present a fragment of mountain rain forests is limited mostly to the southwest (Welega, Ilubabor, Jimma, Kafa, Bench Maji) and southeast (Bale) administrative zones.

In altitude, the rain forest ranges from approximately 1200 to 2100 meters above sea level. The Ethiopian rain forest consists of four stories of vegetation, three woody-emergent, canopy, and shrub plus herbaceous layer (Meyer, 1965). The coffee plant is a component of the shrub under-story, which is regularly harvested by the local people who usually carry out some form of maintenance by clearing the bush to facilitate picking of the ripe cherries. Both inside and outside of Ethiopia the term forest (wild) coffee is used indiscriminately. It is often practically impossible to distinguish old plantations abandoned for years from coffee growing spontaneously. However, the presence of a large number of old plantations does not preclude the existence of spontaneous formations. If coffee is found more often associated with secondary growth forest, it may be due to the custom existing in some places of removing part or the entire upper story of the forest on the assumption that yields may be thereby increased. A thorough study of some of the original forests would probably reveal the existence of true wild coffee trees but perhaps only in small numbers (Sylvain, 1955).

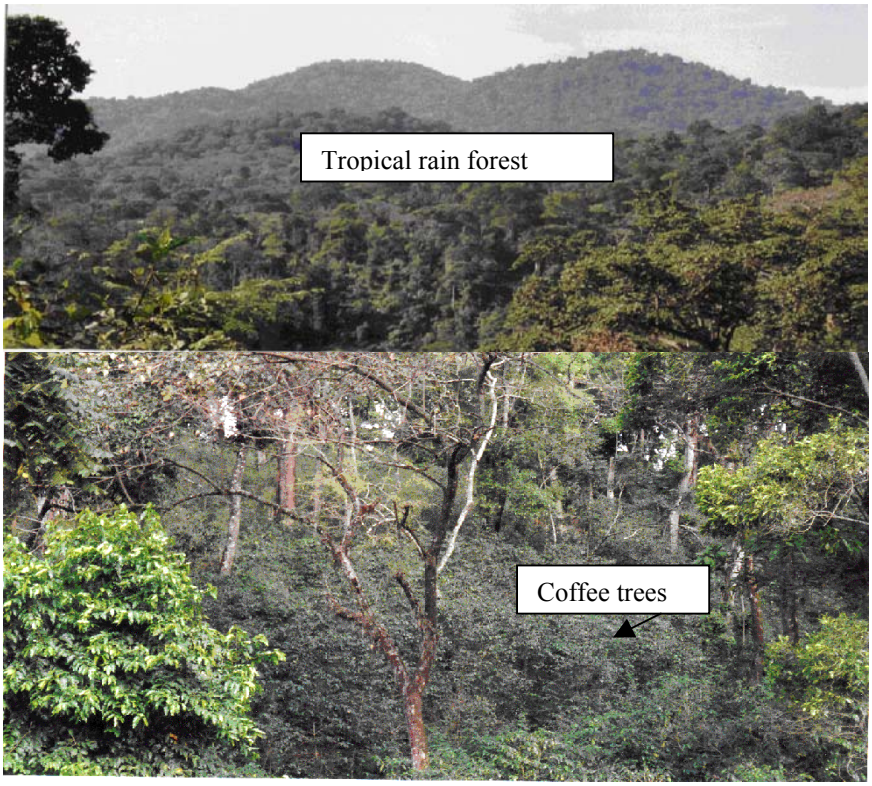

Figure 3. A rain forest area in the southwest Ethiopia. Note the lower story coffee trees.

Genetic diversity of forest coffee (*C. arabica*) in Ethiopia

The variability in natural populations of *C. arabica* has been apparent to most botanists and geneticists, who visited and explored the southwestern highlands of Ethiopia. Morphological variability exists within locations and between regions. For instance, the ratio of trees with different leaf tip colour (green or bronze) varies between locations. *C. arabica* is characterized by low genetic diversity attributed to its narrow genetic base, allopolyploidy, evolution and the predominantly self-pollinating nature of the species (Lashermes *et al*., 2000). However, a higher level of genetic variability with molecular markers was observed among spontaneous and sub spontaneous accessions of this species collected from Ethiopia (Lashermes *et al*., 1996a; Anthony *et al*., 2001). The existence of two subgroups of partial genetic differentiation within germplasm of *C. arabica* into accessions collected from east (Kaffa, Ilubabor, Welega) and west (Sidamo, Harrarge) of Great Rift Valley was established by an analysis with molecular markers (RAPD) (Lashermes *et al*., 1996a) and also by a multivariate analysis of phenotypic characters (Montagnon & Bouharmont, 1996). In this perspective it would appear that the coffee cultivated in Yemen,

from where almost all cultivated *C. arabica* derive, had its origin in Ethiopia east of the Rift Valley (Van der Vossen, 2001).

Methods of detecting genetic diversity

Morphological characters

Morphological characteristics were among the earliest genetic markers used for assessment of variation and are still of great importance. Usually, these characters are inexpensive and simple to score. 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, 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). Besides, assessment of morphological characters in perennial plants such as coffee, often require a lengthy and expensive evaluation during the whole vegetative growth.

Molecular markers

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. 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.

Protein-based molecular markers

Both enzymes and non-enzyme proteins have been used as genetic markers. 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.

Different forms of an enzyme may be expressed either by different loci or by different alleles at the same locus. The term allozyme has therefore been accepted to describe forms of an enzyme associated with different alleles at the same locus and isozyme to describe different forms from distinct loci. Differences 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. Protein based markers have the advantages that they are cheap and need no sophisticated equipment; they are usually co-dominant making them appropriate for heterozygocity 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 polymorphism. In coffee, an attempt to characterize *C. arabica* germplasm using a number of enzyme systems revealed no polymorphism (personal observation).

DNA-based molecular markers

Plant DNA polymorphism assays are powerful tools for characterizing and investigating germplasm resources and genetic relatedness (Powell *et al*., 1996). These 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. 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), 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 polymorhism or to a change in fragment size (Weising *et al*., 2005).

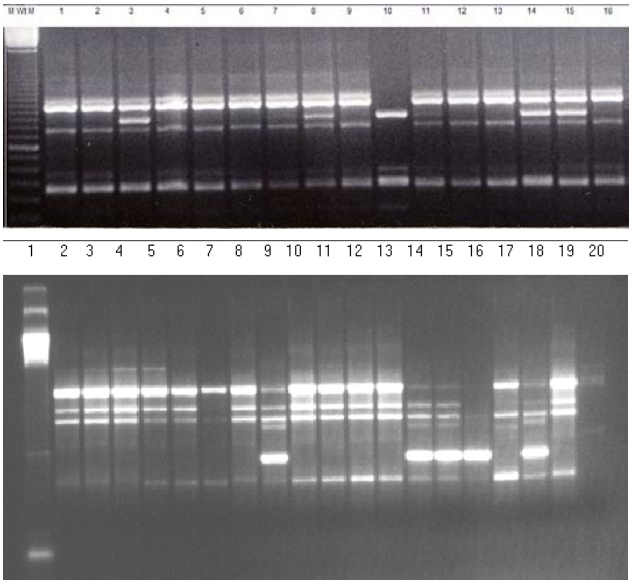

Figure 4. RAPD primers OPM4 (upper panel) and OPC15 (lower panel) banding patterns of forest coffee DNA samples.

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).

Inverse sequence-tagged repeats (ISTR)

In addition to simple sequence repeat motifs, plant and animal genomes contain extended repetitive elements, many of which are mobile genetic elements (transposons) that are capable to change their position within the genome. According to the mechanism of transposition, mobile genetic elements of eukaryotes can be divided into class I transposons (retrotransposons) capable of transposition via an RNA intermediate, and class II transposons capable of propagating via a DNA intermediate (Weising *et al*., 2005).

Among these transposable elements, the retrotransposons constitute the largest group, and are further grouped into two main classes depending on the presence or absence of flanking long terminal direct repeats (LTR) (Rohde, 1996). The LTR retrotransposons are characterized by the presence of about 300-500 base pair long direct repeats at both ends of the element. These LTR are highly conserved, and are exploited for primer design in the development of retrotransposon-based markers. A wide variety of molecular marker techniques use PCR primers directed toward transposable elements (Flavell *et al*., 1998; Kalendar *et al*., 1999; Chang *et al*., 2001; Edwards *et al*., 2002). One of these techniques include inverse sequence-tagged repeat (ISTR) (Rohde, 1996). ISTR markers are universally applicable by using identical primers derived from coconut copia-like sequences in the fingerprinting of eukaryotic genomes (Rohde *et al*. 1995; Rohde, 1996).

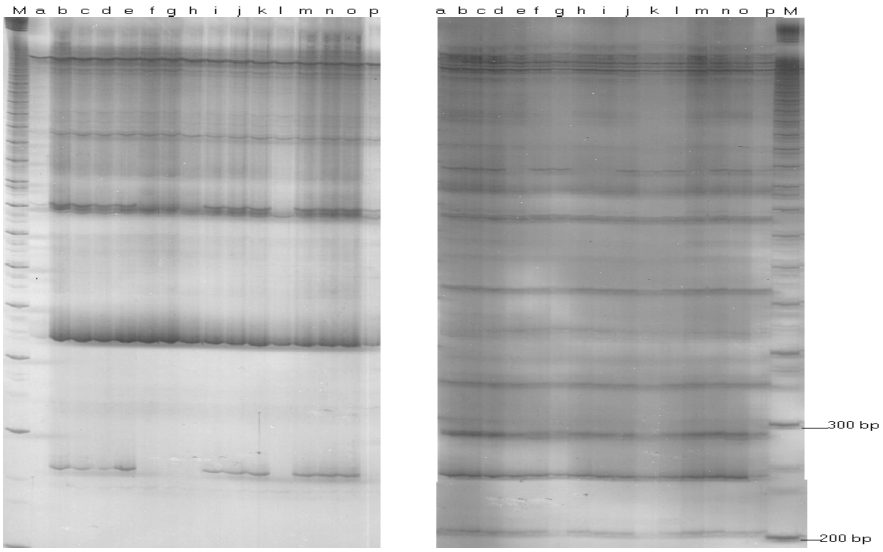

Figure 5. Banding pattern of forest coffee DNA samples amplified with ISTR primers F5/B10 (left panel) and F8/B8 (right panel).

The technique is composed of forward and backward primers of 15 to 21 base pairs in length. ISTR markers are comparable to AFLP markers with respect to the number of loci detected and their resolving power (Rohde, 1996).

Inter-simple sequence repeat (ISSR)

Inter simple sequence repeat (ISSR) is a PCR based method, which involves amplification of DNA segments present at an amplifiable distance between two identical microsatellite repeat regions oriented in opposite direction. The technique uses microsatellites, usually 16–25 base pair long primers in a single primer PCR reaction targeting multiple genomic loci to amplify mainly the inter-SSR sequences of different sizes. The primers used can be either unanchored (Gupta *et al*., 1994; Wu *et al*., 1994) or anchored at the 3' or 5' end with 1 to 4 degenerate bases extended into the flanking sequences (Zietkiewicz *et al*., 1994). The microsatellite repeats used as primers can be di-nucleotide, tri-nucleotide, tetra-nucleotide or penta-nucleotide. 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.

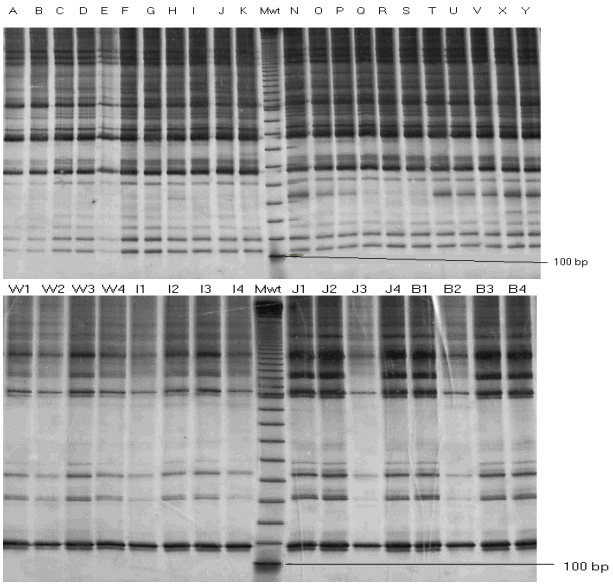

Figure 6. Banding patterns of forest coffee DNA samples amplified with ISSR primers UBC841 (upper panel) and UBC834 (lower panel).

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 is simple, quick, and efficient. The primers are long resulting in high stringency and hence reproducibility. The amplified products are usually 200-2000 bp long and amenable to detection by both agarose and polyacrylamide gel electrophoresis. 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).

Microsatellites (simple sequence repeats)

Microsatellites or simple sequence repeats (SSRs) 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).

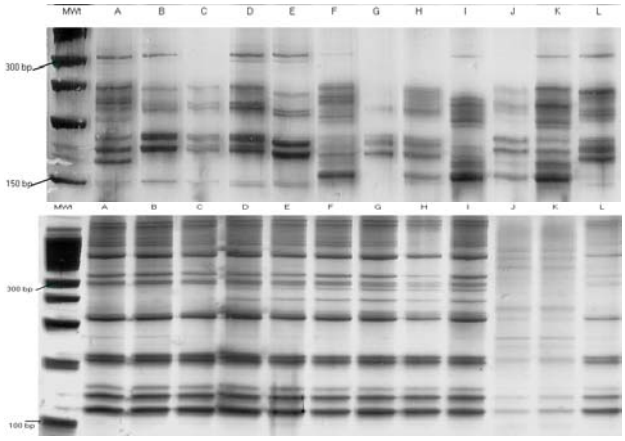

Figure 7. SSR primers M_{24} (upper panel) and M_{42} (Lower panel) banding patterns of forest coffee DNA samples.

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).

The background and objectives of the study

Background

Coffee is an important agricultural export commodity in more than 50 developing countries of Africa, Asia and Latin America (Rani *et al*., 2000; Orozco-Castillo *et al*., 1994; Dublin *et al*., 1991). In Ethiopia coffee (*C. arabica*) is a major source of foreign currency. More than 60% of the country's foreign exchange income comes from this single commodity (Belachew, 1997). Commercial coffee production relies on two species of coffee (*C. arabica* L. and *C. canephora* Pierre). Higher quality coffee is associated with *C. arabica*, which contributes over 70 percent of the world's coffee production. *C. arabica* is characterized by a low genetic diversity, which is attributable to its allotetraploid origin, reproductive biology (self compatibility) and evolution. Its highest genetic diversity was observed among spontaneous and sub-spontaneous sample collections from Ethiopia. It is generally believed that Ethiopia is the center of origin and primary center of genetic diversity of *C. arabica*. Information on crop plant species genetic diversity in their center of origin and of the relationships among elite breeding material has a significant impact on the improvement of crop plants (Hallauer *et al*., 1988).

Many characteristics of agronomic interest were observed in Ethiopian coffee, such as the incomplete resistance to orange leaf rust (*Hemileia vastatrix*), resistance to the nematode (*Meloidogyne incognita*) and to coffee berry disease caused by *Colletotrichum kahawae* (Van der Vossen, 2001). Ethiopian coffee is also a potential source of caffeine free *C. arabica* genotypes, which is an alternative to transgenic caffeine free coffee plants proposed by Moisyadi *et al*.,

1999). However, the use of these genetic resources has been hampered by the limited knowledge of the structure of genetic diversity in its native habitat (Anthony *et al*., 1995). Besides, during the last 40–50 years, a significant reduction of genetic diversity has occurred in the Ethiopian coffee due to deforestation, competition from other cash crops and replacement of the coffee trees by invariable disease resistant cultivars (Ameha, 1990). Therefore, the genetic diversity of *C. arabica* in Ethiopia requires a systematic approach by taxonomists and geneticists in order to develop a practical classification and conservation strategy before it is totally lost (Ameha, 1991).

Estimation of genetic diversity in cultivated crops is essential for breeding programs and for the conservation of genetic resources (Soleimani *et al*., 2002). All genetic resource conservation activities require characterization of the diversity present in both the gene pools and the gene banks (Karp *et al*., 1997). Assessment of the genetic diversity in *C. arabica* is, therefore, of crucial importance for developing conservation strategies for this economically important crop species. In the first instance, this usually involves description of variation for morphological traits. However, morphological variation is often found to be restricted and genotype expression is often influenced by environmental conditions, thereby, constraining the analysis of variation. These limitations have resulted in the deployment of biochemical techniques such as isozyme and protein electrophoresis (Hunter & Markert, 1957) and molecular techniques that analyse polymorphism at the DNA level directly.

With the development of the polymerase chain reaction (PCR), many molecular techniques have been, and still are being developed for plant genome analysis. This study focused on forest coffee populations in Ethiopia to determine the extent and distribution of its genetic diversity using PCR based DNA marker techniques such as random amplified polymorphic DNA (RAPD), inverse sequence-tagged repeat (ISTR), inter-simple sequence repeats (ISSR) and simple sequence repeat (SSR) or microsatellites. These molecular marker techniques are relatively simple and cost effective that can screen a larger portion of the genome, and are thus promising marker types for the identification of variation among the forest coffee samples collected from Ethiopia. The information will be used in designing conservation strategies and identification of genetically diverse populations in order to cross genetically diverse parents and maximize the level of variation present in segregating populations.

Objectives

The specific objectives of this study are:
1) To estimate the extent and distribution of molecular genetic diversity of forest coffee (*C. arabica* L.) in Ethiopia.
2) To disseminate preliminary information on molecular genetic diversity of forest coffee (*C. arabica*) in Ethiopia.
3) To determine genetic relationships between genotypes of various coffee populations.
4) To design conservation strategies which could act as reservoirs of coffee germplasm for further selection of genotypes in future coffee breeding.

Materials and methods

Plant material

The coffee germplasm material used in this study were collected from southwest (Welega, Ilubabor, Jimma, Kafa and Bench Maji) regions and southeast (Bale) region of Ethiopia during the period of November 1999 to January 2000, and January to February 2004. In each population seed samples were collected randomly from 20 individuals across a transverse pass through an area of approximately 400–900 m^2 of the forest coffee. The seed samples were sown in the greenhouse and leaf samples from young coffee plants were used for genomic DNA extraction.

Research methodologies and data analysis

Research methodologies

Polymerase chain reaction (PCR) based DNA markers such as RAPD (randomly amplified polymorphic DNA), ISTR (inverse sequence tagged repeat), ISSR (inter-simple sequence repeat) and microsatellites or SSR (simple sequence repeat) were chosen to investigate the extent of genetic variability within and among the forest coffee sample populations under study. The selections of these marker types were based on their relative technical simplicity, level of polymorphism they detect and cost effectiveness. The technical details of the markers used are illustrated in papers I-IV, respectively.

The development of the polymerase chain reaction (PCR) was a technological breakthrough in genome analysis since it enabled the amplification of specific fragments from the total genomic DNA. The principle of PCR is based on the function of a copying enzyme, DNA polymerase, which is able to synthesize a duplicate molecule of DNA from a DNA template. The product of duplication of the

original template DNA becomes a second template for another round of duplication. Repeated duplications thus lead to an exponential increase in DNA product accumulation. The DNA target is defined by primer annealing sites. Primers are short stretches of single stranded DNA, which are complementary to the opposite ends of the sequence of the target DNA. Since both strands of a DNA molecule run in opposite orientation, the primer sequences point to each other. The usual distance between the priming sites is between 100 base pairs and a few kilo-bases.

Amplification from virtually any region of a DNA molecule is possible by selecting specific sequences as primers at both flanks of the target region as in the case of ISTR and microsatellite markers (papers II & IV). For such direct-targeted PCR the sequences of these flanks must be known. Whether or not a unique and specific product is obtained depends on the selectivity of the primers that are designed based upon the sequences flanking the target. PCR reactions can also be carried out with single primers, which have not been designed on a known target flanking sequence as in the case of RAPD (paper I). In this case, amplification will occur wherever the primer is able to anneal to complementary sequence within the genome. Data were scored as present (1) and absent (0) of a particular character state (amplification product).

Data analysis

It is imperative to understand the different ways that the data generated by molecular techniques be analysed before considering their application to diversity studies (Hillis & Moritiz, 1990; Soltis *et al*., 1992; Avise, 1994). Two main types of analysis are relevant to diversity studies. 1) Analysis of genetic relationships (genetic distance or similarity) among samples or populations. 2) Calculation of population genetic parameters, in particular population diversity and its partitioning at different levels (Karp *et al*., 1997).

The analysis of genetic relationships among samples or populations starts with the construction of a matrix specifying the character-state of each marker for each sample or population. In the context of this study a sample is DNA from an individual coffee plant, and a marker state is binary data as presence or absence of a specific amplification product. This sample × marker matrix of character-states was then used to construct a sample × sample matrix of pair-wise genetic distances or similarities. In this study simple matching, Jaccard and Dice similarity coefficients were used. The resulting similarity matrices were analysed to produce dendrogram (tree diagram) linking together in clusters samples that are more genetically similar to each other than to samples in other clusters. Clusters are linked to each other at progressively lower levels of

similarity until all the samples being analysed are included in a single cluster.

Turning to the measurement of genetic diversity and genetic structure among and within populations, Percent polymorphic bands, Nei's (1973) genetic diversity statistics and genetic differentiation based on Shannon's information index were employed. The Nei's (1973) estimates of gene diversity and Shannon's information index were computed assuming Hardy Weinberg Disequilibrium with fixation indices of 0.95. The total genetic diversity (H_T), the within population genetic diversity (H_S), among populations genetic diversity (D_{ST}), and the coefficient of genetic differentiation (G_{ST}) are related by the expression $H_T = H_S + D_{ST}$, and $G_{ST} = D_{ST}/H_T$. An estimate of overall population differentiation is obtained from the G_{ST} value. The values of G_{ST} vary from 0 to 1, with $G_{ST} = 1$ indicating that the populations are fixed for different alleles. G_{ST} is equivalent to Wrights between populations differentiation coefficient (F_{ST}) (Nei, 1973).

Summary of the results and discussion

The pair wise genetic similarity between all possible pairs of samples ranged from 0.12–0.88, 0.27–0.97, 0.41–1.00 and 0.41–0.98 for RAPD, ISTR, ISSR and SSR markers, respectively (Table 1). Percent polymorphic bands varied from 25% for ISTR and ISSR markers to 100% for RAPD markers. The average number of bands detected per primer for each marker type was 6, 12, 11 and 12 for RAPD, ISTR, ISSR and SSR, respectively. The low number of bands/primer detected by RAPD markers may partially attributed to the electrophoresis gels and band visualization techniques used. Based on the coefficient of similarity range and percent polymorphic bands alone, RAPD marker revealed relatively more variability among the forest coffee samples considered in this study. The correlation between the different marker types varied from 0.10 to 0.59 with statistically significant correlation observed only between ISTR and RAPD, and ISTR and ISSR markers with equal magnitude (0.59). SSR markers are least correlated to the other markers, which may be attributed to differences in mechanism of variation involved. SSR variation is mainly caused by slippage of DNA polymerase during replication. Nevertheless, all the four PCR-based DNA markers used in this study were proved to be useful for the characterization of *C. arabica* germplasm.

To study the distribution of genetic variation in the populations, Nei's (1973) gene diversity statistics was used. The results indicate that most of the variation is found among populations. This observation corresponds well to the genetic structure of

predominantly self-pollinating populations of a species, which are characterized by a relatively high value of total gene diversity (H_T), a low value of gene diversity within populations (H_S), high value of gene diversity among populations (D_{ST}) and a high value of the coefficient of gene differentiation (G_{ST}). Self-pollinating species maintain high genetic diversity at their polymorphic loci, and most of this variation is found among populations. The results of this study fit well with the assumption that *C. arabica* is predominantly a self-pollinating species with most of its variation residing between populations. Comparable pattern of genetic differentiation were reported for other self-pollinating species (*Elymus fibrosis*) (Díaz *et al*., 2000).

Table 1. Comparison of the different molecular markers used in the study.

	N^{0} of primers used	Mean N^{0} of bands/primer detected	Total N^{0} of bands detected	Percent polymorphic bands	Range of Jaccard similarity coefficient
RAPD	12	06	075	100%	0.12–0.88
ISTR	12	12	144	25%	0.27–0.97
ISSR	11	11	123	25%	0.41–1.00
SSR	08	12	097	71%	0.41–0.98

Table 2. Correlations among matrices of the different marker systems used.

Markers	RAPD	ISTR	ISSR	SSR
RAPD	-	**0.59**	0.47	0.16
ISTR		-	**0.59**	0.10
ISSR			-	0.11
SSR				-

Table 5. Results of the Nei's (1973) genetic diversity statistics of forest coffee samples analysed with the four different marker types.

Marker type	H_T	H_S	D_{ST}	G_{ST}
RAPD	0.28	0.12	0.16	0.57
ISTR	0.31	0.14	0.17	0.55
ISSR	0.29	0.09	0.19	0.68
SSR	0.25	0.08	0.17	0.68

H_T = total gene diversity; H_S = gene diversity within populations; D_{ST} = gene diversity among populations; and G_{ST} = coefficient of gene differentiation.

In the cluster analysis performed most the samples were clustered on the basis of their geographical origin, which could be attributed to a few region specific banding patterns detected. However, within each region most of the samples failed to cluster on the basis of their respective populations, which may be due to the closeness of the populations in each region and the presence of substantial gene flow between adjacent populations mainly in the form of young plants by

man. The transportation of young coffee plants on horseback by local farmers for transplantation in coffee tree deprived forests is a common phenomenon in coffee growing regions of the country. In addition wild animals such as monkeys also play a significant role in coffee trees gene flow between adjacent populations by eating the berries at one place and defecating the seeds in an other place, which can germinate and give rise to a new coffee plant.

The cluster analysis also revealed a slight differentiation of samples into east and west of the Great Rift Valley. Comparable differentiation of *C. arabica* germplasm into east and west of Great Rift Valley was previously reported with RAPD marker analysis (Lashermes *et al*., 1996a). However, the presence of tectonic fault "Great Rift Valley" does not account for the genetic differentiation observed because phylogenetic studies based on cpDNA variation suggested a recent origin of the genus *Coffea* (Lashermes *et al*., 1996b; Cros *et al*., 1998). This may indicate that the coffee colonization of Ethiopia probably occurred after the formation of the Rift valley (Anthony *et al*., 2001), and the southeastern coffee trees could have been introduced from southwest by man or collected from southeastern forests before their destruction (Montagnon & Bouharmont, 1996). Some remaining pocket primary forest ecosystem, for instance Harana forest, may still contain the true wild forms of *C. arabica* from which local selection might have occurred.

Conclusion

Extent of distribution, areas sampled and plant characteristics such as mode of reproduction, breeding behaviour and generation time are some of the important parameters that determine the level of genetic variability revealed in a species (Bhat *et al*., 1999). Because *C. arabica* is an allotetraploid and a predominantly self-pollinated species a high degree of genetic uniformity is expected (Lashermes *et al*., 1996a). However, the present study indicate the presence of moderate genetic diversity within or between some forest coffee populations, which need due attention from conservation and breeding perspectives. The variations mainly reside in few populations from different regions. Therefore, considering few populations from each region for either *in situ* or *ex situ* conservations may preserve most of the variation of the species.

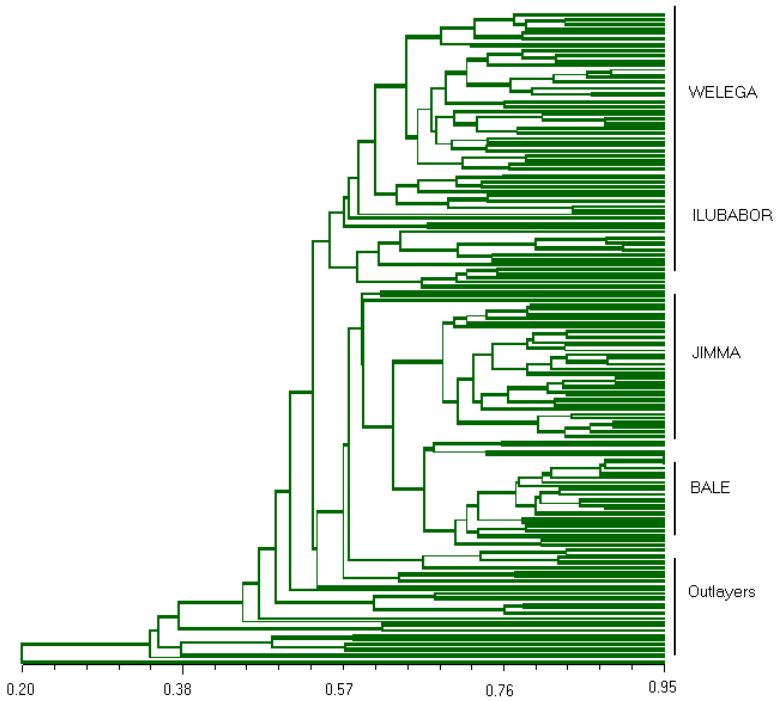

Figure 8. Dendrogram depicting relationships among 144 coffee samples based on simple matching similarity coefficients of RAPD data.

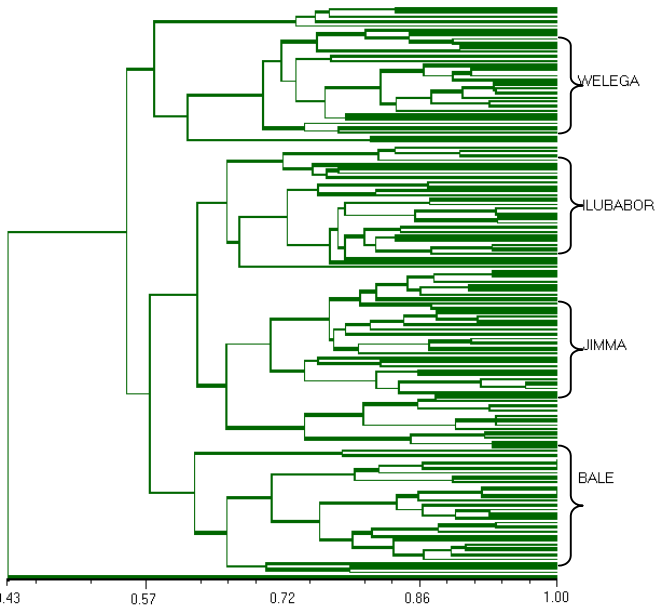

Figure 9. Dendrogram depicting relationships among 128 coffee samples based on Jaccard's similarity coefficients of RAPD data.

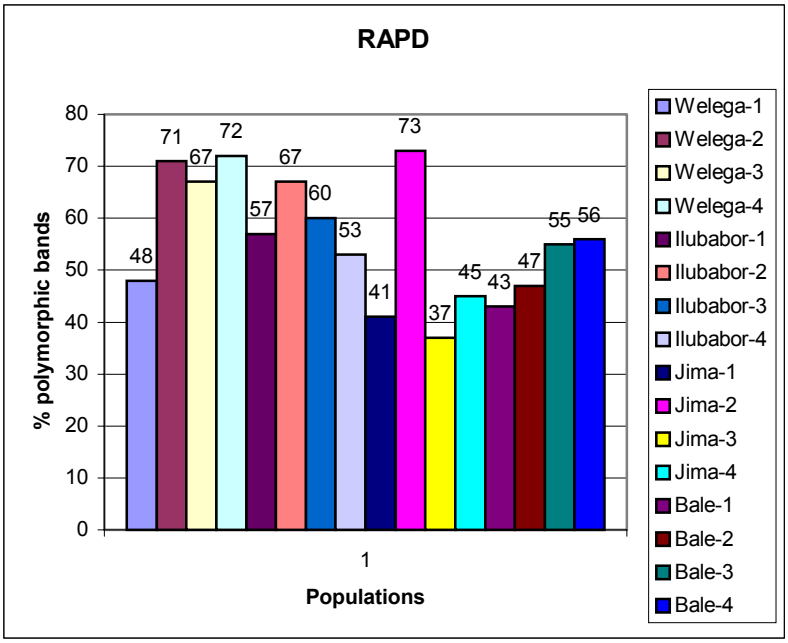

Figure 10. Histogram depicting population diversity based on percent polymorphic bands of the RAPD data.

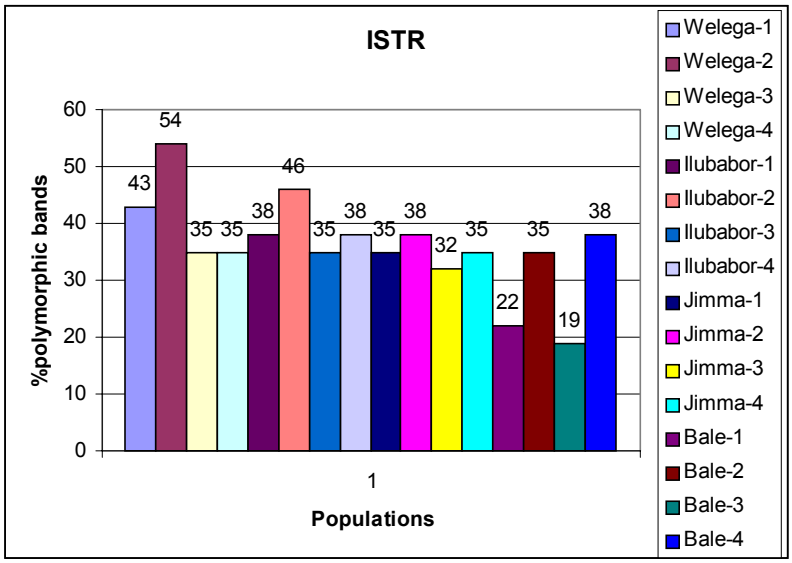

Figure 11. Histogram depicting population diversity based on percent polymorphic bands of the ISTR data.

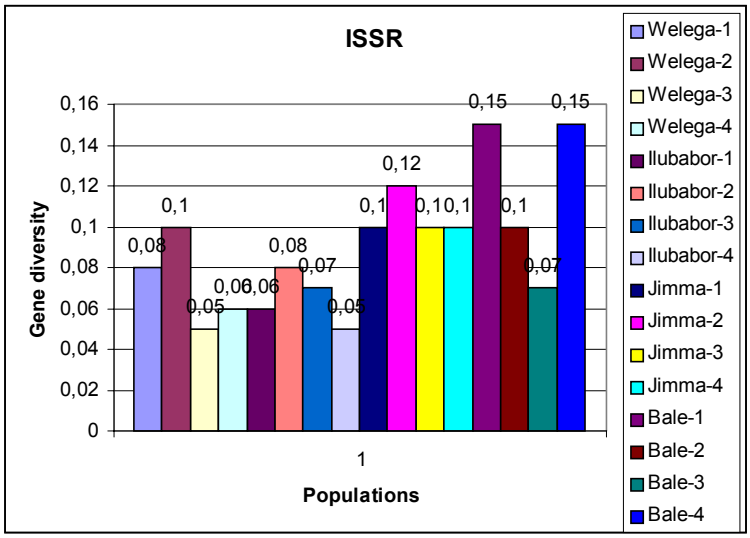

Figure 12. Histogram depicting population diversity based on gene diversity values of the ISSR data.

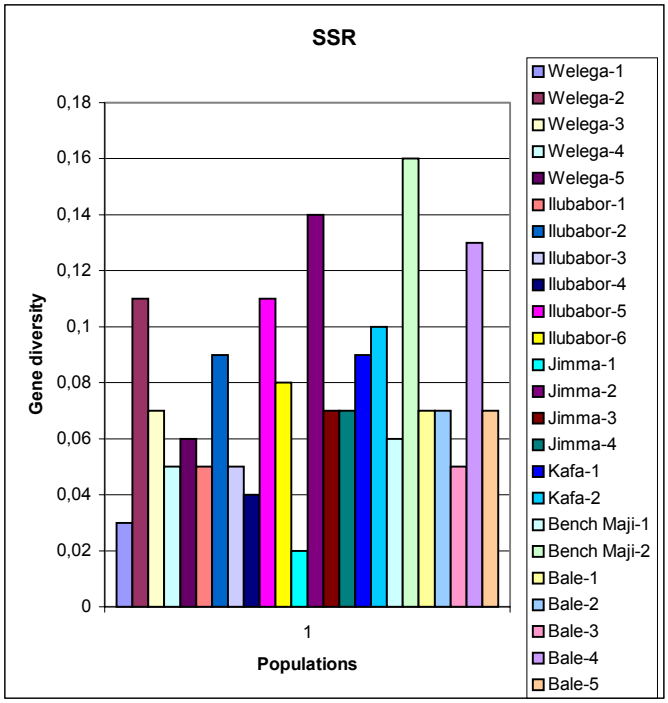

Figure 13. Histogram depicting population diversity based on gene diversity values of the SSR data.

Implication for conservation

Coffee seeds in general cannot be stored for long-term conservation in seed gene banks. Though coffee seeds can withstand desiccation down to 6–8% water content (fresh weight basis), they cannot be considered orthodox because they remain cold sensitive and desiccation does not improve their longevity (Ellis *et al*., 1990). When storing fully hydrated seeds at 19° C and under 100% relative humidity the maximal storage duration obtained was 36 months for *C. arabica* and 15 months for *C. canephora* and *C. stenophylla*.

The collections of coffee genetic resources are conventionally maintained as living trees or shrubs in field gene banks (Berthaud & Charrier, 1988). There is a considerable amount of of *C. arabica* germplasm collected and maintained in *ex situ* field gene banks in Ethiopia since 1966 (Belachew, 1997). *Ex situ* field gene bank is a commonly used alternative for conserving germplasm, which is difficult to conserve as seed. It offers easy access to the plant material for characterization, evaluation and subsequent utilization. However, the number of genotypes, which a field gene bank can hold is restricted by financial and land resources available, and this in turn limits the genetic diversity, which it can conserve. In addition, *ex situ* field gene bank conservation is characterized by risks of losing valuable germplasm due to diseases and pests as well as to poor adaptation of certain genotypes of the species to the local environment.

In contrast to the *ex situ* conservation, *in situ* conservation still remains an important component of the overall strategy for the long-term conservation of the coffee gene pool. The *in situ* method allows the dynamic evolutionary processes of the species in its natural habitat, especially in producing new resistance to pests and diseases. On the other hand, the *ex situ* approach safeguards the species genetic diversity in case of possible habitat destruction and represent a readily available source of germplasm for research and breeding. *In situ* conservation is the most widely used strategy for conserving wild species, recalcitrant seed species and those species, which depend on other organisms within the ecosystem (Maxted *et al*., 1997). One of the objectives of the present study was to obtain genetic diversity information to develop appropriate strategies for the conservation of forest coffee in Ethiopia. This study helps to identify candidate populations for *in situ* conservation. Selection of few populations from each region for *in situ* conservation or collection of samples for *ex situ* field gene bank may preserve most of the genetic diversity of this species. For instance, Welega-2, Ilubabor-2, Jima-2 and Bench Maji-2 populations should be given high priority. The comparatively distinct populations of the Bale (Harana forest) deserve particular attention due to the fact that 1) the

Harana forest is a part of the primary forest ecosystem, and could harbor the true natural forms of forest coffee. 2) Forest coffee samples from this region displayed some unique amplification profiles. 3) Genetic material from this region is not currently represented in either *ex situ* field gene banks or in active breeding programs in Ethiopia.

Implication for breeding

Besides providing information to design conservation strategies, the genetic diversity information obtained in this study could also be used in breeding programs in order to cross genetically diverse parents and maximize the level of variation present in segregating populations.

Some coffee breeding centers now emphasize hybrid varieties as the best strategy for further and more rapid increase of plant productivity. In *C. arabica*, 30 to 60% heterosis in yield over the better parent has been observed in Ethiopia (Ameha, 1990). Coffee hybrids were also found to have greater yield stability over location and time (fewer genotype × environment interaction effects). Combining parents selected from genetically diverse subpopulations increase chances of substantial hybrid vigour. Hybrid vigour for yield noticed in crosses between parents of different origin appears to be the result of accumulation of complementary polygenes dispersed over subpopulations (Van der Vossen, 2001). Molecular markers play an important role in identifying genetically diverse populations or genotypes. In addition, the development of molecular markers could serve in marker-assisted selection, provided that the markers are linked to any of the important agronomic traits. For instance, RAPD markers have been found to be closely linked to some genes conditioning CBD (coffee berry disease) resistance (Agwanda *et al*., 1997; Cristancho, 1999). Development of molecular markers for resistance genes to CBD and other major coffee diseases would enable breeders to ensure effective host resistance by gene pyramiding. In this study some populations or genotypes have displayed unique amplification profiles particularly for RAPD and ISTR markers. Whether these unique bands are linked to any of the important agronomic traits and serve in marker assisted selection in future coffee breeding requires further investigations.

References

Agwanda, C., Lashermes, P., Trouslot, P., Combes, P. & Charrier, A. 1997. Identification of RAPD markers for resistance to Coffee Berry Disease, *Colletotrichum kahawae*, in arabica coffee. *Euphytica* 97, 241–248.

Ameha, M. 1990. Heterosis and arabica coffee breeding in Ethiopia. *Plant Breeding Abstr*. 60, 594–598.
Ameha, M. 1991. Significance of Ethiopian coffee genetic resources to coffee Improvement. In: J.M.M. Engels, J.G. Hawkes & M. Worede (eds) Plant genetic resources of Ethiopia, Cambridge University Press, pp. 354–359.
Anthony, F., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J. & Charrier, A. 2001. Genetic diversity of wild coffee trees (*Coffea arabica* L.) using molecular markers. *Euphytica* 118, 53–65.
Avise, J.C. 1994. Molecular markers, Natural History and Evolution. Kluwer Academic Publishers, Boston/Dordrecht/London.
Bhat, K.V., Babrekar, P.P & Lakhanpaul, S. 1999. Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* 110, 21–23.
Bellachew, B. 1997. Arabica coffee breeding in Ethiopia: a review. In: Proceedings of the 17th ASIC Colloquium (Nairobi), pp.406–416. ASIC Paris France.
Berthaud J & Charrier A (1988). Genetic resources of *Coffea*. In: R.J. Clarke & R. Macrae (eds) Coffee Vol. 4 Agronomy, Elsevier Applied Science, London & New York, pp. 1–42.
Bridson, D.M. & Vercourt, B. 1988. Rubiaceae (Part 2). In: Polhill R.M. (ed) Flora of Tropical East Africa, Balkema, Rotterdom/Brookfield, pp. 703–723.
Carvalho, A., Ferwerda, F.P., Frahm-Leliveld, J.A., Medina, D.M., Mendes, A.J.T. & Monaco, L.C. 1969. Coffee (*Coffea arabica* L. & *C. canephora* Pierre ex Froehner). In: F.P. Ferwerda & F. Wit (eds.) Outlines of perennial crop breeding in the tropics, Veenman & Zonen, Wageningen, pp. 189–241.
Chang, R.Y., O'Donoughue, L.S. & Bureau, T.E. 2001. Inter-MITE polymorphism (IMP): a high throughput transposon based genome mapping and fingerprinting approach. *Theor. Appl. Genet.* 102, 773–781.
Coulibaly, I., Noirot, M., Lorieux, M. & Charrier, A. 2002. Introgression of self-compatibility from *Coffea heterocalyx* to cultivated species *Coffea canephora*. *Theor. Appl. Genet.* 105, 994–999.
Cramer, P.J.S. 1957. Review of literature of coffee research in Indonesia. SIC Editorial International American Institute of Agricultural Sciences, Turrialba, Costa Rica.
Cristancho, M. 1999. Genetic diversity of the cenicafe germplasm collection. Third International Seminar on Biotechnology in the Coffee Agroindustry, Londrina, Brazil.
Cros, J., Combes, M.C., Trouslot, P., Anthony, F., Hamon, S., Charrier, A. & Lashermes, P. 1998. Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. *Mol. Phylogenet. Evol.* 9, 109–117.
Díaz, O., Sun, G.L., Salomón, B. & von Bothmer, R. 2000. Levels and distribution of allozyme and RAPD variation in populations of *Elymus fibrosus* (Poaceae). *Genetic Resour. Crop Evol.* 47, 11–24.
Dublin, P., Enjalric, F., Lardet, F., Carron, M.P., Trolinder, N. & Pannetier, C. 1991. Estate crops. 1. *Coffea* species. In: P.C. Debergh & R.H. Zimmerman (eds) Micropropagation, Kluwer Academic Publ, Dordrecht, the Netherlands, pp. 337–361.
Edwards, D., Coghill, J., Batley, J., Holdsworth, M. & Edwards, K.J. 2002. Amplification and detection of transposon insertion flanking

sequences using fluorescent MuAFLP. *Biotechniques* 32, 1090–1097.
Ellis, R.H., Hong, T.S. & Roberts, E.H. 1990. An intermediate category of seed storage behaviour? I. Coffee. *Journal of Experimental Botany* 41, 1167-1174.
Ellsworth, D.L., Rittenhouse, K.D. & Honeycutt, R.L. 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* 14, 214–217.
Flavell, A.J., Knox, M.R., Pearce, S.R. & Ellis, T.H.N. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant. J.* 16, 643–650.
Gupta, M., Chyi, Y.S., Romero-Sevenson, J. & Owen, J.L. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.* 89, 998–1006.
Hallauer, A.R., Russel, W.A. & Lampkey, K.R. 1988. Corn breeding. In: G.F. Sprague & J.W. Dudley (eds) Corn and corn improvement, ASA, CSSA, and SSSA, Madison, Wisconsin, USA.
Haarer, A.E. 1956. Modern coffee production. Leonard Hill, London.
Hillis, D.M.& Moritiz, C. 1990. An overview of applications of molecular systematics, Sinauer Associates, Sunderland, MA, USA, pp. 502-515.
Hunter, R.L. & Markert, C.L. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125, 1294–1295.
Kalendar, R., Grob, T., Regina, M., Suoniemi, A. & Schulman A. 1999. IRAP & REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.* 98, 704–711.
Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G. & 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.
Lashermes, P., Comes, M.C., Cros, J., Trouslot, P., Anthony, F., & Charrier, A. 1995. Origin and genetic diversity of *Coffea arabica* L. based on DNA molecular markers. In: Proceedings of the 16th ASIC Colloquium (Koyoto), pp. 528-535. ASIC, Paris, France.
Lashermes, P., Trouslot, P., Anthony, F., Comes, M.C. & Charrier, A. 1996a. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* 87, 59–64.
Lashermes, P., Cros, J., Combes, M.C., Trouslot, P., Anthony, F., Hamon, S. & Charrier, A. 1996b. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theor. Appl. Genet.* 93, 626–632.
Lashermes, P., Comes, M.C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F. & Charrier, A. 1999. Molecular characterization and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet.* 261, 259–266.
Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M.C., Dussert, S., Graziosi, G., Trouslot, P.& Anthony, F. 2000. Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *Theor. Appl. Genet.* 100, 139–146.
Levinson, G. & Gutman, G.A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol*. 4, 203–221.

Litt, M. & Lutty, 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.
Lynch, M. & Milligan, B.G. 1994. Analysis of population structure with RAPD markers. *Mol. Ecol.* 3, 91–99.
Maxted, N., Ford-Lloyd, B.V. & Hawkes, J.C. 1997. Plant Genetic Conservation. The *in situ* Approach. Chapman and Hall, London.
Meyer, G.F. 1965. Notes on wild *Coffea arabica* from Southwestern Ethiopia, with some historical considerations. *Econ. Bot.* 19, 136–151.
Meyer, A.J.T., Fernie, L.M., Narasimhaswami, R.L., Monaco, L.C. & Greathead, D.J. 1968. FAO Coffee Mission to Ethiopia 1964–65. FAO, Rome.
Moisyadi, S., Neupane, K.R. & Stiles, J.I. 1999. Cloning and characterization of zanthosine-N^7-methyltransferase pathway. In: Proceedings of the 18th ASIC Colloquium (Helsinki), pp.327–331. ASIC Paris France.
Montagnon, C. & Bouharmont, P. 1996. Multivariate analysis of phenotypic diversity of *Coffea arabica*. *Genetic Resour. Crop Evol.* 43, 221–227.
Morgante, M. & Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3, 175–182.
Muralidharan, K. & Wakeland, E.K. 1993. Concentration of primer and template qualitatively affects products in random amplified polymorphic DNA PCR. *Biotechniques* 14, 362–364.
Narasimhaswamy, R.L. 1962. Some thoughts on the origin of *Coffea arabica* L. Coffee 4(12), 1-5. Turrialba, Costa Rica.
Nei, M., 1973. Analysis of genetic diversity in subdivided Populations (population structure/ genetic variability/ hetrozygosity/ gene differentiation). *Proc. Nat.l Acad. Sci. USA* 70, 3321–3323.
Orozaco-Castillo, C., Chalmers, K.J., Waugh, R. & Powell, W. 1994. Detection of genetic diversity and selective gene introgration in coffee using RAPD markers. *Theor. Appl. Genet.* 87, 934–940.
Peakall, R., Gilmore, S., Keys, W., Morgante, M. & Rafalski, A. 1998. Cross-species amplification of soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera – implication for the transferability of SSRs in plants. *Mol. Biol. Evol.* 15, 1275–1287.
Penner, G.A., Bush, A., Wise, R., Kim, W., Domier, L., Kasha, K., Laroche, A., Scoles, G., Molnar, S.J. & Fedak, G. 1993. Reproducibility of Random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods Appl*. 2, 341–345.
Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2, 225–238.
Raina, S.N., Mukai, Y & Yamamato, M. 1998. In situ hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae). *Theor. Appl. Genet.* 97, 1204–1209.
Rani, V., Singh, K.P., Shiran, B., Nandy, S., Goel, S., Devarumath, R.M., Sreenath, H.L. & Raina, S.N. 2000. Evidence for new nuclear and mitochondrial genome organizations among high-frequency somatic embryogenesis-derived plants of allotetraploid *Coffea arabica* L. (Rubiaceae). *Plant Cell Reports* 19, 1013–1020.
Rohde, W., Kullaya, A., Rodriguez, J. & Ritter, E. 1995. Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer

sequences separating a subset of copia-like *EcoR*I repetitive elements. *Journal of Genetics & Breeding* 49, 179–186.
Rohde, W. 1996. Inverse sequence-tagged repeat (ISTR) analysis, a novel and universal PCR-based technique for genome analysis in the plant and animal kingdom. *Journal of Genetics & Breeding* 50, 249–261.
Sanchez, de la Hoz, M.P., Dávila, J.A., Loarce, Y. & Ferrer, E. 1996. Simple sequence repeat primers used in polymerase chain reaction amplifications to study genetic diversity in barley. *Genome* 39, 112–117.
Soltis, P.S., Soltis, D.E & Smiley, C.J. 1992. An rbcL sequence from a Miocene Taxodium (bald cypress). *Proc. Natl. Acad. Sci. USA* 89, 449–451.
Strenge, H. 1956. Wild Coffee in Kaffa Province of Ethiopia. *Tropical Agriculture Trindad* 33, 297–301.
Sylvain, P.C. 1955. Some observation on *Coffea arabica* L. in Ethiopia. *Turrialba* 5, 37–53.
Sylvain, P.G., 1958. Some observations on *Coffea arabica* L. in Ethiopia. *Turrialba* 5, 37–53.
Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17, 6463–6471.
Thomas A.S. 1942. The wild arabica coffee on the Boma plateau, Anglo-Egyptian Sudan. *The Empire J. Exp. Agri*. 10, 207–212.
Urbaneja, G., Ferrer, J., Paez, G., Arenas, L. & Colina, G. 1996. Acid hydrolysis and carbohydrates characterization of coffee pulp. *Renewable Energy* 9, 1041–1044.
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.
Van der Vossen, H.A.H. 1979. Methods of preserving the viability of coffee in storage. *Seed Science & Technology* 7, 65–74.
Van der Vossen, H.A.H. 1985. Coffee selection and breeding. In M.N. Clifford & K.C. Willson (eds) Coffee: Botany, Biochemistry and production of beans and beverage, Croom Helm, London, New York & Sydney, pp.48–96.
Van der Vossen, H.A.H. 2001. Agronomy I: Coffee breeding Practices. In: R:J. Clake & O:G. Vitzthum (eds) Coffee Recent Development, Blackwell Science Ltd, London, pp.184–201.
Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Plot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414.
Wang, Z.Y. & Tanksley, S.D. 1989. Restriction length polymorphism in *Oryza sativa* L. *Genome* 32, 1113–1118.
Weber, J.L.& 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.
Weising, K., Nybom, H., Wolff, K. & Kahl G. 2005. DNA Fingerprinting in Plants Principles, Methods, and Applications Second edition. Taylor & Francis Group, CRS Press, New York.
Welsh, J. & 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. & Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531–6535.
Wu, K., Jones, R., Dannaeberger, L. & Scolnik, P.A. 1994. Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Res.* 22, 3257–3258.
Zietkiewicz, E., Rafalski, A. & Labuda, D 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20, 176–183.

Acknowledgements

First and foremost I praise and glorify the only Almighty God (Jesus) for his mercy, the patience and endurance he bestowed on me during the long study period I was away from my family.

Next, I would like to express my sincere gratitude to my main supervisor, Professor Tomas Bryngelsson for accepting me to work in his laboratory, the excellent guidance, critical reading and valuable suggestions on the manuscripts and the thesis.

My warmest thanks is due to my Ethiopian co-supervisor Professor Endashaw Bekele for his role in establishing collaborative study programme with Swedish Institution and choosing me to be engaged in this study programme.

My thanks also go to my Swedish co-supervisor Dr. Björn Salomon for his valuable comments on the manuscripts.

I would like to express my heartfelt thanks to my wife Zenebech Wakweya for her patience, love and determination to take care of our children alone. I appreciate and thank my children, especially Bayeti, Ayantu, Bashatu and Amarti for their patience and good conduct during the long study period I was away from them.

My sincere thanks also go to the Office of Sayo District Schools Administration (Waajjira Manneetii Barnootaa Aanaa Sayyoo) for assigning my wife in nearby elementary school by taking into consideration her family burden, which in fact eased my worries to say the least.

I thank the Administration of the Department of Biology, Addis Ababa University, particularly Dr. Kifle Dagne, Mrs. Mulu-Alem, and Banchu for their kind assistance, and Mr. Getachew Taye for driving me by car to the different regions of Ethiopia for sample collection.

My special thanks also go to the Institute of Biodiversity Conservation of Ethiopia for allowing me to take coffee seed samples abroad for the purpose of this study.

I thank all the staff of the Department of Crop Science at Alnarp, Sweden for their contribution to the success of this work in one ways or another.

My special thanks go to Mrs. Britt Greén for her assistance in the laboratory work.

Mrs. Helén Lindgren for her help in ordering chemicals instantly when Britt was not around.

Mrs. Monica Lotifinia for her kind assistance in administrative matter.

Mr. Goran Olsson and Mr. Lars Hagtorn for their help with the computer problems.

Ms. Phuong Nguyen for her help with taking several pictures of my coffee samples.

Dr. Erland Liljeroth and Ms. Therése Bengtsson for their assistance in scanning gel photos.

Dr. Agnese Kolodinska for showing me how to use POPGENE software and answering me several questions regarding the formatting of this thesis.

I thank the Ethiopian colleagues, particularly Dr. Amsalu Ayana for showing me how to use NTSYS Pc software. Faris Hailu and Yohannes Petros are always remembered for their genuine friendship and feel of sympathy in time of my difficulties.

Finally, I would like to acknowledge the financial support from SIDA/SAREC, Sweden for BIO-EARN (East African Regional Programme and Research Network for Biotechnology, Bio-safety and Biotechnology Policy Development) through the administration of Stockholm Environment Institute. My particular thanks go to Ms. Benita Forsman and Dr. Ivar Virgin for their excellent administration of the BIO-EARN programme. My sincere thanks also go to the BIO-EARN regional and national co-ordinators, and the entire BIO-EARN programme family. Thank you all! God bless you!