Ecology and Development Series No. 44, 2006

Kassahun Tesfaye Geletu

Genetic diversity of wild *Coffea arabica* populations in Ethiopia as a contribution to conservation and use planning

Cuvillier Verlag Göttingen

*Dedicated to the memory of my late high school friend**, Genen Gebre-Wold**.*

ABSTRACT

Compiling information on the extent and distribution of the existing genetic diversity of wild coffee (*Coffea arabica)* populations of the Afromontane rainforest of Ethiopia is crucial not only for planning *in situ* conservation measures but also for furthering their utilization in breeding programs. Extensive fieldwork led to a representative sampling of Ethiopian *C. arabica* populations that also allows placing the main study sites of CoCE project with respect to the overall *C. arabica* genetic diversity.

Comparative sequencing was done for highly variable spacers (*atpB-rbcL*, *rps19-rpl2*, *rpl22-rps19, trnS-G*), introns (*atpF*, *rpl16, trnG*, *trnK*) and *matK* gene. Totally ca.7.2 kb was sequenced for 5 individuals of *C. arabica* representing geographically distant wild populations, two commercial cultivars, eight diploid species and *Psilanthus leroyi and Ixora coccinea* as outgroup. Phylogeny reconstruction using Maximum Parsimony and Bayesian Inference resulted in congruent topologies with high support of *C. arabica* and *C. eugenioides* being sisters. Seven microsatellites were characterized, which normally are variable within other species of flowering plants, but surprisingly do not show any variability at all in *Coffea arabica*. All individuals of *C. arabica* have a single chloroplast haplotype, which is inline with assuming a unique allopolyploidization event for the origin of *C. arabica*. The low level of variation observed in the genome of *C. arabica* is a strong evidence for the recent origin of *C. arabica*.

The comparison of wild *Coffea arabica* populations based on ISSR fingerprints yielded a complex geographical distribution pattern of genotypes with most regions possessing their own genotypes and providing evidence for the need of a multi-site *in situ* conservation approach. The observed hierarchical-geographic patterning is obscured by gene flow and resulting in rather high levels of genetic diversity within regions. This could be attributed to high levels of gene flow that are connected to substantial outcrossing rates (insects as pollinators) and seed dispersal (animals such as monkeys and birds and humans) among wild populations of arabica coffee. The addition of landraces clearly showed that wild populations of *C. arabica* are genetically different and can be distinguished from semi-domesticated plants or farmers' variety (landrace). In previous studies, all of these plants were subsumed under "subspontaneous material".

The detailed intraregional analysis of wild coffee in Berhane Kontir and Yayu (Geba Dogi) with denser samples showed predominant fine-scale spatial genetic structuring of wild coffee populations. This confirmed that genetic similarity is higher among neighboring individuals than among more distant individuals. A striking diversity contrast among di- and tri- as compared to tetra-nucleotide repeat primers is observed. This indicates that frequency and localization of tetranucleotide repeat microsatellites is extremely variable in the *Coffea* genome, and is almost unique in every individual plant.

Overall, the results offer not only the extent and patterns of genetic diversity of wild *C. arabica* but also provide information on evolution and origin of the species. The study indicates the importance of establishing microsatellite (co-dominant) marker system that allows assessing heterozygosity and gene flow in wild *Coffee*. Furthermore, detailed discussion of the results with relevant recommendations and further research directions were forwarded.

Genetische Diversität der Wild-Populationen von *Coffea arabica* in Äthiopien als Grundlage für die Planung von Schutzmaßnahmen und nachhaltiger Nutzung

KURZFASSUNG

Kenntnisse über das Ausmaß und die Verteilung genetischer Diversität der Wildpopulationen von Arabica-Kaffee (*Coffea arabica*) in den afromontanen Regenwäldern Äthiopiens sind Grundlage nicht nur für die Planung von *in situ* Schutzmaßnahmen sondern auch für die Nutzung der genetischen Resourcen in Züchtungsprogrammen. Die vorliegende Arbeit wurde im Rahmen des CoCE-Projektes angefertigt und basiert auf einem repräsentativen Sampling von Populationen des tetraploiden *C. arabica* in Äthiopien.

Mit dem Ziel infraspezifisch variable Chloroplasten-Marker zu finden und die Herkunft des mütterlichen Eltern-Genoms zu klären wurden schnell-evolvierende Spacer (*atpB-rbcL*, *rps19-rpl2*, *rpl22-rps19, trnS-G*) und Introns (*atpF*, *rpl16, trnG*, *trnK*) sowie das *matK* Gen vergleichend sequenziert (zusammen > 7 kb). Von *C. arabica* wurden fünf Individuen aus geographisch unterschiedlichen Wild-Populationen, zwei Kultivare, acht diploide Arten von *Coffea* sowie *Psilanthus* und *Ixora* als Außengruppen einbezogen. Phylogenie-Rekonstruktion (Parsimonie, Bayesianische Inferenz) ergab kongruente Topologien mit hoher Stützung von *C. arabica* und *C. eugenioides* als Schwestergruppe. Beide Arten zeigen jeweils mehrere autapomorphe Mutationen. Sieben Mikrosatelliten wurden charakterisiert, die bei anderen Blütenpflanzen innerartlich variabel sind, nicht aber bei *C. arabica*. Basierend auf den vorliegenden Sequenzdaten haben alle Individuen von *C. arabica* einen einzigen Chloroplasten-Haplotyp. Dies spricht für ein einziges Allopolyploidisierungs-Ereignis sowie für ein junges Alter von *C. arabica* bzw. für Flaschenhals-Effekte in der Evolution der Art.

ISSRs (Inter Simple Sequence Repeats) wurden für das Fingerprinting einer großen Zahl von Individuen (insgesamt 314) von *C. arabica* eingesetzt. In einer Interregionalen Analyse wurden Individuen aus Wäldern in den CoCE-Gebieten (= Regionen) mit Individuen aus Wäldern in den übrigen Landesteilen sowie mit Landrassen verglichen. Es zeigte sich ein komplexes Muster der Verbreitung von Genotypen. Für *in situ*-Conservation wird daher ein Multi-Site Ansatz benötigt. Die genetische Diversität innerhalb von Regionen ist hoch, wobei aber alle Regionen ihre eigenen Genotypen besitzen. Ein hierarchisch-räumliches Diversitätsmuster wird offenbar durch starken Gen-Fluß verwischt. Es wird daher angenommen, dass im Gegensatz zu bisherigen Daten aus Kaffee-Plantagen, Outcrossing in den Wildpopulationen von C. arabica eine große Rolle spielt. Dies kann durch Insekten (Bestäuber) und Affen sowie Vögel (Vektoren für Samen) geschehen. Der weitergehende Vergleich mit Landrassen aus verschiedenen Regionen Äthiopiens zeigt, das wilde Populationen von *C. arabica* genetisch von sogenannten Farmer's Varieties verschieden sind. Dies ist eine erhebliche Neuerung gegenüber früheren Studien, bei denen das Material nicht exakt dokumentiert und unter „subspontan" subsummiert wurde.

Für die detaillierte Intraregionale Analyse wurden jeweils 25 Individuen zufällig aus 50x50 m Wald-Plots in Berhane Kontir und Yayu (Geba Dogi) analysiert. Dichteres Sampling ergab innerhalb der Regionen eine räumlich differenzierte Verbreitung von Genotypen, die auf nahe Verwandtschaft von Individuen innerhalb der Plots schließen lässt.

Im Vergleich zwischen Di- und Tri- mit Tetranucleotid-Primern zeigt sich ein extremer Diversitäts-Kontrast. Offenbar sind Lokalisierung und Häufigkeit von Tetranukleotid-Repat Mikrosatelliten im Kerngenom von *C. arabica* extrem variabel von Individuum zu Individuum. Mit Tetranukleotid-ISSRs lassen sich daher einzelne Individuen von *C. arabica* erkennen.

Die Ergebnisse bieten einen ausführlichen Überblick über den Grad die Verteilung der genetischen Diversität der Wildpopulationen von *C. arabica*, sowie neue Erkenntnisse über die Evolution der Art. Die bisher vorliegenden Daten weisen aber auch darauf hin, dass weitere Forschungen mit co-dominanten Kern-Mikrosatelliten nötig sind, um das Ausmaß der Heterozygosität der Individuen von *C. arabica* zu verstehen und Genfluß in natürlichen Populationen abzuschätzen.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

(in alphabetic order)

AMOVA	=	Analysis of Molecular Variance
a.s.l.	=	Above sea level
BI	=	Bayesian Inference
bp	=	base pairs
FCE	=	Forest Coffee Ecosystem
CI	=	Consistancy Index
cp	=	Chloroplast
FISH	=	Fluorescent *in situ* Hybridization
JK	=	Jackknife Support
GISH	=	Genomic *in situ* Hybridization
ISSR	=	Inter Simple Sequence Repeat
MAS	=	Marker assisted selection
MP	=	Maximum Parsimony
Mt	=	Mitochondrial
Nt	=	Nucleotide
PCO	=	Principal Coordinate Analysis
PP	=	Bayesian Posterior Probability
RAPD	=	Random Amplification of Polymorphic DNA
RFLP	=	Restriction Fragment Length Polymorphism
RC	=	Rescaled Consistency Index
RI	=	Retention Index
SD	=	Standard Deviation
SE	=	South East
SSR	=	Simple Sequence Repeat
SW	=	South West

1 GENERAL INTRODUCTION

1.1 *Coffea arabica* as an economic plant in Ethiopia and beyond

The genus *Coffea* L. comprises approximately 100 species. However, only *C. arabica* L., *C. canephora* Pierre ex Froehner and *C. liberica* Bull ex Hiern are the economically important species of the genus. Eighty percent of the world coffee production comes from *C. arabica*, because of better cup quality and low bitterness and a good flavor. Nearly 20% of coffee production comes from *C. canephora. C. liberica* has minor importance and restricted to some localities (Purseglove 1968; Bridson and Verdcourt 1988; Puff and Chamchumroon 2003; Omolaja et al. 2006).

Coffee is one of the most important commercial commodity and foreign currency earnings for 80 developing countries (Cannell 1983; Ponte 2001). It is also considered as the most important tropical product that contributes almost half of total net exports of tropical products (Hallam 2003). Total worldwide exports (75 % of production) go beyond $9 billion, and the sector employs more than 25 million people globally on more than 5 million farms (Kaplinsky 2004). It is anticipated to be regularly consumed by more than 40 percent of the world's population and fills about 400 billion cups a year (Fitter and Kaplinsky 2001). The current statistics showed that coffee ranked only fifth among internationally traded commodities, after oil, aluminium, wheat and coal (Ponte 2001).

In Ethiopia, coffee plays a significant role in the regional and national economies, and also contributes to the country's foreign currency earning by more than 60% (Woldetsadik and Kebede 2000). Coffee also contributes from 4% to 5% to the national GDP (Gross Domestic Product) and generates 20% of government revenue (Asres 1996). Moreover, the processing and marketing of coffee creates employment opportunities for many people, thus making considerable contributions to the economy (Abebe 2005). National production levels are estimated to vary between 140,000-180,000 tons and an estimated 700,000 households nationally are involved in coffee production (Petty 2004).

Generally, the majority (95%) of coffee production in Ethiopia is produced by smallholder farms (Awoke 1997; Grundy 2005). In Ethiopia coffee is found at different levels of domestication and production systems. The intensity and level of management also varies accordingly. There are four major production systems of coffee in Ethiopia

(e.g., Dubale and Tektay 2000; Woldetsadik and Kebede 2000; Gole 2002; Senbeta 2006). These are forest, semiforest, garden and plantation production systems. The first two production systems are regard as a part of forest coffee ecosystem (FCE). In the forest coffee which is also referred as wild coffee, coffee regenerates in natural forests without human intervention as an understory plant. It grows in Afromontane rain forests of West, Southwest and Southeastern Ethiopia. This production system represents about 9% of the total land covered of coffee and also contributes about 5-6% of the national coffee production. The productivity of this production system is very low, and has been estimated to be 200-250 kg ha^{-1}.

The semiforest production system, which is also referred as semiwild coffee, evolved from forest coffee production system with intervention of humans. In this production system, the overstory forest trees are thinned and the ground vegetation also removed about two times a year. The natural forests are maneuvered to create microenvironments for recruitment and establishments of young coffee seedlings and also regeneration of coffee by removing the undergrowth. This production system occupies nearly 24% of the total land covered by coffee and contributes about 20% of the total coffee production in the country. The coffee yield per unit area of the semiforest production system is low ranging from 200-400 kg ha^{-1}. The forest coffee ecosystem (FCE) in total occupies nearly 33% of land given for coffee production and contributes 25% of the national coffee production.

The garden coffee production system is characterized by holding coffee at the farmer backyard and coffee farms with an area of less than 0.5 hectares. This is the main production system in southern and eastern part of the country. The enset-coffee homegardens agroforestry systems where coffee and enset are grown in association with other crops and trees are the main characteristic features of the home gardens in Southern Ethiopia (Figure 1.1d; Abebe 2005). In the eastern part of Ethiopia coffee is intercropped with the mild stimulant perennial crop “chat” (*Chata edulis*), sorghum, maize, beans and sweet potato. In most cases the farmers in both localities used to grow coffee landraces having its own characteristic features and Coffee Berry Disease (CBD) resistant cultivars released by Jimma Agricultural Research Center (Teketay and Tegineh 1991; Bellachew et al. 2000; Woldetsadik and Kebede 2000; Gole et al. 2001; Abebe 2005). The yield for traditional coffee farms is estimated to be 550 kg ha^{-1} with

moderate management of the field but could possibly be increased to 700 kg ha^{-1} with intensive management according to research recommendation (Woldetsadik and Kebede 2000).

The plantation production system of coffee in Ethiopia is mainly observed in the southwestern part of the country under heavy shade. This production system is largely based on the released CBD resistant selection and improved agronomic practices. However the yield obtained in this production system ranges from moderate to high yield (450-880 kg ha^{-1}). Higher average yield observed for the State Coffee farms since they run intensified management practices (Woldesadik and Kebede 2000).

1.2 Distribution and diversity of coffee forests in Ethiopia

1.2.1 Distribution of forest with wild *Coffea arabica*

Arabica coffee is an afromontane rainforest species and occurs naturally in the SW highlands and on the Bale Mountains in the SE highlands of Ethiopia (Figure 1.1; Gole 2003; Senbeta 2006). It is the only naturally occurring species of *Coffea* in Ethiopia and occurs in the undergrowth of the montane rainforest at altitudes between 1,400 and 1,900 m a.s.l. (Berthaud and Charrier 1988; Geber-Egziabher 1990; Gole et al. 2001; Senbeta 2006). Moreover, highest densities of coffee were recorded between 1300 and 1600 m a.s.l. suggesting the optimum altitude of wild coffee (Senbeta 2006). Friis (1979) reported the existence of wild coffee populations in the Boma plateau in SE Sudan and on Mount Marsabit in northern Kenya. A recent expedition into the Southern part of Ethiopia also showed the existence of additional wild coffee in Banja forest in the Dawro highland at the altitude of 1620 m a.s.l. (personal observation).

Generally, the occurrence and abundance of wild coffee populations differ among different regions of wild coffee. The environmental factors and level of interference by humans could be the main factors that affect the patterns of distribution with in the forest. Moreover, on flat to gentle slops highest abundance of wild coffee plants observed (Gole submitted; Senbeta 2006).

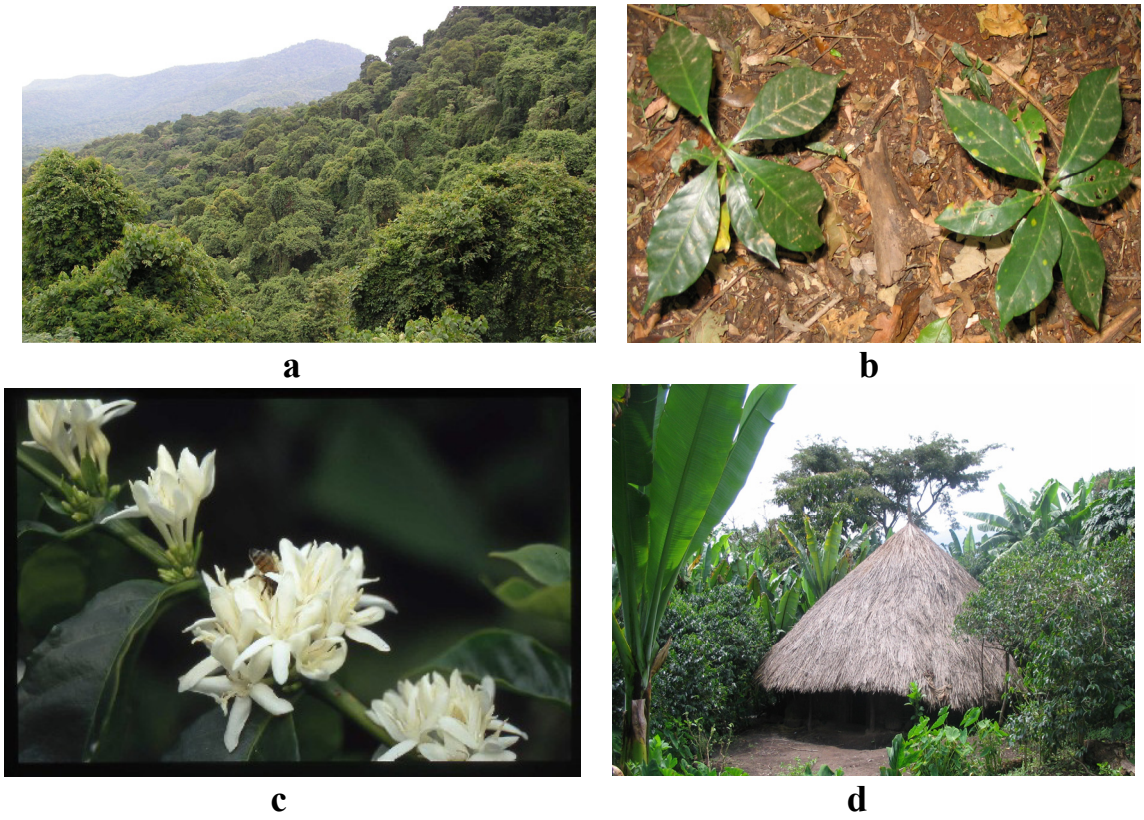


Figure 1.1 Some pictures of coffee from SW, S and SE Ethiopia. (a) Afromontane rainforest with wild coffee Berhane Kontir forest, (b) Naturally regenerated seedling of C. arabica, Bale Mountain (Harrena forest, SE), c) flowers of C. arabica with natural pollinators (honey bee) in Bench Maji, SW, (d) Garden coffee in Dawro Zone, Essara, Southern, Ethiopia. (Photos: a, Kim Govers; b and d, Kassahun Tesfay; c, Christine Schmitt).

1.2.2 Spread of coffee from Ethiopia

The first use of coffee and history of domestication of *Coffea arabica* is not very clear except the most commonly cited legend of goat-herd named Kaldi, who noticed that his goats cavorting excitedly after chewing berries and branch-tips of coffee bushes that he also tasted and enjoyed their stimulating effect. However, early reports show that the roasted and powered coffee were an important travel diet after mixing with butter and fat for the Oromos, one of the ethnic groups in Ethiopia, during long safaris since ancient times (Wellman 1961; Persglove 1968).

No one knows exactly when the first coffee was introduced to Yemen, but it has been estimated at about 575 A.D.. However, the spread of *C. arabica* from Yemen all the way through the world is well documented (Wellman 1961; Meyer 1965). The plant was taken from Yemen to Java (Typica coffee variety) in late 17th century and then to the botanical garden Amsterdam in 1706 and introduced to Latin America early

in the 18th century (Wellman 1961; Meyer 1965; Purseglove 1968). Nowadays, Latin American countries are the major producers of arabica coffee. The spread of cultivation of coffee is shown in Figure 1.2. The variety Bourbon was first taken from Yemen to Bourbon Island (now Reunion) by the French about 1718 and then to countries in Latin America (Persglove 1968).

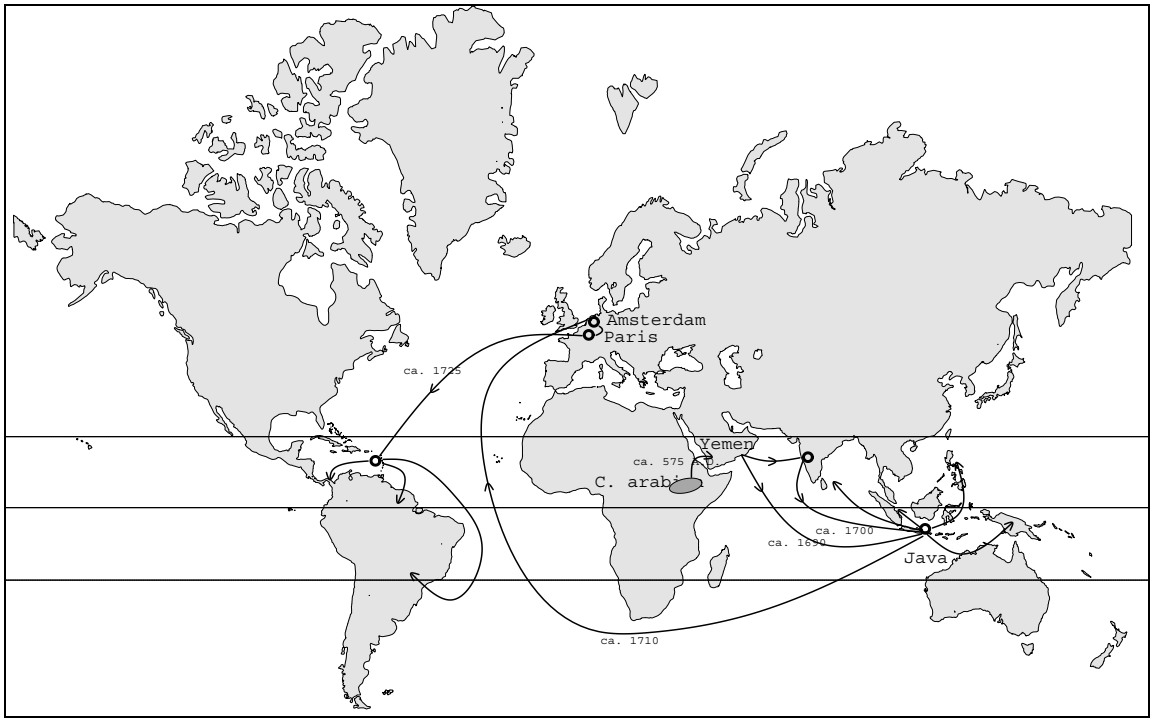


Figure 1.2 Distribution routes for cultivated *C. arabica* in the tropics (Ferwerda 1976). The numbers are the approximate years of introduction.

Coffee as a drink was disseminated directly from Yemen to Europe via Greece and Italy. The first drinking of coffee in Aden took place about the middle of the 15th century and then spread all over the world and became a popular non-alcoholic beverage (Wellman 1961). Apart from the popular coffee drink coupled with a traditional ceremony, coffee is consumed in various forms in Ethiopia and is locally named as *Buna Kella, Buna Besso, Buna Keshir (Hoja), Kuti, Buna Areki, Cheme* (Amaha 1991; Teketay 1999).

1.3 Evolution and origin of *C. arabica*

1.3.1 Hybridization and allopolyploid species evolution

Allopolyploidisation (i.e., genome duplication via hybrid evolution) is a common phenomenon and significant force in the evolution of plants. It is estimated that 50 to 70% of angiosperms are of polyploid origin (Grant 1981; Soltis et al. 1992; Wendel 2000). The genome doubling via autopolyploidy or allopolyploidy has been continuing since angiosperms first appeared. This remains an active and ongoing process and many angiosperm genomes have experienced several cycles of polyploidization at various times in the past. Because of the potentially rapid evolutionary restoration of diploid-like chromosomal behavior it may be difficult to distinguish the ancient polyploidization events (Soltis et al. 1992; Soltis and Soltis 2000; Wendel 2000). Well studied polyploid species are mainly agriculturally important crops. However, the domestication of crops has not favored polyploids over plants with diploid genomes (Soltis et al. 1992; Hilu 1993).

Allopolyploidy is a polyploidization event involving interspecific hybridization and chromosome doubling of fully differentiated parental genomes. It is characterized by permanent heterozygosity resulting from the combination of divergent parental genomes (Roose and Gottlieb 1976; Soltis and Soltis 2000). It is considered to be much more common in nature than autopolyploidy and also, the majority of polyploid cultivated plants are allopolyploids (Soltis et al. 1992; Hilu 1993; Soltis and Soltis 2000).

Chloroplast DNA and nuclear markers can be used to elucidate the genome donor of the cytoplasm as well as the nucleus, and also clarify the mode of allopolyploidization (Soltis and Soltis 1989; Soltis et al. 1992; Widmer and Baltisberger 1999a; 1999b). Restriction fragment analysis of chloroplast DNA of *Tragopogon*, for instance, suggest that *T. porrifolius* has consistently been the maternal parent for *T. mirus* and also reveal a minimum of two independent origins of *T. miscellus* (Soltis and Soltis 1989). The analysis of rDNA showed that *T. mirus* combines the rDNA profiles of the diploids *T. dubius* and *T. porrifolius* (Soltis and Soltis 1991; Soltis et al. 1992). CpDNA has been observed to be an important marker in solving paternity analysis in hybrid speciation in particular and of the maternal lineage in general in angiosperm (Soltis et al. 1992; Soltis et al. 1998). The analysis of 4.3 kb of cpDNA of the

allopolyploid *Arabidopsis suecica*, and its two parental species *A. thaliana* and *A. arenosa* showed that *A. thaliana* is the maternal parent of *A. suecica*, since the sequence were identical in all the cpDNA regions studied. Furthermore, low levels of variation in the allopolyploid *A. suecica* are a strong indication that *A. suecica* has a unique origin in rather recent times (Säll et al. 2003). Moreover, the presence of hypervariable microsatellite sequences in cpDNA makes it useful for the study of genetic relationships and population genetic analyses of plants (Provan et al. 1999; Lira et al. 2003) and studies on the origin of cultivated crops (Ishii et al. 2001; Molina-Cano et al. 2005).

1.3.2 Origin of *C. arabica*

Coffees are members of the tribe *Coffeae* of the large family Rubiaceae and are classified into two genera, *Coffea* and *Psilanthus*. All *Coffea* species are native to the tropical forests of Africa, Madagascar and islands of the Indian Ocean (Mascarene islands), while species of *Psilanthus* occur in Asia and tropical Africa (Bridson and Verdcourt 1988). Most species of *Coffea* are shrubs or small trees with evergreen opposite, petiolate and glabrous leaves. The flowers usually white, with in axillary clusters; calyx shortly tubular, truncate or more or less toothed above; corolla 4-9 loabed, salver-shaped, tube short or elongated; anthers inserted at throat, long, linear, subsessile. Their branching pattern is regular with two opposite lateral branch and one main branch. The anthers are inserted at the throat of the corolla tube by a short filament. Fruits are ellipsoidal, obovate drupes and usually fleshy (Bridson and Verdcourt 1988; Stoffelen 1998; Puff 2003). The genus *Coffea* is subdivided into the two subgenera *Coffea* (*Eucoffea*) and Mascarocoffea (Charrier and Berthaud 1985).The caffeine-containing coffee-trees belong to the subgenera *Coffea* (Charrier and Berthaud 1985*)*. The distribution of this subgenus is ranging from East to Central-West Africa (Charrier and Berthaud 1985; Stoffelen 1998).

All *Coffea* species are diploid (2n=2x=22), except *Coffea arabica* (2n=4x=44) which is autogamous (self-fertile) and considered as allotetraploid (Charrier and Berthaud 1985). Among ca. 100 *Coffea* species in the genus *Coffea*; *Coffea arabica* is the only species occurring in Ethiopia and geographically isolated from the rest *Coffea* species and naturally restricted in to two isolated mountain forests on the western and

eastern sides of the Great Rift Valley in the southern Ethiopia (Mayer 1968; Bridson and Verdcourt 1988; Stoffelen 1998; Gole et al. 2003; Senbeta 2006).

The limited number of phylogenetic studies on *Coffea* genus using molecular markers only allows to infer a group of diploid species as putative closest relatives of *C. arabica* (Lashermes et al. 1997; Cros et al. 1998; Raina et al. 1998; Lashermes et al. 1999). A RFLP (Restriction Fragment Length Polymorphism) analysis of the total cpDNA and the analysis of the *atp*B-*rbc*L intergenic spacer of *Coffea* and *Psilanthus* resulted in only 12 variable characters suggesting exclusively maternal inheritance of *Coffea* cpDNA (Lashermes et al. 1996a). Based on the ITS2 analysis *C. canephora, C. brevipes, C. congensis*, and *C. kapakata* are suggested as the progenitors of *C. arabica* (Lashermes et al. 1997). The lack of resolution among this group of species (the canephoroid group) is most likely caused by the too small number of characters. The analysis of GISH (Genomic *in situ* hybridization) and RFLP data by Lashermes et al. (1999) suggested *C. arabica* as an allopolyploid formed by hybridization between *C. canephora* and *C. eugenioides*. Another study by Raina et al. (1998) on the other hand concludes that *C. congensis* and *C. eugenioides* are the diploid progenitors of *C. arabica*. The *trn*L-*trn*F intergenic spacer sequence analysis also supports *C. eugenioides* as the maternal progenitor; however, the clade was supported only with one restriction site characters and one substitution (Cros et al. 1998). However, in many of these analyses the sample size in terms of genome coverage and number of informative characters was very low which resulted in lower resolution.

1.4 Analyses of genetic diversity within species

The understanding of the amount, the extent and the distribution of genetic variation is vital to the development of effective conservation strategies and use plans. The amount of variation can be very different between species and between different populations of a species (Hodgkin 1997). Understanding the genetic structure of natural populations is one of the central issues in population genetic studies (Epperson and Li 1996). Knowledge about the genetic structure is a fundamental aspect for the understanding of speciation, adaptation or genetic change in plant populations and species (Syamsuardi and Okada 2002). The development and utilization of different marker systems have paramount importance to asses the genetic diversity of a plant species at different levels.

1.4.1 Information from morphological characters

Characterization of diversity has long been based on phenotypic traits mainly. For instance, differentiation among populations was evidenced based on forty-eight morphological traits of *Quercus petraea* (Fagaceae) from five populations in Italy. Furthermore, correlation of morphological variation among population with ecological conditions in the regions of origin was also observed (Bruschi et al. 2003)**.** Substantial phenotypic diversity of tef (*Eragrostis tef*) germplasm from Ethiopia was evidenced recently, which can be utilized in the genetic improvement of the crop. Moreover, analysis of variance using the Shannon Weaver diversity index showed significant regional variation (Assefa 2003). Moderate level of genetic diversity was also observed with six qualitative traits of emmer wheat *(Triticum dicoccum*) collected from Ethiopia (Tesfaye 2000).

The assessment of genetic diversity with morphological traits was also done on Ethiopian tetraploid and hexaploid wheats (Bekele 1984), barley (Demisse 1996), coffee (Montagnon and Bouharmont 1996) and sorghum (Ayana and Bekele 1998). However, morphological variability is often limited since the characters are mainly affected by environment. Moreover, morphological traits might be insufficient to differentiate among pairs of closely related species and ecotypes since not all genetic differentiation results in morphological differentiation (Siva and Krishnamurthy 2005). Currently, different molecular markers have been proposed to assess genetic variability as a complementary strategy to more traditional approaches in genetic resources management (Terzi et al. 1999).

1.4.2 Information from biochemical data

Storage protein and isozymes are the main biochemical markers used for characterization of plant genetic resources, relationships at lower taxonomic levels as well as to detect geographic origin. Seed protein profile studies have been done with various crop plants, such as *Coffea* (Bau et al 2001), Tef (*Erogrostis;* Bekele and Lester 1981) , *Oryza* (Poaceae; Montalvan et al. 1998), *Capsicum* (Solanaceae; Panda et al. 1986), *Arachis* (Fabaceae; Lanham et al. 1994) and emmer wheat (Poaceae; Tesfaye 2000). Allozymes have also been utilized to understand patterns of genetic variation and

the structure of plant populations in many taxonomic groups including *Coffea* (Berthou et al. 1980; Hamrick and Godt 1990).

The technique is rapid and economical, and codominant nature of allozyme data makes it useful for the characterisation of genetic variation in plant species (Weising et al. 2005). However, compared to DNA markers, the limitation is that it provides only poor coverage of genome because of small numbers of marker loci available. Moreover, it underestimates the variation since it needs ca. 30% substitutions for resulting in polymorphic fragment patterns, and it is difficult to interpret allozyme patterns of polyploid species (Liu et al 2002; Weising et al. 2005)

1.4.3 Information from molecular markers

Development of DNA-based genetic markers for assessing levels of genetic diversity, and its patterns of distribution within a phylogenic context have made it possible to understand factors influencing genetic diversity and its patterns of distribution. The DNA based marker system has evolved from hybridization-based to PCR-based techniques (Weising et al. 1995; Hoisington 2001).

RFLPs are the most widely used hybridization-based DNA marker system and were initially developed for human-genome mapping (Botstein et al. 1980). Since then it has been employed to study the relationships in plant species with fragments derived from the nuclear, chloroplast and mitochondrial genome (Berthou et al. 1983; Debener 1990; Lashermes 1996b; Weising et al. 2005). The DNA profiles of the restriction enzyme-digested DNA are visualized by first separating the fragments according to their size by gel electrophoresis and then hybridizing to a labeled probe, which is a DNA fragment of known sequence (Weising et al. 1995; 2005). The high reproducibility and codominant nature of the RFLP marker system is considered as the main advantage. However, main drawbacks are the labor intensive and time consuming nature of the techniques.

PCR-based markers involve *in vitro* amplification of particular DNA sequences or loci, with the help of specifically or arbitrarily chosen oligonucleotide sequences (primers). The amplified fragments are separated electrophoretically and banding patterns are visualized by different methods such as staining and autoradiography. There are a variety of PCR-based marker systems currently available

for the analysis of relationships in plants and also to study the genetic diversity and its patterns of distribution.

The dominant marker systems are multi-locus approaches with no possibility of discrimination between homozygous and heterozygous individuals. Some of these marker systems are Random Amplified Polymorphic DNA (RAPD), Inter-Simple Sequence Repeat (ISSR) and Amplified Fragment Length Polymorphisms (AFLP). The nuclear microsatellites are codominat marker systems and amplify single loci with multiple alleles depending on the homo- or heterozygous state in the diploid individuals (Sergio and Gianni 2005; Weising et al. 2005). All the PCR-based DNA marker systems have their own advantages and limitations, hence the choice of markers can relay up on the research question to be addressed, cost, level of polymorphism expected and the taxonomic unit considered (Karp et al. 1997). The usefulness and efficiency of each marker system are also very different in terms of the number of loci and polymorphisms detected (Nagaraju et al. 2001; Tosti 2002; Bolibok et al. 2005).

DNA sequencing from nuclear, chloroplast and mitochondria genomes also provides information on the intra- and inter-specific phylogenetic relationships of plants (Hillis et al. 1996; Soltis and Soltis 1998; Weising et al. 2005). Moreover, it is also utilized for population genetic studies and haplotyping (Widmer and Baltisberger 1999b; Bakker et al. 2000; Zhang and Hewitt 2003). The low rate of sequence evolution of cpDNA and uniparental modes of inheritance makes it useful to study relationships of plants at different taxonomic levels (Palmer et al. 1988; Soltis et al. 1992). Furthermore, the presence of hyper-variable microsatellites in cpDNA that is characteristically composed of mononucleotide repeats increases the utility at lower taxonomic levels, e.g. to study genetic diversity and differentiation at regional scales (Bryan et al. 1999; Weising and Gardner 1999; Weising et al. 2005).

Molecular marker systems contribute to a better understanding of the factors influencing genetic diversity and its patterns of distribution. Molecular markers have also been used to describe mating patterns within populations (Cruzan 1998). A review of studies of RAPD markers and comparison with allozyme data of Hamrick and Godt (1990) was made on life history traits such as taxonomic status, life form, geographic range, breeding system, seed dispersal and successional status by Nybom and Bartish (2000). It was observed that long-lived, outcrossing, late successional taxa retain most

of their genetic variability within populations. However, annual, selfing and/or early successional taxa allocate most of the genetic diversity among populations (Hamrick and Godt 1990; Nybom and Bartish 2000). Population history was also observed to influence genetic diversity and its patterns of distribution in *Vincetoxicum hirundinaria* Medik (Asclepiadaceae; Leimu and Mutikainen 2005). For maintenance of genetic diversity with the appropriate conservation strategy and sustainable utilization measures these information have paramount importance.

1.5 The gene pool of wild *C. arabica* in Ethiopia

Ethiopia is considered to be one of the important centers of origin and diversity for many cultivated crops and their wild relatives (Vavilov 1951; Engels and Hawkes 1991). The entire diversity of arabica coffee also originated mainly from SW and SE mountain rain forest (Geber-Egziabher 1990; Tadesse and Nigatu 1994; Gole et al. 2001; Senbeta 2006). Early travels and scientist including botanist and geneticist agree that arabica coffee is native to the highlands of Southwestern Ethiopia (Sylvian 1958; Mayer 1965; Gebre-Egziabher 1990; Friis 1979).

Most of the existing genetic diversity studies of *C. arabica* were based on materials outside Ethiopia and describe so called “spontaneous” or “subspontaneous” materials. The study of phenotypic variability of *C. arabica* indicated the existence of variation and grouping based east-west of Great Rift Valley (Montagnon and Bouharmont 1996). The RAPD analysis of *Coffea* species detects intraspecific variation in *C. canephora* and *C. liberica* but not in *C. arabica* (Lashermes et al. 1993). However, analysis with the same marker system by Anthony et al. (2001) point toward the existence of higher genetic diversity among samples collected from SW part of Ethiopia. The combined analysis of AFLPs and microsatellites pointed toward the lower genetic diversity in cultivated samples than the “subspontaneous” samples from Ethiopia (Anthony et al. 2002). The analysis of genetic diversity with nuclear microsatellites showed the existence of higher allelic diversity in wild gene pools than the cultivated. This further suggested that wild coffee in Ethiopia can serve as a source of novel genetic variation to broaden the genetic base of *C. arabica* cultivars (Moncada and McCouch 2004).

The recent study of morphological characters in *C. arabica* in Ethiopia confirms the presence of higher phenotypic diversity among germplasm collected and maintained in the *ex situ* gene bank in the Jimma and Awada Agricultural Research Centers. These could be utilize for the improvement of *C. arabica* in Ethiopia (Kebede and Bellachew 2004a; Kebede and Bellachew 2004b; Seifu et al. 2004). The existence of moderate genetic diversity also is indicated with molecular markers in the samples from SW and SE forests coffee in Ethiopia (Aga 2005). However, most of the above mentioned studies were not exhaustive in terms of both spatial coverage of the range of *C. arabica* and forest coffee priority areas of Ethiopia (Teketay et al. 1998; Dubale and Tektay 2000). Moreover, the documentation of exact geographical location of the samples was also not very clear.

The existence of wider genetic diversity also observed with the presence of diseases in the wild coffee with out having serious damage. The disease outbreak of Coffee Berry Disease (CBD) in Ethiopia was confirmed in 1971 and since then the disease became an important production constraint of coffee in Ethiopia (Ameha and Belachew 1983, 1984; Melaku 1984). However, the successful screening of *C. arabica* selections for CBD resistance in relatively short time showed the existence of broad genetic diversity in the wild gene pool (Ameha and Belachew 1984; Van Der Graaf 1981; Bellachew et al. 2000). Studies also indicated the existence of variations in resistance levels of coffee genotypes in Ethiopia for coffee wilt disease of coffee caused by a fungus *Gibberella xylarioides* (*Fusarium xylarioides*; Adugna and Hindorf 2001). The existence of physiological races of coffee leaf rust (*Hemileia vastatrix*) and variation in the level of resistance noticed for coffee from southwestern part of Ethiopia which indicates the presence of much larger diversity in wild coffee in south west than the farmer's variety in SE (Sylvain 1955, 1958; Wondimu 1987; Montagnon and Bouharmont 1996). Moreover, the resistance study of coffee trees to *Meloidogyne incognita* demonstrated that 40% of the semiwild Ethiopian accessions tested were totally resistant (Anzueto et al. 2001).

The plant samples from Kaffa and Illubabor showed lower caffeine content as compared to the reference commercial cultivar of "Catuaí Vermelho" and "Mundo Novo" (Silvarolla et al. 2000). The recent discovery of the naturally decaffeinated arabica coffee from the materials collected from Ethiopia and the possibility of

transferring the trait via intraspecific hybridization to high yielding commercial variety also an indication for the value of the Ethiopian wild gene pool for further improvement of quality *C. arabica* (Silvarolla et al. 2004). Generally, high economic value of wild *C. arabica* genetic resources in Ethiopian were confirmed based on an assessment of the potential benefits and costs of the use in breeding programs for enhanced coffee cultivars (Heina and Gatzweiler 2005).

1.6 Importance of genetic diversity analyses for genetic resource management

The characterization of genetic diversity of plants provides insights into the evolutionary history of a taxon, on geographic and ecological aspects of the extent and distribution of genetic diversity and the ecological aspects of the processes that have given rise to observed patterns of variation (Hodgkin et al. 2001; Neel and Ellstrand 2003; Weising et al. 2005). Understanding the genetic diversity in natural populations will have paramount importance for a better management of the resources. To withstand the changing environment and biotic stress like disease and pest conserving the genetic diversity within individual species is crucial factor (Faith 1994; Amos and Balmford 2001; Rauch and Bar-Yam 2004).

The genetic diversity within species is distributed unevenly (Rauch and Bar-Yam 2004). Hence, the prior knowledge of the nature, extent and distribution of genetic variation is crucial for successful conservation (*in situ* and *ex situ*) and sustainable utilization of germplasm. Moreover, the number of populations necessary to conserve genetic diversity within a species and choice of sites for *in situ* conservation depends on the measure of diversity and its patterns of partition within and among populations (Petit et al. 1998; Neel and Cummings 2003). The analysis of genetic diversity using molecular marker is also essential to detect population bottlenecks in threatened species since bottlenecks can increase the risk of population extinction (Luikart et al. 1998). Furthermore, the information on the genetic diversity will also be important for the choice of management alternative for capturing and maintaining the available genetic variation. This can reduce the number of populations needed to commit to conservation, thus reducing costs and conflicts with competing land uses (Neel and Ellstrand 2003).

1.7 The CoCE project (Conservation and use of wild populations of *Coffea arabica* in the montane rainforests of Ethiopia)

The CoCE project is an interdisciplinary research approach designed to allow the development of conservation and use concepts that are ecologically sustainable, economically efficient and socially acceptable. It involves the biological field such as comparative vegetation analysis of the forest, molecular genetic diversity of wild coffee, ecophysiological analysis and phytopathological response of wild coffee. Furthermore, assessments were made on economics of conserving wild coffee in the montane rain forest with its optimized institutional structure arrangement.

Generally the CoCE research project aims to assess the diversity and the economic value of the Ethiopian coffee gene pool and develop concepts of model character for *in situ* conservation and use of the genetic resources of *Coffea arabica*. Moreover, expanding the laboratory capacities at the Department of Biology (Addis Ababa University, Ethiopia) for routine screening and evaluation of the genetic diversity of *C. arabica* was also one of the aims as well. The conservation concept is to safeguard wild coffee populations with its natural habitat and also in the traditional forest coffee systems. This PhD research work is formulated with in the frame work of CoCE project to execute the genetic diversity analysis of wild *C. arabica* population in montane rainforests (www.coffee.uni-bonn.de). Some of the research activities of the authors are shown in Figure 1.3.

a

b

c

Figure 1.3 Some of the field and lab research activities in Ethiopia. (a) sample collection in one of the core CoCE site, Yayu/Geba Dogi forest, (b) labeled coffee tree in Bonga forest (Core CoCE site), and (c) Genomic DNA extraction for silica gel dried leaves of C. arabica in Genetic Research lab of Department of Biology, Addis Ababa University. (Photos: a, Thomas Borsch; b and c, Kassahun Tesfaye).

1.8 Aims and scope of this study

Generally, man's dependence on centers of genetic diversity for his plant and animal germplasm resource is becoming very acute because of the high rate of genetic erosion (Bekele 1983). The forest coffee ecosystem of Ethiopia is also highly threatened by settlement and land-use pressures. Establishment and expansion of big farms, migration and human settlements are the main causes for deforestation. The montane rain forests in SW and SE parts of Ethiopia, which is the natural home of the last wild *C. arabica,* are declining at an alarming rate. The endemic wild coffee populations of *Coffea arabica* are in the risk of being extinct (Gole *et. al*, 2001; Richerzhagen and Virchow 2002).

Suggestions were made by different researchers for the need of conservation of wild populations of *C. arabica* in Ethiopia since the risk of loss of the last remaining wild coffee populations is currently high (Melaku 1984; Geber-Egziabher 1990; Amaha 1991; Teketay et al. 1998; Dubale and Tektay 2000; Gole et al. 2002; Aga 2005; Senbeta 2006). For any conservation measures, knowledge of the amount of variability and its patterns of distribution should be obtained. Hence, the systematics, and the extent and distribution of genetic diversity of wild coffee using cpDNA and ISSR (Inter Simple Sequence Repeat) markers were carried out to identify sites with higher diversity and appropriate for *in situ* conservation. The main objectives of the study are:

1. To detect highly variable regions of cpDNA so as to observe infraspecific variation among wild populations *C. arabica*. Then, to characterize cpDNA microsatellites and to detect SNPs (Single Nucleotide Polymorphisms) within *C. arabica*. Moreover, a phylogenetic approach should allow clarifying the maternal parentship among the putatively closely related *Coffea* species.

2. Using an ISSR marker system it is aimed to evaluate the relationships of *Coffea arabica* populations throughout Ethiopia. Moreover, it should be evaluated if individuals from wild populations in forests are genetically different from landraces selected by farmers.

3. To evaluate the molecular diversity and detailed patterns of distribution within the regions Berhane Kontir and Yayu/Geba Dogi. Thereby, a dense population sampling should be used.

4. To apply molecular data on the distribution of genetic diversity in between and within regions for backing up conservation and use strategy.

2 EVOLUTION OF *COFFEA* CHLOROPLAST MICROSATELLITES AND EVIDENCE FOR THE RECENT DIVERGENCE OF *C. ARABICA* AND *C. EUGENIOIDES* CP GENOMES

2.1 Introduction

The genera *Coffea* and *Psilanthus* constitute the tribe Coffeae of the large angiosperm family Rubiaceae (e.g., Ehrendorfer et al. 1994; Stoffelen 1998). *Coffea* is further divided into two subgenera, *Coffea* (which sometimes has been called '*Eucoffea*`) and *Mascarocoffea* (Bridson 1982; Charrier and Berthaud 1985; Bridson and Verdcourt 1988). All caffeine-containing species of *Coffea* belong to subgenus *Coffea* and are distributed in East, Central and West Africa, while *Mascarocoffea* is native to Madagascar, the Mascareign and Comores Islands in the Indian Ocean (Berthout et al. 1983; Charrier and Berthaud 1985; Cros et al. 1998).

Coffea L. includes ±100 species (Bridson 1982; Charrier and Berthaud 1985; Bridson and Verdcourt 1988; Stoffelen 1998) all of which are diploid (2n = 2x = 22) except *Coffea arabica* L. (2n=4x=44) that is an allotetraploid (Charrier and Berthaud 1985; Raina et. al 1998; Lashermes et al. 1999). *C. arabica* is a small tree up to 4 - 6 m tall. The species has been considered as being generally autogamous (Charrier and Berthaud 1985; Bridson and Verdcourt 1988; Free 1993). *Coffea arabica* is naturally occurring in the highlands of Ethiopia, and some populations were also reported in south-eastern Sudan and northern Kenya (Friis 1979; Charrier and Berthaud 1985; Gole 2003; Puff 2003). The wild Ethiopian populations occur in Afromontane rain forests between 1400 and 1900 m (Geber-Egzabeher 1990; Gole et al. 2001; Gole 2003), where *Coffea* is self regenerating but cherries are frequently collected by the local communities. In addition, a high number of regionally selected farmer's varieties (landraces) exists (Geber-Egzabeher 1990; Gole et al. 2001; Gole and Teketay 2001). The natural extent range of *Coffea arabica* does not overlap with any other species of *Coffea.* Geographically closest are *C. eugenioides* Moore occurring in Uganda, Rwanda, Burundi, Congo, western Kenya, Tanzania, and *C. canephora* Pierre ranging from Uganda to central and West Africa (Hutchinson and Dalziel 1963; Bridson and Verdcourt 1988). The commercial coffee production relies on two species, *C. arabica* and *C. canephora*. From these two the better quality is usually associated with *Coffea*

arabica, contributing more than 80 percent of the world's coffee production (Coste 1992; Purseglove 1968; Raina et al. 1998).

So far published phylogenetic analyses of *Coffea* based on DNA sequences revealed several species groups that largely correspond to certain geographical regions. However, trees suffer from rather low resolution and support what may be attributed to a recent origin of the genus *Coffea* (Lashermes et al. 1997; Cros et al. 1998; Mvungi pers.comm.). Parsimony analysis of ITS2 sequences resolved *C. arabica* within a canephoroid group (*C. brevipes*, *C. canephora* and *C. congensis)* but all species appeared in a polytomy, and the group itself gained only 53% bootstrap support (Lashermes et al. 1997). Cros et al. (1998) sequenced the *trn*L-F spacer, and analyzed restriction site data from the whole cp genome. They found *C. arabica* in a clade with *C. eugenioides* and '*C. sp. Moloundou*`. Although this clade received a bootstrap value of 100% it is supported by only a single substitution in the *trn*L-F spacer and one restriction site change. Interestingly, the two samples of *C. eugenioides* differed by an indel in the *trn*L-F spacer. Mvungi (pers.comm.) did not include *C. arabica*. Based on GISH and FISH analyses Raina et al. (1998) suggested that *C. congensis* and *C. eugenioides* are the diploid progenitors of *C. arabica.* Moreover, Lashermes et al. (1999) compared the chromosomes of *C. arabica*, *C. canephora* and *C. eugenioides* using GISH, and found that the genome of *C. arabica* is an amphidiploid formed by hybridization. The origin of *Coffea arabica* and the exact kind of its parental genomes therefore remains to be substantiated. This extends to the question whether *Coffea arabica* is the result of a single or of multiple allopolyploidization events, as they have been reported for a number of other polyploid species (Soltis and Soltis 1989; Soltis et al. 1992; Soltis et al. 1998; Xu et al. 2000; Soltis 2005). Lashermes et al. (1999) argue against that possibility of multiple origins considering the low amounts of genetic variation observed among accessions of *C. arabica*. However, neither a representative sampling of documented wild populations of *C. arabica* nor putative closely related species were included, and therefore the study was limited with respect to find alleles that could hint to differing parental genomes.

Samples used in so far published analyses on genetic diversity of *C. arabica* in Ethiopia mostly are based on "subspontaneous" materials which are not exactly traceable in terms of geography and status of their respective populations (whether

being from authochtonous self regenerating forest populations or representing farmer´s varieties). Also, it appears that several geographically distinct Ethiopian populations have not yet been considered such as those of the Harrena forest (Bale) east of the Rift Valley. Present assessments of genetic diversity based on RAPD (Lashermes et al. 1996; Aga et al. 2003), RFLP (Lashermes et al. 1999), AFLP (Anthony et al. 2002) and SSR data (Anthony et al. 2002) nevertheless indicate that the commercially grown Bourbon and Typica cultivars are derived from a common ancestor. These analyses further provide evidence for higher genetic diversity among "subspontaneous" individual materials than individuals from commercial cultivars.

As part of an ongoing effort to develop strategies for conservation and sustainable use of wild *Coffea arabica* genetic resources in the forests of Ethiopia (The CoCE Project, www.coffee.uni-bonn.de) the genetic diversity of *Coffea arabica* in its native range is being evaluated. To achieve this, different molecular marker systems are analyzed. Such different marker systems together provide wider genome coverage. The analysis of cpDNA in forests tree species also exhibited diversity and differentiation among populations in space and time (Vendramin et al. 1999; Caron et al. 2000; Grivet and Petit 2002; Petit et al. 2002). Chloroplast genome variation will unravel maternal lines of inheritance (Provan et al. 2001; Säll et al. 2003; Patterson et al. 2005) and allow to detect putative introgression patterns (Palme and Vendramin 2002; Xu et al. 2000). This is particularly important in the context of allopolyploid speciation. Plastome haplotypes are also useful markers for revealing infraspecific phylogeographic patterns (Vendramin et al. 1999; Grivet and Petit 2002; Dane and Lang 2004).

Introns and spacers of the rapidly evolving SSC chloroplast genome regions are effective in providing data for relationships among closely related species (Kelchner 2000; Shaw et al. 2005). Moreover, these introns and spacers often have variable sites within species (Soltis et al. 1992; Church and Taylor 2005). The chloroplast genome has also microsatellites, for which universal primers have been developed (Weising and Gardner 1999). In a number of studies chloroplast microsatellites provided insights into infraspecific phylogeographic variability (Echt et al. 1998; Provan et al. 1999; Deguilloux et al. 2004) and allowed to characterize different cultivars of crops (Arroyo-García et al. 2002; Molina-Cano et al 2005; Sukhotu et al. 2005). It needs to be emphasized, however, that chloroplast infraspecific variability is not restricted to length

variable satellites, and that single nucleotide polymorphisms (SNPs) have been found to characterized geographically distinct populations for example in *Nymphaea odorata* (Woods et al. 2005).

Aims of this study are twofold. The first aim is to detect regions in the chloroplast genome with infraspecific variation in *C. arabica,* and the second aim is to increase the number of characters for clarifying the maternal parentship among the putatively closely related *Coffea* species. A sequencing approach seemed appropriate which at the same time could serve to characterize microsatellites and to detect SNPs within *C. arabica* as well as to generate data for inferring plastome trees.

2.2 Materials and Methods

2.2.1 Plant material and sampling

This study includes seven diploid and one tetraploid species (*C. arabica*) as well as one hybrid (*C. liberica* x *C. arabica*) of the genus *Coffea.* Samples from *C. arabica* were collected from forest populations throughout Ethiopia and from two commercial cultivars representing the Typica (cv. Blue Mountain) and Bourbon (cv. Caturra) lines, respectively. The forest populations are located in the Bale (Harrena forest), Bonga, Berhane Kontir, Yayu (Geba Dogi forest), and Boginda regions and are study sites of the CoCE project. Individuals sampled are from wild, self-regenerating populations, and are different from framer's varieties (Tesfaye et al. submitted (b)) based on fingerprint data. Considering analyses of relationships in *Coffea* (e.g., Cros et al. 1998; Lashermes et al. 1996; Mvungi pers. comm.) a species of *Psilanthus* is used as outgroup for *Coffea*. Moreover, *Ixora coccinea* representing the tribe Ixoreae of Rubiaceae has been sequenced as a more distant outgroup. Taxa and sources of materials are listed in Table 2.1.

2.2.2 DNA isolation

Genomic DNA was isolated from silica-gel-dried leaf tissue with a modified CTAB method employing triple extractions to yield optimal amounts of DNA (Borsch et al. 2003). Since polymerase chain reaction (PCR) was inhibited in some samples of *Coffea* due to the presence of secondary compounds such as alkaloids (caffeine, trigonelline),

we further purified genomic DNA using the QIAquick PCR purification kit (Qiagen Inc., Valencia, California).

Table 2.1 List of Coffea species and outgroup used in this study with its country of origin.

Taxon	**Code**	**Field /Garden origin**	**Accession numbers**
C. arabica 'Blue Mountain'	C002	Tropenzentrum Witzenhausen Uni Kassel	
C. arabica x *C. liberica*	C003	Tropenzentrum Witzenhausen Uni Kassel	-
C. arabica 'Caturra Yellow-Bourbon'	C007	Tropenzentrum Witzenhausen Uni Kassel	-
C. liberica	C008	Tropenzentrum Witzenhausen Uni Kassel	-
C. congensis	C009	JARC, Ethiopia, [Basins of Nile -Congo River Congo]	-
C. eugenioides	C010	JARC, Ethiopia, [Congo]	-
C. salvatrix	C011	JARC, Ethiopia, [Origin-Mozambique]	-
C. canephora	C013	JARC, Ethiopia, [Central & West Africa - Congo]	-
C. kapakata	C014	JARC, Ethiopia	-
C. stenophylla	C015	JARC, Ethiopia EARO, JARC, [French Guinea]	-
C. arabica wild	C016	Ethiopia, Wild, Kaffa, Bonga Forest	2716, AAU*, National Herbarium
C. arabica wild	C034	Ethiopia, Wild, Bale, Harrena Forest	2707A, AAU*, National Herbarium
C. arabica wild	C035	Ethiopia, Wild, Berhane kontir Forest	2714, AAU*, National Herbarium
C. arabica wild	C036	Ethiopia, Wild, Yayu, Geba Dogi Forest	2711, AAU*, National Herbarium
C. arabica wild	C037	Ethiopia, Wild, Keffa, Boginda Forest	2715, AAU*, National Herbarium
Psilanthus leroyi Bridson	C033	Herbarium sample, Nees Institute for Biodiversity of Plants, University Bonn [Omo Nat. Park, ETH]	Schloer & Jacobs 424
Ixora coccinea	C017	Botanical Garden, University of Bonn	BG Bonn 5063

**Addis Ababa University*

2.2.3 Selection of genomic regions and primer design

The cp genome regions analysed in this study include both introns and intergenic spacers (Figure 2.1). These were selected based on previous reports of microsatellites in other angiosperms (Weising and Gardner 1999) or because variability was anticipated to be high among closely related species based on other projects. In addition, the fast evolving *mat*K gene was included because it could easily be amplified and sequenced together with its flanking *trn*K intron parts. If not yet available, amplification primers were designed using the complete cp genome sequences of *Nicotiana tabacum* (GenBank accession no. NC001879), *Arabidopsis thaliana* (NC000932), *Atropa belladonna* (NC004561), and *Spinacia oleracea* (NC002202). A number of internal

sequencing primers were designed based on *Coffea* sequences obtained in this study. Sequencing longer microsatellites (poly A/T's) usually required to place primers about 300 bp away and to sequence both strands of DNA.

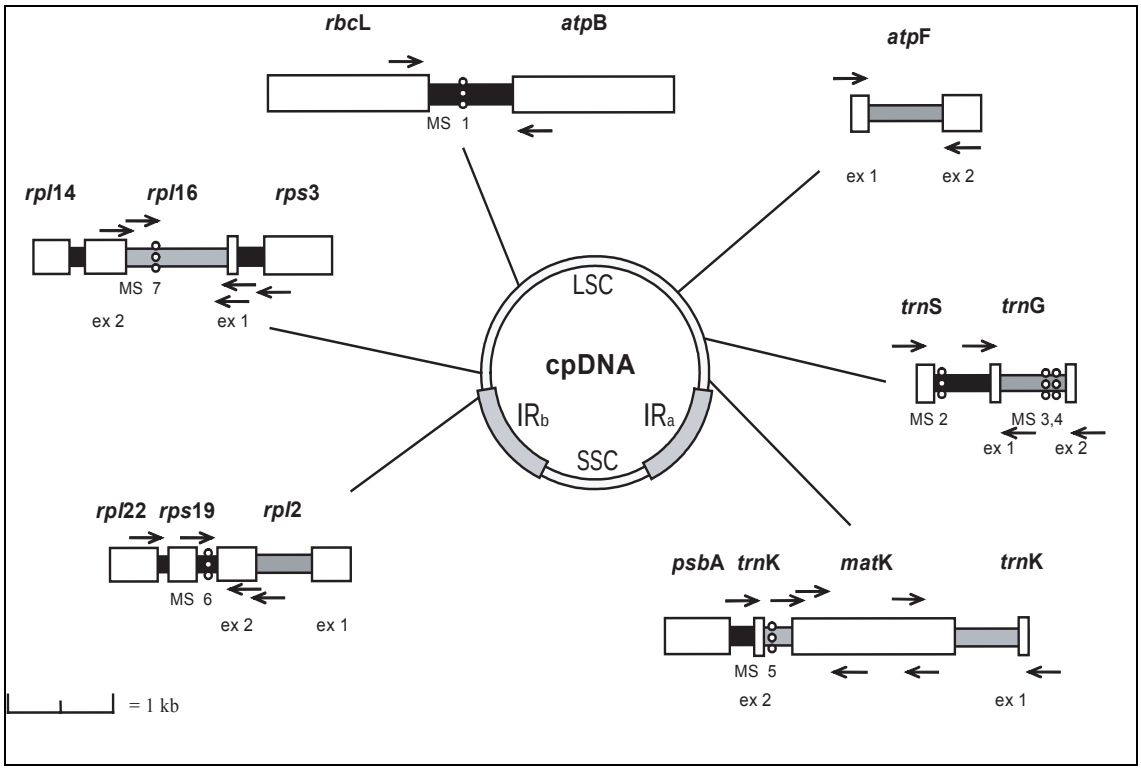


Figure 2.1 Structure and position of the rapidly evolving chloroplast genome regions sequenced. Genes or exons are illustrated as empty boxes, introns as grey and spacers as black bars. Presence of microsatellites (MS) is indicated by small circles, and their position is indicated relative to the extension of the respective genomic regions. Arrows indicate primers used for amplification and sequencing.

2.2.4 Amplification and sequencing of cp genomic regions

Two different PCR protocols were used on a Biometra T3 Thermocycler. For all regions except *mat*K/*trn*K reactions were carried out in 50µl volumes containing 4µl genomic DNA diluted to 1:10, 23.7µl H_2O, 5µl 10x Taq buffer incl. 15 mM $MgCl_2$, 3µl 25mM $MgCl_2$, 2µl each of forward and reverse primers (20 pm/µl), 10µl dNTP (each 1.25mM), and 0.3µl Taq (5 units/µl). To amplify *mat*K/*trn*K the amount of Taq was 0.2µl (5 unit/µl) without additional $MgCl_2$. Amplification conditions were: 34 cycles of 94°C (1 min) denaturation, 48°C (1 min) annealing , 72°C (2 min) extension and 72°C (15 min) final extension for *rpl*16; 96 °C (1min, 30 sec), 50 °C (1 min), 72 °C (1 min 30 sec), 95

°C (30 sec), 50 (1 min), 72 (1 min 30 sec), 34 cycles of steps 4-6, 72 °C (20 min) for *atp*B-*rbc*L, *atp*F, *trn*S-G and *trn*G, *rpl*2-*rpl*22; 96 °C (3min), 50 °C (3 min), 72 °C (3 min), 94 °C (30 sec), 48 (1 min 30 sec), 72 (3 min), 39 cycles of steps 4-6, 72 °C (20 min) for *mat*K/*trn*K. PCR products were run on agarose gels, and bands were excised and further purified with columns (QIAquick gel extraction kit, QIAGEN, Hilden/Germany). Purified products were directly sequenced with the ABI BigDye Cycle Sequencing Ready Reaction Kit, V1.1. Sequencing products were electrophoresed on a Herkules-1412-020 automated sequencer.

The *atp*B-*rbc*L spacer was amplified and sequenced with primers atpB-ol2 and rbcL-ol5 (Manen et al. 1994) annealing to the *atp*B and *rbc*L genes, respectively. The *trn*S-G spacer and the *trn*G intron were amplified in two overlapping halves using primers trnS-40F (5' CAT TAC AAA TGC GAT GCT CT 3') and CotrnG-113R (5'ATG TAT ATT TTT CTC GAA TCT G 3') for *trn*S-G and trnG-1F (5' TAG CGG GTA TAG TTT AGT GG 3') and trnG-725R (5'ATC GTT AGC TTG GAA GGC T 3') for the *trn*G intron. The primers were designed in this study and also used for sequencing. For the *atp*F-intron primers atpF-97F (5' AAT CT[A/C] AGT GTA GT[G/C] CTT G 3') and atpF-914R (5' TTG TTC AAT AGC CCC T[C/T]C 3'), designed in this study, were used for amplification and sequencing. The *rpl*2-*rps*19-*rpl*22 region was amplified as a whole, with two alternative forward primers rpl2-1090F (5' AAT CTA CTT CAA CCG ATA TG 3') and rpl2-720F (5' A[C/T]A TAG AAA TCA CAC TTG G 3') and the reverse primer rpl22-250R (5' GCT TTA CTA ATG ACT AAA TTG G 3'). Additional internal sequencing primes were rpl2-351F (5' TTA TCC TGC ACT TGG AAG AA 3') and rps19-106R (5' TAT AAT GGT [A/G]GA TGC CCG 3'). All primers were designed in this study. The *trn*K intron was amplified including *mat*K using primers trnKFbryo (Quandt in press) and trnK2r (Johnson and Soltis 1995) that anneal to the *trn*K exons. Further internal sequencing primers were NYmatK480F (Borsch 2000) and ACmatK700R (Müller and Borsch 2005), as well as ComatK1118F (5' CAC TTA TTC CCC TTA TTG G 3'), ComatK969R (5' AAG TTA AAG GAA TAA TTG GG 3'), ComatK-670R (5' CTC CAA AAG ATG TTG ATC G 3'; the latter three designed in this study. For the *rpl*16 intron new amplification primers were designed, annealing in greater distance to the intron as compared to previously published primers (e.g., Kelchner and Clark 1997). These are rps3F (5'- ATC TAT GGA GT[A\T] TTA GT-3')

and rpl16R (5'- TCT TCC TCT ATG TTG TTT ACG-3'). Additional internal sequencing primers also designed in this study are Corpl16-67F (5' CTA ATA ACC AAC CCA TCA C 3'), Corpl16-130F (5' TAT CAC ATC CTT GCA GC 3') and Corpl16-940R (5'GAG GTA AGA TAC AAT TTT C 3').

2.2.5 Alignment

Alignment followed the rules layed out in Borsch et al. (2003) and Löhne and Borsch (2005), based on the assumption that length variability is created by microstructural mutation events that can comprise one to several nucleotides at once. Alignments were carried out through visual observation using QUICKALIGN 1.6, designed for optimal manual sequence modification (Müller and Müller 2003). Since distances of sequences were generally low, no mutational hotspots other than microsatellites had to be considered as regions of uncertain homology and excluded from phylogenetic analysis (for utilization of microsatellite variation in separate matrices see Figure 2.2).

2.2.6 Coding of indels

Indels were coded according to the simple indel coding approach (Simmonds and Ochoterena 2001), and considering adjacent independent gaps (Löhne and Borsch 2005) as separate characters. Character state ‘1’ was assigned when sequence was present in the respective taxon, and character state ‘0’ if there was a gap. All characters were analysed as unordered.

2.2.7 Analysis of microsatellites

Microsatellites were consecutively numbered (MS1, MS2, MS3; Figures 2.1, 2.2). Because of the difficulty of assessing homology of individual repeats within length variable microsatellites, neither simple nor complex indel coding methods could be applied. We therefore interpreted a satellite region as a single locus, which translates to a single character, and scored the repeat stretches of different lengths as different multiple states. To facilitate the analysis of microsatellite evolution, state numbers were given consecutively with increasing number of repeat units. The character states were analysed as ordered (additive) based on the hypothesis that microsatellites evolve in a

stepwise process involving A/T or AT/TA as repeat units (Zhu et. al. 2000; Petit et al. 2005).

2.2.8 Phylogenetic inference

Phylogenetic analyses were carried out for data partitions (individual genomic regions, all spacers, all introns, or all genes) or all partitions combined. Analyses were carried out based on substitutions only, or adding indel and microsatellite matrices to the dataset combining all partions. All characters were equally weighted and gaps were treated as missing characters. Maximum parsimony (MP) tree inference used PAUP* 4.0b10 (Swofford 2002) with 1000 random addition replicates and TBR branch swapping. PAUP also served to calculate consistency (CI), retention (RI) and rescaled consistency indices (RC). Jackknifing was done with 10.000 replicates (38% deletion), and searches with 10 random addition replicates in PAUP. Decay values were calculated with PRAP (Müller 2004).

Bayesian Inference (BI) was performed with the overall combined substitution based matrix using MrBayes V3.0 (Huelsenbeck and Ronquist 2001). To measure the support of individual nodes posterior probabilities were calculated with MrBayes. Based on the Akaike Information Criterion (AIC) in Modeltest 3.06 (Posada and Crandall 1998) a "GTR + I + G" model of molecular evolution was implemented. Four runs of Metropolis-coupled Markov Chain Monte Carlo were done, each with four chains and saving one tree every 100 generations for 1,000,000 generations, starting with a random tree. The temperature was set to 0.2 and the burn in at 5000 generations. Initial experiments to root trees including *Coffea* species and *Psilanthus* with *Ixora coccinea* showed strong long branch attraction effects of the outgroup to the *C. arabica* clade. The highly distant sequences of *Ixora* were therefore only used for comparison of microsatellite structure. Trees of *Coffea* were then rooted with *Psilanthus*.

2.2.9 Reconstruction of character evolution

Indel and microsatellite matrices (Tables 2.3, 2.4) were optimized with Winclada 1.00.08 (Nixon 2002) using the tree based on substitutions of all partitions combined as shown in Fig. 3. Assuming that genomic relationships are best reflected by combined analyses (De Queiroz et al. 1995), and that the addition of indels to the matrix did not

change the inferred topology, this tree is considered as the currently best approximation for the origin of the *Coffea arabica* cp genome. Unambiguous transformation is shown, whereas results from ACCTRAN or DELTRAN optimization are solely indicated in the text when different. By coding satellites of different length from different plastomes as multiple states of one character, ancestral states of the satellite locus were reconstructed, thus illuminating the history of microstructural changes in this microsatellite locus.

2.3 Results

2.3.1 Chloroplast DNA sequence datasets

Over 7.2 kilo-base pairs (kb) of cpDNA (4.8 kb of non-coding sequences) from twelve different regions were sequenced for each of the 8 individuals of *C. arabica,* seven diploid species, and *Psilanthus leroyi* (Table 2.1). These represent ca. 5.5% relative to the unique sequence (counting the inverted repeat only once) of the tobacco plastome, and ca. 7.5% of its unique non-coding portion. Location and size of genomic regions sequenced are shown in Figure 2.1.

The overall cpDNA sequence divergence between species of *Coffea* was 0.47%, varying from 0.36%, 0.56% and 0.46% for protein coding genes, introns and intergenic spacers, respectively. Sequence characteristics of individual regions and partitions are provided in Table 2.2. The number of variable substitutions among *Coffea* species was 59 (62 including *Psilanthus*), and a total of 23 indels was observed (no specific indel occurs in *Psilanthus)*. The frequency of indels was highest in the intron partition (Table 2.2). Most of the microstructural changes are simple sequence repeats (SSRs) of 1 to 13 bp in length, with the highest frequency of 1-6 bp (Table 2.3). Among the five individuals of *C. arabica* collected from wild populations in the different regions in Ethiopia and the two cultivars no sequence variability was encountered.

2.3.2 Chloroplast microsatellites in *Coffea*

A total of seven microsatellite loci (MS1-7; for definition see discussion) was found in introns and spacers (Figures 2.1, 2.2). In contrast to the observation of Weising and Gardner (1999) in several angiosperms and of Bryan et al. (1999) in Solanaceae, no microsatellite is present in the *atp*F intron in *Coffea*. The longest microsatellite was observed in the *rpl*2-*rps*19 spacer with a stretch of up to 20 repeat units, whereas the

most polymorphic were found in the *rpl*16 intron (MS7) and the *rpl*2-*rps*19 spacer (MS6). All microsatellites are mononucleotide repeats of [(A/T)*n*] as also observed in plastomes of other plants (Powell et al. 1995; Bryan et al. 1999; Ishii and McCouch 2000) with the exception of MS2. This satellite in the *trn*S-*trn*G spacer is composed of di-nucleotide repeats [(AT/TA)*n*]. These repeat units are generally maintained in all individual sequences. Individuals of *C. arabica* and *C. eugenioides* have the same length of microsatellites except in MS 7 (within the *rpl*16 intron).

Table 2.2 Sequence characteristics of the twelve individual genomic regions sequenced.

Type of Regions	Cp Regions Analyzed	Samples	Length of the Region (SD)	GC Content % (SD)	Micro satellite	Indels*	Ts:Tv	Substitutions (pot. pars. inf.)	Variability % (Corrected)**
Intergenic Spacers	*atp*B-*rbc*L	*C. arabica*	759	29.32	1	0	0	0	0
		Coffea spp.	759-768 (3.94)	29.32-30.60 (0.46)		3	0.49	5	0.42
		Total	759-768 (3.91)	29.32-30.60 (0.46)		3	0.48	5	0.40
	*trn*S-*trn*G	*C. arabica*	516	29.3	1	0	0	0	0
		Coffea spp.	484-518 (8.40)	28.7-29.3 (0.24)		3	0.44	7	0.57
		Total	484-518 (8.12)	28.7-29.3 (0.23)		4	0.41	7	0.59
	*rpl*2-*rps*19	*C. arabica*	56	25	1	0	0	0	0
		Coffea spp.	56-62 (2.66)	21.3-25.0 (1.26)		0	0	0	0.30
		Total	56-65 (3.07)	21.3–25.0 (1.37)		0	0	0	0.28
	*rps*19-*rpl*22	*C. arabica*	74	17.6	0	0	0	0	0
		Coffea spp.	74	17.6		0	0	0	0.18
		Total	74	17.6		0	0	0	0.17
	Combined spacers	*C. arabica*		28.98	3	0	0	0	0
		Coffea spp.		28.98-29.46 (0.18)		6	0.52	12	0.46
		Total		28.91-29.46 (0.19)		7	0.49	12	0.45
Gene	*mat*K	*C. arabica*	1518	32.35	0	0	0	0	0
		Coffea spp.	1518	32.15-32.41 (0.10)		0	0.90	11	0.40
		Total	1518	32.15-32.41 (0.10)		0	1.08	11	0.43
	*rpl*2 partial	*C. arabica*	434	45.4	0	0	0	0	0
		Coffea spp.	434	45.4-45.7 (0.16)		0	0	1	0.13
		Total	434	45.4-45.7 (0.16)		0	0	1	0.13
	*rps*19	*C. arabica*	279	34.4	0	0	0	0	0
		Coffea spp.	279	34.4-35.1 (0.24)		0	0.41	1	0.36
		Total	279	34.4-35.1 (0.24)		0	0.38	2	0.35
	*rpl*22 partial	*C. arabica*	202	35.9	0	0	0	0	0
		Coffea spp.	202	34.1-35.9 (0.49)		0	0.57	1	0.36
		Total	202	34.1-35.9 (0.47)		0	0.53	1	0.36
	Combined Gene	*C. arabica*		34.76	0	0	0		0
		Coffea spp.		24.63-34.89 (0.06)		0	1.25	14	0.36
		Total		34.72-34.89 (0.06)		0	1.43	16	0.38
Introns	*atp*F	*C. arabica*	707	32.82	0	0	0	0	0
		Coffea spp.	707	32.66-32.85 (0.10)		0	2.43	4	0.32
		Total	707	32.53-32.85 (0.12)		0	2.47	4	0.33
	*trn*G	*C. arabica*	683	32.06	2	0	0	0	0
		Coffea spp.	678-683 (1.77)	31.77-32.41 (0.14)		3	1.50	8	0.62
		Total	678-683 (1.73)	31.66-32.41 (0.16)		3	1.60	8	0.64
	*trn*K	*C. arabica*	951	32.70	1	0	0	0	0
		Coffea spp.	938-951 (6.16)	32.7-33.74 (0.43)		5	1.33	9	0.61
		Total	938-951 (5.97)	32.7-33.74 (0.41)		5	1.20	9	0.67
	*rpl*16	*C. arabica*	999	29.53	1	0	0	0	0
		Coffea spp.	971-1005 (8.43)	29.38– 30.15 (0.21)		9	0.32	12	0.64
		Total	971-1005 (8.99)	29.38– 30.15 (0.31)		9	0.37	12	0.63
	Combined Intron	*C. arabica*		32.02	4	0	0	0	0
		Coffea spp.		32.02-32.25 (0.09)		17	0.94	33	0.56
		Total		32.02-32.25 (0.15)		17	0.95	34	0.58
All Combined		*C. arabica*			7	0	0	0	0
		Coffea spp.				23	0.83	59	0.47
		Total				24	0.87	62	0.49

**) Indels do not include length variability at microsatellite loci, **) Corrected values are calculated on the basis of mean actual length of sequences*

2.3.3 Sequence characteristics of individual regions

The length of the *atp*B-*rbc*L spacer in *Coffea* varies from 759-768 bp. All 3 indels found in this region were observed to be homoplastic (Figure 2.4). There is one microsatellite in the *atpB-rbcL* spacer at positions 399-407, exhibiting 3 different lengths (Figure 2.2). The *trn*S-*trn*G spacer showed the highest variability among spacers within *Coffea*, comparable to the *rpl*16 intron. The length of this spacer ranges from 448bp in *C. stenophylla* to 518bp in *C. kapakata, C. congensis* and *C. salvatrix*. A long deletion (32bp) is found in *C. stenophylla,* and a further autapomorphic deletion in *C. eugenioides* (Figures 2.3, 2.4). The number of potentially informative substitutions is highest among all spacers. The single microsatellite locus is at positions 90-99. The spacers between *rpl*22*, rps*19 and *rpl*2 are very short (74bp, 56-62bp) and have a very low GC content, ranging from 17.6% - 18.0% in *rpl22-rps19* to 21.3%-25.0% in *rpl*2-*rps1*9. Both spacers are more conserved than other spacers studied and exhibit no variation except a microsatellite in positions 34-53 in *rpl*2-*rps*19 (MS6). This locus is the second most variable microsatellite and, in *Psilanthus,* contains the longest poly A/T-stretch encountered (Figure 2.2).

The *mat*K coding region is 1518 bp in *Coffea* and does not show any length variation, but contains 23 substitutions (0.4% variability) of which about half are potentially informative characters (Table 2.2). The *rps*19 gene was also completely sequenced with the strategy employed here and is 279bp in all taxa with a sequence variability of 0.36%. This also applies to the 3' exon of *rpl*2 that is 434bp in *Coffea* and *Psilanthus,* and to the 5' half of the *rpl*22 gene (202 of 478 bp). The first exhibits the lowest sequence variability (0.13%) of all sequenced regions, whereas the latter shows 0.36% variability. The GC contents of most genes is between 32.2 and 35.9% except in the *rpl*2-3' exon with a CG content of > 45.4%.

Taxon name	Number	MS 1	State	MS 2	State	MS 3	State	MS 4	State
		atpB -rbcL		trnS - trnG		trnG		trnG	
		396 - 410		87 - 102		519 - 533		572 - 588	
C. arabica	C002	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C007	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C016	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C034	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C035	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C036	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica	C037	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. eugenioides	C015	GCCTTTTTTT--CAT	1	ATTATATAT----GTA	1	CGATTTTTTTTTGTC	3	GACTTTTTTTTT--GTA	2
C. arabica x liberica	C003	GCCTTTTTTTT-CAT	2	ATTATATATAT--GTA	2	CGATTTTTTTT-GTC	2	GACTTTTTTTTTTTGTA	4
C. canephora	C013	GCCTTTTTTTT-CAT	2	ATTATATATAT--GTA	2	CGATTTTTTTT-GTC	2	GACTTTTTTTTTTTGTA	4
C. congensis	C009	TCCTTTTTTT--CAT	1	ATTATATATATATGTA	3	CGATTTTTTTT-GTC	2	GACTTTTTTTT--GTA	2
C. kapakata	C014	GCCTTTATTTTTCAT	3	ATTATATATATATGTA	3	CGATTTTTTTT-GTC	2	GACTTTTTTTTTT-GTA	3
C. liberica	C008	GCCTTTTTTT--CAT	1	ATTATATATAT--GTA	2	CGATTTTTTTT-GTC	2	GACTTTTTTTTTT-GTA	3
C. salvatrix	C011	GCCTTTTTTT--CAT	1	ATTATATATATATGTA	3	CTATTTTTTTT-GTC	2	GACTTTTTTTTT--GTA	2
C. stenophylla	C015	GCCTTTTTTT--CAT	1	ATTATATATAT--GTA	2	CGATTTTTTTT-GTC	2	GACTTTTTTT----GTA	1
P. leroyi	C033	GCCTTTTTTT--CAT	1	ATTATATATATATGTA	3	CGATTTTTTT--GTC	1	GACTTTTTTTTT--GTA	2
I. coccinea	C017	GCCTTTTTTT--CAT	-	ATTATATAGATATGTA	-	CGATTTGTTTT-GTC	-	GGCTTTTTTTTT--GTA	-

(continued)

Taxon name	Number	MS 5	State	MS 6	State	MS 7	State
		trnK		rpl2 - rps19		rpl16	
		863 - 887		31 - 56		718 - 737	
C. arabica	C007	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C002	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C037	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C036	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C035	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C034	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. arabica	C016	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAAA-CAA	6
C. eugenioides	C015	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAA--------[-]TAG	1	GATAAAAAAAAAAAA--CAA	5
C. liberica	C008	TCTAAAAAAAA-----------GAA	1	TACAAAAAAAAAAAAAAAA---[-]TAG	3	GATAAAAAAAAAAA------	4
C. congensis	C009	TCTAAAAAAAA-----------GAA	1	TACAAAAAAAAAAAAAA-----[-]TAG	2	GATAAAAAAAAAAAA----	6
C. arabica x liberica	C003	TCTAAAAAAAA-----------GAA	1	TACAAAAAAAAAAAAAAAAA--[-]TAG	4	GATAAAAAAAAAAAAACAA	7
C. canephora	C013	TCTAAAAAAAA-----------GAA	1	TACAAAAAAAAAAAAAAAAA--[-]TAG	4	GATAAAAAAAAAAAAACAA	7
C. kapakata	C014	TCTAAAAAAAA-----------GAA	1	TACAAAAAAAAAAAAAA-----[-]TAG	2	GATAAAAAAAA-----CAA	2
C. salvatrix	C011	TCTAAAAAAAAA----------GAA	2	TACAAAAAAAAAAAAAAAAA--[-]TAG	4	GATAAAAAAA------CAA	1
C. stenophylla	C015	TCTAAAAAAAAAA---------GAA	3	TACAAAAAAAAAAAAAAAA---[-]TAG	3	GATAAAAAAAA-----CAA	2
P. leroyi	C033	TCTAAAAAAA------------GAA	1	TACAAAAAAAAAAAAAAAAAAA[A]TAG	5	GATAAAAAAAAA----CAA	3
I. coccinea	C017	TCTAAAAATAAAAAAAAAAAAAGAA	-	TACAAAAAAAAAAAAAAA----[-]TAG	-	GATAAAAAAA------CAA	-

Figure 2.2 Sequences of seven chloroplast microsatellites in wild individuals of *C. arabica* representing five geographically distant populations and two commercial cultivars. In addition, the same microsatellite loci are shown in eight diploid species of the genus *Coffea* plus one representative each of *Psilanthus* and *Ixora*. Positions of satellites are indicated with respect to alignments of individual genomic regions. Character state numbers are given right to the sequence for each sampled individual.

Table 2.3 Indels found in individual genomic regions. List of indels found in the tree regions cpDNA. The position in the alignment and the extension of each indel are noted (bp). SSR = simple sequence repeat**.**

No.	Chloroplast regions	Motif Ranging from / to (bp)	Sequences type
	***atp*B-*rbc*L spacer**		
1	"CTATGGAATTCG"	301-312	SSR
2	"G"	531-535	SSR
3	"AATAAT"	618-629	SSR
	***trn*S-*trn*G spacer**		
4	"AAAGA"	166-170	Indel
5	"C"	184-188	SSR
6	"A"	420-429	SSR
7	"ACTCTAATTGTGATGGTTTTTTGATGATCCTA"	439-470	Indel
	***trn*G intron**		
8	"A"	94-100	SSR
9	"TA"	449-450	Indel
10	"TAAAAA"	449-454*	SSR
	***trn*K intron**		
11	"T"	389-394	SSR
12	"ACTAAAAATG"	550-569	SSR
13	"C"	613-617	SSR
14	"T"	769-774	SSR
15	"TTAAA"	848-857	SSR
	***rpl*16 intron**		
16	"T"	148-151*	SSR
17	"AATGACTCTTCTT "	343-368	SSR
18	"AATGACTCTTCTTATAGAACATTCGTTATAGTAGAACATT"	356-395	Indel
19	"A"	408-417	SSR
20	"A"	714-715*	SSR
21	"CAA"	735-737	Indel
22	"CAATTTATAAAAAAAT"	735-755	Indel

** Overlapping indels or SSR*

From the introns the largest and also most variable (0.64 %) is the *rpl*16 intron and contains one satellite (MS 7). In *Coffea* it ranges from 971 (*C. stenophylla*) to 1005 bp (*C. salvatrix*). Resulting high standard deviations of (8.4) otherwise only occur in the *trn*S-*trn*G spacer (Table 2.2). The *rpl*16 intron contains the longest indel observed in *Coffea*, an at least 25 bp deletion and also deviates from other introns by its much lower Ts:Tv ratio (0.3-0.4) only comparable to spacers and the *rps*19 gene (Table 2.2). The *trn*G intron is the smallest (678-683 bp) of all introns sequenced but contains two microsatellite loci in close proximity (MS3 spanning positions 522-530 and MS4 positions 575-585).

Table 2.4 Indel character state for all tax, for description of indel characters, see Table 2.3

Taxa	Number	*atp*B-*rbc*L			*trn*S-G spacer				*trn*G intron			*trn*K intron					*rpl*16 intron						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
C. arabica wild Bonga	C016	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica wild Bale	C034	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica wild B Kontir	C035	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica wild Yayu	C036	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica wild Boginda	C037	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica Blue Mountain	C002	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. arabica Catura Yellow	C007	0	0	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. eugenioides	C010	0	0	1	1	0	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1
C. liberica	C008	1	1	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	1	0	0	0	1
C. congensis	C009	1	1	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	1	0	1	?	0
C. arabica x liberica	C003	1	1	0	1	1	0	1	1	?	0	1	0	1	1	0	1	0	1	0	0	1	1
C. canephora	C013	1	1	0	1	1	0	1	1	?	0	1	0	1	1	0	1	0	1	0	0	1	1
C. kapakata	C014	1	1	0	1	1	0	1	0	1	1	1	0	1	1	0	1	0	1	0	0	1	1
C. salvatrix	C011	1	1	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	0	0	1	1
C. stenophylla	C015	1	1	0	1	1	0	0	0	0	1	0	1	1	1	0	1	?	?	1	0	1	1
Psilanthus leroyi	C033	0	0	1	0	1	0	1	0	1	1	1	1	1	1	0	1	0	0	0	0	1	1

The *atp*F intron is length conserved in *Coffea* (707 bp). It exhibits only half of the variability (0.33%) found in other introns but has a high (2.4) Ts:Tv ratio (Table 2.2). The two parts of the *trn*K intron flanking the *mat*K gene are of different size and variability with the 5' part being large (~720bp) and rather conserved and the 3' part small (~220bp) and more variable (Table 2.2). The latter comprises one microsatellite (MS 5). In sum, the *trn*K intron is the second largest and second most variable of the introns studied (Table 2.2).

In general, introns sequenced for *Coffea* showed higher sequence divergence (0.56%) than spacers (0.46) and coding regions. Among introns a divergence gradient was found from *rpl*16 (0.64%) to *trn*G (0.62%) > to *trn*K (0.61%) to *atp*F (0.32%; Table 2). Among spacers the gradient is from *trn*S-*trn*G (0.57%), to *atp*B-*rbc*L (0.42%), to *rpl*2-*rps*19 (0.30%), to *rps*19-*rpl*22 (0.18%).

2.3.4 Phylogenetic Inference

Parsimony analysis using the whole dataset of 7.2 kb, combining sequences from the twelve different coding and non-coding regions, recovered a single shortest tree (Figure 2.3). Based on substitutions the tree length was 152 steps (CI=0.84, RI=0.90, RC=0.76). Adding indels (Table 2.3, 2.4) and/or microsatellite characters (Figure 3.4) yielded the same topology but increased the length to 178, 180 and 206 steps, respectively. Homoplasy indices were not significantly different, nodes were equally or slightly better supported (data not shown). Bayesian inference using the substitution based combined dataset sampled 159,200 trees in total. The consensus depicts the same topology than MP but posterior probabilities (PP) are generally higher than JK percentages (≥ 0.90; Figure 2.3).

Two major clades are found in *Coffea,* one with East African species and *C. stenophylla* (91% JK, 1.00 PP) and the other with west and central African species (97% JK, 1.00 PP). The sister group of *C. arabica* and *C. eugenioides* gains maximum support in both of MP and BI and is placed with the East African clade. It may be noted that the respective node has the highest decay index (=14) in the whole topology, indicating a high number of synapomorphies of *C. arabica* and *C. eugenioides* (Figure 2.3) based on substitutions. Mapping revealed five synapomorphic indels for *C. arabica* and *C. eugenioides,* plus further two deletions and one insertion shared by the two species, which are globally homoplasious. A further autapomorphic indel is observed for *C. eugenioides* (Figure 2.5). A clade consistently recovered from all regions is *C. canephora* and the hybrid cultivar *C. arabica* X *C. liberica* with support > 85%, except for the *trn*K and *atp*F intron partitions (62% JK, results not shown). *Coffea congensis* was also consistently found as successive sister except for the *atp*B-*rbc*L spacer (trees not shown).

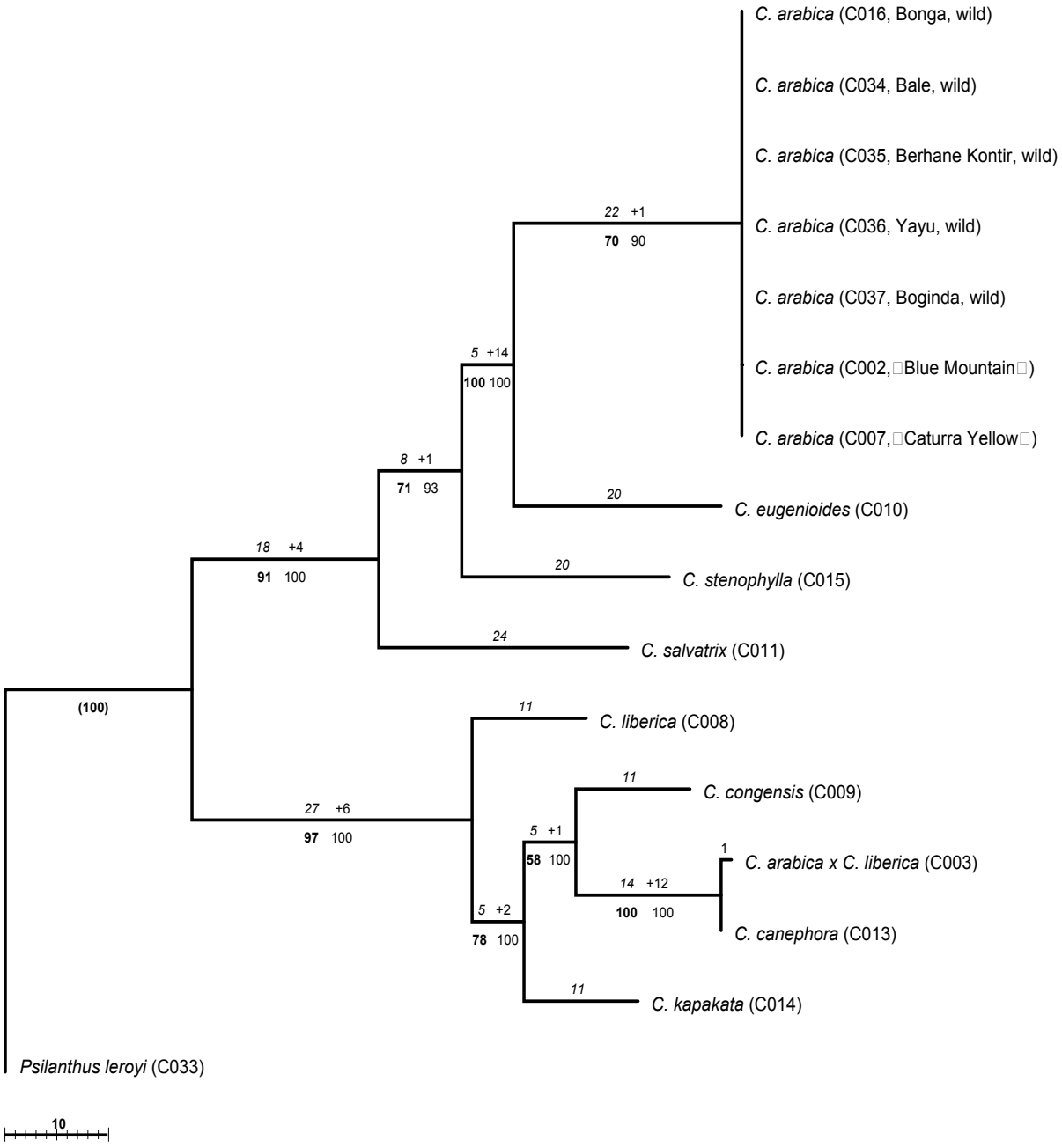


Figure 2.3 The single shortest tree inferred from parsimony analysis (length = 152 steps, CI = 0.84, RI = 0.90, RC = 0.76) of the combined dataset using substitutions only, shown as phylogram. Values above branches refer to branch lengths (left, in italics) and decay values Support (right). Values below branches depict Jackknife support (left, bold) and posterior probabilities (right) from the Bayesian analysis that gave the same topology.

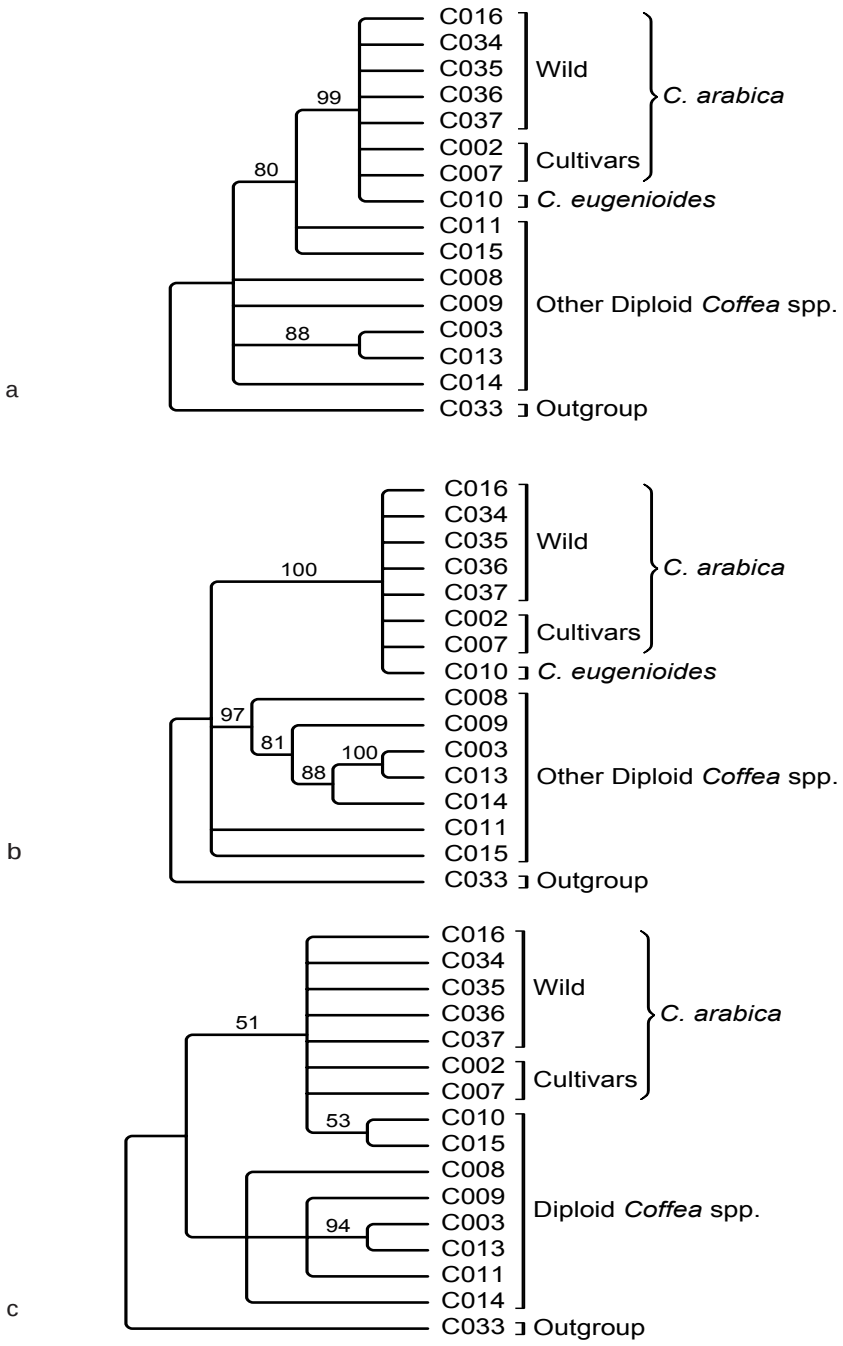


Figure 2.4 Tree topologies based on individual chloroplast partitions (substitution). A (all coding sequences combined) and B (all introns combined) infer *C. arabica* and *C. eugenioides* in a clade but reveal no resolution among individuals of the respective two species. Partition C (all spacers combined) depicts *C. eugenioides* and *C. stenophylla* as sisters but without statistical support.

Partitioned analysis (Figure 2.4) and analysis based on individual genomic regions were generally much less resolved and supported, thus indicating a great benefit of combining multiple spacers, introns and genes. Trees based on all spacers combined (Figure 2.4C) showed low support except for the sister group of *C. canephora* and *C. arabica* X *C. liberica*. Moreover, the spacer topology differs by *C. eugenioindes* forming a separate clade with *C. stenophylla* albeit with low JK (53%). Individuals of *Coffea arabica* and *C. eugenioides* were consistently found in a clade except in the tree inferred from the *atp*B-*rbc*L spacer (there, *C. eugenioides* appeared in a weekly supported clade with *C. liberica*, *C. canephora*, *C. stenophylla*, and *C. arabica* x *C. liberica*). Signal from the *atp*B-*rbc*L is therefore responsible for the different tree of the spacer partition. In the analysis of the dataset containing all introns (Figure 2.4B *atp*F, *trn*G, *trn*K, *rpl*16) and all genes (Figure 2.4A) a clade of *C. eugenioides* and *C. arabica* is supported. However, signal for this clade differed in individual genomic regions as evidenced by JK support ranging from 61% for the *atp*F intron to 99% for the *trn*K intron. Only the overall combined tree based on substitutions resolves all individuals of *C. arabica* in a clade different from *C. eugenioides* (Figure 2.3).

2.4 Discussion

2.4.1 Structure and evolution of chloroplast microsatellites

Microsatellites are repeats of several to many mono-, di-, tri- and tetra-nucleotide motifs with a high level of length variation. In cpDNA usually only homonucleotide strands of A's or T's occur (Bryan et al. 1999; Provan et al. 1999). Like other indels, length variable satellites are the result of microstructural mutations. If a satellite locus entails only two length variants among the sequences in an alignment, an interpretation as entire indel (Borsch et al. 2003) is most parsimonious and allows unambiguous coding in any indel coding scheme. Such situations were considered as "emerging satellites" by Levinson and Gutman (1987), since they were assumed to accelerate in size and variability through increased probability of replication slippage. Because of their straightforward interpretation emerging satellites were included in phylogenetic analysis in this study.

Homologous regions with more than two different numbers of repeated motifs are microsatellites. In their case, more than one overlapping deletion or insertion event

has to be assumed to explain length variability of the aligned sequences (i.e. length variable microsatellites can be described as overlapping indels; Borsch et al. 2003). In contrast to other complex situations, homology assessment within microsatellites is usually ambiguous because very short repeats that appear adjacent to each other hinder motif recognition (Löhne and Borsch 2005). This is particularly true for the homonucleotide strands encountered in cp DNA microsatellites. As a consequence, presence-absence coding of assumed deletion or insertion events as implemented in simple and complex indel coding methods (e.g. Simmons and Ochoterena 2001; Müller 2005) will also lead to unreliable results and cannot be used. In this study, we therefore used an approach of defining the whole length variable satellite region as one single character, and coded each of the different lengths as individual states (multi-state approach). Thereby, no assumptions on actual overlapping microstructural mutations had to be made.

Through optimizing these multiple states corresponding to different repeat numbers ordered (additive) onto the tree (Figures 2.6, 2.7) not only the distribution of states but also the evolution of repeat numbers is reconstructed. Many states have evolved several times independently in most satellites (empty boxes in Figures 2.6, 2.7), whereas nearly all indels outside of satellite loci are apomorphic (Figure 2.5). This confirms assumptions of higher rates of microstructural mutations (Ishii and McCouch 2000; Zhu et al. 2000) and high levels of homoplasy due to frequent slippage in satellite loci (Provan et al. 2004; Estoup and Cornute 1999). For example, state two of the highly variable MS 7 has arisen independently in *C. kapakata* and *C. stenophylla*. State six of MS 7 is reconstructed to have arisen by the addition of one repeat unit in *C. arabica*, and by the addition of two repeat units in the common ancestor of *C. congensis*, *C. canephora*, and *C. arabica* x *C. liberica*, after a different mutational history reconstructed for deeper nodes. As Figure 2.7 shows, microsatellites of different lengths are irregularly distributed among terminals, and gains and losses of repeat units occur within lineages of *Coffea*. Information from microsatellites is therefore likely to be misleading in phylogenetic inference, in particular if used across more distantly related species.

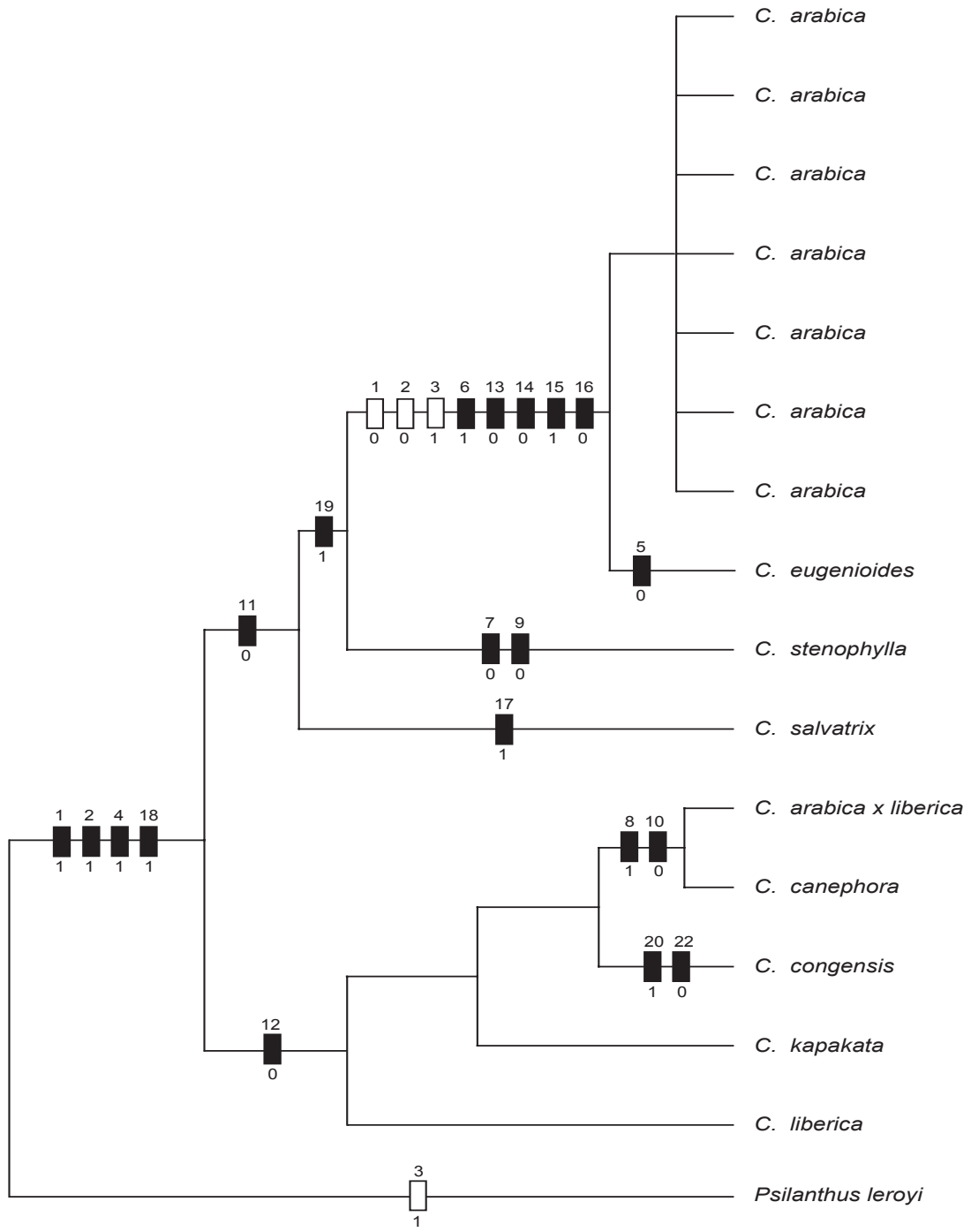


Figure 2.5 Single shortest tree found in the analysis of all partitions combined (substitutions, see figure 2.3) showing the evolution of microstructural mutations outside microsatellites (unambiguous transformation). The numbers below indicate gain and loss ('0' indicate losses whereas '1' boxes indicate gains). Numbers above boxes refer to individual indels as described in Table 2.4. The empty boxes indicate homoplastic indels.

However, locally the gain of one nucleotide in MS7 can be inferred as evolutionary step that distinguishes *C. arabica* from the common ancestor with *C. eugenioides* (Figure 2.7). A simple stepwise model is not inferred for the microsatellite loci investigated in this study, in contrast to theoretical considerations (Kimura and Ohta et al. 1978). Mutational steps involving two and three repeat units at once are inferred (Figure 2.7), although it cannot be completely excluded that intermediate states were present in species now extinct or not sampled. It has been observed that mutation rates in microsatellite loci tend to increase with the number of repeats (Brohede et al. 2002; Petit et al. 2005). Higher levels of variability (i.e. a larger number of different states and more mutational steps in the actual evolutionary history; Figures 2.2, 2.6) are also found in longer microsatellites in *Coffea*. The increase of mutational rates with increased size of microsatellites seems to be a general trend that appears to be caused by a higher probability for slipped strand mispairing (Ortí et al. 1997; Vigouroux et al. 2002)

Microsatellites present in *Coffea* in the *atp*B-*rbc*L, *trn*S-*trn*G and *rpl*2-*rps*19 spacers and the *trn*G, *trn*K, and *rpl*16 introns also occur in other angiosperms (Bryan et al. 1999; Weising and Gardner 1999; Ishii et al. 2001; Xu et al. 2000; Provan et al. 2004). The longest and most variable microsatellites in *Coffea* are in the *rpl*16 intron and the *rpl*2-*rps*19 spacer. The latter locus has also been found highly variable in other angiosperms (Weising and Gardner 1999). A satellite in the *atp*F intron is not present in *Coffea* although it was found in *Nicotiana*, *Lycopersicon* and *Actinidia* by Weising and Gardner (1999). Additional knowledge of the presence of microsatellites gained through sequencing *atp*B-*rbc*L spacer (MS1), *trn*S-G spacer (MS2), *trn*G intron (MS3), *trn*K intron (MS5) and *rpl*16 intron (MS7) (Weising and Gardner 1999). The listed chloroplast microsatellites reveal intraspecific variability in a number of plants. In 30 accessions of *Solanum tuberosum* ssp. *tuberosum* there were 19 distinct haplotypes (Bryan et al. 1999). Echt et al. (1998) observed more than 20 haplotypes among 159 individuals of *Pinus resinosa* surveying 9 cpSSR loci (mostly identical to the loci screened here) examined. A larger number of haplotypes was also detected using cpDNA SSRs among wild and cultivated soybeans (Xu et al. 2000). Dane and Lang (2004) observed five different haplotypes in a 4 kb section of the *Citrullus colcynthis* (Cucurbitaceae) chloroplast genome that differentiate accessions from Ethiopia,

Afghanistan and Pakistan. However, in the case of *C. arabica* our extensive survey of fast evolving cpDNA regions yielded no variation.

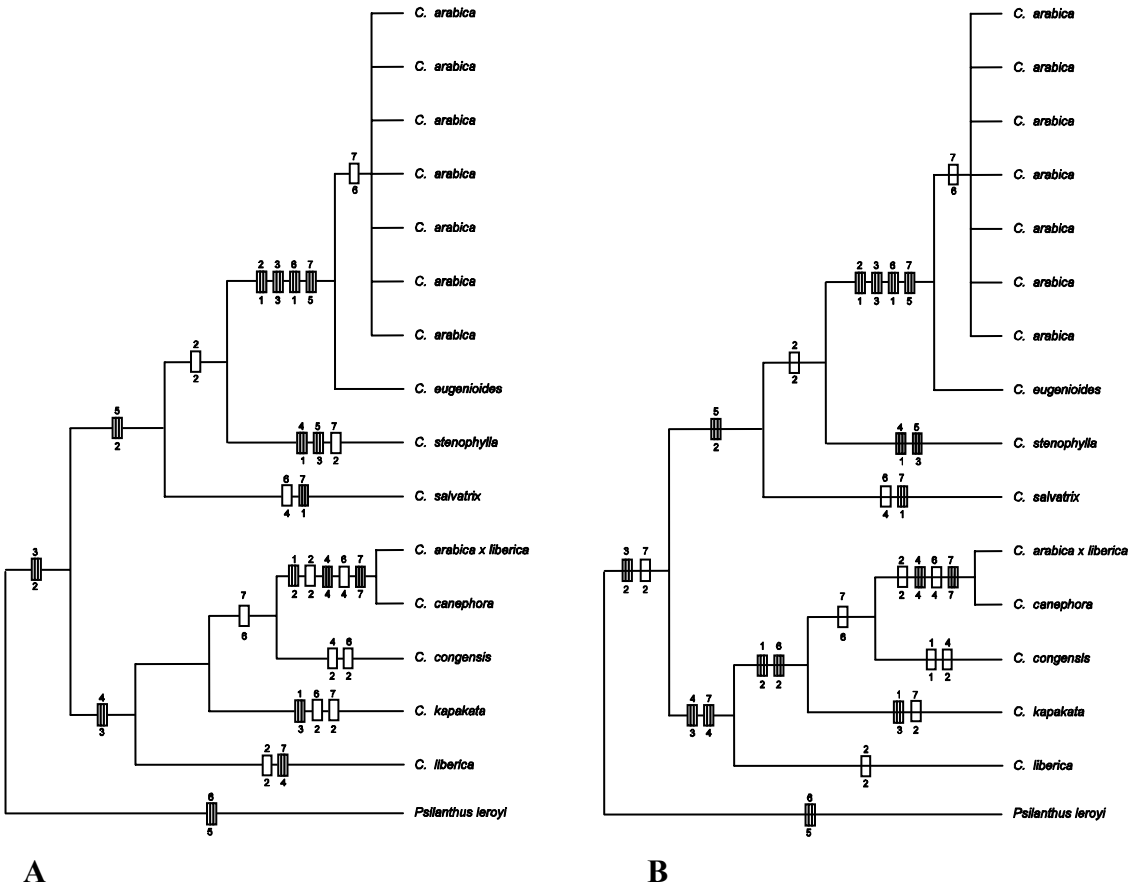


Figure 2.6 The seven microsatellite loci found in *Coffea* mapped onto the single shortest tree of the combined dataset based on substitutions (MP; see Figure 2.2). The tree left (A) refers to slow (DELTRAN) and right (B) to fast optimization (ACCTRAN). Numbers above boxes refer to loci (MS 1 to 7) and below boxes to states. Hatched boxes indicate syn- or autapomorphic states whereas empty boxes indicate homoplastic states.

2.4.2 Molecular evolution of different partitions

Striking differences in variability and molecular evolutionary patterns were found among the spacers, introns and the *mat*K gene analysed. The comparison of different sequence datasets of *Coffea* shows that introns (except the *atp*F intron) are distinctly more variable (0.60-0.65%) than spacers (0.46% on average; Table 2.2). This is surprising given that all introns are group II introns with rather strong structural

constraints in the stem elements of their six domains (Kelchner 2000; Löhne and Borsch 2005). A possible explanation for high average variability of the larger sized group II introns could be that their stem loops which are the relatively least constrained elements account for higher sequence proportions. This would be in line with ideas expressed by Kelchner (2002) for the large *rpl*16 intron in *Myoporum* (Lamiales) and by Löhne and Borsch (2005) for the short *pet*D intron in basal angiosperms. However, a dataset of four introns and four spacers in Nymphaeales revealed spacers as the most variable partition (Löhne et al. submitted). More detailed comparisons of different group II introns in identical taxon sets need to be made in order to get more insights. The *rpl*16 intron provides the largest number (12) of potentially informative characters in the data set among all introns, which in absolute terms may also be explained by its greater length. The *rpl*16 intron is followed by the *trn*K intron in information content (0.61%; 9 potentially informative characters). The *atp*F intron provides only few (4) potentially informative characters in *Coffea* and, contrary to other angiosperms (Weising and Gardner 1999) no microsatellite was detected in this region.

This study analyses the first more comprehensive sequence dataset of the *atp*F intron, and suggests that it is one of the more slowly evolving introns in the chloroplast large single copy region. Spacers vary considerably in size and information content, although none of the spacers exceeds the three larger introns in variability (Table 2.2). The *trn*S-*trn*G spacer, although being of only medium size, provides the largest number of potentially informative characters (7) and is clearly the most variable spacer (0.57% in *Coffea*). The *trn*S-*tr*nG spacer was one of the regions with most phylogenetically informative characters among 21 chloroplast non-coding regions screened for infrageneric 3-taxon datasets across angiosperms (Shaw et al. 2005). The short *rpl*2-*rps*19 and *rps*19-*rpl*22 spacers are rather conserved and of limited utility apart from the microsatellite in the first spacer. The *atp*B-*rbc*L spacer shows the same variability in *Coffea* than the *mat*K coding region and also contains a microsatellite. Shaw et al. (2005) excluded the *atp*B-*rbc*L-spacer from their analysis because they considered it to be of little infrageneric utility. In terms of total information content the *mat*K gene (0.40 %, 11 potentially informative characters) is the second most useful region in this *Coffea* dataset although it is about 1.5 times in size as compared to *rpl*16.

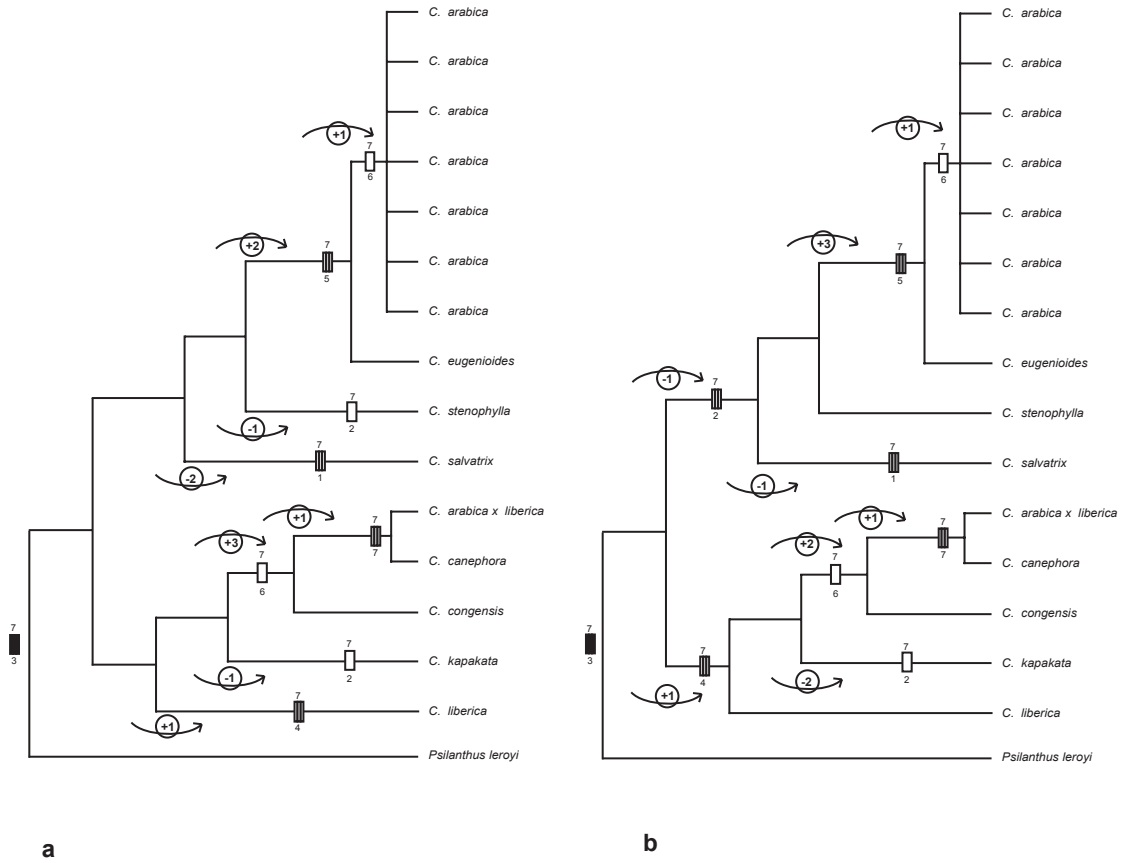


Figure 2.7 Evolution of the most variable microsatellite (MS 7) mapped onto the single shortest tree of the combined dataset based on substitutions (MP; see Figure 2.2). The tree left (a) refers to slow (DELTRAN) and right (b) to fast optimization (ACCTRAN). Gains or losses of the respective number of repeat units are indicated with respect to all inferred ancestors. The numbers above boxes refer to actual numbers of repeat units of the respected satellite sequence, the numbers below boxes to states.

The evolution of microstructural mutations in *Coffea* is shown in Figure 2.5 based on unambiguous optimization. ACCTRAN and DELTRAN optimizations gave the same result (data not shown). In line with the general trend found in analyses of evolution of microstructural changes (e.g. Graham et al. 2000; Müller and Borsch 2005) indels have high phylogenetic utility. For this *Coffea* dataset it is noteworthy that the branch leading to the common ancestor of *C. arabica* and *C. eugenioides* is the longest and best supported branch in terms of indels. Five indels are synapomorphic, and three are locally apomorphic (Figure 2.5). A 1-nt deletion in the *trn*S-*trn*G spacer is unique to *C. eugenioides* and supports the notion that the plastome in *C. eugenioides* is further diverged from the ancestor that provided one genome of *C. arabica*.

2.4.3 Relationships of *Coffea* and origin of *C. arabica*

A multilocus approach was crucial to obtain a sufficient number of phylogenetically informative characters in *Coffea*. Neither combined sequences of four spacers or four introns yielded sufficiently resolved trees, and some individual genomic regions (*atp*B-*rbc*L spacer) even showed different topologies for relationships of *C. arabica*. These differences were inconsistent (very weakly supported nodes) or taxa were not resolved at all by individual positions. Similar difficulties to resolve relationships were made in other groups of recently diverged species (Small et al. 1998; Kusumi et al. 2000). This study shows that combining a large number of sequence data from rapidly evolving and non-coding genomic regions can offer a solution. For *Coffea*, over 7 kb of sequence from twelve-chloroplast regions were analysed for each sample, so that our dataset represents one of the most extensive infrageneric multilocus datasets existing so far.

The sister group relationship of *C. arabica* and *C. eugenioides* chloroplast genomes gains maximum support with both parsimony and Bayesian approaches. Considering the number of synapomorphies and the amount of different partitions sequenced in this study we assume that this picture is representative for the plastome. Earlier hypotheses of *C. eugenioides* being a maternal parent of *C. arabica* based on *trn*L-*trn*F spacer sequences (Cros et al. 1998) and evidence based on cytological data (Lashermes et al 1999; Raina et al. 1998; Charrier and Berthaud 1985) are therefore substantiated. The phylogenetic analysis by Cros et al. (1998) found only a single substitution and one restriction site supporting a clade of *C. arabica*, *C. eugenioides* and *C. sp. Moloundou*. The much more representative dataset of this study does not only substantiate the existence of a clade comprising *C. arabica* and *C. eugenioides* but also infers *C. arabica* as monophyletic (Figures 2.2, 2.3). It therefore shows that chloroplast genomes of both species are significantly different from each other (two substitutions in *atp*B-*rbc*L, one indel in *trn*S-*trn*G and, one conserved state of microsatellite MS 7 in *rpl*16 intron). The sequence divergence between *C. arabica* and *C. eugenioides* using the combined substitution dataset is 0.015% (but 0.103% for *atp*B-*rbc*L and 0.026% for *rpl*16). There are two possible scenarios to explain this divergence. Either, one or both genomes have diverged significantly after the hybridization event with an ancestral species that led to the origin of *C. arabica* or, the actual parental genome of *C. arabica* is from a taxon close to (but not exactly) *C. eugenioides* which has not been sampled.

To evaluate the latter case it will be interesting to study the diversity of wild populations of *C. eugenioides*. The Ethiopian wild populations of *C. arabica* seem to be well represented, but this is not yet the case for *C. eugenioides*. The recent description of new taxa of *Coffea* from the Eastern Arc Mountains, Tanzania by Davis and Mvungi (2004) implies that there still might be undiscovered taxa, even close to *C. eugenioides*.

In addition, *C. congensis* is sister to *C. canephora* and similar phenomenons were also observed by Lashermes et al (1997) with ITS2. *C. arabica* and *C. eugenioides* also form a separate clade in partition analysis and combined analysis having good support. However, the analysis of all spacer C. *arabica*, *C. eugenioides* and *C. stenophylla* form a separate major clade having lower support (51%) and the latter two form a sub group with 53% support.

2.4.4 *Coffea arabica* exhibits a single cp haplotype

The data presented here reveal no variability within *C. arabica* at any of the otherwise variable microsatellite loci analysed, nor could any SNP be detected. The existing genotypic diversity of the species is represented well in this study because wild *C. arabica* populations from a wide range of geographical localities in Ethiopia are included that have been shown to represent extremes in genetic diversity based on ISSR fingerprinting (Tesfaye et al. submitted (b)). Moreover, the included commercial cultivars do not show any difference to individuals from wild populations. In comparison to available assessments of infraspecific chloroplast variability within other species such as *Argania* (Petit et al. 1998), *Quercus* (Deguilloux et al. 2004) or *Vitis* (Arroyo-Garcia et al. 2002) the proportion of the chloroplast genome screened in this study is several folds. Moreover, deviating haplotypes were found within other species based on the same regions as studied here, for example the *trn*K and *rpl*16 introns in *Oryza* (Ishii and McCouch 2000) or the *trn*S-*trn*G spacer within species of *Glycine* (Xu et al. 2000). The majority of microsatellites studied here have been applied successfully through population genetic analyses through the universal primers designed by Weising and Gardner (1999). It therefore becomes clear that chloroplast variation within *C. arabica* is extremely low, and it may even be hypothesized that the species exhibits only a single chloroplast haplotype.

The reasons for the absence of chloroplast variation are therefore to be sought in the origin and population history of *C. arabica*. As a tetraploid it may be of very recent origin, so that time was not sufficient for accumulating mutations. Similar results were obtained for the young allotetraploid *Arabidopsis suecica*, for which Säll et al. (2003) analyszed a large number of chloroplast regions. However, although the present pattern in *C. arabica* is likely to be associated with its allopolyploid origin, it is also possible that wild populations went through a severe bottleneck in their history, causing a much narrower genetic base in as compared to other species. It is well known that genetic diversity may vary considerably among species of a genus (e.g. Hamrick and Godt 1996), and there are also other examples for low genetic diversity as a consequence of founder effects (Schwaegerle and Schaal 1979; Parisod et al. 2005) or in narrow endemics (e.g. Gengler and Crawford 2000; Sgorbati et al. 2004).

Allopolyploids can either be of single or of multiple origins due to several independent hybridization events. Multiple origins of allopolyploids from genetically distinct individuals of diploid progenitor species will result in respective patterns of genetic variability among the allopolyploid populations as observed for example in species of *Spiranthes* (Orchidaceae; Arft and Ranker 1998), *Tragopogon* (Asteraceae; Cook et al. 1998; Soltis et al.2004) and *Sorbus* (Rosaceae; Robertson et al. 2004). Single origins of allopolyploids were suggested for *Dabar ladina* and *Spartina anglica,* both of which completely lack infraspecific variation in the chloroplast regions screened (Widmer and Baltsiberger 1999; Baumel et al. 2001). With respect to the origin of *C. arabica* the present findings are in line with the assumption of a single allopolyploidization event.

2.5 Conclusions and future directions

Generally chloroplast genome is maternally inherited in *Coffea* (Berthu et al. 1983; Lashermes et al 1996a) and the cpDNA data precisely suggested that *C. eugenioides* or its ancestor as maternal parents of *C. arabica* and also point towards single evolutionary event of *C. arabica* polyploidisation in recent times. In addition, very little cpDNA sequence variation has accumulated in genus *Coffea.* The low sequence divergence in *Coffea,* morphological overlap and intergeneric successful hybridisations show rather the genus is recently evolved and may be not fully differentiated (Charrier and Berthaud

1985). The data hint to recent divergence of both *C. arabica* and *C. eugenioides* from *Coffea.* Introns exhibited more variability in terms of substitution as compared to spacers since the first most variable region in this analysis were introns however in other plant group like subgenus Soja is in the intergenic spacer (Xu et al. 2000; Sakai et al. 2003). Moreover, the indel showed more phylogentic utility as compared to the homoplastic microsatelite of *Coffea*.

As a future direction, the low cpDNA polymorphism in general and also absence of haplotypes in wild *C. arabica* population could led us for further analysis of mitochondria genome of *Coffea* to look for most variable regions to map haplotypes. There might also be few geographically scattered mutations occur in different population so further analysis is needed to explore this question with the inclusion of more regions of genome and samples. The paternal donor of *C.arabica* genome should be identified to fully understand evolution of the species. CpDNA analyis of co-occuring forest tree species, probably having joint evolutionary history in Forest Coffee Ecosystem (FCE), would be worth while to understand the evolution and diversification of FCE in general and *C. arabica* in particular. This is particularly important to design conservation strategy for wild *C.arabica* as well as for co-occurring important forest tree species in Afromontane rainforests of Ethiopia.

3 GENETIC DIVERSITY OF *COFFEA ARABICA* THROUGHOUT ITS NATIVE RANGE IN ETHIOPIA BASED ON ISSR FINGERPRINT DATA

3.1 Introduction

The genus *Coffea* L. (Rubiaceae) comprises about 100 species which are native to forests and scrublands of tropical Africa, Madagascar and the Mascarene Islands in the Indian Ocean (Purseglove 1968; Bridson and Verdcourt 1988). The economically most important species of the genus are *C. arabica* L. with more than 80 per cent of the world's coffee production and *C. canephora* with nearly 20 percent, while *C. liberica* has minor importance as a crop. Out of the two major crop species better quality coffee (low content of caffeine and fine aroma) is associated with *Coffea arabica* (Raina *et. al* 1998).

Coffea arabica is native to the Ethiopian highlands. Wild populations have also been reported from the Boma plateau in SE Sudan and on Mount Marsabit in northern Kenya although very little is known about their actual distribution and ecology (Friis 1979). In Ethiopia wild *C. arabica* occurs in mountain forests on both the western and eastern sides of the Great Rift Valley in the southern part of the country. The populations of *C. arabica* grow naturally in the undergrowth of montane rain forests at altitudes between 1,400 and 1,900 m a.s.l. (Meyer 1968; Gebere-Egziabher 1990; Tadess and Nigatu 1996; Gole et al. 2001; Gole et al. 2002). However, Senbeta (2006) described the altitudinal range between 1,300-1,600 m a.s.l. as limit for the occurrence wild coffee.

Coffee is a major source of income for Ethiopia and 67% of the country's foreign exchange income comes from this single commodity (Oxfam 2002; Tafesse 1996). There are four different production systems, two of which are actually based on collecting coffee from autochthonous populations (Gole et al. submitted). Autochthonous populations are used in the forest coffee system, where coffee cherries are harvested directly from naturally regenerating populations without managing the forest. In the case of the semi-forest coffee production system the density of wild *Coffea arabica* individuals is increased through thinning of understorey trees and shrubs, thereby giving more space and light to *Coffea* (Dubale and Teketay 2000; Gole et. al, 2001). Gole et al. (submitted) and Senbeta (2006) showed in ordination analyses of

species inventories of afromontane forests from throughout the southwest of Ethiopia that there is a coffee forest community with a distinct composition of plant species. In the garden coffee system, varieties of *C. arabica* that were selected locally by farmers and also distributed by district agricultural offices are grown in small stands in farmers backyared and small coffee farms (<0.5 hectare; Teketay and Tigneh 1994; Woldetsadik and Kebede 2000). The fourth production systems are large scale plantations, in which commercial cultivars are grown and managed intensively (Van der Graaff 1981; Bellachew 1997; Bellachew et al. 2000; Woldetsadik and Kebede 2000).

The forest coffee ecosystem (forest and semi-forest coffee) is under serious threat due to deforestation caused by the establishment and expansion of big farms, human settlements, and replacement of the forest ecosystem with other cash crops (Ameha 1991; Getahun and Krikorian 1973; Gole et al. 2002). Effective conservation strategies for the endangered wild coffee with its forest ecosystem are therefore urgently needed. At the same time, amounts and geographical distribution of genetic diversity within and among *in situ* populations needs to be known to backup conservation strategies. Two lines of information are very important: the first regards to the kind and geographic distribution of genotypes, and will allow to select sites that include a representative spectrum of genotypes and also unique genotypes; the second regards to effects of vegetation management and coffee cherry collecting on genetic diversity, and to questions such as how big a forest stand needs to be to effectively conserve the genetic diversity of the coffee populations at the site. Moreover, the utilization of wild germplasm in breeding efforts can greatly benefit from the molecular characterization of genetic diversity (Tesemma and Belay 1991; Conner and Wood 2001).

Coffea arabica (2n=4x=44) is an allotetraploid, whereas all other *Coffea* species are diploid (2n=2x=22; Bridson and Verdcourt 1988; Charrier and Berthaud 1985; Bridson 1982; Stoffelen 1998). The reconstruction of phylogenetic relationships within the genus *Coffea* appeared difficult because of low amounts of variability in the genomic regions sequenced such as ITS (Lashermes et al. 1997) and *trn*L-F (Cros et al. 1998), and the resulting trees suffered from a lack in resolution and support. The recent analysis of a combined data set of 8 non-coding chloroplast genome regions and the *mat*K gene suggests that the cp genomes of *C. arabica* and *C. eugenioides* have diverged recently, clearly pointing to an ancestor of *C. eugenioides* as maternal parent

of *C. arabica* (Tesfaye et al. submitted (a)). Sampling individuals of *C. arabica* from forest populations throughout its native range in Ethiopia yielded only a single chloroplast haplotype (Tesfaye et al. submitted (a)). This is inline with the assumption of a single allopolyploidization event for the origin of *C. arabica.* Also there could have been a severe bottleneck situation in the recent evolutionary history of *C. arabica*. Moreover, phylogenetic analysis of East African species of *Coffea* using *trn*L-F, *rpl*16 and ITS sequences is underway (Mvungi pers. comm.), although not giving a definite answer on the origin of *C. arabica*.

RAPD analysis made by Lashermes et al. (1993) on *ex situ* collections of *C. arabica* from various sources revealed no variability among arabica samples. Similarly, RFLP analyses by Lashermes et al. (1996a) detected rather low levels of polymorphism between *C. arabica* accessions. However, later studies on genetic diversity of plants from Ethiopia using RAPD markers revealed higher amounts of polymorphism in coffee trees collected from Kaffa and Illuababor provinces (Anthony et al. 2001; Chaparro 2004). Furthermore, Anthony et al. (2002) provided evidence that there is high polymorphism among "subspontaneous" materials from Ethiopia as compared to commercial cultivars using AFLP (Amplified Fragment Length Polymorphic) and SSR (Simple sequence repeat) analysis. However, the dendrograms generated based on AFLPs and SSRs are not consistent with respect to the groups recognized. Genetic diversity analyses of *C. arabica* cultivated in Tanzania based on ISSR and RAPD markers also showed limited variability (Masumbuko et al. 2003; Masumbuko and Bryngelsson 2004). More recently, Aga et al. (2005) indicated low to moderate level of polymorphisms within and among forest *Coffea arabica* populations in Ethiopia, and were able to differentiate individuals from Bale, Jimma, Welage, and Illubabor. However, their sampling was not exhaustive interms of representing the whole wild coffee diversity in Ethiopia. The identification of additional forest coffee sites such as forests at Essera (Banja forest in Dawro Zone), where we even noted the mixture of different *C. arabica* phenotypes (light green and bronze tip leaves) not present anywhere else, underscores the need for additional sampling.

Generally most of the published studies on the diversification of *C. arabica* mainly considered RAPD and AFLP data from gene bank materials outside of Ethiopia (Lashermes et al.1996a, 1996b, Anthony et al. 2002), so that it was not possible to apply

a clear sampling scheme with exact documentation of the geographical origin of the analysed plants. There is thus a need to investigate the genetic diversity of wild populations of *C. arabica*, using other markers which will provide better resolution in terms of diversification within and between populations of wild C. *arabica*. In spite of its commercial importance, very few studies have so far been carried out on the wild populations in Ethiopia.

The generation of Inter-Simple Sequence Repeats (ISSR) is a quick and cost effective technique that is based on PCR amplification of inter-microsatellite sequences to target multiple loci in the genome. These markers are found to be useful in the broad application for analyses of genetic variation below the species level, mainly in studying population structure and differentiation. ISSRs were proven to be reproducible markers (Zietkiewicz et al. 1994; Wolfe and Liston 1998; Prevost and Wilkinson 1999; Camacho and Liston 2001). In coffee, ISSR markers were applied to study diversity of *C. arabica* cultivars and to assess relationships among *Coffea* species (Ruas et al. 2003; Masumbuko and Bryngelsson 2004; Aga et al. 2005).

The work presented here is part of an integrative project aiming at developing strategies for conservation and sustainable use of wild *C. arabica* genetic resources in Ethiopia. The CoCE (Conservation and Use of Wild Populations of *Coffea arabica* in the Montane Rainforests of Ethiopia) research project aims to assess the diversity and the economic value of the Ethiopian coffee gene pool, and to develop concepts of model character for conservation and use of the genetic resources of *Coffea arabica* in its centre of diversity in Ethiopia.

Using genotypes that are based on ISSR banding patterns, the aims of this chapter are twofold: (1) To assess relationships of *Coffea arabica* populations throughout its native range in Ethiopia, and to evaluate the position of the core CoCE sites within overall patterns of genetic diversity; and (2) to evaluate if individuals from "wild" populations in forests are genetically different from landraces selected by farmers in the different geographical regions of Ethiopia.

3.2 Materials and Methods

3.2.1 Sampling strategy and populations studied

Sampling was carried out to cover all wild coffee areas and coffee producing regions of Ethiopia. As main CoCE project sites (in these sites multidisciplinary studies are carried out) Bale (Harrena Forest), Bonga, Berehane Kontir and Yayu (Geba Dogi Forest) were intially considered (Figure 3.1). These main sites were earlier considered as areas for forest coffee conservation in Ethiopia (Dubale and Tektay 2000). Moreover, individuals from a single plot of additional coffee forests (Anfilo, Bench Maji and Mankira) and two plots from Boginda were collected in order to have a better representation of the overall genetic diversity of *C. arabica* in Ethiopia. This was particularly important for evaluating the position of the CoCE main sites in the context of the overall genetic diversity of *C. arabica*, and to understand patterns of gene flow within and between regions. In addition, five individuals from Daphe, located close to Sudan border which is geographically isolated with no human activity were included. The leave samples we received from Daphe were from seeds collected randomly bulked and raised in *ex situ* maintenance site of Mugi district Ministry of agriculture development office (West Wollega).

Twenty five individual coffee trees from two years to very old were randomly sampled from 50 x 50 m plots (Figure 3.2). For this interregional analysis six individuals were randomly chosen out of the twenty-five. However, there were some samples that did not yield ISSR-PCR products for several of the primers, and were therefore excluded from data analysis. Representative selections of landraces were also sampled from Southwest Ethiopia (Wellaga, Jimma and Kaffa) and from Sidamo and Hararge to be compared with individuals from forest populations. As these plants are not forming natural populations, only one sample per landrace was collected from each locality. Generally, all the trees were labeled in the forest and passport data along with GPS coordinates were recorded. A list of sampled populations and localities is given in Table 3.1. The collecting sites are also shown on the map of Ethiopia (Figure 3.1).

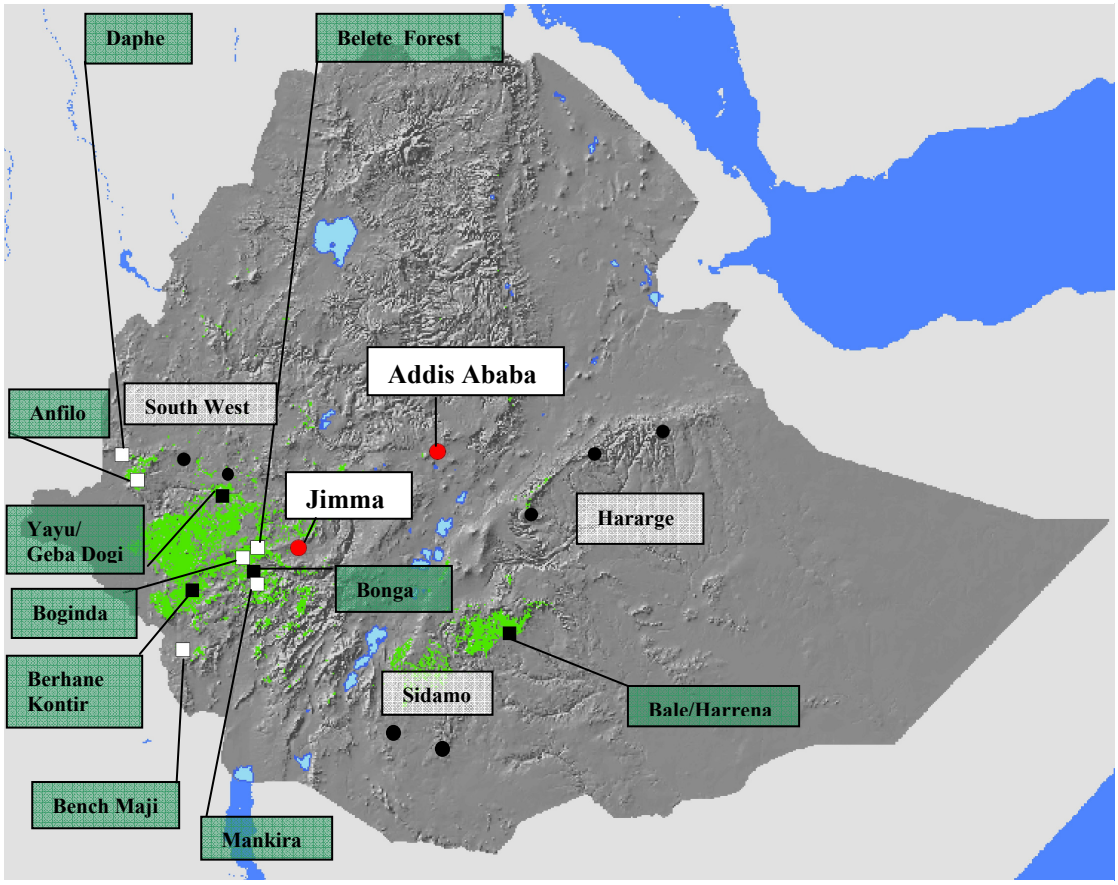


Figure 3.1 Map of Ethiopia and locations of where the sample collection made. Green is the forest area, dark boxes are core CoCE project sites, white boxes and dark dots are additional wild coffee and landraces included for genetic diversity analysis, respectively.

3.2.2 Plant materials and DNA extraction

A total of 125 samples (in total 93 wild *C. arabica* samples, 27 landraces and 5 commercial cultivars) were analyzed. An individual of diploid *C. eugenioides* was also included for reference. Young leaves were collected and dried in silica-gel and genomic DNA was extracted from silica-gel-dried leaf tissue using a modified CTAB method employing triple extractions to yield optimal amounts of DNA (Borsch et al. 2003). The genomic DNAs were further purified using the QIAquick PCR purification kit (Qiagen GmbH, Hilden, Germany) since ISSR-PCR of some of the samples was inhibited due to secondary compounds such as alkaloids (caffeine, trigonelline). Quality of the genomic DNA was checked by ethidium bromide staining on 0.9% agarose gels. DNAs were

then diluted for ISSR-PCR and adjusted in concentration. The adjusted genomic DNA dilutions were further tested on 0.9 % agarose gels for comparable concentration.

Table 3.1 Populations of wild *C. arabica* and landraces examined for inter simple sequence repeat (ISSR) variations. Most individuals were selected randomly as described in the methods from 50m X 50m plots. [1] indicates relaxed sampling, where sampling of individuals was carried out from an area bigger than the plots. Landraces and commercial cultivars are only represented by one individual and their origin is described with the respective names of the regions with their local name.

Region	**Plot codes**	**N_P**	**N_R**	**Location**	**Altitude (m a.s.l.)**	**Habitat Conditions**
Bale (Harrena)	I-2	6	15	06° 29' 14.1" N / 39° 45' 09.5" E	1513	Semidisturbed
	I-4	6				Semidisturbed
	I-5	3				Undisturbed
Bonga	II-1	4	16	07° 19' 57.3" N/ 36° 13' 31.0" E	1760	Semidisturbed
	II-3	3				Semidisturbed
	II-4	3				Undisturbed
	II-5	6				Semidisturbed
Berhane Kontir	III-1	5	13	07° 04' 801" N/ 35° 22' 563" E	1200	Semidisturbed
	III-5	5				Undisturbed
	III-6	3				Semidisturbed
Yayu /Geba Dogi	IV-1	5	16	08° 23' 994" N/ 35° 47' 718" E	1500	Semidisturbed
	IV-4	5				Undisturbed
	IV-5	6				Undisturbed
Boginda	V-1	5	10	07° 31' 572" N/ 36° 04' 270" E	1740	Undisturbed
	V-2	5				Semidisturbed
Bench Maji	VII-1	6	6	06° 14' 503" N/ 35° 15' 365" E	1640	Undisturbed
Anfilo	VIII-1	6	6	08° 36' 05.6" N/ 34° 37' 02.7" E	1617	Semidisturbed
Daphe*	IX-1	5	5	-	910	*
Mankira	X-1	6	6	07° 09' 49.8" N/ 36° 16' 23.2" E	1624	Semidisturbed
Sidamo landrace[1]	-	-	5	Kurme, Walisho, Dega, Qonga		
Hararge landrace[1]	-	-	13	Aruso, Bukuri, Abadro, Shumbure, Muyera, Cherchero, Fendisha, Buna Gurach, Buna Dima		
South West landrace[1]	-	-	9	Kuburi, Sendi, Beddessa, Buna Magala, Buna Abasha		
Other Cultivars[1]	-	-	5	Blue Mountain, San Ramon		
C. eugenioides	-		1	JARC/Ethiopia, [originally brought from Congo]		

**) Samples collected from ex situ maintenance site in Mugi District agricultural field. N_P= number of individual tree sampled per plot, N_R = Number of individual trees samples per region*

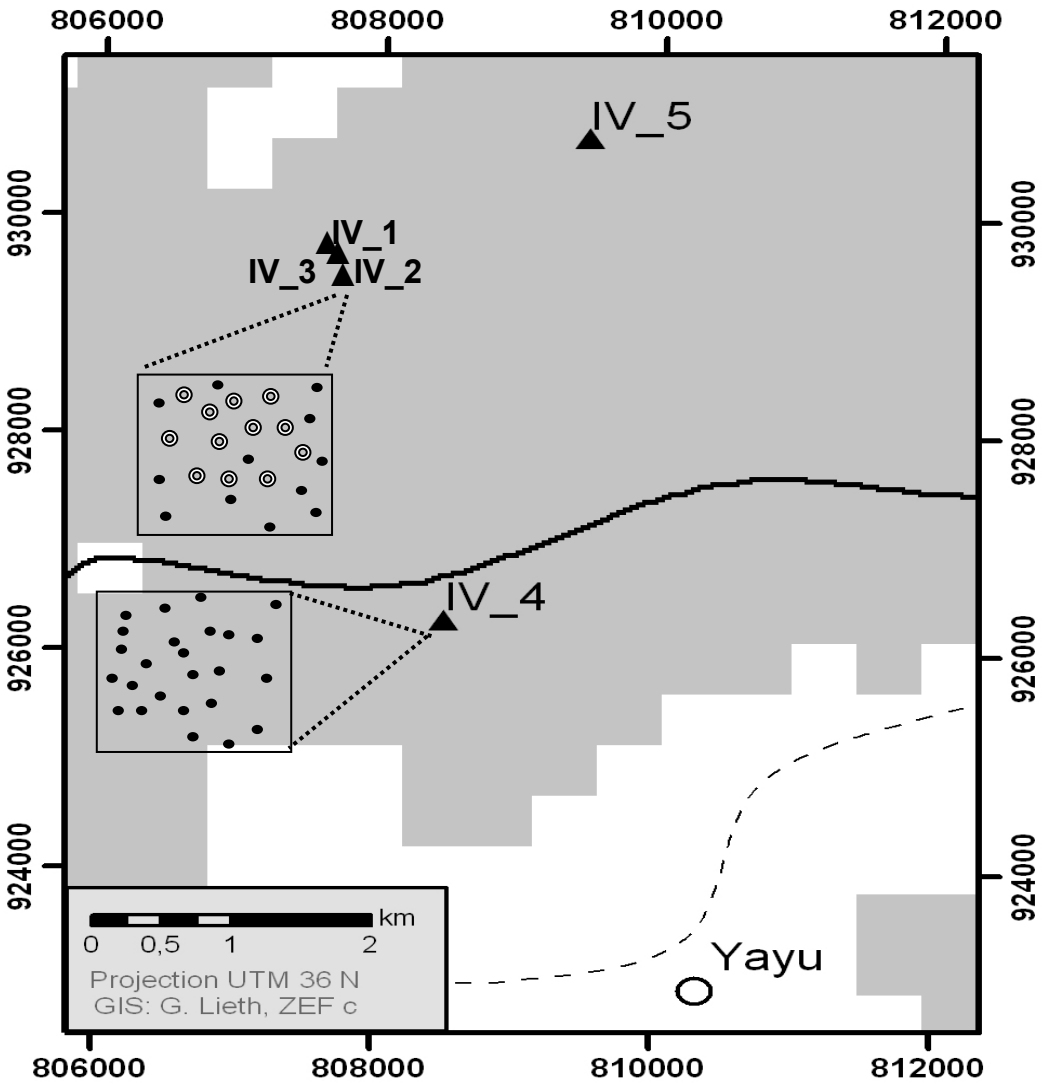


Figure 3.2 Sampled plots of Yayu/Geba Dogi Forest for the interregional analysis. The map is drawn using GPS coordinates taken from the plot. The grey areas mark the position of forest with coffee. The two boxes are a schematic overview of 50m X 50m sample plots IV-2 and IV-4, indicating the positions of individual coffee plants. Twenty five individuals collected per plot for entire analysis which is partitioned as inter and intraregional analysis. Only six individual included per plot for interregional analysis. The twelve samples shared among subprojects are illustrated by concentric circles. Plots 1, 2 and 3 are shared among the CoCE subprojects, whereas plots 4 and 5 are only included in the analysis of this project and considered latter in the intraregional analysis. The doted line shows the road (Bedele to Metu) and the continuous line stands for river (Geba River).

3.2.3 Primer selection and PCR conditions

A set of 120 simple sequence repeat (SSR) primers obtained from the University of British Colombia (Primer kit UBC 900) and primers used by Ruas et al. (2003) were used for initial testing of variability. These primers were di-, tri-, tetra- and penta-nucleotide repeat motives. The primers were screened against 18 samples of *Coffea* (11 samples from a single plot of 50 X 50m in Bonga forest (Kaffa/Ethiopia), one sample each from the CoCE-regions Bale/Harrena, Berhane Kontir, Yayu/Geba Dogi, and Boginda, two cultivars, and diploid *C. eugenioides*. These samples thus should represent a diversity gradient. HPLC (high performance liquid chromatography) purified eight di-nucleotides, one tri-nucleotide, and one tetra-nucleotide primers were finally selected for further analysis based on their polymorphism and reproducibility (Table 3.2).

The polymerase chain reactions were conducted in Biometra T3 Thermocycler. The amplification program was 4 min at 94^{o}C; 39 x 15s at 94°C, 1min at the primer annealing optimal temperature (45^{o}C or 48^{o}C), 1.30min at 72 ^{o}C, and at 72 ^{o}C for 5 min final extension. PCR amplification products were loaded onto a 1.7% agarose gel in TBE buffer, and then stained with ethidium bromide.

3.2.4 Data analysis

Unequivocally reproducible bands were scored manually as present ‘1’ or absent ‘0’ after photographing the gel with the Biodoc Analyze (035-300).The resulting binary character matrix was used for calculations to describe the similarity between individual samples as well as regions. The Jaccard’s coefficient was calculated using NTSYS-PC-version 2.1 (Rohlf, 2000) and Free Tree 0.9.1.50 (Pavlicek et al., 1999) softwares. The unweighted pair group method with arithmetic mean (UPGMA) (Sneath and Sokal 1973) was used to analyses and compares the individual genotypes as well as regions and generates phenogram using NTSYS-PC- version 2.1 software (Rohlf 2000). The neighbor-joining (NJ) method (Saitou and Nei 1987; Studier and Keppler 1988) was used to compare individual genotypes and evaluate patterns of genotype clustering using Free Tree 0.9.1.50 software (Pavlicek et al. 1999). The major difference between the two algorithms is that UPGMA assumes equal rates of evolution (molecular clock assumption) along all branches, whereas neighbor-joining assume variations in the rate

of change (Saitou and Nei 1987; Studier and Keppler 1988; Nei and Kumar 2000; Lan and Reeves 2002).

To further examine the patterns of variation among individual samples, a principal coordinated analysis (PCO) was performed based on Jaccard's coefficient (Jaccard 1908). The calculation of Jaccard's coefficient was made with PAST version 1.18 software (Hammer et al. 2001). The first three axes were latter used to plot with STATISTICA version 6. software (Hammer et al. 2001; Statistica Soft, Inc. 2001).

The genetic diversity as percent polymorphism was calculated for each population/plot and region based on the banding profile using POPGENE 1.31 (Yeh and Boyle 1997). The Shannon index was calculated as $H = -\sum piLog2\ pi$, in which pi is the frequency of a given ISSR fragment, for each population (plot) or region (Lewontin, 1972). Shannon's index of diversity was used to measure the total diversity (Hsp) as well as the mean intra-population/-region diversity (Hpop). The proportion of diversity between populations/regions was then calculated as (Hsp – Hpop/Hsp). Correlation analysis Shannon's index verses abiotic and geographical factors were also done. An analysis of molecular variance (AMOVA) (Excoffier et al. 2005) of regions of wild coffee which was used to estimate variance components of the ISSR data and partitioning the total variation to different hierarchical level were done using Arlequin version 3.0b software (Excoffier et al. 2005).

3.3 Results

3.3.1 Genetic diversity

PCR amplification with SSR primers yielded from 7 bands (for dinucleotide repeat primer 814) to 22 bands (for dinucleotide primer 844), and 31 bands for the tetranucleotide repeat primer (CoIS001) (Table 3.2). The scored DNA fragments varied in size from approximately 300bp to 2500bp. Among the total of 148 characters coded per individual sample the eight di-, one tri- and a tetra-nucleotide primers yield, 56.8% (84 characters) and 33.8% (50 characters) polymorphism for wild and landraces respectively.

Table 3.2 Sequences of primers used in this analysis and numbers of loci detected with this 10 primers and percent polymorphism in comparison between the *C. arabica* individuals from all wild population and from all landraces.

Primer	Sequences, 5' to 3'	No. of scorable bands	Wild /Semiwild		Landraces	
			P	%P	P	%P
810	GAGAGAGAGAGAGAGAT	9	2	22.2	1	11.1
812	GAGAGAGAGAGAGAGAA	20	10	50.0	5	25.0
813	CTCTCTCTCTCTCTCTT	9	6	66.7	2	22.2
814	CTCTCTCTCTCTCTCTA	7	4	57.1	3	42.9
818	CACACACACACACACAG	16	11	68.8	1	6.3
834	AGAGAGAGAGAGAGAGYT	12	2	16.7	1	8.3
844	CTCTCTCTCTCTCTCTRC	22	15	68.2	9	40.9
860	TGTGTGTGTGTGTGTGRA	8	2	25	2	25.0
866	CTCCTCCTCCTCCTCCTC	14	1	7.1	0	0
CoIS001	CCTACCTACCTACCTA	31	31	100	26	83.9
Total	Total loci found (Total polymorphism)	148	84	56.8	50	33.8

P = number of polymorphic bands, %P = percent polymorphism

The eight di- and trinucleotide primers contributed 63.1% and 48.0% of the total polymorphism observed with entire data set in the wild and landraces individuals, respectively. Out of the eight di- and a trinucleotide primers used for this study, primers 844 (68.2% and 40.9%) and 812 (50% and 25%) showed the maximum polymorphism (wild individuals and landraces). The smallest amount of polymorphism resulted from trinucleotide primer 866. Among wild individuals polymorphism observed was 7.1% and no variation was also detected among landraces with primer 866. Based on only di- and tri-nucleotide primers, higher polymorphism was detected in Yayu/Geba Dogi population IV-5 (16.2%), where as population from Daphe (2.6 %) showed the least polymorphism observed among the population analyses.

However, the single tetranucleotide primer (CoIS001) with 31 bands contributed 36.9% and 52% polymorphism observed in wild and landraces, respectively. Generally, the polymorphism generated by the tetranucleotide primer was considerably higher, being 83.9% for Mankira and 77.4 % for populations IV-4 and IV-5 from Yayu (Geba Dogi).

The total locus diversity across populations/plots and regions was determined using data from eight di- and a tri-nucleotide (combined) and one tetranucleotide primer (Table 3.3).The highest amounts of polymorphism were observed for plots IV-5 and IV-

4 from the Yayu (Geba Dogi) region (29.1% and 26.4%, respectively) followed by plots II-5 from Bonga and the Mankira population (25.7% and 23.7%, respectively).

On the other hand, Shannon's diversity index values (data generated by all the primers) for Bonga populations/plots (II-3 and II-5) and Yayu (Geba Dogi) population/plot IV-4 were highest with the value of 0.45. At regional level also Bonga and Yayu (Geba Dogi) showed higher diversity (0.40), followed by the Berhane Kontir forest with 0.34. The lowest Shannon's indices are observed in Bench Maji (0.22) and among the landraces (0.17-0.22; Table 3.3). Generally, there is a trend towards higher diversity in the semidisturbed plots as compared to the undisturbed plots, except for Bale plot II-5 and Yayu plot IV-4 which has slightly higher diversity than the semidisturbed plots. Moreover, the landraces (average on all samples) showed generally lower variability as compared to individuals from the wild. The South west landraces (0.22) showed a slightly higher diversity than landraces from Sidamo (0.21) and Hararge (0.17; Table 3.3).

The average Jaccard's coefficient among wild coffee regions (calculation based on eight di-, and a trinucleotide primers) ranged from 0.954 (Mankira and Daphe) to 0.883 (Berhane Kontir and Mankira). From all regions Berhane Kontir stands out by benign genetically more distant from the other regions than each of these regions to each other. The only wild coffee forests found in the southeast of the Great Rift Valley (Bale/Harrena) were observed to be close to Bonga population (Table 3.4, Figure 3.7). The landraces studied showed the highest genetic similarity with samples from Mankira and Daphe forests (Table 3.4, Figure 3.7). Moreover, no correlation detected with pair-wise comparisons of genetic distance (Nei 1972) and geographic distance among Bale, Bonga, Berhane Kontir, Yayu and Boginda (Figure 3.10). Based on seedling parameters of coffee Seifu et al. (2005) also observed no correspondence between the geographic and genetic distance.

Table 3.3 Shannon's diversity index (H) and percentage of ISSR band polymorphism (P) of *C. arabica* in Ethiopia. Subtotals are provided for different regions of wild population and landraces by calculating index values for the respective set of individual.

Region	Plot codes	Percent Polymorphism (%P)		Over all %P	Shannon's diversity Index (H)		Over all H
		Di-/tri-Primers	Tetra-Primer	All primers	Di-/tri-Primers	Tetra-Primer	All primers
Bale (Harrena)	I-2	9.4	35.5	14.9	0.12	0.47	0.20
	I-4	5.1	41.9	12.8	0.14	0.32	0.18
	I-5	6.0	58.1	16.9	0.13	0.51	0.21
Over all Bale/Harrena		13.7	80.7	27.7	0.18	0.53	0.25
Bonga	II-1	9.4	67.7	21.6	0.13	0.61	0.23
	II-3	9.4	61.3	20.3	0.37	0.74	0.45
	II-4	7.7	19.4	10.1	0.14	0.54	0.22
	II-5	12.8	64.5	23.7	0.31	0.71	0.45
Over all Bonga		28.2	93.6	41.9	0.21	0.60	0.40
Berhane Kontir	III-1	13.7	51.6	21.6	0.21	0.61	0.30
	III-5	14.5	51.6	22.3	0.22	0.39	0.25
	III-6	6.8	41.9	14.2	0.09	0.35	0.14
Over all Berhane Kontir		20.5	87.1	34.5	0.28	0.56	0.34
Yayu/Geba Dogi	IV-1	10.3	71.0	23.0	0.20	0.76	0.32
	IV-4	12.8	77.4	26.4	0.26	0.70	0.35
	IV-5	16.2	77.4	29.1	0.24	0.58	0.31
Over all Yayu/Geba Dogi		20.5	93.6	35.8	0.30	0.76	0.40
Boginda	V-1	12.0	51.6	20.3	0.18	0.44	0.23
	V-2	10.3	67.7	22.3	0.13	0.57	0.23
Over all Boginda		16.3	83.9	30.4	0.18	0.59	0.27
Bench Maji	VII-1	5.1	67.7	18.2	0.13	0.54	0.22
Anfilo	VIII-1	7.7	67.7	20.3	0.21	0.53	0.28
Daphe	IX-1	2.6	71.0	16.9	0.17	0.60	0.25
Mankira	X-1	10.3	83.9	25.7	0.20	0.75	0.31
Sidamo Landraces	-	7.7	54.8	17.6	0.14	0.46	0.21
Hararge Landraces	-	10.3	80.7	25.0	0.08	0.49	0.17
South West Landraces	-	15.4	58.1	24.3	0.17	0.40	0.22

3.3.2 Cluster analysis

Dendrograms constructed on the basis of neighbor joining and UPGMA of 126 individuals of *C. arabica* using Jaccard's coefficients of similarity revealed different groups (Figures 3.3, 3.4, 3.5, and 3.7). The group on the top of the tree (Figure 3.3) is dominated by landraces and cultivars (including the important commercial cultivar 'Blue Mountain'). In additions two landraces from Eritrea, several individuals from Hararge, and Sidamo also grouped together with landraces-cultivar group. The rest groups are either entirely wild or dominated by wild. Some regions (Boginda, Bale/Harrena, Mankria, Bench Maji and Daphe) show closely related genotypes whereas individual plants of others (e.g., Bonga) can be found almost throughout the whole range of genotypes and are extremely heterogeneous genotypes. The group which was dominated by landraces appear to have shorter branches as compared to the wild groups (dominated by individuals from wild populations). However, the clustering algorithms were not able to describe the full variation in the datasets because this variation was not only linear. Generally, the neighbor joining and UPGMA clustering methods of the di- and trinucleotide did not produce exactly the same tree topology. Nevertheless, wild versus landrace grouping is observed to be distinct in both and the grouping of the wild individuals based on population/plot of origin was better observed in UPGMA. This kind of phenomenon also observed with in the genus Focus (Phaophyceae) in which UPGMA and neighbor joining methods resulted in different tree topology (Billard et al. 2005).

The neighbor-joining and UPGMA trees of one hundred twenty six individuals of coffee with all the data set (including tetra primer) were also shown in Figures 3.5 and 3.6. The landraces group is also recovered in both in neighbor-joining and UPGMA trees of all data set (Figures 3.5 and 3.6). Furthermore, some tendency of grouping among wild individuals collected from the same regions also observed (Bale/Harrena and Daphe) (Figure 3.6). However, the trees based on the tetranucleotide repeat seems to be not resolved well since the level of variation is so high where the individual in the wilds were observed to be unique in the case tetranucleotide primer.

Figure 3.3 Neighbour joining analysis of complete interregional dataset (126 individuals) based on eight dinucleotide (810, 812, 813, 814, 818, 834, 844 and 860) and one trinucleotide primers (866). The algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

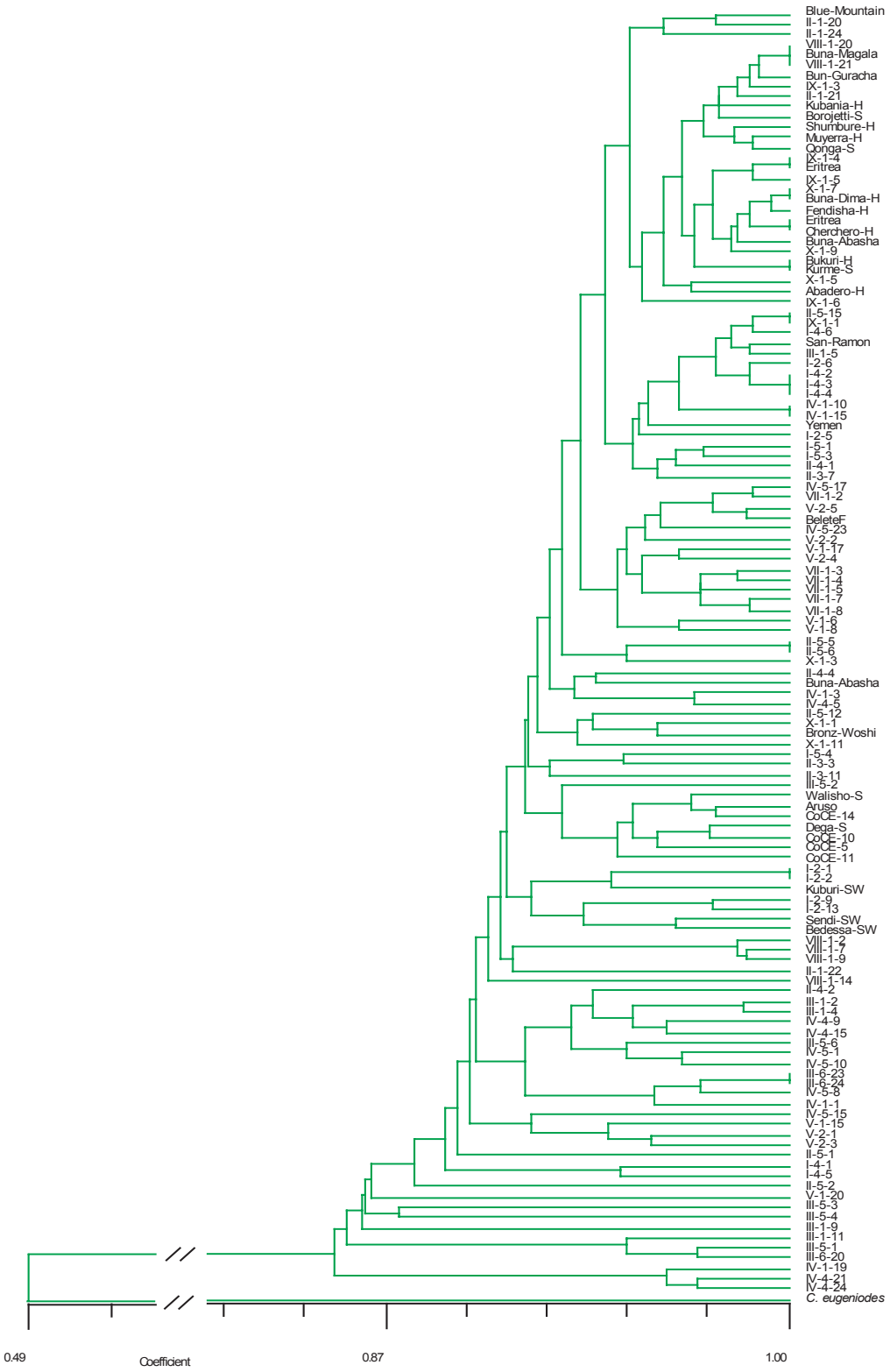


Figure 3.4 UPGMA analysis of complete interregional dataset (126 individuals) based on eight dinucleotide (810, 812, 813, 814, 818, 834, 844 and 860) and one trinucleotide primers (866). The algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

3.3.3 Principal coordinate analysis (PCO)

To explain the variation in the dataset in more than two dimensions, the data from the di- and trinucleotide of all genotype were subjected to PCO analysis using Jaccard's coefficient of similarity. The first three coordinates of the PCO having eigenvalues ranging from 2.5 to 3.9 accounts for 12.2%, 8.6% and 6.1% of the total variance (26.9% cumulative; Figure 3.5). PCO revealed basically the same major clusters of individuals. It showed also the same patterns, landraces and cultivars observed to concentrate in one place but the wild materials covered most of the PCO space.

Table 3.5 Analysis of molecular variance (AMOVA) for nine regions of wild *C. arabica*. The first three rows are AMOVA with geographical region structuring. The last two rows are AMOVA for all regions of wild coffee with out any grouping.

Source of variation	d.f.	Sum of Squares	Variance Components	Percent of Variation	Fixation Indices	P
Among Geographical Region	3	0.82	0.006	5.5	0.017	0.07
Among regions within Geographic groups	5	0.60	0.002	1.6	0.071	0.16
Within Regions	89	9.19	0.103	92.9	0.056	0.19
Among Regions	8	8.07	0.072	22.6	0.23	0.00
Within Regions	89	21.92	0.246	77.4		

3.3.4 Genetic differentiation at different levels in space

The group dominated by wild (Group 2, 3 and 4) on the hierarchical phenogram of Figure 3.3 tends to cluster together but occupy different position in PCO plot space. In general, PCO showed high all over variability of versus small scale hierarchical patterning (Figure 3.8). However, samples from Bonga and very few from other showed some tendency of dispersed all over (Figure 3.3). In some cases the samples collected from the same population grouped together, this is evidenced in the case of Boginda, Mankira, Bale/Harrena (I-4), Bench Maji and Daphe (Figure 3.3).

Table 3.6 Partitioning of the genetic variation into within and between regions of the nine wild coffee regions of Ethiopia based on Shannon's information index.

Parameter	**Mean**
Hpop	0.30
Hsp	0.39
Hpop/Hsp	0.78
1-Hpop/Hsp	0.22

Hpop = mean genetic variation for the regions; Hsp = mean genetic variation for the entire data; Hpop/Hsp = proportion of genetic variations within wild coffee regions (Hsp-Hpop)/Hsp = proportion of genetic variations between wild coffee regions

Analysis of molecular variance (AMOVA) which considered the entire native range of *C. arabica* with all data set found that 77.4% of the variance can be attributed to variation within regions and 22.6% to variation among regions (Table 3.5).When regions are further restructure into groups based on geographic proximity, 5.5% between geographic regions, 1.6% among regions of wild coffee with in the geographic regions, 92.9% within regions. This result were also incongruence with the result of Shannon's estimate of partitioning variation and demonstrating more variations (78%) within regions as compared to between regions of wild coffee variation (22%) (Table 3.6). AMOVA analysis revealed highly significant differences among wild coffee regions (P = 0.00) (Table 3.5).

Figure 3.5 Neighbour joining analysis of complete interregional dataset (126 individuals) based on eight dinucleotide (810, 812, 813, 814, 818, 834, 844 and 860), one trinucleotide primers (866) and a tetra nucleotide (CoIS001) primers. The Neighbour joining algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

Figure 3.6 UPGMA analysis of complete interregional dataset (126 individuals) based on eight dinucleotide (810, 812, 813, 814, 818, 834, 844 and 860), one trinucleotide primers (866) and a tetranucleotide (CoIS001) primers. The UPGMA algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

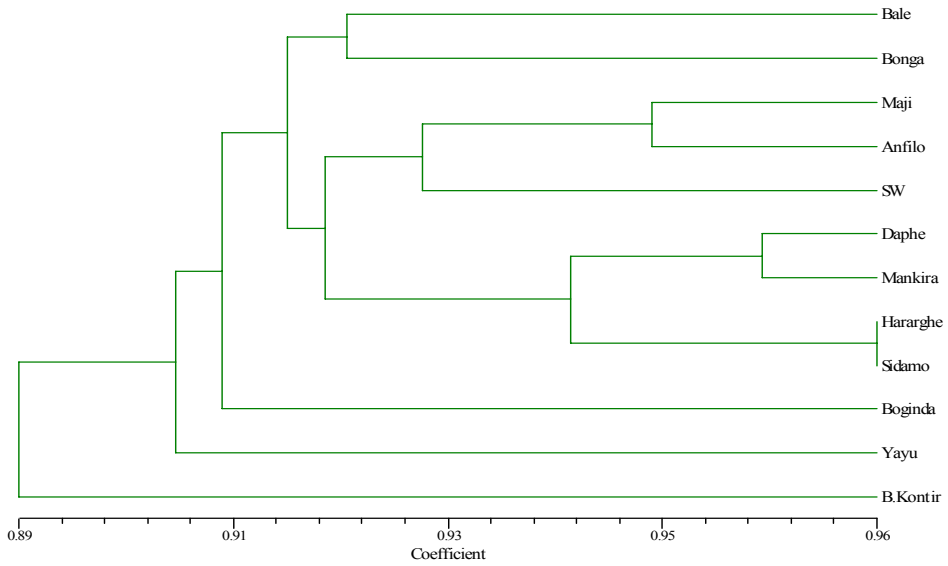


Figure 3.7 Individual genotypes were constrained to regional groups based on this constrained Jaccards coefficient were calculated. The dendrogram were generated by UPGMA cluster analysis using the sequential agglomerative hierarchical nested cluster analysis (SAHN). The individual genotypes were grouped in to nine wild coffee regions and three landraces groups collected from south west and south east Ethiopia.

3.3.5 Abiotical factors and genetic diversity

The interregional ISSR dataset was also used to gain some insights on possible effects of abiotic and geographical factors on amounts of genetic diversity within populations. The correlation analysis between genetic diversity (Shannon's diversity index and percent polymorphism) and precipitation and latitude were positive. It is observed that Shannon's diversity index barely significant for both average perceptions ($R^2 = 0.24$, $P = 0.048$) and latitude ($R^2 = 0.27$, $P = 0.031$), however, percent polymorphism ration showed highly significant correlation with both average perceptions ($R^2 = 0.55$, $P = 0.001$) and latitude ($R^2 = 0.45$, $P = 0.003$) (Figure 3.6). However, the exclusion Bonga II-3 population from the correlation analysis of Shannon's index of both resulted in a significant correlation of the remaining populations with increasing diversity at higher latitudes ($R^2 = 0.49$, $P = 0.003$) and monthly precipitation ($R^2 = 0.25$, $P = 0.046$) (data not shown).The study showed that a clear relationship between genetic diversity within populations is higher in the more humid north western plots and decreases with decreasing latitudes and annual precipitation levels. However, one semidisturbed plot from Bonga (II-3) shows exceptionally high diversity, where human influence (transfer

of plant material) is a likely cause. The other parameters mean monthly temperature, elevation and longitude showed no strong correlation with diversity index (Figure 3.6).

3.4 Discussion

3.4.1 Utility of ISSR markers in *Coffea*

ISSR markers have proven to be effective in detecting very low levels of genetic variation within different plant species (Zietkiewicz et al. 1994). It has also been reported that ISSR markers reveal genetic diversity with in the forest plant species as well as in wild and cultivated coffee (Deshpande et al. 2001; Raus et al. 2003; Aga et al. 2005). The number of polymorphic band ranged from 7.1% for trinucleotide primer (866) to 100% for the tetranucleotide primer (CoIS001). This is much higher than what was reported by Aga et al. (2005) where the number of polymorphic bands was ranged from 8% for pentanucleotide primer to 64% for anchored dinucleotide primer. The tetranucleotide - CoIS001 $(CCTA)_4$ in this study showed 31 bands in which all of them were polymorphic among wild coffee and also showed the highest polymorphism among the landraces. Ruas et al. (2003) also got 100% polymorphism with the same primer among eight *Coffea* species and six interspecific hybrids. In addition, the tetranucleotide primer amplified a large number of bands in the present study, in contrast to earlier studies by Ruas et al. (2003) on *Coffea* species and hybrids. However, previous studies based on *ex situ* materials found outside Ethiopia showed no variability using RAPD marker (Lashermes et al. 1993).

3.4.2 Support for wild *C. arabica* as distinct from semi-domesticated plants

The existence of truly wild *C. arabica* in contrast to semi-domesticated plants was evidenced on the basis of the occurrence of the plants in more or less undisturbed forests and by information from the local communities. Wild populations are self regenerating and not selected by human activity for any trait. Previous studies using genetic markers have largely used so called spontaneous and subspontaneous materials that could not clearly be assigned to a certain forest stand or farmer's garden (Lashermes et al. 1999; Anthony et al. 2002). Montagnon and Bouharmont (1996) based on morphology observed grouping among coffee from West of Great Rift Valley and East of Great Rift Valley, and suggest that the sub-spontaneous coffee in the West has not been involved in the domestication of *C. arabica*. However, with this analysis it is

possible to show that truly wild *Coffee* growing in more or less undisturbed forests (forest and semi-forest coffee systems) is genetically distinct from semi-domesticated *C. arabica*. Landraces or farmer's varieties have been selected by local farmers over hundreds and perhaps thousands of years (Andrea 1999; Gepts 2004). This study can show that there are distinct groups of landraces, which have originated in close geographical proximity to their wild relatives. The wild and cultivated enset samples from SW Ethiopia also observed to be genetically distinct from each other as indicated by their considerable differentiation in RAPD profiles (Birmeta et al. 2004).

On the neighbour joining tree of the individual samples, the landraces and cultivars group appear as having shorter branches compared to the groups that are dominated by individuals from wild populations. The results were a first indication for the existence of several lineages of landraces, which probably arose in geographically different regions. On the 3D plot of PCO it is clearly evident that the wild speared all over the 3D space while landraces were concentrated in one place. Moreover, ISSR data point to lower genetic distances among individuals from landraces as compared to individuals from wild populations, which could probably indicate reduction of variability during domestication and subsequent selection of landraces by farmers (Gepts 2004; Casa et al. 2005). The genetic diversity of tropical tree (*Inga edulis*, Leguminosae) with nuclear microsatellite marker indicates lower allelic variation in planted stand that the natural forest stands of the species (Hollingsworth et al. 2005).

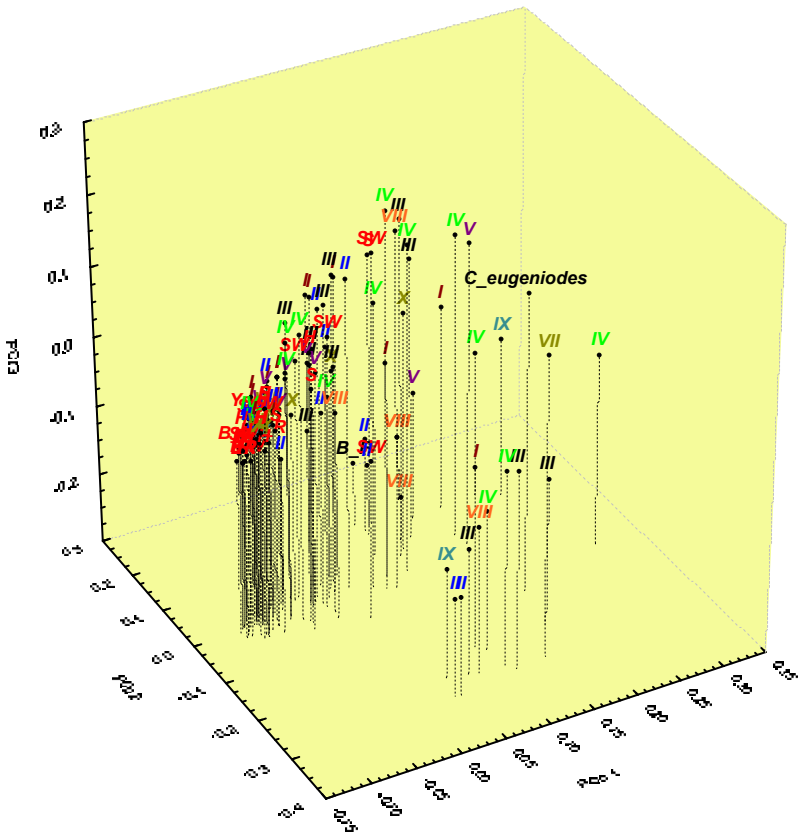


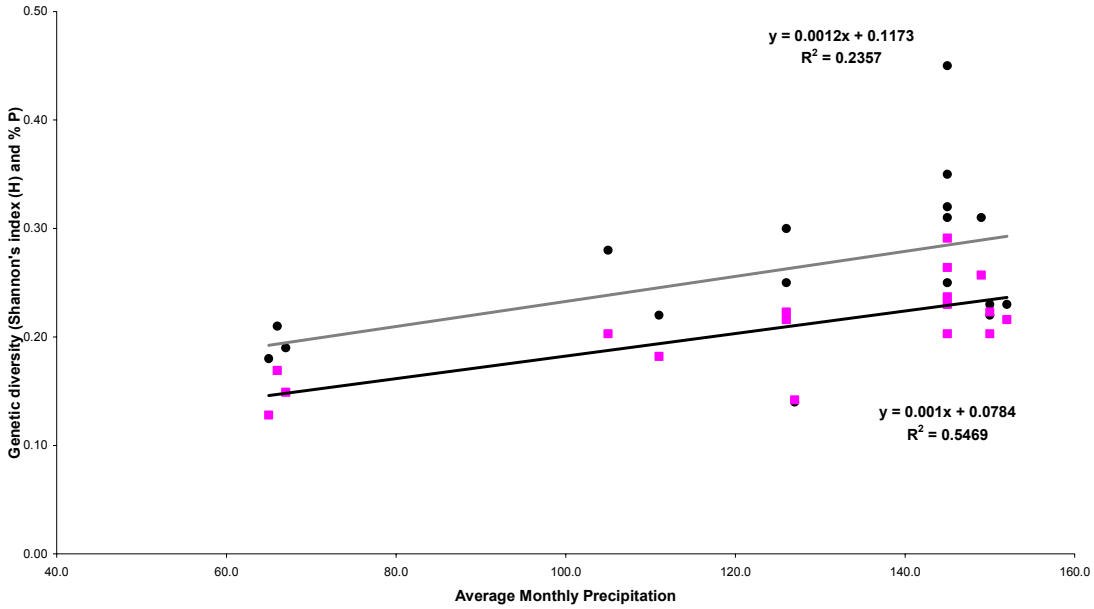
Figure 3.8 Three-dimensional representation of a principal coordinate analysis of the genetic relationships among 126 individuals of wild *C. arabica*, inferred from a distance matrix using the Jaccards index. Roman numbers indicate the region where individuals collected (Table 3.1).

In this analysis of dendrogram and PCO analysis of individual samples clearly separate landraces from wild materials. The addition of landraces noticeably showed that wild populations of *C. arabica* are genetically different and can be distinguished from semi-domesticated plants. In previous studies, all of these plants were subsumed under "spontaneous and subspontaneous materials" (Lashermes et al. 1993, 1999, 2000; Anthony et al. 2001, 2002). Similar result was also obtained with the analysis of cultivated and wild sorghums (*Sorghum bicolor*) using simple sequence repeat marker. The neighbor-joining analysis of sorghum showed that wild sorghums generally formed a distinct group, and about half the landraces tended to cluster by race (Casa et al. 2005). In addition, RAPD analysis of enset (*Ensete ventricosum*) collected from Ethiopia showed distinction of wild from cultivated (Birmeta et al. 2004). The analysis of diversity of wild and semi-wild *C. arabica* materials with RAPD marker exhibited

higher diversity for the materials collected from Kaffa region (Chaparro et al. 2004).The same result also observed in this study with the population from Bonga and Mankira populations.

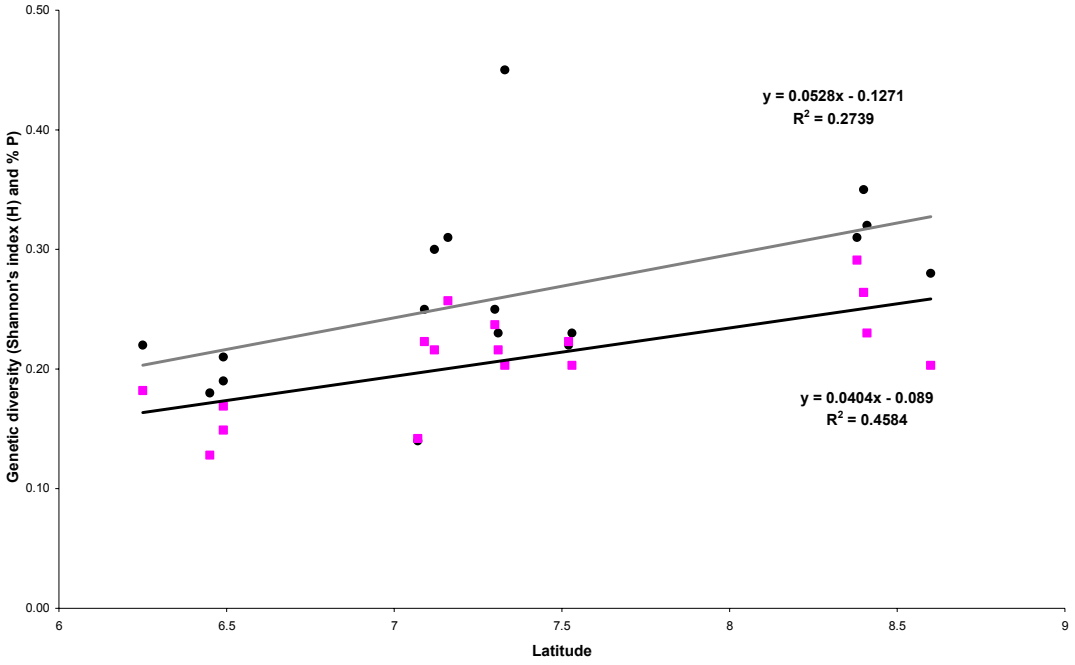
3.4.3 Relationships among wild *C. arabica* populations in Ethiopia

Conversely, Anthony et al. (2001) using RAPD marker observed genetic differentiation between samples collected from Ethiopia into four groups; one from southwestern and three from southern and southeastern Ethiopia with the majority of the markers present in southwester group.

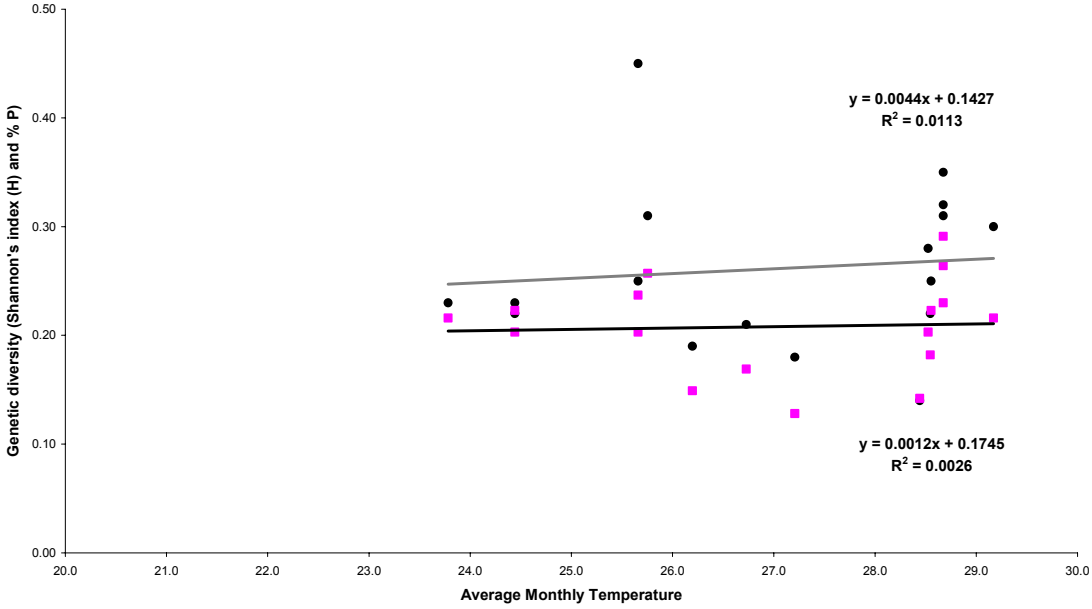


A

Figure 3.9 Correlation of genetic diversity based on Shannon's index and percent polymorphism(%P) to average monthly precipitation (A), latitude (B), temperature (C), elevation (D), and Longitude (E). Circles stand for Shannon's diversity index of individual plots/populations and rectangular point stands for ration of percent polymorphism. The upper trend line with the upper equation stands for Shanonn's index and the lower trend line with equation stands for percent polymorphism. The ration of percent polymorphism used directly (without multiplying with hundred) in order to plot along with Shannon's index.

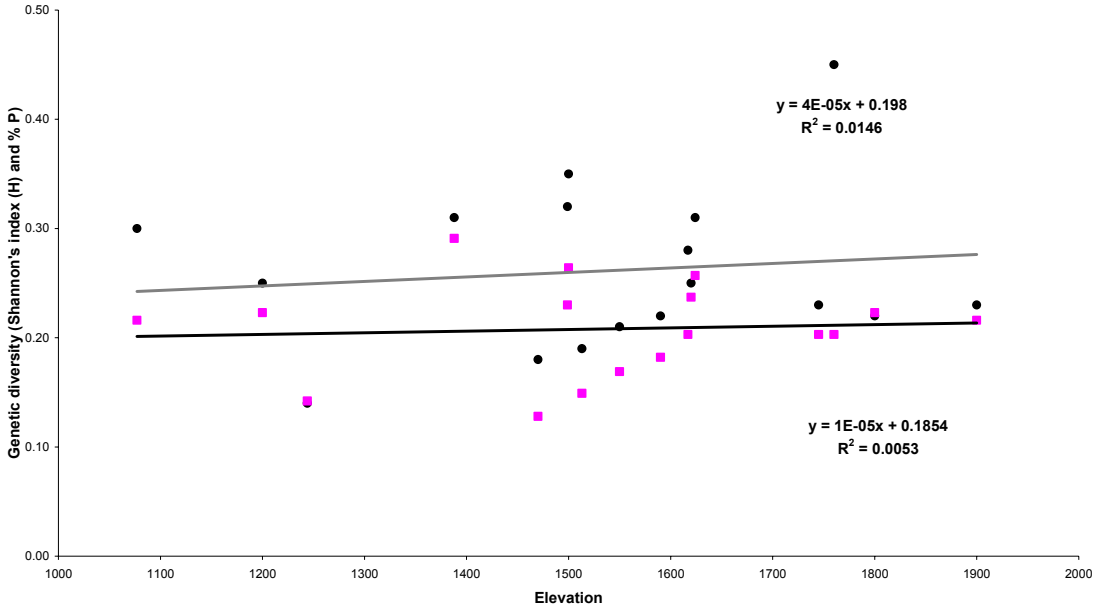


B

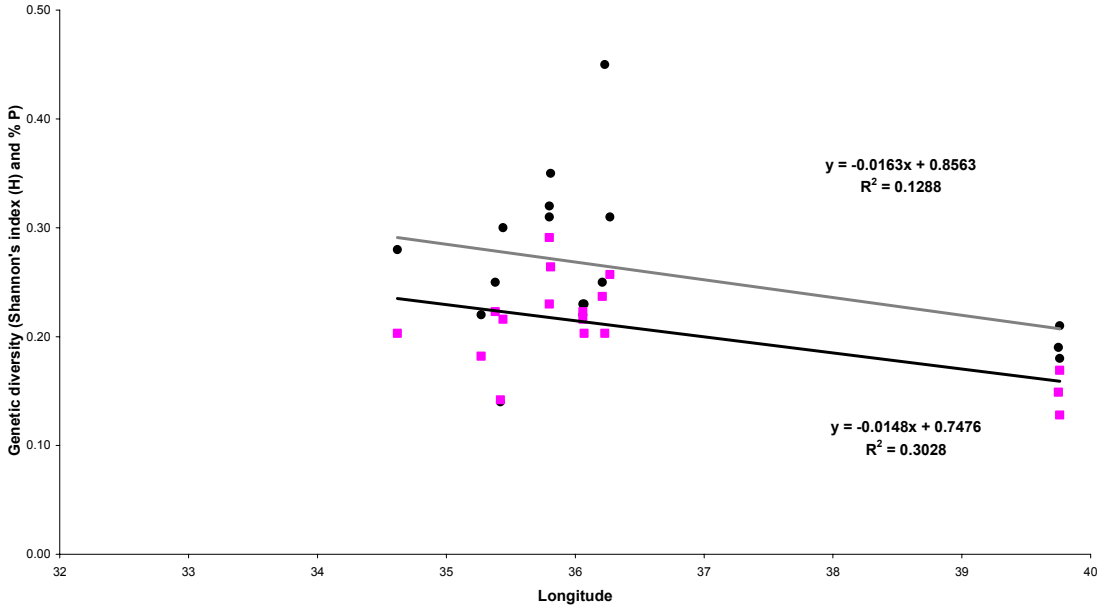


C

Figure 3.9 continued



D



E

Figure 3.9 continued

Furthermore, trees based the Jaccard's similarity coefficient of regions showed Hararge and Sidamo landraces grouped with Daphe and Mankira while south west

landraces form a group with wild populations of Anfilo and Bench Maji (Table 3.4; Figure 3.7). The former grouping might be the result of historical human activity around Bonga and Mankira due to the ancient kingdom of Kaffa since coffee were also treaded along with ivory, gold, and musk in the second half of nineteenth century (Zewede 2002; Wold-Mariam 2004). The latter grouping also showed the presence gene flow from wild to the landraces in Western Wollega where Anfilo forest is located. This is also evidenced in individual trees (Figures 3.3, 3.4, 3.5 and 3.6) since the majority of the landraces collected from South west were confined in wild groups. Some evidences of gene flow between wild and domesticated populations were also observed in the Argentinean common bean germplasm studied (Santalla et al. 2004). Regardless of the geographic distance Bonga and Bale region observed to be grouped together with Jaccard's coefficient similarity. High floristic similarities were observed between Bale and Bonga than Berhane Kontir and Bonga (Senbeta 2006).

3.4.4 Patterns of genetic diversity in wild *C. arabica*

The mating system of a species has implications for the patterns of intraspecific genetic diversity (Hamrick and Godt 1990, 1996; Nybom and Bartish 2000). Evidence for this has been provided through the observation of high sequence similarity among populations of selfing *Mimulus nasutus* (phrymaceae) as compared to the outcrossing *M. guttatus* (Sweigart and Willis 2003). Selfing is also known to reduce effective population size and therefore expected to result in lower levels of genetic variation than in comparable outcrossing taxa (Sgorbati et al. 2004). Common bean is generally considered to be an autogamous species but outcrossing rates as high as 60–70% has been reported (Wells et al. 1988). In addition, it seems that even the lowest rates of outcrossing reported are sufficient to generate broad variability over hundred or thousands of years (Beebe et al. 1997). In addition, together with metapopulation dynamics (i.e., population turnover) selfing can even lead to further reductions of genetic diversity within inbreeding populations (Ingvarsson 2002).

Genetic structure within a population or between populations is induced when gene flow by pollen and seed dispersal is limited. Thus, the genetic structuring observed among populations of wild coffee observed to be obscured because of the existence of gene flow between different population and regions. The exchange of one or more

individuals between two populations will prevent different neutral alleles at the same locus from being fixed in the populations (Slatkin 1987; Slatkin and Barton 1989). In the recent study of Aga, (2005) most of the coffee tree samples were clustered on the bases of their geographic origin but failed to cluster according to their respective populations. This could be due to the presence of substantial gene flow between local populations in the form of young coffee plants.

The presence of long-distance gene flow was also observed with low level of genetic differentiation in *Cercidiphyllum japonicum* (Cercidiphyllaceae) populations of riparian forest in Japan (Sato et al. 2006). Geleta et al. (2006) also observed grouping based on regions of origin for the majority of *Guizotia abyssinica* (Asteraceae) population collected from Ethiopia. High degree of genetic differentiation also confirmed by UPGMA tree topology of *Asparagus acutifolius* L. (Liliaceae) as a result of poor gene flow (Sica et al. 2005). These indicate geographical hierarchical patterning would therefore be expected in the absence of long distances gene flow. In this study also wild individuals from some regions (Boginda, Bale/Harrena, Bench Maji, Mankira and Daphe) tend to form distinct group based on plot and geographic origin (Figures 3.3, 3.4 and 3.5). However, the observed geographical hierarchical patterning was obscured with the long distance gene flow. The genetic diversity partitioning at different level also account 78% of the total diversity for the intraregional diversity. Recent study of forest coffee in Ethiopia using RAPD marker also showed more genetic diversity observed within population (65%) than between population (35%) (Aga et al. 2003). Based on this, the patterns of genetic diversity in *C. arabica* in Ethiopia could probably be strongly influenced by long distance gene flow.

Long distance gene flow depends on the performance of pollinators and seed dispersers. The behaviour of the pollinator is likely affected by interflowering-tree distance and also flowering tree density and performance of pollinators are probably responsible for the mating distance of tropical trees (Bawa 1998; Konuma et al. 2000). Konuma et al., (2000), detected poor genetic structure within the population *Neobalanocarpus heimii* (Dipterocarpaceae) in a lowland tropical rainforest of Malaysia as a result of both long-distance pollen flow and seed migration.

Current knowledge says that *C. arabica* is an in breeder (Carvalho et al. 1969; Charrier and Berthaud 1985; Purseglove 1968). However, these data were obtained

from commercial cultivars in plantations in Brazil. The results obtained on wild populations in Ethiopia are contradictory to what has been reported and showed significant amount of gene flow among populations and regions of wild *C. arabica*. On the other hand, Meyer (1965) reported 40% to 60% cross pollination in wild population of *C. arabica* in Jimma/Ethiopia. The study of *C. arabica* cultivars in different agroforestry systems of Indonesia showed that due to strong smell of *C. arabica* flowers there was frequent visit of coffee flowers by social and solitary bees that resulted in a significant increase of fruit set (Klein et al. 2003). African honey bees (*Apis mellifera scutellata*) were also the predominant floral visitors in fragmented habitats of Amazon region in Brazil. Bees were also observed to be important pollinators in degraded tropical forest and could also alter the genetic structure of remnant population of Amazonian tree *Dinizia excelsa* (Fabaceae), through frequent long-distance gene flow (Dick 2001).

In Ethiopia arabica coffee is well known to be a very important nectar source and provide pollen for African bees (*Apis m. Scutellata*; Fichtl and Adi 1994). Hence, honey bees and other bees might also be responsible for the observed gene flow among wild *C. arabica* populations in Ethiopia as the natural occurrence of these bees (500-2400 m a. s. l.) overlaps with the range of wild coffee (1,400-1,900 m a.s.l.; Geber-Egziabher 1990; Fichtl and Adi 1994; Gole 2003; Senbeta 2006). In addition, wild animals such as monkeys, baboons and hornbills birds may play a role in disseminating seeds through eating the berries at one place and defecting the seeds in other wild coffee populations (Aga et al. 2005; Senbeta 2006). Most importantly, man could also be one of the agents of gene flow via seedling exchange of preferred wild populations and also during thinning and filling up gaps in the forest with seedling collected from other forest patches (Gole 2003; Aga 2005; Senbeta 2006). However, this seems to be particularly relevant in certain geographical regions such as the Bonga forest and would explain why there are a few sites where a particular diversity of genotypes can be observed, with similar individuals otherwise only growing in other parts of Ethiopia. The very old human activities were to be more pronounced in the case of Bonga and Makira forest because of the presence of ancient kingdom of Kaffa in Bonga. Furthermore, the establishment of ancient market place near Bonga and the trade root that linked to NW and E were probably be an indication for early utilization of forest and forest product

including coffee (Zewede 2002; Wold-Mariam 2004). This historical phenomenon might also be relevant to explaining the patterns observed.

Moreover, the evolutionary history and a recent origin of *C. arabica* from single allopolyploidization event (Lashermes 1999; Tesfaye et al. submitted (a)) could in addition be responsible for a rather low genetic differentiation of *C. arabica* within its native range. The fact that only a single chloroplast haplotype was found in *C. arabica* indicates that all current populations have originated from a common ancestor in such short time, that no mutations could accumulate. This could either be due to a very recent allopolyploidization event or a strong bottleneck situation in the evolutionary history of *C. arabica*, perhaps caused by habitat changes. So far there is very little knowledge on the infraspecific differentiation and phylogeography of species in Afromontane forests (Dawson and Powell 1999).

Table 3.4 Jaccard's coefficients of similarity between the nine regions of wild coffee and landraces collected from three regions. Individuals from respective regions were group together and the presence absence data used to calculate Jaccard's coefficient.

	Bale	**Bonga**	**B.Kontir**	**Yayu**	**Boginda**	**Maji**	**Anfilo**	**Daphe**	**Mankira**	**Hararge**	**Sidamo**	**SW**
Bale	1											
Bonga	0.917	1										
B.Kontir	0.897	0.884	1									
Yayu	0.908	0.895	0.898	1								
Boginda	0.896	0.895	0.875	0.897	1							
Maji	0.908	0.929	0.891	0.913	0.931	1						
Anfilo	0.907	0.901	0.889	0.896	0.893	0.944	1					
Daphe	0.913	0.914	0.886	0.914	0.916	0.934	0.922	1				
Mankira	0.903	0.937	0.883	0.894	0.898	0.929	0.926	0.954	1			
Hararge	0.915	0.916	0.886	0.893	0.906	0.903	0.901	0.936	0.943	1		
Sidamo	0.906	0.890	0.889	0.901	0.907	0.904	0.892	0.928	0.941	0.964	1	
SW	0.918	0.906	0.889	0.909	0.912	0.921	0.927	0.927	0.927	0.922	0.896	1

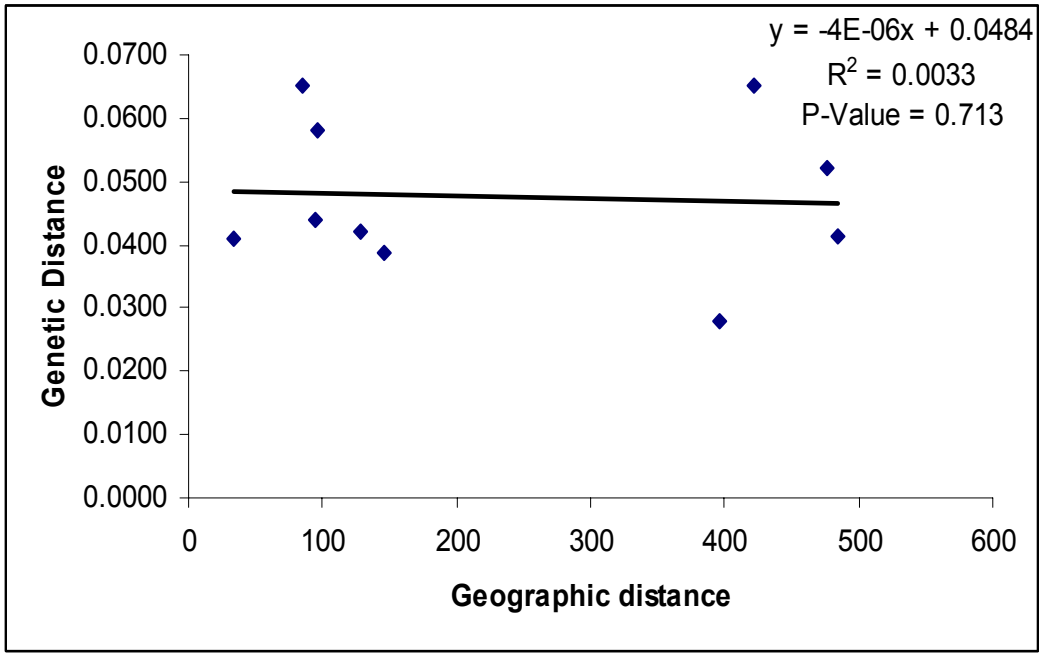


Figure 3.10 Pair-wise comparison and correlation of geographic distance (air distance) and Nei (1972) genetic distance of Bale, Bonga, Berhane Kontir, Yayu and Boginda. The others are excluded because of the different sample number and sampling scheme followed.

3.4.5 Differences in levels of genetic diversity among different plots

A range of deviating levels of genetic diversity as measured by Shannon's index (Table 3.3) was found within different plots. Shannon indices (H) show some interesting patterns that could provide answers as to how *Coffea* genetic diversity changes along geographical and climatological gradients in Ethiopia, and how genetic diversity could be influenced by forest management (forest coffee versus semi-forest coffee systems). Shannon's index for plots based on the same number of individuals (5 to 6 individuals) were used to do correlation analysis, and gradients of precipitation and latitude demonstrate some increasing trends with diversity in the plots. This is rather low sample number per plot to represent the genetic diversity in a population. Hence, the trend should be tested with denser samples per population as well as per region.

Most of the genetically variable regions were Yayu, Bonga, Berahne Kontir and Mankira, all of them are found in SW part of the Rift Valley. The highest percent polymorphism of wild population of *C. arabica* was observed in Yayu population/plot IV-5 (29.1%) and IV-4 (26.4%), on the other hand, the least was in observed in the other side of Great Rift Valley in Bale (Harrena) I-4 (12.8%). This could probable be because of the absence of close coffee forest in the other side of Rift Valley for possible

gene flow. However, Bonga was having the highest percent polymorphism (41.9%) at regional level analysis and followed by Yayu (35.8%) and Berhane Kontir (34.4%). Furthermore, Bonga population II-3 and II-5 (both 0.45) showed the highest Shannon's diversity index and followed by Yayu populations IV-4 (0.35) and IV-1 (0.32). At region level also Bonga and Yayu showed the highest regional diversity indices (0.40) followed by Berhane Kontir (I = 0.34) and Mankira (I = 0.31) demonstrated better diversity. The ecological study also showed that the gamma and beta diversity, Berhane Kontir forest ranks first could probably because of high habitat diversity (Senbeta 2006).

The comparison of genetic diversity between undisturbed and semidisturbed plots of the same region generally showed that the latter showed slightly higher diversity estimate. This is evidenced for example in populations of Bonga II-3 and II-5, Berhane Kontir III-1. Ecological studies of wild populations of coffee also showed that the intermediate disturbance of the habitat resulted in enhanced regeneration of coffee plant and increase the abundance of wild coffee (Gole 2003; Senbeta 2006). In the semi-forest coffee system which involves some management of thinning out trees, and shrubs competing with coffee in the lower story raises the frequency coffee by reducing the competition at early age of coffee (Gole 2003; Senbeta 2006). In this study gradient of genetic diversity observed with average monthly precipitation and latitudes (Figure 3.6). As an explanation for this genetic diversity gradient, a different flowering behavior of *C. arabica* is hypothesized. The presence of additional but low numbers of flowers over long periods of the year in addition to the flowering peak, leads to pollination by insects that have to travel far because of encountering fewer flowers in the humid forests. This, as a consequence, may lead to increased gene flow. Alternatively, historical reasons (diversity close to its centre of origin of *C. arabica* is higher; presence of refugia at certain latitudes) need to be tested. In longleaf pine (*Pinus palustris*) all the diversity parameters were correlated significantly with longitude (diversity decrease from west to east) and it was proposed that migration of this *Pinus* species occurred from a single refugium in the west after the Pleistocene glaciations (Schmidtling and Hipkins 1998). The reduction in genetic diversity with increasing distance from a refgium is a general phenomenon to be expected from repeated population bottlenecks at an advancing edge of a range in any species during postglacial expansion (Comes and Kadereit 1998). A

clear relationship between rainfall and genetic diversity was also evidenced in a study on African wild rice (IPGRI 2001).

3.5 Conclusions

The samples analysed here covers the entire distribution of both wild populations and landraces of *C. arabica,* which provides extensive information on *C. arabica* genetic diversity in Ethiopia. Gradients of genetic diversity were observed along abiotic and geographical factor which could be suggested the flowering initiations due to an extended rain and/or also closer to refugia centre. However, the hypothesis should be tested with denser sampling per plot and region.

The interregional comparison of wild *Coffea arabica* populations in Ethiopia also yielded a complex geographical distribution pattern of genotypes, with higher diversity within region/population. The genetic structure based on plot of origin seems obscured with long-distance gene flow. Hence, it is indeed important to evaluate the genetic diversity within region/population and genetic structure with denser samples. Furthermore, the co-dominant marker system will be particularly relevant in this case where high levels of gene flow are evident. Therefore, the nuclear microsatellites as an allelic marker system that allow analysing degrees of heterozygosity should be used to evaluate the gene flow and also level of inbreeding. Besides, this marker could also give insight on historical and current gene flow and understand the type and patterns of gene flow.

4 INTRAREGIONAL GENETIC DIVERSITY AND POPULATION STRUCTURE OF WILD *C. ARABICA* FROM BERHANE KONTIR AND YAYU (GEBA DOGI)

4.1 Introduction

Studying the genetic diversity of economically important species is essential for genetic resource management, improvement and sustainable use of the plant. Furthermore, the assessment of genetic diversity and relationships of populations in different localities can be useful for planning appropriate germplasm conservation strategies and for the selection of parents for hybridization (Dwivedi et al. 2001). Kanowski and Boshier (1997) and Newbury and Ford-Lloyd (1997) pointed out that analysis of genetic diversity is a prerequisite for planning *in situ* conservation measures. Distribution of genotypes may show complex patterns in space, which are characteristic at different levels, such as region, population, subpopulation or among neighboring individuals. Mutation, genetic drift and natural selection will also lead to the genetic differentiation of local population and gene flow will counter act differentiation (Slatkin 1987; Slatkin and Barton 1989). Moreover, levels of genetic diversity depend on the mating system, with higher levels in the majority of predominantly outcrossing taxa. In contrast, inbreeding species show lower levels of diversity, but greater interpopulationl variation (Hamrick and Godt 1990; Newbury and Ford-Lloyd 1997). Life forms and breeding system observed to have significant influence on genetic diversity and its patterns of distribution (Hamrick and Godt 1996; Nybom and Bartish 2000). Patterns of spatial distribution of genetic diversity are also strongly influenced by habitat heterogeneity and seed dispersal, including human activities (Knowles et al. 1992; Escudero et al. 2003).

Coffee is one of the most important commodities for many developing countries. It has an annual production around 4 million tons of green beans with sale between 6 to 12 billion dollars (Viniegra-Gonzalez 2000). The most important economic species are *C. arabica* which produces about 80% percent of the world's coffee, *C. canephora* nearly 20% percent and *C. liberica* has only minor importance (Purseglove 1968). The trade of coffee in Ethiopia is the largest export, which generates more than 60% of its total export earnings and the national production levels are estimated to vary between 140,000-180,000 tonnes (Asres 1996; Petty et al. 2004).

Arabic coffee has its origins in SW and SE mountain rainforest of Ethiopia and occurs at altitudes between 1,400 and 1,900 m a.s.l. (Meyer1968; Gebre-Egziabher 1990; Tadesse and Nigatu 1996; Gole et al. 2001; Gole et al. 2002). The altitude between 1300-1600 m also considered as a critical altitude for the occurrence and abundance wild coffee (Senbeta 2006).

C. arabica is the only allotetraploid (2n = 2x = 44) and autogamous species in the genus *Coffea*. The phylogenetic analysis of genus *Coffea* based on cpDNA and mtDNA suggests recent origin of the genus and rapid radial mode of speciation and also indicated that *C. eugenioides* as maternal progenitor of *C. arabica* (Berthou et al. 1983, Lashermes et al 1996; Cros et al. 1998). However, Lasherms et al. (1997) didn't detect additivity of parental rDNA (ITS2) types of the allotetraploid species *C. arabica* with the analysis of ITS2 sequence of *C. arabica* and other diploid taxa of *Coffea*. Furthermore, based on genomic *in situ* hybridization (GISH) and fluorescent *in situ* hybridization (FISH) analysis of *C. arabica* and other diploids showed *C. congensis* and *C. eugenioides* as progenitors of *C. arabica*, however, *C. canephora* also probed with *C. arabica* (Raina et al.1998). In many of these cases the conclusion made based on a limited data set and limited number of samples with narrow regional representation. However, recent analysis of chloroplast (cp) genome sequence data (ca. 7.2kb) by Tesfaye et al. submitted (a), *C. arabica* appears as a species that arose in recent time through a single allopolyploidization event, involving *C. eugenioides* or its ancestor as a mother. Rapid subsequent spreading then generated today's geographical range. Although large fractions of the cp genome have been screened, including the characterization of a number of microsatellites, no deviating cp haplotypes were encountered so far in *C. arabica*. This is uncommon compared to other plants and crops (Caron et al. 2000; Grivet and Petit 2002; Romero-Severson 2003; Dorken and Barrett 2004; Dobes et al. 2004; Molina_Cano et al. 2005).

The genetic variation of spontaneous and subspontaneous *C. arabica* from rainforests of Ethiopia is much greater than in the cultivated material (Mayer 1965; Narasimhaswamy 1968). Edjamo et al. (1996) described the three dominant coffee canopy types with corresponding seed and leave size and their agronomic and management requirements. The morphological characterization of Hararge coffee showed the existence of a wide range of phenotypically different types among

accessions with respect to majority of the characters studied such as growth habit, branching habit, leaf tip color and leaf shape (Kebede 2003). Furthermore, the analysis of morphological variability based on seedling parameters such as seedling height, leaf length, leaf width, stem diameter, leaf area, shoot fresh weight and root fresh weight, of eighty one accessions of Ethiopia coffee germplasm collected from all over the coffee growing area (both wild and garden coffee), confirm the presence of wide genetic variation among accessions which could be exploited in the genetic improvement of the crop through hybridization and selection (Seifu et al. 2004). The co-existence of the coffee rust fungus (*Hemileia vastatrix*) and *C. arabica* in the mountain rain forests of Ethiopia without serious damages being observed in the coffee trees are probably an indication of local adaptation in host-parasite interactions and also feature broader genetic diversity with resistant/ tolerant types in the wild populations of arabica (Meyer 1965; Van der Graaff 1981; Bergelson et al. 2001). Based on evaluating different morphological and agronomic traits Montagnon and Bouharmont (1996) observed that the sub-spontaneous genotypes from west of the Rift Valley appear genetically different from the commercial cultivars and individuals collected from Hararge and Sidamo. They are also divers and contain interesting agronomic features such as resistance to coffee berry disease and coffee leaf rust. Moreover, the recent analysis of wild coffee in Ethiopia revealed immense variability in morpho-physiological characteristics and hydraulic properties (leaf water potential, root and shoot hydraulic conductivity) among wild coffee trees and their progenies (Kufa 2006).

The study of genetic diversity based on RAPD marker demonstrated the existence of high diversity in the spontaneous and sub-spontaneous materials of Ethiopia with the existence of more bands (markers) with in this materials collected from SW (Anthony et al. 2001). Furthermore, microsatellite marker reveled low levels of diversity within currently existing *Coffea* cultivars in Latin America and ca. 55% of the alleles found in wild *C. arabica* from Ethiopia were not shared with cultivated *C. arabica* genotypes (Moncada and McCouch 2004). Recently, Aga (2005) observed the existence of moderate genetic diversity within and among forest coffee populations in Ethiopia. However, except Agas (2005) most of these studies were done on materials found out side Ethiopia with no clear indication of the origin and status of the materials, except the two group of spontaneous and subspontaneous. In addition, the samples did

not well considered those forests that are considered as *in situ* reserves of forest coffee in the country (Teketay et al. 1998; Dubale and Teketay 2000). However, recent analysis of the interregional genetic diversity of *C. arabica* with the materials collected from the entire native range in Ethiopia (Tefaye et al. (Submited (b)); Chapter 3), provide evidence for high genetic diversity within geographical regions. Such patterns of genetic diversity, as found in *C. arabica*, are normally attributed to high levels of gene flow that are connected to considerable outcrossing rates (insects as pollinators) and may further be influenced by modes of seed dispersal (animals such as monkeys and birds).

Current data provide evidence for a genetic diversity gradient in wild coffee with abiotic and geographic factors. The correlation of genetic diversity with latitude and average monthly precipitation could be explained with the presence of an additional flowers and behavior of pollinators of *C. arabica.* The presences of low numbers of flowers due to extended rain over long periods of the year in addition to the flowering peak, leads to pollination by insects that have to travel further because of encountering fewer flowers in the humid forests, and thus to increased gene flow. A historical reason such as diversity close to its centre of origin of *C. arabica* is higher due to the presence of refugia at certain latitudes, however, these needs to be tested with denser sampling. Moreover, it is possible to show that truly wild coffee growing in SW and SE forests (forest and semi-forest coffee) is genetically distinct from semi-domesticated *C. arabica* (Tesfaye et al. submitted (b); Chapter 3).

Inter simple sequence repeat markers are an important marker system interms of revealing patterns of genetic diversity in *C. arabica* (Masumbuko and Bryngelsson 2004; Ruas et al. 2003; Aga 2005; Tesfaye et al. submitted (b)). ISSRs are a quick and cost effective technique that is based on PCR amplification of large numbers of DNA fragments per reaction, representing multiple loci from across the genome (Zietkiewicz et al. 1994; Goodwin et al. 1997).

The aims of this chapter is particularly design to test the hypotheses put forward in the interregional analysis (Tesfaye et al submitted (b), Chapter 3) with higher number of denser sampling of wild coffee per plot (population) and per regions; high levels of diversity within populations but individuals from the same plot are often similar or closely related. Moreover, it is aimed to evaluate the molecular diversity and

its patterns of distribution within and among naturally existing wild populations of *C. arabica* in Berhane Kontir and Yayu/Geba Dogi.

4.2 Materials and Methods

4.2.1 Study population and plant materials

This study was carried out building upon the results of the interregional analysis of wild coffee in Ethiopia. Hence, two of the CoCE study regions, Berhane Kontir and Yayu/Geba Dogi, were considered for further in-depth analysis with denser sampling. Four and five plots with a size of 50m x 50m were selected for Berhane Kontir and Yayu forest, respectively. Furthermore, three samples were also included from Sorr river water fall forest which located close to Yayu/Geba Dogi. A clear definition of the population is important to avoid inappropriate extrapolation of results. In the absence of knowledge on effective population size, for practical reasons all the trees in 50m x 50m plot was regarded as a population. Twenty-five individuals were collected from each plot/population and in total that makes 125 and 100 individuals from Berhane Kontir and Yayu forest, respectively. However, samples which didn't amplify well excluded from the final analysis.

Coffee trees from two years to very old within the plots/population were chosen at random throughout each plot. Young leaves were collected in the field and dried directly using silica gel for later DNA extraction. Each of the individual trees sampled was labeled in the forest and passport data along with GPS coordinates were collected. The map of the two study regions and location of the populations/plots examined from Berhane Kontir and Yayu/Geba Dogi is shown in Figure 4.1 and 4.2 respectively. For practical reasons (better recording and communication with in CoCE project) hierarchical coding of regions, populations and individuals were used. The regions were coded with Roman numbers, while Arabic numeral used for populations and individual coffee trees.

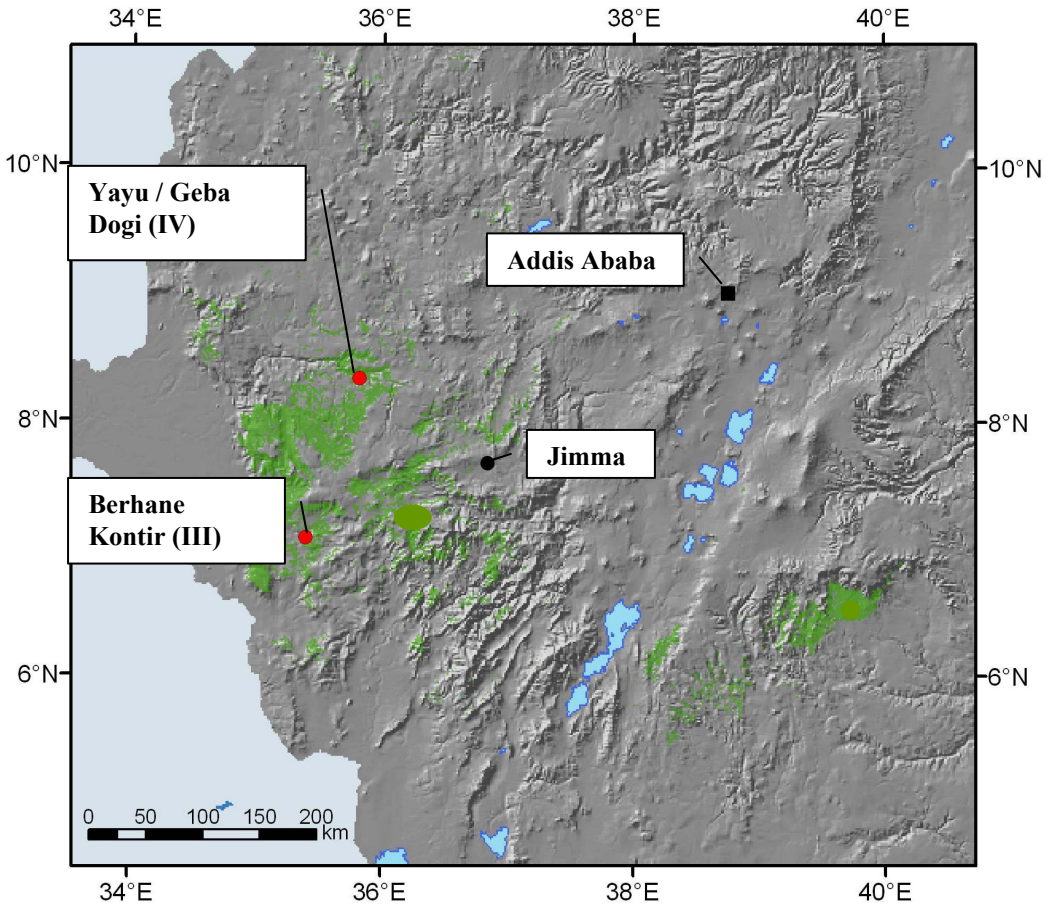


Figure 4.1 Map of SW Ethiopia with regions (Berhane Kontir and Yayu/Geba Dogi) considered for intraregional analysis.

4.2.2 DNA isolation

DNA was extracted in a similar manner to that described in Borsch et al. 2003 and Tesfaye et al. submitted (b) (see chapter 3), and the samples were further cleaned using QIAquick PCR purification kit (Qiagen GmbH, Hilden, Germany) according to the manufacture's instructions. The quality and concentration of DNA was assessed by gel electrophoresis using 0.9% agarose with known concentration of DNA marker. The concentrations of DNAs were then adjusted with dilution for ISSR-PCR. The adjusted genomic DNA dilutions were further tested on 0.9% agarose gels and visualized with ethidium bromide for comparable concentration.

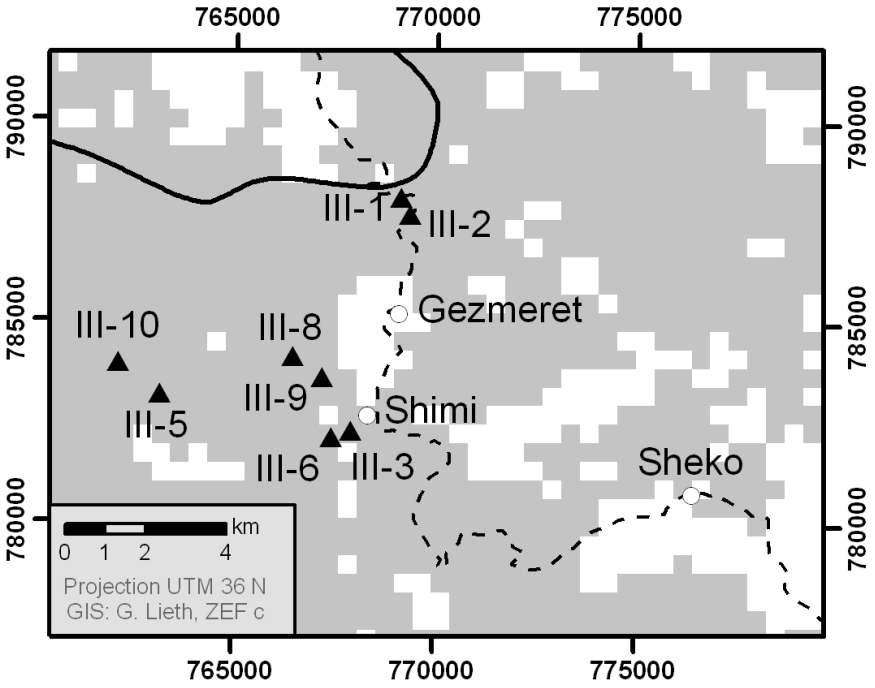


Figure 4.2 Sampled plots of Berhane Kontir forest for the Intraregional analysis. The map is drawn using GPS coordinates taken from the plot. The grey areas mark the position of forest with coffee. The schematic overview of the sampled individual coffee trees in 50m X 50m plots showed in Figure 3.2. Plots 1, 2 and 3 are shared among the CoCE subprojects, whereas plots 5, 6, 8, 9 and 10 are only included in the analysis of this project. Twenty five individuals collected per plot for entire analysis which is partitioned as interregional and intraregional analysis. Only six individuals included per plot for interregional analysis but all the twenty five individuals per plot used for intraregional analysis. Plot 4 and 7 were not included in this analysis. Doted line stands for road (Mezane Teferi to Tapi) and the continue line stands for river.

Table 4.1 Populations of wild *C. arabica* from Berhane Kontir and Yayu/Geba Dogi forest examined for inter simple sequence repeat (ISSR) variations. Most individuals were selected randomly as described in the methods from 50m X 50m plots.

Region	Plot codes	N_P	N_R	Location	Altitude (m a.s.l.)	Habitat
Berhane Kontir	III-1	19	107	07° 07' 23.3" N/35° 26' 11.4" E	1077	Semidisturbed
	III-3	19		07° 04' 13.4" N/35° 25' 29.0" E	1244	Semidisturbed
	III-8	24		07° 05' 13.5" N/35° 24' 44.3" E	1825	Undisturbed
	III-9	22		07° 04' 56.3" N/35° 25' 07.8" E	1560	Semidisturbed
	III-10	23		07° 05' 08.2" N/35° 22' 22.7" E	1104	Undisturbed
Yayu / Geba Dogi	III-1	18	81	08° 23' 99.4" N/35° 47' 71.8" E	1499	Semidisturbed
	III-2	23		08° 23' 71.1" N/35° 47' 72.8" E	1490	Semidisturbed
	III-4	19		08° 23' 48.4" N/35° 47' 88.2" E	1500	Undisturbed
	III-5	18		08° 22' 37.9" N/35° 48' 02.9" E	1388	Undisturbed
	Sorr Water Fall	3		08° 19' 14.8" N/35° 42' 20.8" E	1330	Semidisturbed
Total			188			

N_P= number of individual tree sampled per plot, N_R = Number of individual trees samples per region

4.2.3 ISSR-PCR amplification

ISSR-PCR involves PCR amplification of genomic DNA with a single primer composed of SSRs (Zietkiewicz et al. 1994; Wolfe and Liston 1998). Initial screening of the SSR primers set obtained from the University of British Colombia (Primer kit UBC 900) and primers used by Ruas et al. (2003) was done against 17 individuals of *C. arabica* and one individual of the diploid *C. eugenioides*. Ten primers were selected for the interregional analysis to evaluate the genetic diversity in spatial scale in Ethiopia (Tesfaye et al. submitted (b), Chapter 3). For further in-depth analysis of Berhane Kontir and Yayu populations, six dinucleotide (810, 812, 814, 818, 834, 844) and two tetranucleotide (CoIS001 and CoIS002) primers were initially chosen. The analysis of individuals from Yayu/Geba Dogi with five dinucleotide (810, 812, 814, 818, 844) and one tetranucleotide (CoIS001) primers were done by Tamiru Oligira, M.Sc. student from Addis Ababa University, within the frame of CoCE project. Additional two primers, di-(834) and tetranucleotide (CoIS002), were employed by the author and the results of all primers for Berhane Kontir and Yayu/Geba Dogi are summarized in this chapter.

Polymerase chain reactions (PCRs) were performed as described in Tesfaye et al (submitted (b), Chapter 3) with a Biometra Thermocycler. The samples were subjected to the following cycle: 4 min at 94^{o}C; 39 x 15s at 94°C, 1 min at the primer annealing optimal temperature (45^{o}C or 48^{o}C), 1.30 min at 72 ^{o}C, and at 72 ^{o}C for 5 min

final extension. The amplified products were electrophoresed in 1.7% agarose gels in TBE 1X buffer at 100V for 2h, and then stained and visualized with stained with ethidium bromide. The gels were photographed with Biodoc Analyze system from Biometra (035-300).

4.2.4 Data analysis

Only bands that could be unequivocally scored across all the sampled populations were used in this study. The amplified fragments on the gel images scored manually as '1' for presence and '0' for absence of equally sized DNA bands. Fragments with the same size were considered as identical and scoring was made from higher to lower molecular weight without considering the qualitative difference in band intensity. Binary characters matrix (1 and 0) was created from these data.

Genetic diversity was calculated by the percentage of polymorphic bands (P), which was calculated by dividing the number of polymorphic bands at population or regional levels by the total number of bands surveyed was done using POPGENE 1.31 (Yeh and Boyle 1997).The Shannon index was also calculated as $H = -\sum piLog2\ pi$, in which pi is the frequency of the presence or absence a given ISSR fragment, for each population or region (Lewontin 1972). Shannon's index is less biased since it does not rely on Hardy-Weinberg equilibrium and was thus used to calculate the total diversity (Hsp) as well as the mean intra-population/region diversity (Hpop). The amount of diversity among populations/regions was then calculated as (Hsp - Hpop/Hsp). The software Arlequin version 3.0b was employed to calculate AMOVA (Analysis of Molecular Variance) and estimate variance components of the ISSR data. AMOVA also computed to determine and partitioning the total variation to different hierarchical level and estimated the variation among individuals/within population, among population/within region (Excoffier et al. 2005).

Similarity matrices were constructed using Jaccard's coefficient (Jaccard 1908) based on the binary data of 0 and 1. The coefficient was calculated as: a/(a+b+c), where 'a' represents the presence of a given band in both individuals, 'b' represents the presence of the bands in the first individual, but not in the second, and 'c' represents the absence of the bands in the first individual and the presence in the second. The similarity matrixes and the trees were constructed using NTSYS-PC- version 2.1 (Rohlf

2000) and Free Tree 0.9.1.50 (Pavlicek et al. 1999) softwares. Both UPGMA (unweighted pair group method with arithmetic mean) and neighbor-joining (NJ) algorithms of tree construction were used (Sneath and Sokal 1973; Saitou and Nei 1987; Studier and Keppler 1988). Furthermore, principal coordinated analysis (PCO) was carried out based on Jaccard's coefficient. The softwares PAST version 1.18 was used to calculate Jaccard's coefficient and STATISTICA version 6.l to plot on the first three axes of the result (Hammer et al. 2001; Statistica Soft, Inc. 2001).

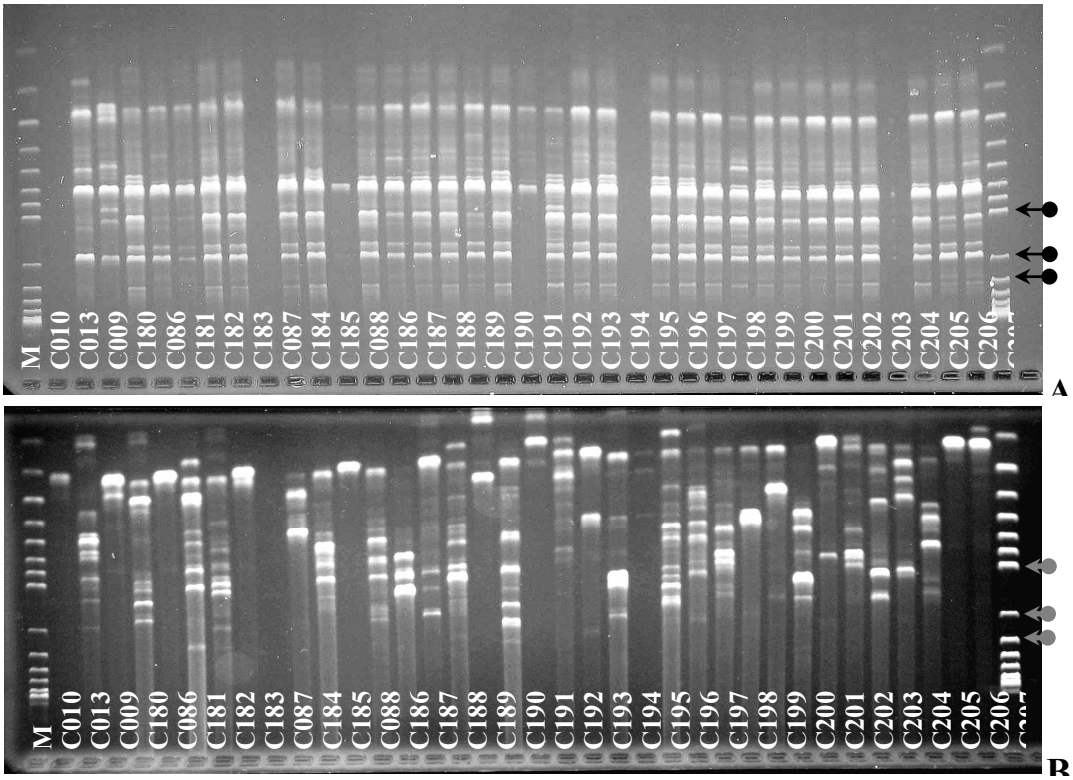


Figure 4.3 Ethidium bromide-stained gels of ISSR banding patterns generated for 36 individual coffee trees from Yayu (Geba Dogi). A is based on a dinucleotide [$(GA)_8T$; 810] and B on a tetranucleotide primer [$(GGTA)_4$, CoIS002]. Outside lanes (M) show an extended 100-bp ladder. The arrows on the right side of both gel pictures indicate fragment sizes (arrows correspond to 1000 bp, 1500 bp and 2000 bp from top to bottom).

4.3 Results

4.3.1 Genetic diversity

The selection of ISSR primers was based on the previous analysis and results of preliminary testing and screening using 18 *Coffea* samples. In addition, primers which gave good amplification and polymorphism for the interregional analysis were further promoted to be used for detailed analysis of wild coffee from Berhane Kontir and Yayu/Geba Dogi. The six selected primers generated a total of 63 well-amplified and scorable bands, ranging in size from 300 to 2500 bp among 188 individuals of wild coffee. The number of bands scored for each primer ranged between 7 for 814 to 13 for 810 and 812 (Table 4.2). In general, the dinucleotide primers demonstrated slight higher polymorphism for Berhane Kontir population (36.5%) than Yayu (34.9%) and the combined analysis with all data set showed 44.4% polymorphism for the two regions (Table 4.3).

Table 4.2 Sequences of primers used in this analysis and numbers of loci/bands (P) detected with this six primers and percent polymorphism (%P). * The tetranucleotide primers were included initially, however, finally exclude from the analysis since the level of polymorphism were higher and each individuals were unique. Further more, scoring of such gel could probably lead to scoring error.

Primer	**Sequences 5' to 3'**	**No. of scorable bands**	**Berhane Kontir**		**Yayu / Geba Dogi**	
			P	**%P**	**P**	**%P**
810	GAGAGAGAGAGAGAGAT	13	8	61.5	7	53.9
812	GAGAGAGAGAGAGAGAA	13	1	7.7	1	7.7
814	CTCTCTCTCTCTCTCTA	7	4	57.2	2	28.6
818	CACACACACACACACAG	10	4	40.0	3	30.0
834	AGAGAGAGAGAGAGAGYT	8	2	25.0	1	12.5
844	CTCTCTCTCTCTCTCTRC	12	4	33.3	8	66.7
*CoIS001	CCTACCTACCTACCTA	-				
*CoIS002	GGTAGGTAGGTAGGTA	-				
Total	Total loci found (Total polymorphism)	63	23	36.5	22	34.9

P = number of polymorphic bands, %P = percent polymorphism

The two tetranucleotide CoIS001 and CoIS002 used for this study were excluded from data analysis since the numbers of bands produced and level of polymorphism were extremely higher and also hardly possible to make any scoring. This reflects the high level of genetic polymorphism and also the high frequency and distribution of the terta repeat motives in *Coffea* genome. The preliminary assessment of the two primers showed that almost no individual had the same banding pattern. A

representative gel picture of the electrophoresis of the 810 dinucleotide $(GA)_8T$ and CoIS002 tetranucleotide $(GGTA)_4$ primers ISSR-PCR is shown in Figure 4.3 (A and B).

Table 4.3 Shannon's diversity index (H) and percentage of ISSR band polymorphism (P) of *C. arabica* in Berhane Kontir and Yayu. Subtotals are provided with mean values of populations for regions of wild coffee.

Region	**Plots/ populations**	**Polymorphism**		**Shannon's diversity Index (H)**
		P	**%P**	
Berhane Kontire	III-1	13	20.6	0.33
	III-3	10	15.9	0.30
	III-8	9	14.3	0.18
	III-9	16	25.4	0.35
	III-10	12	19.1	0.22
Over all Berhane Kontir		23	36.5	0.28*
Yayu	IV-1	15	23.8	0.19
	IV-2	16	25.4	0.30
	IV-4	6	9.5	0.17
	IV-5	10	15.0	0.15
Over all Yayu		22	34.9	0.20*
Over all Berhane Kontir and Yayu		28	44.4	0.24*

*P = number of polymorphic bands, %P = percent polymorphism, * Mean value of H of populations and regions*

The percent polymorphic bands for Berhane Kontir populations (plots) ranged from 14.3% to 25.4% for population -8 and 9 respectively. Similarly, Shannon's diversity index also showed higher diversity for population -9 (0.35) and the least divers was population -8 (0.18). In the case of Yayu, the higher percent polymorphic band was exhibited by population -2 (25.4%) and the least (9.5%) was by population -4. Besides, the Shannon's diversity index also showed higher diversity for population -2 (0.30) and lower diversity for population -5 (0.15).The semidisturbed population observed to show slightly higher diversity as compared to the undisturbed population with in respective regions of Berhane Kontir and Yayu (Table 4.3 and Table 4.1). The Shannon's diversity index for the region Berhane Kontir (0.28) showed higher average diversity as compared to that of Yayu (0.20). In addition, the Hsp of regions (where all the individuals with in regions used for calculating regional Shannon's index) also showed higher diversity for Berhane Kontir (0.41) than Yayu (0.30).

4.3.2 Partitioning genetic diversity and population structure

Shannon's diversity index and analysis of molecular variance (AMOVA) were used to partition the available genetic diversity. The result of locus by locus analysis of AMOVA was taken in this case since it is observed to be more accurate in the presence of missing data (Excoffier et al. 2005). Shannon's diversity index for each population (Hpop) and entire data (Hsp) were calculated for each one of the regions separately (Berhane Kontir and Yayu) (Table 4.4) and also all the populations together (Table 4.5). More genetic diversity, 67% and 66% were observed within population of Berhane Kontir and Yayu. However, a considerable amount of diversity also present among populations of Berhane Kontir (33%) and Yayu (34%). In general, the overall analysis of all data sets revealed that 59% of the variation is found within populations and the remaining 41% between populations (Table 4.5).

Table 4.4 Partitioning of the genetic variation into within and between populations of the wild coffee populations in Berhane Kontir and Yayu based on Shannon's information index.

Populations from	**Hpop**	**Hsp**	**Hpop/Hsp**	**1-Hpop/Hsp**
Berhane Kontir	0.28	0.41	0.67	0.33
Yayu / Geba Dogi	0.20	0.30	0.66	0.34

Hpop = mean genetic variation for the populations; Hsp = mean genetic variation for the entire data; Hpop/Hsp = proportion of genetic variations within wild coffee population; (Hsp-Hpop)/Hsp = proportion of genetic variations between wild coffee populations

The analysis of molecular variance also reveled that Berhane Kontir wild coffee accumulate much of its diversity within populations (63%) than among population (37%). Similarly, the populations in Yayu also accumulate slight higher percent variation for within population variation (53%) than among populations (47%). The analyses of molecular variance for over all data were done with and without grouping the populations with their respective region (Table 4.6). Very little variation (2%) accounted for among regions variation, while a significant amount of variation partitioned among population with in regions (45%) and within populations (53%). A similar result was obtained without structuring populations to its regions of origin, where slightly higher variations account for within population variation (53%) than among populations variations (47%).

Table 4.5 Partitioning of the genetic variation into within and between populations of the two wild coffee populations (Berhane Kontir and Yayu) of Ethiopia based on Shannon's information index.

Parameter	Mean
Hpop	0.24
Hsp	0.41
Hpop/Hsp	0.59
1-Hpop/Hsp	0.41

Hpop = mean genetic variation for the population; Hsp = mean genetic variation for the entire data; Hpop/Hsp = proportion of genetic variations within wild coffee population
(Hsp-Hpop)/Hsp = proportion of genetic variations between wild coffee population

Table 4.6 The Analysis of molecular variance (AMOVA) for two regions of wild *C. arabica.* The analysis were done with structuring the population with respective regions (the first three rows), with out structuring all the populations (fourth and fifth row) and separately for each of the regions (the last four rows)

Source of variation	Sum of Squares	Variance Components	Percent of Variation	Fixation Indices	P
Among Regions	4.52	0.008	2.0	0.02	0.00
Among Populations within regions	26.83	0.177	45.2	0.46	0.00
Within populations	36.37	0.207	52.8	0.47	0.00
Among all populations	31.35	0.181	46.7	0.47	0.00
Within populations	36.37	0.207	53.3		
Among populations of Berhane Kontir	23.83	0.259	37.4	0.37	0.00
Within Populations of Berhane Kontir	44.21	0.433	62.6		
Among populations of Yayu	85.60	1.466	47.0	0.47	0.00
Within Populations of Yayu	114.26	1.656	53.0		

4.3.3 Cluster analysis

The UPGMA and the neighbor-joining tree of all the individual samples from both Berhane Kontir and Yayu calculated based on Jaccard's coefficients are shown in Figures 4.4 - 4.10. The neighbor-joining trees recovered the majority of the groupings of UPGMA. The individual from the same plot/population Berhane Kontir were largely grouping together (Figures 4.6, 4.7). Furthermore, it is observed that the samples from undisturbed populations tend to show a stronger grouping according to their populations of origin, while the semidisturbed populations tend to be scattered and form groups with individuals from other population. The same pattern of clustering is also observed in the UPGMA tree of Yayu, where the individuals from the same population group together (Figure 4.8).

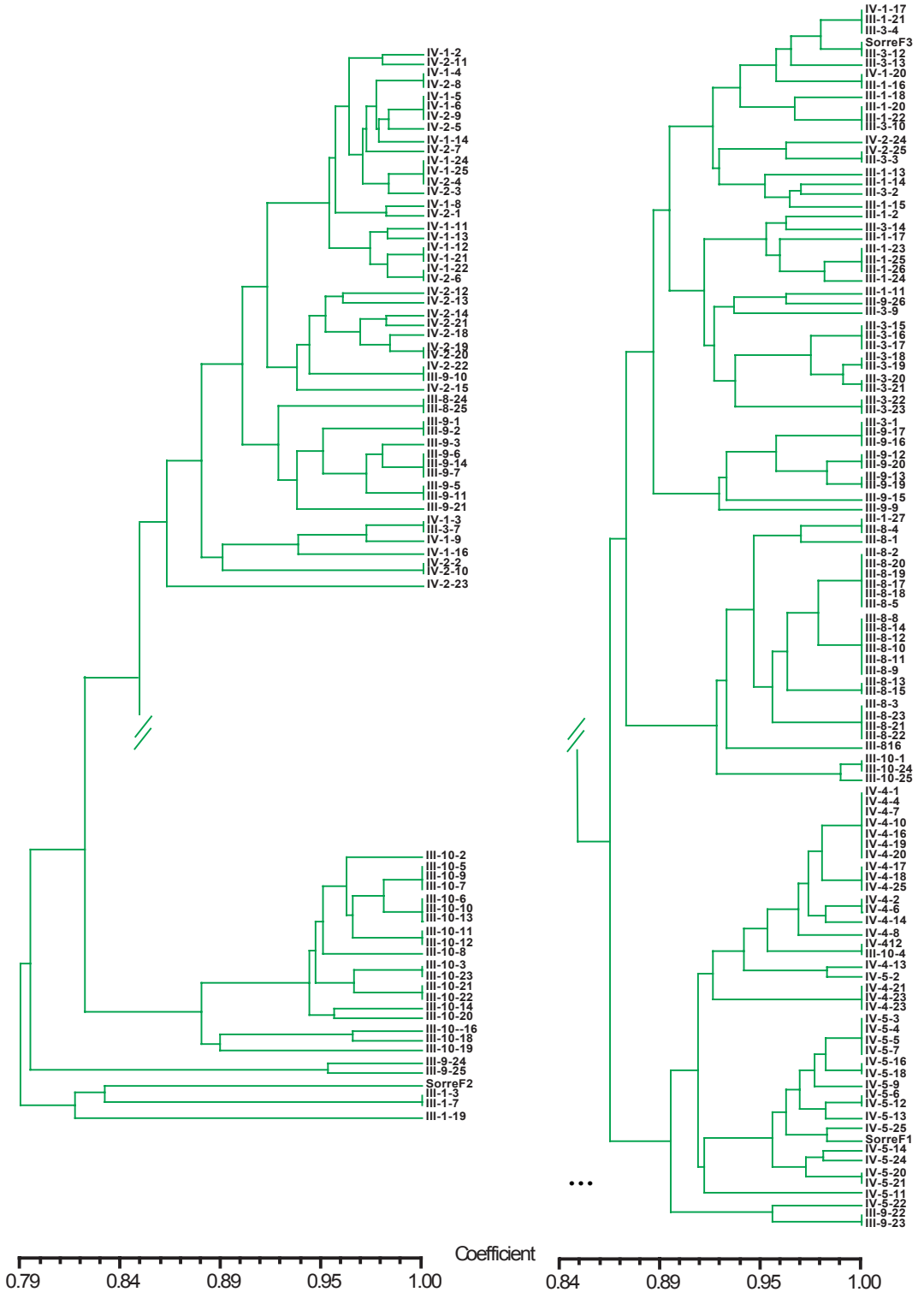


Figure 4.4 UPGMA analysis of complete intraregional dataset (188 individuals) based on six dinucleotide (810, 812, 814, 818, 834 and 844). The algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint. The tree splited in to two for better graphical display (//).

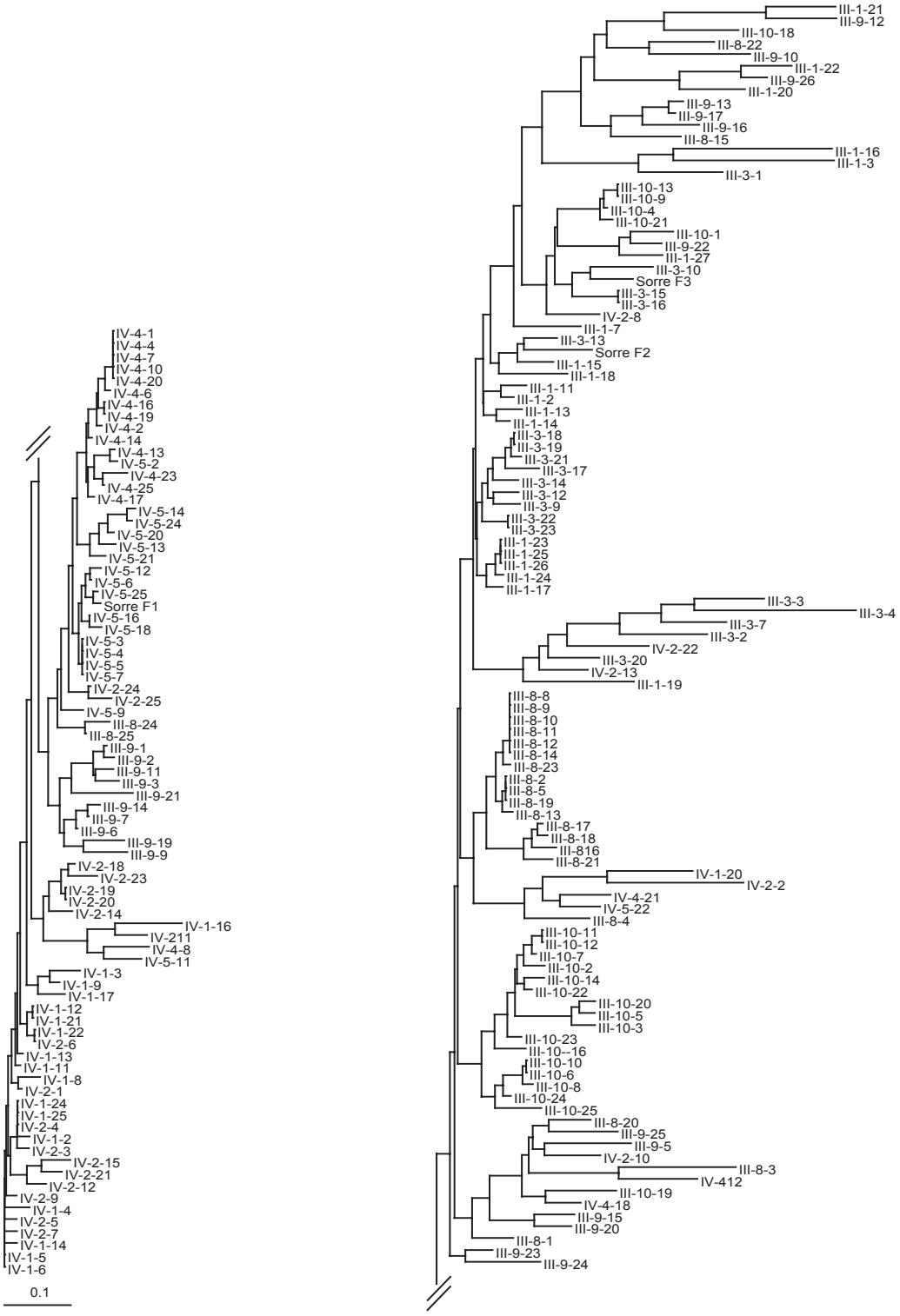


Figure 4.5 Neighbour joining analysis of complete intraregional dataset (188 individuals) based on six dinucleotide (810, 812, 814, 818, 834 and 844). The algorithm of neighbour joining is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint. The tree splited in to two for better graphical display (//).

Similarly, the UPGMA tree for all the samples (both from Berhane Kontir and Yayu) showed similar patterns of grouping, where all the majority of the individuals from the same population grouped together (Figure 4.4). However, individuals from semidisturbed plots of both Berhane Kontir and Yayu are observed to be intermixed.

Individual genotypes were constrained to their population; groups based on this constrained Jaccard`s coefficient were calculated. The dendrograms were generated by UPGMA cluster analysis using the sequential agglomerative hierarchical nested cluster analysis (SAHN) (Figure 4.10). The tree demonstrates differentiation among Bernahe Kontir and Yayu populations. The semidisturbed populations of Yayu (IV-1 and IV-2) were grouped together whereas the undisturbed populations (IV-4 and IV-5) form a separate group. Similarly, the populations from the semidisturbed plots (III-1 and III-3) and undisturbed plots (III-8 and III-10) of Berhane Kontir form their own group. However, the semidisturbed Berhane Kontir population III-9 grouped with Yayu populations (IV-1 and IV-2).

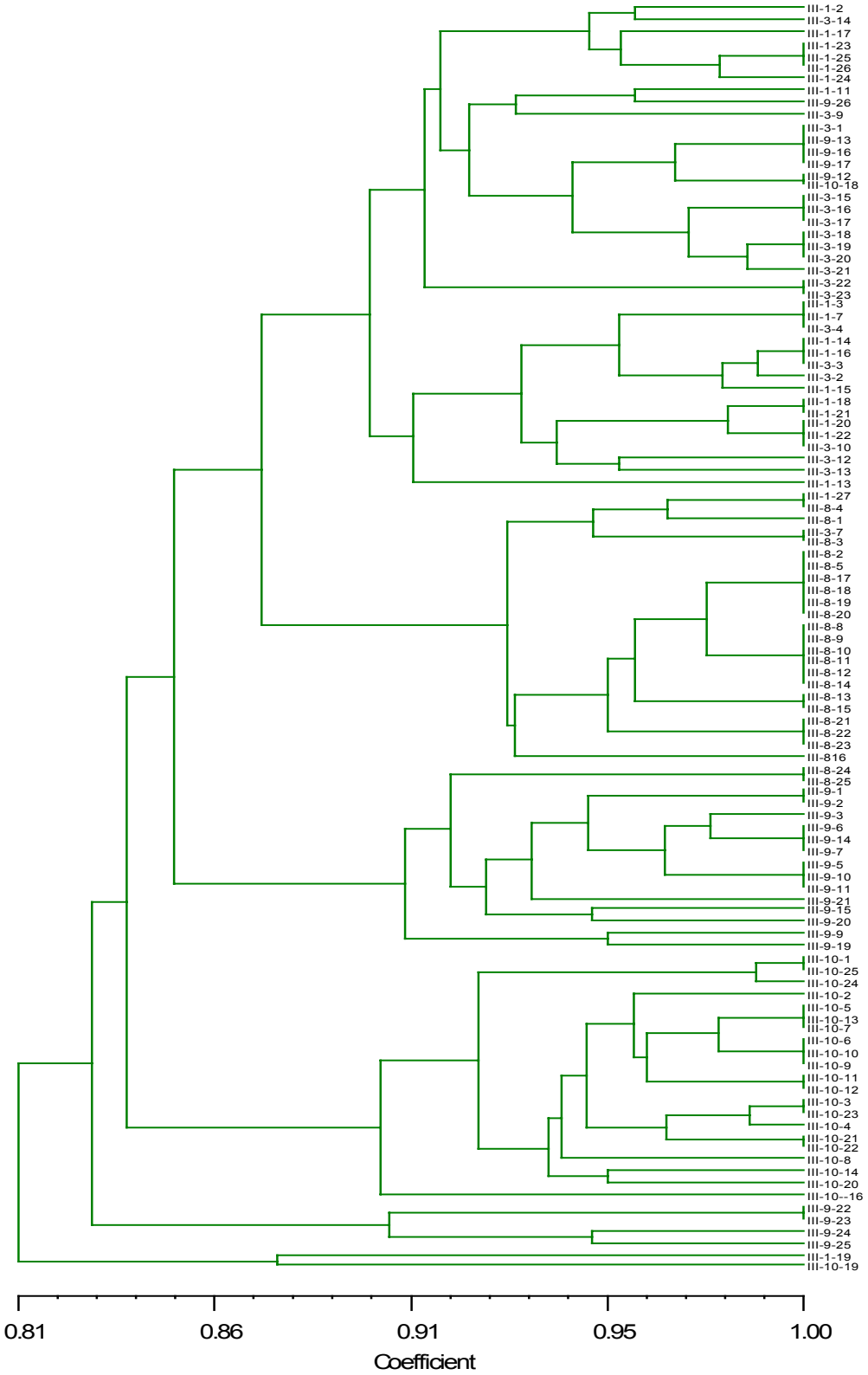


Figure 4.6 UPGMA analysis of one hundred seven 107 individuals from Berhane Kontir based on six dinucleotide (810, 812, 814, 818, 834 and 844). The algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

Figure 4.7 Neighbour joining analysis of one hindered seven individuals collected from Berhane Kontir forest based on six dinucleotides (810, 812, 814, 818, 834 and 844). The algorithm of neighbour joining is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

4.3.4 Principal coordinate analysis (PCO)

The plot of the first three principal coordinate analysis of Berhane Kontir with eigen-values ranging from 3.9 to 9.0 accounts for 15.8%, 10.2% and 6.8% of the total variance (32.8% cumulative; Figure 4.11). PCO revealed that population-10 was clearly separated from the rest, while the population -1 and -3 observed to occupy parallel position on the 3D spaces of the PCO. Furthermore, population III-8 also tends to form separate group from population -1 and -3, however, population -9 tends to spread between III-1, III-3 and III-8. This is also the case on the dendrograme (see Figure 4.6).

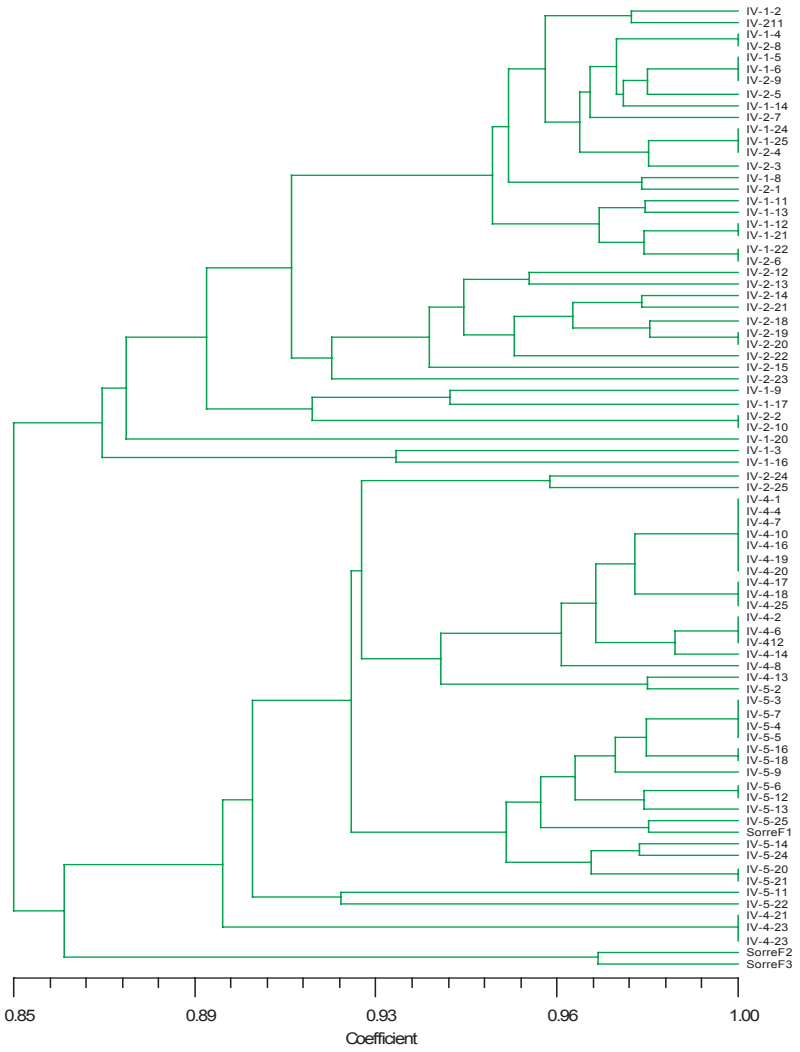


Figure 4.8 UPGMA analysis eighty one individuals from Yayu/Geba Dogi based on six dinucleotides (810, 812, 814, 818, 834 and 844). The algorithm is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint.

The first three axes with eigenvalues ranging from 1.9 to 6.0 describe 21.2%, 12.7% and 6.8% of the total variance (40.8% cumulative) of the PCO analysis of 81 individual coffee samples from Yayu (Figure 4.12). The samples from undisturbed populations (IV-4 and IV-5) observed to aside together on the PCO whereas individuals from the semidisturbed populations (IV-1 and IV-2) are intermixed with each other and tend to form separate groups.

The first three axes of the PCO analysis including all individuals both from Berhane Kontir and Yayu with eigenvalues ranging from 6.6 to 13.0 illustrate 14.1%, 7.8% and 7.1% of the total variance (29.0% cumulative; Figure 4.13). The individuals from the respective regions tend to form groups based on their regions of origin, in addition samples from Berhane Kontir are observed to be spread on PCO plot as compared to samples from Yayu. Generally the PCO clustering demonstrated similar patterns like that of UPGMA and NJ trees.

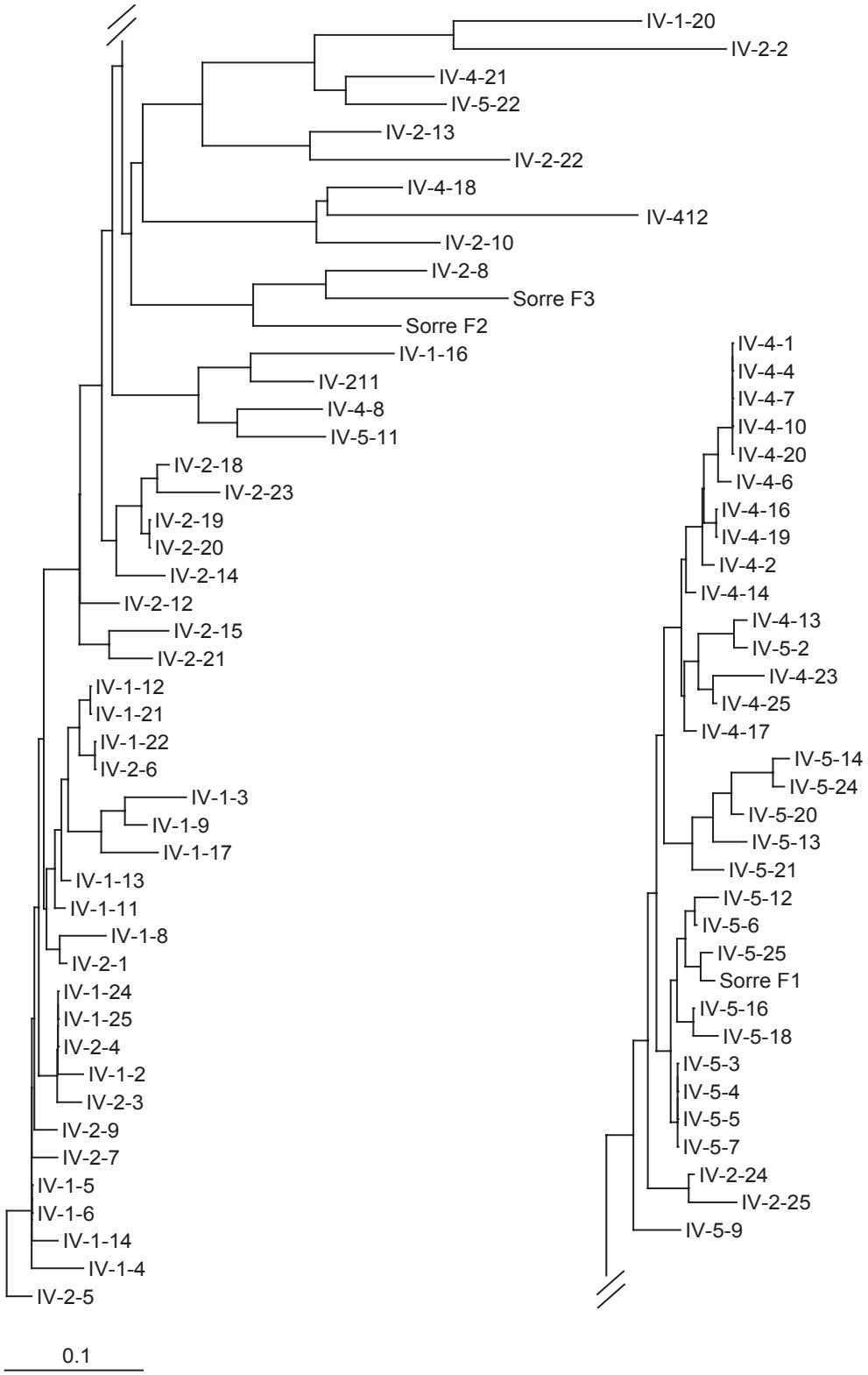


Figure 4.9 Neighbour joining analysis of eighty one individuals collected from Yayu/Geba Dogi forest based on six dinucleotides (810, 812, 814, 818, 834 and 844). The algorithm of neighbour joining is based on Jaccard's coefficients obtained after pairwise comparison of the presence-absence fingerprint. The tree splited in to two for better graphical display (//).

4.4 Discussion

4.4.1 Intraregional diversity

The results of the in-depth genetic diversity study of Berhane Kontir and Yayu confirms that there are high levels of genetic diversity in wild coffee within regions as well as with in populations in a region. This is also supported with the interregional diversity analysis of nine wild coffee regions (Tefaye et al. submitted (b), Chapter 3). The genetic diversity of a species or population is due to the combined effects of genealogical history, evolutionary processes and abiotic factors (Comes and Kadereit 1998).

The genetic diversity within populations ranges from 0.15 to 0.33 for Yayu population-5 and Berhane Kontir population-9. In addition, the second highest diversity is observed in Berhane Kontir population -1. Probably the presence of high gene flow by pollinators and seed disseminating agents in the forest of Berhane Kontir could contribute for the observed high diversity. The indigenous Mezenger community living in the forest of Berhane Kontir is known with hanging bee hives on big trees in the forest and also hunting honey, this could imply that the bee populations might be higher than other forest (Tesfaye 2006). Moreover, the distribution of wild coffee in Berhane Kontir also observed to be clumped pattern as compared to Yayu. Apart from seed dispersal agent, the observed coffee distribution patterns in Berhane Kontir are affected by a competition from other plants species in the forest (Senbeta 2006). Hence, this kind of patch distribution enhances the recruitment of new coffee seedling from neighboring population and increase seed flow among populations in Berhane Kontir.

The study of species diversity and ecology of afromontane rainforest of Berhane Kontir showed high beta and gamma diversity than the others forest, thus, within-species diversity might be mediated by ecosystem networks. Spatial and temporal heterogeneity in the environment may create diversifying selection that is thought to be a powerful mechanism of maintaining both species diversity and genetic diversity and so may generate correlations between them (Vellend and Geber 2005). Furthermore, based on theoretical and empirical studies, Vellend and Geber (2005) hypothesized; species diversity within communities and genetic diversity within populations are co-variable in space and time. Senbeta (2006) also observed the expansion of wild coffee towards the lowland part on the other side of the Berhane Kontir forest and this could also implies the existence of variable genotype that can

adapt to the extreme (lowland) environment and also role of seed dispersal agent in the area.

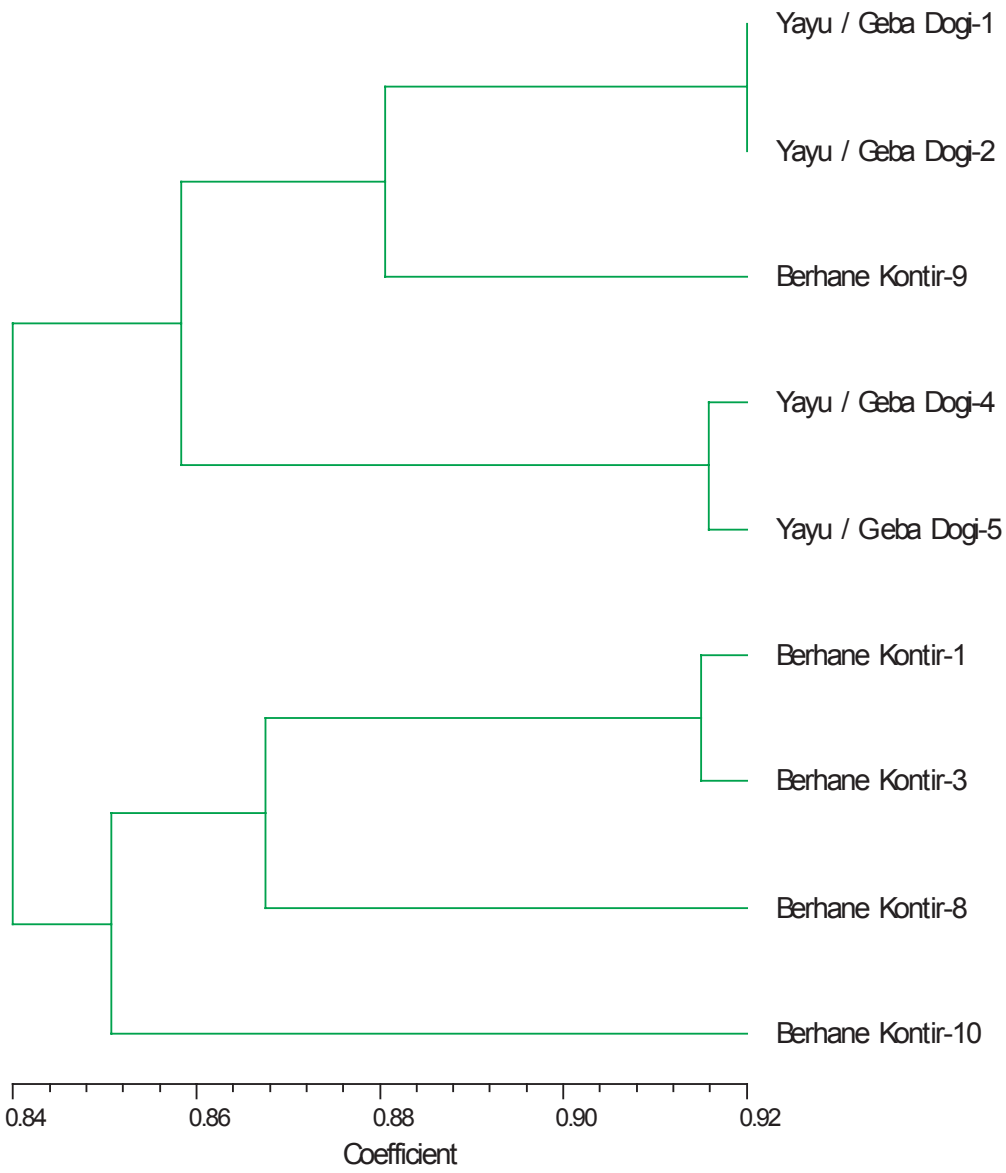


Figure 4.10 Individual genotypes were constrained to population grouping based on this constrained Jaccards coefficient were calculated. The dendrogram were generated by UPGMA cluster analysis using the sequential agglomerative hierarchical nested cluster analysis (SAHN). The individual genotypes were grouped into respective populations within their regions (Berhane Kontir and Yayu).

A striking pattern of genetic diversity observed with the two tetranucleotide primers CoIS001$(CCTA)_4$ and CoIS002 $(GGTA)_4$, where by almost all individuals were observed to have different banding pattern. Raus (2003) also observed 100% polymorphism for all tetranucleotide primers (including CoIS001 and CoIS002) used in the study of genetic relationships and parentage determination of *Coffea* and their interspecific hybrids crosses. This showed that frequency, distribution and degree of polymorphism of tetranucleotide repeat microsatellites in tetraploid genome of *C. arabica* is extremely higher and being almost unique in every individual plant. Moreover, the study made on evolutionarily diverse plant species with di-, tri-, tetra-, and penta-nucleotide microsatellite repeats as primer evidenced that tetranucleotide primers were most effective in amplifying the polymorphic patterns and were able to identify lines or varieties of plants (Gupta et al. 1994). Moreover, the study of microsatellite evolution in Salmonid fishes and human cell are also consistent with the findings of the above studies, suggesting that tetranucleotide repeats have higher mutation rates than dinucleotide repeats (Weber and Wong 1993; Ellegren 2000a, 2000b; Steinberg et al 2002). However, apart from motif/repeat type mutation rates can also depend on allele size, flanking sequence and chromosomal location (e.g., Goldstein and Clark 1995; Harr et al. 1998; Schlotterer et al. 1998; Goldstein and Schlotterer 1999). However, Lee et al. (1999) observed a contrasting trend of variability where by mutation rates of the tetra-nucleotide repeat [(GAAA)(17)] were much lower than those of the dinucleotide repeat [(CA)(17)] mouse cells.

The semi-disturbed wild coffee in respective regions exhibits slightly higher diversity than the one found in the undisturbed plots. The interregional analysis also showed the same patterns with some increment in diversity along with intermediate disturbance (Tesfaye et al submitted (b), Chapter 3). It is also observed that a slight disturbance of the forest where coffee occurs enhances the abundance of coffee since management of thinning out competing tree will give enough space for coffee seedlings to grow in early stage (Gole 2003; Senbeta 2006). An observation were also made that *Cedrus atlantica* (Pinaceae), a conifer native to the Rif and Atlas Mountains of North Africa, harbors similar levels of genetic variation within managed populations in comparison to undisturbed ones (Renau-Morata et al. 2005). However, a contrary phenomenon is observed in wild populations of the medicinal American ginseng, *Panax*

quinquefolius L. (Araliaceae) in which the average expected heterozygosity was significantly greater within protected populations as compared to populations in which harvesting was permitted (Cruse-Sanders and Hamrick 2004).

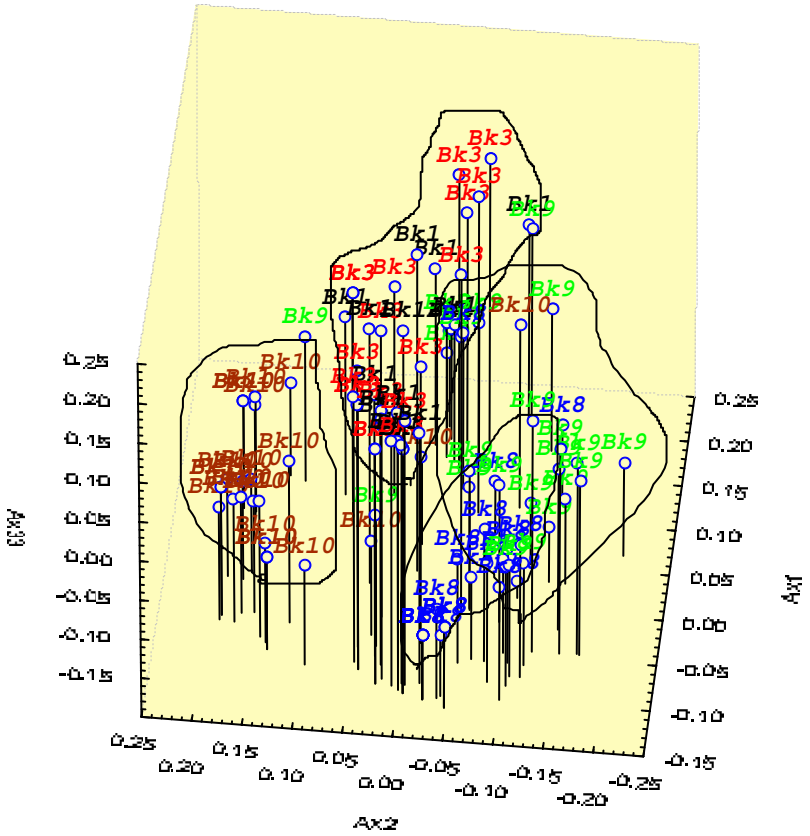


Figure 4.11 Three-dimensional representation of a principal coordinate analysis of the genetic relationships among 107 individuals of *C. arabica* from Berhane Kontir, inferred from a distance matrix using the Jaccards index. Individuals from the same population are labeled with the same population code (Bk1, Bk3, Bk8, Bk9 and Bk10).

4.4.2 Fine-scale spatial genetic structuring

The high Fst value (0.47) of AMOVA for the overall population analyses indicates that there is a clear genetic differentiation within populations of *C. arabica.* Furthermore, the genetic differentiation is confirmed by the UPGMA tree topology and the PCO analysis, in which the majority of individual samples from the same populations/plots grouped together. Nevertheless, adjacent plots show a slight tendency of mix up. Thus, there might be gene flow between plots or effective population size is probably bigger than the area of a plot of 50 m X 50 m size. The estimation of pollen donor of the tropical dry forest tree spices of Costa Rica (*Enterolobium cyclocarpum,* Fabaceae*)* with allozyme indicated of unidentified pollen donors which might be located out side of the study area. This could be because of long gene flow and showed individuals in the study area are part of a large network of reproductively active individuals (Apsit et al. 2001).

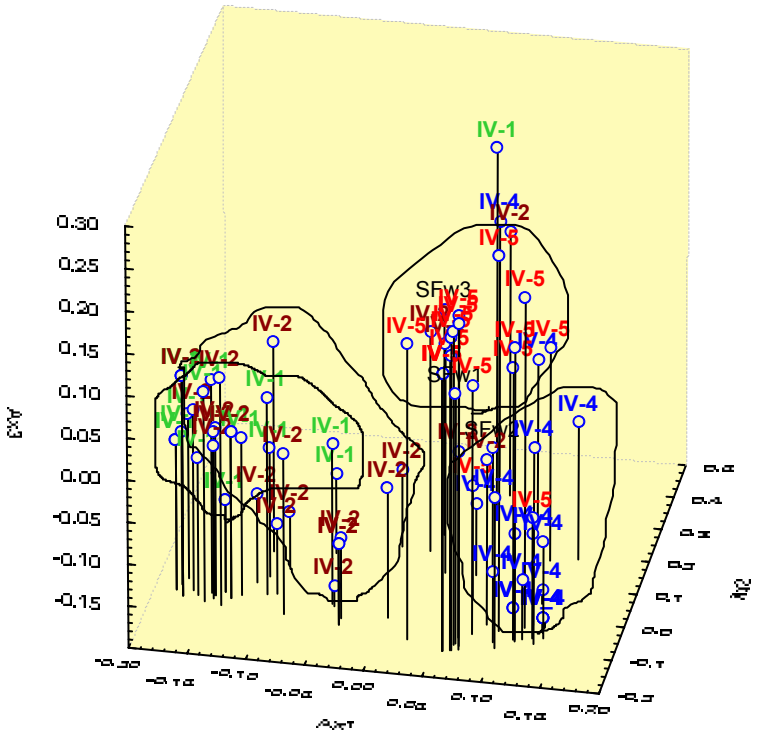


Figure 4.12 Three-dimensional representation of a principal coordinate analysis of the genetic relationships among 81 individuals of *C. arabica* from Yayu, inferred from a distance matrix using the Jaccards index. Individuals from the same population are labeled with the same population code (Y1, Y2, Y4, Y5 and Swf1-3).

The UPGMA clustering of constrained individuals to respective populations (individual from the same plot are considered as the same population and the data merged for pooled analysis of populations) shows tendency that populations found in spatial proximity group together (Figure 4.10). However, population-9 of Berhane Kontir formed a group with population-1 and -2 of Yayu. The ordination analysis species composition of Harrena (Bale), Bonga, Berhane Kontir and Yayu illustrated some overlap of Berhane Kontir with Yayu (Senbeta 2006). This implies the relative degree of similarity or difference in terms of floristic composition or other complex environmental variables; and indicates the history of these plots might be similar. Furthermore, plant communities in the forests of Berhane Kontir and Yayu are linked to Guineo-Congolian floral elements and categorized as "transitional forest" (Gole 2003; Senbeta 2006).

The fine spatial scale hierarchical patterning of genotypes is the predominant pattern observed when dense sampling is carried out within regions (Figure 4.4).This confirmed that genetic similarity is higher among neighboring individuals than among more distant individuals, which could be caused by local pedigree structures as a result of limited gene dispersal (Chung et al. 2004; Vekemans and Hardy 2004). However, since the patterning of genotypes for semidisturbed plots is not totally hierarchical, gene flow does play a role and obscures the hierarchical pattern for III-9, IV-1 and IV-2 plots. This gene flow might be a historical plant migration (e.g. climatic changes) and/or current through insect pollinators and seed dispersal agents (Gole 2003; Senbeta 2006; Tesfaye et al. submitted (b), Chapter 3).

The observed genetic structuring is stronger in Yayu (Fst = 0.47) than in Berhane Kontir (Fst = 0.37). This shows that there is a limited gene flow within Yayu forest as compared to Berhane Kontir. This further supported with the observation of Gole (2003) and Senbeta (2006), where wild coffee is distributed in patchy in Berhane Kontir forest but wide distribution of coffee were observed across the forest in Yayu. The patchiness could probably enhance the gene flow between different populations by leading pollinators to travel further to look for more flowers. In the study of the interregional diversity of *C. arabica* in Ethiopia, the genetic structure of wild population obscured due to the presence of gene flow between regions (Tesfaye et al. submitted (b); Chapter 3). Small scale genetic structuring which was caused by limited gene flow

also reported in *Centaurea corymbosa* (Asteraceae) (Hardy et al. 2004). However, despite free pollen movement across population a significant microsapatial genetic structure was observed in *Silene tatarica* (Caryophyllaceae), which was attributed merely to the isolation-by-distance and restricted seed dispersal (Tero et al. 2005).

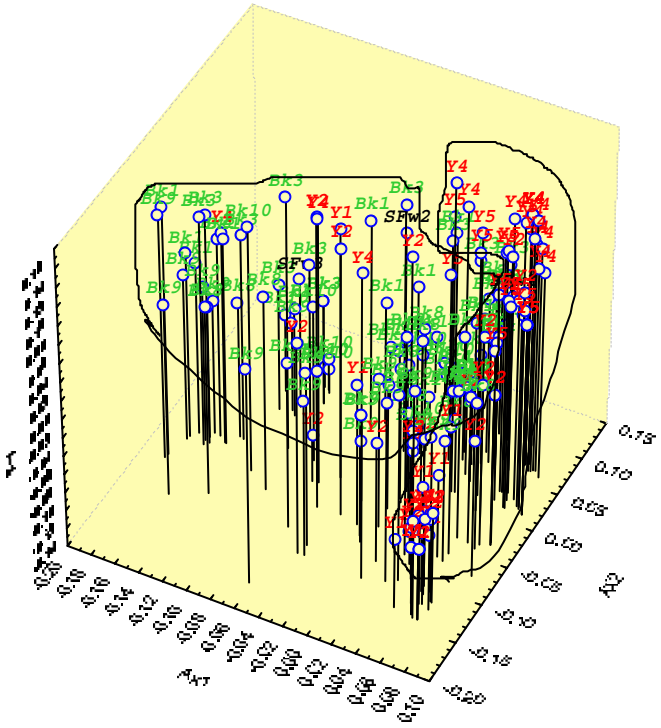


Figure 4.13 The over all analysis of the three-dimensional representation of a principal coordinate analysis of the genetic relationships of among 188 individuals of *C. arabica* from Berhane Kontir and Yayu, inferred from a distance matrix using the Jaccards index. Individuals from the same population are labeled with the same population code, Berhane Kontir (Bk1, Bk3, Bk8, Bk9 and Bk10) and Yayu (Y1, Y2, Y4, Y5 and Swf1-3).

Life form and reproductive biology are the main factor to determine the structure of populations (Hamrick and Godt 1996; Nybom and Bartish 2000). It is also observed that the selfing plant species tend to retain 51% the genetic variation among populations, while outcrossing species present on average 10% of the variation between populations (Hamrick and Godt 1989). Chase et al. (1995) found 12% of the genetic variation among populations of the outcrossing tropical tree species *Cordia alliodora* (Boraginaceae), while 88% of the variation accounts for within population variation.

Similarly, Rossetto *et al*. (1995), also using RAPDs, found that 13% of the genetic variation in the outcrossing plant species *Grevillea scapigera* (Proteaceae) was attributable to population differences and the rest 87% of the variability being attributable to single plant difference in population. Higher outcrossing rates detected in the mainly selfing wild soybean (*Glycine soja*) is supported by the observation of frequent visitors of the flowers by honey bee and carpenter bees (Ohara and Shimamoto 2002). The study of self-pollinated *Mycelis muralis* (Asteraceae) populations in Europe, using microsatellite markers, demonstrated different rates of selfing for populations and contain rather high levels of within-population diversity (50%) compared to other selfing species (Hamrick and Godt 1996, Chauvet 2004).

The interregional diversity study of wild coffee in Ethiopia also indicates 78% of the variation exists within regions and the rest 22% among regions (Tesfaye et al submitted (b), Chapter 3). Similarly, Aga et al. (2003) has got 65% and 80% variation within population and zone respectively. However, Aga et al. (2005) observed a contrasting pattern of genetic diversity, where by 20% of the variation account for within population and 80% accounts for among population variation. The intraregional study of wild coffee with more than 18-24 individuals per populations showed a little more than half (59%) of the total variation account for within population variation and substantial amount of diversity also observed among populations. This also further supported with AMOVA analysis. Too few samples per populations and inappropriate sampling scheme may underestimate the genetic diversity and also lowered the resolution at population-level relationships and genetic structuring (Wolfe and Randle 2001).

In general high local-scale genetic structuring observed in this analysis which could be attributed to limited gene flow which resulted in localized family structures. Spatial scale of genetic structuring was observed with limited number of samples of wild *C. arabica*, however, the pattern is some how obscured with gene flow (Aga et al. 2005 and Tesfaye et al. submitted (b); Chapter 3). It is generally accepted that cultivated *C. arabica* is self fertile and self-pollinating with a very low outcrossing rate (Carvalho et al. 1969; Charrier and Berthaud 1985; Purseglove 1968). Although it is believed that *C. arabica* is predominantly in-breeder, there has not been done any kind of evaluation of the breeding system of the wild population in Ethiopia except the observation made

by Meyer (1965) for the existence of considerable amount of out crossing rate (40%-60%). Insects as pollinators (African honey bees) and seed dispersal animals (such as monkeys and birds, and humans) can also play a significant role in the observed gene flow (Fichtl and Adi 1994; Gole 2003; Aga et al. 2005; Senbeta 2006).

4.5 Conclusions

The ISSR marker was able to show polymorphism within and among population/region of wild coffee. However, this marker system doesn't allow assessing heterozygosity and characterizing gene flow like the codominat markers microsatellite (Butcher et al. 1999; Dow and Ashley 1998). Using ISSR marker high levels of fine-scale genetic structure are observed within and between regions. Moderate gene flow was also observed among populations and regions of wild coffee.

The pattern of genetic diversity observed in wild arabica coffee in Ethiopia need further research using co-dominant marker microsatellite to assess heterozygosity, and measure and charachterise gene flow via paternity analysis. The nuclear microsatellite is a marker system that enables to estimate the rate of selfing in wild coffee population via estimation of inbreeding coefficient and also estimate the effective population size which is central for the development of appropriate conservation strategies. Furthermore, the pollination biology of wild *C. arabica* requires further research in order to confirm whether the wild populations in Ethiopia is dominantly in breeder or an outbreeder and also which factors are responsible for the observed gene flow. This study could also have a practical implication in establishing *in situ* conservation site. Several kinds of animals serving as vectors probably mediate the observed gene flow through both pollen and seed. This illustrates the need to conserve wild coffee with pollinators and animal vectors so as to maintain the natural means of mating and dispersal system in the forest.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

C. arabica is the only allotetraploid species in the genus *Coffea* and also the only species naturally occurring in Ethiopia. The SW and SE Afromontane rainforest of Ethiopia is the natural home of arabica coffee. Most of the montane rain forest in general and wild coffee in particular is under greatest threat of extinct because of deforestation and human settlement. This study was carried out to evaluate the genetic diversity of wild *C. arabica* and its patterns of distribution in Ethiopia. Moreover, it was designed within the frame work of CoCE project as a contribution for conservation and use planning.

5.2 Conclusions for conservation and sustainable use of wild *C. arabica* based on the results of this study

The analysis of genetic diversity of *C. arabica* population in Ethiopia with sequences from coding and non coding regions cpDNA and ISSR marker system gave fundamental insights on the evolution and diversification of *C. arabica.* Moreover, ISSR marker revealed the patterns of genetic diversity of wild *C. arabica* at different levels (interregional and intraregional). The main conclusions and recommendations drown from the result of this study are:-

5.2.1 Evidence for true wild coffee

The interregional genetic diversity study of wild populations of *C. arabica* and landraces from SW and SE gave a comprehensive picture of the diversity of coffee in different level of domestication in Ethiopia (Tesfaye submitted (b); chapter 3). The addition of landraces forms (farmers' variety) the basis to evaluate whether a population may be considered as being wild or not, resulted in clear separation of both groups. Hence, the truly wild coffee growing in forest and semi-forest coffee systems is genetically distinct from landraces (farmers' variety) and can be clearly recognized. Moreover, individuals from wild populations have higher level of genetic diversity as compared to landraces and also indicate multiple origins of lineages of landraces in different geographical areas. This is an important implication with respect to justifying the conservation efforts of wild *C. arabica* in Ethiopia.

5.2.2 Multi-site in situ conservation approach

Fingerprint (ISSR) markers show high levels of genetic diversity not only between but also within regions. Generally Berahne Kontir, Yayu/Geba Dogi, and Bonga all of them are found in SW part of the Rift Valley showed higher diversity as compared to the rest. However, differences in genotype distances among different regions were observed and some regions (Boginda, Bale/Harrena, Bench Maji, Mankira and Daphe) show closely related genotypes whereas individual plants of others (e.g., Bonga and Yayu) can be found almost throughout the whole range of genotypes and are extremely heterogeneous genotypes. A major conclusion for conservation is that a multi-site approach is needed, what is congruent to results on floristic diversity/biodiversity (Senbeta 2006). This approach is the best to capture the diversity and uniqueness of genotype in different forest. An optimal conservation strategy will, therefore, be able to use synergies from integrating *in situ* conservation of coffee genetic diversity and forest biodiversity.

5.2.3 Gene flow and genetic diversity

The genetic structuring observed in the interregional diversity analysis is obscured with the presence of gene flow (Tesfaye et al submitted (b); chapter 3). The in-depth analysis of intraregional wild coffee in Berhane Kontir and Yayu, with denser sample revealed more hierarchical structuring of genotypes based on neighborhood (Chapter 4). However, the fine-scale spatial pattern of grouping in the semidisturbed populations of the intraregional analysis was not totally hierarchical and one way or another obscured by gene flow. Moreover, both inter- and intraregional analysis showed higher genetic diversity within populations and region. Generally gene flow is likely to have a major role in generating and maintaining the genetic diversity of wild *C. arabica.* Hence, attention should be given to insect pollinators and also natural seed dispersal agents (birds, monkey and baboon) of wild coffee. Therefore, conservation effort should also consider insect pollinators and animal vectors responsible for the movement of pollen and seed.

The establishments of coffee farms should also consider some distance from the vicinity of wild coffee localities in general and *in situ* conservation zone in particular. This in particularly has a practical important to control the gene flow from near by coffee farms to the wild gene pool and protect genetic pollution of wild coffee.

5.2.4 Utilization in breeding program

The existing variation in the wild population and divergence among coffee population and regions observed in this study can be utilized through selection and hybridization program. The association between genetic divergence and heterosis was demonstrated in arabica coffee cultivars. The hybridization between the cultivars produced a heterosis of 25% that was probably the result of different allele combination in the hybrid (Silveira et al. 2003). Higher hybrid performance for yield (18%-60%) has been exhibited among the cross of elite breeding materials in Ethiopia (Amaha and Bellachew 1983). Moreover, recently investigation on semiwild collections of Ethio-Sudan showed 22-47% hybrid performance with a cross made with the traditional variety of Latin America (Bertrand et al. 2005). Hence, wild coffee in Ethiopia can be used for the improvement program of *C. arabica* through hybrid performance with out going to interspecific crossing with diploids which is time consuming and labours to get the intended trait in place with good fertility (Herrera et al. 2002a; Herrera et al 2002b; Yapo et al. 2003).

5.2.5 ISSR markers and its implications

The dinucliotides primers observed to show variability among different wild coffee regions and also able to differentiate wild from landraces. This could be used to certify wild (forest) coffee from Ethiopia as organic product in local and international market. Moreover, the tetranucleotide observed to be extremely variable as compared to dinucleotide. This implies that caution is needed interms of comparing results among markers with different frequency and distribution in the genome and also reliable measurements of genetic diversity require different marker systems. The tetranucleotide primers could give preliminary indication of the frequency and distribution of tetra simple sequence repeat in *C. arabica* genome. Hence, this could probable utilized for screening polymorphic tetra microsatellite loci for *C. arabica* since ISSR primers flanking the region of perfect or imperfect repeat of internal microsatellite. This is observed in the cauliflower (*Brassica oleracea* var. *botrytis* L.) where thirty-four of the 44 ISSRs had one or several perfect or imperfect internal microsatellites (Bornet et al. 2002). Therfore, tetranucliotide could be very useful for fingerprinting elite breeding lines and individual coffee trees.

5.2.6 Utility of chloroplast regions

The study of the chloroplast genome of *C. arabica* and *C. eugenioides* showed very low sequence divergence and also sister group relationship. This shows that *C. eugenioides* or its ancestor is maternal parent for *C. arabica* and also both genomes diverge recently from other *Coffea*. Furthermore, the absence of variability among *C. arabica* from a range of geographic origin pointed to ward recent origin of the species with single allopolypodization event.

The optimization of microstructural mutations showed that indels have high phylogenetic utility than microsatellite since microsatellite are observed to be more homoplasy due to frequent slippage in satellite loci. It is also observed that some regions like *atp*B-*rbc*L spacer has different signal than the other regions considered for cpDNA analysis. This study generally shows that the need to combine large sequence data set from rapidly evolving regions of chloroplast to avoid inconsistence of results from different data partitions.

5.3 Further conclusions and recommendations

1. There must be some mechanism of market incentive for semiwild coffee framers for getting premium price that compensate the low yield they are harvesting from semiwild. This is important to protect the introduction of improved coffee variety to Forest Coffee Ecosystem.

2. Certification of wild coffee for best market price should also be backed by low cost fingerprints.

3. The *in situ* conservation strategy should be supplemented with *in situ* gene bank, on-farm conservation and *ex situ* conservation of individuals from wild to save guard the wild gene pool from uncontrolled and unpredicted natural disaster.

5.4 Future research needs

1. Analysis of genetic diversity with nuclear co-dominant marker system (microsatellites) needs to be executed to find explanations for high levels of within region and population diversity. These will be fundamental to evaluate minimum areas

of *in situ* conservation sites, levels of inbreeding and also the effects of management activities. Moreover, microsatellite marker will also be important interms of revealing the type of gene flow (pollen /seed flow).

2. Multiple and/or single copy nuclear genes including microsatellite marker should also be sequenced to clarify the paternal donors of *C. arabica* genome. This is important to determine the relationships between arabica and other diploids and understand more the evolution of allotetraploid genome of *C. arabica.* This is in particular important for improvement of arabica through breeding. Moreover, mitochondrial genome might also be the other possibility to look for haplotype of *C. arabica* with maternally inherited genome.

3. The analysis of genetic diversity of co-existing forest tree species (potentially co-evolving with coffee) needs to assess. This is crucial for *in situ* conservation of not only wild coffee but also associated tree species in terms of evaluating and having a clear picture of the genetic diversity of the biodiversity in the Forest Coffee Ecosystem in Afromontane rainforest of Ethiopia.

4. Generally the existence of high genetic diversity of wild *C. arabica* is confirmed with this study. This can be utilized for the improvement of *C. arabica* and also landrace enhancement programme through hybridization. The existence of variation for agronomically important traits like cup quality, disease resistant and morpho-physiological traits were observed by different authors (Sylvain 1958; Ameha and Bellachew 1984; Van Der Graaf 1981; Wondimu 1987; Silvarolla et al. 2000; Adugna and Hindorf 2001; Anzueto et al. 2001; Silvarolla et al. 2004). Hence, work on marker assisted selection (MAS) should be one area of research in the future to supplement and shorten the time taking national improvement program of coffee in Ethiopia.

6 References

Abebe T. 2005. Diversity in homegarden agroforestry systems of Southern Ethiopia. Doctoral Thesis,Wageningen University, Wageningen, pp 153.

Adugna G. and Hindorf H. 2001. Recent investigation on coffee Tracheomycosis, *Gibberella xylarioides* (*Fusarium xylarioides*) in Ethiopia. 19th International Conference on Coffee Science, Trieste, May 14th-18th, 2001.

Aga E., Bryngelsson T., Bekele E. and Salomon B. 2003. Genetic diversity of forest arabica coffee (*Coffea arabica* L.) in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis. Hereditas 138: 36–46.

Aga E. 2005. Molecular genetic diversity study of forest coffee tree (*Coffee arabica* L.) populations in Ethiopia: Implications for conservation and breeding. Doctoral Thesis, Faculty of Landscape planning, Horticulture and Agricultural Science, Swedish University of Agricultural Sciences (SLU).

Asres T. 1996. Agroecological zones of southwest Ethiopia. Matreialien Zur Ostafrica-Forschung, 13.

Assefa K. 2003. Phenotypic and molecular diversity in the Ethiopian cereal, Tef [*Eragrostis tef* (Zucc.) Trotter]. Doctoral Dissertation, Department of Crop Science, SLU. Acta Universitatis Agriculturae Sueciae. Agraria vol. 426.

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

Amaha M. and Bellachew B. 1983. Hetrosis in cross of indigenous coffee selected for yield and resistance to Coffee Berry Disease II-First three Years. Ethiopian Journal of Agricultural Sciences 1: 13-21.

Amaha M. and Bellachew B. 1984. Resistance of the F1 to coffee berry disease in six parent diallel crosses in coffee. Proceedings of the first regional workshop on coffee berry disease. Association for the advancement of agricultural sciences in Africa, Addis Ababa, pp.167-177.

Amos A. and Balmford A. 2001.When does conservation genetics matter? Heredity 87: 257–265.

Andrea A.C.D., Lyons D.E., Haile M. and Butler E.A. 1999. Ethnoarchaeological approaches to the study of prehistoric agriculture in the Ethiopian highlands. In: Van der Veen (ed.), The Exploitation of Plant Resources in Ancient Africa. Kluwer Academic, Plenum Publishers, New York.

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.

Anzueto F., Bertrand B., Sarah J.L., Eskes A.B. and Decazy B. 2001. Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions. Euphytica 118:1–8.

Apsit V.J., Hamrick J.L. and Nason J.D. 2001. Breeding population size of a fragmented population of a Costa Rican dry forest tree species. J. Hered. 92: 415-420.

Arft A.M and Ranker T.A. 1998.Allopolyploid origin and population genetics of the rare orchids *Spiranthes diluvialis*. Am. J. Bot. 85:110-122.

Arroyo-García R., Lefort F., de Andrés M.T., Ibáñez J., Borrego J., Jouve N., Cabello F. and Martínez-Zapater J.M. 2002. Chloroplast microsatellite polymorphisms in *Vitis* species. Genome 45: 1142-1149.

Awoke T.C. 1997. The culture of coffee in Ethiopia. Agroforestry Today 9:19-21.

Ayana A. and Bekele E. 1998. Geographical patterns of morphological variation in sorghum (*Sorghum bicolor* L.) Moench) germplasm from Ethiopia and Eritrea: qualitative characters. Hereditas 129: 195-205.

Bakker F.T., Culham A., Pankhurst C.E. and Gibby M. 2000. Mitochondrial and chloroplast DNA-based phylogeny of *Pelargonium* (Geraniaceae). Am. J. Bot. 87:727-734.

Bau S.M.T., Mazzafera P. and Santoro L.G. 2001. Seed storage proteins in coffee. Rev. Bras. Fisiol. Veg.13: 33-40.

Baumel A., Ainouche M.L and Levasseur J.E. 2001. Molecular investigations in population of *Spartina anglica* C.E Hubbard (*Poaceae*) invading coastal Brittany (France). Mol. Ecol. 10 : 1689-1701.

Bawa K. 1998. Conservation of genetic resources in the Dipterocarpaceae. In: Appanah A. and Turnbull J.M. (eds.), A Review of Diprerocarps: Taxonomy, Ecology and Silviculture. Center for International Forestry Research, Bogor, Indonesia, pp. 45–55.

Beebe S., Toro O.C., González A.V., Chacón M.I. and Debouck D.G. 1997. Wild-weed-crop complexes of common bean (*Phaseolus vulgaris* L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet. Res. Crop. Evol. 44: 73–91.

Bekele E. 1983. Some measures of gene diversity analysis on landrace populations of Ethiopian barley. Hereditas 98:127-143.

Bekele E. 1984. Analysis of regional patterns of phenotypic diversity in the Ethiopian tetraploid and hexoploid wheats. Hereditas 100: 131-154.

Bekele E. and Lester R.N. 1981. Biochemical assessment of the relationship of *Eragrostis tef* (Zucc) Trotter with some wild *Eragrostis* species (Gramineae). Ann. Bot. 48:717-725.

Bellachew B. 1997. Arabica coffee breeding in Ethiopia: A review. 17th International Scientific Colloquium on Coffee, Nairobi, 20-25 Juillet 1997. Association Scientifique Internationale du Café (ASCI).

Bellachew B., Atero B. and Tefera F. 2000. Breeding for resistance to Coffee Berry Disease in arabica coffee: Progress since 1973. Proceedings of the workshop on control of Coffee Berry Disease in Ethiopia, Addis Ababa (Ghion Hotel), 13-15 August 1999, Addis Ababa, Ethiopia.

Bergelson J., Dwyer G. and Emerson J.J. 2001. Models and data on plant-enemy coevolution**.** Annu. Rev. Genet. 35: 469-499.

Berthaud J. and Charrier A. 1988. Genetic resources of *Coffea*. In: Clarke R.J. and Macrae R. (Eds.) Coffee. Vol. 4. Elsevier Applied Sciences, London, UK, pp. 1-42.

Berthou F., Trouslot P., Hamon S., Vedel F. and Quetier F. 1980. Analyse electrophorese du polmorphisme biochimique des cafeiers: varation enzymatique dans dix-huit population sauvages: *C.canephora*,*C.eugenioides* et *C.arabica*. Café Cacao The 24: 313-26.

Berthou F., Mathieu C. and Vedel F. 1983. Chloroplast and mitochondrial DNA variation as an indicator of phylogenetic relationships in the genous *Coffea* L.Theor. Appl. Genet. 65 :77-84.

Bertrand B., Etienne H., Cilas C., Charrier A. and Baradat P. 2005. *Coffea arabica* hybrid performance for yield, fertility and bean weight. Euphytica 141: 255–262.

Billard E., Daguin C., Pearson G., Ester Serrão E., Engel C. and Valero M. 2005. Genetic isolation between three closely related taxa: *Fucus vesiculosus, F. spiralis* and *F. ceranoides* (Phaophyceae). J. Phycol. 41**:** 900–905.

Birmeta G., Nybom H. and Bekele B. 2004. Distinction between wild and cultivated Enset (*Ensete ventricosum*) gene pools in Ethiopia using RAPD markers. Hereditas 140:139-148.

Bolibok H.B., Monika R-T., Aneta H. and Robert P. 2005. Efficiency of different PCR-based marker systems in assessing genetic diversity among winter rye (*Secale cereale* L.) inbred lines. Euphytica 146: 109-116.

Bornet B., Muller C., Paulus F. and Branchard M. 2002. Highly informative nature of inter simple sequence repeat (ISSR) sequences amplified using tri- and tetra-nucleotide primers from DNA of cauliflower (*Brassica oleracea* var. *botrytis* L.). Genome 45: 890-896.

Borsch T. 2000. Phylogeny and evolution of the genus *Nymphaea* (Nymphaeaceae). *Dissertation*, Rheinische Friedrich-Wilhelms-Universität Bonn.

Borsch T., Hilu K.W., Quandt D., Wilde V., Neinhuis C. and Barthlott W. 2003. Noncoding plastid *trn*T-*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. J. Evol. Biol 16: 558-576.

Botstein D., White R.L., Skolnick M. and Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331.

Bridson D.M. 1982. Studies in *Coffea* and *Psilanthus* (Rubiaceae subfam. Cinchonoideae) for part 2 of 'Flora of Tropical East Africa': Rubiaceae. Kew Bulletin 36**:** 817–859.

Bridson D.M. and Verdcout B. 1988. Rubiaceae, Part 2. In: Polhill R.M., (ed.), Flora of Tropical East Africa. Balkema, Rotterdam, The Netherlands, pp. 703-727.

Bruschi P., Vendramin G.G., Bussotti F. and Grossoni P. 2003. Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). Ann. Botany 91: 707-716.

Brohede J, Primmer C.R., Moller A. And Ellegren H. 2002. Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. Nucleic Acids Res. 30:1997-2003.

Bryan G.J., McNicoll J., Ramsay G., Meyer R.C. and De Jong W.S. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor. Appl. Genet. 99: 859-867.

Butcher P.A., Glaubitz J.C. and Moran G.F. 1999. Applications for microsatellite markers in the domestication and conservation of forest trees. In: Forest Genetic Resources No. 27. Food and Agriculture Organization of the United Nations, FAO.

Camacho F.J. and Liston A. 2001. Population structure and genetic diversity of *Botrychium pumicola* (Ophioglossaceae) based on Inter Simple Sequence Repeats (ISSR). Am. J. Bot. 88:1065-1070.

Cannell M.G.R. 1983. Coffee. Biologist 30:257–263.

Caron H., Dumas S., Marque G., Messier C., Bandou E., Petit R. J. and Kremer A. 2000. Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. Mol. Ecol. 9: 1089–1098.

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: Ferwerda F.P. and Wit F. (eds.), Outlines of perennial crop breeding in the tropics. Wageningen, The Netherlands, pp. 189–241.

Casa A.M., Mitchell S.E., Hamblin M.T., Sun H., Bowers J. E., Paterson A. H., Aquadro C.F. and Kresovich S. 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor. Appl. Genet.111: 23-30.

Chaparro A.P., Cristancho M.A., Cortina H.A. and Gaitan A.L. 2004. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. Genet. Resour. Crop. Evol. 51: 291–297.

Charrier A. and Berthaud J. 1985. Botanical classification of coffee. In: Clifford M.N. and Wilson K.C. (eds.), Coffee Botany, Biochemistry and Production of Beans and Beverage, Croom Helm, London, pp. 13-47.

Chase M.R., Boshier D.H. and Bawa K.S. 1995. Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree: Genetic variation in natural populations. Am. J. Bot. 82: 468–475.

Chauvet S., Van Der Velde M., Imbert E., Guillemin M.L., Mayol M., Riba M., Smulders M.J.M., Vosman B., Ericson L., Bijlsma R. and Giles B.E. 2004. Past and current gene flow in the selfing, wind-dispersed Species *Mycelis muralis* in western Europe. Mol. Ecol. 13**:**1391–1407.

Chung M.Y., Nason J.D. and Chung M.G. 2004. Spatial genetic structure in populations of the terrestrial orchid *Cephalanthera longibracteata* (Orchidaceae). Am. J. Bot. 91:52-57.

Church S.A. and Taylor D.R. 2005. Speciation and hybridization among *houstonia* (Rubiaceae) species: the influence of polyploidy on reticulate evolution. Am. J. Bot. 92: 1372-1380.

Comes H.P. and Kadereit J.W. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Sci. 3:432–438.

Conner P.J. and Wood B.W. 2001. Identification of pecan cultivars and their genetic relatedness as determined by Randomly Amplified Polymorphic DNA analysis. Journal of the American Society for Horticultural Science 126: 474-480.

Cook L.M., Soltis P.S., Brunsfeld S.L. and Soltis D.E. 1998. Multiple independent formations of *Tragopagon* tetraploids (Asteraceae): evidence from RAPD marker. Mol. Ecol. 7:1293-1302.

Coste R. 1992. Coffee; The plant and the product. CTA Waginingen, The Netherlands.

Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S., Charrier A. and Lashermes P. 1998. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. Mol. Phylogenet. Evol. 9:109-117.

Cruse-Sanders J.M. and Hamrick J.L. 2004. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax Quinquefolius* L. (Araliaceae). Am. J. Bot. 91: 540–548.

Cruzan M.B. 1998. Genetic markers in plant evolutionary ecology. Ecology 79:400-412.

Dane F. and Lang P. 2004. Sequence variation at cpDNA regions of watermelon and related wild species: implications for the evolution of *Citrullus* haplotypes. Am. J. Bot. 91: 1922-1929.

Davis A.P. and Mvungi E.F. 2004. Two new and endangered species of *Coffea* (Rubiaceae) from the Eastern Arc Mountains (Tanzania) and notes on associated conservation issues. Botanical Journal of the Linnean Society 146: 237–245.

Dawson I.K. and Powell W. 1999. Genetic variation in the Afromontane tree *Prunus africana*, an endangered medicinal species. Mol. Ecol. 8: 151-156.

Debener T., Salamini E. and Gebhardt C. 1990. Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment length polymorphisms (RFLPs). Theor. Appl. Genet. 79:360-368.

Deguilloux M.F., Pemonge M.H. and Petit R.J. 2004. Use of chloroplast microsatellites to differentiate oak populations. Ann. For. Sci. 61: 825-830.

Demissie A. 1996. Morphological molecular markers diversity in Ethiopian landrace barleys implications to *in situ* and *ex situ* conservation of landrace materials. Ph.D Thesis, Department of Horticulture and Crop Science, Agricultural University of Norway, As, Norway.

De Queiroz A., Donoghue M.J. and Kim J. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.

Deshpande A.U., Apte G.S., Bahulikar R.A., Lagu M.D., Kulkarni B.G., Suresh H.S., Singh N.P., Rao M.K.V., Gupta V.S., Pant A. and Ranjekar P.K. 2001. Genetic diversity across natural populations of three montane plant species from the Western Ghats, India revealed by intersimple sequence repeats. Mol. Ecol. 10: 2397-2408.

Dick C.W. 2001 Genetic rescue of remnant tropical trees by an alien pollinator. Proc. R. Soc. Lond. B 268: 2391-2396.

Dobes C.H., Mitchell-Olds T. and Koch M.A. 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A.× divaricarpa*, and *A. holboellii*(Brassicaceae). Mol. Ecol. 13: 349–37.

Dorken M.E. and Barrett S.C.H. 2004. Chloroplast haplotype variation among monoecious and dioecious populations of *Sagittaria latifolia* (Alismataceae) in eastern North America. Mol. Ecol. 13:2699-707.

Dow B.D. and Ashle M.V. 1998. High levels of gene flow in Bur Oak revealed by paternity analysis using microsatellites. Journal of Heredity 89:62-70.

Dubale P. and Teketay D. 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, 13-15 August 1999, Ethiopian Agricultural Organization (EARO) Addis Ababa, Ethiopia, pp 125-135.

Dwivedi S.L., Gurtu S., Chandra S., Yuejin W. and Nigam S.N. 2001. Assessment of genetic diversity among selected groundnut germplasm. I: RAPD analysis. Plant Breeding 120:345-349.

Echt C.S, Deverno L.L, Anzidei M. and Vendramin G.G. 1998. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Mol. Ecol. 7: 307-316.

Edjamo Y., Shimber T., Yilma A., Kufa T., Netsre A., Negewo T., Abachebsa M.N. and Bogale B. 1996. Advances in coffee agronomy research in Ethiopia. In: Improving Coffee Management Systems in Africa. Proceedings of IACO

workshop, Kampla,Uganda, 4-6 September 1995. The African Crop Science Society, pp 40-55.

Ehrendorfer F., Manen J-F. and Natali A. 1994. CpDNA intergene sequences corroborate restriction site data for reconstructing Rubiaceae phylogeny. Pl. Syst. Evol. 190: 245-248.

Ellegren H. 2000a. Heterogeneous mutation processes in human microsatellite DNA sequences. Nat. Genet. 24:400-402.

Ellegren H. 2000b. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. 16: 551-558.

Engels J.M.M. and Hawkes J.G. 1991. The Ethiopian gene center and its genetic diversity. In: Engels J.M.M., Hawkes J.G. and Worede M. (eds.), Plant Genetic Resources of Ethiopia. Cambridge University press, Cambridge, UK, pp.23-41.

Epperson B.K. and Li T.Q. 1996. Spatial genetic structure within populations using spatial autocorrelations statistics. Proc. Natl. Acad. Sci. USA 93: 10528–10532.

Escudero A., Iriondo J.M. and Torres M.E. 2003. Spatial analysis of genetic diversity as a tool for plant conservation. Biol. Conserv. 113:351-365.

Estoup A. and Cornuet J-M. 1999. Microsatellite evolution: inferences from population data. In: Goldstein D.B. and Schlötterer C. (eds.), Microsatellites: evolution and applications. Oxford University Press, Oxford, England, pp. 49–65.

Excoffier L., Laval G. and Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Faith D.P. 1994. Genetic diversity and taxonomic priorities for conservation. Biol. Conserv. 68:69–74.

Ferwerda F.P. 1976. Coffees *Coffea* spp. (Rubiaceae). In: Simmonds N.W. (ed.), Evolution of Crop Plants, Longman, London, pp. 257–260.

Fichtl R. and Adi A. 1994. Honeybee flora of Ethiopia. DED, Margraf Verlag.

Fitter R. and Kaplinsky R. 2001. "Who gains from product rents as the coffee market becomes more differentiated? A value chain analysis". IDS Bulletin Special Issue on the value of value chains, Vol. 32, No.3; pp. 69-82.

Free J.B. 1993. Insect pollination of crops. London: Academic Press.

Friis I. 1979. The wild populations of *Coffea arabica* L., and cultivated coffee. In: Kunkel (G.) Taxonomic Aspects of African Economic Botany. Proceedings of the IX plenary meeting of AETFAT. Las Palmas de Gran Canaria, 18-23 March, 1978.

Gebre-Egziabher T. 1990. The importance of Ethiopian forests in the conservation of arabica coffee gene-pools. Mitt. Inst. Allg. Bot. Hamburg. Proceedings of the 12 planary meeting of AETFAT, Hamburg.

Geleta M., Bryngelsson T., Bekele E. and Dagne K. 2006. Genetic diversity of *Guizotia abyssinica* (L.F.) Cass. (Asteraceae) from Ethiopia as revealed by random amplified polymorphic DNA (RAPD). Genet. Res. Crop Evol. DOI10.1007/s10722-006-0018-0.

Gengler K.M. and Crawford D.J. 2000.Genetic diversities of four little-known species of *Malesherbia* (Malesherbiaceae) endemic to the arid inter-Andean valleys of Peru. Brittonia 52: 303-310.

Gepts P. 2004. Crop domestication as a long-term selection experiment. In: Janick J. (ed.), Plant Breeding Reviews, Volume 24, part 2, John Wiley & Sons, Inc.

Getahun A. and Krikorian A.D. 1973. Chat: Coffee's rival from Harar, Ethiopia. Botany, cultivation and use. Econ.Bot. 27:357-389.

Goldstein D.B. and. Clark A.G. 1995. Microsatellite variation in North American populations of *Drosophila melanogaster.* Nucleic Acids Res. 23:3882-3886.

Goldstein D.B. and Schlotterer C. 1999. Microsatellites: evolution and applications. Oxford University Press, Oxford.

Gole T.W., Teketay D., Denich M., Borsch T. 2001. Diversity of traditional coffee production systems in Ethiopia and its contribution for the conservation of coffee genetic diversity. Deutscher Tropentagorganized by the Universities of Berlin, Bonn, Göttingen, Hohenheim,GhK Kassel-Witzenhausen and the Arbeitsgemei-nschaft Tropische und Subtropische Agrarforschung (ATSAF) e.V. Deutscher Tropentag – Bonn , 9-11 October 2001.

Gole T.W. and Teketay D. 2001. The forest coffee ecosystems: ongoing crisis, problems and opportunities for coffee gene conservation and sustainable utilization. In: Imperative Problems Associated with Forestry in Ethiopia, Biological society of Ethiopia. Addis Ababa, pp. 131-142.

Gole T.W., Denich M., Teketay D. and Vlek P.L.G. 2002. Human impacts on *Coffea arabica* genetic pool in Ethiopia and the need for its *in situ* conservation. In: Engels J., Rao V.R., Brown A.H.D. and Jackson M. (eds.), Managing Plant Genetic Diversity, CAB International / IPGRI, pp. 237-247.

Gole T.W. 2003. Vegetation of the Yayu forest in SW Ethiopia: impacts of human use and implications for *in situ* conservation of wild *Coffea arabica* L. populations. Ecoloyg and Development Series, No.10, Zentrum fur Entwicklungsforschung, Center for Development Research, University of Bonn.

Gole T.W., Borsch T., Denich M. and Teketay D. (submitted). Floristic composition and environmental factors characterizing coffee forests in southwest Ethiopia. Biol.Cons. (submitted).

Goodwin I.D., Aitken E.A.B. and Smith L.W. 1997. Application of inter simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis 18: 1524-1528.

Graham S.W., Reeves P.A., Burns A.C.E. and Olmstead R.G. 2000. Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int. J. Plant Sci. 161:S83-S96.

Grant V. 1981. Plant Speciation. Columbia Univ. Press, New York, 2nd Ed.

Grivet D. and Petit R.J. 2002. Phylogeography of the common ivy (*Hedera* sp.) in Europe: genetic differentiation through space and time. Mol. Ecol. 11: 1351–1362.

Grundy S. 2005. It's not just about the money: An evaluation of the impact of fair trade on improving the livelihoods of smallholder coffee farmers in Oromia Region, Ethiopia. Dissertation submitted in partial fulfillment of the requirements of the Masters of Science degree in poverty reduction and development management of the University of Birmingham.

Gupta M., Chyi Y-S., Romero-Severson J. and Owen J.L. 1994. Amplification of DNA markers from evolutionary diverse genomes using single primers of single-sequence repeats. Theor. Appl. Genet. 89: 998-1006.

Hallam D. 2003. Falling commodity prices and industry responses: Some lessons from the international coffee crisis. In: Commodity Market Review 2003-2004.

Commodities and trade division Food and Agriculture Organization of the United Nations, Rome, Italy.

Hammer Ø., Harper D.A.T. and Ryan P.D. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4:9, http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Hamrick J.L. and Godt M.J.W. 1990. Allozyme diversity in plant species. In: Brown A.H.D., Clegg M.T., Kahler A.L. and Weir B.S. (eds.), Plant Population Genetics, Breeding, and Genetic Resources. Sinauer, Sunderland, MA, pp. 43–63.

Hamrick J.L. and Godt M.J.W. 1996. Effects of life history traits on genetic diversity in plant species. Phil. Trans. Roy. Soc. L. B 351: 1291–1298.

Hardy O.J., Gonzalez-Martinez S.C., Freville H., Boquien G., Mignot A., Colas B. and Olivier I. 2004. Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae) I. Pattern of pollen dispersal. J. Evol. Biol. 17:795-806.

Harr B., Zangerl B., Brem G. and Schlotterer C. 1998. Conservation of locus specific microsatellite variability across species: a comparison of two *Drosophila* sibling species*D. melanogaster* and *D. simulans.* Mol. Biol. Evol. 15:176-184.

Heina L. and Gatzweiler F. 2005. The economic value of coffee (*Coffea arabica*) genetic resources. Ecological Economics doi:10.1016/j.ecolecon.2005.11.022.

Herrera J.C., Combes M.C., Anthony F., Charrier A. and Lashermes P. 2002a: Introgression into allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica*×*C. canephora*). Theor. Appl. Genet. 104: 661-668.

Herrera J.C., Combes M.C., Cortina H., Alvaredo G. and Lashermes P. 2002b. Gene introgression into *Coffea arabica* by way of triploid hybrids (*C. canephora* x *C. canephora*) Heredity 89: 488-494.

Hillis D.M., Mable B.K., Larson A., Davis S.K. and Zimmer E.A. 1996. Nucleic acids IV: Sequencing and cloning. In: Hillis D.M., Moritz C. and Mable B.K. (eds.), Molecular Systematics, 2nd edtion. Sinauer Associates Inc. Sunderland, MA, USA. pp 321-381.

Hilu K.W. 1993. Polyploidy and the evolution of domesticated plants. Amer. J. Bot. 80:1494-1499.

Hodgkin T. 1997. Some current issues in the conservation and use of plant genetic resources. In: Ayad W.G., Hodgkin T., Jaradat A. and Rao V.R. (eds.), Molecular genetic techniques for plant genetic resources. Report of IPGRI workshop, 9-11 October 1995, Rome, Italy.

Hodgkin T., Roviglioni R., De Vicente M.C. and Dudnik N. 2001. Molecular methods in the conservation and use of plant genetic resources. Acta Hortic. 546:107-118.

Hollingsworth P.M., Dawson I.K., Goodall-Copestake W.P., Richardson J.E., Weber J.C., Sotelo Montes C. and Pennington R.T. 2005. Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. Mol. Ecol. 14: 497–501.

Hoisington D. 2001. Application of biotechnology to maize improvement past, present and future prospects CIMMYT, Mexico. Seventh Eastern and Southern Africa Regional Maize Conference 11th -15th February, 2001, pp.7-11.

Huelsenbeck J.P. and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Hutchinson J. and Dalziel J.M. 1963. Flora of west tropical Africa. 2nd ed., Hepper F.N. (ed.), Vol. III, Part I, Crown Agents for Oversea Governments and Administrations, London. pp.152-157.

Ingvarsson P. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. Evolution 56:2368–2373.

IPGRI 2001. International Plant Genetic Resources Institute Thematic Report 200-2001. Rome, Italy, pp. 8-9.

Ishii T. and McCouch S.R. 2000. Microsatellites and microsynteny in the chloroplast genomes of Oryza and eight other Gramineae species. Theor. Appl. Genet. 100:1257–1266.

Ishii T., Mori N. and Ogihara Y. 2001. Evaluation of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. Theor. Appl. Genet. 103:896–904.

Jaccard P. 1908. Nouvelles rescherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat. 44:223-270.

Johnson L.A. and Soltis D.E. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Ann. Missouri Bot. Gard. 82: 149–175.

Kanowski P. and Boshier D. 1997. Conserving the genetic resources of tree *in situ*. In: Maxted N., Ford-Lloyd B.V. and Hawkes J.G. (eds.), Plant genetic Conservation: The *in situ* approach. Chapman and Hall, UK.

Kaplinsky R. 2004. Competitions policy and the global coffee and cocoa value chains. Institute of development studies university of Sussex, and Centre for research in innovation management, University of Brighton. Paper prepared for United Nations conference for trade and Development (UNCTAD), May 2004.

Karp A, Kresovich S., Bhat K.V., Ayad W.G. and Hodgkin T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2. IPGRI, Rome, Italy.

Kebede M. 2003. Morphological characterization of Hararge Coffee (*C.arabica*) germplasm accessions at prebearing stage. MSc Thesis. School of Graduate studies, Alemeya University. Ethiopia.

Kebede M. and Bellachew B. 2004a. Genetic divergence of Hararge coffee (*Coffea arabica* L) germplasm accessions at pre-bearing stage. 20th International Conference on Coffee Science, 11^{th} –15^{th} October, 2004, Bangalore India.

Kebede M. and Bellachew B. 2004b. Morphological characterization of Hararge coffee (*Coffea arabica* L.) germplasm accessions for qualitative characters. 20th International Conference on Coffee Science, Bangalore, 11^{th} –15^{th} October, 2004.

Kelchner S.A. and Clark L.G. 1997. Molecular evolution and phylogenetic utility of the *rpl*16 intron in *Chusquea* and the Bambusoideae (Poaceae). Mol. Phylogenet. Evol. 8:385-397.

Kelchner S.A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Missouri Bot. Gard. 87:482-498.

Kelchner S.A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am. J. Bot. 89:1651-1669.

Kimura M. and Ohta T. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc. Natl.Acad. Sci. USA 75: 2868-2872.

Klein A-M., Steffan-Dewenter I. and Tscharntke T. 2003. Fruit set of highland coffee increases with the diversity of pollinating bees. *Proc. R. Soc. Lond.* B 270: 95-961.

Knowles P., Perry D.G. and Foster H.A. 1992. Spatial genetic structure in two tamarack (*Larix laricina* Du Roi K. Koch) populations with different establishment history. Evolution 46: 572-576.

Konuma A., Tsumura Y., Ting Lee C., Leong Lee S., and Okuda T. 2000. Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol. Ecol. 9: 1843-1852.

Kufa T. 2006. Ecophysiological diversity of wild Arabica coffee populations in Ethiopia: Growth, water relation and hydraulic characteristics along a climatic gradient. Doctoral Thesis. Center for Development Reseach (ZEF). University of Bonn. Germany.

Kusumi J., Tsumura Y., Yoshimaru H. and Tachida H. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. Am. J. Bot. 87: 1480-1488.

Lan R. and Reeves P.R. 2002. Pandemic Spread of Cholera: Genetic diversity and relationships within the seventh pandemic clone of vibrio cholerae determined by Amplified Fragment Length Polymorphism. J. Clin. Microbiol. 40: 172-181.

Lanham P.G., Foster B.P. and McNicol P. 1994. Seed storage protein variation in *Arachis* species. Genome 37: 487-496.

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. Genet. Res. Crop Evol. 40: 91-99.

Lashermes P., Cros J., Combes M.C., Trouslot P., Anthony F., Hamon S. and Charrier A. 1996a. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Theor. Appl. Genet. 93: 626-632.

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

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., Combes M.C., Robert J., Trouslot P., D'Hont A., Anthony F., and 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., and Anthony F. 2000. Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L). Theor. Appl. Genet. 100: 139-146.

Lee J.S., Hanford M.G., Genova J.L. and Farber R.A. 1999. Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. Hum. Mol. Genet. 8: 2567-2572.

Leimu R. and Mutikainen P. 2005. Population history matting system and fitness variation in a perennial herb with a fragmented distribution. Conserv. Biol. 19: 349-356.

Levinson G. and Gutman G.A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4: 203-221.

Lewontin R.C. 1972. The apportionment of human diversity. Evol. Biol. 6: 381–398.

Lira C.F., Cardoso S.R.S., Ferreira P.C.G., Cardoso M.A. and Provan J. 2003. Long-term population isolation in the endangered tropical tree species *Caesalpinia echinata* Lam. revealed by chloroplast microsatellites. Mol. Ecol. 12:3219-3225.

Liu F., Sun G., Salomon B. and Von Bothmer R. 2002. Characterization of genetic diversity in core collection accessions of wild barley, *Hordeum ulgare* ssp. Spontaneum. Hereditas 136: 67–73.

Löhne C. and Borsch T. 2004. Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. Mol. Biol. Evol. 22: 317-332.

Löhne C., Borsch T. and Wiersema J.H. (subm.) Phylogenetic analysis of Nymphaeales using fast-evolving and non-coding chloroplast markers. Botanical Journal of the Linnean society.

Luikart G., Sherwin W.B., Steele B.M. and Allendorf F.W. 1998. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Mol. Ecol. 7:963-974.

Manen J-F., Natali A. and Ehrendorfer F. 1994. Phylogeny of Rubiaceae-Rubieae inferred from the sequence of a cpDNA intergenic region. Plant Syst. and Evol. 190: 195-211.

Masumbuko L.I., Bryngelsson T., Mneney E.E. and Salomon B. 2003. Genetic diversity in Tanzanian Arabica coffee using random amplified polymorphic DNA (RAPD) markers. Hereditas 139: 56–63.

Masumbuko L.I and Bryngelsson T. 2005. Inter simple sequence repeat (ISSR) analysis of diploid coffee species and cultivated *Coffea arabica* L. from Tanzania. Genet. Res. Crop Evol. 53: 357 – 366.

Melaku W. 1984. Coffee genetic resources in Ethiopia conservation and utilization with particular reference to CBD resistance. Proceeding first regional workshop "Coffee Disease". Addis Ababa, 19-23 July, 1982, pp. 203-211.

Meyer F. 1965. Notes on wild *Coffea arabica* from southwestern Ethiopia, with some historical considerations. Economic Botany 19: 136-151.

Meyer F.G. 1968. Further observation on the history and botany of the Arabica coffee plant, *Coffea arabica* L., in Ethiopia. FAO Mission to Ethiopia 1964-65, Rome.

Molina-Cano J.L., Russell J.R., Moralejo M.A., Escacena J.L., Arias G. and Powell W. 2005. Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley. Theor. Appl. Genet. 110: 613-619.

Moncada P. and McCouch S. 2004. Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. Genome 47: 501-509.

Montagnon C. and Bouharmont P. 1996. Multivariate analysis of phenotypic diversity of *Coffea arabica*. Genet. Res. Crop Evol. 43: 221–227.

Montalvan R., Ando A. and Echeverrigaray S. 1998. Use of seed protein polymorphism for discrimination of improvement level and geographic origin of upland rice cultivars. Genet. Mol. Biol. 21: 1415-4757.

Müller J. and Müller K. 2003. QuickAlign: A new alignment editor. Plant Mol. Biol. Reporter 21:5.

Müller K. 2004. PRAP - computation of Bremer support for large data sets. Mol. Phyl. Evol. 31: 780-782.

Müller K. 2005. Incorporating information from length-mutational events into phylogenetic analysis. Mol. Phylogenet. Evol. 38: 667–676.

Müller K. and Borsch T. 2005. Phylogenetics of Amaranthaceae based on *mat*K/*trn*K sequence data - evidence from parsimony, likelihood, and Bayesian analyses. Ann. Missouri Bot. Gard. 92: 96-102.

Narasimhaswamy R.L. 1968. Report on visit to coffee areas in Ethiopia, 1964-65. FAO Mission to Ethiopia 1964-65, Rome.

Nagaraju J., Reddy K.D., Nagaraja G.M. and Sethuraman B.N. 2001. Comparison of multilocus RFLPs and PCR-based marker systems for genetic analysis of the Silkworm, *Bombyx Mori*. Heredity 86: 588-597.

Neel M.C. and Cummings M.P. 2003. Effectiveness of conservation targets in capturing genetic diversity. Conserv. Biol. 17: 219–229.

Neel M.C and Ellstrand N.C. 2003. Conservation of genetic diversity in the endangered plant Eriogonum *ovalifolium* var. *vineum* (Polygonaceae). Conserv. Genet. 4: 337–352.

Nei M. and Kumar S. 2000. Molecular Evolution and phylogenetics.Oxfored Univrsity press, Inc.

Newbury H.J. and Ford-Lloyd B.V. 1997. Estimation of genetic diversity. In: Maxted N., Ford-Lloyd B.V. and Hawkes J.G. (eds.), Plant genetic Conservation: The *in situ* approach, Chapman and Hall, UK.

Nixon K.C. 2002. WinClada ver. 1.00.08. Published by the author, Ithaca, NY, USA.

Nybom H. and Bartish I.V. 2000. Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspectives in plant ecology, evolution and Systematics 3/2: 93-114.

Ohara M. and Shimamoto Y. 2002. Importance of genetic characterization and conservation of plant genetic resources: The breeding system and genetic diversity of wild soybean (*Glycine soja*). Plant Species Biol. 17: 51–58.

Omolaja S.S., Obatolu C.R. and Williams J.A. 2006. Collecting of *Coffea abeokutae* Cramer and *Coffea liberica* Bull. in southwestern Nigeria. Issue No.122, pp. 29-31.

Ortí G., Pearse D.E. and Avise J.C. 1997. Phylogenetic assessment of length variation at a microsatellite locus. Proc. Natl. Acad. Sci., USA 94: 10745-10749.

Oxfam G.B. 2002. Bitter coffee: how the poor are paying for the slump in coffee prices. Oxfam International, Oxford UK.

Palme A.E. and Venderamin G.G. 2002. Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. Mol. Ecol. 11: 1769-1779.

Palmer J.D., Jansen R.K, Michaels H.J., Chase M.W. and Manhart J.R. 1988. Cholroplast DNA variation and plant phylogeny. Ann. Missouri Bot. Gard. 75: 1180-1206.

Panda R.C., Kumar O.A. and Rao K.G.R. 1986. The use of seed protein electrophoresis in the study of phylogenetic relationship in Chile pepper (*Capsicum* L.). Theor. Appl. Genet. 72: 665-670.

Parisod C., Trippi C. and Galland N. 2005. Genetic variability and founder effect in the Pitcher Plant *Sarracenia purpurea* (Sarraceniaceae) in population introduced into Switzerland: from inbreeding to invasion. Ann. Botany 95:277–286.

Patterson J.T., Larson S.R. and Johnson P.G. 2005. Genome relationships in polyploid *Poa pratensis* and other *Poa* species inferred from phylogenetic analysis of nuclear and chloroplast DNA sequences. Genome 48: 76–87.

Pavlicek A., Hrda S. and Flegr J. 1999. Free-Tree--freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus Frenkelia. Folia Biologica (Praha) 45: 97-99.

Petit R.J., El Mousadik A. and Pons O. 1998 Identifying populations for conservation on the basis of genetic markers. Conserv. Biol. 12**:** 844–855.

Petit R.J, Deguilloux M., Chat J., Grivet D., Garnier-Gere P. and Vendramin G.G. 2005. Standardizing for microsatellite length in comparisons for genetic diversity. Mol. Ecol. 14: 885-890.

Petit R.J., Latouche-Halle C., Pemonge M. and Kremer A. 2002. Chloroplast DNA variation of oaks in France and the influence of forest fragmentation on genetic diversity. Forest Ecol. and Manag. 156: 115–129.

Petty C., Seaman J. and Majid N. 2004.Coffee and household poverty: A study of coffee and household economy in two districts of Ethiopia. Floor Grootenhuis and Save the Children UK.

Posada D. and Crandall K.A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

Powell W., Morgante M., McDevitt R., Vendramin G.G. and Rafalski J.A. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. Proc. Natl. Acad. Sci. USA 92: 7759-7763.

Prevost A. and Wilkinson M.J. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor. Appl. Genet. 98: 107–112.

Provan J., Powell W., Dewar H., Bryan G., Machray G.C. and Waugh R. 1999. An extreme cytoplasmic bottleneck in modern European cultivated potato (*Solanum tubersum*) is not reflected in decreased levels of nuclear diversity. Proc. R. Soc. Lond. B 266: 633-639.

Provan J, Russell J.R, Booth A. and Powell W. 1999. Ploymorphic chloroplast simple sequence repeat primers for systematic and population studies in the genus *Hordeum.* Mol. Ecol. 8: 505-511.

Provan J., Powell W. and Hollingsworth P.M. 2001.Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends Ecol. Evolut. 16: 142-147.

Provan J., Wolters P., Caldwell K.H. and Powell W. 2004. High-resolution organellar genome analysis of *Triticum* and *Aegilops* sheds new light on cytoplasm evolution in wheat. Theor. Appl. Genet. 108: 1182–1190.

Puff C. 2003. Rubiaceae. In: Hedberg I., Edwards S. and Nemomissa S. (eds.), Flora of Ethiopia and Eritrea Apiaceae to Dipscaceae, Vol. 4, Part 1. The National Herbarium, Addis Ababa University, Ethiopia.

Puff C. and Chamchumroon V. 2003. Non-indigenous Rubiaceae grown in Thailand.Thai For. Bull. (Bot.) 31: 75–94.

Purseglove J. 1968. Tropical Crops: Dicotyledons. Harlow, UK: Longman Group Ltd., pp 458-492.

Ponte S. 2001. Coffee markets in East Africa: Local responses to global challenges or global responses to local challenges? Working paper sub-series on Globalisation and Economic Restructuring in Africa No. XV, CDR working paper 01.5 September 2001.Published and distributed by Centre for Development Research, Copenhagen.

Quandt D. Universal primers for amplification of the *trn*K/*mat*K-*psb*A region in land plants. Anales del Jardín Botánico de Madrid xx: in press.

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.

Rauch E.M. and Bar-Yam Y. 2004. Theory predicts the uneven distribution of genetic diversity within species. Nature 431: 449-452.

Renau-Morata B.A., Nebauer S.G., Sales E., Allainguillaume J., Caligari P. and Segura J. 2005. Genetic diversity and structure of natural and managed populations of *Cedrus Atlantica* (Pinaceae) assessed using Random Amplified Polymorphic DNA. Am. J. Bot. 92: 875–884.

Richerzhagen C. and Virchow D. 2002. Sustainable utilization of crop genetic diversity through property rights mechanisms? The case of coffee genetic resources in Ethiopia. The Economics of Incentive Mechanisms for Biodiversity Conservation: Property Rights. BioEcon Workshop May 30 – May 31, 2002.

Robertson A., Newton A.C. and Ennos R.A. 2004. Multiple hybrid origins, genetic diversity and population structure of two endemic *Sorbus* taxa on the Isle of Arran, Scotland. Mol. Ecol. 13: 23-34.

Rohlf F.J. 2000. NTSYS-$_{PC}$, Numerical Taxonomy and Multivariate Analysis System,Version 2.11. Exeter Software New York.

Romero-Severson J., Aldricw P., Feng Y., Sun W. and Michler C. 2003. Chloroplast DNA variation of northern red oak *(Quercus rubra L.)* in Indiana. New For. 26: 43-49.

Roose M.L. and Gottlieb L.D. 1976. Genetic and biochemical consequences of polyploidy in *Tragopogon*. Evolution 30**:** 818–830.

Rossetto M., Weaver P.K. and Dixon K.W. 1995. Use of RAPD analysis in devising conservation strategies for the rare and endangered Grevillea scapigera (*Proteaceae*). Mol. Ecol. 4: 321– 329.

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. Mol. Biol. 26: 319-327.

Saitou N. and Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Sakai M., Kanazawa A., Fujii A., Thseng F.S., Abe J. and Shimamoto Y. 2003. Phylogenetic relationships of the chloroplast genomes in the genus Glycine inferred from intergenic spacer sequences. Plant Syst. Evol. 239: 29-54.

Säll T., Jakobsson M, Lind-Hallden C. and Hallden C. 2003. Chloroplast DNA indicates a single origin of the allotetraploid *Arabidopsis suecica.* J. Evol. Biol. 16: 1019-1029.

Santalla M., Men´endez-Sevillano M.C., Monteagudo A.B. and De Ron A.M. 2004. Genetic diversity of Argentinean common bean and its evolution during domestication. Euphytica 135: 75–87.

Satastica StatSoft, Inc. 2001. STATISTICA (data analysis software system), version 6.C www.statsoft.com.

Sato T., Isagi Y., Sakio H., Osumi K. and Goto S. 2006.Effect of gene flow on spatial genetic structure in the riparian canopy tree *Cercidiphyllum japonicum* revealed by microsatellite analysis. Heredity 96: 78-84.

Seifu S., Singh H. and Bellachew B. 2004. Diversity in the Ethiopian coffee (*Coffea arabica* L.) germplasm. 20th International Conference on Coffee Science 11-15 October 2004, Bangalore, India.

Senbeta W.F. 2006. Biodiversity and ecology of afromontane rainforests with wild *Coffea arabica* L. populations in Ethiopia. Ecology and Development Series No. 38, Zentrum fur Entwicklungsforschung, Center for Development Research. University of Bonn. Germany.

Schlotterer C., Ritter R., Harr B. and Brem G. 1998. High mutation rates of long microsatellite alleles in *Drosophila melanogaster* provide evidence for allele specific mutation rates. Mol. Biol. Evol. 15: 1269-1274.

Schmidtling R.C. and Hipkins V. 1998. Genetic diversity in longleaf pine *(Pinus pahstris):* influence of historical and prehistorical events. Can. J. For. Res. 28: 1135-1145.

Schwaegerle K.E. and Schaal B.A. 1979. Genetic variability and founder effect in pitcher plant *Sarracenia purpurea* L. Evolution 33: 1210-1218.

Sergio L. and Gianni B. 2005. Molecular markers based analysis for crop germplasm preservation. In: The role of biotechnology for the characterisation and conservation of crop, forestry, animal and fishery genetic resources. International workshop, FAO, 5-7 March, 2005.Villa Gualino, Turin, Italy,

Sgorbati S., Labra M., Grugni E., Barcaccia G., Galasso G., Boni U., Mucciarelli M., Citterio S., Iramátegui A.B., Gonzales L.V. and Scannerini S. 2004. A survey of genetic diversity and reproductive biology of *Puya Raimondii* (Bromeliaceae), the endangered queen of the Andes. Plant Biology 6: 222–230.

Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., Siripun K.C.,Winder C.T., Schilling E.E. and Small R.L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92: 142–166.

Sica M., Gamba G., Montieri S., Gaudio L. and Aceto S. 2005. ISSR markers show differentiation among Italian populations of *Asparagus acutifolius* L. BMC Genetics 6: 17.

Silvarolla M.B, Mazzafera P. and De Lima M.M.A. 2000. Caffeine content of Ethiopian *Coffea arabica* beans. Genet. Mol. Biol. 23: 213-215.

Silvarolla M.B., Mazzafera P. and Fazuoli L.C. 2004. A naturally decaffeinated arabica coffee. Nature 429: 826.

Silveira, S. R.; Ruas, P. M.; Ruas, C. F.; Sera, T.; Carvalho, V. P. and Coelho, A. S. G. (2003), Assessment of genetic variability within and among progenies and cultivars of coffee using RAPD markers. Gen. Mol. Biol. 26:329-336.

Simmons M.P. and Ochoterena H. 2001. Incorporation, relative homoplasy and effect of gap characters in sequence-based phylogenetic analyses. Syst. Biol. 50: 454-462.

Siva R. and Krishnamurthy K.V. 2005. Isozyme diversity in *Cassia auriculata* L.Afr. J. Biotechnol. 4: 772-775.

Slatkin M. 1987. Gene flow and geographic structure of natural populations. Science 236:787-792.

Slatkin M. and Barton N.H. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43: 1349-1368.

Small R.L, Ryburn J.A., Cronn R.C., Seelanan T. and Wendel J.F. 1998. The Tortoise and The Hare: Choosing between noncoding plastome and nuclear *ADH*

sequences for phylogeny reconsraction in recently diverged plant group. Am. J. Bot. 85: 1301-1315.

Sneath P.H.A. and Sokal R.R. 1973. Numerical taxonomy. WH Freeman and Company, San Francisco, California, USA.

Soltis P.S. 2005. Ancient and recent polyploidy angiosperms. New phytologist 166: 5-8.

Soltis D.E. and Soltis P.S. 1989. Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. Am. J. Bot. 76**:** 1119–1124.

Soltis P.S., Doyle J.J. and Soltis D.E. 1992. Molecular data and polyploidy evolution in plants. In: Soltis P.S., Soltis D.E. and Doyle J.J. (eds.), Molecular Systematics of Plants, Chapman and Hall, New York, pp 117-150.

Soltis P.S. and Soltis D.E. 1991. Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. Syst. Bot. 16: 407–413.

Soltis D.E. and Soltis P.S. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis D.E., Soltis P.S. and Doyle J.J. (eds.), Molecular systematics of plants II. DNA sequencing, Kluwer Academic Publishers, Boston, pp. 1-42.

Soltist P.S., Doyle J.J. and Soltist D.E. 1992. Molecular data and polyploid evolution in plants. In: Soltis P.S., Soltis D.E. and Doyle J.J. (eds.), Molecular Systematics of Plants, Chapman and Hall, New York, pp. 177-201.

Soltis P.S. and Soltis D.E. 2000.The role of genetic and genomic attributes in the success of polyploids. PNAS 97: 7051–7057.

Soltis D.E, Soltis P.S. and Milligan B.G. 1992. Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: Soltis P.S., Soltis D.E. and Doyle J.J. (eds.), Molecular Systematics of Plants, Chapman and Hall, New York, pp 441–457.

Soltis D.E., Soltis P.S, Pires J.C., Kovarik A., Tate J.A. and Mavrodiev E. 2004. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons**.** Biol. J. Linnean Soc 82**:** 485–501.

Steinberg E.K., Lindner K.R., Gallea J., Maxwell A., Meng J. and Allendorf F.W. 2002. Rates and patterns of microsatellite mutations in pink salmon. Mol. Biol. Evol. 19: 1198–1202.

Stoffelen P. 1998. *Coffea* and *Psilanthus* (Rubiaceae) in tropical Africa: a systematic and palynological study, including a revision of the West and Central African species. PhD Thesis, Leuven: Katholieke Universiteit.

Studier J.A. and Keppler K.J., 1988. A note on the neighbor-joining algorithm of Saitou and Nei. Mol.Bol.Evol 5: 729-731.

Sukhotu T., Kamijima O. and Hosaka K. 2005. Genetic diversity of the Andean tetraploid cultivated potato (*Solanum tuberosum* L. subsp. *andigena* Hawkes) evaluated by chloroplast and nuclear DNA markers. Genome 48: 55–64.

Sweigart A.L. and Willis J.H. 2003. Patterns of nucleotide diversity are affected by mating system and asymmetric introgression in two species of *Mimulus*. Evolution 57: 2490-2506.

Swofford D.L. 2002. PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0b6 (computer program). Sinauer Associates, Sunderland, Massachusetts.

Syamsuardi and Okada H. 2002. Genetic diversity and genetic structure of populations of *Ranunculus japonicus* Thunb. (Ranunculaceae). Plant Species Biol. 17: 59–69.

Sylvian P.G. 1955. Some observations of *Coffea arabica* L. in Ethiopia. Turrialba 5:37-53.

Sylvian P.G. 1958. Ethiopia coffee-its significance to world coffee problems. Economic Botany 12:111-139.

Tadesse M. and Nigatu L. 1996. An ecological and ethnobotanical study of wild or spontaneous coffee, *Coffea arabica* in Ethiopia. In: van der Maesen L.J.G., van der Burgt X.M. and van Medenbach de Rooy J.M. (eds.), The Biodiversity of African Plants, Proceedings XIVth AETFAT Congress, 22–27August 1994, Wageningen, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 277–294.

Teketay D. 1999. History, botany and ecological requirements of Coffee. Walia; Volume page 20:28-50.

Teketay D., Anaga A., Mulat G. and Enyew M. 1998. Study on forest coffee conservation. CIP/CTA, Addis Ababa.

Teketay D. and Tegineh T. 1991. Traditional tree crop based agroforestry in coffee producing areas of Harerge, Eastern Ethiopia. Agroforestry Systems 16: 257-267.

Teketay D. and Tigneh A. 1994. A study on landraces of Harer coffee in Eastern Ethiopia.In: Seyani J.H. and Chikuni A.C. (eds.), Proceedings of the 13th Plenary Meeting of AETFAT, National Herbarium and Botanic Gardens of Malawi, Zomba, Malawi, pp.161-169.

Tesemma T. and Balay G. 1991 Aspects of Ethiopian tetraploid wheats with emphasis on durum wheat genetics and breeding research. In: Gebre-Mariam H., Tanner D.G. and Hulluka M. (eds.), Wheat Research in Ethiopia.. IAR/CIMMYT, Addis Ababa, Ethiopia. pp. 47-71.

Tesfaye G.K. 2000. Morphological and biochemical diversity of emmer wheat (*Triticum dicoccum* (Schrand) Schub.) in Ethiopia. M.Sc Thesis Department of Biology Addis Ababa University/Ethiopia.

Tesfaye G.K., Borsch T., Govers K. and Bekele E. (submitted)a. Characterisation of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. (Submitted).

Tesfaye G.K., Govers K., Bekele E. and Borsch T. (Submitted)b: ISSR fingerprinting of *Coffea arabica* throughout Ethiopia reveals high variability in wild populations and distinguishes them from farmer's varieties. (Submitted)

Tesfaye T. 2006. Coffee forest conservation, local level institutions influencing the conservation and use of coffee forests in SW Ethiopia. Doctoral Thesis.University of Hohenheim.

Tero N., Aspi J., Siikamaki P. and Jakalaniemi A. 2005. Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). Heredity 94: 478-487.

Terzi V., Faccioli P., Pecchioni N., Pisani R., Arru L. and Stanca A.M. 1999. Detection of genetic diversity in barley using tools from DNA technologies. In: Broggio M. (eds.), Pan European Network on the Genetic Indicators of Biodiversity (PEN/GIB): Description of Diversity and Exchange of Information by Network Web. Florence, 25-27 October 1999. Istituto Agronomico per l'Oltremare. Florence, Italy.

Tosti N. and Negri V. 2002. Efficiency of three PCR-based markers in assessing genetic variation among cowpea (*Vigna unguiculata* subsp. unguiculata) landraces. Genome 45: 268-275.

Van der Graaf N.A. 1981. Selection of arabica coffee types resistant to coffee berry disease in Ethiopia. Doctoral Thesis, Uinversity of Wageningen, The Netherlands, pps.110.

Vavilov N.I. 1951. The origin, variation, immunity and breeding of cultivated plants. Chronica Botanica 13: 1-351.

Vekemans X. and Hardy O.J. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. Mol. Ecol. 13: 921-935.

Vellend M. and Geber M.A. 2005. Connections between species diversity and genetic diversity. Ecological Letters 8: 767–781.

Vendramin G.G., Degen B., Petit R.J., Anzidei M., Madaghiele A. and Ziegenhagen B. 1999. High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. Mol. Ecol. 8: 1117–1126.

Vigouroux Y., Jaqueth J.S., Matsuoka Y., Smith O.S., Beavis W.D., Smith J.S.C. and Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in Maize. Mol. Biol. Evol. 19:1251–1260.

Viniegra-Gonzalez G. 2000. Biotechnology and future of coffee production. In: Sera T., Soccol C.R., Pandey A. and Rous S., (eds.), Coffee biotechnology and quality. Proceedings of the 3rd International Seminar on Biotechnology in the Coffee Agro-Industry. Londrina, Brazil, pp. 1-16.

Weber J.L. and Wong C. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. 2: 1123–1128.

Weising K. and Gardner R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42: 9-19.

Weising K., Nybom H., Wolff K. and Meyer W. 1995. DNA Fingerprinting in Plants and Fungi. Arbor A. (ed.), First edition. CRC Press, Boca Raton, Florida.

Weising K., Nybom H., Wolff K. and Kahl G. 2005. DNA fingerprinting in plants: principles, methods and applications, second edition. CRC Press, Taylor and Francis Group.

Wellman F.L. 1961. Coffee: Botany, Cultivation and Utilization. World Crop Series. Inter Science, New York .

Wells W.C., Isom W.H. and Waines J.G. 1988. Outcrossing rates of six common bean lines. Crop Sci 28: 177–178.

Wendel J.F. 2000. Genome evolution in polyploidy. Plant Mol. Biol. 42: 225–249.

Widmer A. and Baltisberger M. 1999a. Molecular evidence for allopolyploid speciation and a single origin of the narrow endemic *Draba ladina* (Brassicaceae). Am. J. Bot. 86:1282–1289.

Widmer A. and Baltisberger M. 1999b. Extensive intraspecific chloroplast DNA (cpDNA) variation in the alpine *Draba aizoides* L. (Brassicaceae): haplotype relationships and population structure. Mol. Ecol. 8:1405-1415.

Woldetsadik W. and Kebede K. 2000. Coffee production systems in Ethiopia. In: Proceedings of the workshop on control of Coffee Berry Disease in Ethiopia Addis Ababa (Ghion Hotel) 13-15 August 1999. Addis Ababa, Ethiopia.

Wold-Mariam A.B. 2004. Kaffa kingdom and their short history. Amaharic Book. Mega printing house. Addis Ababa, Ethiopia.

Wolfe A.D. and Liston A. 1998. Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Soltis D.E., Soltis P.S. and Doyle J.J. (Eds.). Plant Molecular Sstematics II. Kluwer Acadamic Publisher, Boston, MA, pp.43-86.

Wolfe A.D. and Randle C.P. 2001. Relationships within and among species of the Holoparasitic Genus *Hyobanche* (Orobanchaceae) inferred from ISSR banding patterns and nucleotide sequences. Syst. Bot. 26: 120–130.

Wondimu M. 1987. Variation in physiologic groups of *Coffea arabica* L. for resistance to rust *(Hemileia vastatrix* B. and Br.) and their pattern of distribution in Ethiopia. Germplasm Newsletter 15: 8-13.

Woods K., Hilu W.K., John H.W. and Borsch T. 2005. Pattern of variation and systematics of *Nymphaea odorata*: I. Evidence from morphology and inter-simple sequence repeats (ISSRs). Sysematic Botany 30: 471-480.

Yeh F.C. and Boyle T.J.B. 1997. Population genetic analysis of codominant markers and qualitative traits. Belgian J. Bot. 129:157.

Yapo A., Louarn J. and Montagnon C. 2003. Genetic gains for yield after two back-crosses of the interspecific hybrid Libusta (*Coffea canephora* P. X *C. liberica* Bull. Ex Hiern) to *C. canephora* P.. Plant Breeding 122: 288-290.

Xu D.H., Abe J., Sakai M., Kanazawa A. and Shimamoto Y. 2000. Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. Theor. Appl. Genet. 101: 724-732.

Zewde B. 2002. A history of modern Ethiopia, 1855-1991. 2nd edition, Addis Ababa University printing press.

Zhang D.X. and Hewitt G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol. Ecol.12: 563-584.

Zhu Y., Queller D.C. and Strassmann J.E. 2000. A phylogenetic perspective on sequence evolution in microsatellite loci. J. Mol. Evol. 50: 324-338.

Zietkiewicz E., Rafalski A. and Labuda D. 1994. Genome fingerprint by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176-183.

ACKNOWLEDGEMENTS

The present thesis was carried out in the framework of the CoCE project (Conservation and use of wild populations of *Coffea arabica* L. in the montane rainforests of Ethiopia) in the BIOTEAM program, sponsored by the German Federal Ministry of Education and Research (BMBF). This Ph.D work was supervised by PD Dr. Thomas Borsch and Professor Wilhelm Barthlott from the University of Bonn, Germany, and Professor Endashaw Bekele from Addis Ababa University (AAU), Ethiopia.

Many persons have contributed to the realization of this project. First of all, I am extremely grateful for my direct supervisors Dr. Thomas Borsch and Prof. Endashaw Bekele for accepting me as a student to work in their lab and providing me the necessary support during the study period. They have both, with their respective fields of expertise, been instrumental in helping me to accomplish my project. Both have been following my work progress closely during fieldwork as well as in the laboratory. Thank you for your insightful comments and ideas as well as your incredibly positive energy. Here I would like to thank Dr. Thomas again for introducing me to the area of plant phylogenetics, and for scientific guidance, patience, critical reviews of manuscripts and untiring support during the course of the study. Without his support it would have been difficult to achieve this project. I am also indebted to Prof. Wilhelm Barthlott, the Director of Nees-Institute for Biodiversity of Plants for his continuous support and encouragement. I would like to thank and appreciate Dr. Demel Teketay (Former Director General of the Ethiopian Agricultural Research Organization, EARO), Dr. Manfred Denich (Center for Development Research, ZEF), for their support in developing the CoCE-project. I thank the project coordinators based in Addis Ababa and Bonn; Dr. Tadesse Woldemariam, Dr. Franz Gatzweiler and Detlef Virchow (former CoCE coordinator).

I would like to express my heartful thanks and appreciation to Ing. Kim Govers, who has been supervising the lab processes in the project, for his help all the way in the laboratory analysis to the installation of laboratory equipment in Addis Ababa University, and for his continuing support and friendship. I am grateful to all members of the Molecular Systematics Working Group at the Nees-Institute for Biodiversity of Plants, Dr. Kai Müller, Dr. Cornelia Löhne, Andreas Worberg, Dr. Rolf Blöcher, Dr. Dietmar Quandt, Kerstin Wilhelm and Nadja Korotkova– thank you for all the assistance in the lab and at the PC.

I thank the Oromia Agricultural Research Institute and the Sinana Agricultural Research Center and all their respective staff for allowing me the study leave, and providing me with the necessary support and encouragement during my work in Ethiopia. I also thank the Department of Biology of Addis Ababa University and the Ethiopian Institute of Agricultural Research administration for facilitating my work in Ethiopia.

I am grateful to the ZEF doctoral program coordinator Dr. Gunter Manske and his assistant for all their help and support during the study period. I wish to thank all the students in CoCE project for all the nice discussion and working environment we had: Admasu Shibru, Alice Beining, Anke Rojahn, Aseffa Siyume, Christine Schmitt, Fekru Meko, Dr. Feyera Senbeta, Tamiru Oljira, Taye Kufa, Teklu Tesfaye and Till Stellmacher. I also thank Georg Lieth for arranging the GIS map of the study sites. I would also like to thank the staffs in project coordination office of Addis Ababa; Mr. Andinet Ambesse and Wondimagen Gudeta. I appreciate the assistance and help of the

local people and district agricultural development office in all the wild coffee forest I had been.

I am grateful to Mr. Mulatu Gelata, Mr. Yemane Tsehaye and Mr. Medhane Asmelash for the assistant and discussion on different statistical tools at the PC in Addis Ababa. I am grateful to Mr. Million Abebe and Dr. Bayetta Bellachew from the Jimma Agricultural Research Center (Ethiopia) and Institute of Biodiversity Conservation (Ethiopia, IBC) for providing me with leaf samples of diploid *Coffea* and *Coffea arabica* landraces. I appreciate the help of the Tropenzentrum Witzenhausen University of Kassel with samples of *C. arabica* cultivars. Friends and Ethiopian colleagues in Bonn also deserve special thanks; Mr. Alemu his family, Dr. Daniel Tesgai, Dr. Dawit Tesaye and his family, Mr. Dessi, Dr. Dilnessaw Alamirew and his family, and Mr. Gezahegn Bekele and his family, Dr. Lulseged Tamene, Mr. Liykune, Mr. Melkay, and her family, Mr. Sewalem Meberate and his family, Dr. Tsegaye Yilma his family, Frau Workenesh B. Schmitd and Adde Yeshi Maraga and her family. I am also thankful for friends and colleagues in Addis Ababa particularly Mr. Tsegaye Haile and his wife Tsehaye Alemayhu.

I would like to thank and appreciate the countless support from my parents; my mother Birkineshe Waqkene, my father Tesfaye Geletu, my sisterts Fanaye, Meseret and Hirut. They played a major role in encouraging me and pushing me in attaining my goal.

My dearest gratitude goes to my wife Neway Beco for her patience, heartfelt support and encouragement during the course of the study.

Finally, I would like to thank the almighty GOD for helping me throughout my life including in the realization of this thesis.