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

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

Zentrum für Entwicklungsforschung
Center for Development Research
University of Bonn
ZEF Bonn

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).

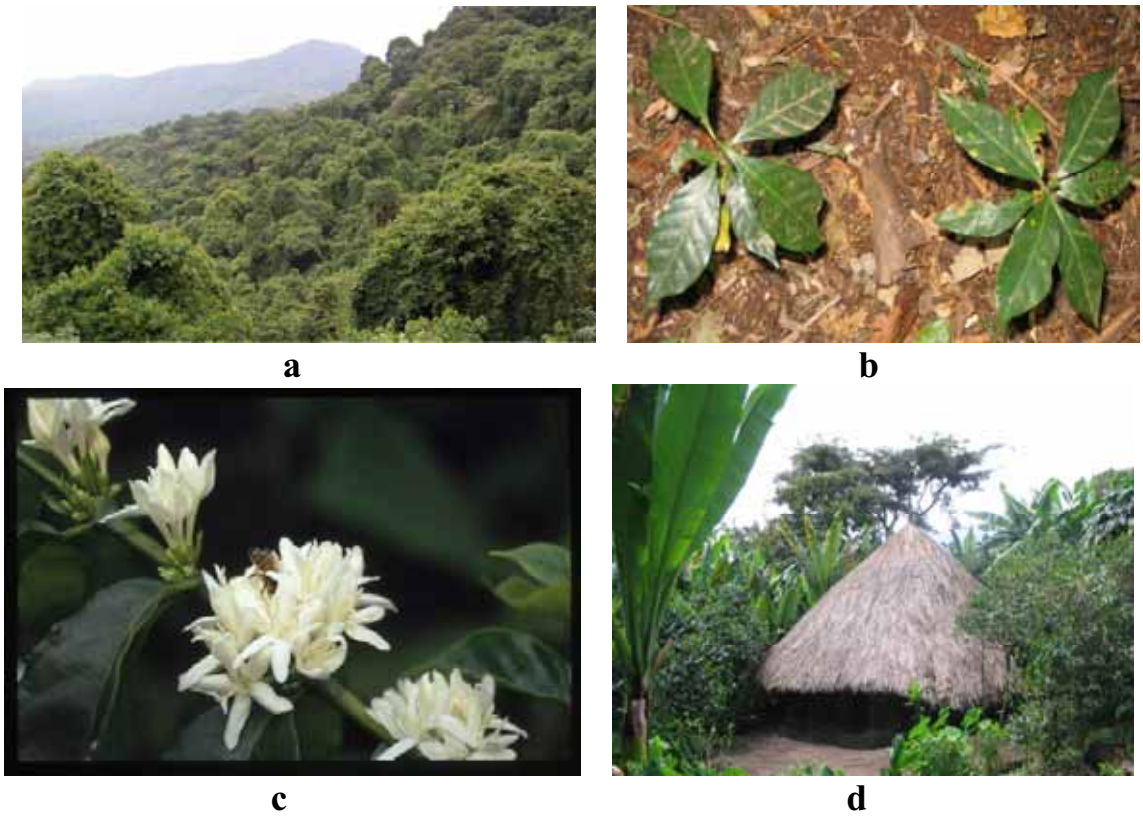

Figure 1.1 Some pictures of coffee from SW, S and SE Ethiopia. (a) Afromontane rainforest with wild coffee Berhane Kontir forest, (b) Naturally regenerated seedling of C. arabica, Bale Mountain (Harrena forest, SE), c) flowers of C. arabica with natural pollinators (honey bee) in Bench Maji, SW, (d) 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).

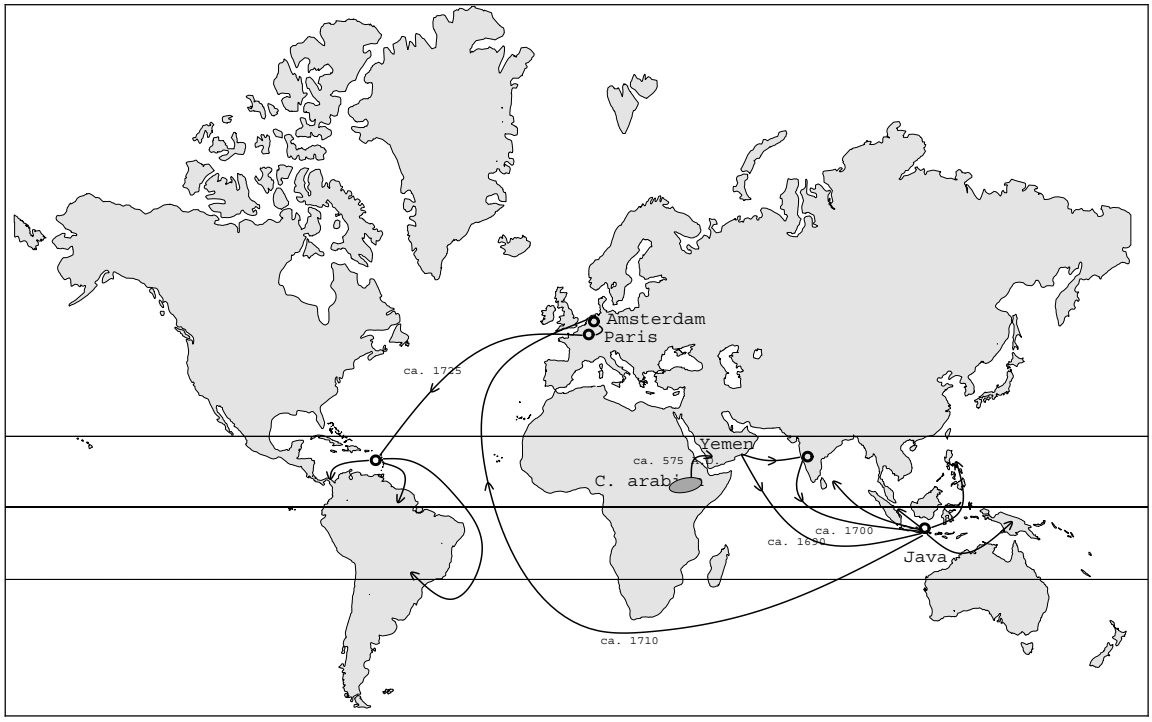

Figure 1.2 Distribution routes for cultivated *C. arabica* in the tropics (Ferwerda 1976). The numbers are the approximate years of introduction.

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