Journal of Science and Sustainable Development (JSSD), 2015, 3(2), 1-17

ISSN: 2304-2702 (print); 2414-4479 (online)
DOI: https://doi.org/10.20372/au.jssd.3.2.2015.044

Simple Sequence Repeat (SSR) Variation of Ethiopian Forest Coffee (*Coffea arabica* L.) Populations

Esayas Aga[1] Endashaw Bekele[2] and Tomas Bryngelsson[3]
[1]*Department of Biology, College of Natural and Computational Sciences, Ambo University,*
[2]*Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia*
[3]*Department of Crop Science, Swedish University of Agricultural Sciences, Alnarp, Sweden*

Abstract

*Genetic diversity and relatedness among 240 individuals of forest coffee trees (***Coffea arabica** ***L.)** *representing 24 populations from Ethiopia were evaluated using 11 simple sequence repeat (SSR) loci. Three of the loci were found to be monomorphic and the other eight loci produced a total of 97 alleles with a mean of 12.1 alleles per locus. Polymorphic information content (PIC) values varied from 0.19 for locus M_{29} to 0.74 for locus M_{47} with an average of 0.48 per locus. Jaccard similarity coefficient between all possible pairs of samples varied from 0.41 to 0.96. Total Nei's genetic diversity (H_T) = 0.25 while the average diversity within populations (H_S) = 0.08. The coefficient of genetic differentiation between populations (G_{ST}) = 0.68. Thus, 68% of the total variation was between populations. The partitioning of genetic variation into within and between populations based on Shannon's information index also revealed more differentiation between populations (0.72) than within populations (0.28), which are in line with the features of predominantly inbreeding populations. The phenogram derived from Jaccard similarity coefficient between populations produced a single tree in which some populations clustered with respect to their region of origin, and a slight differentiation of populations into east and west of the Great Rift Valley was observed except for few populations, which exhibited unique allelic forms. Although,* **C. arabica** *is characterized by low genetic diversity owing to its allopolyploidy and autogamous nature, the SSR markers were able to differentiate the forest coffee samples considered in the study. The results are discussed in relation to conservation of forest coffee trees in Ethiopia.*

Keywords: Conservation, forest coffee, Ethiopia, genetic diversity, simple sequence repeat.

Introduction

Coffee trees belong to the genus Coffea in the family Rubiaceae. Global commercial production relies mainly upon two species, *Coffea arabica* L. and C. canephora Pierre. *C. arabica*, which accounts for 70% of world coffee production is self-fertile and the only known allopolyploid species of the genus. The other Coffea species are dipoid and self-sterile except for C. heterocalyx and C. sp. Moloundou, which are diploid but self-fertile (Lashermes *et al*., 1999; Coulibaly *et al*., 2002). *C. arabica* is characterized by low genetic diversity owing to its allopolyploidy and autogamous nature (Lashermes *et al*., 1996a).

Ethiopia is the origin and primary center of diversity of *C. arabica* (Sylvain, 1955). There is every ground for the belief that Ethiopia is the original home of *C. arabica*, which is confirmed by the fact that within small areas four or more genetically different types of coffee trees can be found, and C. *aethiopica* would therefore have been a more correct name than *C. arabica* (Strenge, 1956). The first belief that this plant was of Arabic origin was due to the fact that the first knowledge of the beverage and the tree was obtained from Arabia, hence the scientific name given by Linnaeus (Sylvain, 1955).

Forest coffee refers to *C. arabica* growing naturally in the forest or the young plants, which have been irregularly transplanted in the forest (Van der Graaf, 1981). However, many of the forest coffee trees now seen in the Southwest and Southeast forest regions of Ethiopia could be remnant of the previous plantings, but this does not preclude the existence of truly wild coffee stands, which have escaped cultivation (Sylvain, 1955). Assessment of the genetic variability within and between forest coffee tree (*C.arabica* L.) populations in Ethiopia is important for coffee breeding and the conservation of species genetic resources. Many characteristics of agronomic interest were observed in Ethiopian coffee, such as the incomplete resistance to orange leaf rust (*Hemileia vastatrix*), resistance to nematode (*Meloidogyne incognita*) and to coffee berry disease caused by Colletotrichum kahawae (Van der Vossen, 2001). In addition, Ethiopian coffee is characterized by good flavor on global scale, and also it is the potential source of low caffeine or caffeine free genotypes. Previous studies with other molecular markers such as RAPD (random amplified polymorphic DNA) revealed 0.46 of mean genetic diversity (Aga *et al.*, 2003), and ISSR (inter simple sequence repeat) revealed 0.55 genetic diversity (Aga *et al*., 2005) among the same materials, indicated the existence substantial genetic variation among the forest coffee samples from Ethiopia.

Plant DNA polymorphism assays are powerful tools for characterizing and

investigating germplasm resources and genetic relatedness (Powell *et al*., 1996). These techniques include RFLP (restriction fragment length polymorpism) and PCR-based molecular markers such as RAPD (random amplified polymorphic DNA) (Welsh & McClelland, 1990; Williams *et al*., 1990), AFLP (amplified fragment length polymorphism) (Vos *et al*., 1995), ISSR (inter-simple sequence repeat) (Zietkiewicz *et al*., 1994) and SSR (simple sequence repeat) (Tautz, 1988; Morgante & Olivieri, 1993).

Simple sequence repeat (SSR) consist of tandem repeat units of short nucleotide motifs (1-6 base pair long) occurring frequently and randomly throughout the genomes of plants and animals, and typically show extensive variation within and among individuals, populations and species (Jarne & Logoda, 1996). As multiallelic single-locus co-dominant genetic marker in which both homozygotes and heterozygotes are detected, the SSR markers improve the information content per locus over the other PCR based dominant markers (RAPD, AFLP and ISSR). Simple sequence repeats have been used for the study of genetic diversity and population differentiation in the genus Eucalyptus (Byrne *et al*., 1996), in rice (Cho *et al*., 2000; Song *et al*., 2003; Yu *et al*., 2003), in cassava (Fregene *et al*., 2003), and in banana (Creste *et al*., 2004).

In coffee, Combes *et al*. (2000) described 11 SSR loci with their respective repeat motifs from *C. arabica* var. Catura and designed 11 primer pairs, which were also found to work well for most of the other diploid coffee species. Out of the 11 pair of primers, six were reported to be polymorphic in *C. arabica* (Combes *et al*., 2000; Anthony *et al*., 2002).

The present study was therefore undertaken with the aimed to assess the extent of genetic diversity (within and between) of forest coffee in Ethiopia using SSR markers to provide additional molecular markers based information in order to design appropriate conservation strategies for its sustainable use.

Materials and Method

Plant material

Seed samples representing 24 populations of natural forest coffee trees (*C. arabica* L.) were collected from two regional states (Oromia and SNNPS) of Ethiopia during January to February 2004. Overall six administrative zones were considered in the study. From the six administrative zones four were from Oromia state (Wollega, Ilubabor, Jimma and Bale zones), and two from SNNPS (Kafa and Bench Maji zones). Within each zone, samples were classified based on the Districts of the sample collection sites and altitude of the site (Table 1). The seed samples were germinated and grown in a

glasshouse, and leaf samples from a month old seedlings were used for genomic DNA extraction following the hexadecyl trimethyl ammonium bromide (CTAB) procedure (Wang, *et al*., 1996) with some modifications.

Table 1. Lists of forest *C. arabica* populations used for the study and specific sites of collection.

Populations	Sample size	District	Altitude (m)	Latitude	Longitude
Welega-1	10	Anfilo	1718	N 08° 37'	E 034° 37'
Welega-2	10	Sayo	1971	N 08° 38'	E 034° 40'
Welega-3	10	Anfilo	1422	N 08° 31'	E 034° 35'
Welega-4	10	Sayo	1925	N 08° 31'	E 034° 44'
Welega-5	10	Sayo	1707	N 08° 36'	E 034° 48'
Ilubabor-1	10	Bedele	1929	N 08° 25'	E 036° 18'
Ilubabor-2	10	Yayo Hurumu	1541	N 08° 21'	E 036° 02'
Ilubabor-3	10	Yayo Hurumu	1430	N 08° 22'	E 035° 52'
Ilubabor-4	10	Alediduu	1796	N 08° 11'	E 035° 34'
Ilubabor-5	10	Chora	1944	N 08° 21'	E 036° 13'
Ilubabor-6	10	Yayo Hurumu	1388	N 08° 19'	E 035° 64'
Jimma-1	10	Seka Chokorsa	1967	N 07° 31'	E 036° 32'
Jimma-2	10	Yabu	1693	N 07° 48'	E 036° 41'
Jimma-3	10	Seka Chokorsa	1499	N 07° 30'	E 036° 31'
Jimma-4	10	Goma	1652	N 07° 55'	E 036° 31'
Kafa-1	10	Gimbo	1887	N 07° 18'	E 036° 03'
Kafa-2	10	Gimbo	1903	N 07° 18'	E 036° 03'
Bench Maji-1	10	Shako	1393	N 07° 02'	E 035° 26'
Benchi-2	10	Shako	1456	N 07° 01'	E 035° 33'
Bale-1	10	Dolo Mana	1498	N 06° 28'	E 039° 45'
Bale-2	10	Dolo Mana	1580	N 06° 30'	E 039° 44'
Bale-3	10	Dolo Mana	1413	N 06° 25'	E 039° 44'
Bale-4	10	Dolo Mana	1395	N 06° 25'	E 039° 52'
Bale-5	10	Dolo Mana	1258	N 06° 28'	E 039° 56'

DNA extraction

Fresh young leaf samples were frozen in liquid nitrogen and ground to powder. Powdered leaves (0.35-0.40g) were collected in eppendorf tubes and 750µl of extraction buffer (0.1M Tris pH 7.5, 0.05M EDTA, 0.5 MNaCl and 100µl 10% w/v SDS) was added. The mixture was incubated for 20 min at 65°C, 250µl of 5MKAc was added and kept on ice for 1 hour before centrifugation at 1400 rpm. The supernatant was precipitated with an equal volume of iso-propanol and centrifuged again at 1400 rpm. The pellet was air dried, dissolved in 250µl of TE (10mM Tris HCl, pH 7.6, 1mM EDTA), 250µl of CTAB buffer (0.2 M Tris pH 7.5, 50mM EDTA) 2MNaCl, 2% CTAB), and incubated for 15 min at 65°C. DNA was extracted twice with chloroform. The final water-phase was precipitated with iso-propanol and the DNA pellet was washed with 70% ethanol. The DNA pellet was air dried, dissolved in 100µl of TE, 5µl RNase (1mg/ml) added, incubated at 37°C for 30 min, and kept at -20°C for later use. The DNA quality was checked by

electrophoresis in 1% agarose gel and the concentration was estimated in relation to co migrating lamda-phage DNA and by repeated measurements with spectrophotometer at 260nm.

Primer selection

The 11 pair of primers described for the coffee genome SSR loci (Combes *et al*., 2000) (Table 2), were tested with representative genomic DNA from different forest coffee tree populations. The microsatellites were derived from a partial genomic library of *C. arabica* var. Caturra enriched for $(TG)_{13}$ motifs (Vascotto *et al*., 1999) (Table 2). Out of the 11 pair of primers, eight revealed polymorphisms and were selected for further analysis (Table3). In the present **study**, we used simple sequence repeats (SSRs) or microsatellites as genetic markers to indicate variation among forest coffee tree (*C. arabica*) populations. The goal was to preserve the maximum genetic diversity across the range of forest coffee (*C. arabica*) populations. The decision for selection of candidate population for in situ field conservation was mainly based on the relative gene diversity values (Figure 2 and Table 6) and the presence of unique allelic forms (Figure.3).

Amplification reactions

The amplification of the SSR loci consists of 25ng of genomic DNA, 1 × PCR reaction buffer without $MgCl_2$ (10mM Tris-HCl, pH 8.3, 50mM KCl), 0.1% Triton X-100, 2.5mM $MgCl_2$, 0.1µM of each primer, 0.2mM of each dNTP and 1 unit of Taq DNA polymerase (Sigma D-4545) in a final reaction volume of 25 µl. Reaction was performed in a Gene AMP® PCR System 9700 thermocycler (Biosystems, California, U.S.A) with amplification cycles consisting of an initial 2 min denaturation at 94° C, followed by 5 cycles of denaturation at 94° C for 45 s, 1 min primer annealing at 60°C with decreasing temperature of one degree at each cycle, 1½ min elongation at 72°C. Then, 30 cycles of 45 s at 90°C, 1 min at 55°C, 1½ min elongation at 72°C were performed. The last cycle was followed by 8 min final elongation at 72°C.

Gel electrophoresis of amplification products

CleanGel HyRes Plus System (CleanGel HyRes plus 36S and DNA-HyRes buffer) was purchased from ETC (Elektrophorese-Technik, Kirchentellinsfurt, Germany) for electrophoresis of microsatellite (SSR) PCR products. The gel was first rehydrated in DNA-HyRes buffer for 90 minutes. Each amplification product was thoroughly mixed with equal volume of loading buffer (25 ml DNA-HyRes buffer, 80 µl 1% (w/v) orange G, 40 µl 1% (w/v) xylene cyanol, and 250 µl 0.2 M EDTA). 7 µl of this mixture was loaded on the rehydrated gel and run on a horizontal Multiphor II Electrophoresis Unit at 200V and 35

mA, 375V and 50 mA, and 500V and 42 mA for 45, 15 and 95 minutes, respectively. A 50 base pair DNA ladder from Life Sciences was used to estimate the molecular size of the bands (alleles).

Band visualization

The gel was silver stained using a Hoefer Automated Gel Stainer (Pharmacia Biotech.) to visualize the bands with initial 40 min fixing in 0.6% benzene sulfonic acid and 24% ethanol, 3 × 10 min washing in 0.07% benzene sulfonic acid, 40 min silvering in 0.2% $AgNO_{3,}$ 0.07% benzene sulfonic acid and 0.05% formaldehyde, 2 min washing in distilled water, 5-6 min developing in 2.5% Na_2CO_3, 0.05% formaldehyde and 0.002% sodium thiosulfate followed by 3 × 10 min stopping and preserving in 10% acetic acid and 10% glycerol. The gel was then air dried for one to two hours and sealed in plastic to avoid gel desiccation and cracking.

Data analysis

Each single fragment generated by a locus specific primer pair was considered as an allele, and scored as presence (1) and absence (0) for binary data matrix. The binary data were used to compute all possible pair wise similarity matrices between samples and populations following Jaccard's similarity coefficient (Jaccard, 1908). The similarity matrices were subjected to cluster analysis, using UPGMA (unweighted pair-group method with arithmetic average) and plotted in a phenogram using NTSYS-pc, version 2.1 (Rohlf, 2000). Polymorphic information content (PIC) at each locus, or expected heterozygocity at each locus (Nei, 1987) was calculated according to Anderson *et al*. (1993). $PIC = 1-\Sigma p_i^2$, where p_i is the frequency of i^{th} allele of a locus. The summation extends and averaged over all alleles of a locus. Effective number of alleles per locus (A_{ep}) were also calculated according to Weir (1989) given as: $A_{ep} = 1/(1-H_{ep})$, where H_{ep} is the genetic diversity (expected heterozygocity) per locus, which is also equivalent to polymorphic information content (PIC) per locus.

Table 2. Repeat motifs, forward and reverse oligonucleotide primer sequences of 11 SSR loci

Locus	Repeat motifs	Forward (F) and Reverse (R) primer sequences from 5' to 3' ends
M_{2A}	$(GT)_8/(GT)_8/(GT)_7$	F: AGTGGTAAAAGCCGTTGGTG
		R. GCGGTTGTTGGTGAGTTGAA
M_3*	$(CA)_6/(CA)_3/(CA)_3/(CA)_3/(CA)_4/(CA)_3/(CA)_3/(CA)_3$	F: ATTCTCTCCCCCTCTCTGC
		R: TGTGTGCGCGTTTTCTTG
M_{11}*	$(GT)_4/(GA)_4/(GT)_4/(GT)_6$	F: CCCGAAAGAAAGAACCAAG
		R: CCACACAACTCTCCTCATTC
M_{20}	$(GA)_5/(GT)_8TT(GT)_4TT(GT)_7(GA)_{11}(TC)_2(CT)_3GT$	F: CTTGTTTGAGTCTGTCGCTG
		R: TTTCCCTCCCAATGTCTGTA
M_{24}*	$(CA)_{15}(CG)_4CA$	F: GGCTCGAGATATCTGTTTAG
		R: TTTAATGGGCATAGGGTCC
M_{25}	$(GT)_5CT(GT)_2/(GT)_{12}$	F: CCCTCCCTGCCAGAAGAAGC
		R: AACCACCGTCCTTTTCCTCG
M_{27}*	$(GT)_{11}$	F: GAGGGAGGTGTGGGTGAAG
		R: AGGGGAGTGGATAAGAAGG
M_{29}*	$(CTCACA)_4/(CA)_9$	F: GACCATTACATTTCACACAC
		R: GCATTTTGTTGCACACTGTA
M_{32}*	$(CA)_3/(CA)_3/(CA)_{18}$	F: AACTCTCCATTCCCGCATTC
		R: CTGGGTTTTCTGTGTTCTCG
M_{42}*	$(GT)_3/(GT)_7$	F: ATCCGTCATAATCCAGCGTC
		R: GGCCAGGAAGCATGAAAGG
M_{47}*	$(CT)_9(CA)_8/(CT)4/(CA)_5$	F: TGATGGACAGGAGTTGATGG
		R: TGCCAATCTACCTACCCCTT

Source: Comb *et al.,* 2000 (Asterisks are those loci, which showed polymorphisms in the present report.

Nei (1973) genetic diversity statistics (average genetic diversity for each population (H), the total genetic diversity (H_T), the within population genetic diversity (H_S), among populations genetic diversity (D_{ST}), and the coefficient of genetic differentiation (G_{ST}) were computed using POPGENE software 1.32 (Yeh *et al*., 1997). Shannon's information index was also computed for each population and the entire data according to Lewontin (1972) using POPGENE software 1.32 (Yeh *et al*., 1997). The partitioning of genetic differentiations into within population and between populations components were calculated based on Shannon's information content according to King & Schaal (1989).

Results

Out of the 11 primer pairs described by Combes *et al*. (2000) for coffee SSR loci analysis (Table 2), eight produced polymorphic products across 240 samples analyzed. The remaining three loci (M_{2A}, M_{20} and M_{25}) produced monomorphic amplification products. The eight SSR loci revealed a total of 97 alleles out of which 69 (71%) were polymorphic with a frequency range of 0.03–0.98. Monomorphic alleles were also scored for the SSR data analyzed. The number of alleles per locus ranged from 8 for M_3 to 18 for

M_{27} with a mean of 12.1 alleles per locus. Polymorphic information content (PIC) values for the eight loci varied from 0.19 for M_{29} to 0.74 for M_{27} with an average of 0.48 per locus (Table 3). Examples of SSR banding patterns are shown in fig.1.

Table 3. Number of alleles per locus, PCR product size range, effective number of alleles per locus (A_{ep}) and polymorphic information content (PIC) of the 8 SSR loci of *C. arabica*, which showed polymorphism in the present study.

Locus	Poly-, Mono- and total number of alleles.			Molecular size range		
Poly-	Mon-	Poly-	total	(b p)	(A_{ep})	(PIC)
M_3	07	01	08	280–400	1.8	0.45
M_{11}	06	04	10	050–350	1.7	0.42
M_{24}	10	-	10	150–350	2.6	0.62
M_{27}	08	10	18	045–350	1.5	0.32
M_{29}	05	06	11	130–400	1.2	0.19
M_{32}	09	-	09	075–300	2.9	0.66
M_{42}	10	06	16	120≥400	1.7	0.41
M_{47}	14	01	15	045–200	3.8	0.74
Mean	8.6	3.5	12.1	–	2.1	0.48

Table 4. Nei's (1973) genetic diversity statistics assuming Hardy Weinberg Disequilibrium with fixation index = 0.95.

Parameter	**Entire data**			
	H_T	H_S	D_{ST}	G_{ST}
Hardy Weinberg Disequilibrium	0.25	0.08	0.17	0.68

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

The results of Nei's (1973) genetic diversity statistics indicated that most of the variation resides between populations rather than within populations, which is in line with the genetic structure of inbreeding populations (Table 4). Self-pollinating species are characterized by a relatively high value of total gene diversity (H_T = 0.33), a low value of gene diversity within populations (H_S = 0.159), and a high value of the coefficient of gene differentiation (G_{ST} = 0.51) (Hamrick & Godt, 1989).

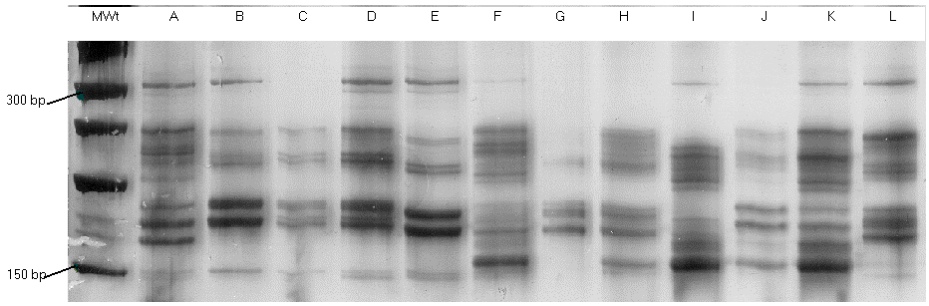

Fig.1. Banding patterns of SSR loci M_{24}.

Table 5. Partitioning of the genetic variation into within and between populations based on Shannon's information index.

Parameter	Mean
Hpop	0.11
Hsp	0.39
Hpop/Hsp	0.28
(Hsp − Hpop)/Hsp	0.72

Hpop = Mean genetic variation for the populations. Hsp = Mean genetic variation for the entire data. Hpop/Hsp = Proportion of genetic variation within populations. (Hsp − Hpop)/Hsp = Proportion of genetic variation between populations.

Gene diversity analysis with Shannon's information index for each population ranged from 0.03 for Jimma-1 population to 0.22 for Bench Maji-2 population with a mean of 0.11. The gene diversity index for the entire data was 0.39 (Table 6).

Shannon's information index revealed that proportion of genetic variation between populations (0.72) is greater than proportion of genetic variation within population (0.28) (Table 5). This result is in line with that obtained with Nei's (1973) genetic diversity statistics, which showed that most variation resides between populations than within population (Table 4).

Table 6. Measures of genetic diversity estimate in each population and the entire data of forest coffee tree populations, assuming Hardy Weinberg Disequilibrium with fixation index of 0.95.

Populations	Polymorphic alleles		Gene diversity	Shannon's index
Number	Number	Percent		
Welega-1	05	05	0.03	0.04
Welega-2	21	22	0.11	**0.15**
Welega-3	13	13	0.07	0.10
Welega-4	10	10	0.05	0.07
Welega-5	12	12	0.06	0.09
Ilubabor-1	09	09	0.05	0.07
Ilubabor-2	18	19	0.09	**0.13**
Ilubabor-3	09	09	0.05	0.07
Ilubabor-4	08	08	0.04	0.06
Ilubabor-5	21	22	0.11	**0.15**
Ilubabor-6	16	16	0.08	0.12
Jimma-1	04	04	0.02	0.03
Jimma-2	26	27	0.14	**0.19**
Jimma-3	14	14	0.07	0.10
Jimma-4	13	13	0.07	0.10
Kafa-1	17	18	0.09	**0.13**
Kafa-2	19	20	0.10	**0.14**
Bench Maji-1	12	12	0.06	0.09
Bench Maji-2	30	31	0.16	**0.22**
Bale-1	14	14	0.07	0.10
Bale-2	14	14	0.07	0.10
Bale-3	10	10	0.05	0.07
Bale-4	25	26	0.13	**0.18**
Bale-5	13	13	0.07	0.10
Mean	15	15	0.08	0.11
Entire data	69	71	0.25	0.39

Gene diversity and Shannon's information index (Figure 2 and Table 6) indicated that Bench Maji-2, Jimma-2, Bale-4, Welega-2, Ilubabor-5, Ilubabor-2 and Kafa -1 are candidate populations for *in situ* field conservation programmes, consideration respectively.

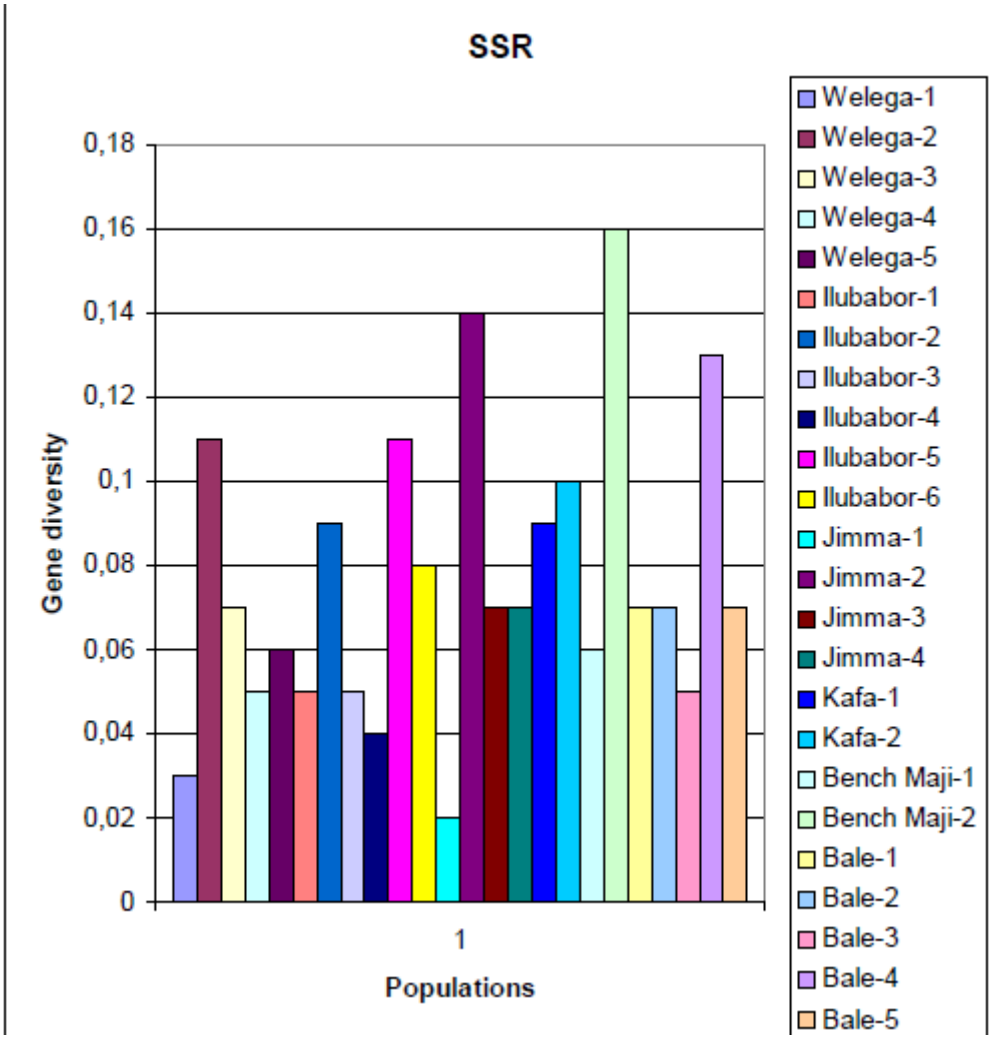

Fig. 2. The level of gene diversity within each population.

The phenogram derived from Jaccard's similarity coefficient between populations is shown in Figure 3. The phenohram produced a single tree in which some populations tend to cluster with respect to their region of origin, and slightly differentiated into east and west of the Great Rift Valley (GRV) except for few populations, which exhibited unique allelic forms.

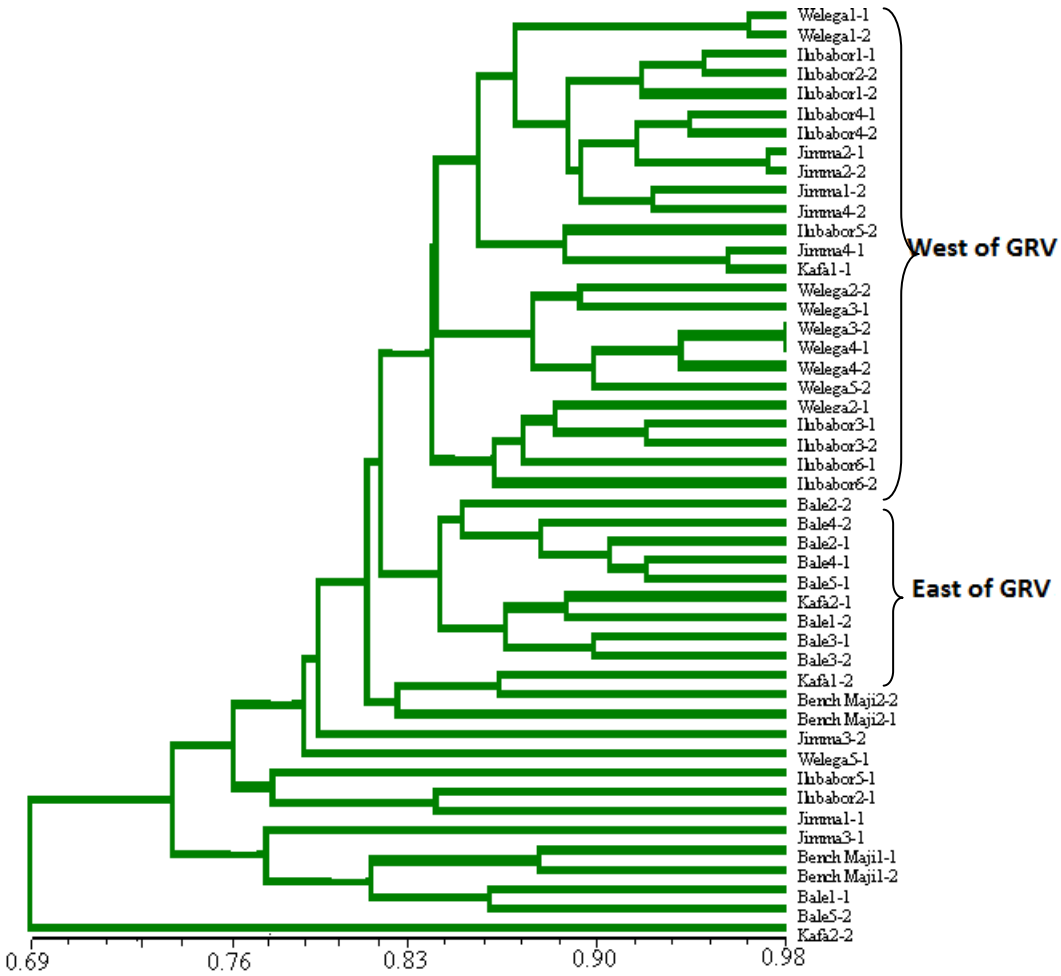

Fig.3. Phenerogram depicting relationships amongst the two samples from each of the forest coffee tree populations considered in the study. GRV = Great Rift Valley.

Discussion

The co-dominant nature of the SSR markers is one of the major advantages when compared to other PCR based markers, because it allows an estimation of allele relationship between individual genotypes. In the present report, the number of alleles per locus ranged from 8 to 18 with a mean of 12.1 alleles per locus. This value was considerably higher than the alleles per locus previously reported by Combes *et al*. (2000) and Anthony *et al*. (2002). One possible explanation could be because of differences in the number of samples analyzed and sample sources used. It is likely that sample size influences the number of alleles and genotypes observed, and the sample sources (forest coffee) used in the present

study represents unique sources of genetic diversities that were not observed in both cultivated accessions and other cultivated varieties of this species.

In the present analysis, the values of H_T, G_{ST} and H_S are 26, 0.65 and 0.09, respectively (Table 5). A comparable total gene diversity and Coefficient of gene differentiation were reported for similar materials with inter-simple sequence repeat (ISSR) markers (Aga *et al*., 2005). Anthony *et al*. (2001) also reported high level of total gene diversity (H_T = 0.25) and high coefficient of genetic differentiation (G_{ST} > 0.50) among groups of spontaneous and sub-spontaneous accessions of *C. arabica* from Ethiopia and the cultivated forms of the species (Typica and Bourbon) using RAPD (random amplified polymorphic DNA) markers.

Phenogram derived from Jaccard similarity coefficient among two samples from each population produced a single tree in which some samples tend to cluster with respect to their region of origin and a slight differentiation into east and west of the Great Rift Valley was observed except for few samples, which exhibited unique allelic forms (Figure 3). Comparable differentiation of *C. arabica* germplasm into east and west of Great Rift Valley was previously reported with RAPD marker analysis (Lashermes *et al*., 1996a; Aga *et al*., 2003, 2005; Aga and Bryngelsson, 2006). However, the presence of tectonic fault "Great Rift Valley", which crosses the country from South West to North East, does not account for the genetic differentiation into Western and Eastern C. arabica populations, and the southeastern coffee trees could have been introduced from southwest by humans or collected from southeastern forests (Montagnon & Bouharmont, 1996). Some remaining pocket primary forest ecosystems (Harana forest) may still contain the true wild forms of *C. arabica* from which local selection might have occurred.

Although ex situ field gene bank is a commonly used **as** alternative for conserving germplasm, of *C. arabica*, which is difficult to conserve as seed for a long period of time, it is expensive to maintain and the number of genotypes, which a field gene bank can hold is restricted by financial and land resources available, and this in turn limits the genetic diversity, which it can conserve. In addition, ex situ field gene bank conservation is characterized by risks of losing valuable germplasm due to diseases and pests as well as to poor adaptation of certain genotypes of the species to the local environment. In contrast to the ex situ field gene bank conservation, in situ conservation still remains an important component of the overall strategy for the long-term conservation of the coffee gene pool. *Ex situ* and *in situ* conservation strategies are complementary and should not be viewed as antagonistic

(Nevo, 1998). Conservation of genetic resources under *in situ* conditions would insure that evolutionary dynamic forces continue to influence plant adaptation and survival (Holden *et al*., 1993; Cohen *et al*., 1997; Maxted *et al*., 1997).

In situ conservation of forest or wild coffee (*C. arabica* L.) is important due to their agronomic values and the current critical deterioration of their habitat. Not all populations of a species are necessarily designated as in situ conservation sites. Characterization of the inter- and intra-population genetic variation may be used to determine which populations require special conservation status.

Conclusion

This study showed the usefulness of microsatellite markers for the analysis of genetic variability and the distribution of this variability among the 24 forest *C. arabica* populations from Ethiopia. Relatively large numbers of alleles were detected with few SSR markers, suggesting the multi-allelic nature of this marker. Besides providing information to design conservation strategies for forest *C. arabica* in Ethiopia, the results could be used in coffee breeding programs in order to cross genetically diverse parents and maximize the level of variation present in segregating populations. The SSR allelic variation detected in the present investigation could also serve in marker-assisted selection, in case it is linked to any of the important agronomic traits.

Acknowledgments

The authors thanked the Swedish International Development Agency (SIDA /SAREC) and BIO-EARN (East African Regional Program and Research Network for Biotechnology, Bio-safety and Biotechnology Policy Development) for sponsoring this project

References

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

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

Aga, E. & Bryngelsson T. 2006. Inverse sequence-tagged repeat (ISTR) analysis of genetic variability in forest coffee (*Coffea arabica* L.) from Ethiopia. Genetic Resources and crop evolution **53**: 7221-728.

Anderson, J.A., Churchil, G.A., Autrique, J.E. Tanksley, S.D. & Sorrells, M.E. 1993. Optimizing parental selection for genetic linkage maps. Genome **36**: 181– 86.

Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G. & Lashermes, P. 2002. The origin of cultivated Coffea arabica L. varieties reveled by AFLP and SSR markers. Theor. Appl. Genet. **104**, 894–900.

Byrne, M., Marquez-Garcia, M.I., Uren, T., Smith, D.S. & Moran, G.F. 1996. Conservation and genetic diversity of microsatellite loci in the genus Eucalyptus. Aus. J. Bot. **44:** 331–341.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., Cartinhour, S. 2000. Diversity of microsatellites derived from genomic libraries and Gene Bank sequences in rice (Oryza sativa L.). Theor. Appl. Genet. 100, 713–722.

Cohen, J.I., Williams, J.T., Plucknett; D.L. & Shands, H. 1997. Ex situ conservation of plant genetic resources: global development and environmental concerns. Science **253**, 866–872.

Combes, M.C., Andrzejewski, S., Anthony F., Bertrand, B., Rovelli, P., Graziosi, G. & Lashermes, P. 2000. Characterization of microsatellite loci in Coffea arabica and related coffee species. Mol. Ecol**. 9,** 1178–1180.

Coulibaly, I., Noirot, M., Lorieux, M. & Charrier, A. 2002. Introgression of self-compatibility from Coffea heterocalyx to cultivated species Coffea canephora. Theor. Appl. Genet. **105**, 994–999.

Creste, S., Tulmann-Neto, A., Vencovsky, R., D'Oliveira-Silva, S. & Figueira, A. 2004. Genetic diversity of musa diploid and triploid accessions from Brazilian banana breeding program estimated by microsatellite markers. Genetic Resour. Crop Evol. 51, 723–733.

Fregene, M.A., Suarez, M., Mkumbira, J., Kulembeka, H., Ndedya, E., Kulaya, A., Mitchel, S., Gullberg, U., Rosling, H., Dixon, A.G.O., Dean, R. & Kresovich, S. 2003. Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. Theor. Appl. Genet. **107**, 1083–1093.

Hamrick, J.L. & Godt, M.J.W. 1989. Allozyme diversity in plant species. In: A.H.D. Brown, M.T. Clegg, A.L. Kahler & B.S. Weir (eds) Plant population genetics, breeding & genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, USA. pp. 43–63.

Holden, J., Peacock, J. & Williams, T. 1993. Genes, Crops, and the Environment. Cambridge University Press, Cambridge, United Kingdom.

Jaccard, P. 1908. Novelles recherches sur la distribution florale. Bul. Soc. Vaud. Sci. Nat. **44**: 223–270.

Jarne, P. and Lagoda, P.J.L. 1996. Microsatellites from molecules to populations and back. Trends Ecol. Evol. **11**: 424–429.

King, L.M. & Schaal, B.A. 1989. Ribosomal-DNA variation and distribution in Rudbeckia missouriensis. Evolution **43**: 1117–1119.

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

Lashermes, P., Comes, M.C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F. & Charrier, A. 1999. Molecular characterization and origin of the Coffea arabica L. genome. Mol. Gen. Genet. **261**: 259–266.

Lewontin, R.C. 1972. The genetic basis of Evolutionary Change. Columbia Univ. Press.

Maxted, N., Ford-Lloyd, B.V. & Hawkes, J.C. 1997. Plant Genetic Conservation. The in situ Approach. Chapman and Hall, London.

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

Morgante, M. & Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**: 175–182.

Nei, M. 1973. Analysis of genetic diversity in subdivided populations (population structure/ genetic variability/ hetrozygosity/ gene differentiation). Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

Nei, M. 1987. Molecular evolutionary genetics. Colombia University Press. New York.

Nevo, E. 1998. Genetic diversity in wild cereals: regional and local studies and their bearing on conservation ex situ and in situ. Genetic Resour. Crop Evol. **45**, 355–370.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. & Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. **2**, 225–238.

Rohlf, F.J. 2000. NTSYS-pc, Numerical taxonomy and multivariate analysis system, version 2.1, Exeter Software, New York.

Song, Z.P., Xu, X., Wang, B., Chen, J.K. & Lu, B.-R. 2003. Genetic diversity in the northernmost Oryza rufipogon populations estimated by SSR markers. Theor. Appl. Genet. **107**, 1492–1499.

Strenge, von H. 1956. Wild Coffee in Kaffa Province. Tropical Agriculture Trindad 33, 297–301.

Sylvain, P.G. 1955. Some observations on Coffea arabica L. in Ethiopia. Turrialba **5**, 37–53.

Tautz, D. 1988. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17**: 6463–6471.

Van der Graaff, N.A. 1981. Selection of Arabica coffee types resistant to coffee berry disease in Ethiopia. Doctoral thesis, University of Wageningen, Netherlands, pp. 110.

Vascotto, F., Gegli Ivanissevich, S., Rovelli, P., Anthony, F., Anzueto, F., Lashermes, P. & Graziosi, P. 1999. Microsatellites in Coffea arabica: construction and selection of two genomic libraries. In: Proc 3rd Int Biotechnol Coffee Agroind. Londrina, Brazil, pp. 125–128.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Plot, J., Peleman, J., Kuiper, M. & Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23**, 4407 – 4414.

Wang, M., Oppedijk B.J., Lu, X., Duijn, B.V. & Schilperoort, R.A. 1996. Apoptosis in barley aleurone during germination and its inhibition by abscisic acid. Plant Mol. Biol. 32, 1125–1134.

Weir, B.S. 1989. Sampling properties of gene diversity. In: A.H.D. Brown, M.T. Clegg, A.L. Kahler & B.S. Weir (eds) Plant population genetics, breeding & genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, USA. pp 43–63.

Welsh, J. & McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18, 7213–7218.

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey. S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18**, 6531–6535.

Yeh, F.C., Yang, R.C., Boyle, T.J., Ye, Z.H. & Mao, J.X. 1997. POPGENE 1.32, the user friendly shareware for population genetic analysis, Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, Canada.

Yu, S.B., Xu, W.J., Vijayakumar, C.H.M., Ali, J., Fu, B.Y., Xu, J.L., Jiang, Y.Z., Marghirang, R., Domingo, J., Aquino, C., Virmani, S.S. & Li, Z.K. 2003. Molecular diversity and multi locus organization of the parental lines used in the International Rice Molecular Breeding Program. Theor. Appl. Genet. 108, 131–140.

Zietkiewicz, E., Rafalski, A. & Labuda, D 1994. Genome fingerprinting by simple sequence repeat (SSR) anchored polymerase chain reaction amplification.