Egypt. J. Plant Breed. 21(4):713– 736 (2017)

GENETIC DIVERSITY, STRUCTURE AND DNA FINGERPRINT FOR DEVELOPING MOLECULAR IDs OF YEMENI COFFEE (*Coffea arabica* L.) GERMPLASM ASSESSED BY SSR MARKERS

M.A.A. Hussein[1], A.A.A. Al-Azab[2, 4], S.S. Habib[2], Fadia M. El Sherif[2] and Hoda A.S. El-Garhy [3]

1. Dept. of Botany, Fac. of Agri., Suez Canal Univ., Egypt.
2. Dept. of Horticulture, Fac. of Agri., Suez Canal Univ., Egypt.
3. Dept. of Genetics, Fac. of Agri., Benha Univ., Egypt.
4. Ministry of Education, Republic of Yemen.

ABSTRACT

This study was carried out to determine the genetic diversity degree and genetic relationships among seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (Coffea arabica L.) germplasm collected from different Governorates in Yemen, and analysis of the DNA fingerprinting data for creating molecular IDs for conservation and protection of these genotypes. These goals were done using 15 previously described SSR primer pairs, in addition, evaluating of the efficiency and performance of these loci to achieve these objectives. The SSR loci were very highly polymorphic with an average of 100% polymorphism, the scored alleles were high and ranged from 4 to 25 with a mean value of 10.7 per locus. Heterozygosity values per locus and per genotype were low with an average of 0.21., the (PIC) values or gene diversity differed from 0.57 to 0.98. Also, the discriminating power for all loci was high with a mean value of 0.81; the most informative primer pairs was gSSRCa 021 with DP value of 0.94. However, the probability of matching fingerprints was low with an average value of 0.19. In this study, it was impossible to obtain identical DNA fingerprints for any of genotype pairs at all loci, even the genotypes with the same name but were collected from different geographical regions. However, the 15 SSR loci could differentiate between 17 Yemeni coffee genotypes through detecting of 86 clear specific/exclusive alleles. All Cultivars and accessions had unique alleles to that genotype alone with a mean value of 5.06. The cluster analysis grouped the 17 genotypes into three main clusters with a genetic similarity degree ranged from 0.00 to 0.486 with an average of 0.243. The genotypes also were classified according to fruit color and geographical regions. Results of this study confirmed the presence of genetic diversity among Yemeni coffee genotypes ranging between a moderate to high, indicating their importance as source of genetic variability for the purposes of coffee improvement, especially with the suffering of Arabica coffee from the narrowness of its genetic base.

Key words: ***Yemeni Arabica coffee, Coffea arabica, Genetic diversity, Conservation, DNA fingerprint, SSR markers.***

INTRODUCTION

Coffee is an important crop, which is the main source of income for millions of people living in developing countries and the second most traded commodity in the world after petroleum oil (ICO 2013). Increasing the productivity of coffee cultivars with improved the quality is the main objective of coffee genetic improvement programmes. Coffee belongs to the Rubiaceae family, which is mostly grown in tropical and subtropical regions

(Berthaud and Charrier 1998) and consists of 124 species spread across two genera, *Coffea* L. and *Psilanthus*.

Commercial coffee production depends mainly on two species belonging to the Coffea genus, *Coffea arabica* L. (Arabica coffee) which is responsible for the production of at least 70 % of the world's production and 30% from *Coffea canephora* (Robusta coffee). Also *Coffea arabica* is considered to be of excellent quality depending on the low caffeine content and taste than C. *canephora.* In spite of that, improving the quality of both species is a target for coffee improvement programmes by searching for genes responsible for control of the levels of the major biochemical components (Tran *et al* 2016). All species of Coffee are diploid (2n =2x=22 chromosomes) and self-incompatible, excepting *Coffea arabica* which is a self pollinated and allotetraploid (2n = 4x = 44 chromosomes) derived from hybridization between *Coffea canephora* and

Coffea eugenioides (Lashermes *et al* 1999). The genome size of the two main domesticated coffee species has been predestined to be 1.3 Gb for *C. arabica* (Kochko *et al* 2010) and 710 Mb for *C. canephora* (Denoeud *et al* 2014) and the DNA content of the *Coffea arabica* was 2.47 pg while *canephora* was 1.43 pg. Many of the previous studies suggest that the origin *of* Arabica coffee and its primary centre of genetic diversity in highlands of South West Ethiopia (Sylvain 1955), also there are records of wild Arabica coffee in the Boma Plateau of Sudan and Mount Marsabit of Kenya (Thomas 1942 and Anthony *et al* 1987). However, a few of *Coffea arabica* plants could have been introduced to Yemen from Central Africa, either early in 575 A.D. (Wellman 1961) or since three to four centuries ago (Eskes 1989). Then progenies of single plant introduced from Indonesia and cultivated in Europe in the Amsterdam botanical gardens were spread out to America, and it was the genetic basis for typica variety. In the same context, historical data indicated that the genetic base for the second botanical variety called *Coffea arabica* var. *bourbon* also originated from a few coffee plants introduced from Mocha in Yemen to Bourbon Island between 1715 and 1718, subsequently, increasing the number of commercial cultivars of Arabic coffee in Brazil and around the world (Chevailier and Dagron 1928, Carvalho 1946 and Haarer 1956). Noteworthy, that the Bourbon-derived cultivars are characterized by higher yield and better quality than the Typica-derived cultivars (Carvalho *et al* 1969). It was clear from many of the previous studies that the Arabica coffee plants in the main coffee producing countries such as Asian countries (Yemen) and Latin America have a narrow genetic basis. It is possible that attributed to the few seeds per plants used for dissemination, whereas the reduction of successive genetic variation due to human impacts and reproduction nature, especially with the nature of self-pollination in the Arabica coffee (Anthony *et al* 2002 and Lashermes *et al* 1996). Through comparisons the genetic diversity for

the populations of *Coffea arabica*, it is less polymorphic compared to diploid species. However, the populations of Arabica coffee in its place of origin and centre of genetic diversity in the southwestern highlands of Ethiopia have a lot of genetic variability for many of characters compared to, the cultivated materials which spread world-wide from Yemen. This is agreed by many previous studies through different ways such as morphological (Carvalho 1988), biochemical (Sivarolla *et al* 2000) and based on the DNA-based molecular markers techniques (Lashermes *et al* 1996). All these findings support hypothetical that Arabica Coffee plants transferred to Yemen for cultivation by Arabs have been originated from the parts of south eastern of evergreen mountainous region of Ethiopia. Native cultivars of Arabica coffee in Ethiopia are clearly adapted to different places for planting with existence of wide genetic variability. Increasing the genetic base has become a priority for further crop improvement and has induced several collection missions. Despite attempts by the FAO mission through the period 1964 – 1965 to collect spontaneous and subspontaneous coffee germplasm in probable native regions of the species, in order to face this narrow genetic basis and for genetic resources conservation (FAO 1968). In spite of, all these previous studies and attempts to evaluation coffee germplasm collection regarding botanic, agronomic and technologic sides, the genetic diversity and structure of this collection is incomprehensible. While the use of spontaneous and subspontaneous plants will play important rule in the future of *Coffea arabica* breeding programs.

A lot of previous studies have already evaluated the genetic diversity and structure of some Ethiopian and FAO accessions of Coffee with a few accessions from Yemen. These studies were conducted using different techniques that ranged from traditional morphological techniques to the modern DNA-based molecular markers. However, use of morphological techniques through morphologic and agronomic characteristics (Montagnon and Bouharmont 1996) is restricted by the influence of the environmental factors and growth stage. While, the more effective techniques dependent on the protein and isozymes were developed but, their application was limited, due to the fact that these markers were not efficient to detect the genetic differences within the species of Arabica coffee (Berthaud and Charrier 1998). Therefore, the importance of a number of DNA-based molecular markers in genetic studies of coffee including RFLP and different PCR-based methods such RAPD (Lashermes *et al* 1996 and Anthony *et al* 2001) AFLP (Anthony *et a*. 2001a and Steiger *et al* 2002), ISSR markers (Aga *et al* 2005) and microsatellites markers (SSRs) (Moncada and McCouch 2004). However, such studies reported different outcomes regarding grouping and genetic relationships of those accessions which collected in different regions from Yemen and Ethiopia. These recent molecular markers have many advantages; they are not affected by the

environmental factors and potential of existing in unlimited numbers, covering the whole genomes. Of these techniques, microsatellite (SSRs) markers, which, are used recently in the genetic studies. It is short sequence repeats of DNA from one to six base pairs. It is also characterized by advantages more than other molecular markers such as; it is highly polymorphic, reproducible, specific markers and co-dominant. Each of these features, today SSRs are the markers of choice for many genetic studies. Recently, the SSRs markers were used to estimate the genetic structure and diversity within the gene pool of cultivated Arabica coffee, because it has shown to be useful for breeding purposes even though small. Good to mention that, SSRs are present in the coding regions (c DNA sequences); it was marked with an EST-SSR (expressed sequence tag-SSR). While, SSRs which present in non-coding regions (g DNA sequences), named g SSR (Gur-Arie *et al* 2000). Previous studies have been conducted to estimate the level of polymorphism within *C.arabica* accessions using EST-SSR markers. Low level of polymorphism has been shown compared to *C. canephora.* accessions. It was explained by the narrow genetic base and autogamy in *C.arabica* (Missio *et al* 2009b). In addition to, as mentioned previously those EST-SSR markers are designed from highly conserved genomic regions. It is expected to show low level of polymorphism compared to gSSR markers, which are scattered in abundance in the genome (Varshney *et al* 2005). By using molecular markers developed for *C. arabica*, as quicker and accurate tools, in addition to high-resolution for discrimination between the individual genotype with reliable results. Moncada and McCouth (2004) showed the particular value of microsatellite markers for DNA fingerprinting and differentiate between closely related commercial cultivars of Coffee.

For more economic benefits, that's where the Yemeni coffee is of high quality. So it is very important to the future strategic planning, for the purpose of conservation and protection of these cultivars. Especially, it is the genetic basis for the origin of the cultivated cultivars around the world. Hence, one of the goals of the current study is to analyze of the DNA fingerprinting data for creating molecular IDs. This could be important to registration, protection and conservation of Yemeni coffee germplasm.

Since, the optimum utilization of available genetic differences in Yemeni coffee for breeding programs is depended on the availability of genetic diversity information. Consequently, any effort towards creation information on the genetic diversity and genetic relationships among Yemeni coffee are very important, especially using DNA molecular markers. Despite their cost, it is highly desirable in coffee genetic improvement programs, especially Arabica coffee that suffers with low genetic base.

Thus, the second purpose of this study was to evaluate the effectiveness of SSR markers, as accurate method in detecting the genetic diversity among 17 cultivars and accession of Yemeni coffee collected from different Governorates in Yemen and analysis of population structure to classify them according to genetic relationships and genetic similarity degrees.

MATERIALS AND METHODS

Plant materials and DNA extraction

A total of seventeen cultivars and accessions of Yemeni coffee (*Coffea arabica* L.) were obtained from Al-mahaweet, Sana'a, Sadah, Taiz, Lahj and Ibb Governorates, and used in the present study (Table 1). These genotypes included sixteen cultivars and one accession and characterized by two colors of fruit. These genotypes are representing most of the main cultivars grown in Yemen and then this study included seventeen of DNA samples, three samples of the same cultivar which is called "Odayni" and two DNA samples from the same cultivar which is called "Odayni- Bayad", But from different geographical locations to help in pointing to the influence of diverse geographical location on the emergence of differences at the molecular level using specific SSR markers. The total DNA of each sample was extracted from frozen young leaves of a single plant, according to the CTAB method (Rogers and Bendich 1985). with slight modifications according to (Porebski *et al* 1997) for obtaining good quality total DNA as follows: 2 g of young leaves were ground in liquid nitrogen to fine powder and extracted by the addition of 10 ml preheated (65ºC) from cetylhexadecyl-trimethyl ammonium bromide (CTAB) extraction buffer [3% CTAB (w/v), 100 mM Tris- HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2% (w/v) PVP (Polyvinyl pyrrolidone)], and furthermore 1% (v/v) of β-mercaptoethanol (15 mM) with further grinding. The mixture was incubated at 65ºC for 60 min, then two extractions with chloroform/isoamyl alcohol (24:1). After which the nucleic acids precipitated with Cold isopropanol, and the pellet was dissolved in 1 mL TE 0.1X (Tris-EDTA) buffer. RNA was removed with RNaseA 1 µl (10 mg/mL). The DNA was purified further by 300 µl phenol: chloroform: isoamyl alcohol (25:24:1), then overnight at (-20 ºC) using 1/10 vol. from 2 M Na acetate (pH 8.0) and one volume of cold isopropanol alcohol. The precipitated was washed twice, and the pellet was dissolved in 0.1X TE buffer. The purified total DNA was quantified by gel electrophoresis, and its quality assured by spectrophotometer.

SSRs markers selection and PCR amplification

The fifteen SSR primer pairs used for PCR amplification in this study, belongs to two sets (Table 2). The first, containing 14 primers pairs and each described previously by (Missio *et al* 2009a), and others were designed by (Vieira *et al* 2010, Combes *et al* 2000, Rovelli *et al* 2000 and Baruah *et al* 2003).

Table 1. List the seventeen cultivars and accessions of Yemeni coffee (*Coffea arabica* L.) with their Collection area and the color of fruit.

Genotypes	Collection area	Source*	Type	Fruit color
Odayni – IbbG[a]	Ibb Governorate	ARABSUR	Cultivated-cultivar	Red
Odayni – AlmG. [c]	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Red
Odayni – SadG. [b]	Sadah Governorate	YGRCSU	Cultivated-cultivar	Red
Burai	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Red
Benan	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Red
YEMAM1[d]	Al-Mahaweet Governorate	YFO	Cultivated-local type	Red
Odayni - Bayad L1[e]	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Yellow
Odayni - Bayad L2[f]	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Yellow
Benan - Bayad	Al-Mahaweet Governorate	YFO	Cultivated-cultivar	Yellow
Shubriqi	Sana'a Governorate	YFO	Cultivated-cultivar	Red
Ismaili	Sana'a Governorate	YFO	Cultivated-cultivar	Red
Dawairy	Sana'a Governorate	YFO	Cultivated-cultivar	Red
Tofahy	Sana'a Governorate	YGRCSU	Cultivated-cultivar	Red
Gadi	Sana'a Governorate	YGRCSU	Cultivated-cultivar	Red
Oubale	Taiz Governorate	ARABSUR	Cultivated-cultivar	Red
Abu Sura	Taiz Governorate	ARABSUR	Cultivated-cultivar	Red
Tasawai	Lahj Governorate	ARABSUR	Cultivated-cultivar	Red

*** Refers to source of samples from where the sample materials were obtained.** Whereas **ARABSUR = Agricultural Research Authority Branch for Southern Uplands (TAIZ-IBB in Yemen), YGRCSU = Yemeni Genetic Resources Center of Sana'a University and YFO = from** Yemeni Farmers' **Orchards.**

[a, b] and [c] refers to that, "Odayni" cultivar samples was collected from three locations: Ibb, Al-Mahaweet and Sadah Governorates.

[d] refers to the Yemeni coffee accession which are cultivated in Al-Mahaweet Governorate, we classify it as Cultivated- local type.

[e] and [f] refers to that, "Odayni- Bayad" cultivar were collected from two different locations in Al-Mahaweet Governorate.

Fig. 1. A general map of the Republic of Yemen showing collecting sites of the samples in the current study, which referred to as black circles.

This first set was obtained from an enriched genomic library of *Coffea arabica* Bourbon-derived genotype Amarelo. While, the second set included one primer pair only, was isolated from *Coffea arabica* ESTs of the Brazilian coffee Genome Project by (Missio *et al* 2009b). These fifteen SSR primers were synthesized by Oligo Macrogen, Seoul, korea. PCR reactions were performed according to published procedures by (Plechakova *et al* 2009). PCR amplification was done on a total volume of 25 µl reaction mixture composed of 2 µl DNA template (50 ng/µl), 5 µl of 5x PCR Buffer (Promega), 2.5 µl dNTP`s (2 mM), 2 µl of each primer (10 pmol/µl), and 0.2 µl Taq DNA polymerase (5 U/µl) and then the final volume was adjusted using dd.H2O. The amplification was performed on a thermocycler (Eppendorf Master Cycler Gradient Eppendorf) as follows: initial denaturation step at 94°C for 4 min, 35 cycles of amplification (40s of denaturation at 94°C, 60s of annealing at 54°C to 69°C based on the optimal annealing temperature for the used primer, 60s of extension at 72°C), and followed by final extension at 72°C for 10 min.

Data scoring and analysis

The gels were scored as presence or absence of alleles or bands. Because of, the Arabica coffee was described as allotetraploid species, the

SSR markers can amplify more than two alleles per locus, where it is able to amplify alleles from the two origins: *C. eugenioides* and *C. canephora* genome. Hence, these alleles in the Arabica coffee can be used to determine the origin of this species depending on the size of alleles in the two diploid species.

The analyses were repeated at least twice to assure the reproducibility of the products. The PCR products of the Microsatellite markers were detected by polyacrylamide gel (8%) to separate PCR products with different sizes, and then stained with ethidium bromide solution and visualized under gel documentation model (Gel-Doc 2000 with Diversity Database software Ver. 2.1, Bio-Rad Laboratories, Hercules, California, USA) for gel analysis. With the observation that in this case, it is impossible to differentiate between diallelic duplex, simplex and among different types of triallelic combination of SSR loci. Consequently, the fragment frequencies were analyzed as multiloci and each allele was scored present or absent (1/0). All the scored alleles were used to calculate the following genetic parameters: the number of alleles per SSRs locus without Zero alleles (the alleles in case that no amplification product in one or more genotypes), maximum number of alleles per genotype, percent of polymorphic bands per SSR locus, average polymorphism was calculated by dividing the number of polymorphic bands by the total of amplified bands for all loci, the number of exclusive (unique) alleles per genotype, observed heterozygosity (Ho, direct count) was calculated by dividing the number of heterozygous genotypes per locus to the total number of genotypes, the polymorphism information contents (PIC), which also known as heterozygosity index, was calculated for each locus depends on number of alleles and the allele frequency from the formula: $PIC = 1- \sum pi^2$ where pi is the frequency of each allele, with taking into consideration, the alleles which have introduced from diploid species into tertraploid Arabica coffee.

If PIC value calculated in this way, it is similar to the expression 'gene diversity' as described by (Nei 1973). The power of discrimination (PD) was calculated as above formula for each locus, but the allele frequency was replaced by the genotype frequency, according to (kloosterman *et al* 1993). The number of effective alleles (NE) was calculated for each marker according to (Hart and Clark 1997) using the formula: $NE = 1/\sum (E/F)^2$, where E refers to the total number of genotypes at each allele of locus *i*, and F the total number of alleles of the locus *i* in all genotypes. The probability of matching fingerprints was estimated according to (Jones 1972). In addition, the heterozygosity level within each genotype was used. All previous calculations of genetic values were performed with the programs GENEPOP version 1.31 (Raymond and Rouset 1995), Quantity one and Microsoft Excel.

Recorded data as a presence/absence (1/0) were used to calculate the similarity degree according to Dice coefficient (Sneath and Sokal 1973), using the SPSS software ver. 16. The dendrogram was produced by clustering the data with the average Linkage (Between Groups) utilizing all detected alleles to estimate the genetic relationship and similarity among the Yemeni coffee genotypes under study.

RESULTS AND DISCUSSIONS

Genetic diversity and characterization of SSR loci

Fifteen SSR loci produced a total of 160 bands or alleles across 17 Yemeni coffee (Coffea *arabica* L.) genotypes (Table 3). All tested SSR markers were polymorphic and detected percentage of 100% polymorphism among the evaluated cultivars and accessions. The number of observed alleles per locus ranged from four in loci gCarM048 and gE8-3CTG to twenty five in locus gSSRCa 021 with an average of 10.7 alleles per SSRs locus. However, different mean numbers of alleles were previously reported in different *Coffea* species: 2.2 alleles per primer with sixteen SSR primers pairs among 13 accessions of *C. arabica* from Yemen by (Silvestrini *et al* 2007), 10 alleles per primer with 11 gSSR among 10 genotypes of *C. canephora* by (Combes *et al* 2000) and 2.87 alleles using 23 gSSR in 28 inbred lines of Arabica coffee and *C. canephora* by (Maluf *et al* 2005). In general, average numbers of alleles in diploid coffee species were similar or often lower than those observed in *Coffea arabica* as reported (Silvestrini *et al* 2007), possibly due to the allotetraploid nature of Arabica coffee which leads to increase the number of alleles per genotype. In a related context, the present study is in agreement with previously reported by (Misso *et al* 2011), of that gSSR showed a higher number of alleles than did the EST-SSR loci. The number of SSR alleles assessed in any study depends on the sample size, genetic variability of the samples and the number of SSRs used in addition their genome coverage. Where, the larger coverage of the whole genome is one of the limiting factors in diversity study of coffee due to lack of sufficient number of SSR markers. Also the variation in the number of alleles in an SSR locus be a result of the differences in mutation rates on each locus as reported (Metais *et al*., 2002), which are influenced by the structure, length of the repeated tandem and type of the locus whether it is gSSR or EST-SSR. In this study, locus gSSRCa 021 was produced high number of alleles (25 alleles) compared to the same locus in the previous study by (Missio *et al* 2009a) which was produced only four alleles. One reason for such difference could be due to the small sample size and type of used coffee genotypes in the previous study.

In this study, the SSR markers detected high percentage of polymorphism (100 %) within the studied genotypes. It may occur due to a low conservation and high variability of the DNA sequences in non coding regions of the genome (Varsheny *et al* 2005). High percentages of

polymorphism with gSSR markers in this study was also reported for coffee by (Misso *et al* 2011), may also have influenced by the number of repeated nucleotides where most SSR loci in our study are consist of repeated dinucleotide (Table 2). It was reported that SSR markers with trinucleotides were less polymorphic than with dinucleotide in wheat (Gadaleta *et al* 2007). The high polymorphism which was found here also observed in previous studies (Baruah *et al* 2003 and Silvestrini *et al* 2007), and it may be mainly attributed to the diverse parentage background. From the total of 15 SSR markers three (gSSRCa 018, gSSRCa 021 and EST-SSR 073) showed more than two alleles (4, 3, 3) per genotype, respectively most of them are specific alleles for those cultivar. Hence, these alleles in the allotetraploid Arabica coffee can be used to determine the origin of this species, depending on their size comparison with the size of alleles in the two diploid species *C. eugenioides* and *C. canephora* genome from previous studies. However, it was difficult due to size overlapping for the alleles of the two genomes. In this study locus (EST-SSR 073) was produced 12 alleles from the 17 Yemeni coffee genotypes (Table 3). Of the total 12 alleles, 3 alleles were detected in Tasawai cultivar and were specific for this cultivar. One size only (84 bp) of these introgressed specific alleles is similar for those amplified alleles in *C.canephora* species with the same locus in previous study by (Missio *et al* 2009b). While the rest two alleles which was produced from this locus (256 and 400 bp) or even the alleles from the other two loci (gSSRCa 018, gSSRCa 021) showed different size range compared to the published size range (Table 2). In this regard, locus gSSRCa 018 was produced four and three alleles with Ismaili cultivar and the accession respectively, of these alleles three and one is described as specific alleles respectively. However, locus gSSRCa 021 was produced three alleles with Oubale and Odayni AlmG cultivars. Maybe these specific alleles have introduced from *Canephora* (Robusta) species to Yemeni coffee genotypes during the introgression process to development of their agronomic characters. This result is in agreement with historical data which indicated that the Arabica coffee derived from hybridization between *Coffea canephora* and *Coffea eugenioides* (Lashermes *et al* 1999). Effective number of alleles (Ne) is the measure of allelic evenness. In our study, the results showed that the effective number of alleles (Ne) ranged from 2.34 to 49.8 with average value of 11.09 (Table 3). According to the selective standard of the microsatellite loci, it ought to have at least four alleles to be considered useful for the evaluation of genetic diversity. Bases on this criterion, the 15 microsatellite loci in our study were useful for the evaluation of genetic diversity in the 17 Yemeni Coffee genotypes. This high (ne) values from the reality that more than two alleles were observed in some individual genotypes with some loci.

Table 2. List of 15 microsatellite (SSRs) markers used, locus code and their nucleotide sequence.

SSR locus code	Primer sequence (5' -3')	Repeated sequence	Published size range (bp)*	Fragment size range (bp)	Tm (C°)
gE8-3CTG	F: CACTGGCATTAGAAAGCACC R: GGCAAAGTCAATGATGACTC	$(CA)_{14}$	198	170 - 237	56
gM24	F: GGCTCGAGATATCTGTTTAG R: TTTAATGGGCATAGGGTCC	$(CA)_{15}(CG)_4$ CA	132 - 166	204 - 292	55
gCM2	F: TGTGATGCCATTAGCCTAGC R: TCCAACATGTGCTGGTGATT	$(AC)_{10}$ $(AT)_9$	177 - 229	146 - 288	56
gSSR Ca 006	F: CTTGCTCAGTGAACCATCC R: TGCCTCTTATGCCACTACTAAA	$(CT)_6$	209	250 - 340	57
gSSR Ca 018	F: GTCTCGTTTCACGCTCTCTC R: ATTTTTGGCACGGTATGTTC	$(GT)_{18}$ $GA)_{10}$	115	70 - 110	54
gSSR Ca 021	F: GCTGAGAGTTTTGAGGGAAA R: CCGACGTAGTTGATGATTGA	$(GCA)_3$ N_4 $(AAG)_2$	232	276 - 600	56
gSSR Ca 034	F: TGGACAAGAAATTGAAGTGG R: GGGTTTAAATTATCGGGTGT	$(CT)_3$ N_3 $(CT)_5$ $(TC)4(AC)_2$	257	115 - 434	54
gSSR Ca 055	F: AAGGAAAACAACACCCAAGA R: CGAGACAAGAGAGGGGAAA	$(ATC)_3$.$(AAGG)_2N_9(CT)_4N_6(CT)_3$	294	290 - 486	54
gSSR Ca 065	F: ATCTAACAAAATCCCCGTCA R: ATCGGTCGCCCTTCTAAT	$(AG)_2(AAG)_3$	142	50 - 182	54
gSSR Ca 085	ATGTGAAAATGGGAAGGATG CACAGGAAAGTGACACGAAG	$(TC)_{24}$	105	190 -276	54
gSSR Ca 087	F: TCACTCTCGCAGACACACTAC R: GCAGAGATGATCACAAGTCC	$(TC)_{22}$	143	86 - 162	58
EST-SSR 073	F: GAGGTCTTCCCACCACAACA R: GGATACGAGAGTCCCTTCCA	$(CGC)_2(GGC)_5$	160 - 175	66 - 400	60
gCar M048	F: CCAGCAATCCTCCCTCCCACCAC R: TACCGTATGCAGAGACAACAATG	$(GA)_{24}$	280 - 310	82 - 146	61
gCar M051	F:GATGTGGAGGAGGCTGCTGCTGAA R: TAGGGCGCCATCTGGTAGGGTTGT	$(CT)_{10}$	250 - 300	58 - 330	69
gCar M101	F: TATGTCTCTAACTTTCCTATTTT R: AGAGACTACATTTACACAGAAGA	$(CT)_{28}$	170 - 210	138 - 220	54

***Published size range refers to the published microsatellite sequence in (Rovelli *et al* 2000, Combes *et al* 2000, Baruah *et al* 2003, Missio *et al* 2009a, Missio *et al* 2009b and Vieira *et al* 2010).**

Table 3. Characterization and comparison of the PCR products for 15 microsatellites (SSRs) markers in 17 of Yemeni coffee genotypes.

NO	SSR locus code	Total alleles	polymorphic	Maximum Number of alleles/Genotype	Ne[a]	Ho[b]	PIC[c]	Matching of fingerprints[d]	DP[e]	Number of exclusive alleles
1	gE8-3CTG	4	4	1	3.30	0.00	0.70	0.30	0.70	1
2	gM24	9	9	1	5.45	0.00	0.82	0.40	0.60	5
3	gCM2	7	7	1	5.25	0.00	0.81	0.19	0.81	3
4	gSSR Ca 006	7	7	1	4.74	0.00	0.79	0.39	0.61	3
5	gSSR Ca 018	11	11	4	6.09	0.35	0.84	0.16	0.84	5
6	gSSR Ca 021	25	25	3	49.8	0.94	0.98	0.05	0.95	17
7	gSSR Ca 034	15	15	2	17.3	0.23	0.94	0.07	0.93	9
8	gSSR Ca 055	14	14	1	19.3	0.05	0.95	0.08	0.92	10
9	gSSR Ca 065	9	9	2	6.15	0.76	0.84	0.19	0.81	2
10	gSSR Ca 085	12	12	1	13.8	0.00	0.93	0.15	0.85	7
11	gSSR Ca 087	10	10	1	8.76	0.00	0.89	0.13	0.87	5
12	EST-SSR 073	12	12	3	6.75	0.23	0.85	0.14	0.86	7
13	gCar M048	4	4	2	2.34	0.05	0.57	0.40	0.60	2
14	gCar M051	12	12	2	11.6	0.35	0.91	0.08	0.92	5
15	gCar M101	9	9	2	5.78	0.12	0.83	0.16	0.84	5
Total		160	160							86
% polymorphism		100 %								
Average		10.7	10.7		11.09	0.21	0.84	0.19	0.81	5.7

[a] number of effective alleles (NE) for each marker was calculated according to (Hartl and Clark 1997).

[b] Observed heterozygosity and is calculated as direct count. 1.00 indicates that the locus is heterozygous over all the studied genotypes.

[c] refers to the polymorphism information content or gene diversity and it is calculated according to Nei's (1973) and reflect the ability of **a marker for detecting polymorphism between the genotypes, depending on the numbers of detectable alleles and their frequency.**

[d] refers to, the probability of two genotypes or more having similar SSR profiles.

[e] refers to, the probability of discriminating between two genotypes or more with every locus. It is calculated as 1-P (P = probability of matching fingerprints).

This could be explained by the truth that Arabica coffee is an allotetraploid species, and could be detected alleles from both parents. The heterozygosity (Ho) within the loci was calculated from direct count and ranged from 0.00 to 0.94 with a mean value of 0.21 (Table 3). In the present study, heterozygosity values were low in the most of SSR loci and this is consistent to the results obtained in other studies on genetic diversity of Coffee by (Anthony *et al*. 2001a). It could be due to low mutational rate at the SSR loci, whereas the loci with few numbers of repeated units tend to show weakly mutational rate. Polymorphism Information Content (PIC) or gene diversity was calculated by using marker allelic frequencies.

The results showed that the PIC values (Table 3) were quite high and varied from 0.57 to 0.98, with average value of 0.84. All loci in this study have showed PIC values above 0.50, hence, it could be described a highly informative to study the genetic diversity between Yemeni coffee genotypes. This wide range of PIC values was indicative of the presence of unique alleles in some cultivars which simplify their differentiation from another. The mean PIC value of 0.84 reflected the high level of polymorphisms of the used set of microsatellites and heterogeneity in 17 coffee genotypes. These PIC values are higher compared to other studies which showed a mean of PIC value 0.319 as reported by (Al-Murish *et al* 2013) and 0.043 according to (Missio *et al* 2010) .This difference might be linked with selection of different markers and more diverse set of genotypes in present study. In a related context, based upon discriminating power of the used 15 SSR loci in this study, the most informative primer pairs was gSSRCa 021 with DP value of 0.94 (Table 3). The power of discrimination was varied from 0.60 to 0.95 with an average value of 0.81. The results was showed that the DP values were consistent with the PIC values for the 15 SSR loci, due to computation of the DP depends on the number of alleles at a locus and the relative frequencies of genotypes instead of frequencies of alleles in the PIC values.

Heterozygosity level is an appropriate measure of genetic variability within cultivars because genetic diversity can be measured as the amount of actual or potential heterozygosity. Level of heterozygosity within the genotypes in this study was ranged from 0.067 to 0.333 with a mean of 0.212 (Table 4). The average of heterozygosity degree in the amplified SSR loci is very low compared to similar analysis (Lee 1995, Sabry *et al* 2015 and Hussein 2009) but in other species. This result seems to indicate that these genotypes maybe have not received enough external genes flow, which is likely the reason of the excess heterozygosity (Bodian *et al* 2012). Also, *C. arabica* is an autogamous species, which give contribute to low heterozygosity levels. On the other hand, our results were relatively consistently with (Combes *et al* 2000) which in general, reported low

genetic diversity of coffee. This is probably, as a result of the narrow genetic basis of the cultivated Coffee.

Genetic fingerprint for discrimination and creating of molecular ID for Yemeni coffee germplasm

The profiles of all the fifteen used loci in present study were combined and compared to elucidate whether any genotypes were genetically identical (Figure 2). The results showed that none of the genotypes pairs are identical at all loci, whereas all the fifteen SSR loci were polymorphic (100%). Two genotypes only (0.12) Shubriqi and Ismaili cultivars had 5 identical loci. However, from two to three genotypes were genetically identical at 2 or 3 loci. Regarding this, the three samples of the "Odayni" cultivar, which were collected from three different Governorates (Figure 1). These genotypes (Odayni _ IbbG, Odayni _ AmlG and Odayni _ SadG.) were identical in two loci only despite being the same cultivar. Where these two loci (g E8_3CTG and g CarM048) are showed highest value of matching fingerprints, 0.30 and 0.40 respectively. Similar result with "Odayni - Bayad" cultivar which its two samples were collected from two different places in Al-Mahaweet Governorate and were identical in one locus (gSSRCa 006) only. In this regard, the probability of matching fingerprints is indicating to, two or more of the genotypes having similar SSR profiles (Figure 2).

Table 4. Heterozygosity levels within the studied Yemeni coffee genotypes and assayed with 15 SSR loci.

No.	Genotypes	Heterozygosity	No.	genotypes	Heterozygosity
1	Odayni _ IbbG.	0.200	10	Shubriqi	0.133
2	Odayni _ AlmG.	0.067	11	Ismaili	0.200
3	Odayni _ SadG.	0.200	12	Dawairy	0.200
4	Burai	0.067	13	Tofahy	0.267
5	Benan	0.267	14	Gadi	0.200
6	YEMAM1	0.333	15	Oubale	0.267
7	Odayni- Bayad L1	0.267	16	Abu Sura	0.200
8	Odayni- Bayad L2	0.267	17	Tasawai	0.333
9	Benan- Bayad	0.133			
Mean	0.212				

Value of matching fingerprints was ranged from 0.05 with gSSRCa 021 locus to 0.40 with gM24 and g CarM048 loci with an average of 0.19 (Table 3). The low average of matching fingerprints is confirming to the low degree of similarity between the Yemeni coffee germplasm in this study. It is possible due to the nature of the used SSR loci and the selected group of genotypes in the present study is more diverse. In addition, this is

consistent with registration a high number of specific/ exclusive alleles in these genotypes (Table 3).

The difference between the number of alleles in each locus and number of effective alleles showed the existences of exclusive/specific alleles in the genotypes. However, private alleles occurred in one or more genotypes, as reported (Kohpayegani and Behbahani 2008).

The genetic analysis of the 17 Yemeni Coffee genotypes based on 15 polymorphic SSR markers is detected 86 clear specific alleles across all loci. The number of specific alleles was varied from one with gE8-3CTG locus to seventeen with gSSRCa 021 locus with an average value of 5.7 (Table 3). All cultivars and accessions in this study had alleles were unique to that genotype alone (Figure 2). Tasawai cultivar had the most unique alleles (10 alleles), while the lowest number of genotypes-specific alleles (2 alleles) was scored with "Benan-Bayad", "Dawairy" and "Odayni- Bayad L2" cultivars. Also to increase the long-term usefulness of the DNA fingerprinting analysis, all the fifteen SSR loci data in this study were used to create so-called 'reference DNA polymorphism panel' (Figure 2) as reported (Aggarwal *et al.*, 2004). This panel could be used to generate of molecular ID for genotype identification and for other advantages in breeding programs of genetic improvement, for example, selection of most informative SSR marker (in our study was gSSRCa 21) and also selection of appropriate or diverse parental materials for breeding purpose and linkage analysis.

The exclusive/specific alleles may only be inferred for the sampled gene pool. Also the specific alleles could use to identify the genotypes by means of combination of some genetic loci. For example, in (Table 3), gSSRCa 021 locus can be used as primer to identify genotypes which the gene alleles were varied from each other. In addition, "Benan-bayad", "Odayni-bayad L2" and "Odayni-bayad L1" cultivars, although they are similar morphologically in the fruit color but it could be discriminated from each other using a set of (2, 2 and 7) primers which have produced (2, 2 and 7) unique alleles respectively. Thus each sequence would work as a molecular signature for each cultivar. In a related context, the rest of genotypes under study are characterized by the same color of red fruits. These genotypes could be distinguished genetically from each other, through recording a number of unique alleles that ranged from 2 to 10 specific alleles. While, YEMAM1 accession is produced high number of unique alleles (7 specific alleles) from five different loci (Figure 2), where it is differs from the rest of the cultivars under study in the nature of growth and the parpule color of leaves.

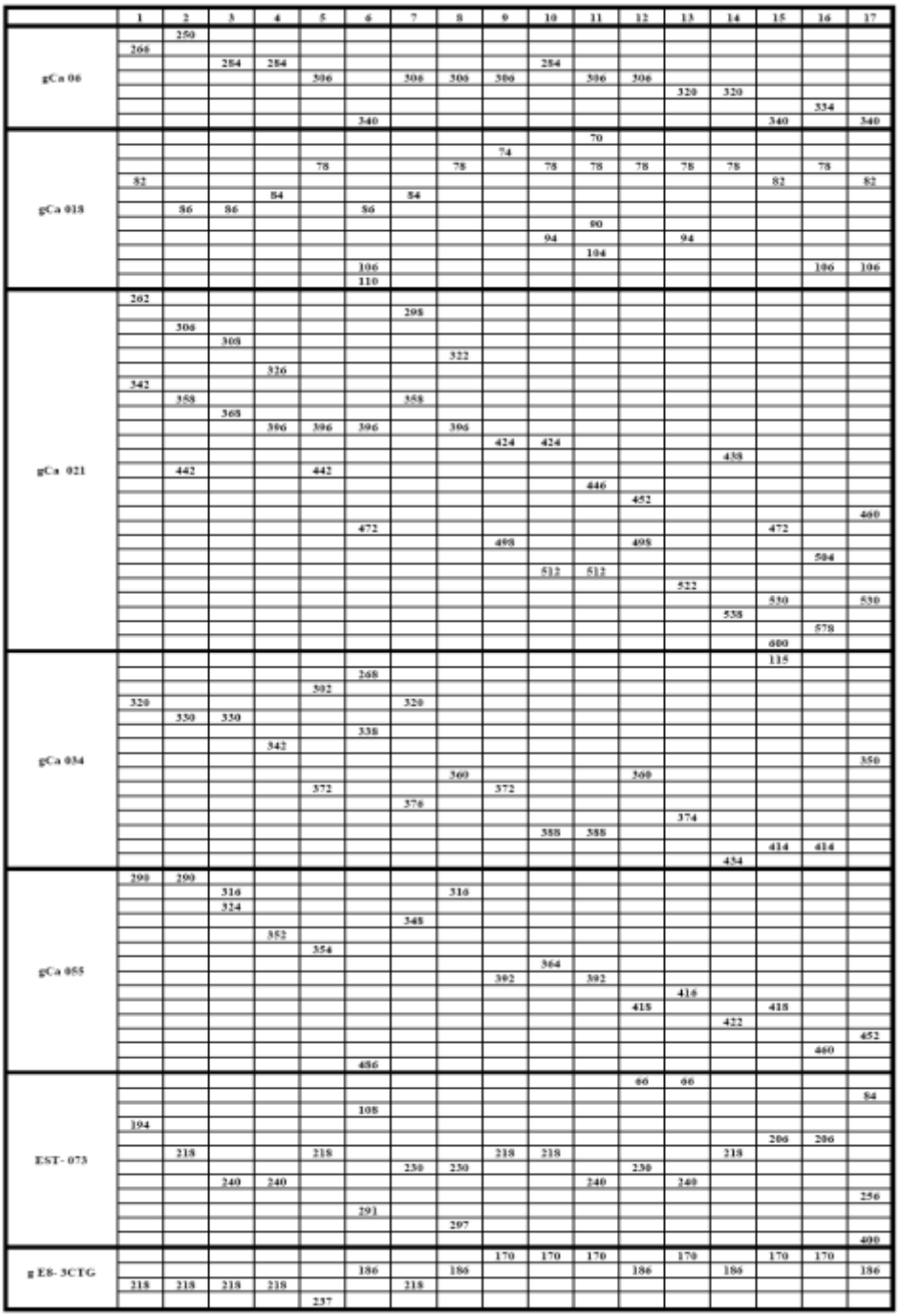

Figure 2. Part of reference DNA polymorphism panel, it is showing allelic distribution across seven out of the fifteen amplified SSR loci with 17 Yemeni coffee genotypes referred to in Table1. **It is noted that each genotype distinguished from the other in this panel.**

Unexpectedly, it was possible to record genetic differences between samples of "Odayni" cultivar, which collected from three different governorates. In this regard, Odayni _ IbbG, Odayni _ AlmG and Odayni _ SadG, registered (7, 5 and 3) specific alleles respectively, despite being the same cultivar. Similar result with "Odayni-Bayad" cultivar which its samples (Odayni-Bayat_L1 and Odayni-Bayat_L2) were collected from two different places in Al-Mahaweet Governorate and had registered (7 and 2) specific alleles respectively. It is possibly result from mutations that occurred in the genome region for which the primers are designed, or genomic rearrangement, including insertion – deletion events within the region to be amplified, as reported by (Leigh *et al* 2003). This difference also could be arise, from the fact that most of the used SSR markers in the present study are genomic SSR and these loci are high evolutionary regions and located in non-coding regions of the genome. Therefore it be limited use in predict the phenotypic diversity of individuals, especially in complex traits such as yield and this agree with (Collard *et al* 2005). The results presented here very useful for identification the genotypes and fingerprinting of Yemeni coffee cultivars for breeding programs purposes

The results are confirmed that SSR markers could be used reliably for identification and fingerprinting of the Yemeni coffee germplasm for their conservation and that the SSR markers are the most efficient and desirable markers for these purposes.

Genetic structure of Yemeni coffee Genotypes

The genetic structure reflects the interactions among cultivars with regard to their long-term evolutionary history, mutations, recombination, genetic drift, reproductive system, genes flow, and natural selection (Schaal *et al* 1998). Thus, an understanding of the level and structure of the genetic diversity of Yemeni coffee is a necessary for the conservation and efficient use of the germplasm available for breeding purposes (Laidò *et al* 2013).

In addition to their usefulness for genetic fingerprinting of coffee cultivars, the co dominant nature of SSRs allows a better understanding of the pedigree relationships and level of diversity among the studied cultivars. The data were analyzed to determine the genetic relationships among the 17 coffee cultivars and accessions and to compute the similarity matrices according to Dice coefficient (data not shown).

The genetic similarity degree ranged from 0.00 to 0.486 between all possible pairs of genotypes with an average of 0.243. The highest similarity was observed between “Dawairy” and “Odayni-Bayad-L2” (0.486), means that the two genotypes are genetically the closest despite the different color of the fruits in each of them. This is probably the results of a narrow genetic basis of the cultivated Arabic coffee, as reported by Maluf *et al* (2005). It could be also explained on the basis that the two cultivars maybe have been originated in the same geographical region, since the onset of the coffee-

growing in Yemen. Then it is possible over the years many have been transplanted to areas other (Sana'a and Al-Mahaweet Governorates) than the area of their origin, and there may have been adapted. Which was led to form of some translated genetic differences, and then it has new names. While “Tasawai” cultivar is showed the smallest similarity value (Zero) or slightly higher with a large number of genotypes in the present study, means that they are the most genetically distance. Whereas, “Tasawai” cultivar is the only in this study which is cultivated in southern Yemen (Lahj Governorate) as shown in (Figure 1). Therefore, it is far from the geographic regions of the rest genotypes. So maybe it is retains the unique genetic characteristics since the Arabica coffee had introduced from Ethiopia to Yemen. This is consistent with the recording of the highest number of exclusive alleles of this cultivar.

Despite contrasting with some previous studies concerning higher of the genetic similarity in Arabica coffee, the genetic similarity between the coffee cultivars in the present study was not comparatively high (0.243). This due to the nature of molecular characterization of the used SSR loci and the selected group of genotypes are more diverse. These results are in agreement with (Lashermes *et al* 1996 and Al- Murish *et al* 2013). Whereas. polymorphism between genotypes can bring about through nucleotide changes that prevent amplification by introducing a mismatch at one primer binding site, deletion of a primer binding site, insertions that render primer binding site too distant to support amplification and insertions or deletions that change the size of the amplified product as reported (Williams *et al* 1990) Thus, SSR markers provide enough power of resolution to discriminate between coffee cultivars and it could serve as a potential tool in the identification and characterization of genetically distant cultivars from various sources.

Using Average Linkage (Between Groups), the similarity degree based on Dice coefficient and the dendrogram was obtained. Three main groups can be observed in the dendrogram (Figure 3), based on the genetic similarity degree and the geographical distribution of Yemeni coffee. The first group has included five genotypes, two cultivars from Al- mahaweet governorate (Burai and Odayni-bayad L1); in addition to three samples of "Odayni" cultivar (Odayni _ IbbG, Odayni _ AlmG and Odayni _ SadG) and the three from different governorates. This cluster could be divided into two sub clusters, the first contained Odayni _ IbbG, Odayni _ AlmG and Odayni _ SadG, which were obtained from three different regions (Ibb, Al-Mahaweet and Sadah Governorates). It is accepted that the three samples have the same color of red fruits. Nevertheless, it could be distinguishing them from each other, through recording (7, 5 and 3) specific alleles respectively, despite being the same cultivar.

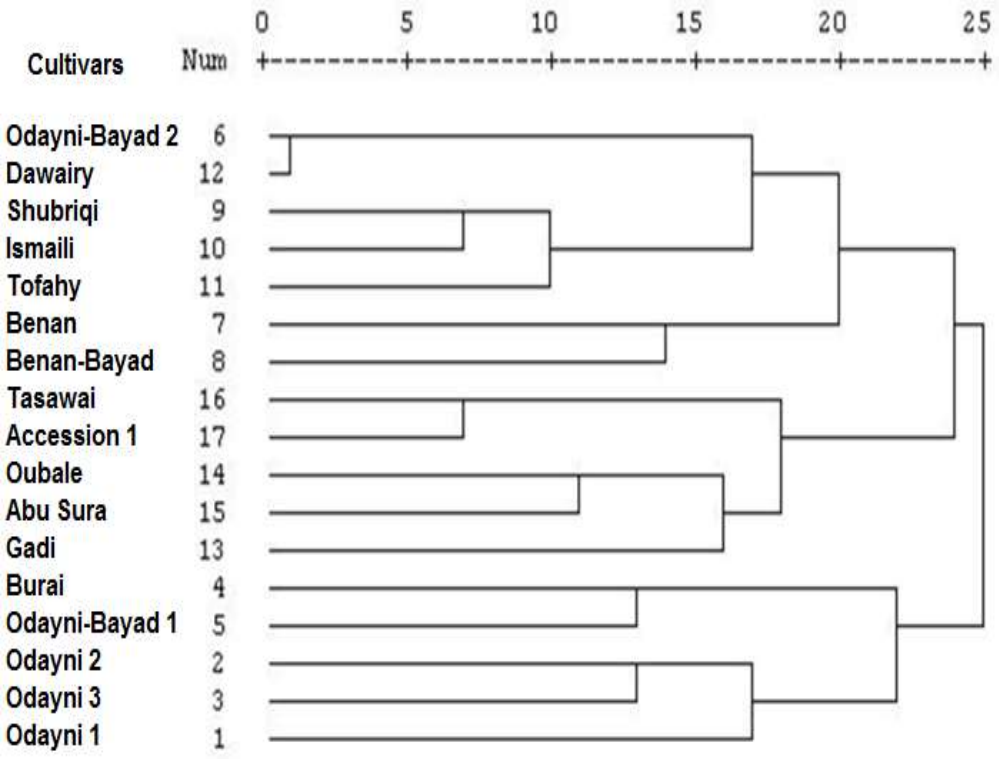

Figure 3. Dendrogram of the seventeen Yemeni coffee genotypes constructed from SSRs data based on Average Linkage (Between Groups), using Similarity computed according to Dice coefficient, with 160 fragments.

The second cluster is included also five genotypes (Gadi, Abu Sura, Oubale, YEMAM1 and Tasawai) and divided to two sub cluster. The first sub cluster is included (Gadi, Abu Sura and Oubale) and the three cultivars have the same color of red fruits. Abu Sura and Oubale cultivars are clustered together with a high similarity degree (0.316) compared to the rest cultivars and the both are cultivated at (Taiz Governorate). This result consistent with what is declared in the Coffee Research Unit of the Ministry of Agriculture in Taiz, that Abu Sura variety is one of the oldest cultivars in Yemen. It is possibly that this variety is the origin of the rest cultivars. The second sub cluster is included "Tasawai" cultivar and "YEMAM1", whereas "Tasawai" cultivar is cultivated at (Lahj Governorate), which is far from" YEMAM1" at (Al-Mahaweet Governorate). Nevertheless, the similarity degree between them was high (0.381) relatively. Where, it is observed that

both them have the same color of red fruits. May this due to that YEMAM1 was elected from "Tasawai" cultivar.

The third group could be divided into two sub clusters and includes the rest of the cultivars under study. First sub cluster included Benan- Bayad and Benan cultivars with similarity degree (0.278).

On the other hand, the second sub cluster is included (Tofahy, Ismaili, Shubriqi, Dawairy and Odayni- Bayad L2) cultivars, first four cultivars are cultivated at the same Governorate (Sana'a) and have red fruits. Since, the Tofahy and Dawairy cultivars are the oldest cultivars of coffee in Yemen as reported (Qaid 1993). It is clear that they clustered together. Where, the fifth cultivar "Odayni- Bayad L2" has yellow fruit and it is showed highest degree of similarity (0.486) with "Dawairy" cultivar in the present study. These results are in agreement with the narrow genetic basis of cultivated plants of Arabica coffee and historical data. Which is supported the hypothesis that the Yemeni coffee has been originated from a few plants transferred to Yemen as early as 575 AD for cultivation by the Arabs (Wellman 1961).

Results of the present study showed the presence of genetic difference among Yemeni coffee genotypes within and between Governorates. And despite the narrow genetic basis of the Arabica coffee and being self-pollinated. The genetic diversity was ranged between a moderate to high among the seventeen genotypes in this study. It is possible due to the nature of the used SSR loci and the selected group of genotypes is more diverse. In addition, the genotypes were classified according to similarity degree, fruit color and geographical regions. SSR markers confirmed previously studies for genetic relationships among *Coffea* species and they verified to be efficient and desirable markers to analysis the genetic diversity and structure within *C. Arabica* species. These results indicated the importance of the genetic fingerprint for conservation of Yemeni coffee germplasm. Which are characterized that it is more diverse, higher yield and better quality indicating their importance as a source of genetic variability for the purposes of coffee improvement.

REFERENCES

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

Aggarwal, R., R. Rajkumar, P. Rajendrakumar, P. Hendre, A. Baruah, R. Phanindranath, V. Annapurna, N. Prakash, A. Santaram and L. Singh (2004). Fingerprinting of Indian coffee selections and development of reference DNA polymorphism panels for creating molecular IDs for variety identification. In Proceedings of 20th International Conference on Coffee Science (ASIC) Bangalore, India, PP 751-755.

Al-Murish, T. M., A. A. Elshafei, A.A. Al-Doss and M. N. Barakat (2013). Genetic diversity of coffee (*Coffea arabica* L.) in Yemen via SRAP, TRAP and SSR markers. Journal of Food, Agriculture & Environment, 11(2): 411-416.

Anthony, F., J. Berthaud, J.L. Guillaument and M. Lourd (1987). Collecting wild coffea species in Kenya and Tanzania. Plant Genet Resour News 1, 69: 23-29.

Anthony, F., B. Bertrand, O. Quiros, A. Wilches, P. Lashermes, J. Berthaud and A.Charrier (2001). Genetic diversity of wild coffee (*Coffea arabic*a L.) using molecular markers. Euphytica, 118:53-65.

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

Berthaud, J. and A. Charrier (1998). Genetic resources of coffea pp. 1-42 In Clarke RJ and Marcrae R (eds). Coffee Vol. 4 Agronomy, Elsevier applied Science, London and New York.

Bodian A., A. Mohamed and N. Khadidiatou (2012). Genetic diversity analysis of date palm (*Phoenix datylifera* L.) cultivars from Figuig oasis (Morocco) using SSR markers. International Journal of Science and Adavnced Technology, 2 (3): 96-104.

Baruah, A., V. P. S. NaikHendre, R. Rajkumar, P. Rajendrakumar and K. Aggarwal (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications, Molecular Ecology Notes, 3: 647–650

Carvalho, A. (1946). Distribuição geografica e classificação botânica do gênero Coffea com referência especial à espécie Arabica. V. Origem e classifica çãbotãnica do *C. arabica* L. Bol Super Serv Caf, 21:174–180

Carvalho, A., F. P. Ferwerda, J. A. Frahm-Lelield, P. M. Medina, A. J. Mendres and L.C. Monaco (1969). Coffee. In: Ferwerda, F. P., and Wit, F. (eds) Outlines of Perennial Crop Breeding in the Tropics. Veenman & Zonen, Wageningen.

Carvalho, V.P. (1988). Principles and Practices of coffee plant breeding for productivity and quality factors: *Caffea arabica* PP. 1 - 42. In Clarke, R. J. and Marcrae, R. (eds). Coffee Vol. 4 Agronomy, Elsevier applied Science, London and New York.

Chevalier, A. and M. Dagron (1928). Recherches historiques sur les débuts de la culture du caféier en Amérique. Commun Actes Acad Sci Colonial Paris.

Collard, B., M. Jahufer, J. Brouwer and E. Pang (2005). An introduction to markers quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica, 142: 169-196.

Combes, M. C., S. Andrzejewski, F. Anthony, B. Bertrand, P. rovelli, G. Graziosi and P Lashermes (2000). Characterization of microsatellites loci in *Coffea arabica* and related coffee species. Molecular Ecology, 9: 1171-1193.

Denoeud, F., L. Carretero-Paulet, A. Dereeper and G. Droc (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science, 345: 1181-1184.

Eskes, A. B. (1989). Identification, description and collection of coffee types in P.D.R. Yemen. IPGRI, Rome

FAO (1968). FAO Coffee mission to Ethiopia 1964 – 1965. FAO, Rome.

Gadaleta, A., G. Mangini, G. Mulè and A. Blanco (2007). Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome. Euphytica, 153: 73-85.

Gur-Arie, R., C. J. Cohen, Y. Eitan and L. Shelef (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res., 10: 62-71.

Haarer, A.E. (1956). Modern Coffee Production. Leonard Hill, London.

Hart, L. and A.G. Clark (1997). Principles of population Genetics. 3rd edition. Sinauer Associates, Inc, Sunderland, U. S. A.

Hussein, M. A. (2009). Identification and estimation of genetic relationships among some mango cultivars grown in Egypt using DNA techniques. Ph.D. Thesis, Suez Canal University, Egypt.

International Coffee Organization (ICO), ICO Annual Review 2011/12, ICO, London, England (2013).

Jones, D.A. (1972). Blood Samples: probability of discrimination. Journal of the Forensic science Society, 12: 355-359.

Kloosterman, A.D., B. Budowlw and P. Daselaar (1993). PCR-amplification and detection of human DIS80 VNTR locus-Amplification condition, population genetics and application in forensic analysis. International Journal of legal Medicine, 105: 257-264.

Kochko, A.D., S. S. Akaffou, A. C. Andrade, C. Campa, D. Crouzillat and R. Guyot (2010). Advances in *Coffea* genomics. Adv Botanical Res., 53: 23-63.

Kohpayegani, J. and M. Behbahani (2008). Genetic diversity of some populations of Iranian melon using SSR markers. Biotechnology, 7 (1): 19-26.

Laidò, G., G. Mangini, F. Taranto, A. Gadaleta, A. Blanco, L. Cattivelli, D. Marone, A. Mastrangelo, R. Papa and P. de-Vita (2013). Genetic diversity and population structure of tetraploid Wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. PLOS, 8 (6): 1-14.

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

Lashermes, P, M. C. Combes, J. Robert, P. Trouslot, D. A. Hont, F. Anthony (1999). Molecular characterization and origin of the *coffea arabica* L. genome. Molecular Genomics and Genetics , 261: 259-266.

Lee, M. (1995). DNA markers and plant breeding programs. Advances in agronomy, 55: 265-344.

Leigh, F., V. Lea, J. Law, P. Wolters, W. Powell and P. Donini (2003). Assessment of EST-and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica, 133(3): 359-366.

Maluf, M. P., M. Silvestrini, L. Ruggiero, O. Guerreiro-Filho and C. Colombo (2005). Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Science Agriculture (Braz.), 62(4), 366-373.

Metais, I., B. Hamon, R. Jalouzot and D. Peltier (2002). Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor. Appl. Genet., 104: 1346-1352.

Missio, R.F., E. T. Caixeta, E. M. Zambolim and L. Zambolim (2009a). Development and validation of SSR markers for *Coffea arabica* L. Crop Breed. Appl. Biotechnol. 9: 361-371.

Missio, R. F., E. T. Caixeta, E. M. Zambolim and G. F. Pena (2009b). Assessment of EST-SSR markers for genetic analysis on coffee. Bragantia, 68: 573-581.

Missio, R. F., E. T. Caixera, E. M. Zambolim, L. Zambolim, C. D. Cruz and N. S. Sakiyama (2010). Polymorphic information content of SSR markers for *Coffea* spp.Crop Breed. Appl. Biotechnology, 10: 89-94.

Missio, R. F., E.T. Caixeta, E. M. Zambolim, G. F. Pena, L. Zambolim, L. A. Dias and N. S. Sakiyama (2011). Genetic characterization of elite coffee germplasm assessed by g SSR and EST-SSR markers. Genetic and Molecular Research, 10 (4): 2366-2381.

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

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

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA, 70: 3321-3323.

Plechakova, O., C. Tranchant-Dubreuil, F. Benedet; M. Couderc, A. Tinaut; V. Viader, P. Block, P. Hamon; C. Campa A. Kochko and S. Hamon (2009). MoccaDB-an integrative database for functional, comparative and diversity studies in the Rubiaceae family. BMC Plant Biology, 9(1): 123.

Porebski S., L. Grant Bailey and B. R. Baum (1997). Modification of a CTAP DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter, 15 (1): 8-15.

Qaid, A. (1993). Surveying and Classifying Coffee in Yemen. Available through Ministry of agriculture and irrigation, Sana'a, Yemen.

Raymond, M. and F. Rouset (1995). GENEPOP (version 1.31): population genetics software for exact test and ecumenicism. J. Hered 86:248-249. http:// whbiomed.curtin.edu.au/gene-pop/.

Rogers, S.O. and A. J. Bendich (1985). Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues, **Plant** Molecular Biology, 5: 69-76.

Rovelli, P., R. Mettulio, F. Anthony, F. Anzueto, P. Lashermes and G. Graziosi (2000). Microsatellites in *Coffea arabica* L. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) Coffee Biotechnology and Quality. Proceedings of the 3rd International Seminar on Biotechnology in the Coffee Agroindustry, Londrina, Brazil. Kluwer. Academic Publishers, Dordrecht, The Netherlands.

Sabry, S., M. Eid; M. A. Hussein and S. Greish. (2015). Genetic diversity analysis in date palm (*Phoenix dactylifera* L) germplasm using microsatellite markers . Journal of Plant Production Sciences; Suez Canal University, 3: 19- 29.

Schaal, B., D. Hayworth, K. Olsen, J. Rauscher and W. Smith (1998). Phylogeographic studies in plants: problems and prospects. Molecular Ecology, 7 (4): 465-474.

Silvarolla, M.B., P. Mazzafera and M.M. Alves de lima (2000). Caffeine content of Ethiopian *Coffea arabica* beans. Genetics and Molecular Biology, 23 (1): 213-215.

Silvestrini, M., M. G. Junqueira, A. C. Favarin, O. Guerreiro-Filho, M. P. Maluf and M. B. Sivarolla (2007). Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. Genet Res Crop Evol., 54: 1367-1379.

Sneath, P.H. and R. R. Sokal (1973). Numerical taxonomy. Freeman. San Francisco.

Steiger, D.L., C. Nagai, P. H. Moore, C.W. Morden, R. V. Osgoog and R. Ming (2002). AFLP analysis of genetic diversity cultivars. Theor. Appl. Genet., 185: 189-215.

Sylvain, P.G. (1955). Some observations on *Coffea arabica* L. in Ethiopia. Turrialba, 5: 37-53.

Thomas, A.S. (1942). The wild Arabica coffee on the Boma Plateau, Anglo-Egyptian Sudan. Emp. Exp. Agric., 10:207-212.

Tran, H. T., L. S. Sade, A. furtado, H. Smyth and R. J. henry (2016). Advances in genomics for the improvement of quality in coffee. J. Sci. Food Agric., 96: 3300-3312.

Varshney, R.K., A. Graner and M. E. Sorrells (2005). Genic microsatellite markers in plants: features and applications. Trends Biotechnol., 23: 48-55.

Vieira E.S., E. V. Von Pinho, M. G. Carvalho, D.G. Esselink and B. Vosman (2010). Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. Genetics and Molecular Biology, 33 (3): 507-514.

Wellman, F.L. (1961). Coffee, Botany, Cultivation and Utilization. Leonard Hill Books, London.

Williams, J., A. Kubelik, K. Livak, J. Rafalski and S. Tingey (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research, 18 (22): 6531-6535.

الاختلافات والعلاقات الوراثية وتحديد البصمة الوراثية لإنشاء تعريف جزيئي للأصول الوراثية للبن اليمني العربي بواسطة معلمات التتابعات المكررة البسيطة (الميكروستاللیت)

محمد عبد الجواد عبد الكريم حسين[١] وأمين على أحمد العزب[٢،٤] و سندالسيد حبيب[٢]،

فاديه محمد عسكر الشريف[٢] و هدى الجارحي[٣]

١. قسم النبات الزراعي– كلية الزراعة – جامعة قناة السويس

٢. قسم البساتين – كلية الزراعة – جامعة قناة السويس

٣. قسم الوراثة و الهندسة الوراثية – كلية الزراعة – جامعة بنها

٤. وزارة التربية والتعليم – الجمهورية اليمنية

أجريت هذه الدراسة بهدف تقدير مستوى الاختلافات و العلاقات الوراثية بين سبعة عشر تركيب وراثي تشمل ١٦ صنف تجاري معروف بالإضافة الى تركيب وراثي واحد غير مسمى من الأصول الوراثية للبن اليمني العربي، والتى تم تجميعها من محافظات مختلفة في اليمن. وهدفت الدراسة ايضا لتحليل بيانات البصمة الوراثية بهدف إنشاء هوية جزيئية بغرض حماية وصيانة هذه الأصول الوراثية عالية الجودة. وقد تم تحقيق اهداف الدراسة بمساعدة واستخدام ١٥ زوج من بوادئ معلمات الميكروستاللیت والتي سبق تصميمها وتوصيفها لتقييم كفائتها وأدائها في تحقيق أهداف الدراسة. اظهرت النتائج وجود تباين عالي لهذه المواقع الوراثية بنسبة ١٠٠% حيث تراوح عدد الاليلات المضخمة من ٤ إلى ٢٥ وبمتوسط ١٠.٧ لكل موقع وراثي، وقد أظهرت نسبة الخلط سواء بداخل كل موقع او بكل تركيب وراثي على مستوى كل المواقع الوراثية انخفاضا ملحوظا بنسبة ٠,٢١ في حين لوحظ أرتفاع مقدار ما يتيحة كل موقع وراثي من معلومات عن الاختلافات متراوحا ما بين ٠,٥٧ الى ٠,٩٨ وقد ارتفعت أيضا الطاقة التمييزية لكل المواقع بمتوسط ٠,٨١ حيث أظهر موقع gSSR Ca 021 أكبر طاقة تمييزية وكفاءة في تحقيق أهدف الدراسة بقيمة ٠,٩٤ . وقد أظهر احتمال تطابق البصمة الوراثية لزوج أو أكثر من التراكيب والوراثية انخفاضا ملحوظا بمتوسط ٠,١٩ حيث لم يظهر اي تطابق في البصمة الوراثية لأي زوج من الأصناف على مستوى كل المواقع الوراثية المضخمة حتى مابين التراكيب الوراثية التي تحمل نفس الأسم ولكنها منزرعة في مناطق جغرافية مختلفة. على الجانب الأخر، أمكن التفريق والتمييز بين جميع التراكيب الوراثية بتسجيل عدد أجمالي ٨٦ حزمة منفردة أو حصرية واضحة حيث تم تمييز كل تركيب عن الأخر بوجود حزمة منفردة او أكثر بمتوسط ٥,٠٦ حزمة حصرية/ تركيب وراثي. وفيما يتعلق بتحليل العلاقات الوراثية ودرجة التشابه، قسمت النتائج التراكيب الوراثية الى ثلاثة مجاميع رئيسية وفقا لدرجة التشابه والتي ظهرت منخفضة و تراوحت من صفر الى ٠,٤٨ بمتوسط ٠,٢٤ . ويشكل عام أثبتت النتائج وجود مستوى عالى من الاختلافات الوراثية بين التراكيب الوراثية في هذه الدراسة تراوحت ما بين المتوسطة والعالية، مما يشير لأهمية هذه المجموعة من التراكيب الوراثية كمصدر جيد للاختلافات الوراثية لأغراض تحسين وتربية البن العربي والذي يعاني من قاعدة وراثية ضيقة .

المجلة المصرية لتربية النبات ٢١ (٤) : ٧١٣ - ٧٣٦ (٢٠١٧)