



Molecular characterization of Arabica Coffee (*Coffea arabica* L.) germplasms and their contribution to biodiversity in Ethiopia

Gudeta Dida¹ · Kassahun Bantte² · Tesfaye Disasa¹

Received: 26 July 2021 / Revised: 26 October 2021 / Accepted: 28 October 2021 / Published online: 1 November 2021
© Korean Society for Plant Biotechnology 2021

Abstract

Ethiopia is the center of origin and diversity for Arabica coffee with high morphological diversity between accessions as compared to commercial cultivars. Coffee germplasm collection and molecular characterization are crucial steps towards its conservation, breeding, and development of superior genotypes for various end uses. Hence, this study was initiated with the objective of studying the genetic diversity of *Coffea arabica* accessions collected from different regions of Ethiopia, using SSR markers. A total of 20 SSR markers were used to genotype 86 accessions and produced a total of 112 alleles, with an average of 5.6 alleles per locus. All the loci across the entire populations were found to be highly polymorphic with a mean of 0.6 PIC value. Average observed heterozygosity and allelic richness across all populations ranged from 0.22 to 0.27 and 3.52–4.26, with a mean of 2.43 and 3.97, respectively. AMOVA showed high variation within population based on geographical origin. The smaller F_{st} (0.037) observed indicates the presence of lower population genetic differentiation as a result of higher gene flow ($N_m = 2.45$) across populations and the lowest mean genetic distance (0.21) observed between populations. The UPGMA, PCoA, and structure analysis poorly grouped the individuals into distinct clusters indicating the presence of population admixture. The observed higher genetic variability in all populations indicates that the country has huge coffee genetic diversity which can be used for future coffee improvement. Our results revealed an unexploited highly diverse genetic resource particularly from Omo, Ilubabor, and Benchi Maji that should be considered in future coffee breeding program and germplasm conservation.

Keywords *Coffea arabica* · Genetic diversity · Gene flow · Heterozygosity · SSR marker

Introduction

Coffee is one of the most valuable commodity crops in the world trade. More than 90% of its production comes from developing countries, providing an income for millions of smallholder farmers around the world that are dependent on coffee for their subsistence (Tran et al. 2016). It contributes largely to the economy of more than 50 countries in Asia, Latin America, and Africa (Davis et al. 2011). The share of coffee in total export earnings has a positive and significant

reflection on economic growth for developing countries (Al-Abdulkader et al. 2018).

Since Ethiopia is the origin and the primary center of diversity for *Coffea arabica*, its production in Ethiopia has a long tradition which dates back to dozens of centuries. The country possesses a diverse genetic base for Arabica coffee and its wild relatives with considerable heterogeneity (Gebreselassie et al. 2018). The coffee grown in the country showed wide genetic variations within and between populations of different regions for yield, quality disease resistance and other traits. The availability of such genetic variations provides immense possibilities for improvement of the crop for any desirable traits of interest (Benti 2017; da Silva et al. 2019). Coffee is by far Ethiopia's most important export crop which accounts the lion's share of the foreign exchange earnings of the country. Furthermore, the livelihood of some 15 million people directly or indirectly depends on coffee. Ethiopia accounts for around 3% of the global coffee market

✉ Gudeta Dida
gudetadida2014@gmail.com

¹ Ethiopian Institute of Agricultural Research, National Agricultural Biotechnology Research Center, P.O. Box 249, Holeta, Ethiopia

² College of Agriculture and Veterinary Medicine, Jimma University, P.O. Box 307, Jimma, Ethiopia

and around 35% of foreign income comes from coffee (Benti 2017).

The quality standards of Ethiopian coffee are classified according to their origin of production. Among the best-known coffee varieties in Ethiopia are Yirgacheffe, Harar, Wolegga, Limu, Jimma, Sidama, and others take the priority. Within Ethiopia, there are some distinctive varieties that are highly sought after. The majority of cultivated varieties in Ethiopia comes from Harar, where the Long berry variety is the most popular, having a wine-like flavor and tasting slightly acidic. Another Ethiopian coffee from Yirgacheffe is the highest premium coffee in Ethiopia as well in the world (Gebreselassie et al. 2018). However, presently Ethiopian coffee genetic resource is under greatest threat, mostly due to deforestation, replacement of farmer's variety by a few high yielding and disease-resistant varieties, establishment and expansion of modern plantation and illegal and legal settlements, drought, the spread, and escalating severity of devastating fungal pathogens (da Silva et al. 2019).

Coffee breeding programs invested intense efforts to release cultivars with high productivity, with climate change and simultaneous change in biotic and abiotic pressures stress tolerance and high biochemical quality of the beans (Tran et al. 2016). However, several factors are limiting the genetic gains in breeding programs (Vieira et al. 2010). Commercial coffee varieties are originated from a limited number of cultivars; consequently, only a narrow genetic base is available to support breeding programs. Susceptibility of commercial coffee cultivars to pests and diseases, climate change, and emerging pest and outbreak of diseases are important challenges to global coffee production (Bunn et al. 2015).

Many studies reported wide phenotypic diversity of Arabica coffee accessions collected from Ethiopia regarding leaf size, height, biotic and abiotic stress tolerance, and yield (Tran et al. 2016) and the presence of higher genetic variability compared with cultivars, demonstrating the potential of these accessions for breeding purposes (Aerts et al. 2013; Sant'Ana et al. 2018). The accessions also showed a great variability for chlorogenic acids, lipids, sucrose, and diterpenes contents of coffee beans and the discovery of decaffeinated coffee varieties (Silvarola et al. 2004) which underlines the great potential existing in Ethiopian coffee accessions for Arabica coffee future breeding programs.

A number of DNA-based techniques were used in different coffee genetic studies. These include RAPD (Diniz et al. 2005), AFLP (Steiger et al. 2002), ISSR (Tesfaye et al. 2014), and microsatellite (SSR) markers (Lashermes et al. 1995; Teressa et al. 2010; Geleta et al. 2012; Motta et al. 2014; Sousa et al. 2017; da Silva et al. 2019). Among the several types of molecular markers, microsatellites are most commonly used in genetic diversity study because of several advantages, including a high degree of polymorphism,

repeatability, reproducibility, codominance, technical simplicity, speed, and multi allelism (Vieira et al. 2010).

Most of the studies, so far using these technologies, were done on commercial cultivars of Arabica coffee. Because of this, little is known about the genetic structure and a pattern of Arabica coffee accessions. However, the genetic diversity of a few number of coffee accession in Ethiopia has been studied (Teressa et al. 2010). The works have shown that there is sufficient diversity in Ethiopia, especially the South-Western part of the country. However, each coffee producing region has its own unique trait which should be studied independently and can be used for future breeding efforts. Therefore, analyzing the genetic diversity using different marker systems is vital to minimize the loss of coffee genetic resources and the best strategy for conservation and sustainable utilization for future breeding efforts. Hence, in this study, we analyze the genetic diversity among and within populations of *Coffea arabica* accessions collected from a range of eco-geographical zones of Ethiopia.

Materials and methods

Plant materials

In this study, a total of 86 *Coffea arabica* accessions were used. The accessions were obtained from Ethiopian Biodiversity Institute (EBI) and they were collected from different agro-ecologies varying in altitude, soil type, rainfall, temperature, and represent the major coffee growing regions of Ethiopia. The list of accessions used is given in Table 1.

DNA extraction

The entire experiment was conducted at the National Agricultural Biotechnology Research Center (NABRC), located at Holeta, 29 km west of Addis Ababa. Young leaf samples that are free from the disease were collected from the actively growing tips of the coffee branches and preserved and dried using silica gel (Blulux laboratory (p) Ltd.-121,005). The leaves were transported to NABRC for DNA extraction. Genomic DNA was extracted from dried leaf samples according to Khanuja et al. (1999) with some modifications. About 0.5 g of dried leaf tissue was placed in 2-ml Eppendorf tube and ground using a geno-grinder (Retsch MM200, Germany). Nearly 1 ml of freshly prepared extraction buffer (100 mM Tris-Cl (pH 8.0), 25 mM EDTA, 1.5 M NaCl, 2.5% CTAB, 0.2% β -mercaptoethanol (v/v), and 1% PVP (w/v) was added to the grinded material and mixed inversely. The mixed solution was incubated at 60 °C in a shaking water bath (100 rpm) for 2 h followed by addition of 1 ml chloroform: isoamyl alcohol (24:1). The solution was mixed by inverting the tubes for about 15 min and then

Table 1 List of *Coffea arabica* accessions used in the study

S/No.	Code	Ac. No.	Region	Zone/region	Latitude	Longitude	Altitude
1	HG01	244,460	Oromia	Hararge	N09° 04.95	E41° 21.68	1700
2	HG02	244,335	Oromia	Hararge	N09° 11.12	E41° 30.01	1960
3	HG03	244,438	Oromia	Hararge	N09° 05.95	E41° 21.24	1780
4	HG04	244,355	Oromia	Hararge	N09° 17.37	E41° 31.47	1880
5	HG05	244,430	Oromia	Hararge	N09°04.99	E41° 22.53	1720
6	HG06	244,434	Oromia	Hararge	N09° 04.92	E41° 22.52	1700
7	HG07	244,386	Oromia	Hararge	N09° 02.02	E41° 19.45	1700
8	HG08	244,341	Oromia	Hararge	N09° 11.36	E41° 29.90	1960
9	HR09	244,216	Harari	Harar	N09° 23.20	E42° 14.73	1970
10	HR10	244,230	Harari	Harar	N09° 15.34	E42° 08.99	1840
11	HR11	244,227	Harari	Harar	N09° 22.53	E42° 14.99	2980
12	HR12	244,224	Harari	Harar	N09° 23.34	E42° 14.63	2010
13	HR13	244,232	Harari	Harar	N09° 15.27	E42° 08.93	1830
14	HR14	244,229	Harari	Harar	N09° 15.43	E42° 08.99	1830
15	HR15	244,233	Harari	Harar	N09° 15.23	E42° 08.94	1830
16	HR16	244,239	Harari	Harar	N09° 15.85	E42° 08.60	1910
17	IB17	30,744	Oromia	Ilubabor	N09° 27.18	E81° 45.43	1424
18	IB 18	30,731	Oromia	Ilubabor	N09° 26.97	E81° 57.86	1461
19	IB 19	30,745	Oromia	Ilubabor	N09° 27.16	E81° 45.53	1423
20	IB 20	30,655	Oromia	Ilubabor	N08° 11.13	E35° 43.03	1706
21	IB 21	30,638	Oromia	Ilubabor	N08° 30.10	E35° 42.39	1320
22	IB 22	30,655	Oromia	Ilubabor	N08° 11.13	E35° 43.03	1706
23	IB 23	212,085	Oromia	Ilubabor	N08° 34.12	E35° 47.18	1500
24	IB 24	212,095	Oromia	Ilubabor	N08° 18.16	E35° 34.36	1530
25	IB25	22,873	Oromia	Ilubabor	N07° 11.13	E61° 45.53	1950
26	YC26	24,909	SNNP	Gedeo/Yirgachefe	N06° 06.19	E38° 12.04	2036
27	YC27	24,909	SNNP	Gedeo/Yirgachefe	N06° 06.19	E38° 12.04	2036
28	YC28	24,902	SNNP	Gedeo/Yirgachefe	N06° 05.59	E38° 12.11	2100
29	YC29	211,947	SNNP	Gedeo/Yirgachefe	N06° 08.76	E38° 12.39	1880
30	YC30	24,910	SNNP	Gedeo/Yirgachefe	N06° 09.55	E38° 12.54	2030
31	YC31	24,907	SNNP	Gedeo/Yirgachefe	N06° 79.03	E38° 22.00	2222
32	YC32	24,912	SNNP	Gedeo/Yirgachefe	N06° 04.14	E38° 09.21	1920
33	YC33	24,887	SNNP	Gedeo/Yirgachefe	N06° 02.11	E38° 08.56	1718
34	YC34	24,900	SNNP	Gedeo/Yirgachefe	N06° 06.08	E38° 12.09	2076
35	WG35	8714	Oromia	Wolegga	N09° 13.25	E36° 30.15	1900
36	WG36	8721	Oromia	Wolegga	N09° 08.58	E36° 24.24	1882
37	WG37	8730	Oromia	Wolegga	N09° 12.10	E36° 22.06	1579
38	WG38	8744	Oromia	Wolegga	N09° 11.11	E36° 27.00	1664
39	WG39	8750	Oromia	Wolegga	N09° 11.37	E36° 30.28	1706
40	WG40	212,403	Oromia	Wolegga	NA	NA	1500
41	WG41	8753	Oromia	Wolegga	N09° 09.17	E36° 31.02	1888
42	WG42	8760	Oromia	Wolegga	N10° 30.37	E35° 40.02	1745
43	WG43	8754	Oromia	Wolegga	N09° 08.35	E36° 31.23	1868
44	KF44	211,971	SNNP	Kefficho/Shekicho	NA	NA	1700
45	KF45	211,973	SNNP	Kefficho/Shekicho	NA	NA	1750
46	KF46	211,978	SNNP	Kefficho/Shekicho	NA	NA	1650
47	KF47	211,975	SNNP	Kefficho/Shekicho	NA	NA	1720
48	KF48	211,927	SNNP	Kefficho/Shekicho	NA	NA	1640
49	KF49	219,392	SNNP	Kefficho/Shekicho	NA	NA	1650
50	KF50	212,250	SNNP	Kefficho/Shekicho	NA	NA	1940
51	KF51	219,394	SNNP	Kefficho/Shekicho	NA	NA	1840

Table 1 (continued)

S/No.	Code	Ac. No.	Region	Zone/region	Latitude	Longitude	Altitude
52	KF52	211,926	SNNP	Kefficho/Shekicho	NA	NA	1700
53	BM53	212,033	SNNP	Bench Maji	N06° 59.66	E035° 34.11	1520
54	BM54	212,030	SNNP	Bench Maji	N07° 02.03	E035° 32.77	1440
55	BM55	212,032	SNNP	Bench Maji	N07° 02.91	E035° 29.74	1320
56	BM56	212,032	SNNP	Bench Maji	N07° 02.88	E035° 29.74	1650
57	BM57	212,039	SNNP	Bench Maji	N07° 02.91	E035° 29.75	1350
58	BM58	212,036	SNNP	Bench Maji	N07° 04.13	E035° 37.74	1400
59	BM59	212,039	SNNP	Bench Maji	N07° 02.88	E035° 29.74	1600
60	BM60	212,048	SNNP	Bench Maji	N06° 59.69	E035° 34.9	1650
61	OM61	8654	SNNP	Omo	N05° 58.39	E36° 34.27	1643
62	OM62	8662	SNNP	Omo	N05° 47.30	E36° 32.51	1500
63	OM63	8678	SNNP	Omo	N05° 49.6	E037° 49.11	1360
64	OM64	8693	SNNP	Omo	N06° 26.44	E36° 32.7	1500
65	OM65	8711	SNNP	Omo	N06° 18.9	E36° 37.26	1860
66	OM66	8812	SNNP	Omo	N07° 11.53	E37° 14.24	1465
67	OM67	8789	SNNP	Omo	N07° 09.41	E37° 03.55	1642
68	OM68	8800	SNNP	Omo	N07° 08.41	E37° 10.43	1426
69	OM69	8820	SNNP	Omo	N06° 58.09	E37° 19.43	1484
70	JM70	218,931	Oromia	Jimma	N07° 40.43	E036° 48.76	1740
71	JM71	218,940	Oromia	Jimma	N07° 40.46	E036° 48.79	1640
72	JM72	211,967	Oromia	Jimma	N07° 40.58	E036° 48.75	1750
73	JM73	218,930	Oromia	Jimma	N07° 41.87	E036° 48.13	1950
74	JM74	212,330	Oromia	Jimma	N07° 31.37	E036° 53.44	1900
75	JM75	218,921	Oromia	Jimma	N07° 31.28	E036° 53.63	1610
76	JM76	211,932	Oromia	Jimma	N07° 46.80	E036° 48.84	1800
77	JM77	211,966	Oromia	Jimma	N07° 44.89	E036° 46.87	1740
78	JM78	212,335	Oromia	Jimma	N07° 41.89	E036° 49.94	1760
79	BL79	23,376	Oromia	Bale	N06° 24.00	E39° 35.00	1520
80	BL80	23,384	Oromia	Bale	N06° 20.00	E39° 37.00	1540
81	BL81	23,368	Oromia	Bale	N06° 24.00	E39° 35.00	1540
82	BL82	23,427	Oromia	Bale	N06° 24.00	E39° 35.00	1560
83	BL83	23,440	Oromia	Bale	N06° 24.00	E39° 35.00	1560
84	BL84	23,447	Oromia	Bale	N07° 08.00	E40° 48.00	1510
85	BL85	23,453	Oromia	Bale	N07° 08.00	E40° 48.00	1510
86	BL86	23,484	Oromia	Bale	N07° 08.00	E40° 48.00	1670

SNNP South Nation Nationalities and Peoples, *NA* absent

centrifuged at 8000 rpm for 10 min at 25–30 °C. Carefully, the upper clear aqueous layer was transferred to another 2-ml Eppendorf tube and properly mixed with 0.75 ml of 5 M NaCl. About 0.6 volume of isopropanol was added and placed at room temperature for 1 h. After 1 h, the mixture was centrifuged at 10,000 rpm for 10 min at 25–30 °C and the supernatant was discarded and the pellet washed with 80% ethanol. The pellet was dried in a vacuum for 15 min and dissolved in 0.5 ml of high salt TE buffer (1 M NaCl, 10 mM Tris–Cl (pH 8.0) and 1 mM EDTA). After the pellet was dissolved in high salt TE buffer, 2 µl of RNase was added and incubated at 37 °C for 30 min to degrade RNA. An equal volume of chloroform: isoamyl alcohol (24:1)

was added and the aqueous layer was transferred to fresh 1.5 ml microfuge tube and 2 volumes of cold ethanol was added to precipitate the DNA. The mixture was centrifuged at 10,000 rpm for 10 min at 25–30 °C and the pellet was washed with 80% ethanol. Finally, the pellet was dried in a vacuum and dissolved in 100 µl of nuclease free water. The genomic DNA concentration was estimated at 260 and 280 nm using the Nanodrop (Nano Drop® ND-800). Similarly, the quality of DNA was checked using 0.8% agarose gel by loading 5 µl genomic DNA after mixing with 2 µl 6X loading dye with gel red. The band of DNA was visualized using gel documentation (Bio-Doc-It™ imaging system) under UV light.

SSR markers used

Initially 24 SSR markers previously described as being polymorphic in *C. arabica* were selected from published literature (Rovelli et al. 2000; Combes et al. 2000; Moncada and McCouch 2004; Cubry et al. 2008). The annealing temperatures of the primers were optimized by performing gradient PCR. Finally, 20 polymorphic SSR markers were selected by excluding four markers that failed to amplify properly. Detail of the SSR markers used for this study is given in Table 2.

PCR amplification

PCR was conducted in 12.5 µl reaction volume containing: One Taq® 2X Master Mix with Standard Buffer, 10 µM of forward primers, 10 µM of reverse primers, Nuclease free water, and 20 ng/µl genomic DNA template. DNA free 10.5 µl reaction volume was used as a control. The PCR Conducted in Bio-Rad T100™ thermal cycler.

Table 2 List of selected SSR markers with their marker name, forward/reverse primer sequences and repeat motifs

S/No.	Locus	5'3' Forward/5'3' Reverse primer	Repeat motifs	Annealing T°	References
1	AJ308782	F:AAAGGAAAATTGTTGGCTCTGA R:TCCACATACATTCCCGCA	(GT)15	58.1	Rovelli et al. (2000)
2	AJ308779	F:TCCCGATCTTTTCTTTCC/ R:GGGAGTGTTTTGTGTTGCTT	(TG)17	57.8	Rovelli et al. (2000)
3	AJ308753	F:CTTGTGTTGAGTCTGTGCTG/ R:TCCAGAAGTCTTGGGTT	(CA)15	56.1	Rovelli et al. (2000)
4	AJ250251	F:ATTCTCTCCCTCTCTGC/ R:TGTGTGCGGTTTTCTTG	(CA)8	58.1	Combes et al. (2000)
5	CFGA92	F:TGAGGGCAAAGGAGTAAGAAAG/ R:TCAAACCTCAACAATCAAATACCC	(AG)10	60.1	Moncada and McCouch (2004)
6	AJ250253	F:GGAGACGCAGGTGGTAGAAG/ R:TTTCCCTCCCAATGTCTGTA	(GA)5(GT)8TT(GT)4T T(GT)7(GA)11(TC)2 (CT)GT	59.2	Combes et al. (2000)
7	CFGA91	F:CTTCTCCAGCTTTAGGTTCACTTTG/ R:TTTTGAATACTGGCTCGTGAACCT	(AG)17	55	Moncada and McCouch (2004)
8	CFGA69	F:TGGTGGAGTGGCTTTGATTGATG/ R:GCAACTTATGAGCCTAATCC	AG)14(GT)11	55.8	Moncada and McCouch (2004)
9	CFGA502	F:AAGCCACCCAGAAAACAGCACATC/ R:ATTTGCTTCTCATGTTCCCTTTCA	(AG)27	62	Moncada and McCouch (2004)
10	AJ250255	F:CCCTCCCTGCCAGAAGAAGC/ R:AACCACCGTCCTTTTCCTCG	(GT)5CT(GT)2/(GT)12	57.8	Combes et al. (2000)
11	CFGA465	F:ACATCCCCTTGCCATTTCTCT/ R:ACCCTTACTCATTATTTACTCTC	(AG)18	55.7	Moncada and McCouch (2004)
12	AJ250257	F:GACCATTACATTTACACAC/ R:GCATTTTGTGCACACTGTA	(CTCACA)4/(CA)9	55.7	Combes et al.(2000)
13	AJ308774	F:GCCACAAGTTTCGTGCTTTT/ R:GGGTGTGCGGTGTAGGTGTATG	(CT)5 (CA)7	55	Cubry et al. (2008)
14	AJ250258	F:AACTCTCCATTCCCGCATTC/ R:CTGGGTTTTCTGTGTTCTCG	(CA)3/(CA)3/(CA)18	56.9	Combes et al.(2000)
15	CFGA100	F:TGACTCTTTTCTCTCCCAA/ R:ATTTAGCAGGCTTGGCATTTTT	(AG)15	56.1	Moncada and McCouch (2004)
16	AJ308809	F:AGCAAGTGGAGCAGAAGAAG/ R:CGGTGAATAAGTCGCAGT	(CA)15(CG)4CA	60.1	Cubry et al. (2008)
17	AJ250260	F:TGATGGACAGGAGTTGATGG/ R:TGCCAATCTACCTACCCCTT	(CT)9(CA)8/(CT)4/(CA)5	58.1	Combes et al.(2000)
18	AJ308790	F:TTTTCTGGGTTTTCTGTGTTCTC R:TAACCTCTCCATTCCCGCATT	(GT)21	57.8	Rovelli et al. (2000)
19	AJ308755	F:CCCTCCCTCTTTCTCCTCTC R:TCTGGGTTTTCTGTGTTCTCG	(CA)20	56.9	Cubry et al. (2008)
20	AJ308837	F:CTCGCTTTCACGCTCTCTCT R:CGGTATGTTCTCCTGTTCTCT	(GT)16(GA)11	61.1	Rovelli et al. (2000)

F forward primer, R reverse primer

Gel electrophoresis

The amplified PCR products were separated using 3.5% (w/v) agarose gel by loading 10 µl of each of the PCR product mixed with 2 µl of loading dye and gel red. The gels were run in 1X TAE (40 mM Tris, 40 mM Acetate, 1 mM EDTA, pH 8) buffer and at 85 voltages, 100 mA for 3.5 h. Finally, the amplified bands were documented under UV light in gel documentation.

Data scoring and analysis

The amplified SSR fragment sizes on the gel were estimated using PyElph 1.4 software package. Locus based diversity indices including major allele frequency (MAF), number of alleles (Na), gene diversity; polymorphic information contents (PIC) were computed using Power Marker v3.25 software (Liu and Muse 2005). The effective number of alleles, Shannon's Information index (I), and Gene flow (Nm) were calculated using POPGENE version 1.31 (Yeh and Yang 1999). Allelic frequency, observed heterozygosity (Ho), expected heterozygosity (He), fixation index (F), and the estimate of the deviation from Hardy–Weinberg equilibrium (HWE) over the entire populations were computed with GenAlEx version 6.501 software (White and peakall 2015). The same software package was used to compute population differentiation test: Wrights fixation indices and pairwise population differentiation (Fst), analysis of molecular variance (AMOVA) and the estimate of the variance components. Moreover, rarefied allelic richness (Ar) and private rarefied allelic richness (Arp) were computed using HP-Rare 1.1 software (Kalinowski 2005). Genetic relationships among the different populations and genetic dissimilarity matrix were calculated using Jaccard's coefficient and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering was generated using DARwin var. 6.0.14 (Perrier and Jacquemoud-Collet 2006). Principal coordinate analysis (PCoA) was performed using the GenALEX to produce Eigen values and Eigen vectors which reveal both total variance and the loci that are important to the variance. Biplots were generated to reveal associations among accessions. Population structure and admixture patterns were determined using STRUCTURE software ver. 2.3.4 based on Bayesian algorithm (Pritchard et al. 2000). To estimate the true number of population cluster (K), a burn-in period of 50,000 was used in each run, and data were collected over 500,000 Markov Chain Monte Carlo (MCMC) replications for $K=1$ to $K=10$ using 20 iterations for each K. The optimum K value was predicted following the simulation method of Evanno et al. (2005) using the web-based STRUCTURE HARVESTER ver. 0.6.92 (Earl and vonHoldt 2012). Bar plot for the optimum K was determined using Clumpak beta version (Kopelman et al. 2015).

Results and discussion

Polymorphism and allelic diversity among SSR markers

A total of 112 SSR alleles were produced across 20 loci, with the estimated product size range of 100 to 385 base pairs (Table 3). The mean number of alleles per locus was 5.6. The detected number of alleles ranged from 3 (AJ250253 and CFGA69) to 11 (CFGA465). This result is higher than that of a previous study by Anthony et al. (2002) who reported an average of 4.7 alleles per SSR using six SSR in Arabica coffee collections. Moncada and McCouch (2004) reported an average of 2.5 alleles per SSR in 11 wild Arabica coffee genotypes and 12 cultivated genotypes, respectively, with the number of alleles ranging from one to eight. Maluf et al. (2005) also reported an average number of 2.87 alleles in 28 cultivated Arabica lines using 23 SSR markers. Bigirimana et al. (2013) reported an average of 3.2 alleles in 14 genotypes using six SSR markers. In the present study, the higher average number of alleles (5.6) may be due to high diversity in Ethiopian *C. arabica* accessions. In another study, Teressa et al. (2010) reported a mean of 6.5 alleles per locus that ranged from 2 to 14 using 133 coffee genotypes. In another study, da Silva et al. (2019) reported 6.9 mean number of allele per locus in 36 *Coffea* genotypes using 30 SSR markers that is higher than the present study. Hussein et al. (2017) also reported higher mean number of allele per locus (10.7) using 15 SSR markers. Generally, the authors suggested the number of alleles depends on markers, the platform used for resolution of amplified products and the sources of genotypes used.

The polymorphic information content (PIC) value, ranged from 0.45 to 0.75 with a mean value of 0.6. About 16 (80%) of the markers were highly informative (i.e. $PIC > 0.50$). The rest four markers were moderately informative with PIC values less than 0.5. Al-Murish et al. (2013) reported 0.43 mean PIC value which is a lower than that of the present study. Similarly, Moncada and McCouch (2004) reported lower average PIC value of 0.35 among 11 *Coffea arabica* genotypes. In the present study, the higher mean PIC value (0.6) could be attributed to the diversity of genotypes used. In other studies, da Silva et al. (2019) reported 0.72 mean PIC values per locus in 36 *Coffea arabica* using 30 SSR markers that is higher than the present study. Such PIC value that observed in present and past studies provides an estimate of the discriminatory power of a locus (Marulanda et al. 2014).

About 17 (85%) of the markers showed significant ($p < 0.05$) deviation from Hardy–Weinberg equilibrium (HWE). Only three markers (CFGA465 $p=0.17$),

Table 3 In-formativeness and levels of different diversity indices of the SSR loci across *Coffea arabica* populations

Locus	EPS	MAF	GD	PIC	Na	Nm	I	Ar	Arp	He	Ho	Ne	F	P_{HWE}^a
Aj308782	130–150	0.25	0.79	0.67	5	0.82	1.08	3.56	0.00	0.61	0.00	2.65	1.00	0.001**
Aj308779	130–180	0.27	0.79	0.66	6	2.78	1.40	4.61	0.01	0.72	0.99	3.86	−0.37	0.008**
Aj308753	340–310	0.17	0.87	0.75	8	1.25	1.45	4.91	0.00	0.72	0.00	4.10	1.00	0.001**
Aj250251	300–370	0.57	0.61	0.48	5	1.38	0.94	3.26	0.00	0.52	0.00	2.41	1.00	0.001***
CFGA92	345–385	0.28	0.78	0.65	6	2.05	1.30	4.22	0.78	0.69	0.01	3.35	0.98	0.001***
AJ250253	320–350	0.42	0.63	0.48	3	1.51	0.87	2.69	0.00	0.55	0.00	2.24	1.00	0.001**
CFGA91	270–300	0.40	0.69	0.55	7	2.24	1.10	3.49	0.19	0.62	0.06	2.90	0.91	0.001**
CFGA69	190–220	0.50	0.62	0.45	3	2.10	0.89	2.79	0.00	0.55	0.00	2.27	1.00	0.005**
CFGA502	285–320	0.52	0.64	0.50	4	2.10	0.98	3.18	0.00	0.57	0.00	2.47	1.00	0.001**
AJ250255	200–230	0.42	0.71	0.57	4	1.39	1.09	3.56	0.00	0.60	0.00	2.76	1.00	0.001***
CFGA465	180–280	0.33	0.83	0.73	11	3.36	1.71	6.54	0.03	0.77	0.88	4.62	−0.14	0.170 ^{ns}
AJ250257	130–170	0.38	0.76	0.62	5	6.98	1.44	4.81	0.00	0.73	0.98	3.82	−0.34	0.008**
AJ308774	160–190	0.42	0.70	0.55	5	8.33	1.25	4.03	0.00	0.68	0.97	3.17	−0.42	0.088 ^{ns}
AJ250258	100–120	0.40	0.65	0.40	4	1.18	0.88	2.78	0.00	0.54	0.00	2.32	1.00	0.001**
CFGA100	300–325	0.27	0.81	0.69	6	1.85	1.34	4.24	0.00	0.71	0.00	3.67	1.00	0.003**
AJ308809	140–170	0.27	0.83	0.71	7	1.98	1.41	4.46	0.00	0.73	0.00	3.86	1.00	0.001***
AJ250260	180–220	0.33	0.77	0.63	6	1.03	1.13	3.74	0.88	0.62	0.01	2.82	0.98	0.013*
AJ308790	140–160	0.41	0.70	0.56	5	1.61	1.04	3.18	0.00	0.61	0.01	2.69	0.98	0.001**
AJ308755	110–150	0.21	0.84	0.72	7	3.24	1.64	5.77	0.00	0.78	0.92	4.55	−0.18	0.062 ^{ns}
AJ308837	150–175	0.35	0.74	0.61	5	1.76	1.16	3.65	0.01	0.65	0.03	3.06	0.96	0.001**
Mean	–	0.36	0.72	0.60	5	2.45	1.21	3.97	0.10	0.65	0.24	3.18	0.67	–

EPS estimated product size in base pair, MAF major allele frequency, Na number of allele, PIC polymorphism information content, Ne number of effective alleles, Ho observed heterozygosity, He expected heterozygosity, F fixation index, I Shannon's diversity index, ns not significant, P_{HWE}^a P value for deviation from Hardy–Weinberg equilibrium, Nm gene flow estimated from $F_{st}=0.25(1-F_{st})/F_{st}$, GD gene diversity, Arp private allelic richness, Ar = allelic richness

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ = significant

(AJ308774 $p = 0.088$) and (AJ308755 $p = 0.062$) showed non-significant deviation from HWE. The loci studied displayed differences between Ho and He in which some of them showed excess heterozygosity that led to a significant departure from HWE across populations. Such excess heterozygosity is expected in sexually reproducing organisms that can maintain their heterozygosity through sexual recombination, or other factors such as natural selection pressure and excess gene flow (Gadissa et al. 2018). In other ways, a high number of alleles and large gene diversity consequently give a high value of expected heterozygosity (Liu et al. 2005).

On the other hand, the lowest fixation index (−0.42) was observed for AJ308774 marker (Table 3). The fixation index of the markers (F) ranged from 0 to 1 with a mean of 0.67 (Table 3). Five markers had negative fixation index values, indicating an excess of heterozygotes. For loci Aj308779 (72%), CFGA465 (77.4%), AJ250257 (73.2%), AJ308774 (68%), and AJ308755 (77.8%), genotypes were expected to be heterozygous under the specific locus under random mating conditions. However, 98.8%, 88.3%, 97.8%, 96.5%, and 91.9% of the genotypes were heterozygotes for Aj308779, CFGA465, AJ250257, AJ308774, and AJ308755,

respectively. The F value revealed that five of the twenty markers showed an excess of heterozygotes (negative F value) which might be due to mutation at specific loci.

Genetic diversity within and among populations

Analysis of genetic diversity parameters

The comparative analysis showed that there is no much difference among the ten populations with regard to genetic diversity indices. Comparatively, the population from Omo had the higher genetic diversity indices. Next to population from Omo, the population from Ilubabor and Bench Maji scored the highest number of genetic diversity indices parameters. They are genetically more diverse than the other populations as estimated by genetic diversity indices parameters. Hence, the areas, where these populations were collected, could be considered as genetic diversity hot spots and potential *in-situ* conservation sites for *C. arabica* germplasm. Similarly, Tadele et al. (2014) reported higher genetic diversity in the population that was collected from South Omo. In addition to this, in terms of allelic richness the populations from Omo, Ilubabor, Bench Maji and Kefficho

are the top four in that order, and hence they are more interesting in terms of genetic and evolutionary studies (Leberg 1991). On other hand, the populations from Bale, Hararge and Harar scored the least number of genetic diversity indices parameters in order of their magnitude, which might suggest current rapid genetic erosion in the area.

The mean observed heterozygosity (0.24) observed in the present study is lower than the half mean of expected heterozygosity (0.65) observed ($H_o < H_e$) indicating that the population is inbreeding (Govindraj et al. 2015). The mean value of observed heterozygosity (H_o) observed in the present study is higher than that of Hussein et al. (2017) who reported 0.212 mean observed heterozygosity of Yemeni coffee (*Coffea arabica* L.). Combes et al. (2000) also reported 0.043 average H_o value in *Coffea arabica* using eleven SSR primers that is lower than the present study. In other studies, Al-murish et al. (2013) reported 0.32 mean observed heterozygosity (H_o) that higher than the present study. Overall the lower observed heterozygosity level in this study and other studies might be due to the autogamous nature of *C. arabica* which contributes to low heterozygosity levels.

The polymorphism detected in the present study was similar to the result reported by Hussein et al. (2017) who observed 100% polymorphism of markers in Yemeni coffee (*Coffea arabica* L.). This higher degree of polymorphism might be due to the difference in discrimination power of the genetic marker and the source of genotypes used in the present study. Hence, SSR markers are more powerful and very informative for genetic diversity analysis than any other markers (Vieira et al. 2010; Pestiana et al. 2015). In contrast to the present study, in several studies, a small number of polymorphic loci were observed for *C. arabica* (Combes et al. 2000; Vieira et al. 2010; Pestiana et al. 2015). These

might be due to the low genetic variability of species and narrow genetic base of origin.

The mean diversity indices parameters obtained in the present study showed a high level of genetic variation within populations of Ethiopian coffee. This might be due to a relatively different genetic basis of the populations included in this study that resulted from different germplasm resources accessible to farmers, or due to addition in population size, both natural as well as human factors such as, sexual reproduction of the *C. arabica* and free movement of germplasm between different geographic areas (Table 4).

Population genetic differentiation and gene flow

Analysis of molecular variance (AMOVA) partitioned the total molecular variance within and among the populations based on their area of collections. There were highly significant ($P < 0.001$) molecular variances among populations, among individuals within populations and within individuals. The highest proportion, 63% of the variation was attributed to genetic variability among individuals within populations, while 33% was due to variation within individuals in the population. In contrast, a smaller portion of the total variation (4%) was among populations suggesting that all populations are genetically similar (Table 5). In several studies, Aga et al. (2003), Tesfaye (2006), Balemi (2007), and Tadele et al. (2014) observed high variation within the population as compared to among population variation, which is in agreement with the results of this study.

The higher variation within populations' genetic diversity might be accounted to contrary reasons. *Coffea arabica* is affected by multiple evolutionary forces which operate within historical and biological context of the plant species. This includes the mating types, gene flow, mode of

Table 4 Summary of genetic diversity indices of Arabica coffee populations classified by area of collection

Populations	Genetic parameter								
	<i>N</i>	<i>Na</i>	<i>Ne</i>	<i>Ar</i>	<i>Arp</i>	<i>I</i>	<i>He</i>	<i>Ho</i>	<i>%p</i>
Hararge	8.00	3.550	2.92	3.52	0.00	1.11	0.63	0.24	100
Harar	8.00	3.750	2.90	3.71	0.05	1.12	0.62	0.26	100
Ilubabor	9.00	4.350	3.45	4.21	0.00	1.28	0.68	0.23	100
Yirgachefe	9.00	4.300	3.09	4.15	0.00	1.21	0.63	0.24	100
Wolegga	9.00	3.950	3.13	3.85	0.01	1.19	0.65	0.22	100
Kefficho	9.00	4.350	3.18	4.16	0.05	1.24	0.66	0.24	100
Benchi Maji	8.00	4.150	3.27	4.09	0.04	1.25	0.67	0.24	100
Omo	9.00	4.400	3.60	4.26	0.04	1.31	0.69	0.24	100
Jimma	9.00	4.250	3.34	4.14	0.01	1.23	0.65	0.238	100
Bale	8.00	3.700	2.91	3.65	0.04	1.11	0.62	0.25	100
Mean	-	4.08	3.18	3.97	0.024	1.21	0.65	0.24	100

N number of individual within each population, *Na* Number of different alleles, *Ne* number of effective alleles, *Ar* allelic richness, *Arp* private allelic richness, *I* Shannon diversity index, *He* expected heterozygosity, *Ho* observed heterozygosity, *%p* percentage of polymorphic loci

Table 5 Analysis of molecular variance (AMOVA) across the full data set of 86 *Coffea arabica* accessions according to area of collection

Source	DF	SS	MS	Estimated variance	Percentage variation (%)	P value	F statistics
Among populations	9	149.878	16.653	0.276	4	0.001	Fst = 0.037
(among individuals within populations)	76	905.237	11.911	4.717	63	0.001	Fis = 0.656
Within individuals	86	213.048	2.477	2.477	33	0.001	Fit = 0.668
Total	171	1268.16		7.470	100	–	

DF degree of freedom, SS sum of squares, MS mean squares

reproduction and natural selection. In addition, the high genetic diversity observed within populations of the *Coffea arabica* might be due to preferential adaptive gene complexes adapted to environmental changes being evolved during the long evolutionary period in a given region (Loveless and Hamrik 1984; Tadele et al. 2014).

The overall observed gene flow (Nm) or gene migration value observed in this study on Ethiopian Arabica coffee was 2.45. Based on Slatkin (1985) and Waples (1987), Nm values grouped into three categories: Nm > 1.00 high, 0.25–0.99 intermediate, and 0.000–0.249 low. Therefore, the Nm value observed in this study indicates high gene flow between populations. Gene flow between population could occur through either seeds or seedlings exchange. Moreover, coffee farmers could also contribute to gene flow by exchanging seeds and seedlings of enhanced landraces among nearby districts. This leads to maximize the diversity of local genotypes and increase the distribution of alleles among different populations regardless of their geographical distance.

The magnitude between and within population differentiation was quantified using F-statistics (Fst, Fis, and Fit) also known as “Wright 1951”. According to Wright (1951), the threshold to determine the level of Fst value ranges from 0 to 0.05 considered as low, 0.05–0.15 moderate, 0.15–0.25 large and those greater than 0.25 mean

very large genetic differentiation among populations. In the current study the overall F statistics used as a measure of population differentiation was low (Fst = 0.037 which is < 0.05), which indicates that *C. arabica* has very low genetic differentiation among populations, which accounted only for 3.7% of the total genetic variation (Table 5). This low genetic differentiation among population may be due to high gene flow that resulted from the movement of seeds from one location to another.

The pairwise genetic differentiation among populations within the source of origin ranged from 0.041 to 0.108 (Table 6). The lowest population differentiation was observed between Benchi Maji and Wolegga, Omo and Wolegga, and Jimma and Wolegga, whereas the highest population differentiation observed between Bale and Hararge, Kefficho and Hararge, and Jimma and Bale populations in order of their magnitude. The low value of Fst implies that there is high frequency of identical alleles among *Coffea arabica* accessions. The lowest Fst value observed between a pair of populations can be explained by high level of gene flow, which leads to genetic similarity of populations. The low population genetic differentiation is supported by high gene flow owing to step-wise pollen movement across *C. arabica* populations. This study showed the minimal effects of regions or geographic origins on genetic variation among *Coffea arabica* populations.

Table 6 Pairwise population matrix Fst values for ten populations of *C. arabica* population

	HG	HR	IB	YC	WG	KF	BM	OM	JM	BL
HG	0.000									
HR	0.051	0.000								
IB	0.085	0.055	0.000							
YC	0.082	0.066	0.045	0.000						
WG	0.087	0.088	0.060	0.063	0.000					
KF	0.105	0.100	0.065	0.056	0.047	0.000				
BM	0.075	0.085	0.059	0.053	0.041	0.050	0.000			
OM	0.088	0.079	0.052	0.062	0.041	0.047	0.045	0.000		
JM	0.097	0.085	0.057	0.051	0.041	0.049	0.051	0.055	0.000	
BL	0.108	0.098	0.097	0.104	0.090	0.089	0.099	0.078	0.105	0.000

HG Hararge, HR Harar, IB Ilubabor, YC Yirgachefe, WG Wolegga, KF Kefficho/Shekicho, BM Bench Maji, OM Omo, JM Jimma, BL Bale, all pair wise Fst values are significant at P = 0.05

Genetic distance among populations

The magnitude of genetic distance between *Coffea arabica* populations was estimated based on Nei's (1978). The pairwise Nei's unbiased genetic distance value ranged from 0.048 to 0.426 whereas the mean Nei's unbiased genetic distance of the populations ranged from 0.16 to 0.35 with an overall mean of 0.21 (Table 6). The Population from Hararge showed the highest (0.426) pairwise Nei's unbiased genetic distance with Kefficho population. The population from Bale also, separated from Jimma and Hararge populations at higher pairwise Nei's unbiased genetic distance of 0.406 and 0.4, respectively. The population from Bale separated from all populations at highest mean Nei's unbiased genetic distance (0.35). This population was the most genetically distinct population with the highest mean Nei's unbiased genetic distance. The population from Hararge also scored the higher mean Nei's unbiased genetic distance (0.29) next to Bale population. Hence, these populations may serve as potential sources of new genetic variation of important traits that can be used in breeding programs.

Additionally, the smallest pairwise Nei's unbiased genetic distance was observed between Jimma and Wolegga and between Benchi Maji and Wolegga populations in order of their magnitude. Thus, the population from Wolegga showed the smallest pairwise Nei's unbiased genetic distance with Jimma and Benchi Maji populations with Nei's unbiased genetic distance of 0.048 and 0.051, respectively (Table 6). The result of the present study indicates the availability of low pairwise genetic distance and low overall genetic distance among all populations of *Coffea arabica* (Table 7).

Genetic relationships among accessions

Cluster analysis of 86 *Coffea arabica* genotypes revealed three major clusters, consisting of 2.3%, 58%, and 39.5% of the total populations in clusters I, II, and III, respectively (Fig. 1). Similarly, Hussein et al. (2017) grouped Yemeni coffee into three major clusters that contain seventeen genotypes involving 16 commercial cultivars and one accession of Yemeni coffee (*Coffea arabica* L.) germplasm using 16 SSR markers. In another study, Tadele et al. (2014) reported two major clusters of 87 *Coffea arabica* that collected from southern Ethiopia using five ISSR markers. The first cluster (Cluster I) contains only two accessions (YC30 and WG40) from Yirgachefe and Wolegga populations. These two accessions grouped together in cluster I regardless of their geographical origin. Clusters II and III, classified in to different hierarchical sub-groups. The second cluster included most of the genotypes (50) from all populations. Cluster III included 34 genotypes from nine populations. Only the Accessions from Hararge were not included in cluster III.

Accessions from different populations were clustered together, which may imply the existence of gene flow between and within populations. Hararge accessions were grouped with geographically distant accessions from Omo in cluster II. Accessions from Bale were also grouped with Ilubabor accessions. In all of the clusters many accessions are grouped with geographically distant populations. This indicates accessions in one cluster might be evolved from different lines of ancestry. In addition, the independent events of evolutionary forces such as genetic drift, mutation, migration, natural and artificial selection and germplasm exchange might be separated them into related but different gene pools (Slatkin 1985).

Generally, the cluster revealed a weak clustering pattern confirming low genetic differentiation among the

Table 7 Pairwise population matrix of Nei's unbiased genetic distance of *Coffea arabica* populations from different source of origin

	HG	HR	IB	YC	WG	KF	BM	OM	JM	BL
HG	0.000									
HR	0.074	0.000								
IB	0.287	0.116	0.000							
YC	0.247	0.170	0.073	0.000						
WG	0.295	0.292	0.164	0.161	0.000					
KF	0.426	0.396	0.204	0.120	0.087	0.000				
BM	0.225	0.272	0.155	0.104	0.051	0.102	0.000			
OM	0.333	0.245	0.125	0.150	0.059	0.085	0.075	0.000		
JM	0.336	0.259	0.123	0.095	0.048	0.093	0.097	0.118	0.000	
BL	0.4	0.325	0.377	0.383	0.315	0.315	0.383	0.254	0.406	0.000
Mean	0.29	0.24	0.17	0.17	0.16	0.2	0.16	0.16	0.175	0.35
Overall mean = 0.21										

HG Hararge, HR Harar, IB Ilubabor, YC Yirgachefe, WG Wolegga, KF Kefficho/Shekicho, BM Bench Maji, OM Omo, JM Jimma, BL = Bale

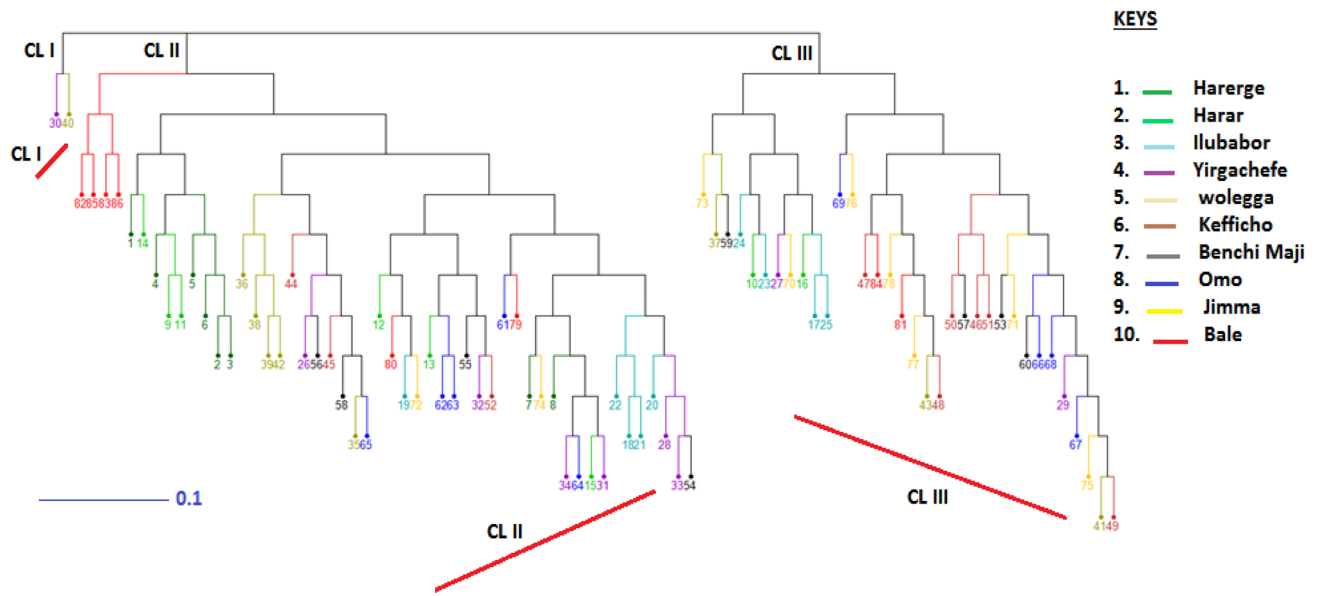


Fig. 1 UPGMA dendrogram for 86 *Coffea arabica* accessions based on the Jaccard's coefficient as revealed using 20 microsatellite markers

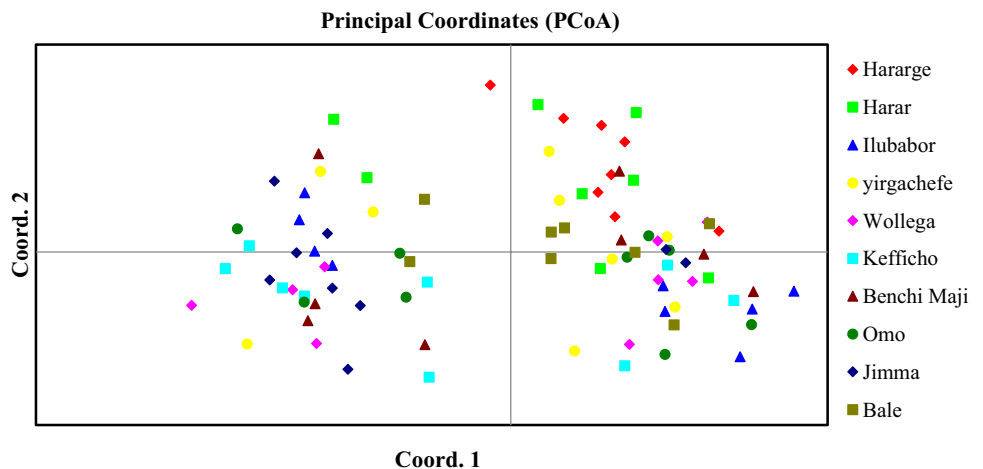
populations, suggesting the genetic background of *Coffea arabica* populations do not correlate with their geographical origin. Hence, the inter-mixed UPGMA dendrogram of 86 *Coffea arabica* individual plants indicate high level of genetic variation among individual genotypes investigated and the clustering pattern is weak to support the concept of “isolation by distance” (Fig. 1).

Principal coordinates analysis and population genetic structure

The PCoA analysis in the two dimensional plot displayed in Fig. 2 showed that accessions from different collection sites often grouped together. There was no separate group formed by a single population. This, in turn, agrees

with the results of the UPGMA dendrogram in that there was no unique clustering among accessions from the same population. In some cases, accessions of the same population such as Hararge and Ilubabor formed sub cluster in the major groups. Even though, some of the accessions forming a sub cluster in their specific group, there was no separate group formed by a single population. The overall grouping pattern of PCoA corresponds with the clustering of UPGMA dendrogram (Fig. 1) which explains about conformity of the results obtained from the cluster analysis. The presence of seed exchange and high gene flow between and within populations or collection sites may be the probable explanation behind the mixed clustering of accessions from different populations.

Fig. 2 Principal coordinates analysis (PCoA) bi-plot showing the clustering pattern of 86 *Coffea arabica* accessions based on 20 SSR. Accessions coded with the same symbol and color belongs to the same population



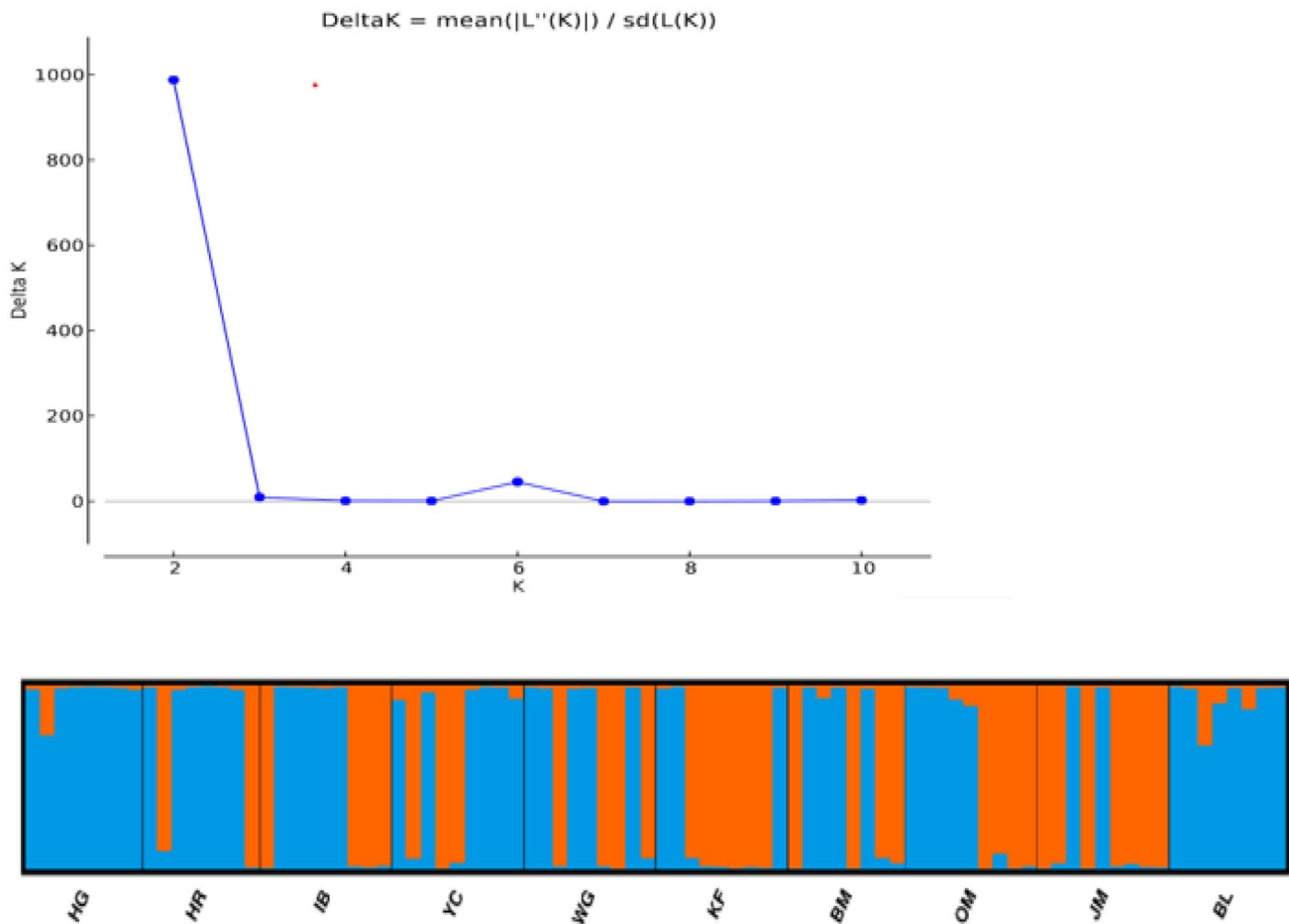


Fig. 3 Delta K value estimated using Evano et al. (2005) method (a) and Bayesian model-based estimation of population structure ($K=2$) (b) for the *Coffea arabica* L. individual plants in ten pre-determined populations

The Bayesian approach based assignment of the 86 individual plants to different populations and determination of their population structure, using STRUCTURE outputs, predicted $K=2$ to be the most likely number of clusters (Fig. 3a). Based on this value, Clumpak result (bar plot) showed admixtures and hence there was no clear geographic origin based structuring of populations (Fig. 3b). Structure analysis revealed a close relationship (weak sub-division) among the samples from the ten collection zones and in general, two inferred groups ($K=2$) with potential admixtures have been observed. It is interesting to indicate that all individual plants analyzed have alleles originated from the two clusters, which supports the presence of a strong gene flow that led to poor population genetic differentiation.

Acknowledgements The study was conducted by financial support provided by Ethiopian Institute of Agricultural Research Institute, National Agricultural Biotechnology Research Center. The authors are grateful to the Ethiopian Biodiversity Institute for providing the coffee germplasm.

Author contributions Material preparation, data collection, and analysis were performed by GD and TD. The first draft of the manuscript was written by GD and all authors commented on previous versions of the manuscript. All authors participated in designing the experiment, interpreting the data, drafting and revising the manuscript and approved the final manuscript.

Funding This work was supported by grants from the Ethiopian Institute of Agricultural Research Institute, National Agricultural Biotechnology Research Center.

Availability of data and materials Full passport data of the 86 coffee samples representing the 10 populations used in the present study is provided in Table 1. The other datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval and consent to participate Not Applicable.

Consent for publication Not Applicable.

References

- Aerts R, Berecha G, Gijbels P, Hundera K, Van Glabeke S, Vandepitte K, Muys B, Roldan Ruiz I, Honnay O (2013) Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rain forests. *Evol Appl* 6(2):243–252. <https://doi.org/10.1111/j.1752-4571.2012.00285.x>
- Aga E, Bryngelsson T, Bekele E, Balemi S (2003) Genetic diversity of forest arabica coffee (*Coffea arabica* L.) in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis. *Hereditas* 138:36–46. <https://doi.org/10.1034/j.1601-5223.2003.01636.x>
- Al-Abdulkader AM, Al-Namazi AA, Al Turki TA, Al-Khuraish MM, Al-Dakhil AI (2018) Optimizing coffee cultivation and its impact on economic growth and export earnings of the producing countries: the case of Saudi Arabia. *Saudi J Biol Sci* 25(4):776–782. <https://doi.org/10.1016/j.sjbs.2017.08.016>
- Al-Murish TM, Elshafei AA, Al-Doss AA, Barakat MN (2013) Genetic diversity of coffee (*Coffea arabica* L.) in Yemen via SRAP, TRAP and SSR markers. *J Food Agric Environ* 11(2):411–416. <https://doi.org/10.1234/4.2013.4315>
- Anthony F, Combes MC, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) the origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* 104:894–900. <https://doi.org/10.1007/s00122-001-0798-8>
- Balemi S (2007) Genetic diversity analysis of the wild *Coffea arabica* L. populations from Harena forest, Bale Mountains of Ethiopia, using Inter Simple Sequence Repeats (ISSR) marker. MSc thesis, Addis Ababa University, Addis Ababa, Ethiopia. P 89
- Benti T (2017) Progress in Arabica coffee breeding in Ethiopia: achievements, challenges and prospects. *Int J Sci Basic Appl Res* 33(2):15–25
- Bigirimana J, Njoroge K, Muthomi JW, Gahakwa D, Phiri NA, Gichuru EK, Walyaro DJ (2013) Genetic diversity among disease resistant coffee varieties and cultivars in Rwanda based on RAPD and SSR markers. *J Renew Agric* 1(6):106–112. <https://doi.org/10.12966/jra.09.01.2013>
- Bunn C, Läderach P, Ovalle-Rivera O, Kirschke D (2015) a bitter cup: climate change profile of global production of Arabica and Robusta coffee. *Clim Change* 129:89–101. <https://doi.org/10.1007/s10584-014-1306-x>
- Combes MC, Anderzejewski F, Anthony B, Bertrand P, Rovelli G, Graziosi and Lashermes, (2000) characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Mol Ecol* 9:1178–1180. <https://doi.org/10.1046/j.1365-294x.2000.00954-5.x>
- Cubry PP, Musali H, Legnate D, Pot FD, Bellis V, Poncet F, Anthony M, Dufour M, Leroy T (2008) diversity in Coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome* 51:50–63. <https://doi.org/10.1139/G07-096>
- da Silva BSR, Sant’Ana GC, Chaves CL, Androcioli LG, Ferreira RV, Sera GH, Charmetant P, Leroy T, Pot D, Domingues DS, Pereira LFP (2019) population structure and genetic relationships between Ethiopian and Brazilian *Coffea arabica* genotypes revealed by SSR markers. *Genetica* 147(2):205–216. <https://doi.org/10.1007/s10709-019-00064-4>
- Davis AP, Tosh J, Ruch N, Fay MF (2011) growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot J Linn Soc* 167(4):357–377. <https://doi.org/10.1111/j.1095-8339.2011.01177.x>
- Diniz LEC, Ruas CDF, Carvalho VDP, Torres FM, Ruas EA, Santos MDO, Sera T, Ruas PM (2005) Genetic diversity among forty coffee varieties assessed by RAPD markers associated with restriction digestion. *Braz Arch Biol Technol* 48(4):511–521. <https://doi.org/10.1590/S1516-89132005000500002>
- Earl DA, Von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Gadissa F, Tesfaye K, Dagne K, Geleta M (2018) Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC Genet* 19(1):1–15. <https://doi.org/10.1186/s12863-018-0682-z>
- Gebreselassie H, Atinafu G, Degefa M, Ayano A (2018) Arabica Coffee (*Coffea arabica* L) Hybrid genotypes evaluation for growth characteristics and yield performance under Southern Ethiopian growing condition. *Acad Res J Agric Sci Res* 6(2):89–96. <https://doi.org/10.14662/ARJASR2017.071>
- Geleta M, Herrera I, Monzón A, Bryngelsson T (2012) Genetic diversity of Arabica coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. *Sci World J* 2012:1–11. <https://doi.org/10.1100/2012/939820>
- Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity Assessment in crop plants and its recent advances: an overview of its analytical perspectives. A review article. *Genet Res Int* 2015:1–14. <https://doi.org/10.1155/2015/431487>
- Hussein MAA, Al-Azab AAA, Habib SS, El Sherif FM, El-Garhy HA (2017) Genetic diversity, structure and dna fingerprint for developing molecular ids of Yemeni Coffee (*Coffea arabica* L) germplasm assessed by SSR markers. *Egypt J Plant Breed* 21(4):713–736
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5:187–189. <https://doi.org/10.1111/j.1471-8286.2004.00845.x>
- Khanuja SP, Shasany AK, Darokar MP, Kumar S (1999) Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Mol Biol Rep* 17(1):74–74. <https://doi.org/10.1023/A:1007528101452>
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour* 15:1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Lashermes P, Combes MC, Cros J, Trouslot P, Anthony F, Charrier A (1995). Origin and genetic diversity of *Coffea arabica* L. based on DNA molecular markers. In: 16th Conference of ASIC, Kyoto (Japan), pp 528–535
- Leberg PL (1991) Influence of fragmentation and bottlenecks on genetic divergence of wild turkey populations. *Conserv Biol* 5(4):522–530. <https://doi.org/10.1111/j.1523-1739.1991.tb00359.x>
- Liu K, Muse SV (2005) Power marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>
- Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annu Rev Ecol Syst* 15:65–95
- Maluf MP, Silvestrini M, Ruggiero LM, Guerreiro Filho O, Colombo C (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola* 62(4):366–373. <https://doi.org/10.1590/S010390162005000400010>

- Marulanda ML, Lopez AM, Isuzu L, Lopez P (2014) Microsatellite isolation and characterization for *Colletotrichum* spp. causal agent of anthracnose in Andean black berry. *Genet Mol Res* 13:7673–7685. <https://doi.org/10.4238/2014.September.26.5>
- Moncada P, Couch M (2004) Simple sequence repeats diversity in diploid and tetraploid *Coffea* species. *Genome* 47:501–509. <https://doi.org/10.1139/g03-129>
- Motta LB, Soares TCB, Ferrao MAG, Caixeta ET, Lorenzoni RM, Souza Neto JDD (2014) Molecular characterization of arabica and Conilon coffee plants genotypes by SSR and ISSR markers. *Braz Arch Biol Technol* 57(5):728–735. <https://doi.org/10.1590/S1516-8913201402071>
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/>
- Pestana KN, Capucho AS, Caixeta ET, de Almeida DP, Zambolim EM, Cruz CD, Zambolim L, Pereira AA, de Oliveira ACB, Sakiyama NS (2015) Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443–03. *Tree Genet Genomes* 11(4):72. <https://doi.org/10.1007/s11295-015-0903-9>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Rovelli P, Mettullio R, Anthony F, Anzueto F, Lashermes P, Graziosi G (2000) Microsatellites in *Coffea arabica* L. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) *Coffee biotechnology and quality*. Springer, Dordrecht, pp 123–133
- Sant'Ana GC, Pereira LF, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, Pagiatto NF, da Silva BSR, Nogueira LM, Kitzberger CS, Scholz MB, de Oliveira FF, Sera GH, Padilha LL, Labouisse JP, Guyot R, Charmetant P, Leroy T (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Sci Rep* 8(1):465. <https://doi.org/10.1038/s41598-017-18800-1>
- Silvarolla MB, Mazzafera P, Fazuoli LC (2004) Plant biochemistry: a naturally decaffeinated Arabica coffee. *Nature* 429(6994):826. <https://doi.org/10.1038/429826a>
- Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution* 39(1):53–65. <https://doi.org/10.1111/j.1558-5646.1985.tb04079.x>
- Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS (2017) Molecular markers useful to discriminate *Coffea arabica* L cultivars with high genetic similarity. *Euphytica* 213(3):75. <https://doi.org/10.1007/s10681-017-1865-9>
- Steiger D, Nagai C, Moore P, Morden C, Osgood R, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor Appl Genet* 105(2–3):209–215. <https://doi.org/10.1007/s00122-002-0939-8>
- Tadele S, Mekbib F, Tesfaye K (2014) Genetic diversity of coffee (*Coffea arabica* L.) landraces from southern Ethiopia as revealed by inter simple sequence repeat marker. *Glob Adv Res J Agric Sci* 3(1):24–34
- Teressa A, Crouzillat D, Petiard V, Brouhan P (2010) Genetic diversity of Arabica coffee (*Coffea arabica* L) collections. *EJAST* 1(1):63–79
- Tesfaye K (2006) Genetic diversity of wild *Coffea arabica* populations in Ethiopia as a contribution to conservation and use planning. Ecology and development series. University of Bonn, Germany
- Tesfaye K, Govers K, Bekele E, Borsch T (2014) ISSR fingerprinting of *Coffea arabica* throughout Ethiopia reveals high variability in wild populations and distinguishes them from landraces. *Plant Syst Evol* 300(5):881–897. <https://doi.org/10.1007/s00606-013-0927-2>
- Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ (2016) Advances in genomics for the improvement of quality in coffee. *J Sci Food Agric* 96:3300–3312. <https://doi.org/10.1002/jsfa.7692>
- Vieira ESN, von Pinho EVR, Carvalho MG, Esselink GD, Vosman B (2010) Development of microsatellite markers for the identification of Brazilian *Coffea arabica* varieties. *Genet Mol Biol* 33(3):507–514. <https://doi.org/10.1590/S1415-47572010005000055>
- Waples RS (1987) A multi species approach to the analysis of gene flow in marine shore fishes. *Evolution* 41(2):385–400. <https://doi.org/10.1111/j.1558-5646.1987.tb05805.x>
- White Head and Peakall (2015) GENALEX 6.502: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295. <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
- Wright S (1951) The genetic structure of populations. *Annu Eugen* 15:323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- Yeh FC, Yang R (1999) POPGENE VERSION 1.31: Microsoft Window-based Freeware for Population Genetic Analysis. Department of Renewable Resources University of Alberta, Canada. pp 1–29

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.