

Research Article

Genetic Diversity and Population Structure of South Ethiopian Arabica Coffee [*Coffea arabica* L.] Genotypes Using ISSR Markers

Habtamu Gebreselassie^{1,2,*} , Bizuayehu Tesfaye² , Andargachewu Gedebo², Yayis Rezene³

¹Crop Research Directorate, Wondo Genet Agricultural Research Center, Wondogenet, Ethiopia

²School of Plant and Horticultural Science, Hawassa University, Hawassa, Ethiopia

³Molecular Biology Research, Hawassa Agricultural Research Center, Hawassa, Ethiopia

Abstract

Arabica coffee originated and diversified in Ethiopia, yet its considerable genetic diversity remains underutilized. This study assessed the genetic diversity and population structure of 50 Arabica coffee genotypes representing five populations (Sidama, Amaro, Jinka, Guji, and improved varieties) using inter-simple sequence repeat (ISSR) markers. The populations produced 74 distinct bands, with improved varieties showing the highest number of private bands (8) and lowest common bands ($\leq 50\%$) at 6. Band frequency ranged from 8.62% (Guji) to 25.86% (improved varieties), averaging 17.93%. Genetic diversity parameters, including number of alleles per population, effective alleles, Shannon's information index, observed diversity, and unbiased diversity, ranged from 0.276-0.672, 1.063-1.149, 0.052-0.12, 0.036-0.082, and 0.039-0.092, respectively. AMOVA revealed significant genetic variability, with 67% among populations and 33% within. Principal coordinate analysis explained 42.96% of total variation across three axes. UPGMA cluster analysis grouped the genotypes into four clusters (I-IV) containing 20%, 28%, 12%, and 40% of the genotypes, respectively, with genotypes from the same populations clustering together. Overall, the study demonstrated substantial genetic variation and population structure among South Ethiopian Arabica coffee genotypes, highlighting the potential for conservation and breeding efforts. Future studies should incorporate high-resolution markers and broader accession sets to better capture the genetic landscape of Ethiopian Arabica coffee.

Keywords

Arabica Coffee, Genetic Polymorphism, Population Structure, ISSR Markers

1. Introduction

Coffea arabica is the only allotetraploid ($2n=4x=44$) and self-fertile species in the *Coffea* genus, whereas the rest species are diploid, highly self-incompatible and allogamous (out

crossing [26, 28]. It is evergreen plant, which grows within the belt lies equidistant from the equator (23.5°N) in the north and (23.5°S) in the south, typically in tropical regions of

*Corresponding author: habteline@gmail.com (Habtamu Gebreselassie)

Received: 15 September 2025; **Accepted:** 29 September 2025; **Published:** 30 October 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Africa, Asia, Oceania, South America (Brazil and Colombia), Central America and Mexico [23, 24]. It is the most important agricultural commodity upon which the economy of more than 80 producer countries depend on [6] and a foreign currency earner. It plays a key role in creating direct employment for 125 million people in worldwide [14]. It is a vital and delightful beverage that is frequently drunk due to its stimulatory effects, desirable aroma, and flavor. Its beverage contains antioxidant phenols, which may provide health benefits such as a lower risk of cancer, cardiovascular disease, diabetes, and Parkinson's disease [11, 19, 24, 40].

Coffee Arabica has enormous economic, tourism, social, and cultural value in Ethiopia [17, 22]. From smallholder coffee growers to coffee exporters, the coffee industry provides employment and boost economic growth [37]. Coffee tourism allows the tourists to explore the birthplace of coffee, strongly connect with the country's rich cultural heritage, witness vibrant and ritualistic traditional coffee ceremonies, and drink the flavorsome brews captured the world's attention [59, 60]. The Ethiopian coffee ceremonies bring people together; promote unity, and create a sense of togetherness among individuals [17].

Ethiopia is the original homeland and center of genetic diversity of coffee Arabica [12, 27, 57]. Many different types of coffee trees can be found within small areas [62]. Even though Arabica coffee has much important and it has high genetic diversity, Ethiopian coffee improvement program is solely dependents on conventional breeding approach and needed to be supplemented with modern molecular breeding techniques [9, 13].

The genetic diversity of *Coffea arabica* arises from its origin as an allotetraploid hybrid between *C. canephora* and *C. eugenoides*, which introduced a novel genetic combination [28]. Although *C. arabica* is predominantly self-pollinating, significant variation persists, especially in Ethiopia, might be due to its outcrossing by insects and humanmade activities. This diversity is maintained by natural selection across diverse agro-ecological zones, traditional farming practices that promote seed propagation and genetic recombination, and long-term in situ conservation by smallholder farmers [25]. Human-mediated selection for yield, quality, and resilience has further shaped this diversity over generations [4, 5]. Despite the species' narrow genetic base compared to outcrossing crops, these combined factors contribute to the rich genetic variability observed in Arabica coffee.

Conventional coffee breeding in Ethiopia was started in 1952 with the major objectives of collection and conservation of coffee germplasm, development of cultivars that combine high yield, disease resistance, improved cup quality, and wider environmental adaptable [6]. Since then, a number of achievements reported in germplasm collection and conservation, development of high yielder and CBD resistant having

good cup quality selections and heterotic hybrids varieties. Despite its many merits, conventional coffee breeding in Ethiopia faced a number of challenges. Firstly, it takes long time (at least 21 years) to develop a new coffee cultivar [9]. Secondly, highly influenced by growing environmental factors [1, 7, 15, 32, 33] and growth stage of the plant [58]. Thirdly, the coffee Arabica genetic diversity is low due to its narrow genetic base associated with autogamy, evolutionary history, and domestication [5, 34, 48, 51, 52, 59]. Fourthly, phenotypically similar individuals may be genetically distinct, which reduces the selective efficiency.

Use of molecular markers associated with important traits can make coffee breeding more precise, rapid, allow the detection of variations in DNA level, overcome environmental factors and growth stage of the plant, and probably cost effective in comparison to phenotypic selection [21, 39, 53]. Some of the DNA based markers that have been extensively used in coffee genetic diversity studies includes Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Randomly Amplified Polymorphic DNAs (RADP), Sequence-related Amplified Polymorphism (SRAP) and Simple Sequence Repeats (SSR), single-nucleotide polymorphisms (SNPs) markers [10, 29, 35, 42, 47, 61]. Inter Simple Sequence Repeat (ISSR) markers are widely used in assessing genetic diversity in *Coffea arabica* due to their simplicity, cost-effectiveness, and ability to generate high polymorphism without prior genomic information [46].

Eventhough many studies were conducted on Arabica coffee genetic diversity, the coffee genetic study program, particularly the south Ethiopian coffee populations, has not benefited a lot from the development of the recent molecular markers. Because of this, little is known about the genetic structure and pattern of south Ethiopian coffee [13], which in turn has limited the use of its diverse gene pool in the improvement program [13]. Accordingly, the present study was conducted estimate the genetic diversity and population genetic structure of South Ethiopian coffee Arabica genotypes for successful conservation of genetic resources for future improvement using ISSR markers.

2. Materials and Methods

2.1. Plant Material

The study employed fifty Arabica coffee genotypes that were collected from four growing zones (Sidama, Gujji, Amaro, Jinka) and improved varieties (Table 1). The coffee trees of the genotypes were *ex-situ* conserved at the field gen bank of Awada Agricultural research sub-center, Ethiopia.

Table 1. Description of Arabica coffee germplasm used for molecular genetic diversity study.

Acc.	Population	Region	Loction	Specific location	Altitude	Mainainer
Ak-4	Gujji	Oromiya	Dole	Mendere	1746	AARSC/EIAR
Ak-5	Gujji	Oromiya	Dole	Mendere	1741	AARSC/EIAR
Ak-6	Gujji	Oromiya	Dole	Akoku	1755	AARSC/EIAR
Ak-10	Gujji	Oromiya	Obadola	kumina	1698	AARSC/EIAR
Ak-12	Gujji	Oromiya	Obadola	kumina	1720	AARSC/EIAR
Ak-13	Gujji	Oromiya	Obadola	kumina	1709	AARSC/EIAR
Ak-23	Gujji	Oromiya	Chembe	keca eba	1655	AARSC/EIAR
Ak-25	Gujji	Oromiya	Dida guda	rohe	1519	AARSC/EIAR
Ak-27	Gujji	Oromiya	Chembe	keca eba	1563	AARSC/EIAR
Ak-23	Gujji	Oromiya	Qebesa	Sageji Ela Bedessa	1962	AARSC/EIAR
Ak-9	Amaro	South Ethiopia	Golbe	Golbe	1880	AARSC/EIAR
Ak-10	Amaro	South Ethiopia	Golbe	Golbe	1880	AARSC/EIAR
Ak-16	Amaro	South Ethiopia	Kerma	Dogodo-2	1660	AARSC/EIAR
Ak-23	Amaro	South Ethiopia	Danobulto	Shashe	1500	AARSC/EIAR
Ak-37	Amaro	South Ethiopia	Kele town	Kele-02	1600	AARSC/EIAR
Ak-38	Amaro	South Ethiopia	Kele town	Kele	1600	AARSC/EIAR
Ak-44	Amaro	South Ethiopia	Sharo	Angushi	1600	AARSC/EIAR
Ak 47	Amaro	South Ethiopia	Darba	mane na	1700	AARSC/EIAR
Ak 52	Amaro	South Ethiopia	Tifata	Tsilalo omo	1680	AARSC/EIAR
Ak 57	Amaro	South Ethiopia	Tifata	Abetu kotsare	1700	AARSC/EIAR
Ak-1	Gamogofa	South Ethiopia	Genamer	Woyiki	1640	AARSC/EIAR
AK-5	Gamogofa	South Ethiopia	Genamer	Genamer	1600	AARSC/EIAR
AK-8	Gamogofa	South Ethiopia	Tolta	Tolta	1630	AARSC/EIAR
AK-14	Gamogofa	South Ethiopia	Metser	Zelet	1643	AARSC/EIAR
AK-42	Gamogofa	South Ethiopia	Alga	Wozader	1500	AARSC/EIAR
AK-43	Gamogofa	South Ethiopia	Koibe	Alkaido	1390	AARSC/EIAR
AK-58	Gamogofa	South Ethiopia	Banata	Meso	1400	AARSC/EIAR
AK-59	Gamogofa	South Ethiopia	Aykamer	Beso	1400	AARSC/EIAR
AK-61	Gamogofa	South Ethiopia	Aykamer	Beso	1400	AARSC/EIAR
AK-63	Gamogofa	South Ethiopia	Aykamer	Beso	1450	TARC/EIAR
Aw-31	Sidama	Sidama	Dara	Aleme Kancha	1800	AARSC/EIAR
Aw-79	Sidama	Sidama	Bensa	Hedamo	1800	TARC/EIAR
AK-77	Sidama	Sidama	Dale	Loya	1750	ARC/EIAR
AK-43	Sidama	Sidama	Chicho	Bedecha	1830	AARSC/EIAR
AK-53	Sidama	Sidama	Debeka	Gerari	1770	MARSC
AK-14	Sidama	Sidama	Weto	Metero	1800	JARC/EIAR
AK-91	Sidama	Sidama	Mamena	Gona	1760	JARC/EIAR

Acc.	Population	Region	Loction	Specific location	Altitude	Mainainer
AK-121	Sidama	Sidama	Bedelicho	Bedelicha	1980	JARC/EIAR
AK-123	Sidama	Sidama	Bedelicho	Bedelicha	1800	JARC/EIAR
AK-122	Sidama	Sidama	Foka	Kisho	1770	AARSC/EIAR
Koti	Improved Variety					AARSC/EIAR
74112	Improved Variety					JARC/EIAR
Angafa	Improved Variety					AARSC/EIAR
CH-1	Improved Variety					AARSC/EIAR
Rori	Improved Variety					AARSC/EIAR
Mechara-1	Improved Variety					MARSC
J-19	Improved Variety					TARC/EIAR
J-21	Improved Variety					TARC/EIAR
74140	Improved Variety					AARSC/EIAR
74110	Improved Variety					JARC/EIAR
Harussa	Improved Variety					JARC/EIAR
Dessu	Improved Variety					JARC/EIAR
Gesha	Improved Variety					JARC/EIAR
Mocha	Improved Variety					MARSC
75227	Improved Variety					JARC/EIAR
Odicha	Improved Variety					AARSC/EIAR

JARC =Jimma Agricultural Research Center; EIAR=Ethiopian Institute of Agricultural Research; AARSC=Awada Agricultural Research Center; TARC= Tepi Agricultural Research Center; MARSC= Mechara Agricultural Research Center.

2.2. Sample Collection

Young and healthy leaf samples were collected from growing tips of a single tree of each genotype using Eppendorf tube. Four discs leaf samples were directly placed in to an Eppendorf tube, and transported to a molecular laboratory of South Agricultural Research Institute, Hawassa, Ethiopia and stored under refrigerator until DNA extraction was done.

2.3. DNA Extraction

The genomic DNA extraction of coffee samples was performed using a modified cetyltrimethylammonium bromide (CTAB) protocol [49]. The leaf samples were crushed using a plastic micro pestle and 400 ul of CTAB buffer added to the Eppendorf containing the crushed samples and vortex for 10 seconds before incubating under a water bath at 60 °C for 30 minutes. The solution was incubated in a water bath at 60 °C for 30 minutes. During the incubation process the solution was homogenized every 10 minutes. Following incubation, 60 ul of chloroform-isoamyl alcohol was added to the mixture and centrifuged at 10,000 rpm for 5 minutes. The supernatant

was transferred to a new Eppendorf tube and the chloroform isoamyl alcohol step was repeated and centrifuged again at 10,000 rpm for 5 minutes. Stages using chloroform isoamyl alcohol were performed three times to separate DNA from proteins, polysaccharides, and other impurity compounds.

The supernatant was transferred to a new 1.5 mL tube and cold isopropanol was added to 2/3 of the supernatant. The solution was homogenized and incubated at -25 °C for 30 minutes. The solution was then centrifuged at 10,000 rpm for 5 minutes to form a pellet and supernatant. The pellet containing DNA was washed three times using 70% ethanol and centrifuged at 10,000 rpm for 5 minutes. The DNA pellet was re-suspended with TE-RNase and stored at -25 °C. The DNA quantity and purity was measured by both a nano spectrophotometer (IMPLENNP80) and gel electrophoresis. The concentration (ng/μL) and purity of DNA was observed by nanophotometer at A260/A280. The quality of extracted DNA was observed on a 1% agarose gel. DNA was visualized under UV light using gel documentation (UVITEC).

2.4. PCR Amplification with ISSR Primers

The polymerase chain reactions (PCR) were performed to

investigate the suitability of the extracted DNA for amplification. The PCR reaction was carried out using 1 µg of primer, 1.5 µg of template DNA, and 10.5 µl of H₂O (grade), for a total volume of 13 µl. The amplifications were carried out for 35 cycles of 96 °C for 3 minutes, 96 °C for 30 seconds, 49 °C for 30 seconds, 72 °C for 30 seconds, followed by a final extension step of 72 °C for 5 min [50]. The amplified DNA

products were loaded in agarose gel. Gel electrophoresis was carried out on a 1.2% agarose gel in 100 ml of 1 × TAE buffer at 80 volts for 90 minutes. Amplified products were visualized using gel documentation system. A total of 9 universal ISSR primers were used to determine the molecular diversity of the genotypes under examination; of which 4 primers showed high percentage of polymorphism (Table 2).

Table 2. List of ISSR primers used in this study.

Sample No	ISSR primers name	Primer Sequences	Length of (bp)	Annealing T (°C)
1	UBC825	(AC) 8T	17	52
2	UBC826	(AC) 8C	17	47
3	UBC842	(GA) 8CG	18	52
4	UBC854	(TC) 8RG	18	49
5	UBC835	(AG) 8 YC	18	48
6	UBC868	(GAA) 6	18	55
7	UBC879	(CTTCA) 3	15	51
8	UBC888	BDB (CA) 7	17	52
9	UBC889	DBD (AC) 7	17	48

The reproducible profiles with template DNA's are detailed below; B = (C, G, T), i.e., not A; D = (A, G, T), i.e., not C; R = (A, G), Y = (C, T); UBC = University of British Columbia, Canada

2.5. Agarose Gel Analysis and ISSR Data Scoring

The integrity of DNA was judged by agarose gel analysis using the following steps: 0.8% agarose gel was prepared in 100 ml 1X TBE (Tris-borate EDTA) buffer containing 2.5 µl of ethidium bromide. The gel was cast in the tray and allowed to solidify until the wells were properly formed. 20 µl of DNA samples along with 5 µl a DNA ladder were individually loaded into each well after mixing them with a drop of gel tracking dye. The gel was run at 80 V for 90 minutes. After the gel was run to about 3/4th of the tray, the gel was removed from the cast and visualized under UV light. The presence of a single compact band at the corresponding band of the DNA ladder indicates that the DNA isolated is of high molecular weight.

2.6. Data Analysis

Statistical Analysis

Among the tested nine ISSR markers data only four polymorphic molecular markers were considered and fragments were scored based on a binary scoring matrix for the presence (1) or absence (0) visually based on the presence and absence of bands. Number of alleles (Na), number of effective allele

(Ne), observed (Ho) and expected (He) heterozygosity, Shannon's information index (I), Diversity (h), Unbiased Diversity (uh), Band Frequency (p), No. Bands, No. Bands Freq. >= 5%, No. Private Bands, No. LComm Bands (<=25%), and No. LComm Bands (<=50%) as well as analysis of molecular variance (AMOVA) were computed with GenAlExver.6.503 software [44].

Cluster analysis was carried out using DARwin software version 6.0.021 [45]. A dendrogram was generated based on the dissimilarity matrix as input data to visualize pattern of cluster within and among the genotypes. Principal coordinate analysis (PCoA) was computed by GenAlex version 6.503 software [45].

3. Results and Discussions

3.1. ISSR Analysis

Among the nine ISSR primers included in this study, the four universal primers, UBC 826, UBC 835, UBC 842, and UBC 889, provided a clearly amplified product and conspicuous banding pattern whereas the remaining primers were considered unsuitable due to poor amplification. The four ISSR markers are efficient in estimating the genetic diversity of coffee Arabica genotypes. The ISSR markers have been

also applied to characterize coffee germplasm in Ethiopia [2, 8, 55, 56].

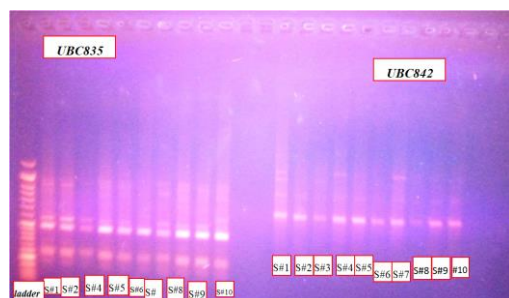


Figure 1. Sample picture of banding pattern of UBC835 and UBC842 primers on improved varieties.

3.2. Polymorphism of ISSR and Genetic Diversity

The 4 ISSR markers used in this study detected 74 different bands with a frequency $\geq 5\%$ for Gujji, Amaro, Jinka, Improved variety, and Sidama collections (Figure 2). The number of private bands (unique alleles) was higher for improved variety and Sidama collection 8 and 4 respectively. Private bands (unique alleles) are genetic variants found exclusively within a single population, not present in other populations being analyzed [54]. It can be crucial for a species' long-term survival and adaptation to changing environments. They might carry genes that confer resistance to new diseases, tolerance to drought, or other beneficial traits that are not found in the more common alleles [20, 30]. A high number of private bands among improved variety and Sidama

collection suggests that the presence of many rare and unique alleles within the improved variety and Sidama coffee population being studied. Accordingly, improved variety and Sidama coffee populations with a high number of private alleles are considered particularly important for conservation efforts.

Improved variety and Sidama collection were found to show highest value of the lowest number of common bands ($\leq 50\%$), 6 and 5 respectively, whereas the other populations showed the lowest value (1) (Figure 2). Common bands refer to alleles that are present at a high frequency across the entire sample set of populations being analyzed. They are alleles that are widespread within the population. The highest value of lowest number of common bands suggests a higher overall genetic diversity, presence of many rare alleles, reduced impact of selection bottlenecks, and heterogeneous population structure [24, 46]. Accordingly, improved variety and Sidama collection having highest value of lowest number of common bands points to a more diverse and less homogenous coffee population, with a greater presence of rare or less frequent alleles.

Band frequency varied from 8.62% (Gujji) to 25.86% (Improved varieties) with average of 17.93% (Table 3). The observed high band frequency for improved varieties indicates the presence of higher genetic diversity and a wider range of allele combinations among improved varieties than the other populations. The detected polymorphic band in this investigation is relatively low as compared to [3, 36] report; where as high as compared to Gichuru (2012) and Panaligan *et al.*, (2020) [18, 43]. Even though the genetic variability of Arabica coffee has been low due to the narrow genetic basis and self-fertile nature of the species, the existed genetic variation among the studied population is a promising result for coffee improvement program through selection of parental lines and hybridization.

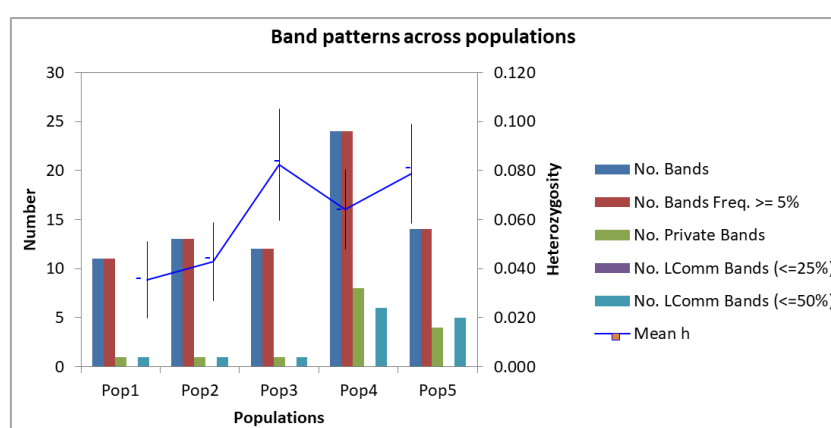


Figure 2. Band patterns across populations.

Where; No. Bands = No. of Different Bands; No. Bands Freq. >= 5% = No. of Different Bands with a Frequency >= 5%; No. Private Bands = No. of Bands Unique to a Single Population; No. LComm Bands (<=25%) = No. of Locally Common Bands (Freq. >= 5%) Found in 25% or Fewer Populations; No. LComm Bands (<=50%) = No. of Locally Common Bands (Freq. >= 5%) Found in 50% or Fewer Populations; h = Diversity = $1 - (p^2 + q^2)$.

The number of alleles per populations (N_a) varied from 0.276 (Guji) to 0.672 (improved varieties) with an average of

0.434 (Table 3). Number of effective alleles (N_e) varied from 1.063 (Gujji) to 1.149 (Jinka). Shannon's information index (I) varied from 0.052 (Gujji) to 0.12 (Sidama). Likewise, observed diversity (h) varied among populations from 0.036 (Gujji) to 0.082 (Jinka) with an average value of 0.061. On the

other side, unbiased diversity (uh) ranged from 0.039 (Gujji) to 0.092 (Jinka) (Table 3). Based on N_e , I , h , and uh values, Jinka and Sidama population showed relatively high level of genetic diversity, whereas the results for the genetic diversity in Gujji and Amaro were convergent.

Table 3. Genetic parameter estimates based on ISSR Marker among collection regions.

Population	No. of genotypes	N_a	N_e	I	h	uh	%P
Gujji	10	0.276	1.063	0.052	0.036	0.039	8.62%
Amaro	10	0.345	1.07	0.065	0.043	0.048	12.07%
Jinka	10	0.397	1.149	0.118	0.082	0.092	18.97%
Improved variety	10	0.672	1.095	0.105	0.064	0.071	25.86%
Sidama	10	0.483	1.129	0.12	0.079	0.087	24.14%
Grand Mean	10	0.434	1.101	0.092	0.061	0.067	17.93%
Sum	50	2.173	5.506	0.46	0.304	0.337	89.66%

Where:- N_a = number of allele; N_e = number of effective allele; I = Shannon's information index; h = Diversity = $1 - (p^2 + q^2)$; uh = Unbiased Diversity = $(N / (N-1)) * h$; p = Percent of polymorphic loci.

3.3. Analysis of Molecular Variance

Analysis of molecular variance (AMOVA) partitioned the total genetic variation into among and within populations (Table 4). Of the total genetic variation, 67% of variation was attributed to the variability among populations and 33% accounted for variability within populations, indicating that the

genetic variation among populations contributed more to genetic diversity (Table 4). The result is in agreement with previous works by Aga, [2] who reported that most of the variability was observed between populations than within populations. However, other scholars reported the higher proportion of genetic diversity was observed within the population rather than between populations [8, 16, 31, 38, 41, 56].

Table 4. Analysis of molecular variance between population of Arabica coffee genotypes.

Sources of variation	Df	SS	MS	Estimated variance	Percentage of variation
Among populations	4	165.240	41.310	3.935**	67%
Within populations	45	88.000	1.956	1.956**	33%
Total	49	253.240		5.891	100%

df=degree of freedom; SS=sum of square; MS=mean square

3.4. Principal Component Analysis

The principal coordinate analysis (PCoA) generated by using genetic distance showed the presence of weak population structure and explained 42.96% of the cumulative variation (Figure 3). The first, second and third axis explained 19.45%, 14.73%, and 8.78% of the total genetic variation, respectively. As it is observed in two-dimensional plot of

density versus discriminant function, genotypes collected from the same population often grouped together, and genotypes from different population differently (Figure 4). In both analyses, there was no definite clustering of populations, and showed consistent pattern of genetic relationship and differentiation among the accessions. This suggests gene flow or shared ancestry among populations, consistent with a weakly structured but genetically diverse population.

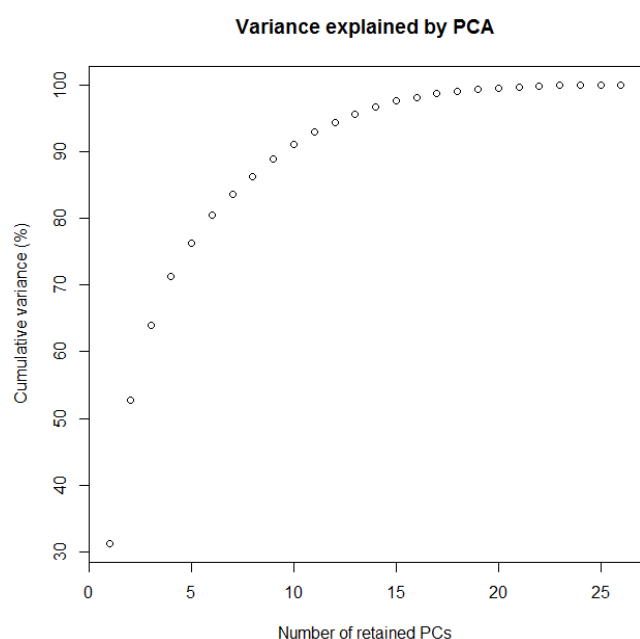


Figure 3. Variance explained by principal component analysis.

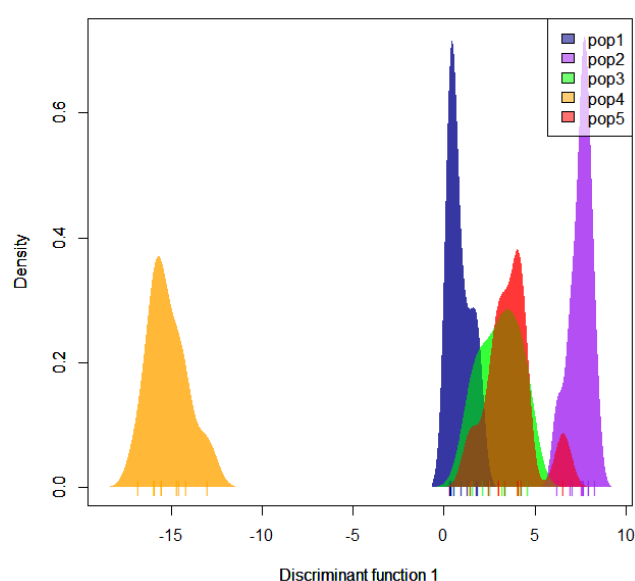


Figure 4. Density vs discriminant function bi-plot.

To infer in the genetic structure of the populations, a discriminant analysis of principal component (DAPC) analysis was conducted using the 40 accessions and 10 improved varieties. The DAPC analysis revealed five genetic clusters and four discriminant eigenvalues (Figure 5). Each of these clusters corresponded with genotypes derived from population collected from each collection regions. The DAPC identified five genetic clusters, with four discriminant eigenvalues capturing between-group variation. These clusters generally aligned with regional origins, providing further insight into the underlying genetic structure shaped by geographic and environmental factors.

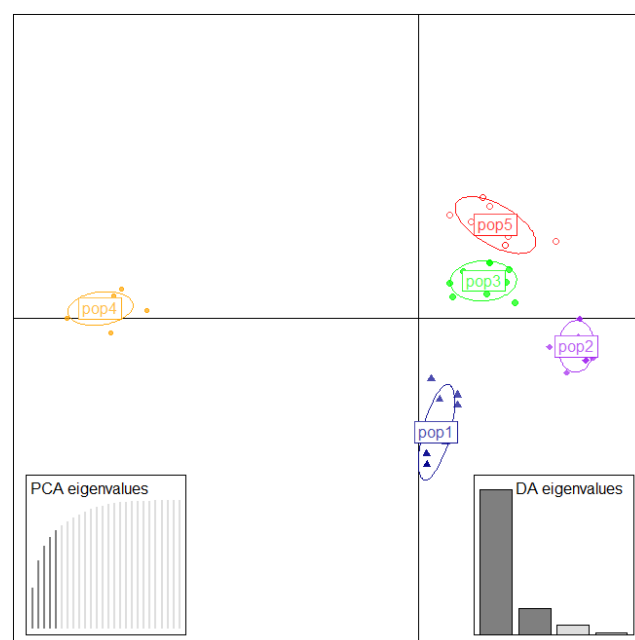


Figure 5. Discriminant analysis of principal components (DAPC) for the 50 arabica coffee genotypes. Dots represent individuals whereas colours denoting sampling origin.

3.5. Cluster Analysis

The cluster analysis based UPGMA method grouped the 40 Arabica coffee accessions and 10 improved varieties into four major clusters, indicating the presence of significant genetic diversity among them (Figure 6). Cluster I consisted 10 genotypes (20%), suggesting a moderately distinct group with shared genetic characteristics. Cluster II comprised 14 genotypes (28%), representing the second-largest cluster, possibly reflecting a group with closer genetic similarity. Cluster III was the smallest with 6 genotypes (12%), implying a more genetically distinct or less common group of genotypes that might possess unique traits. Cluster IV, the largest group with 20 genotypes (40%), likely includes more genetically related genotypes. The formation of these distinct clusters and sub-groups suggests a wide range of genetic variation within the Arabica coffee genotypes studied.

The clustering of accessions and improved varieties into four distinct groups has important implications for breeding. It highlights the presence of significant genetic diversity, which can be exploited to select genetically distant parents for crossing, thereby enhancing hybrid vigor and trait improvement. The smaller and more distinct Cluster III may harbor unique alleles valuable for introducing novel traits. The grouping also suggests potential adaptation to specific environments, guiding breeders in developing varieties suited to diverse agro-ecologies. Moreover, understanding the genetic structure helps in conserving valuable germplasm and avoiding redundancy in breeding programs, ultimately supporting the development of improved, resilient, and high-performing coffee cultivars.

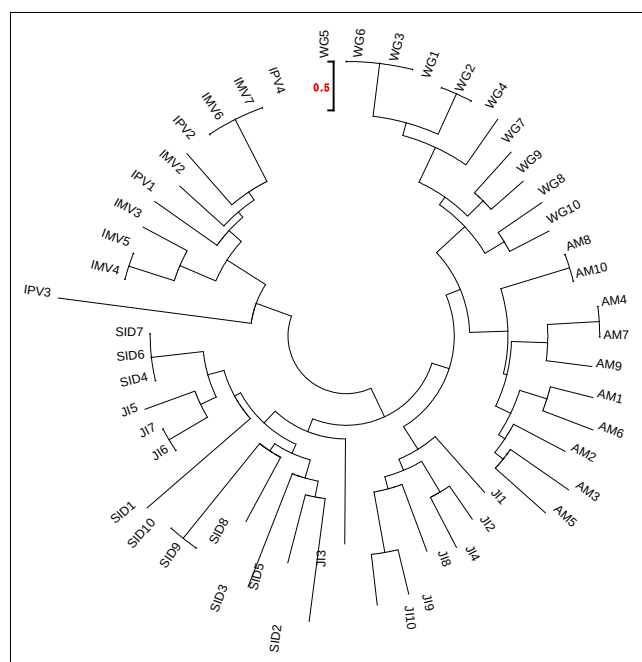


Figure 6. UPGMA dendrogram representing the genetic relationships among the indicated Arabica coffee genotypes.

4. Summery

Genetic diversity analysis of 50 arabica coffee genotypes using 4 SSR markers showed the presence of genetic vdiversity among genotypes. The different diversity indices obtained in this study showed that the ISSR markers used were efficient and informative for Ethiopian Arabica coffee diversity study. High level of genetic diversity among population of Jinka and Sidama was identified; suggesting that the germplasm has enormous opportunity in the improvement program through direct selection. The analysis of molecular variance (AMOVA) revealed the presence of significant genetic variability; with diversities mainly distributed among populations. The generated UPGMA dendrogram along with PCoA confirmed the result from AMOVA. The findings indicate a moderate level of genetic diversity within and among populations, highlighting the necessity for conservation measures to preserve unique genotypes that are crucial for sustainable coffee production. In these population clustering analyses, genotypes from the same populations were found to be clustered similarly and vise-versa.

5. Recommendation

Although ISSR markers utilized and proved useful genetic diversity information in this study, it is limited due to its dominant, lower reproducibility, and difficulty in pinpointing specific genomic locations. Therefore, it is strongly recommended that future studies adopt co-dominant marker systems such as SSR (Simple Sequence Repeats) or SNP (Single Nucleotide Polymorphisms), which are a co-dominant, highly

reproducible, and can provide precise information about gene locations, enabling more effective association mapping and gene discovery for south Ethiopian coffee yield and quality traits in future research.

While this study provided useful information on genetic diversity, it lacks the resolution necessary to pinpoint specific genes influencing important agronomic traits to facilitate a more direct and efficient link between phenotypes and genotypes. Therefore, I strongly recommend a transition to association studies breeding strategies to link phenotypes with genotypes for yield and quality improvement of coffee in south Ethiopia. The association study will facilitate the identification of candidate genes underlying important agronomic traits.

Acknowledgments

The author wishes to express sincere gratitude to the Ethiopian Institute of Agriculture, Awada Sub-Center, for providing the genetic materials used in this study. Special thanks are also extended to the Molecular Biology Laboratory technical staff at Hawassa Agricultural Research Center for their valuable technical assistance and cooperation.

Author Contributions

Habtmu Gebreselassie: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft

Bizuayehu Tesfaye: Methodology, Supervision, Writing – review & editing

Andargachewu Gedebo: Project administration, Resources, Supervision, Writing – review & editing

Funding

This research work was financed by Ethiopian Institute of Agricultural Research and the institutional collaboration program between Hawassa University (Ethiopia) and the Norwegian University of Life Science.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Adem, A., Mohammed, H., & Ayana, A. (2020). Phenotypic Diversity in Arabica Coffee Genotypes from Eastern Ethiopia. 5(4), 42-47. <https://doi.org/10.11648/j.ijee.20200504.11>
- [2] Aga (2005). Molecular genetic diversity study of forest coffee tree (*Coffea arabica* L.) populations in Ethiopia: implications for conservation and breeding (Vol. 2005, No.79).

- [3] Aga E, Bryngelsson T, Bekele E, Salomon B (2003). Genetic diversity of forest arabica coffee (*Coffea arabica* L.) in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis. *Hereditas*. 138: 36-46.
- [4] Aga, E., Bekele, E., & Bryngelsson, T. (2005). Genetic diversity of forest Arabica coffee (*Coffea arabica* L.) populations in Ethiopia as revealed by RAPD analysis. *Hereditas*, 142(1), 47-52.
- [5] Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104(5), 894-900. <https://doi.org/10.1007/s00122-001-0798-8>
- [6] Bayetta Belachew, (2015). ACRN activity report for the period 2013 - 2015. IACO Abidjan, 2015.
- [7] Beksisa, L. (2021). GGE Biplot Analysis of Genotype x Environment Interaction and Bean Yield Stability of Arabica Coffee (*Coffea Arabica* L.) Genotypes in Southwestern Ethiopia. 9(3), 110-115. <https://doi.org/10.11648/j.ajbio.20210903.16>
- [8] Belami, S. (2007) Genetic diversity analysis of the wild *Coffea arabica* L. populations from Harennaforest, Bale Mountains of Ethiopia, using inter simple sequence repeats (ISSR) marker, AAU.
- [9] Benti, T. (2017). Progress in Arabica Coffee Breeding in Ethiopia: Achievements, Challenges and Prospects. *International Journal of Sciences: Basic and Applied Research*, 33(2), 15-25. <http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>
- [10] Benti, T., Gebre, E., Tesfaye, K., Berecha, G., Kyallo, M., & Yao, N. K. (2021). Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. *Journal of Crop Improvement*, 35(2), 147-168. <https://doi.org/10.1080/15427528.2020.1803169>
- [11] Böhn, S. K., N. C. Ward, J. M. Hodgson, and K. D. Croft. 2012. Effects of tea and coffee on cardiovascular disease risk. *Food & function* 3(6): 575-591.
- [12] Cristancho, MA, Chaparro, AP, Cortina, HA and Gaità, AL. 2004a. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. ASIC 20th. Bangalore, India.
- [13] EIAR (Ethiopian Institute of Agricultural Research), 2017. National Coffee Commodity Research Strategy of Fifteen Years (2016 -2030). Addis Ababa, Ethiopia. 124: 213-221.
- [14] Fairtrade Foundation, (2022). Coffee farmers. <https://www.fairtrade.org.uk/farmers-and-workers/coffee/>
- [15] Gebreselassie, H., Tesfaye, B., Gedebo, A. (2024): Genetic diversity of Arabica coffee genotypes in south Ethiopia using quantitative agro-morphological traits. *Genet Resour Crop Evol* 71, 3485-3506. <https://doi.org/10.1007/s10722-023-01853-9>
- [16] Geleta M., Isabel H., Arnulfo M. and Tomas B., (2012). Genetic diversity of arabica coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. *The Scientific World Journal*.
- [17] Geremew, M., Satheesh, N. and Fanta, S. W. (2022). Role of coffee in ethiopian ethnic culture-a coffee festival 2nd international conference on coffee and cocoa. Available at: <https://www.researchgate.net/publication/360385395>
- [18] Gichuru, E. K. (2012). Genetic diversity among commercial coffee varieties, advanced selections and museum collections in Kenya using molecular markers. 4(February), 39-46. <https://doi.org/10.5897/IJBC11.231>
- [19] Gökçen, B. B., & Şanlıer, N. (2019). Coffee consumption and disease correlations. *Critical Reviews in Food Science and Nutrition*, 59(2), 336-348. <https://doi.org/10.1080/10408398.2017.1369391>
- [20] 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. *Genet Res Int*. <https://doi.org/10.1155/2015/431487>
- [21] Henry R. J., 1997. Molecular markers in plant improvement. In: *Practical application of plant molecular biology*. Chapman and Hall, New York, pp 101-134.
- [22] Higdon, J. V., and B. Frei. 2006. Coffee and health: a review of recent human research. *Critical reviews in food science and nutrition* 46(2): 101-123.
- [23] ICO (2016). Historical Data on the Global Coffee Trade; <http://www.ico.org/new-historical.asp>, accessed on 20 July 2022.
- [24] Kanaka, K. K., Sukhija, N., Goli, R. C., Singh, S., Ganguly, I., Dixit, S. P., Dash, A., & Malik, A. A. (2023). On the concepts and measures of diversity in the genomics era. *Current Plant Biology*, 33(March), 100278. <https://doi.org/10.1016/j.cpb.2023.100278>
- [25] Labouisse, J. P., Bellachew, B., Kotecha, S., & Bertrand, B. (2008). Current status of coffee (*Coffea arabica* L.) genetic resources in Ethiopia: implications for conservation. *Genetic Resources and Crop Evolution*, 55, 1079-1093. <https://doi.org/10.1007/s10722-008-9321-7>
- [26] Lachenmeier, D. W. (2023). Identification of Coffee Species, Varieties, Origins, and Processing and Preparation Methods—A Status Report. 9. <https://doi.org/10.3390/icc2023-14824>
- [27] Lashermes P, Trouslot P., Anthony C. and Charrier A., 1996. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* 87, 59-64.
- [28] Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., & Charrier, A. (1999). Molecular characterization and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics*, 261(2), 259-266. <https://doi.org/10.1007/s004380050956>

- [29] Maluf M. P., Silvestrini M., Ruggiero L. C. M., Guerreiro Filho O., Colombo C. A., 2005. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Sci. Agric.* 62: 366-373.
- [30] Maruki T, Ye Z, Lynch M., (2022). Evolutionary Genomics of a Subdivided Species. *Mol Biol Evol.* 3; 39(8): msac152. <https://doi.org/10.1093/molbev/msac152>
- [31] Masumbuko L. I and Bryngelsson T., 2006. Inter simple sequence repeat (ISSR) analysis of diploid coffee species and cultivated *Coffea arabica* L. from Tanzania. *Gen. Res. Crop Evol.*, 53: 357-366.
- [32] Merga, D., Mohammed, H., & Ayano, A. (2021). Estimation of Genetic Variability, Heritability and Genetic Advance of Some Wollega Coffee (*Coffea arabica* L.) Landrace in Western Ethiopia Using Quantitative Traits. 9(4), 182-191. <https://doi.org/10.11648/j.jps.20210904.18>
- [33] Merga, W. (2022). Bean yield stability analysis of coffee (*Coffea arabica* L) genotypes in south western Ethiopia. *International Journal of Recent Scientific Research*, 13(ISSN: 0976-3031), 1-8. <https://doi.org/10.24327/IJRSR>
- [34] Meyer R. S., Purugganan M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat Rev Genet.* 14(12): 840-52. <https://doi.org/10.1038/nrg3605>.
- [35] Mishra M. K., Tornincasa P., De Nardi B., Asquini E., Dreos R., Del Terra L., Rathinavelu R., Rovelli P., Pallavicini A. and Graziosi G., (2011). Genome organization in coffee as revealed by EST PCR-RFLP, SNP and SSR analysis. *J. Crop Sci. Biotech.* 14: 17-24.
- [36] Mishra, M. K., Nishani, S., Gowda, M., Padmajyothi, D., Suresh, N., Sreenath, H., & Raghuramulu, Y. (2018). Genetic Diversity Among Ethiopian Coffee (*Coffea Arabica* L.) Collections Available In Indian Gene Bank Using Sequence Related Amplified Polymorphism Markers. *Plant Breeding and Seed Science*, 70(1), 29-40. <https://doi.org/10.1515/plass-2015-0011>
- [37] Muhie, S. H. (2022). Strategies to improve the quantity and quality of export coffee in Ethiopia, a look at multiple opportunities. *Journal of Agriculture and Food Research*, 10(April), 100372. <https://doi.org/10.1016/j.jafr.2022.100372>
- [38] Mulatu Geleta, Isabel Herrera, Arnulfo Monz, and T. B. (2012). Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers. 2012. <https://doi.org/10.1100/2012/939820>
- [39] Ni S., Yaho M., Chen L., Zhao L. and Wang X. C., (2008). Germplasm and breeding research of tea plant based on DNA marker approaches. *Front. Agric. China.* 2, 200-2007.
- [40] Nkondjock, A. 2009. Coffee consumption and the risk of cancer: an overview. *Cancer letters* 277(2): 121-125.
- [41] Ogutu, C., Fang, T., Yan, L., Wang, L., Huang, L., Wang, X., ... Han, Y. (2016). Characterization and utilization of microsatellites in the *Coffea canephora* genome to assess genetic association between wild species in Kenya and cultivated coffee. *Tree Genetics & Genomes*, 12(3), 54.
- [42] Oktavioni, M., Mada, U. G., Irawati, I., Andalas, U., Syukriani, L., Andalas, U., Setiawan, M. A., & Andalas, U. (2020). Analysis of genetic diversity of Arabica coffee [*Coffea arabica* L.] in Solok Regency by SRAP molecular markers
- [43] Panaligan, A. C., Baltazar, M. D., & Alejandro, G. J. D. (2020). Short communication: Genetic polymorphism of registered and popularly cultivated coffee (*coffea* spp.) in the philippines using inter-simple sequence repeats markers. *Biodiversitas*, 21(9), 4228-4233. <https://doi.org/10.13057/biodiv/d210938>
- [44] Peakall, R. and P. E. Smouse. 2016. GENALEX6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology*. 6: 288-295.
- [45] Perrier, X. and J. P. Jacquemoud-Collet. 2006. DARwin software. <http://darwin.cirad.fr/Darwin>
- [46] Raychaudhuri S., (2011). Mapping rare and common causal alleles for complex human diseases. 147(1): 57-69. <https://doi.org/10.1016/j.cell.2011.09.011>
- [47] Reddy, M. P., Sarla, N., & Siddiq, E. A. (2002). Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica*, 128, 9-17. <https://doi.org/10.1023/A:1020691618797>
- [48] Scalabrin, S., Toniutti, L., Di Gaspero, G., Scaglione, D., Magris, G., Vidotto, M., Pinosio, S., Cattonaro, F., Magni, F., Jurman, I., Cerutti, M., Suggi Liverani, F., Navarini, L., Del Terra, L., Pellegrino, G., Ruosi, M. R., Vitulo, N., Valle, G., Pallavicini, ... A., Bertrand, B. (2020). A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Scientific Reports*, 10(1), 1-13. <https://doi.org/10.1038/s41598-020-61216-7>
- [49] Schenk, J. J., Becklund, L. E., Carey, S. J., & Fabre, P. P. (2023). What is the “modified” CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Sciences*, 11(3), 1-11. <https://doi.org/10.1002/aps3.11517>
- [50] Segarra-Moragú J.G. and G. Gleiser. 2009. Isolation and characterization of di and tri nucleotide microsatellite loci in *Rosmarinus officinalis* (Lamiaceae), using enriched genomic libraries. *Conservation Genetics* 10: 571-575.
- [51] Setotaw TA, Caixeta ET., Pereira AA. *et al.*, 2013. Coefficient of parentage in *Coffea Arabica* L. cultivars grown in Brazil. *Crop Sci* 53: 1237-1247. <https://doi.org/10.2135/cropsci2012.09.0541>
- [52] Silvestrini M., Junqueira M. G., Favarin A. C., Guerreiro-Filho O., Maluf M. P., Silvarolla M. B. and Colombo C. A., 2007. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellite markers. *Genet Resour Crop Evol* 54: 1367-1379.
- [53] Sousa, T. V., Caixeta, E. T., Alkimim, E. R., de Oliveira, A. C. B., Pereira, A. A., Zambolim, L., & Sakiyama, N. S. (2017). Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica*, 213(3). <https://doi.org/10.1007/s10681-017-1865-9>

- [54] Szpiech, Z. A., & Rosenberg, N. A. (2011). On the size distribution of private microsatellite alleles. *Theoretical Population Biology*, 80(2), 100-113. <https://doi.org/10.1016/j.tpb.2011.03.006>
- [55] Tadele, S., 2012. Genetic diversity of coffee (*Coffea arabica* L.) landraces from major coffee growing areas of southern Ethiopia as revealed by inter simple sequence repeat marker. An M. Sc. Thesis submitted to the school of graduate studies of Haramaya University, 10-45PP
- [56] Tesfaye K., 2006. Genetic Diversity of Wild *Coffea arabica* Populations in Ethiopia as a contribution to conservation and use planning. PhD thesis, Bonn University, Germany. Pp 142.
- [57] Vavilov, N. I., 1951. The origin, variation, immunity and breeding of cultivated plants. (Translated by S. K. Chestitee). *Chronica Botanica*, 13, 1-366
- [58] Weising, K, Nybom, H, Wolff, K and Kahl, G. 2005. DNA Fingerprinting in Plants: Principles, Methods, and Applications. 2nd ed. CRC Press, Taylor and Francis Group. <http://dx.doi.org/10.1201/9781420040043>
- [59] Weldemichael, G. (2023). Review on Coffee (*Coffea arabica* L.) Genetic Diversity Studies Using Molecular Markers. 11(5), 164-170. <https://doi.org/10.11648/j.jps.20231105.14>
- [60] Woyesa, T., & Kumar, S. (2021). Potential of coffee tourism for rural development in Ethiopia: a sustainable livelihood approach. *Environment, Development and Sustainability*, 23(1), 815-832. <https://doi.org/10.1007/s10668-020-00610-7>
- [61] Yan W, Hunt LA, Sheng Q, Szlavniks Z (2000). Cultivar evaluation and mega-environment investigation based on the GGE biplot. *J Crop Sci* 40(3): 597-605.
- [62] Yilma Yemane Teferaa, 2017. A guide to coffee production in Ethiopia. Addis Ababa Ethiopia. P 1-280.