Original Research Paper

Assessment of Genetic Diversity Among Elite Breeding Lines of Arabica Coffee (*Coffea arabica* L.) in Ethiopia using Simple Sequence Repeats Markers

[1]Tadesse Benti, [2]Endale Gebre, [3,8]Kassahun Tesfaye, [4]Gezahegn Berecha,
[5]Philippe Lashermes, [6]Martina Kyalo and [7]Nasser Kouadio Yao

[1]*Department of Coffee Breeding and Genetics, Jimma Agricultural Research Center (JARC), Ethiopian Institute of Agricultural Research (EIAR), Ethiopia*
[2]*College of Agricultural and Natural resources, Michigan State University East Lansing, Michigan, USA*
[3]*Department of Agricultural Biotechnology, Ethiopian Biotechnology Institute, Ethiopia*
[4]*Department of Horticulture and Plant Science, Jimma University, College of Agriculture and Veterinary Medicine, Ethiopia*
[5]*Diversity, Adaptation and development of Cultivated Plants (DIADE), IRD, University of Montpellier, France*
[6]*Department of Biosciences eastern and central Africa (BecA-ILRi) Hub, International Livestock Research Institute, Kenya*
[7]*Alliance Bioversity International-Center for Tropical Agriculture (CIAT), Nairobi, Kenya*
[8]*Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Ethiopia*

Abstract: Despite being the center of origin and genetic diversity, there has been limited application of molecular markers in arabica coffee breeding in Ethiopia. In this study, the extent of genetic diversity and relationships of 48 elite breeding lines of arabica coffee from different geographical origins were evaluated using 14 SSR markers. The SSR markers amplified 105 alleles of which 104 are polymorphic with an average of 7.4 polymorphic alleles per locus. The average percent of polymorphism and Polymorphic Information Content (PIC) of the SSR markers were 98.81 and 0.79, respectively. The average genetic similarity coefficient among all possible pairs was 0.37 and ranged from 0.39 to 0.46 within the same geographical origin. The Unweighted Pair Group Method Using Arithmetic averages (UPGMA) cluster analysis placed the studied breeding lines into seven clusters with Jaccard's similarity coefficient of ≤0.44. The results of this study demonstrated the powerfulness of the SSR markers and the presence of a high level of genetic diversity and wide genetic dissimilarity within and among elite breeding lines from the same and different geographical origins. The observed diversity could be utilized in the breeding program to improve productivity and maintain the typical quality profile of origin-based landrace coffees. The SSR markers could be used in future genetic analysis of arabica coffee germplasm to increase the efficiency of the breeding program in Ethiopia. The studied breeding lines could also be conserved and managed as an important source of genetic variability to be used in the future coffee improvement program.

Keywords: Genetic Parameters, Genetic Similarity, Landrace Coffee, SSR Polymorphism, Specialty Coffee

Introduction

The beverage quality, aromatic characteristics and low caffeine content of arabica coffee make it the most preferred over other coffee species in the global coffee market. In addition, the ever-increasing consumers demand for differentiated and high-quality coffee contributed to the double-digit growth of the specialty coffee market (Giovannucci *et al.*, 2008). This necessitates the coffee breeding program elsewhere to focus on developing productive and genetically superior quality coffee varieties with specific flavors for the sustainability of the market (Montagnon *et al.* 2012). Ethiopia has a great potential to supply such differentiated high-quality coffee, due to the presence of immense genetic diversity and different geographical regions that produce arabica coffee with unique flavors recognized in the international trade (Gallu, 2006; Labouisse *et al.*, 2008). However, quality was not such a major area of focus of the coffee breeding program until very recently. To exploit all the available

opportunities of agro-ecological and genetic diversity, a new coffee breeding strategy known as 'Local coffee landrace variety development program' was designed and adopted recently by the coffee breeding program of Jimma Agricultural Research Center (JARC) in Ethiopia (Bellachew and Labouisse, 2007). The new strategy aimed at developing improved coffee varieties that possess a typical quality profile independently for each geographical region using coffee germplasm accessions collected from and evaluated in the respective major coffee-producing areas.

In effect, a large number of coffee accessions were collected and planted in field genebank of different research sub-centers of JARC established in major coffee-producing geographical regions (Labouisse *et al*., 2008; Benti, 2017). Eventually, the evaluation of some batches of the germplasm collections, as per the step-wise procedures described in the new breeding strategy (Bellachew and Labouisse, 2007), has resulted in the selection of 48 elite/promising breeding lines for Limmu (11), Illubabor (13), Wollega (12) and Bale (12) geographical regions that are among the areas representing primary coffee-producing regions in Ethiopia. These regions are also known to produce arabica coffee with a unique quality profile. Currently, the elite breeding lines are under the final breeding stage (verification trial) in the respective regions to confirm their performance before recommendation for release as a new pure-line landrace coffee variety.

It is well known that a better understanding of the extent of genetic diversity of breeding materials could help in the efficient management and utilization of genetic resources in the breeding program. It has also been recognized that molecular markers-based genetic diversity study, among others, provides useful information that helps to understand the extent of genetic diversity of available genetic resources and improves selection efficiency to maximize genetic gain (Tesfaye *et al*., 2014). Among DNA markers, Simple Sequence Repeats (SSRs) markers are considered ideal to study genetic diversity and relationships among closely related crop species such as arabica coffee due to their abundance, hyper-variability, multi-allelic and co-dominant nature (Powell *et al*., 1996). Comparative studies using RFLP, RAPD, AFLP and SSRs have also indicated that SSRs markers detect a higher level of polymorphism than other markers (Powell *et al*., 1996; Dessalegn *et al*., 2009). Simple sequence repeats markers were also found to be efficient in discriminating within and among selected varieties of arabica, robusa and liberica coffee species (Santos *et al*., 2016; Pruvot-Woehl *et al*., 2020; Montagnon *et al*., 2021). Emanuelli *et al*., (2013), also confirmed the high discriminating capacity of SSR markers in grapevine varieties as well as being as efficient as SNPs to establish the genetic diversity of the evaluated

varieties. In another study, SSR markers were also found to be more effective than SNP markers when the objective was mainly focused on the study of genetic diversity (Singh *et al*., 2013). Moreover, the high level of polymorphism in SSRs also makes it a dependable, practical and cost-effective choice (Hodel *et al*., 2016).

Previous studies on coffee using various molecular markers reported the narrow genetic base and the low level of genetic diversity in *ex-situ* conserved arabica coffee germplasm accessions, cultivated varieties and elite breeding lines of arabica coffee outside of Ethiopia (Anthony *et al*., 2002; Maluf *et al*., 2005; Missio *et al*., 2011; Geleta *et al*., 2012; Al-Murish *et al*., 2013; Scalabrin *et al*., 2020). On the other hand, moderate to a high level of genetic diversity was reported among arabica coffee germplasm accessions established in *ex-situ* field genebank in different countries including Ethiopia (Anthony *et al*., 2001; Moncada and McCouch, 2004; López-Gartner *et al*., 2009; Mekbib *et al*., 2022) and specifically in forest coffee populations and farmers' cultivars in Ethiopia (Aga *et al*., 2005; Al-Murish *et al*., 2013; Tesfaye *et al*., 2014) using various DNA markers.

Although outstanding achievements have been recorded following the conventional breeding approach, the lack of application of molecular markers has long been recognized as one of the research gaps in the coffee breeding program in Ethiopia (Labouisse *et al*., 2008). Among the very limited works done so far, (Teressa *et al*., 2010) reported the presence of a high level of genetic diversity in 57 coffee accessions obtained from the research plot of the main breeding program using SSR markers. Similarly, SSR markers revealed a high level of genetic diversity among 40 commercial arabica coffee varieties developed from the coffee breeding program in Ethiopia (Benti *et al*., 2021). However, the extent of genetic diversity and relationships within and among the aforementioned elite breeding lines selected for the respective geographical regions has not been studied using DNA-based molecular markers. As new varieties are continuously developed from any breeding program, assessment of genetic diversity among elite breeding lines and selection of genetically divergent genotypes is very crucial for the development and release of improved pure line varieties and exploiting hybrid vigor in any crop including coffee (López-Gartner *et al*., 2009). Moreover, analysis of genetic diversity among genetic resources obtained from different geographical origins is vital to understand the level of diversity present within the same and among different regions that would help to design appropriate breeding and conservation strategies. Therefore, in the present study, SSR markers were used to detect polymorphism and generate information on the level of genetic diversity and relationships within and among 48 elite breeding lines (upcoming new varieties) of arabica coffee originating from different geographical regions.

Materials and Methods

Genetic Materials and DNA Extraction

A total of 48 elite breeding lines selected from four geographical regions were used in this study. The detailed information is summarized in Table 1. These breeding lines are pure-line selections advanced from different batches of germplasm collections based on coffee production areas where the local landrace variety development is targeted. The DNA of each breeding line was extracted from silica-gel dried young leaves of a single plant following a modified version of the CTAB method (Borsch *et al*., 2003). The DNA purification, quality, concentration and dilution were performed as described by (Benti *et al*., 2021).

SSR Markers and PCR Amplification

In this study, 14 SSR primer pairs, labeled with 6FAM, NED, PET, or VIC fluorescent dye at the 5′-end of the forward primers, were used (Table 2). The SSRs markers were selected based on their high polymorphism and discrimination capability reported in previous studies (Combes *et al*., 2000; López-Gartner *et al*., 2009) and by Institute for Research and Development (IRD), France (personal communication). The annealing temperature of each primer, PCR amplification, SSR multiplexing and allele calling was performed as described in detail by (Benti *et al*., 2021).

Genotyping and Data Analysis

The allele peaks were visually inspected and then analyzed using GeneMapper Software 4.0 (Applied Biosystems) based on the GeneScan™ 500 LIZ® Size Standard. Each peak was considered as an allele of an appropriate microsatellite locus according to size (bp) and area of the peak. Accordingly, every allele of each primer across all breeding lines was scored and used for the analysis. To calculate the basic genetic parameters, scored alleles were formatted to present (1) or absent (0) for polymorphic marker loci (Medini *et al.*, 2005). Number of total (Na) and polymorphic alleles (Pa) per locus and rate of polymorphism (Pr) were calculated as described by (Morgante *et al*., 1994). The Polymorphism information content (PIC) values were also determined based on allelic frequency using PowerMarker version 3.25 (Liu and Muse, 2005). A binary data matrix was generated from the allelic data. Pair-wise genetic similarity was also estimated between the breeding lines using Jaccard's similarity coefficient (Jaccard, 1908). The resulting similarity matrix was used for cluster analysis and the relationships among breeding lines were displayed as a dendrogram constructed based on the UPGMA method (Sneath and Sokal, 1973). Correlation between the underlying similarity matrices and the relationships between breeding lines in the dendrogram, the cophenetic correlation coefficient (r), was computed with 1000 permutations. The significance of the cophenetic correlation was also tested by the Mantel correspondence test (Mantel, 1967). All analyses were performed using NTSYS–PC 2.11 software (Rohlf, 2000).

Table 1: List of and information on Elite Breeding Lines (EBL) of arabica coffee used in this study

EBL	Origin	EBL	Origin	EBL	Origin
L1	Limmu	mk1	Illubabor	B1	Bale
L2	Limmu	mk2	Illubabor	B2	Bale
L3	Limmu	mk3	Illubabor	B3	Bale
L4	Limmu	mk4	Illubabor	B4	Bale
L5	Limmu	mk5	Illubabor	B5	Bale
L6	Limmu	mk6	Illubabor	B6	Bale
L7	Limmu	w1	Wollega	B7	Bale
L8	Limmu	w2	Wollega	B8	Bale
L9	Limmu	w3	Wollega	B9	Bale
L10	Limmu	w4	Wollega	B10	Bale
L11	Limmu	w5	Wollega	B11	Bale
mm1	Illubabor	w6	Wollega	B12	Bale
mm2	Illubabor	w7	Wollega		
mm3	Illubabor	w8	Wollega		
mm4	Illubabor	w9	Wollega		
mm5	Illubabor	w10	Wollega		

Table 2: List of primer pairs used to amplify the SSR loci of the elite breeding lines of arabica coffee along with their annealing temperature (Ta)

Locus Name	Primer sequence (5' to 3')		Ta (ºC)
	Forward + Dye	Reverse	
AJ-250254[a]	GGCTCGAGATAT CTGTTTAG-**VIC**	TTTAATGGGCA TAGGGTCC	58
AJ-250255[a]	CCCTCCCTGCCA GAAGAAGC-**NED**	AACCACCGTCC TTTTCCTCG	58
AJ-250260[a]	TGATGGACAGGA GTTGATGG-**6-FAM**	TGCCAATCTAC CTACCCCTT	58
Sat-237[b]	CAAGAGCAGACGA TTCTCAATCT-**6-FAM**	TTGGGGTTAG GAAATCACAAT	58
Sat-171[b]	TTCCCCCATCT TTTTCTTTC-**VIC**	TTGTATACGGCT CGTCAGGT	58
CFGA-465[b]	ACCCTTTACTACTTA TTTACTCTC-**6-FAM**	ACATCCCCTTG CCATTTCTTC	62
AJ-250257[a]	GACCATTACATTTC ACACAC-**NED**	GCATTTTGTTGC ACACTGTAGCAA	58
Sat-235[b]	TCGTTCTGTCATT AAATCGTCAA-**PET**	ATCATGAAA ATAGTTGGTG	58
MR-054[b]	TGATGTGGAAG GCCATTG-**VIC**	GCCCCTATTATG ACCCATGC	62
Sat-180[b]	CATGTGTAATACA TTCAACAGTGA-**NED**	GCAATAGTGGT TGTCATCCTT	60
AJ-250258[a]	AAC TCT CCA TTC CCG CAT TC-**PET**	CTG GGT TTT CTG TGT TCT CG	62
Sat-41[b]	AGTGTAACTTTAG TTCTTGC-**PET**	ATTTAATGGGC ATAGGGTC	58
AJ-250253[a]	CTTGTTTGAGTCT GTCGCTG-**VIC**	TTTCCCTCCC AATGTCTGTA	58
MR-336[b]	GAGTCGTCCACAC TGCTTGA-**6-FAM**	CATCTGCTTTG GTCCCTGAT	60

[a] Combes *et al*. (2000). [b] Institute for Research and Development (IRD) France (personal communication)

Results

SSR Polymorphism

The genetic parameters (measures of genetic diversity) calculated for the 14 SSR markers are shown in Table 3. A total of 105 alleles, of which 104 are polymorphic were amplified by the markers across the evaluated breeding lines. The allelic richness of the SSR markers varied from 3 (AJ-250255) to 10 (AJ-250260), with an average of 7.5 total and 7.4 polymorphic alleles per locus. The Polymorphism rate (Pr) ranged from 83.3% to 100%, with an average of 98.8 percent. The PIC values also ranged from 0.26 (Sat-180) to 0.94 (AJ-250260), with an average of 0.79 per locus.

Genetic Similarity

The differences among the evaluated breeding lines at the DNA level were determined by comparing the genetic similarity. The Jaccard's similarity coefficient values among all possible pair-wise combinations ranged from 0.15 to 0.78, with an overall mean of 0.37 (data not presented). In general, 92% of the total pair-wise combinations exhibited genetic similarity values ranging from 0.15 to 0.50. Moreover, the genetic similarity coefficient between elite breeding lines from the same geographical origin was also evaluated independently. Accordingly, the similarity coefficient values within elite breeding lines from Limmu, Ilubabor,

Wollega and Bale origins ranged from 0.20 to 0.78 with average values of 0.42, 0.39, 0.46 and 0.45, respectively (Table 4A-4D). In the same order, 78.2, 89, 68.2 and 79% of the pair-wise combinations among individuals of the same origin were ≤ 0.50.

Clustering Patterns

The Jaccard's genetic similarity-based cluster analysis of the breeding lines revealed seven clusters with genetic distance values ranging from 0.56 to 0.73 (Fig. 1). The cophenetic correlation value between the dendrogram and the original similarity matrix was 0.88 with correspondence being significant at $p = 0.05$ by the Mantel test. Cluster I consisted of five genotypes from Bale and one from Limmu origin. Cluster II comprised five genotypes originating from Illubabor group. Similarly, Cluster III contained only two genotypes from Illubabor.

Cluster IV contained 18 genotypes of different geographical origins viz., Limmu (eight), Bale (six) and Illubabor (four). This cluster divided in to three sub clusters where some of the breeding lines from Limmu, Bale and Illubabor were grouped closely to each other in the sub clusters. Cluster V consisted of 11 genotypes, all of which were from Wollega. Cluster VI contained only two genotypes each were from Bale and Wollega. The last cluster (VII) consists of four genotypes that belong to Illubabor (two) and Limmu (two).

Table 3: Indices of genetic diversity in elite breeding lines of arabica coffee using 14 SSR markers

Locus name	Na	Pa	Pr (%)	PIC
AJ-250254	9.0	9.0	100.00	0.88
AJ-250255	3.0	3.0	100.00	0.79
AJ-250260	10.0	10.0	100.00	0.94
Sat-237	6.0	5.0	83.33	0.52
Sat-171	6.0	6.0	100.00	0.76
CFGA-465	8.0	8.0	100.00	0.87
AJ-250257	9.0	9.0	100.00	0.85
Sat-235	9.0	9.0	100.00	0.87
MR-054	9.0	9.0	100.00	0.78
Sat-180	4.0	4.0	100.00	0.26
AJ-250258	9.0	9.0	100.00	0.86
Sat-41	9.0	9.0	100.00	0.89
AJ-250253	9.0	9.0	100.00	0.90
MR-336	5.0	5.0	100.00	0.82
Total	105.0	104.0		
Average	7.5	7.4	98.81	0.79

Note: Na = Number of total alleles, Pa = polymorphic alleles,

Pr = polymorphism rate, PIC = polymorphic information Content

Table 4: Jackard's similarity coefficients among elite breeding lines of arabica coffee originated from Limmu (A), Ilubabor (B), Wollega (W) and Bale (B)

A

Gen.	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	1.00									
L2	0.50	1.00					**Range = 0.24-0.68**			
L3	0.41	0.51	1.00				**Average = 0.42**			
L4	0.51	0.68	0.56	1.00			**78.2% ≤ 0.50**			
L5	0.25	0.33	0.39	0.32	1.00					
L6	0.26	0.31	0.25	0.27	0.30	1.00				
L7	0.38	0.39	0.54	0.45	0.38	0.24	1.00			
L8	0.45	0.56	0.43	0.53	0.38	0.44	0.38	1.00		
L9	0.50	0.45	0.51	0.43	0.37	0.27	0.55	0.44	1.00	
L10	0.34	0.42	0.33	0.40	0.24	0.36	0.44	0.41	0.40	1.00
L11	0.49	0.51	0.46	0.53	0.35	0.41	0.42	0.59	0.37	0.44

B

Gen.	mm1	mm2	mm3	mm4	mm5	mm6	mm7	mk1	mk2	mk3	mk4	mk5
mm1	1.00											
mm2	0.29	1.00						**Range = 0.20-0.65**				
mm3	0.45	0.45	1.00					**Average = 0.39**				
mm4	0.58	0.36	0.64	1.00				**89.2% ≤ 0.50**				
mm5	0.22	0.31	0.20	0.27	1.00							
mm6	0.61	0.33	0.46	0.59	0.23	1.00						
mm7	0.33	0.43	0.36	0.37	0.32	0.49	1.00					
mk1	0.38	0.26	0.23	0.28	0.28	0.36	0.26	1.00				
mk2	0.34	0.38	0.37	0.33	0.27	0.42	0.38	0.28	1.00			
mk3	0.29	0.50	0.26	0.28	0.25	0.39	0.47	0.32	0.41	1.00		
mk4	0.43	0.40	0.39	0.43	0.24	0.59	0.48	0.36	0.49	0.43	1.00	
mk5	0.36	0.53	0.41	0.43	0.32	0.43	0.46	0.30	0.35	0.45	0.43	1.00
mk6	0.62	0.31	0.48	0.65	0.24	0.64	0.42	0.28	0.33	0.34	0.45	0.41

C

Gen.	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11
w1	1.00										
w2	0.64	1.00						**Range = 0.24-0.68**			
w3	0.62	0.39	1.00					**Average = 0.46**			
w4	0.54	0.41	0.56	1.00				**68.2% ≤ 0.50**			
w5	0.43	0.35	0.44	0.49	1.00						
w6	0.38	0.29	0.47	0.44	0.49	1.00					
w7	0.59	0.41	0.56	0.53	0.41	0.53	1.00				
w8	0.36	0.34	0.24	0.29	0.27	0.26	0.28	1.00			
w9	0.44	0.39	0.45	0.43	0.34	0.42	0.50	0.37	1.00		
w10	0.54	0.41	0.51	0.53	0.35	0.40	0.68	0.34	0.54	1.00	
w11	0.59	0.37	0.61	0.53	0.32	0.44	0.63	0.31	0.54	0.68	1.00
w12	0.47	0.38	0.41	0.43	0.43	0.31	0.59	0.33	0.48	0.59	0.50

D

Gen.	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11
B1	1.00										
B2	0.53	1.00						**Range = 0.23-0.78**			
B3	0.57	0.78	1.00					**Average = 0.45**			
B4	0.38	0.46	0.46	1.00				**79.2% ≤ 0.50**			
B5	0.36	0.53	0.45	0.34	1.00						
B6	0.33	0.53	0.45	0.38	0.43	1.00					
B7	0.42	0.50	0.58	0.34	0.53	0.42	1.00				
B8	0.41	0.50	0.46	0.36	0.31	0.53	0.40	1.00			
B9	0.35	0.50	0.43	0.28	0.34	0.53	0.40	0.62	1.00		
B10	0.36	0.50	0.43	0.29	0.36	0.49	0.41	0.47	0.69	1.00	
B11	0.41	0.39	0.42	0.25	0.31	0.41	0.37	0.60	0.52	0.49	1.00
B12	0.34	0.32	0.35	0.23	0.28	0.26	0.33	0.35	0.38	0.39	0.57
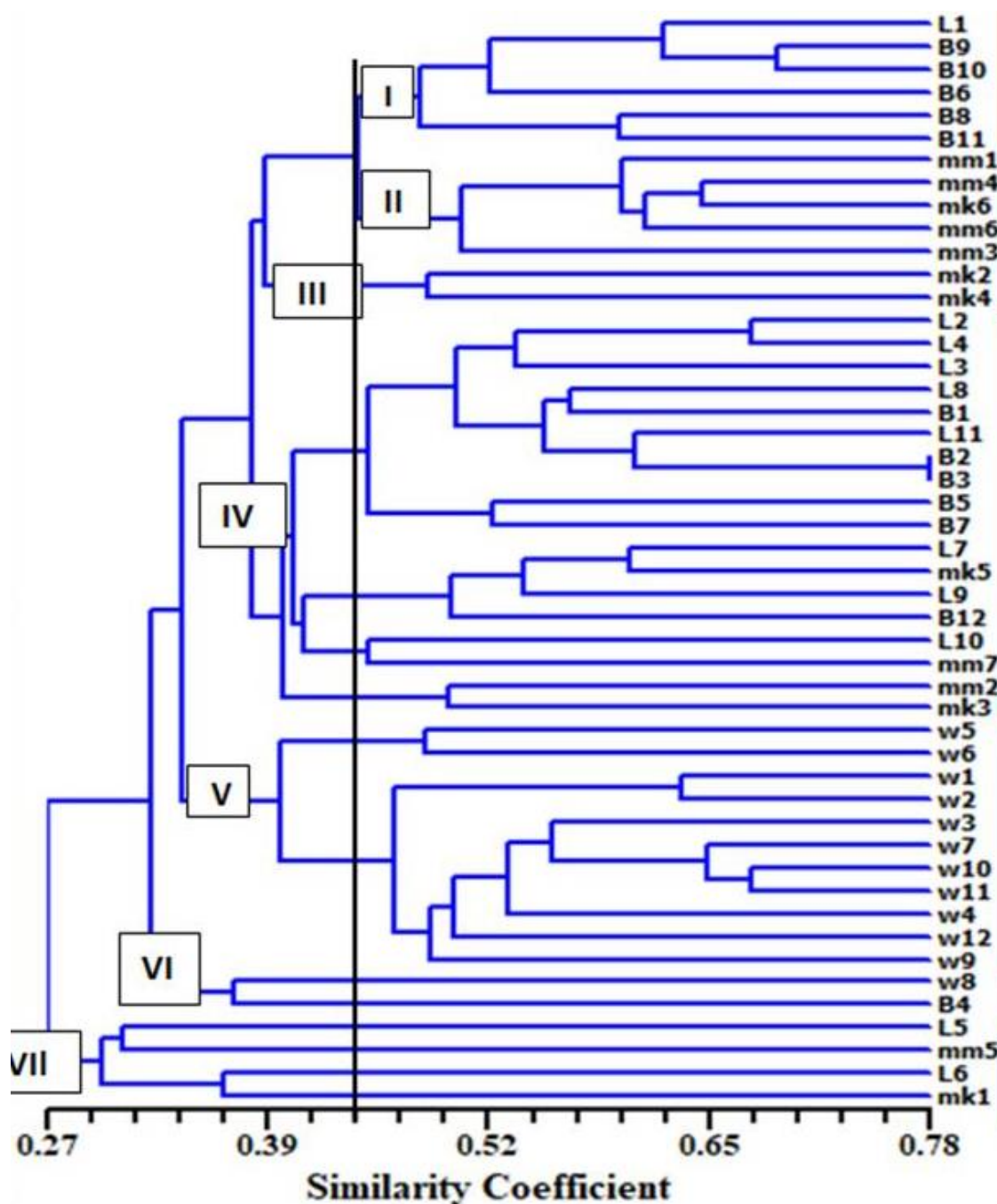
Fig. 1: Dendrogram generated after UPGMA based on Jaccard's similarity coefficient of the 48 elite breeding lines using 14 SSR markers

Discussion

SSR Polymorphism Among Elite Breeding Lines

The SSR markers used in this study were found to be highly polymorphic (Table 3). Except for primer set Sat-180 which showed a moderate (0.26) PIC value all the remaining 13 SSR markers detected high (>0.50) PIC values indicating high discriminating power of the markers. Moreover, none of the evaluated breeding lines exhibited the same profile across the 14 loci assessed. The results demonstrate the potential of the markers for use in genotyping, quantification of genetic diversity and differentiating between elite breeding lines by fingerprinting. This agrees with previous results of several

authors who reported the successful use of SSR markers in diversity analysis of different genetic categories of *Coffea arabica* (Teressa *et al*., 2010; Sousa *et al*., 2017; Pruvot-Woehl *et al*., 2020; Benti *et al*., 2021; Montagnon *et al*., 2021). The mean values of genetic parameters exhibited by the SSR markers were also high, demonstrating the presence of a high level of genetic diversity among the elite breeding lines. Similar results were reported in commercial verities of arabica coffee in Ethiopia, with mean values of number of polymorphic alleles (7.4) per locus, rate of polymorphism (98.1%) and PIC (0.80) using the same sets of SSR markers (Benti *et al*., 2021). In agreement with our study, (Pruvot-Woehl *et al*., 2020; Montagnon *et al*., 2021) also reported a high level of genetic diversity in arabica coffee comprised of different breeding categories, with 11.9 and 7.4 alleles per marker respectively, using eight SSR markers.

In contrast, (Missio *et al*., 2011) reported the presence of a low level of genetic diversity with a mean number of 4.0 and 1.9 alleles per locus and 11.1 and 28.6% polymorphism in elite breeding lines of arabica coffee developed from Brazilian coffee breeding program using 17 EST-SSR and 18 g SSR markers, respectively. Values of genetic parameters detected in our study are also higher than those reported among cultivated commercial varieties and *ex-situ* conserved genetic resources of arabica coffee outside of Ethiopia using SSR markers (Anthony *et al*., 2001; Teressa *et al*., 2010; Geleta *et al*., 2012; Al-Murish *et al*., 2013). In most of these studies, low levels of genetic diversity with mean values ranging from 2.0-2.87 alleles per locus, 32-42% polymorphism and 0.22-0.33 PIC were reported across the evaluated coffee samples. These differences compared to the results of the present and the previous studies could partly be explained by the variation in the number and repeat motifs of the SSRs primer sets used for the analysis as well as the sample size and genetic background of the studied genetic materials. The elite breeding line analyzed in our study were selected from populations with broad genetic bases assembled from diverse geographical origins whereas most of the coffee samples analyzed in previous studies were traditional cultivars developed by line selection or crosses between parental lines derived from bourbon and Typica that are known to have a narrow genetic base (Van der Vossen, 1985).

Genetic Similarity Among Evaluated Breeding Lines

The genetic similarity coefficient values for the majority (92 %) of the pair-wise combinations and the overall average (0.37) were very low, indicating the presence of wide genetic diversity among the studied breeding lines. A similar result was reported by Benti *et al*. (2021) in commercial coffee varieties in Ethiopia using the same stets of SSR markers. Such results are expected as the coffee breeding program

has been relying on a source population having a broad genepool assembled from fairly distant geographical regions (Benti, 2017). The present study also revealed the presence of a high level of genetic diversity among individuals of the same geographical origin. This can be noted from Table 4A-D where 68.2 to 89% of the pair-wise as well as the average similarity coefficient values detected within the regions were less or equal to 0.50 which corroborates with values observed in local landrace varieties developed independently for Hararge, Sidama/Yirga cheffe and Wollega coffee-producing regions (Benti *et al*., 2021). However, several authors reported contrasting results and very close genetic relationships with similarity coefficients ranging from 0.90 to 0.96 in previous studies using SSRs and other marker systems (Dessalegn *et al*., 2009; Mishra *et al*., 2012; Al-Murish *et al*., 2013). The presence of distantly related elite breeding lines within the same geographical regions is of particular importance in the development of improved locally adapted landrace coffee varieties through pure-line selection or hybridization without affecting the typical quality profile of the targeted coffee-producing areas viz., Limmu, Illubabor, Wollega and Bale.

Clustering Patterns of Elite Breeding Lines

The results of cluster analyses also agree with the results of genetic parameters and the estimates of genetic similarity. The cophenetic correlation coefficient (0.88) was high, indicating that the dendrogram corresponded to 88% of the similarity matrices. Accordingly, most of the breeding lines with low pair-wise genetic similarity coefficients (0.20-0.24), for instance, L6xL7 (in clusters VII and IV), L5xL10 (clusters VII and VI), mm3xmm5 (clusters II and VII), w3xw8 (clusters V and VI) and B4xB12 (clusters VI and VII), belong to different clusters (Table 4A-D; Fig.1). Whereas those pairs of lines exhibited the highest coefficients, as in the case between L2 x L4 (0.68), m4 x m6 (0.65), w7 x w10 (0.68) and the most genetically similar lines B2 x B3 (0.78), grouped very closely in the same cluster viz., clusters IV, II, V and IV, respectively. Moreover, two clustering patterns were also noted on the denerogram. In the first case, breeding lines from the same geographical origin were distributed into different clusters, except those from Wollega where almost all are assigned to a single cluster (Fig. 1).

The distribution of the breeding lines into different clusters regardless of their geographical origin could facilitate easy identification of distantly related parental lines for crossing to develop new productive landrace hybrid coffee varieties. This, in turn, would account for broadening the genetic base of released varieties, as well as contribute towards a sustainable supply of origin-based high-quality coffee from each region to the specialty coffee market. Similar results were reported by (López-Gartner *et al*., 2009) who observed the distribution of 68 arabica coffee

accessions collected from the same geographical regions into different clusters using SSR markers. Furthermore, distantly related elite breeding lines from different geographical origins assigned into different clusters could also be crossed to study the effects of genetic and/or geographical variation on the typical quality profile as well as to develop productive hybrid coffee varieties that can be grown in wider agro-ecologies and changing environmental conditions in the current climate change scenario.

On the other hand, there was a clear relationship between geographic origin and the resulting clusters as observed in clusters I (Bale), II and III (Ilubabor) and V (Wollega), where each of them is dominated by genetic materials from the same origin. Similar results were reported previously by (Aga *et al.*, 2003; Tesfaye *et al.,* 2014) using ISSR markers, where coffee tree samples comprising forest, semi-forest and farmers' cultivars collected from different parts of Ethiopia were clustered based on their geographic origin. Such results were also reported from Indonesia, where 73 to 92 % genetic similarity was detected by eight SRAP markers in populations of five arabica coffee varieties collected from different locations (Yunita *et al.*, 2020). This indicates that some genotypes from the same geographic origin are more closely related at the DNA level than those from different origins. This could be due to a large number of shared alleles that resulted from gene flow within the population of the same region than between different regions. Such clustering pattern is more reflected by coffee types from Wellega (Cluster V) where proximity among all the breeding lines was observed on the denderogam. This could also be explained in part by the domestication process of Wollega coffee might have been initiated from source populations that differ in origin from the other groups of breeding lines evaluated in this study. In such cases, coffee breeders need to focus on the presence of morphological variation as a complement to genetic distance while selecting parental lines for hetersosis breeding or direct release as a new pure line variety. The importance of variation in morphology and/or geographical origin among parental lines in exploiting hybrid vigor in arabica coffee was reported by Bellachew *et al.* (2008).

Conclusion

In this study, we demonstrated the potential of SSR markers in depicting genetic diversity among elite breeding lines of arabica coffee. Hence, 13 of the markers that showed high PIC (above 0.50) values could be used for future genetic analysis in advanced breeding lines of arabica coffee. The use of SSR markers also verified considerable genetic diversity and distant relatedness within and among the elite breeding lines selected for each geographical origin. The information generated from this study could also be used by coffee breeders to design appropriate breeding strategies for regional and/or wider adaptable variety development considering origin-based quality profiles and the present climate change scenario, respectively. The studied breeding lines could also be conserved and managed as an important source of genetic variability to be used in the future coffee improvement program.

Conflict of Interest

The authors declare that they have no any conflict of interest.

Author's Contributions

Tadesse Benti: Conceived the idea, designed the analysis, collected data, performed data analysis and wrote the manuscript.

Endale Gebre, Nasser Kouadio Yao, Kassahun Tesfaye, Gezahegn Berecha, Philippe Lashermes and Martina Kyalo: Assisted in conception and design of the idea, interpretation of data, reviewed and approved the submission of the manuscript as well as the revised version.

Ethics

I declare that this work is fully incompliance with the ethical rules and standards of the 'American journal of Biochemistry and biotechnology'. I also declare that this manuscript was not submitted to another journal as well as all authors read and approved the final manuscript.

References

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.
https://doi.org/10.1034/j.1601-5223.2003.01636.x

Aga, E., Bekele, E., & Bryngelsson, T. (2005). Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. *Genetica*, 124(2), 213-221.
https://doi.org/10.1007/s10709-005-1484-6

Al-Murish, T. Elshafei, M., Al-Doss, A. A., & Barakat, M. N. (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., Bertrand, B., Quiros, O., Wilches, A., Lashermes, P., Berthaud, J., & Charrier, A. (2001). Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*, 118(1), 53-65.
https://doi.org/10.1023/A:1004013815166

Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties was revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104(5), 894-900.
https://doi.org/10.1007/s00122-001-0798-8

Bellachew, B., & Labouisse, J. P. (2007). Arabica coffee (*Coffea arabica* L.) local landrace development strategy in its center of origin and diversity. ASIC.
https://agritrop.cirad.fr/540084/7/ASIC2006-Communication-Bayetta-Labouisse.pdf

Bellachew, B., Atero, B., Tefera, F., Ayano, A., & Benti, T. (2008). Genetic diversity and heterosis in arabica coffee. In: Girma Adugna, Bayetta Belachew, Tesfaye Shimber, Endale Taye and Taye Kufa (eds.). Coffee Diversity and Knowledge. Proceedings of a National Workshop, Four Decades of Coffee Research and Development in Ethiopia, 14-17 August 2007, Addis Ababa, Ethiopia.

Benti, T. (2017). Progress in Arabica coffee breeding in Ethiopia: Achievements, challenges, and prospects. *Int J Sci Basic Appl Res*, 33(2), 15-25.
https://core.ac.uk/download/pdf/249335806.

Benti, T., Gebre, E., Tesfaye, K., Berecha, G., Lashermes, P., Kyallo, M., & Kouadio Yao, N. (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

Borsch, T., Hilu, K.W., Quandt, D., Wilde, V., Neinhuis, C., & Barthlott, W. (2003). Noncoding plastid trnT-trnF sequences reveal a well-resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology*, 16(4), 558-576.
https://doi.org/10.1046/j.1420- 9101.2003.00577.x

Combes, M. C. andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., & Lashermes, P. (2000). Characterization of microsatellite loci in *Coffea arabica* and related coffe species. *Molecular Ecology*, 9(8), 1178-1180.
https://doi.org/10.1046/j.1365-294x.2000.00954-5.x

Dessalegn, Y., Herselman, L., & Labuschagne, M. (2009). Comparison of SSR and AFLP analysis for genetic diversity assessment of Ethiopian arabica coffee genotypes. *South African Journal of Plant and Soil*, 26(2), 119-125.
https://doi.org/10.1080/02571862.2009.10639943

Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M., & Grando, M. S. (2013). Genetic diversity and population structure were assessed by SSR and SNP markers in a large germplasm collection of grapes. *BMC Plant Biology*,13(1),1-17.
http://www.biomedcentral.com/1471–2229/13/39

Gallu, J. (2006). A Hot Cup of Money, Starbucks, Ethiopia, and the Coffee Branding Wars. *Der Spiegel Online*, *16*.

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. *The Scientific World Journal*, *2012*. https://doi.org/10.1100/2012/939820

Giovannucci, D., Byers, A., & Liu, P. (2008). Adding value: Certified coffee trade in North America.
https://mpra.ub.uni-muenchen.de/17174/

Hodel, R. G., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., & Soltis, P. S. (2016). The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, *4*(6), 1600025.https://doi.org/10.3732/apps.1600025

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale Bulletin de la Société Vaudoise des Sciences Naturelles.
https://doi.org/10.5169/seals-268384

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(7), 1079-1093.
https://doi.org/10.1007/s10722-008-9361-7

Liu, K., & Muse, S. V. (2005). Power Marker integrated analysis environment for genetic marker data. *Bioinformatics*, 21(9), 2128-2129. https://doi.org/10.1093/bioinformatics/bti282

López-Gartner, G., Cortina, H., McCouch, S. R., & Moncada, M. D. P. (2009). Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree Genetics & Genomes*, 5(3), 435-446. https://doi.org/10.1007/s11295-008-0197-2

Maluf, M. P., Silvestrini, M., Ruggiero, L. M. D. C., Guerreiro Filho, O., & Colombo, C. A. (2005). Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP, and SSR marker systems. *Scientia Agricola*, 62, 366-373. https://doi.org/10.1590/S0103- 90162005000400010

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*,27(2_Part_1), 209-220.

Medini, M., Hamza, S., Rebai, A., & Baum, M. (2005). Analysis of genetic diversity in Tunisian durum wheat cultivars and related wild species by SSR and AFLP markers. *Genetic Resources and Crop Evolution*, 52 (1), 21-31. https://doi.org/10.1007/s10722-005-0225-0

Mekbib, Y., Tesfaye, K., Dong, X., Saina, J. K., Hu, G. W., & Wang, Q. F. (2022). Whole-genome resequencing of *Coffea arabica* L. (Rubiaceae) genotypes identify SNP and unravels distinct groups showing a strong geographical pattern. *BMC Plant Biology*, 22(1), 1-9. https://doi.org/10.1186/s12870-022-03449-4

Mishra, M. K., Sandhyarani, N., Suresh, N., Satheesh Kumar, S., Soumya, P. R., Yashodha, M. H., & Jayarama. (2012). Genetic diversity among Indian coffee cultivars determined via molecular markers. *Journal of Crop Improvement*, 26(6), 727-750. https://doi.org/10.1080/15427528.2012.696085

Missio, R. F., Caixeta, E. T., Zambolim, E. M., Pena, G. F., Zambolim, L., Dias, L. A. S., & Sakiyama, N. S. (2011). Genetic characterization of elite coffee germplasm assessed by gSSR and EST-SSR markers. In *Embrapa Café-Artigo em anais de congresso (ALICE). Genetics and Molecular Research*, 10 (4), 2366-2381, 2011. https://doi.org/10.4238/

Moncada, P., & McCouch, S. (2004). Simple sequence repeats diversity in diploid and tetraploid Coffea species. *Genome*, 47 (3), 501-509. https://cdnsciencepub.com/doi/abs/10.1139/g03-129

Montagnon, C., Mahyoub, A., Solano, W., & Sheibani, F. (2021). Unveiling a unique genetic diversity of cultivated *Coffea arabica* L. in its main domestication center: Yemen. *Genetic Resources and Crop Evolution*, 68 (6), 2411-2422. https://doi.org/10.1007/s10722-021-01139-y

Montagnon, C., Marraccini, P., & Bertrand, B. (2012). Breeding for coffee quality. *Specialty Coffee: Managing Quality. International Plant Nutrition Institute, Southeast Asia Program*, 93-122. http://www.researchgate.net/publication/23410 354

Morgante, M., Hanafey, M., & Powell, W. (1994). "Microsatellites are preferentially associated with non-repetitive DNA in plant genomes." *Nature Genetic* 30 (2), 194-200. https://doi.org/10.1038/ng82

Powell, W., Morgante, M. andre, C., Hanafey, M., Vogel, J., Tingey, S., & Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2(3), 225-238. https://doi.org/10.1016/S1360-1385(96) 86898-0

Pruvot-Woehl, S., Krishnan, S., Solano, W., Schilling, T., Toniutti, L., Bertrand, B., & Montagnon, C. (2020). Authentication of *Coffea arabica* Varieties through DNA Fingerprinting and its Significance for the Coffee Sector. *Journal of AOAC International*, *103*(2), 325-334. https://doi.org/10.1093/jaocint/qsz003

Rohlf, F. J. (2000). NTSYS 2.1: Numerical taxonomic and multivariate analysis system. *New York, Exeter Software*.

Santos, D. M. C., Besa, C. F. F., Victoria, A. J. A., & Cao, E. P. (2016). Simple Sequence Repeat Analysis of Selected NSIC-registered Coffee Varieties in the Philippines. *Science Diliman*, *28*(1). https://www.researchgate.net/publication/32147999

Scalabrin, S., Toniutti, L., Di Gaspero, G., Scaglione, D., Magris, G., Vidotto, M., Pinosio, S., Cattonaro, F., Magni, F., Jurman, I., Cerutti, M., Liverani, F.S. Navarini, L., Del Terra, L., Pellegrino, G., M.R. Ruosi, N. Vitulo, G. Valle, A. Pallavicini, Graziosi, G., Klein, P.E., Bentley, N., Murray, S., Solano, Al Hakimi,W. A., Schilling, Montagnon, T. C., Morgante, M. & Bertrand B. (2020). A single polyploidization event at the origin of the tetraploid genome of Cofea arabica is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Jurnal of Scientific Reports* 10:4642. https://doi.org/10.1038/s41598-020-61216-7

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., & Singh, R. (2013). Comparison of SSR and SNP markers in the estimation of genetic diversity and population structure of Indian rice varieties. *PloS one*, *8*(12), e84136. https://doi.org/10.1371

Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy. The principles and practice of numerical classification.*

Teressa, A., Crouzillat, D., Petiard, V., & Brouhan, P. (2010). Genetic diversity of Arabica coffee (*Coffea arabica* L.) collections. *Ethiopian Journal of Applied Science and Technology*, *1*(1), 63-79. http;//www.researchgate.net/publication/228798624

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 Systematics and Evolution*, *300* (5), 881-897. https://doi.org/10.1007/s00606-013-0927-2

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), 1-15. https://doi.org/10.1007/s10681-017-1865-9

Van der Vossen, H. A. M. (1985). Coffee selection and breeding. In: Clifford C, Willson J (eds) Coffee botany, biochemistry, and production of beans and beverage. Croom Helm, London, pp, 48-96.

Yunita, R., Oktavioni, M., Chaniago, I., Syukriani, L., Setiawan, M. A., & Jamsari, J. (2020), April). Analysis of genetic diversity of Arabica coffee [*Coffea arabica* L.] in Solok Regency by SRAP molecular markers. In *IOP Conference Series: Earth and Environmental Science* 497(1), IOP Publishing. https://iopscience.iop.org/article/10.1088/1755-1315/497/1/012018/meta