

RESEARCH ARTICLE

Open Access



Genetic structure and relationships within and between cultivated and wild korarima [*Aframomum corrorima* (Braun) P.C.M. Jansen] in Ethiopia as revealed by simple sequence repeat (SSR) markers

Dagmawit Chombe^{1*} , Endashaw Bekele¹, Tomas Bryngelsson², Abel Teshome² and Mulatu Geleta²

Abstract

Background: Korarima [*Aframomum corrorima* (Braun) P.C.M. Jansen] is a spice crop native to Ethiopia. Understanding the extent and partitioning of diversity within and among crop landraces and their wild relatives is among the first steps in conserving and measuring their genetic potential. The present study is aimed at characterizing the population genetic structure and relationships between cultivated and wild korarima in the southwestern part of Ethiopia.

Results: We analyzed a total of 195 individuals representing seven wild and fourteen cultivated populations. Eleven polymorphic simple sequence repeat (SSR) markers were used. We observed a total of 53 alleles across the eleven loci and individuals. In total, 32 alleles were detected in the cultivated populations, whereas 49 alleles were detected in the wild populations. We found higher genetic diversity in wild populations than in the cultivated counterpart. This result implies the potential of wild korarima as a possible source for novel alleles contributing to the improvement of cultivated korarima. Analysis of molecular variance (AMOVA) showed significant but low differentiation between cultivated and wild korarima populations. Similarly, neighbour-joining and STRUCTURE analyses did not group cultivated and wild populations into two distinct clusters. The lack of clear differentiation between cultivated and wild populations could be explained by historical and contemporary gene flow between the two gene pools.

Conclusion: The 11 SSR loci developed in this study could be employed to examine genetic diversity and population structure of korarima in other countries as well as other *Aframomum* species. From the five administrative zones considered in this study, the Bench-Magi and Sheka zone showed populations with high genetic diversity, and these populations could be used as a potential starting point for in-situ and ex-situ germplasm conservation and korarima improvement through breeding programs after proper agronomic evaluation.

Keywords: *Aframomum corrorima*, Korarima, Genetic diversity, Simple sequence repeats, Genetic structure, Genetic differentiation, Gene flow

* Correspondence: dagechombe@yahoo.com

¹Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, P. O. Box, 1176 Addis Ababa, Ethiopia

Full list of author information is available at the end of the article



Background

Korarima (*Aframomum corrorima* (Braun) P.C.M. Jansen) belongs to the monocotyledonous family Zingiberaceae [1]. It is an herbaceous, perennial and aromatic species native to Ethiopia [2, 3]. Korarima grows usually with strong fibrous subterranean scaly rhizomes and leafy stems reaching 1–2 m in length. The position of stigma in the flower is below or against the base of the thecae of the anther. Although it is usually self-pollinated, occasional cross-pollination by insects is possible due to the presence of large nectaries at the top of the ovaries [4].

Korarima, also called “false cardamom”, is a part of daily Ethiopian dishes as it is used for preparation of curry powder for culinary purposes. The seeds are used to flavour coffee, special kinds of bread, butter and all kinds of sauces [3]. From a survey by [5], korarima seeds, pods, leaves, rhizomes and flowers are all used in Southern Ethiopia as traditional medicine for different kinds of human and animal ailments. Korarima also plays a role in soil conservation as the rhizomes and leaves spread on the ground covering and protecting the soil from erosion in hilly areas year-around [5]. There is also a demand for korarima in the neighboring countries to Ethiopia as well as in Arabia and Europe where it has long been highly prized as a spice [6]. Therefore, this spice could be developed into an important commodity if necessary attention is given to its research and genetic improvement. The beneficial characteristics of the species can be further enhanced through plant breeding. However, data on genetic characterization of this species is scarce and hence considered necessary before any breeding work can commence. Determination of genetic diversity and population structure are prerequisites of breeding programs and a first step in the development and evaluation of plant genotypes. Phenotypic traits may not give reliable estimates of genetic diversity as these traits are influenced by environmental factors and are limited in number [7]. On the other hand, genetic diversity based on molecular data can potentially facilitate conservation and can be employed as a tool for mining germplasm collections for genomic regions associated with adaptive or agronomically important traits [8]. Microsatellites or simple sequence repeats (SSRs) are tandemly repeated motifs of two to six nucleotides found in all prokaryotic and eukaryotic genomes [9]. Microsatellites are co-dominantly inherited, highly abundant, polymorphic, multi-allelic and reproducible. Hence they have become one of the most desirable molecular markers for use in genetic studies [10]. Interestingly, SSR has been the marker of choice for assessment of genetic diversity in many plant species such as field pea [11], *Sorghum bicolor* [12], Arabica Coffee [13] and chickpea [14]. SSR markers were also used for analysis of phylogenetic relationships [15], marker assisted

selection [16], construction of genetic linkage maps [17] and quantitative trait loci maps [18].

The genus *Aframomum* lacks molecular markers that can be used for assessment of its genetic diversity and for other applications. It was, therefore, essential to develop new molecular markers such as microsatellites for *A. corrorima*. Microsatellites can be identified by screening DNA sequence databases for target or closely related species. In the present study, eleven new SSR markers were developed and utilized (1) to evaluate the genetic variation of wild and cultivated korarima populations in southwestern Ethiopia (2) to determine the presence of indirect evidence for crop-wild hybridization by testing for admixture of SSR alleles in cultivated and wild korarima populations from the same regions (3) and to determine potential factors shaping the population genetic structure of cultivated and wild korarima in Ethiopia.

Methods

Plant material

For this study, 195 individuals from 21 korarima populations (14 cultivated and 7 wild) were sampled across five zones/areas (Illubabour, Jimma, Sheka, Kefa and Bench-Maji) in the southwestern part of Ethiopia (Fig 1). Cultivated populations include korarima samples collected directly from the farmers filed and the wild populations were sampled from the forest. Formal identification of the samples was undertaken based on the description listed on flora of Ethiopia and Eritrea [19]. To collect korarima plant samples from the wild and farmers filed, permission was obtained from the local managers and the farmers respectively. All populations were represented by 10 individuals except three (Mizan-Teferi_C1, Masha_C1 and Metu_W) which were represented by 9, 3 and 3 individuals, respectively (Additional file 1). The distance between sampled plants within each population was at least 20 m. This was done in order to increase the likelihood of representing the genetic variation of each population with limited number of individual plants. Two young leaves were collected from each plant and sealed within plastic bags containing silica gel for DNA extraction. The samples were taken to genetics research laboratory at Addis Ababa University, Ethiopia, and stored at room temperature until the DNA extraction was conducted. Standard material transfer agreement to Sweden was obtained from Ethiopian Institute of Biodiversity conservation (IBC). A copy of each sample gathered was deposited in IBC gene bank for future utilization.

DNA extraction

Total genomic DNA was isolated from about 1 gram of pulverized leaf sample following a modified CTAB method employing triple extractions to yield optimal amounts of high quality DNA [20]. Genomic DNA from

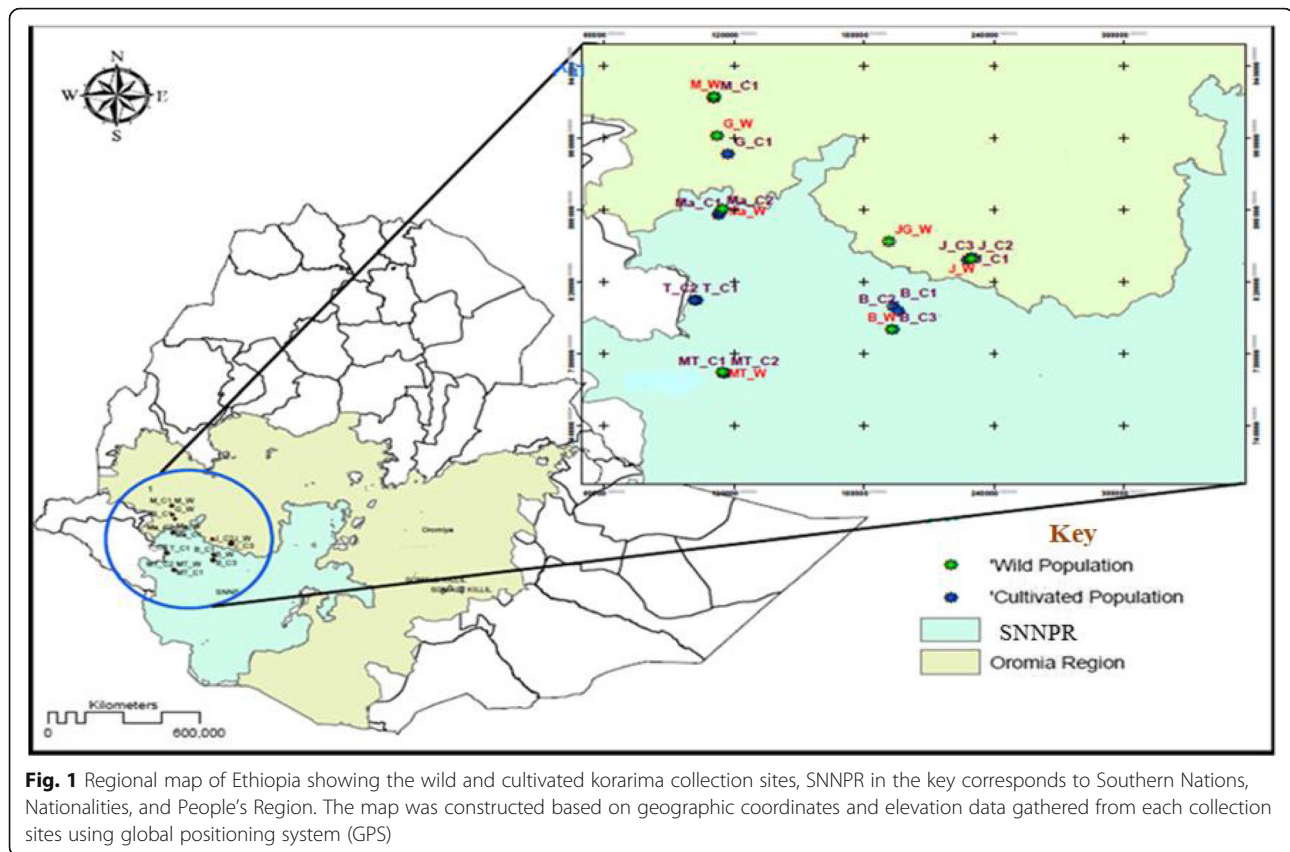


Fig. 1 Regional map of Ethiopia showing the wild and cultivated korima collection sites, SNNPR in the key corresponds to Southern Nations, Nationalities, and People's Region. The map was constructed based on geographic coordinates and elevation data gathered from each collection sites using global positioning system (GPS)

the second extractions was used for PCR amplification, as it was high in both quality and quantity when analyzed using a Nano Drop[®] ND-1000 spectrophotometer (Saveen Werner, Sweden). QIAamp[®] 96 DNA QIAcube[®] HT Kit was used to extract DNA from samples that we failed to isolate in desired quality and quantity using the CTAB method.

DNA database searches and development of genomic-SSR markers

Since there are no microsatellite containing DNA sequences of *Aframomum* in the NCBI nucleotide database, we used DNA sequences of *Alpinia* in GenBank to develop genomic SSR markers for *Aframomum*. *Alpinia* was used as it is the most closely related genus to *Aframomum* within the Zingiberaceae family [21] which has genomic resources at NCBI. We used 1498 DNA sequences of various *Alpinia* species in NCBI for mining SSRs. DNA sequences that contain SSR with two to six nucleotide repeat motif were identified using WebSat web-based software [22]. The analysis revealed that only 1.5% of these sequences contain SSRs and they all were from *Alpinia oxyphylla*. After evaluating the suitability of these sequences for designing primers, 23 SSR primer-pairs were designed using the Primer3 primer designing program [23].

SSR PCR amplification

The 23 SSR primer-pairs were initially tested for successful amplification of target site and capacity to detect polymorphism among genotypes (Additional file 2). Among these primers, 12 failed to amplify the target loci in *Aframomum corrorima* cultivated and wild populations. The remaining 11 primer-pairs successfully amplified their target loci in *A. corrorima* populations and hence used for genetic diversity analysis (Table 1). The forward primers of these primer-pairs were labeled at the 5'-end with either 6-FAM[™] or HEX[™] fluorescent dyes. In order to avert Taq DNA polymerase tendency to add non templated nucleotide to the PCR product as described in Ballard et al. [24], the reverse primers were PIG-tailed with "GCTTCT".

A total volume of 25 μ l containing 25 ng genomic DNA, 1 \times PCR buffer (10 mM Tris-HCl, pH 8.3 and 50 mM KCl), 0.3 mM dNTPs, 0.3 μ M forward and reverse primers, 1.5 mM MgCl₂ and 1 U Dream Taq DNA polymerase (Sigma, Germany) was used for PCR reactions. The reactions were performed using a GeneAMP PCR system 9700 thermo cycler (Applied Biosystems Inc. USA) in 96-well plates. PCR amplification reaction consisted of 3 min preliminary denaturation at 95 $^{\circ}$ C, 30 s touchdown denaturation at 94 $^{\circ}$ C for nine cycle, 30 s annealing at 58 $^{\circ}$ C (in every cycle annealing

Table 1 List of primer-pairs developed and used to amplify the SSR loci in this study

Locus	SSAN	Primer sequence (5'-3')	Repeat motifs	Expected size	Observed repeat Motifs	Allele size range ^a
Afc0_2	JX422069.1	F TTGACTTGGGTATGGCAAAA	(AG) ₁₃	230	-	212–242
		R AAGGTCGAGCAGGAGTAGCA				
Afc0_3	JX422068.1	F GAATTCATGTTCTTGAGAAAAGTTTG	(AG) ₇	198	(GA) ₂	191–200
		R GCCAAATGAACGGACAGATT				
Afc0_5	JX422066.1	F TGACTCCAAACTGCAGGAG	(CT) ₈	160	(CT) ₁₂	160–171
		R AGCAGATCAATGCACGTGAG				
Afc0_6	JX422065.1	F TCGACATGAAATCCCTACGAGA	(AG) ₁₅	243	(AG) ₁₂	230–253
		R GAGCTGTGAAGTGAAGGGC				
Afc0_8	JX422063.1	F GCTAACTGTCTTCTATTTCTCC	(CT) ₁₃	239	*	229–245
		R TGGAAGCTGCATTCACTGAG				
Afc0_11	JX422060.1	F AATGCTTCTAGCTGGTTCGAC	(GT) ₇	241	(GT) ₆	240–260
		R CCTTGAATTTATATTTCTCCAGATG				
Afc0_14	JX422057.1	F CCTTCCACGGTGTCTCATT	(GA) ₁₉	281	(GA) ₁₀	280–285
		R TCATCCAAAACCTCAATCATGG				
Afc0_15	JX422056.1	F ATCGATGGGATCGCCTTAC	(GA) ₁₉	292	(GA) ₉	256–262
		R GACGTCACGAATGTTGGTTG				
Afc0_19	JX422052.1	F CAGACGAGAGGAGGAGATG	(GA) ₁₇	373	(GA) ₉	356–364
		R CTCTGTGAGCCGTTCATCC				
Afc0_21	JX422050.1	F CGACAAGGAGGAGAAGAGGT	(GA) ₁₄	250	(GA) ₆ GAAAGG (GA) ₇	250–260
		R CCAACAGCCCTTCTTTTGA				
Afc0_22	JX422049.1	F GAAGAAGCGTTGGTGAGAGG	(TC) ₂₀	468	(TC) ₉	449–455
		R CTGTGTCGTCCAGCCGTATT				

^a= refers to allele size across all individuals included in the study; SSAN source sequence accession number; *= The SSR are located in the trimmed (low quality) part of the sequence and hence the length of the repeat motif could not be determined; – = sequencing failed

temperature was decreased by -1 °C) and elongation for 45 s at 72 °C, afterward 29 cycles of denaturation at 94 °C for 30 s, annealing at 48 °C for 30 s, and elongation at 72 °C for 45 s. Following completion of the 29 cycles, a 20 min final elongation at 72 °C was included to allow completion of the reactions. Quality check on the PCR products were conducted with gel electrophoresis and PCR products were stored at 4 °C until further use.

Gel electrophoresis and genotyping

For each locus, amplification and quality was confirmed by running 7 µl of the PCR products mixed with 2 µl of 6× loading dye on ethidium bromide containing 1.5% agarose gels. A DNA ladder with the size of 50 bp (GeneRuler™ Fermentas Life Sciences) was used as fragment size marker. The PCR products were multiplexed into three panels where markers with different fluorescent dyes and sizes placed together. ABI Prism 3730 DNA Analyzer (Applied Biosystems) was used for subsequent analysis of the multiplexed PCR products at Swedish University of Agricultural Sciences, Alnarp, Sweden.

Confirmation of SSRs using DNA sequencing and analysis

To confirm that the amplified products are the target microsatellite loci, the PCR products of two individuals of *A. corrorima* (one cultivated and one wild) amplified by the 11 primer-pairs were sequenced. Before the PCR products were sent for sequencing to Eurofins Genomics (Germany), they were purified using the Qiaquick PCR purification kit. Then, each purified PCR product was mixed with corresponding forward primer, which was used as sequencing primer. The final edited sequences were searched in NCBI through Basic Local Alignment Search Tool (BLAST) to find out if it hits with the original SSR containing microsatellite sequence or any other similar sequence in the database. Similar sequences were aligned using Clustal X version 2.1 software [25], and then sequences were edited using BIOEDIT version 7.0.5 [26].

Data analysis

Allele peak and its size was identified using GeneMarker® V2.2.0 software (Soft Genetics, LLS, State College, Pennsylvania) based on the Genescan-500 LIZ internal size standard. After pre-analysis of peaks using default settings

in GeneMarker and visual examination, peaks with threshold intensity of 200 were accepted for scoring.

For estimating observed heterozygosity, expected heterozygosity, gene flow and allele frequency based genetic distance analyses, POPGENE software version 1.31 [27] was used. The estimate of gene flow (N_m) was generated based on equation provided in [28]. GenAlEx 6.5 software V3.25 [29] was used to compute private alleles by population and locus and to calculate pair-wise F_{ST} . Analysis of molecular variance (AMOVA) was conducted using Arlequin 3.0 according to Excoffier et al. [30]. A neighbor-joining (NJ) and UPGMA (unweighted pair group method using arithmetic average) dendrogram was constructed based on pair-wise simple matching dissimilarity index using DARwin6 software [31]. In the NJ tree construction a random re sampling of five individuals from 19 population and three individuals from two populations [(5 × 19) + (3 × 2) = 101] were done in order to obtain a better viewing of clusters due to reduced number of samples.

The pattern of population structure and detection of probable introgression was visualized using a Bayesian model based clustering method implemented in STRUCTURE software version 2.2 [32]. The site of collection was assumed as the putative population origin for each individual during analysis. The admixture model with a burn-in period of 10,000 and 100,000 Markov Chain Monte Carlo (MCMC) replications was implemented for gathering data from $K = 1$ to $K = 21$ groups. Five independent replicates were run for each K value. In this way the proportion of membership of each individual in each population was estimated as described in Falush et al. [33]. Following the simulation method of Evanno et al. [34], the optimum number of clusters was estimated between $K = 1$ and $K = 21$ using the web based software STRUCTURE HARVESTER v0.6.8 [35]. The clusters across the replicates were aligned using CLUMPP software [36] and then DISTRUCT software [37] was used to display population clusters.

Results

In the present study, a total of 11 genomic-SSR loci was developed. Detailed information of these SSR- loci such as source sequence accession numbers, forward and reverse primers, repeat motifs, expected fragment size and observed fragment size range was described in Table 1. All of these SSR loci have dinucleotide repeat motifs and were amplified in all populations of korarima. At loci *Afco_8* and *Afco_14*, only one allele was detected within cultivated populations (across the 132 individuals) and within wild populations (across 63 individuals), respectively, whereas the remaining nine loci were polymorphic within both groups of populations. Therefore, only the 10 polymorphic loci within each group were used to

compute genetic diversity parameters when cultivated and wild populations were analyzed separately.

The sequences of amplified products matched the original *A. oxyphylla* sequences used for designing the SSR primer-pairs except in the case of *Afco_2*. The sequencing of the PCR product amplified by *Afco_2* failed. All sequences that matched the original *A. oxyphylla* sequences contained the target microsatellites except in the case of *Afco_8* which had poor sequence quality in the microsatellite containing region and hence trimmed off (Fig 2).

Loci like *Afco_6*, *Afco_8*, *Afco_11*, *Afco_15*, *Afco_19* and *Afco_22* produced private allele(s) that are unique to a single population. In the case of the cultivated populations, only one population (Mizan-Teferi_C1) produced a private allele, which was at locus *Afco_11* (allele frequency = 6.3%), whereas private alleles were detected in three wild populations (Jimma-Gera_W, Metu_W and Mizan-Teferi_W). The highest number of private alleles (9) was detected in the Mizan-Teferi_W population across six loci but with a low frequency ranging from 6.3% to 33.3% (Fig. 3).

In total across the 10 loci, 32 alleles were detected in the 14 cultivated populations, whereas 49 alleles were detected in the seven wild populations. The observed number of alleles (n_a) per locus was from two to five in cultivated populations and from four to eight in wild populations. The effective number of alleles (n_e) per locus varied from 1.02 (*Afco_19*) to 2.39 (*Afco_11*) for cultivated populations and from 1.42 (*Afco_8*) to 4.87 (*Afco_2*) for wild populations. The Shannon diversity index (I) per locus ranged from 0.06 (*Afco_19*) to 1.02 (*Afco_2* and *Afco_11*) for cultivated populations and from 0.63 (*Afco_8*) to 1.72 (*Afco_2*) for wild populations. The average Shannon diversity index for all loci was 0.76 and 1.06 for cultivated and wild populations, respectively. Observed heterozygosity (H_o) ranged from zero at locus *Afco_19* to 0.88 at locus *Afco_11* in cultivated populations and from zero at locus *Afco_19* to 0.86 at locus *Afco_11* in wild populations. The average observed heterozygosity across all loci was 0.63 and 0.48 for cultivated and wild populations, respectively. Expected heterozygosity (H_e) ranged from 0.02 at locus *Afco_19* to 0.58 at locus *Afco_11* in cultivated populations and from 0.30 at locus *Afco_8* to 0.80 at locus *Afco_2* in wild populations. The average expected heterozygosity for all loci was 0.48 and 0.57 for cultivated and wild populations, respectively (Table 2). The estimate of gene flow (N_m) based on F_{ST} across all populations was 0.94 (Table 2).

Percentage of polymorphic loci (PPL), observed and expected heterozygosity, Shannon diversity index was determined for each population. The maximum PPL per population was 100%, which was observed in the Mizan-Teferi_W from the Bench-Maji Zone, while the lowest

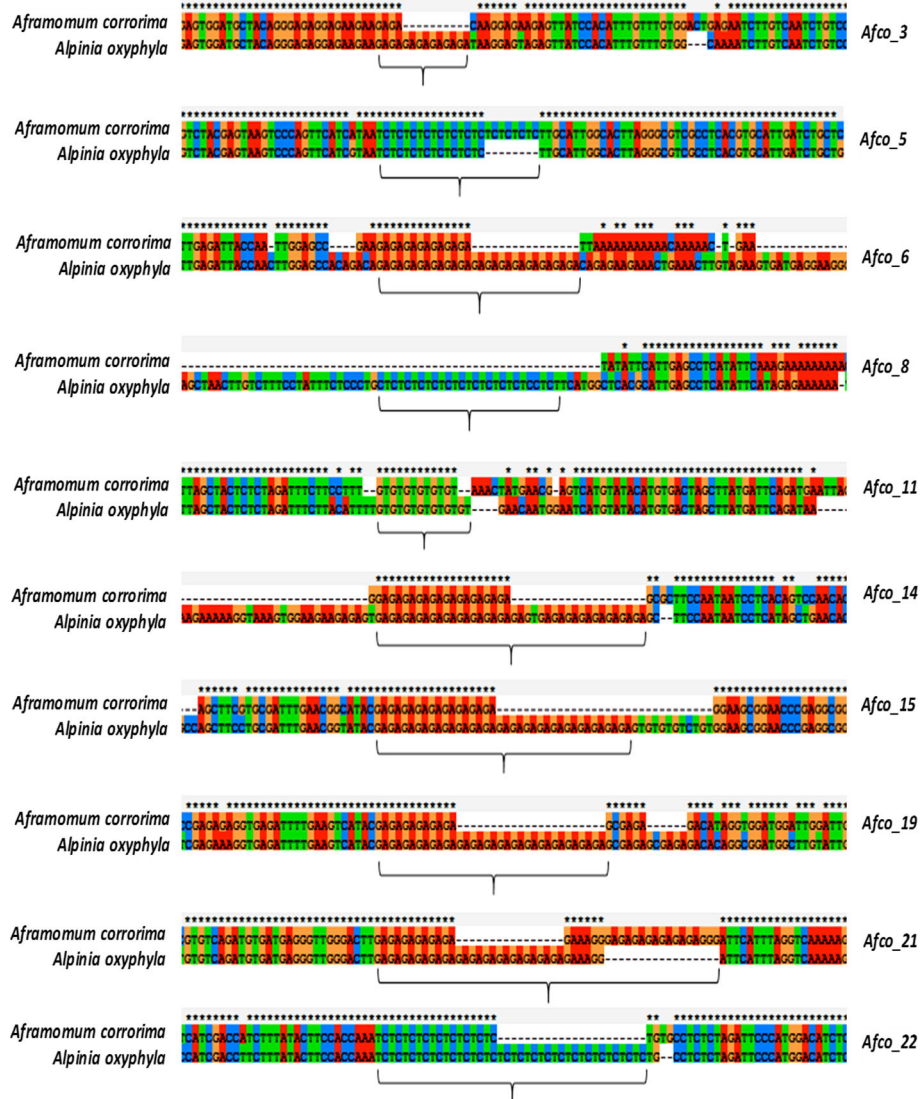


Fig. 2 Short segments of the alignment of microsatellite containing sequences of *Aframomum corrorima* and *Alpinia oxyphylla* species for 10 of the 11 SSR loci used in the present study. The locus names are given to the right of each aligned sequences. The SSR region for each locus is shown by curly bracket. Note: sequencing of *Afco_2* failed and hence is not included here

was 60% for Masha_W (Sheka zone), both of which are wild populations. In the cultivated populations, the highest PPL was 90% which was recorded in seven of the 14 populations (Gore_C1, Jimma_C1, Mizan-Teferi_C1, Masha_C2, Tepi_C1, Tepi_C2 and Bonga_C1) and the least was 70% for the Masha_C1 population (Table 3).

Among the wild populations, the highest Shannon diversity index was recorded for Mizan-Teferi_W ($I = 1.03$) and the lowest for Masha_W ($I = 0.52$), whereas in the cultivated populations, the highest was for Masha_C2 ($I = 0.82$) and the lowest was for Masha_C1 ($I = 0.52$). The highest and lowest observed heterozygosity was scored for Bonga_C1 ($H_o = 0.78$) and Tepi_C2 ($H_o = 0.50$) in the cultivated populations, and Metu_W ($H_o = 0.73$)

and Masha_W ($H_o = 0.30$) in the wild populations. The expected heterozygosity in Masha_C2 ($H_e = 0.56$) was the highest and the lowest was 0.40 (for Jimma_C3) among the cultivated populations. Among the wild populations, the highest and the lowest expected heterozygosity was recorded in Mizan-Teferi_W ($H_e = 0.64$) and Masha_W ($H_e = 0.38$), respectively (Table 3).

Two complementary approaches were used to explore the genetic diversity structure: analysis of molecular variance (AMOVA) and pair-wise fixation index (F_{ST}). In the case of AMOVA, the total genetic variation was partitioned into three hierarchical levels: among groups (cultivated vs wild), within groups among populations and within populations. Pair-wise F_{ST} was estimated for

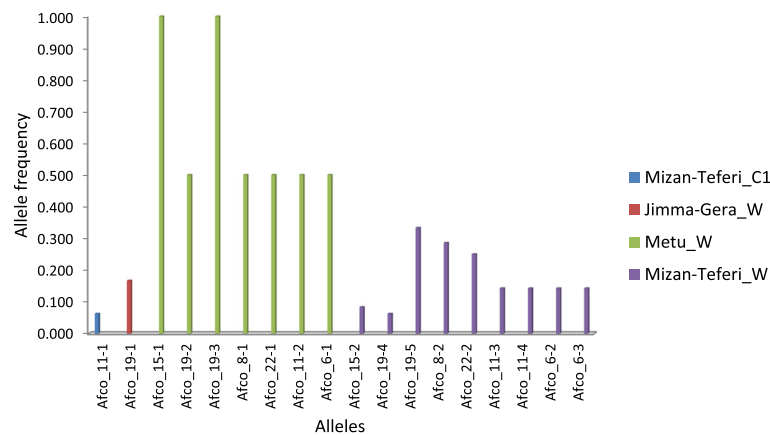


Fig. 3 List of private alleles unique to one of the four populations shown on the right and its frequencies at six SSR loci

all populations as well as for cultivated and wild populations, separately.

AMOVA revealed highly significant ($P < 0.001$) but low level of genetic differentiation ($F_{ST} = 0.04$) between the wild and cultivated populations. Of the total genetic variation, 5.7% was attributable to the variation between cultivated and wild korarima. The analysis revealed absence of differentiation among the cultivated populations as well as wild populations (Table 4).

Both cultivated and wild populations were further divided into smaller groups separately and AMOVA analysis was conducted. First, both cultivated and wild populations were grouped into two: Oromia and SNNP based on their region of origin. Then, the populations were grouped into five groups based on their zones of origin. In cultivated korarima, no genetic differentiation was obtained among populations, regions and zones ($F_{ST} = -0.05$; $P < 0.001$). However, in the case of wild populations significant differentiation was obtained at population ($F_{ST} = 0.07$; $P < 0.001$), region ($F_{ST} = 0.06$; $P < 0.001$) and zone ($F_{ST} = 0.09$; $P < 0.001$) levels (Table 5).

The Nei's standard genetic distance between cultivated and wild korarima populations ranged from 0.06 (T_C1 vs G_W) to 1.56 (B_C1 vs M_W). The genetic distance within cultivated populations ranged from 0.01 (MT_C2 vs M_C1) to 0.19 (B_C3 vs M_C1). The genetic distance within wild populations ranged from 0.04 (Ma_W vs B_W and Ma_W vs JG_W) to 1.64 (M_W vs Ma_W) (Table 6). The Metu wild population (M_W) is distantly related to the rest of the wild populations and cultivated korarima with an observed genetic distance of greater than 1.30 (Table 6).

The comparison of the extent of genetic differentiation within and between cultivated and wild korarima populations (pair-wise F_{ST}) is presented in Table 6. Among cultivated korarima, the most differentiated population is J_C1 which is significantly differentiated from six

cultivated populations (G_C1, MT_C1, Ma_C2, B_C1, B_C3 and M_C1). Populations B_C1 and B_C3 were significantly differentiated from two cultivated populations (M_C1 and J_C1). Among wild korarima, the most differentiated population was Metu wild (M_W) population. Pair-wise F_{ST} showed that M_W was significantly differentiated from all other wild populations as well as from all cultivated populations.

Analysis using STRUCTURE HARVESTER revealed that the maximum peak is at $K = 2$ (Additional file 3), suggesting that two populations can integrate all individuals included in the present study with the highest likelihood. A graphic representation of estimated membership coefficients of each individual at $K = 2$ was shown in Fig. 4a and b. Each color represents the proportion of membership of each individual, represented by a vertical line, to the two clusters. Bayesian model-based cluster analysis at $K = 2$ did not produce distinct differentiation among cultivated and wild individuals (Fig. 4a). In this analysis, we can observe some individuals from both cultivated and wild population jointly assigned to two clusters, following the assumption that an individual was only exclusively assigned to a particular genetic cluster if at least 85% of its genome (i.e. $q_i \geq 0.85$) is found in it. In Fig. 4b there is no distinct structure among populations which shows an admixture of populations.

Discussion

Genomic-SSR markers

Molecular markers have been widely used in studying genetic diversity and relationships, screening candidate genes associated with the target traits and marker assisted breeding. Microsatellite markers derived from genomic DNA regions have been chosen in many studies for their high levels of allelic variability compared to microsatellites derived from expressed sequence tags (EST-SSRs) [38]. Primer pairs intended on the base of

Table 2 Estimates of different genetic diversity parameters and gene flow for *A. corrorima* at each polymorphic microsatellite loci based on (A) all populations, (B) only cultivated populations, and (C) only wild populations

Locus	Sample size	na	ne	I	Ho	He	Av.Het	Nm
(A) All Populations								
Afco_2	270	8	3.16	1.38	0.66	0.69	0.53	0.87
Afco_3	322	5	2.42	1.07	0.59	0.59	0.50	1.60
Afco-5	284	5	2.58	1.10	0.66	0.61	0.55	1.50
Afco-6	272	5	1.94	0.78	0.71	0.49	0.43	2.10
Afco-8	284	4	1.09	0.22	0.02	0.08	0.06	0.19
Afco-11	300	8	2.45	1.10	0.87	0.59	0.55	3.02
Afco-14	58	2	1.62	0.57	0.10	0.39	0.14	0.08
Afco-15	274	4	2.09	0.80	0.84	0.52	0.46	1.45
Afco-19	228	4	1.13	0.30	0.00	0.12	0.05	0.13
Afco-21	296	4	2.64	1.06	0.75	0.62	0.54	1.32
Afco-22	260	4	1.91	0.75	0.68	0.48	0.40	1.25
Mean	259	4.82	2.09	0.83	0.54	0.47	0.38	0.94
St. Dev		1.78	0.64	0.36	0.33	0.20	0.20	
(B) Cultivated Populations								
Afco-2	188	5	2.28	1.02	0.68	0.56	0.47	1.91
Afco-3	224	5	2.27	1.00	0.71	0.56	0.52	3.73
Afco-5	196	4	2.17	0.88	0.71	0.54	0.51	3.20
Afco-6	208	2	1.91	0.67	0.76	0.48	0.47	9.48
Afco-11	228	5	2.39	1.02	0.88	0.58	0.55	6.54
Afco-14	38	2	1.92	0.67	0.16	0.49	0.21	0.14
Afco-15	208	2	1.99	0.69	0.87	0.50	0.49	11.36
Afco-19	170	2	1.02	0.06	0.00	0.02	0.02	1.62
Afco-21	218	3	2.33	0.92	0.79	0.57	0.53	3.80
Afco-22	210	2	1.87	0.66	0.73	0.47	0.46	9.63
Mean	189	3.2	2.01	0.76	0.63	0.48	0.42	1.94
St. Dev		1.40	0.39	0.29	0.29	0.17	0.17	
(C) Wild Populations								
Afco-2	82	8	4.87	1.72	0.61	0.80	0.67	1.21
Afco-3	98	4	2.24	0.95	0.31	0.56	0.47	1.68
Afco-5	88	5	3.40	1.34	0.55	0.71	0.62	1.28
Afco-6	64	5	1.93	0.94	0.56	0.49	0.36	1.18
Afco-8	72	4	1.42	0.63	0.08	0.30	0.18	0.23
Afco-11	72	7	2.64	1.21	0.86	0.63	0.54	1.53
Afco-15	66	4	2.40	0.99	0.76	0.59	0.42	0.54
Afco-19	58	4	1.55	0.72	0.00	0.36	0.13	0.12
Afco-21	78	4	3.29	1.24	0.64	0.71	0.57	0.97
Afco-22	50	4	1.96	0.92	0.44	0.50	0.29	0.38
Mean	73	4.90	2.57	1.06	0.48	0.57	0.42	0.74
St. Dev		1.44	1.04	0.32	0.28	0.15	0.18	

Note: *Afco-8* and *Afco-14* were monomorphic within cultivated and wild populations, respectively, and hence were not included when the data for the cultivated and wild populations were analyzed separately. *na* observed number of alleles, *ne* effective number of alleles, *I* Shannon information index, *Ho* observed heterozygosity, *He* expected heterozygosity, *Av. He* average heterozygosity, *Nm* gene flow estimated from $F_{st} = 0.25(1 - F_{st})/F_{st}$

Table 3 Percentage of polymorphic loci (PPL), Shannon's diversity index (I), observed heterozygosity (Ho) and expected heterozygosity (He) for cultivated and wild populations of *Aframomum corrorima*

Zone	Population	PPL	I	Ho	He	Sample size
Cultivated Populations						
Illubabour	Gore_C1	90	0.75	0.58	0.54	10
	Metu_C1	80	0.71	0.57	0.47	10
	Average	85	0.73	0.58	0.51	
Jimma	Jimma_C1	90	0.56	0.54	0.43	10
	Jimma_C2	80	0.61	0.69	0.45	10
	Jimma_C3	80	0.56	0.55	0.40	10
	Average	83.33	0.58	0.59	0.43	
Bench-Maji	Mizan-Teferi_C1	90	0.66	0.73	0.49	9
	Mizan-Teferi_C2	80	0.63	0.55	0.43	10
	Average	85	0.65	0.64	0.46	
Sheka	Masha_C1	70	0.52	0.72	0.50	3
	Masha_C2	90	0.82	0.62	0.56	10
	Tepi_C1	90	0.68	0.60	0.47	10
	Tepi_C2	90	0.69	0.50	0.48	10
	Average	85	0.68	0.61	0.50	
Kefa	Bonga_C1	90	0.66	0.78	0.53	10
	Bonga_C2	80	0.66	0.65	0.47	10
	Bonga_C3	80	0.58	0.74	0.43	10
	Average	83.33	0.68	0.72	0.48	
Wild Populations						
Illubabour	Metu_W	80	0.56	0.73	0.48	3
	Gumero_W	80	0.65	0.51	0.43	10
	Average	80	0.61	0.62	0.46	
Jimma	Jimma-Gera_W	90	0.73	0.38	0.50	10
	Jimma_W	80	0.56	0.42	0.43	10
	Average	85	0.65	0.40	0.47	
Bench-Maji	Mizan-Teferi_W	100	1.03	0.52	0.64	10
Sheka	Masha_W	60	0.52	0.30	0.38	10
Kefa	Bonga_W	80	0.71	0.45	0.46	10

sequence of one species possibly will be used to develop SSR markers for other related species [39]. We used publicly available microsatellite containing genomic sequences of *Alpinia oxyphylla* to develop 11 new SSRs for *Aframomum corrorima*. All of these microsatellites are dinucleotide repeats, which are the most common types of SSRs developed based on genomic DNA sequences [9]. All of these loci were polymorphic when the whole populations studied were considered. However, the number of polymorphic loci was reduced by one when the loci were considered either for cultivated

Table 4 Analysis of molecular variance (AMOVA) for the cultivated and wild korarima populations

Source of variation	df.	Sum of squares	Variance components	%age of variation	P-value
AG	1	13.555	0.08193 V_a	5.74	Vc and $F_{ST} = 0.000$
APWG	19	16.850	-0.02892 V_b	-2.03	Vb and $F_{SC} = 0.069$
WP	335	460.267	1.37393 V_c	96.29	Va and $F_{CT} = 0.000$
Total	355	490.671	1.42694		
Fixation Index	$F_{ST} = 0.04$				

AG among groups of populations (cultivated vs wild), APWG among populations within groups (cultivated and wild), WP within populations

or wild populations. The high rate of marker transferability obtained in the present study suggest that the flanking regions of these loci are sufficiently conserved, and allows comparative analyses of genetic diversity in the genus *Aframomum* and related genera using these markers. The sequence alignment analysis showed that there was variation in the repeat length of the dinucleotide motifs between the two species *A. corrorima* and *Alpinia oxyphylla* (Fig. 2), which suggests a clear differentiation of these two species at these SSR loci. However, the sequences that cover the SSR containing part of the PCR products were not obtained for *Afco_2* and *Afco_8*, as sequencing failed for *Afco_2*, and the SSR containing part of *Afco_8* had poor sequence quality, which is attributed to the fact that the SSR is located close to the left primer annealing site.

Genetic diversity

Nearly half of the loci developed for the diversity study amplified alleles exclusive to a single population (private alleles). The mainly self-pollinating nature of korarima [4] could be one of the main reasons behind the recognition of private alleles in the majority of the loci studied smoothing the possibility of new mutations to remain in the original population due to reduced gene-flow among populations. Among cultivated populations, only one (MizanTeferi_C1) produced a private allele (frequency = 6.25%; Fig. 3). Differences in agricultural practices in order to strengthen their adaptation to local agro-ecological conditions may have contributed to the genetic distinctiveness of this population. Most of the private alleles were found in three wild populations: Jimma Gera_W, Metu_W and Mizan-Teferi_W having 1, 7 and 9 private alleles, respectively (Fig. 3). The

Table 5 F_{ST} based estimates of genetic differentiation of cultivated and wild korarima populations at various levels

Populations	Differentiation at	Mean F_{ST}	P-value
Cultivated	Population level	-0.05	0.000
	Region level	-0.05	0.000
	Zone level	-0.05	0.000
Wild	Population level	0.07	0.000
	Region level	0.06	0.000
	Zone level	0.09	0.000

presences of these private alleles have important implications because they may be in linkage disequilibrium with genes coding for desirable traits. Thus they could be useful in indicating a particular type of genotype for breeding purposes [40].

When genetic analysis was done for all populations, a total of 53 alleles, which averaged to 4.82 alleles per locus, were detected across 11 loci. The average number of alleles per locus observed in the current study is analogous to a 4.3 alleles/locus that was reported for ginger (*Zingiber officinale*) accessions genotyped with eight SSR markers [41]. However, it is lower than 5.7 alleles/locus reported for *Curcuma longa* [42] and 5.8 alleles/locus for *Piper nigrum* [43] based on data from eighteen and nine microsatellite loci, respectively. The relatively lower number observed in *A. corrorima* and *Z. officinale* could be attributed to their diploid chromosome number, as *C. longa* and *P. nigrum* are triploid and tetraploid, respectively. In the present study, *Afco-2* and *Afco-11* were the most polymorphic loci both containing eight alleles. Comparable number of alleles per locus was reported in *Z. officinale* (seven alleles; [41]), *C. longa* (nine alleles; [42]) and *P. nigrum* (ten alleles; [43]). Based on 23 polymorphic microsatellites, Zou et al. [44] reported up to 13 alleles per locus in *Alpinia oxyphylla* which is higher than that obtained in the present study.

The average observed heterozygosity (H_O) obtained in this study was 0.54 (Table 2). This result is lower than that reported in *P. nigrum* (0.62, [43]) and in *A. oxyphylla*, (0.64, [44]). Likewise, the expected heterozygosity (H_e) reported in this study was lower than that reported in *P. nigrum* (0.72; [43]) and in *A. oxyphylla* (0.62; [44]). The relatively lower genetic diversity observed in this study can be partly explained by combinations of fewer populations and loci studied.

Our study showed higher levels of diversity in wild populations compared to cultivated populations in terms of average number of alleles/locus, Shannon diversity index and expected heterozygosity. The average number of alleles/locus and Shannon diversity index of wild korarima were higher than that of cultivated korarima, 4.9 vs. 3.2 and 1.06 vs 0.76, respectively. The expected heterozygosity (H_e) showed the same trend with 0.57 and 0.48 for wild and cultivated

Table 6 Nei's standard genetic distance (above diagonal) and pair wise F_{ST} significance (below diagonal) between the 21 korarima populations

	G_C1	MT_C1	MT_C2	Ma_C1	Ma_C2	B_C1	B_C2	B_C3	M_C1	T_C1	T_C2	J_C1	J_C2	J_C3	B_W	JG_W	G_W	J_W	Ma_W	M_W	MT_W
G_C1	****	0.06	0.11	0.15	0.03	0.04	0.10	0.09	0.11	0.09	0.07	0.09	0.11	0.08	0.10	0.17	0.13	0.18	0.16	1.52	0.14
MT_C1	-	****	0.08	0.11	0.03	0.04	0.07	0.07	0.08	0.06	0.04	0.04	0.08	0.07	0.12	0.18	0.12	0.16	0.18	1.44	0.12
MT_C2	-	-	****	0.11	0.08	0.06	0.04	0.18	0.01	0.03	0.10	0.09	0.11	0.03	0.10	0.19	0.10	0.16	0.18	1.46	0.18
Ma_C1	-	-	-	****	0.12	0.10	0.13	0.13	0.12	0.13	0.08	0.06	0.02	0.14	0.27	0.36	0.22	0.27	0.35	1.36	0.14
Ma_C2	-	-	-	-	****	0.02	0.06	0.06	0.07	0.07	0.05	0.07	0.09	0.06	0.10	0.18	0.11	0.17	0.17	1.45	0.15
B_C1	-	-	-	-	-	****	0.05	0.04	0.07	0.06	0.05	0.06	0.08	0.06	0.14	0.22	0.14	0.23	0.22	1.56	0.14
B_C2	-	-	-	-	-	-	****	0.18	0.04	0.02	0.11	0.09	0.12	0.03	0.10	0.16	0.07	0.13	0.16	1.48	0.20
B_C3	-	-	-	-	-	-	-	****	0.19	0.18	0.06	0.09	0.11	0.17	0.28	0.39	0.28	0.36	0.37	1.48	0.18
M_C1	-	-	-	-	-	-	-	-	****	0.03	0.11	0.10	0.12	0.05	0.08	0.17	0.09	0.12	0.16	1.53	0.18
T_C1	-	-	-	-	-	-	-	-	-	****	0.09	0.08	0.10	0.02	0.08	0.12	0.06	0.14	0.14	1.40	0.17
T_C2	-	-	-	-	-	-	-	-	-	-	****	0.04	0.08	0.09	0.16	0.21	0.15	0.19	0.19	1.50	0.13
J_C1	+	-	+	-	+	+	-	+	+	-	-	****	0.05	0.08	0.19	0.25	0.16	0.21	0.26	1.51	0.17
J_C2	-	-	-	-	-	-	-	-	-	-	-	-	****	0.11	0.21	0.29	0.18	0.25	0.30	1.34	0.11
J_C3	-	-	-	-	-	-	-	-	-	-	-	-	-	****	0.08	0.15	0.1	0.15	0.14	1.55	0.17
B_W	-	-	+	-	-	+	-	+	-	-	-	+	-	+	****	0.06	0.07	0.09	0.04	1.57	0.19
JG_W	-	-	+	-	-	+	+	+	-	-	+	-	+	+	-	****	0.08	0.12	0.04	1.47	0.23
G_W	-	+	+	-	-	+	-	+	-	-	+	-	+	+	-	-	****	0.12	0.10	1.31	0.22
J_W	-	-	+	-	-	+	+	+	-	-	+	-	+	+	-	-	-	****	0.07	1.50	0.29
Ma_W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	****	1.64	0.22
M_W	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	****	1.37
MT_W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	****

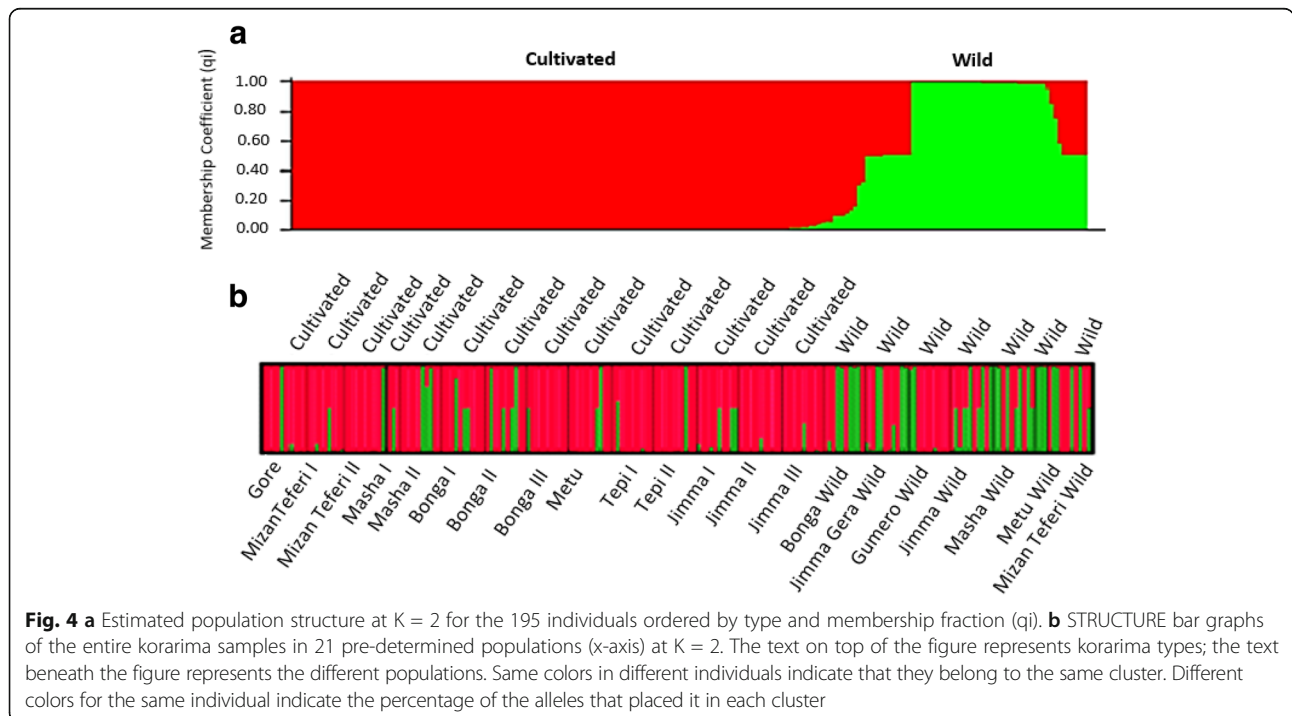
+: Significant differentiation among the joint populations ($P < 0.05$)
 -: No significant differentiation among the joint populations ($P > 0.05$)

Area shaded with gray color represents pair-wise comparison b/n cultivated population; the one shaded with red color represents pair-wise comparison b/n wild populations. The unshaded region is b/n cultivated and wild population. G_C1 = Gore_C1; MT_C1 = Mizan-Teferi_C1; MT_C2 = Mizan-Teferi_C2; Ma_C1 = Masha_C1; Ma_C2 = Masha_C2; B_C1 = Bonga_C1; B_C2 = Bonga_C2; B_C3 = Bonga_C3; M_C1 = Metu_C1; T_C1 = Tepi_C1; T_C2 = Tepi_C2; J_C1 = Jimma_C1; J_C2 = Jimma_C2; J_C3 = Jimma_C3; B_W = Bonga_W; Jg_W = Jimma-Gera_W; G_W = Gumero_W; J_W = Jimma_W; Ma_W = Masha_W; M_W = Metu_W; MT_W = Mizan-Teferi_W

korarima, respectively (Table 3). The high genetic diversity recorded in the wild populations may suggest the possibility of finding novel alleles contributing to the improvement of cultivated korarima and has to be given special attention for conservation and utilization.

Among wild populations, the highest Shannon diversity index, the highest percentage of polymorphic loci and the highest expected heterozygosity was recorded in

Mizan-Teferi_W, a population from the Bench-Maji zone and lowest in Masha_W, a population from the Sheka Zone. The genetic diversity of most cultivated populations is relatively high with all of the populations having a percentage of polymorphic loci higher than or equal to 70% and a Shannon diversity index of more than 0.50. High genetic diversity was observed in Masha_C2 population (Sheka zone) possessing high



percentage of polymorphic loci and high Shannon diversity index compared to other cultivated populations. This may indicate that farmers in this area grow a mixture of landraces of korarima or have their own unique culture of conservation. The lowest diversity found in the other two Masha populations (Masha_C1 and Masha_W) in the Sheka zone, which are cultivated and wild in that order, was rather unexpected. The decreased genetic diversity in cultivated Masha_C1 could be due to strong genetic drift which results in a rapid decline of population size or sampling error during individual collection. However, the reduced diversity observed in Masha_W is most likely due to human interference such as digging the rhizomes for planting and other traditional practices. These will result a decline in population size. There is a real risk that if the trend towards a decline in population size is left unchecked, genetic consequences associated with small isolated populations, including genetic drift and inbreeding, will reduce most of the genetic diversity in the Masha wild population. Thus, there is an urgent need to assess the distribution and size of *A. corrorima* populations in this zone and to adopt appropriate *insitu* conservation strategies and collect germplasm for *exsitu* conservation.

Genetic structure and gene flow

Our analysis did not show distinct divergence between wild and cultivated korarima. Based on the AMOVA analyses, we found a significant but a relatively low level genetic differentiation among populations of cultivated and wild korarima ($F_{ST} = 0.04$; $P < 0.001$). Pair-wise F_{ST} also showed the presence of low level of crop-wild genetic divergence with each wild population showing a significant differentiation from only less than half of the cultivated population. Several factors can be included in attempts to account for the observed genetic structure in cultivated and wild korarima. The present-day cultivated *A. corrorima* may have been directly introduced to cultivation from wild progenitors inhabiting the same localities and thus may have maintained most of their initial genetic diversity. In Southwest Ethiopia, the cultivation area of *A. corrorima* overlaps with the distribution of natural habitats of wild populations. Because of the sympatric distribution pattern of the wild and cultivated populations, migration from wild to cultivated populations by natural or artificial means may be an ongoing process. Therefore, subsequent gene flow from wild ancestors into cultivated populations ($Nm = 0.94$; Table 2) may also contribute to the maintenance of a relatively high genetic diversity in cultivated *A. corrorima*.

Crop cultivation impacts not only the amount of genetic variation contained in cultivated populations but also the structure of this variation [45]. Hamrick and Godt [46] reported that the mean value of genetic differentiation

among populations of crop species is higher than that of non-crop species. However, our results run contrary to this report. The cultivated (crop) *A. corrorima* populations did not show genetic differentiation among populations ($F_{ST} = -0.05$; $P < 0.001$) whereas wild (non-crop) populations showed significant population differentiation ($F_{ST} = 0.07$, $P < 0.001$). The absence of clear differentiation among cultivated populations can be explained by the movement of genetic material between growers within and among the major production areas, which result in a higher degree of gene flow among cultivated populations ($Nm = 1.94$, Table 2) and diminish their genetic differences. For wild populations, dispersal of seeds and natural migration between populations is limited, especially when habitats are isolated by geographic barriers such as mountains, rivers or distance. As a result, wild populations are expected to display a relatively greater proportion of variance among populations.

The main evolutionary forces that shape genetic structure in plant populations are gene flow, selection associated with environmental heterogeneity and/or farmer preferences and random genetic drift [47]. For cultivated korarima, there is no differentiation among populations both at the level of regions and zones. The absence of genetic divergence among regions and zones could be attributed to geographical proximity of the regions and zones where korarima is currently cultivated, resulting in high gene flow in the form of exchanges of germplasm among farmers. The overall population differentiation among regions and among administrative zones for wild korarima was significant ($P < 0.0001$) with F_{ST} value of 0.06 and 0.09 respectively (Table 5). Natural barriers and geographic distance between regions restricts the level of gene flow among wild populations and should therefore contribute significantly to the contemporary genetic structure. In addition, the moderate level of differentiation among administrative zones is expected because different altitudes within a geographical zone would provide different degrees of selection pressure for adaptation and could accelerate population differentiation rather than gene flow [48].

STRUCTURE analysis following Evanno's *ad hoc* ΔK method suggested that the 21 populations included in the present study most likely form two clusters ($K = 2$) (Additional file 3). Similar to the result obtained from AMOVA and pair-wise F_{ST} analysis, divergence between the two types of korarima was not clear at $K = 2$. The mixed clustering pattern of individuals of the cultivated and wild populations was observed in the population structure constructed based on STRUCTURE software. All individual genotypes in each population showed partial membership to both clusters (Fig. 4a and b). In addition, the NJ tree revealed two clusters, each comprising individuals from cultivated and wild populations

(Additional file 4). Moreover, the UPGMA dendrogram, based on pair-wise simple matching dissimilarity index among the 21 cultivated and wild populations, showed two clusters with two wild and one cultivated populations as outliers (Additional file 5). Isolation from cultivars is generally considered to be beneficial from the standpoint of preserving the integrity and diversity of wild gene pools. The fact that two wild populations grouped very closely with cultivars suggests that there may be ongoing gene flow between them via hybridization and/or cultivars becoming naturalized as free-living populations. Therefore, the overall result obtained in this study suggests that there is gene flow among the two korarima types (cultivated and wild) due to their relative geographical proximity to each other.

Conclusions

This study provides a first detailed analysis and estimation of genetic diversity in Ethiopian populations of korarima using 11 genomic SSR markers newly developed based on genomic resources of *Alpinia oxyphylla*. These SSR markers were highly polymorphic within Ethiopian *Aframomum corrorima* used in this study and could be employed to examine genetic diversity and population structure of korarima in other countries as well as other *Aframomum* species. Some of the alleles identified across the 11 SSR loci may be linked to important agronomic traits and have to be further evaluated through their use in korarima mapping populations. Overall, Ethiopian korarima needs broader characterization at both molecular and agro-morphological levels in order to use it in future korarima improvement. The data also shows the power of SSR markers to define how korarima genetic diversity is structured in Ethiopia. This can lead to a better management of the populations in-situ and ex-situ and promote the use of these populations in further breeding programs. Among the five zones in southwestern Ethiopia considered in this study, both Bench-Maji and Sheka harbor populations with high genetic variation and could be used as a starting in-situ conservation sites. We expect that more extensive sampling of korarima in other regions of Ethiopia followed by molecular characterization would reveal additional target sites for in-situ conservation, for development of core collection for long-term ex-situ conservation, as well as possible evidence of crop-wild gene flow. The present finding is a timely contribution, considering korarima's high economic and endogenous value, and current challenges in korarima biodiversity management and conservation.

Additional files

Additional file 1: Passport data of *A. corrorima* samples collected from different sites in Ethiopia. (DOCX 17 kb)

Additional file 2: Characteristics of 23 microsatellite primers developed for *Aframomum corrorima* (The bold one are used for genetic diversity analysis) (DOCX 16 kb)

Additional file 3: A biplot detected the maximum peak at $K = 2$ (the optimum number of clusters) based on Evanno et al. 2005 [35] prediction. (PDF 10 kb)

Additional file 4: Neighbor-joining tree based on pair-wise simple matching dissimilarity index showing the genetic relationships among 101 individuals randomly re sampled from cultivated and wild korarima populations (19 populations were represented by five samples each and two populations were represented by three samples each). Blue color indicates individuals from cultivated populations whereas red represents individuals from wild populations. (DOCX 22 kb)

Additional file 5: UPGMA dendrogram based on pair wise simple matching dissimilarity index showing the genetic relationships among Ethiopian cultivated and wild korarima populations using SSR markers. The analysis resulted in two major clusters shown as "I" and "II". (DOCX 27 kb)

Additional file 6: Allele frequency data for all korarima populations across the 11 SSR loci. (XLSX 18 kb)

Abbreviations

AMOVA: Analysis of molecular variance; CTAB: Cetyl trimethyl ammonium bromide; EST: Expressed sequence tags; IBC: Institute of Biodiversity Conservation; NCBI: National Center for Biotechnology information; SNNP: Southern Nations, Nationalities, and Peoples' Region; SSR: Simple sequence repeats

Acknowledgments

The first author would like to thank Addis Ababa University, the Swedish International Development Agency (Sida) and Jimma University for financial, material and technical supports to this research as a component of her PhD thesis. The authors would like to extend heartfelt gratitude to Ethiopian Institute of Biodiversity conservation (IBC) for allowing plant material transfer to Swedish University of agricultural Science, Alnarp, Sweden. We would like to thank Melese Terefe for his assistance in the map construction presented as Fig. 1.

Funding

Not applicable.

Availability of data and materials

Allele frequency data for all korarima populations across the 11 SSR loci (Additional file 6).

Authors' contributions

DC conducted the laboratory experiments. DC and MG designed the SSR primers and carried out the data analysis. DC, MG, AT, EB and TB participated in experimental design, data interpretation and drafting the manuscript. All authors revised the draft and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, P. O. Box, 1176 Addis Ababa, Ethiopia. ²Department of Plant Breeding, Swedish University of Agricultural Sciences, P.O. Box 101, -23053 Alnarp, SE, Sweden.

Received: 16 January 2017 Accepted: 25 July 2017

Published online: 01 August 2017

References

1. PCM J. Spices, condiments and medicinal plants in Ethiopia. Their taxonomy and agricultural significance. Agricultural research reports 906. Wageningen: Center for agricultural publishing and documentation; 1981. p. p327.
2. Tefera W, Wannakrairoj SA. Micro propagation method for korarima (*Aframomum corrorima* (Braun) Jansen). *Sci Asia*. 2004;30:1–7.
3. Eyob S, Tsegaye A, Appelgren EM. Analysis of korarima (*Aframomum corrorima* (Braun) P.C.M. Jansen) indigenous production practices and farm based biodiversity in southern Ethiopia. *Genet Resour Crop Evol*. 2009;56:573–85.
4. Jansen PCM. *Aframomum corrorima* (Braun) P.C.M. Jansen. In: Oyen LPA, Lemmens RHMJ, editors. PROTA (Plant Resources of Tropical Africa). Wageningen; 2002. [https://www.prota4u.org/database/protav8.asp?g=pe&p=Aframomum+corrorima+\(A.Braun\)+P.C.M.Jansen](https://www.prota4u.org/database/protav8.asp?g=pe&p=Aframomum+corrorima+(A.Braun)+P.C.M.Jansen). Accessed 23 Feb 2016.
5. Eyob S, Martinsen BK, Tsegaye A, Appelgren M, Skrede G. Antioxidant and antimicrobial activities of extract and essential oil of korarima (*Aframomum corrorima* (Braun) P.C.M. Jansen). *Afr J Biotech*. 2008;7(15):2585–92.
6. Escoubas P, Lajide L, Mizutani J. Termite antifedent activity in *Aframomum melegueta*. *Pytochemistry*. 1995;40:1097–9.
7. Van Beuningen LT, Busch RH. Genetic diversity among North American spring wheat cultivars: III. Cluster analysis based on quantitative morphological traits. *Crop Sci*. 1997;37:981–8.
8. Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, et al. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. *Theor Appl Genet*. 2005;111:23–30.
9. Goldstein DB, Pollock DD. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J Hered*. 1997;88:335–42.
10. Zheng Y, Zhang G, Lin F, Wang Z, Jin G, Yang L, et al. Development of microsatellite markers and construction of genetic map in rice blast pathogen *Magna porthegrisea*. *Fungal Genet Biol*. 2008;45:1340–7.
11. Teshome A, Bryngelsson T, Dagne K, Geleta M. Assessment of genetic diversity in Ethiopian field pea (*Pisum sativum* L.) accessions with newly developed EST-SSR markers. *BMC Genet*. 2015;16:102.
12. Adugna A, Snow AA, Sweeney PM, Bekele E, Mutegi E. Population genetic structure of in situ wild Sorghum bicolor in its Ethiopian center of origin based on SSR markers. *Genet Resour Crop Evol*. 2013;60:1313–28.
13. Geleta M, Herrera I, Monz A, Bryngelsson T. Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. *Sci World J*. 2012;1:11.
14. Keneni G, Bekele E, Imtiaz M, Dagne K, Getu E, Assef F. Genetic diversity and population structure of Ethiopian chickpea (*Cicer arietinum* L.) germplasm accessions from different geographical origins as revealed by microsatellite markers. *Plant Mol Biol Rep*. 2012;30:654–65.
15. Biswas MK, Chai L, Mayer C, Xu Q, Guo W, Deng X. Exploiting BAC-end sequences for the mining, characterization and utility of new short sequences repeat (SSR) markers in Citrus. *Mol Biol Rep*. 2011;1–14.
16. Ashkani S, Rafii MY, Rusli I, Sarah M, Abdullah SNA, Abdul RH. SSRs for marker-assisted selection for blast resistance in rice (*Oryza sativa* L.). *Plant Mol Biol Rep*. 2012;30:79–86.
17. Hong Y, Chen X, Liang X, Liu H, Zhou G, Li S, et al. A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. *BMC Plant Biol*. 2010;10:17.
18. Zhang M, Zhang R, Yang J, Luo P. Identification of a new QTL for Fusarium head blight resistance in the wheat genotype “Wang shui-bai”. *Mol Biol Rep*. 2010;37:1031–5.
19. Lock JM. Zingiberaceae. In: Edwards S, Demissew S, Hedberg I, editors. Flora of Ethiopia and Eritrea, vol. 6. Sweden: The national herbarium, Addis Ababa University, Ethiopia and department of systematic botany, Uppsala University; 1997. p. 324–6.
20. Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W. Non-coding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms. *J Evol Biol*. 2003;16:558–76.
21. Kress WJ, Prince LM, Williams KJ. The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am J Bot*. 2002;89:1682–96.
22. Martins WS, Lucas DS, Souza Neves KF, Bertoli DJ. WebSat- A web software for microsatellite marker development. *Bioinformatics*. 2009;3:282–3.
23. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz S, editors. *Bioinformatics methods and protocols*, vol. 132. New Jersey: Humana Press; 1999. p. 365–86.
24. Ballard LW, Adams PS, Bao Y. Strategies for genotyping: effectiveness of tailing primers to increase accuracy in short tandem repeat determinations. *J Biomol Tech*. 2002;13:20–9.
25. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
26. Hall T. BioEdit v. 7.0.5. Biological sequence alignment editor for Windows. Ibis Therapeutics a division of Isis pharmaceuticals. 2005. <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>. Accessed 6 Feb 2016.
27. Yeh F, Yang R, Boyle T. Popgene version 131. 1999.
28. Crow JF, Aoki K. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proc Natl Acad Sci U S A*. 1984;81:6073–7.
29. Peakall R, Smouse PE. GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. 2012;28:2537–9.
30. Excoffier L, Laval G, Schneider S. Arlequin ver3.0: an integrated software package for population genetics data analysis. *Evol Bioinformatics Online*. 2005;1:47–50.
31. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. <http://darwin.cirad.fr/>. Accessed 1 Apr 2016.
32. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
33. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567–87.
34. Evanno S, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611–20.
35. Earl D, von Holdt B. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons Gen Res*. 2012;4:359–61.
36. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801–6.
37. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*. 2004;4:137–8.
38. Chistiakov DA, Hellemans B, Volckaert FAM. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*. 2006;255:1–29.
39. Plieske J, Struss D. Microsatellite markers for genome analysis in Brassica L. development in *Brassica napus* and abundance in Brassicaceae species. *Theor Appl Genet*. 2001;102:689–94.
40. Park Y, Dixit A, Ma K, Lee J, Lee M, Chung C, et al. Evaluation of genetic diversity and relationships within an on-farm collection of *Perilla frutescens* (L.) Britt using microsatellite markers. *Genet Resour Crop Evol*. 2008;55:523–35.
41. Lee S, Fai WK, Zakaria M, Ibrahim H, Othman RY, Gwang J, et al. Characterization of polymorphic microsatellite markers, isolated from ginger (*Zingiber officinale* Rosc.). *Mol Ecol Not*. 2007;7:1009–11.
42. Siju S, Dhanya K, Syamkumar S, Sheeja TE, Sasikumar B, Bhat AI. Development, characterization and utilization of genomic microsatellite markers in turmeric (*Curcuma longa* L.). *Biochem Syst Ecol*. 2010;38:641–6.
43. Menezes IC, Cidade FW, Souza AP, Sampaio IC. Isolation and characterization of microsatellite loci in the black pepper, *Piper nigrum* L. (piperaceae). *Conservation Genet Resour*. 2009;1:209–12.
44. Zou Y, Zou P, Liu H, Liao J. Development and characterization of microsatellite markers for *Alpinia oxyphylla* (Zingiberaceae). *Appl Plant Sci*. 2013;1:1200457.
45. Miller AJ, Schaal BA. Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). *Mol Ecol*. 2006;15:1467–80.
46. Hamrick JL, Godt MJW. Allozyme diversity in cultivated crops. *Crop Sci*. 1997;37:26–30.
47. Hartl DL, Clark G. Principles of population genetics. Sunderland: Sinauer Associates Inc; 1997. p. 565.
48. Lopez-Gartner G, Cortina H, McCouch S, Moncada M. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree Genet Genomes*. 2009;5:435–46.