Simple sequence repeat markers revealed genetic divergence and population structure of okra [*Abelmoschus esculentus*] collections of diverse geographic origin

Wassu Mohammed[1], Beyene Amelework*[2], Hussein Shimelis[3]

[1]Haramaya University, School of Plant Sciences, P.O. Box 219, Dire Dawa, Ethiopia
[2]Agricultural Research Council, Vegetable and Ornamental Plant, Private Bag X293, Pretoria, 0001, South Africa
[3]African Centre for Crop Improvement, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa

*Corresponding author: amele_g@yahoo.com; assefaa@arc.agric.za

Abstract

The study was carried out to assess the genetic diversity and population structure of okra collections from diverse geographic origin using selected polymorphic simple sequence repeat (SSR) markers. Thirty-two okra accessions collected from three geographic regions were genotyped using 16 selected SSR markers. The SSR markers generated a total of 71 loci of which 67% were polymorphic. Genetic distances among accessions ranged from 2.2 to 7.1, with a mean of 4.65. Gene diversity ranged from 0.10 to 0.78 with a mean of 0.52. Genetic clustering classified the accessions into three major clusters and four sub-clusters. Each cluster and sub-cluster consisted of accessions derived from different sources. Analysis of molecular variance revealed that 67%, 81% and 83% of the total genetic variation detected was found within populations of geographic origin, altitude and collection district, respectively. The observed moderate to high population differentiation could partly be attributed to limited germplasm exchange, agro-ecological differences, and partly by selection pressure. The present study revealed the presence of high genetic diversity and population divergence among okra collections from Ethiopia. The study demonstrated that a collection strategy for conservation of okra genetic resources should focus on sampling of representative genotypes covering wide geographic regions and altitudinal ranges of target agro-ecologies.

Keywords: Accessions, *Abelmoschus esculentus*, genetic diversity, population structure, SSR markers.
Abbreviation ACCI_African Centre for Crop Improvement; AFLP_Amplified fragment length polymorphism; AMOVA_ Analysis of molecular variance; CEF_Controlled Research Facility; CTAB_Cetyl trimethylammonium bromide; DNA_Deoxyribonucleic acid; EDTA_Ethylene diamine tetraacetic acid; F_{IS}_Inbreeding coefficient; F_{ST}_Genetic differentiation; IBC_Institute of Biodiversity Conservation; I_Shannon's information index; ISSR_Inter-simple sequence repeat, H_e_Gene diversity; H_o_Observed heterozygosity; m.a.s.l._Meter above sea level; N_a_Alleles per locus; N_e_Number of effective alleles per locus; N_m_Gene flow, PCR_Polymerase chain reaction; %P_Percent polymorphism; PIC_Polymorphic information content; RAPD_Random amplified polymorphic DNA; SIDA_Swedish International Development Cooperation Agency; ASRECA_Association for Strengthening Agricultural Research in Eastern and Central Africa; SSR_Simple sequence repeat markers; UPGMA_Un-weighted pair group method using arithmetic average.

Introduction

Okra [*Abelmoschus esculentus* (L.) Moench] is a vegetable crop grown in many parts of the world, especially in tropical, subtropical and warm temperate regions (Reddy et al., 2012; Magagula and Ossom, 2011). The geographic origin of okra remains uncertain and the literature supports India (Ikram-ul et al., 2013), Ethiopia (Benchasri, 2012; Reddy et al., 2012) or West Africa (Hammon and Van Sloten, 1989) as centers of origin. Okra is an annual tropical herb cultivated mainly for its edible green pod or fruit and for its young and succulent leaves. Fresh tender okra fruit is a good source of dietary fiber, which is fat-free and low in calories. Okra contains protein, vitamins A, C, and B complex, iron and calcium useful in human nutrition (AbdEl-Kader et al., 2010).

Although Ethiopia is considered as one of the centers of origin of okra (Benchasri, 2012; Reddy et al., 2012), there have not been formal research on okra breeding and genetic resources conservation in the country. Cultivation of okra is mainly using landraces maintained by farmers (Binalfew and Alemu, 2016). Wild okra genetic resources are present in its natural population with low human intervention. Limited reports are available that assessed the genetic diversity of the crop mainly using agro-morphological traits (Binalfew and Alemu, 2016; Demelie et al., 2015 & 2016; Yonas et al., 2014a,b). These reports indicated the presence of high morphological diversity of okra. To ehance our understanding of genetic diversity and genetic relatedness among genotypes, the use of molecular markers is feasible.

Molecular markers such as simple sequence repeats (SSRs) (Kumar et al., 2017), inter-simple sequence repeat (ISSR) (Yuan et al., 2014), amplified fragment length polymorphism (AFLP) (Salameh, 2014) and random amplified polymorphic DNA (RAPD) (Kaur et al., 2013) have been applied in okra genetic diversity analysis. SSR markers have been become the markers of choice in genetic studies because they are highly polymorphic, multi-allelic, codominant and have greater reproducibility (Zorrilla-Fontanesi et al., 2011). Recently, SSR markers are developed for okra that can be utilized in genetic diversity studies in this crop (Schafleitener et al., 2013). There is no previous report on genetic diversity assessment of okra collections from Ethiopia using molecular markers.

The present challenges posed by climate change require crop diversification to cope with future risks related to crop production. For this purpose, okra can be considered as an alternative crop for its favorable attributes including its rapid growth cycle, easy cultivation, tolerance to drought, resistance to pests, high yield potential and nutritional values (Çalışır et al., 2005). Therefore, this study is aimed to assess the genetic diversity and population structure of okra collections of diverse geographic origin using selected polymorphic SSR markers.

Results

Polymorphism and allelic diversity of SSR markers

In this study, 19 SSR markers were used to profile the 32 okra accessions including three introduced commercial varieties. Estimates of genetic diversity parameters are presented in Table 1. Three markers with a major allele frequency of ≤ 5% were omitted from the analysis. The remaining 16 selected SSR markers produced a total of 71 polymorphic alleles. Number of alleles scored per markers ranged from 1 to 8, with a mean of 4.44 alleles per locus. The major allele frequency ranged from 0.25 for the marker AVRDC28 to 1 for the marker AVRDC66. Allele availability ranged from 0.16 for the marker AVRDC89 to 1 for the marker AVRDC28, with a mean of 0.67. The observed heterozygosity values ranged from 0.03 to 1.00, with a mean of 0.65. This suggests that 65% of the loci were heterozygous, while the remaining loci reached acceptable level of homozygosity. Marker AVRDC17 had the lowest H_o value (0.03), whereas 44% of the markers revealed the highest H_o value (1.00). The gene diversity (H_e) ranged from 0.10 (marker AVRDC17) to 0.78 (AVRDC28) with a mean of 0.52; suggesting that 52% of the landraces are expected to be heterozygous at a given locus under random mating conditions.

The inbreeding coefficient (F_{IS}) exhibited contrasting values ranging from –0.07 to 0.76 with a mean of -0.14. F_{IS} represents the average deviation of the population's genotypic proportions from Hardy–Weinberg equilibrium for a locus. In this study, 63% of the markers had negative F_{IS} values indicating an excess of heterozygotes. For example, for locus AVRDC89, 50% of the okra genotypes are expected to be heterozygous at the specific locus under random mating conditions; however, 100% of the genotypes at this locus were heterozygotes. The PIC values of the tested SSR markers ranged from 0.00 to 0.75, with a mean of 0.47. Fifty percent of

the markers had PIC value of ≥0.50, indicating a moderate discriminatory power of individual SSR loci used in this study.

Within and among population variation

The number of gene copies observed using the 16 SSR loci ranged from 6 to 24. The highest number of gene copies was recorded from accessions collected from southwestern region and the lowest was from introduced commercial varieties. The maximum gene copies were found from accessions collected from lowland areas (480 – 630 m.a.s.l.) followed by accessions collected from altitudes of 1020 – 1180 m.a.s.l. However, there was non-significant variation in gene copies among districts of collection. The number of usable loci with less than 5% missing data ranged from 4 to 9. All the accession collected from different geographic regions of Ethiopia showed non-significant variation in terms of the number of usable loci. However, the exotic commercial varieties revealed the highest number of usable loci with less than 5% missing data. Accessions collected from the Mandura, Menge and Gambella sites revealed the highest number of usable loci (Table 2).

Genetic parameter estimates of okra populations based on geographic origin, altitude and district of collection are presented in Table 3. The four populations were classified based on geographic origin revealing the highest mean values for most of the genetic parameters followed by classification based on altitudinal ranges. Among the test populations, accessions collected from southwestern region scored significantly higher level of variation for most of the genetic diversity parameters. The mean observed number of detected alleles (N_a) was higher for accessions from northwestern region followed by accessions from southwestern region. Non-significant variation was observed in gene diversity (H_e) among the different geographic origins except for accessions collected from southwestern region. The higher number of N_a and H_e were observed in accessions collected from southwestern region. This suggests that the landraces are genetically diverse and unique. However, exotic accessions revealed the lowest values for all the genetic parameters including N_a and H_e. This could partly be due to small number of accessions sampled or the check improved commercial cultivars have undergone continuous breeding and selection processes resulting in narrow genetic diversity.

Accessions collected from lowland areas (480 - 550 m.a.s.l.) revealed the highest values for all genetic parameters tested in this study followed by mid altitudes (1020 - 1105 m.a.s.l.). The accessions collected from relatively higher altitudinal ranges, on the other hand, displayed low genetic diversity. The high level of genetic diversity displayed by the accessions collected from lowland and midland areas which are adapted to drought stress and other harsh environmental conditions. Among the district of collection, Abobo and Dangura displayed the highest genetic diversity, whereas accessions collected from Gambela displayed the lowest genetic diversity. The higher genetic diversity observed in the present okra populations is mainly attributed to differences in geographic origin and altitudinal ranges.

Significant variation was observed in mean genetic distance (GD) among the 32 okra accessions. The genetic distances estimates revealed a normal distribution curve in which 68% of
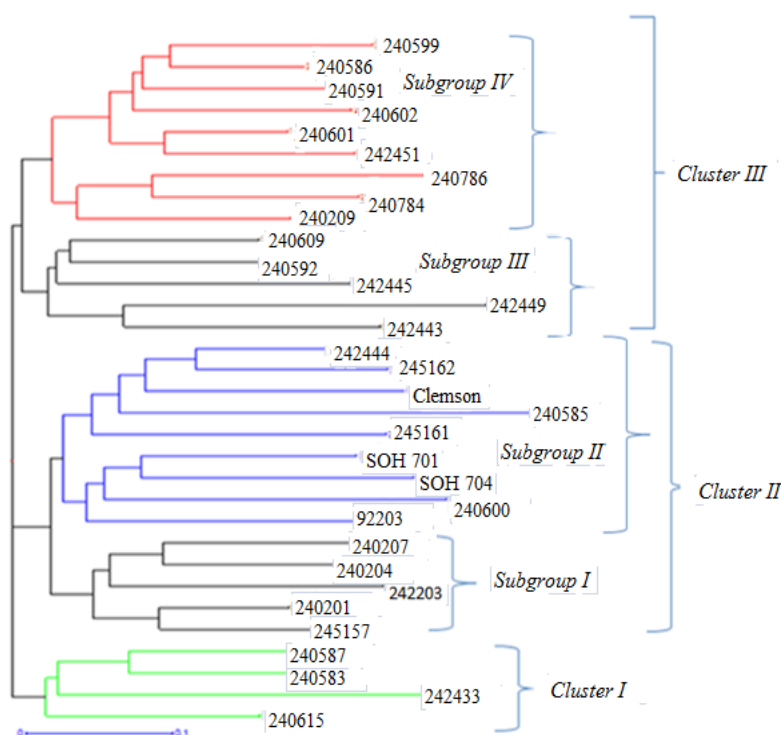
Table 1. Genetic parameters based on 16 SSR markers among 32 Okra accessions.

Marker	M_F	N_a	A	H_e	H_o	PIC	F_{IS}
AVRDC01	0.76	5.00	0.97	0.40	0.26	0.38	0.38
AVRDC09	0.45	7.00	0.94	0.70	0.30	0.65	0.58
AVRDC17	0.95	3.00	0.94	0.10	0.03	0.09	0.66
AVRDC21	0.40	4.00	0.91	0.66	0.97	0.60	-0.44
AVRDC28	0.25	7.00	1.00	0.78	1.00	0.75	-0.26
AVRDC39	0.50	3.00	0.06	0.63	1.00	0.55	-0.33
AVRDC54	0.42	5.00	0.38	0.64	1.00	0.57	-0.53
AVRDC56	0.50	8.00	0.47	0.70	1.00	0.67	-0.41
AVRDC63	0.38	5.00	0.66	0.70	0.76	0.64	-0.07
AVRDC64	0.50	2.00	0.03	0.50	1.00	0.38	-0.17
AVRDC66	0.98	3.43	1.00	0.26	0.26	0.24	0.25
AVRDC70	0.50	3.00	0.56	0.57	1.00	0.48	-0.74
AVRDC77	0.75	4.00	0.88	0.40	0.50	0.36	-0.22
AVRDC78	0.53	7.00	0.97	0.65	0.16	0.61	0.76
AVRDC86	0.80	3.00	0.88	0.33	0.39	0.31	-0.16
AVRDC89	0.50	2.00	0.16	0.50	1.00	0.38	-1.00
Mean	0.57	4.31	0.67	0.52	0.65	0.46	-0.15

M_F= Major allele frequency; N_a = number of alleles amplified with $\geq$5% frequency; A= Availability; H_e= gene diversity; H_o = Observed heterozygosity; PIC= Polymorphic information content; F_{IS}= inbreeding coefficient
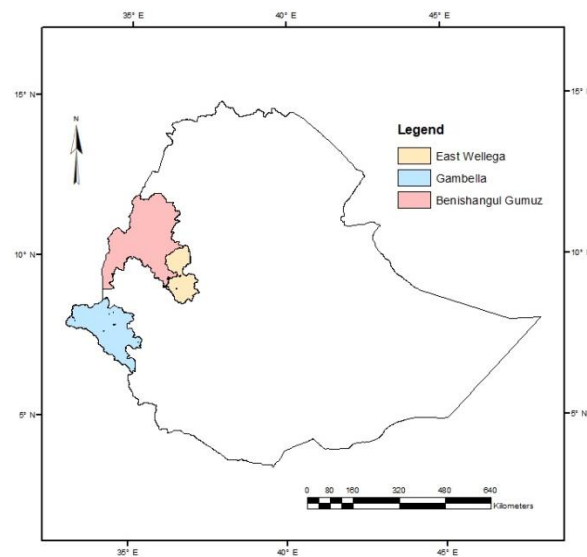
Fig 1. Administrative map of Ethiopia showing the three Zones where the okra accessions were collected.

Table 2. Standard diversity indices from molecular analysis of variance for 32 okra accessions.

| Population	Number of gene copies	Number of loci	Number of usable loci with <5% missing data	Number of polymorphic loci
Geographic origin				
Northwestern Ethiopia	18	16	6	5
Western Ethiopia	16	16	5	5
Southwestern Ethiopia	24	16	4	4
Others	6	16	7	6
Altitudinal ranges (m)				
480-630	20	16	4	4
725-935	12	16	8	5
1020-1180	18	16	6	5
1200-1480	8	16	6	5
Collection districts				
Mandura	8	16	8	6
Dangura	10	16	7	6
Asossa	8	16	5	3
Menge	8	16	8	5
Akobo	8	16	7	7
Abobo	10	16	5	4
Gambella	6	16	8	4

Fig 2. Dendrogram generated based on neighbour-joining clustering using the unweighted pair group method (UPGMA) depicting genetic relationships among 32 okra accessions using16 SSR markers.

Table 3. Genetic parameters of 32 okra populations based on geographic origin, altitude and collection districts.

Population	Genetic parameters						
	N_a	N_e	I	H_e	F_{IS}	P_A	PCI
Geographic origin							
Northwestern Ethiopia	2.81	2.17	0.74	0.50	-0.35	0.50	0.81
Western Ethiopia	2.69	2.03	0.74	0.47	-0.18	0.44	0.75
Southwestern Ethiopia	2.75	2.17	0.83	0.61	-0.32	0.56	0.94
Others	1.75	1.57	0.54	0.48	-0.39	0.19	0.75
Mean	2.56	1.99	0.71	0.52	-0.31	0.42	0.81
SE	0.18	0.13	0.06	0.04	0.08	0.16	0.04
Altitudinal range (m)							
480-630	2.94	2.21	0.84	0.63	-0.33	0.63	0.94
725-935	2.06	1.61	0.56	0.44	-0.23	0.25	0.69
1020-1180	2.94	2.31	0.81	0.52	-0.30	0.69	0.81
1200-1480	1.56	1.30	0.41	0.32	-0.32	0.06	0.56
Mean	2.38	1.86	0.65	0.48	-0.30	0.41	0.75
SE	0.19	0.14	0.06	0.04	0.07	0.30	0.08
Collection district							
Mandura	2.06	1.80	0.56	0.47	-0.56	0.19	0.63
Dangura	2.38	2.05	0.69	0.57	-0.33	0.25	0.81
Asossa	2.00	1.79	0.61	0.45	-0.05	0.19	0.63
Menge	2.06	1.81	0.60	0.47	-0.43	0.19	0.69
Akobo	2.06	1.67	0.61	0.47	-0.24	0.19	0.75
Abobo	2.38	1.99	0.68	0.59	-0.59	0.31	0.88
Gambella	1.50	1.41	0.42	0.40	-0.77	0.00	0.56
Mean	2.06	1.79	0.59	0.49	-0.42	0.19	0.71
SE	0.12	0.10	0.04	0.04	0.06	0.10	0.04

N_a= number of alleles with ≥ 5% allele frequency; N_e= number of effective alleles; I= Shannon diversity index; H_o= observed heterozygosity; H_e= unbiased expected heterozygosity, F_{IS}= fixation index, P_A= number of private alleles; PIC= polymorphic information content.
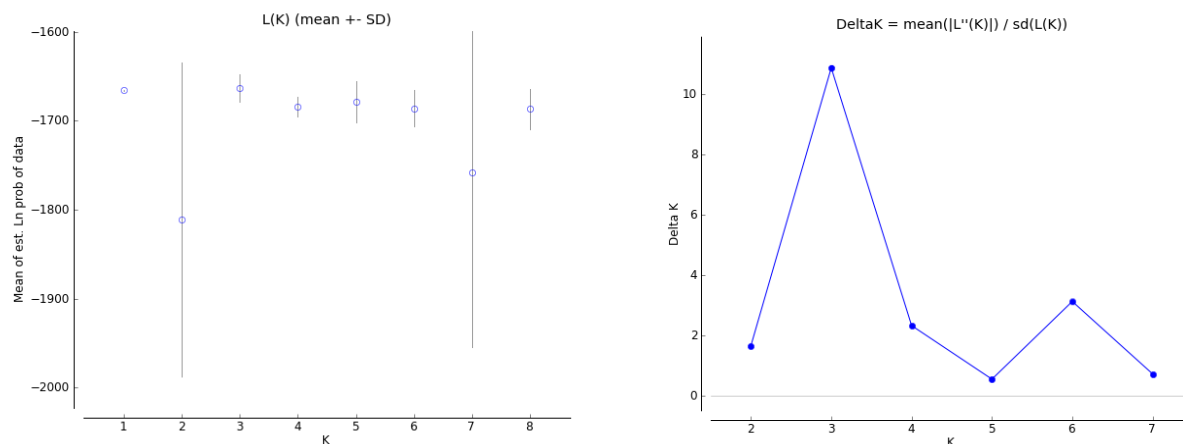
Fig 3. Model-based population structure analysis using 32 okra accessions. Plot of LnP(K) and Delta K values for different numbers of sub-populations (K) assumed in the STRUCTURE software

Table 4. Mean genetic distance of 32 okra accessions from five geographic origins.

Geographic origin	Accession	Min	Max	Mean	SD	CV (%)
Northwestern Ethiopia	240201	3.4	7.1	4.79	0.76	15.98
	242203	3.4	6.1	4.82	0.70	14.61
	240204	3.0	6.4	4.47	0.87	19.51
	240207	3.5	5.9	4.89	0.66	13.58
	240209	3.0	6.0	4.54	0.71	15.60
	240784	4.3	7.1	5.38	0.63	11.63
	240786	4.0	6.2	5.29	0.52	9.79
	245161	3.4	5.7	4.54	0.55	12.12
	245162	3.4	6.3	4.82	0.72	14.85
Mean		3.0	7.1	4.84	0.68	14.18
Western Ethiopia	92203	3.0	5.6	4.54	0.67	14.67
	242433	2.8	5.6	4.27	0.66	15.55
	242443	2.8	5.4	4.24	0.72	16.92
	242444	3.5	6.4	4.99	0.85	17.02
	242445	2.5	5.7	4.59	0.81	17.65
	242449	2.5	5.4	4.32	0.78	18.12
	242451	2.2	5.7	4.48	0.84	18.69
	245157	2.8	5.9	4.51	0.85	18.89
Mean		2.2	6.4	4.49	0.77	17.19
Southwestern Ethiopia	240583	2.8	5.9	4.59	0.78	16.93
	240585	3.7	6.0	4.78	0.54	11.39
	240586	3.4	6.2	5.05	0.75	14.85
	240587	2.9	6.1	4.64	0.85	18.31
	240591	2.2	5.5	4.36	0.70	16.14
	240592	2.8	6.1	4.52	0.92	20.25
	240599	2.8	5.3	4.25	0.63	14.95
	240600	3.3	5.9	4.75	0.65	13.62
	240601	3.1	6.2	5.07	0.70	13.77
	240602	3.0	5.8	4.63	0.69	14.98
	240609	2.6	5.9	4.40	0.71	16.12
	240615	3.0	5.5	4.49	0.72	16.13
Mean		2.2	6.2	4.63	0.72	15.62
Introductions	SOH701	3.5	6.4	4.55	0.73	16.07
	SOH704	3.2	5.3	4.30	0.59	13.67
	Clemson	4.0	6.1	5.04	0.60	11.89
Mean		3.2	6.4	4.67	0.59	12.78
Overall mean		2.2	7.1	4.65	0.71	15.44
SE		0.09	0.08	0.05	0.02	0.44

Min = minimum; Max = maximum; SD = standard deviation, CV = coefficient of variation.

Table 5. Analysis of molecular variance among and within okra populations.

Source of variation	df	SS	Est. Var.	Perc. Var.	Fixation index
Based on geographic origin					
Among groups	1	2.98	0.170	16.25	$F_{CT}=0.162$
Among populations within groups	2	2.60	0.015	1.39	$F_{SC}=0.017$
Among individuals within populations	28	28.65	0.160	15.28	$F_{IS}=0.185$
Within individuals	32	22.50	0.703	67.09	$F_{IT}=0.329$
Total	63	56.73	1.048		
Based on altitudinal range					
Among groups	1	2.00	0.021	2.39	$F_{CT}=0.024$
Among populations within groups	2	2.63	0.027	3.08	$F_{SC}=0.032$
Among individuals within populations	25	23.95	0.117	13.14	$F_{IS}=0.139$
Within individuals	29	21.00	0.724	81.38	$F_{IT}=0.186$
Total	57	49.59	0.890		
Based on collection district					
Among groups	2	2.60	-0.018	-2.14	$F_{CT}=-0.021$
Among populations within groups	4	6.50	0.091	10.37	$F_{SC}=0.102$
Among individuals within populations	22	21.00	0.081	9.21	$F_{IS}=0.100$
Within individuals	29	21.00	0.724	82.56	$F_{IT}=0.174$
Total	57	49.59	0.877		

df= degrees of freedom; SS= sum of squares; Est. var.= Estimated variance, Perc. Var = Percentage variance; F_{CT}= fixation index among groups; F_{SC}= fixation index among populations within groups; F_{IS}= fixation index among individuals within populations; F_{IT} = fixation index within individuals

the distances is within one standard deviation of the mean. Mean GD among accessions ranged from 2.2 (recorded between accessions 240592/242449 and 242451/240591) to 7.1 (240784/240201 and 240201/240784), with the mean distance of 4.65 ± 0.11. Accessions 240784 and 240786 from northwest revealed the highest mean genetic distance of 5.38 and 5.29, respectively. Accessions 242443 with GD of 4.24 from western and 240599 (GD=4.25) from southwestern Ethiopia had the lowest mean genetic distance values (Table 4).

The lowest coefficient of variation was recorded from accession 240786, whereas the highest was scored for accession 240592 (Table 4). Considering the geographic origins, the accessions from northwestern Ethiopia had the highest mean genetic distance followed by exotic commercial varieties introduced from India and the USA. Clemson, a commercial okra variety from the USA revealed similar GD to that of accession 240586 which was collected from southwestern region of Ethiopia suggesting that the two genotypes were genetically closely related.

Molecular analysis of variance (AMOVA)

Table 5 presents the analysis of molecular variance among okra populations based on geographic origin, altitude and district of collection. The variance component computed among different populations based on geographic origin, altitude ranges and district of collection were 16.25%, 2.39% and -2.14%, respectively. The highest variance component observed among population of different geographic origins indicated that accessions within each region had homogeneous allele frequencies, while accessions from different regions differ in allele frequencies. On the other hand, the negative variance component observed among populations based on districts of collection indicated the absence of significant allele frequency variation among the districts. The variance component among populations within groups was the highest in the seven populations classified based on district of collections followed by the 4 populations

classified based on altitudinal range, while it was low in the 4 populations grouped based on geographic origin. This suggested that altitude and collection districts had more contribution to the allele diversity of accessions within a group. Variance component within individuals was significantly higher in the 7 populations based on district (82.56%) followed by the 4 populations based on altitude (81.38%). However, the within individual variation observed among the 4 populations of geographic origin was low (67.09%). This result suggested that the genetic variation observed among okra accessions from Ethiopia was highly influenced by variation due to collection districts and altitudinal ranges.

The mean fixation index within individuals were positive in all classes of populations in which the highest ($F_{IS}=0.185$) was observed among the 4 populations based on geographic origin followed by among the 4 okra populations based on altitude ($F_{IS}=0.139$) suggesting that outcrossing among the tested okra populations was low. The results suggested that the okra population structure was moderately dependent on the variation among geographic origins. Similarly, altitudinal variations also had a considerable contribution to the diversity of accessions irrespective of variations among geographic origins in Ethiopia.

The F_{ST} values among okra populations classified based on geographic origin ranged from 0.105 to 0.251, while F_{ST} values based on altitude ranged from 0.163 to 0.325. Among the three classifications, the highest F_{ST} values were estimated between populations obtained from altitude ranges of 725-935 and 1200-1480 m.a.s.l. (0.325). The highest Nei's unbiased genetic distance (0.421) was observed between accessions collected from altitude ranges of 1020-1180 and 1200-1480 m.a.s.l. suggesting accessions collected from the higher altitude were unique and isolated. Accessions collected from the three regions of Ethiopia were greatly differentiated from exotic commercial varieties ($F_{ST}>0.22$), whereas each region was moderately differentiated from one another. Accessions collected from Gambela was uniquely differentiated from the other districts (data not shown).

Genetic relationships among okra accessions

The population structure of the accessions was assessed using both distance-base and model-based analyses. Genetic relationship among the okra accessions was revealed using neighbour-joining algorithm using the unweighted pair group method (UPGMA). The analyses classified the 32 accessions into three distinct clusters (Fig 2). Similarly, the model-based STRUCTURE analysis using the Evanno criterion (Earl & vonHoldt, 2012), confirmed the presence of three distinct groups at K = 3 (Fig 3). However, genetic clustering based on the 16 SSR markers did not perfectly correspond (p = 0.072) to the pre-defined population structure based on geographic origin, altitudinal ranges or districtsof collection. The first cluster included only four accessions of which three were collected from Gambella and one from BenishangulGumuz. While Clusters II and III consisted of 14 accessions each, which were further subdivided into two subgroups. Cluster II was dominated by accessions collected from northwestern regions, while accessions collected from southwestern region were more common in Cluster III. Cluster II Sub-group I contained five accessions sampled from BenishangulGumuz region, whereas Sub-group II comprised of nine genotypes that were collected from BenishangulGumuz and Gambella regions. The dendrogram also revealed that the three introduced accessions were allocated with collections from the different regions of Ethiopia. Sub-group III and IV comprised of five and nine accessions, respectively. Among the three main clusters, Cluster III revealed the highest mean intra-cluster genetic distance followed by Cluster II. All accessions with the highest genetic distance were allocated in Sub-group IV, while all accessions with the smallest mean genetic distance were grouped in Sub-group II. However, non-significant variation was observed in the mean inter-cluster genetic distance among the clusters. The highest variation within each cluster was observed in Cluster I and Subgroup IV.

Discussion

Knowledge of the genetic variation and inter relationships among genetic resources can help plant breeders to identify accessions of interest and also provide information regarding materials for use in crop improvement programs (Cruz et al., 2007). This has been a major limitation in the utilization and development of improved okra varieties. The present study is the first major effort to analyze the okra germplasm collections from Ethiopia with the aim to measure genetic distinctiveness and population divergence using SSR markers.

Only few okra SSR markers have been developed recently and made available for researchers (Schafleitener et al., 2013). This marker system is most adapted for genetic characterization of accessions (Agrama and Tuinstra, 2003) and genetic studies of species with complex genomes (Zorrilla-Fontanesi et al., 2011). In the current study, 32 okra accessions collected from different regions of Ethiopia, India and the USA were subjected to genetic diversity analysis using 19 SSR markers. However, only16 markers with alleles frequency >5% were sufficient to differentiate the 32 okra accessions. This demonstrates the suitability of SSR markers for the analysis of genetic variation and relationship in okra accessions. Schafleitener et al. (2013) also classified 58 accessions from three okra species collected from wide geographic origins using the same set of SSR markers.

In the present study, the 16 SSR primers generated a total of 71 putative alleles of different fragment size. The number of alleles detected ranged from 2 to 8, with a mean value of 4.3 per locus. Forty six percent of the loci were polymorphic with 67% of the alleles were sufficient to decimate the genotypes on average. In okra genetic diversity studies a greater number of alleles (154) using ISSR were reported (Yuan et al., 2015), while 150 alleles were detected using AFLP markers by Salameh (2014). Eight markers (50%) were found to be highly polymorphic with PIC > 0.50 suggesting their high discriminating power and their utility for genetic diversity studies in this species. The highest number of gene copies were observed when genotypes were classified according to geographic origin as compared to altitude and collection districts. However, considering number of usable loci with <5% missing data, the highest gene copy were observed when genotypes were classified according to collection districts. This signifies that the genetic diversity could partly be attributed to differences in agro-climatic conditions of the collection districts and it may be evolved with environmental stress factors.

In the current study, analysis of statistical genetic parameters across geographic origin, altitude and collection districts revealed a high level of variability in gene diversity, heterozygosity, Shannon's diversity index. The relatively higher mean gene diversity, Shannon's diversity index and heterozygosity suggested the genetic diversity in the populations was a function of geographic region and altitude. Allelic diversity, heterozygosity and Shannon's diversity index are commonly used parameters to understand the genetic diversity in plant genetic resources (Hamilton, 2009; Freeland et al., 2011). Shannon diversity is a population parameter where high values indicatea more diverse the population. Fixation index, genetic diversity and heterozygosity are measures of diversity among varied accessions (Toro and Caballero, 2005). The presence of genetic differentiation is the basis of population structure which is the function of low dispersal rates between local populations (Slatkin, 1987). Distances and geographic barriers to dispersal are among the causes for the occurrence of population structure.

Partitioning of the revealed genetic diversity using analysis of molecular variances revealed that most of the molecular variations in the okra accessions existed within individual (67 - 83%). The highest variance component among groups of geographic origins and altitude and highest positive fixation indices for all sources of diversity indicate the low outcrossing rate among the tested populations. Populations within geographic origin had homogeneous allelic frequencies but populations among the different geographic origin differ in their allelic frequencies. However, the negative variance component among populations of districts of collection and the negative fixation index among groups indicated the absence of significant difference in allele frequencies between populations and the presence of outcrossing among populations. The results indicated that the different population classified on the basis of geographic origin and altitude showed moderate population differentiation as compared to populations of districts of collections. Pluess and Stocklin (2004) indicated spatial isolation of populations adapted to

different environments may result from an increase in gene diversity between populations.The two major drivers of genetic differentiation and thereby genetic diversity among plant populations are geographic distances and the environment (Nosil et al., 2008; Slatkin, 1987). Given the high within individual variance and low outcrossing among populations of varied geographic regions and altitudinal ranges, collection strategy for conservation of okra should focus on sampling of representative genotypes covering wide geographic regions and altitudinal ranges of the target agro-ecologies.

Genetic clustering using neighbour-joining of the 32 okra accessions classified the genotypes into three distinct clusters and four sub-clusters. Although the cluster patterns did not correspond to the pre-defined population structure based on the geographic origin, altitude and districts of collection, accessions from different regions were genetically distant. The higher Nei's genetic distances observed among okra populations based on collection districts and altitude in the same or different geographic origins indicated the movement of okra germplasm is restricted and the environment where they grow has also contributed to genetic differentiation in Ethiopia. Population structure occurs when dispersal rates between local populations are sufficiently low to allow genetic differentiation (Slatkin, 1987). This suggested most of the landraces cultivated in each geographic region were unique and isolated. This may be due to the low outcrossing rate among regions and the low human involvement in germplasm exchange. In the past there has been limited interest in okra cultivation and consumption in Ethiopia (Binalfew and Alemu, 2016). Farmers collect okra fruits from wild plants for local use and market. Consequently, the okra accessions from each geographic region in Ethiopia can be considered as vital genetic resources for future breeding. The presence of random mating in a species make the population a single entity while lack of gene flow between regions allow the formation of sub-populations of geographic regions (Morjan and Rieseberg, 2004). Nonrandom spatial distribution of genetic diversity in a species results in differentiation of populations (Hartl and Clark, 1997). Isolation by geographic distance is one of the major factors for genetic differentiation (Nosil et al., 2008; Slatking, 1987)

Okra is native to North Eastern Africa in the area of Ethiopia and Sudan (Santos, 2012) and it was flourished before date palm in the tropical climate of Ethiopia (William, 1999). In Ethiopia, the environmental factors such as temperature, rainfall, soil physical and chemical properties are variable within short distances. The spatial isolation of populations either in geographic distances or physical barriers such as the great Blue Nile gorge coupled with the different environments where the accessions adapted to may be the major cause for the wide genetic diversity between okra populations. Spatial isolation of populations adapted to different environments increase the genetic diversity between populations (Pluess and Stocklin, 2004). The present results based on SSR analysis indicated the presence of wide genetic base of okra in Ethiopia. The current results agreed with the previous studies that indicated the presence of considerable agro-morphological variability of okra accessions within and between regions in Ethiopia (Demelie et al., 2015 & 2016; Binalfew and Alemu, 2016; Yonas et al., 2014a, b).

Materials and Methods

Plant material

In this study, a total of 32 okra accessions were used. Twenty nine accessions were obtained from the Institute of Biodiversity Conservation (IBC) in Ethiopia. Three commercial introduced varieties (two from India and one from the United States of America) were included in the study as comparative controls. Among the IBC accessions 17 were originally collected from Benishangul-Gumuz Region and 12 from Gambella Region of Ethiopia. The collection regions represent three agro-ecologies viz. southwestern, northwestern and western Ethiopia. Table S1 summarizes the description of the 32 okra accessions used in this study with geographic origins, collection altitude and site (Fig 1). The accessions were collected from wide altitudinal ranges (480 to 1480 meter above sea level). Based on altitude the 29 okra accessions were grouped into four populations. The first group consisted of accessions collected from low altitudes (480 - 630 m.a.s.l.); the second group composed of accessions collected from an altitude range of 725 - 935 m.a.s.l.; the third group was made up of accessions collected from an altitude range of 1020 - 1180 m.a.s.l; and the fourth group contained accessions collected from a relatively higher altitudes (1200 – 1480 m.a.s.l.).

DNA extraction, purification and quantification

Okra seeds were grown in 2 liter capacity polyethylene plastic pots in the Controlled Research Facility (CEF), University of KwaZulu-Natal, Pietermaritzburg, South Africa. To account possible genetic variability within each accession, leaf samples were taken from 10 plants per accession six weeks after planting. The leaf samples were sent to INCOTEC PROTEIOS laboratory (INCOTECH, SA Pty Ltd, Pietermaritzburg, South Africa) for SSR analysis. Genomic DNA was extracted from 20 young leaves (200 - 300 mg per accession) of 32 okra genotypes using modified CTAB method (Ahmed et al., 2013). The quantity and quality of total genomic DNA were determined by 0.7% Tris-Borate-EDTA (TBE) agarose gel electrophoresis and spectrophotometer, respectively. A working concentration of 20 ngμl^{-1} was standardized for all extracted DNA.

Polymerase chain reaction (PCR)

All the 32 samples were used in bulked amplification, using DNA extracted from leaf material per sample. SSR sequences were amplified through PCR using 19 selected SSR primers developed for okra (Table 1). The markers were selected based on their high polymorphic information content (PIC) and that they were developed and recommended for okra germplasm diversity studies (Schafleitner et al., 2013). PCR amplification reaction contained 20 µl of PCR mix. The mix contains 1x PCR buffer, 3 mM MgCl, 1.25 U Taq polymerase, 0.2 mM dNTPs, 4pM each primer and 5 µl genomic DNA. A PCR profile of initial denaturation for 2 min at 94 ^{o}C, and 33 cycles of denaturation for 1 min at 94 ^{o}C, annealing temperature of 63 ^{o}C for 2 min, extension for 2 min at 72 ^{o}C was used. PCR products were fluorescently labeled and separated by capillary

electrophoresis on ABI 3130 automatic sequencer (Applied Biosystems, Johannesburg, South Africa) and analysis was performed using GeneMapper 4.1 (Applied Biosystems, Johannesburg, South Africa).

Data analysis

The analysis of genotypic data in this study was performed using two approaches. In the first approach, polymorphisms were treated as binary data (presence or absence). In this case, each amplified fragment was considered as one locus and evaluated as dominant markers. However, to determine the genetic structure within and among genotypes, a second approach based on the co-dominant nature of the marker was adopted. Genotypic data were subjected to analyses with various measures of genetic diversity within and among genotypes using GenAlex software version 6.5 (Peakall and Smouse, 2006). The χ^2 test was performed to determine the differences in allele frequencies among the SSR markers. Genetic diversity parameters such as total number of alleles per locus (N_a), number of effective alleles per locus (N_e), Shannon's information index (I), observed heterozygosity (H_o), gene diversity (H_e), number of putative alleles (P_a), percent polymorphism (%P), and inbreeding coefficient (F_{IS}) were determined using the protocol of Nei and Li (1979). Population differentiation (pairwise F_{ST}) was determined using PowerMarker 3.25 (Liu and Muse, 2005) and polymorphic information content (PIC) were estimated using GenAlex software.

The program GGT 2.0 (Van Berioo, 2008) was used to calculate the Euclidian distances between bulked samples. Pair-wise comparison of genetic distance were made to determine the genetic relationship of accessions based on mean Euclidian distances, geographic regions, altitudinal ranges and collection districts. The mean intra- and inter-cluster distances were calculated as a mean distance of accession within the cluster and with other accessions in other clusters, respectively.

The Bayesian genotypic clustering approach of STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to determine the population structure existed with the okra accession. An admixture model with independent allele frequencies, without prior population information, was used to simulate the population. Each individual is grouped in a given cluster using 'membership coefficient' for each cluster interpreted as a probability of membership. To assegnin individual genotype to a given population and for optimal alignment of genotypes, 10 replicates structure analysis were conducted. The genotype membership was determined by the computer program CLUMPP (Jakobsson and Rosenberg, 2007). The structure analysis result was vsualized by the online genetic software STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

The binary data were used to obtain a dissimilarity matrix using the Jaccard index. The matrix was used to run a cluster analysis based on neighbour-joining employing the software DARwin 5.0 (Perrier and Jacquemoud-Collet, 2006). A dendrogram was then generated on the dissimilarity matrix. Bootstrap analysis was performed for node construction using 10,000 bootstrap values and phylogenetic tree was constructed. The mean genetic distances of each accession were calculated from the pair-wise genetic distance with the other accessions. Analysis of molecular variance (AMOVA) and fixation indices were computed using Arlequin3.11software (Excoffier et al., 2005).

Conclusions

The present study revealed the presence of high genetic diversity and population divergence among okra collections from Ethiopia when analyzed using selected polymorphic SSR markers. The genetic diversity was largely ascribed by differences in geographic origin, altitude and districts of collection of test genotypes. The study demonstrated the presence of genetic diversity and population differentiation of the okra germplasm from Ethiopia that can be exploited in future breeding programs and conservation strategies. Further, wild relatives and domesticated types of okra prevails in various regions of Ethiopia. Therefore, further studies are required using larger and representative sample of okra employing the presently tested SSR markers to fully explore the genetic potential of the crop for future breeding or strategic conservation.

References

Abd El-Kader AA, Shaaban SM, Abd El-Fattah MS (2010) Effect of irrigation levels and organic compost on okra plants (*Abelmoschus esculentus*) grown in sandy calcareous soil. Agr Bio J North Amer. 1:225-231.

Agrama HA, Tuinstra MR (2003) Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr J Biotech. 2: 334-340.

Ahmed N, Nawaz S, Iqbal A, Mubin M, Butt A, Lightfoot DA, Maekawa M (2013) Extraction of high-quality intact DNA from okra leaves despite their high content of mucilaginous acidic polysaccharides. Biosci Meth. 4:19-22.

Benchasri S (2012) Okra (*Abelmoschus esculentus* (L.) Moench) as a valuable vegetable of the world ratar. Ratarstvo I povrtarstvo. 49:105-112.

Binalfew T, Alemu Y (2016) Characterization of okra (*Abelmoschus esculentus* (L.) Moench) germplasms collected from western Ethiopia. Inter J Res Agric Forest. 3 : 11-17.

Çalışır S, Özcan M, Hacıseferoğullar, Yıldız MU (2005) A study on some physicochemical properties of Turkey okra (*Hibiscus esculenta*) seeds. J Food Engin. 68:73–78.

Cruz, VMV, Luhman R, Marek LF, Rife CL, Shoemaker RC, Brummer EC, Gardner CAC (2007) Characterization of flowering time and SSR marker analysis of spring and winter type *Brassica napus* L. germplasm. Euphytica 153: 43-57.

Demelie M, Mohamed W, Gebre E (2015) Genetic diversity of Ethiopian okra collections through multivariate analysis at Werer, Rift Valley of Ethiopia. Int J Sci Technol. 3: 186-192.

Demelie M, Mohamed W, Gebre E (2016) Variability, heritability and genetic advance in Ethiopian okra [*Abelmoschus esculentus* (L.) Monech] collections for tender fruit yield and other agro-morphological traits. J App Life Sci Int. 4: 1-12.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 4:359-361 doi: 10.1007/s12686-011-9548-7.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinf. 1: 47-50.

Freeland J, Kirk H, Petersen S (2011) Molecular Ecology 2nd edition. John Wiley and Sons. Chichester, UK.

Hamilton M (2009) Population Genetics. Wiley-Blackwell: New York, NY, USA.

Hammon S, Van Sloten DH (1989) Characterization and evaluation of okra. In: Brown AH, Frankel OH, Marshall D, Williams JT (eds.) The Use of Plant Genetic Resources. Cambride University press, NY. pp. 173-174.

Hartl DL, Clark AG (1997) Principles of Population Genetics. Third editon. Sinauer Associates, Sunderland, Massachusetts, USA.

Ikram-ul-Haq, Khan AA, Azmat MA (2011) Assessment of genetic diversity in okra (*Abelmoschus esculentus* L.) using RAPD markers. Pakistan J Agri Sci. 50: 655-662.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-180.

Kaur K, Pathak M, Kaur S, Pathak D, Chawla N (2013) Assessment of morphological and molecular diversity among okra (*Abelmoschus esculentus* (L.) Moench.) germplasm. Afr J Biotech. 12: 3160- 3170.

Kumar S, Parekh MJ, Fougat RS, Patel SK, Patel CB, Kumar M, Patel BR (2017) Assessment of genetic diversity among okra genotypes using SSR markers. J Plant Biochem Biotech. 26:172-178.

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129.

Magagula P, Ossom E (2011) Effects of seed size on seedling vigor of okra (*Abelmoschus esculentus* L.) in Swaziland. Advan Env Bio. 5:180-187.

Morjan C, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol Ecol. 13: 1341–1356.

Nei M (1987) Molecular Evolutionary Genetics.New York: Columbia University Press.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation'' and multiple roles for divergent selection. Evolution 62:316–336.

Peakall R, Smouse PE (2012) GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539.

Perrier X, Jacquemoud-Collet JP (2006) DARwin (Version 6)-Dissmilarity analysis and representation for windows. Available at (http://www.darwin.cirad.fr/darwin.html).

Pluess AR, Stocklin J (2004) Population genetic diversity of the clonal plant *Geumreptans* (*Rosaceae)* in the Swiss. Amer J Bot. 91: 2013-2021.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Reddy TM, Haribhau K, Ganesh M, Chandrasekhar RK, Begum H (2012) Genetic divergence analysis of indigenous and exotic collections of okra [*Abelmoschus esculentus* (L.) Moench]. J Agric Technol. 8:611-623.

Salameh NM (2014) Genetic diversity of okra (*Abelmoschus esculentus* L.) genotypes from different agroecological regions revealed by amplified fragment length polymorphism analysis. Amer J Appl Sci. 11: 1157-1163.

Santos BM, Dittmar PJ, Olson SM, Webb SE, Zhang S (2012) *Okra Production in Florida*. University of Florida IFAS extension, pp 163-171.

Schafleitner R, Kumar S, Lin CY, Hegde SG, Ebert A (2013) The okra (*Abelmoschus esculentus*) transcriptome as a source for gene sequence information and molecular markers for diversity analysis. Gene 517:27–36.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236: 787–792.

Toro MA, Caballero A (2005) Characterization and conservation of genetic diversity in subdivided populations. Philosophical transactions of the Royal Society of London Series B, Biol Sci. 360: 1367–1378.

van Berloo R (2008) GGT 2.0: Versatile software for visualization and analysis of genetic data. J Hered. 99: 232-236.

William JL (1999) Okra a versatile vegetable crop. Horti Technol. 9:179-184.

Yonas M, Garedew W, Debela A (2014a) Multivariate analysis among okra [*Abelmoschus esculentus* (L.) Moench] collection in South Western Ethiopia. J Plant Sci. 9: 43-50.

Yonas M, Garedew W, Debela A (2014b) Variability and association of quantitative characters among okra [*Abelmoschus esculentus* (L.) Moench] collection in South Western Ethiopia. J Biol Sci. 14: 336-342.

Yuan CY, Zhang C, Wang P, Hu S, Chang HP, Xiao WJ, Lu XT, Jiang SB, Ye JZ, Guo XH (2014) Genetic diversity analysis of okra (*Abelmoschus esculentus*L.) by inter-simple sequence repeats (ISSR) markers. Genet Mol Res. 13 : 3165-3175.

Yuan CY, Wang P, Chen PP, Xiao WJ, Zhang C, Hu S, Zhou P, Chang HP, He Z, Hu R, Lu XT, Ye JZ, Guo XH (2015) Genetic diversity revealed by morphological traits and ISSR markers in 48 okras (*Abelmoschus escullentus* L.). Physiol Mol Biol Plants. 21:359-364.

Zorrilla-Fontanesi YZ, Cabeza A, Torres AM, Botella MA, Valpuesta V, Monfort A, Sanchez-Sevilla JF, Amaya I (2011) Development and bin mapping of strawberry genic-SSRs in diploid *Fragaria* and their transferability across the *Rosoideae* sub-family. Mol Breed. 27: 137-156.