Indian J. Genet., (2023); 83(1): 95-102
doi: 10.31742/ISGPB.83.1.12
ISSN: 0975-6906
www.ijgpb.org

RESEARCH ARTICLE

Development of EST-SSR markers using transcriptome data in okra (*Abelmoschus esculentus* L.): a genetically orphan crop

P. V. Puneeth, Ramesh K. Yadav*, Suman Lata, Mir Asif Iquebal[2], Akshay Talukdar[1], H. Choudhary and Avinash Tomer

Abstract

Okra is an important vegetable crop with few genomic resources and only limited simple sequence repeat (SSR) markers. In the present investigation, we have performed RNA sequencing (RNA-Seq) from leaf samples of okra varieties Pusa Sawani and Pusa Bhindi-5. Out of the 106224 EST-SSRs identified, mononucleotide repeats were dominant with a frequency of 72.20% (71699) followed by trinucleotide 15% (14897) and dinucleotide 10.68% (10604) repeats. In total, 183 best primer pairs (156 di and 27 trinucleotides) were selected for validation in 12 okra accessions comprising wild and cultivated *Abelmoschus* species. Overall, 163 primer pairs produced expected PCR amplicons with an amplification efficiency of 89%. A total of 23 primer pairs showed a 14.1% polymorphism percentage with an average value of 0.32 polymorphic information content (PIC) and these polymorphic markers were used further for diversity analysis. The UPGMA dendrogram analysis separated the okra accessions into two main clusters. This study provided insight into the distribution and frequency of EST-SSRs in the okra transcriptome. The EST-SSR markers developed here will be a robust molecular tool for germplasm identification, genetic diversity analysis, and comparative mapping in okra and related species due to their high polymorphism and cross-transferability.

Keywords: *Abelmoschus,* transcriptome, EST-SSR, genetic diversity, molecular markers

Introduction

Okra (*Abelmoschus esculentus* L.), (2n=2x=130) also called bhindi or lady's finger, is an important summer and rainy season vegetable crop which is widely cultivated in tropical and sub-tropical regions of the world. It is a rich source of dietary fibre, vitamins and minerals. Okra pods are used for industrial and consumptive purposes. It is probably a natural amphidiploid (allotetraploid) derived from two wild okra species, *i.e., Abelmoschus tuberculatus* (2n = 58) and (*Abelmoschus ficulneus* (L.) (2n = 72). Most okra production occurs in developing countries in Asia and Africa. With a production of 6.3 million tonnes, India is the world's leading producer (72% of the total world production) of okra (NHB 2019-20). However, utilization of biotechnological and molecular tools in okra improvement programmes has been very limited due to a large number of chromosomes, allopolyploid nature, large genome size (approximately 1.6 Gb), availability of few polymorphic molecular markers and undefined genetic maps, which makes it more complicated for genomic studies (Lata et al. 2021)

RNA sequencing (RNA-Seq) is a useful tool for developing SSR markers in non-model crops like okra where genome sequence is unavailable (Strickler et al. 2012). Only limited transcriptomic and genomic studies are reported and few EST-SSRs have been developed in okra. Schafleitner et al. (2013) reported the first transcriptome assembly in okra in which 935 non-redundant SSR motifs were identified from unigenes data and 161 polymorphic SSR markers were designed and 19 markers were used for diversity analysis. The development of genomic SSR markers in okra was first reported by Ravishankar et al. (2018) where 2708 contigs had microsatellites and 50 randomly selected SSR primers were used for the amplification of okra DNA.

Division of Vegetable Science, [1]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India

[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

***Corresponding Author:** Ramesh K. Yadav, Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India, E-Mail: rkyadavneh@gmail.com

How to cite this article: Puneeth P.V., Yadav R.K., Lata S., Iquebal M.A., Talukdar A., Choudhary H., Tomer A. 2023. Development of EST-SSR markers using transcriptome data in okra (*Abelmoschus esculentus* L.): a genetically orphan crop. Indian J. Genet. Plant Breed., **83**(1): 95-102.

Source of support: ICAR, NAHEP-CAAST

Conflict of interest: None.

Received: Aug. 2022 **Revised:** Dec. 2022 **Accepted:** Jan. 2023

Benefits of SSR markers, such as codominant inheritance, high reproducibility, chromosomal coverage and high polymorphism facilitate their application in evolution and genetic diversity studies in crop plants. Therefore, this work was undertaken to produce novel EST-SSRs in okra. In this study, the Illumina NGS technology was used to generate high-quality transcriptome sequences, and EST-based SSRs for okra were developed. The distribution of EST-SSR motifs in the sequences generated was described and a set of EST-SSRs was identified. Primer pairs were developed to validate their use as microsatellite markers in diversity analysis. We are confident that it will be helpful in linkage map creation, and elucidation of okra's biochemical pathways and physiological processes which aids in okra breeding programs in future.

Materials and methods

Development of EST-SSR markers

Pusa Sawani (presently yellow vein mosaic virus (yvmv) disease susceptible) and Pusa Bhindi-5 (yvmv resistant), widely adopted okra varieties were used for transcriptome sequencing using Illumina technology and *de novo* assembly was constructed. RNA extraction was done according to Barman et al. (2017). RNA Seq data is available at NCBI with BioProject ID: PRJNA819417, BioSample accessions: SAMN26930854 and SAMN26930855. EST-SSRs from the unigenes of okra transcriptome were mined using a microsatellite identification tool (http://pgrc.ipk-gatersleben.de/misa/misa.html). The repeat sequence motifs included mono-, di-, tri-, tetra-, penta- and hexanucleotides with a minimum repeat number of 10, 6, 5, 5, 5 and 5, respectively. All the EST-SSRs developed in this study are freely available in GOAT (Genetic Output Analysis Tool) database (http://backlin.cabgrid.res.in/oyvmvtdb/). Batch Primer 3 (https://primer3.ut.ee/), an online web tool, was used to develop the primers for selected EST-SSR motifs (You et al. 2008). The following parameters were taken into account when designing the primers: (a) primer length of 20 to 25 bp, with 20 bp being the optimum; (b) PCR product size of 150 to 300 bp; (c) melting temperature (Tm) of 55°C to 65°C, with a difference of not more than 5°C between the Tm values of the forward and reverse primers, and (d) GC content of 40% to 60 %, with 50 % being the optimum. For this study, a total of 183 primer pairs were selected and synthesized at Sigma-Aldrich, USA (Datasheet S1).

Plant materials and DNA extraction

A total of 12 okra accessions were used to validate newly designed EST-SSR markers and to perform diversity analysis (Table 1). These accessions were grown on the farm of ICAR-Indian Agricultural Research Institute, New Delhi (28.6377'N, 77.1571'E) under insect-proof net house following the standard cultural practices during *kharif* of 2019-20. Genomic DNA was isolated from fresh leaf tissues of 15 days old okra plants by the CTAB method (Ahmed et al. 2013). The quality and integrity of DNA were checked on 0.8% agarose gel and quantified using Nanodrop. DNAs of all samples were then diluted to 30–50 ng/µL and stored at – 20°C for further PCR analysis.

Polymorphism survey of EST-SSR markers

The PCR mixture (10 µL) contained 5 µL of 2X PCR master mix (Promega, India), 1 µL primers (0.5 µL of each forward and reverse primers), 1 µL of genomic DNA (30–50 ng) and 3 µL of nuclease-free water. PCR program was performed at 94°C for 5 minutes of initial denaturation, followed by 35 cycles of 94°C for 1-minute denaturation, 55 to 65°C for 45 seconds of annealing and final extension for 7 minutes at 72°C using Master cycle ^{R}nexus (Eppendorf, India). PCR amplified products were separated on 2.5% agarose gel (0.6 µg/mL Et Br) and 50 bp DNA ladder (G- Biosciences, India) in 1X TBE buffer at a constant power of 120 V for about 2–3 hours. The gels were visualized and documented by a gel documentation system (Alpha Imager, India).

Functional annotation of EST-SSRs

Only polymorphic EST-SSR containing sequences were subjected to Blast2Go, an online functional annotation tool that performed a BlastX of EST sequences (with an e-value cut-off of 10^{-6}) against the NCBI non-redundant protein sequence database (nr) to understand the distribution of gene functions.

Data processing and genetic analysis

Only polymorphic SSR primer pairs were considered for data analysis. Consistent and reproducible SSR amplicons were scored as fragment size separately. The polymorphic information content (PIC) of polymorphic SSR primers was calculated by Cervous software (Kalinowski et al. 2007). Hierarchical clustering was created using DARwin6 (Perrier and Jacquemoud-Collet 2009) software Unweighted pair group method with arithmetic mean (UPGMA) to show a phylogenetic representation of genetic links indicated by Jaccard's similarity coefficient. The bayesian model-based clustering approach was used to analyse the population structure using STRUCTURE v2.3.4 software to determine the number of accessions in the k cluster and sub-clusters. The software program burn-in period was set at 100,000 lengths followed by 100,000 Markov Chain Monte Carlo (MCMC) repetitions. The optimum value of k for the whole population was determined according to the simulation method of the DeltaK (ΔK) value (Evanno et al. 2005) by using the web-based tool STRUCTURE HARVESTER v0.6.94. Genetic distance was used to calculate the principal coordinate analysis (PCoA). PCoA, number of observed alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), Shannon's information

Table 1. List of okra genotypes used in current study along with species and their source.

S. No.	Genotype	Species	Source
1	DOV-92	*A. esculentus*	IARI, New Delhi
2	DOV-66	*A. esculentus*	IARI, New Delhi
3	DOV-69	*A. esculentus*	VNMAU, Parbhani
4	Parbhani Kranti	*A. esculentus*	VNMAU, Parbhani
5	Pusa Sawani	*A. esculentus*	IARI, New Delhi
6	DOV-3	*A. esculentus*	IARI, New Delhi
7	IC- 599706	*A. moschatus*	NBPGR, New Delhi
8	IC- 470747	*A. manihot var. tetraphyllus*	NBPGR, New Delhi
9	Arka Anamika	*A. esculentus*	IIHR, Bangalore
10	SKM-17	*A. caillei*	NBPGR, New Delhi
11	IC-90476	*A. manihot var. tetraphyllus*	NBPGR, New Delhi
12	IC-470751	*A. ficulneus*	NBPGR, New Delhi

Table 2. Summary of the number and type of repeat units in okra EST-SSR loci identified from transcriptome assembly data.

SSR Motif	Total	SSR Motif	Total
Mono-nucleotide (71699)		Tetra-nucleotide (1545)	
T	37253	TTTTA	176
A	33560	AAAT	130
G	501	ATAA	72
C	385	AAAG	70
Di-nucleotide (10604)		TTTC	66
AT	3378	TTAT	63
TA	2648	others	968
AG	964	Penta-nucleotide (394)	
TC	896	GGAAT	24
CT	683	AAAAT	22
others	2035	TTTTC	20
Tri-nucleotide (14897)		TTTTA	14
TTC	928	AGAAA	11
TCT	730	others	303
ACC	565	Hexa-nucleotide (172)	
TTA	464	ATGTAT	13
CTT	344	GAAATA	6
ATG	307	TCCGGC	6
ATC	299	ACATGG	5
others	11260	others	142
Compound SSR	6913		
Grand Total= 1,06,224			

index (I) and analysis of molecular variance (AMOVA) was calculated using GenAlEx v.6.5 (Peakall and Smouse, 2006).

Results

Characteristics of the EST-SSR loci

A total of 1,06,224 SSR loci were identified from the transcriptome assembly data of okra. Among the SSR motifs identified in the *Abelmoschus* EST-SSRs, mononucleotide repeats were dominant with a frequency of 72.20% (71,699) followed by trinucleotides 15%, (14897) and dinucleotide 10.68%, (10604) repeats (Fig. 1). In contrast, the remaining motifs were noticeably rare. Among the mononucleotide motifs, the most common motif was 'T' (37,253), which accounted for 51.95% of all mononucleotide SSRs, followed

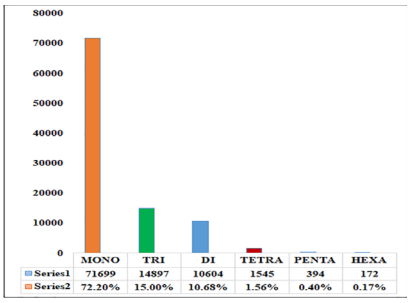


Fig. 1. Distribution of 1, 06,224 EST-SSRs motifs derived from the transcriptome assembly data of okra

by 'A' motifs (33560). Among the dinucleotide motifs, AT (3378) was the richest and most dominant motif, followed by 'TA' (2648) and 'AG' (964). Among the trinucleotides, 'TTC' (928) was the most abundant motif, followed by 'TCT' (730). In addition, 'TTTA' (176), 'GGAAT' (24) and 'ATGTAT' (13) were the major motifs of tetra, penta and hexa nucleotides, respectively (Table 2).

Validation of developed EST-SSR markers

In total, 183 primer pairs (156 di and 27 trinucleotides) were randomly selected for PCR validation purposes in *Abelmoschus* accessions. Overall, 163 primer pairs produced expected PCR amplicons and the amplification efficiency was 89%. Of the 163 primer pairs, 23 pairs showed polymorphisms among the 12 *Abelmoschus* samples, accounting for 14.1% polymorphism percentage (Fig. 2). Out of 23 polymorphic markers, dinucleotides account for 60%, followed by trinucleotides (40%). In dinucleotides, TA motifs amplified significantly, followed by GA motifs. In the trinucleotide motif, CTT repeats amplified in more numbers followed by AGA repeats.

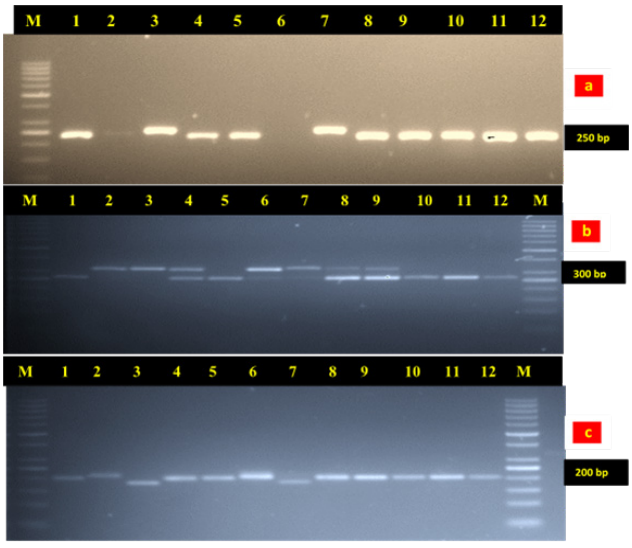


Fig. 2. Representative gel photograph showing PCR amplicons obtained from three okra SSR primers a) OSSR23 b) OSSR150 c) OSSR121 in 12 okra accessions (accessions: listed in table 1; L: 50 bp ladder)

Polymorphic analysis of 23 polymorphic EST-SSRs

Twelve okra accessions were used to assess the molecular diversity using 23 EST SSR primers which showed polymorphism (Table 3). The results showed that 23 markers were 100% polymorphic in the accessions examined. The number of observed alleles ranged from 2 to 4, with the highest value in OSSR138 (4) and a mean of 2.609 alleles. The number of effective alleles ranged from 1.087 (OSSR153) to 3.388 (OSSR138), with a mean of 1.768 alleles. Shannon's information index (I) varied from 0.173 (OSSR60) to 1.281 (OSSR138) with a mean of 0.623. Furthermore, the observed heterozygosity (Ho) and expected heterozygosity (He) ranged from 0.083 to 1.00 (mean 0.460) and 0.080 to 0.705 (mean 0.367), respectively. The PIC values for 23 SSR markers ranged from 0.077 for (OSSR60) to 0.649 for (OSSR138), with an average value of 0.319. These findings indicated a high level of informativeness within these EST-SSRs loci. Furthermore, 5 markers showed PIC values greater than 0.5. In addition, the Shannon index, which can also reflect the polymorphic value of markers, ranged from 0.173 (OSSR60) to 1.281 (OSSR138) and the mean value was 0.623 (Table 3).

Functional annotation of polymorphic EST-SSRs

A total of 57.17% (12) out of 23 polymorphic EST-SSRs were successfully annotated against the database. Phosphoglycerate mutase-like protein AT74H (OSSR23), Nitro reductase family protein isoform 1 (OSSR60),60S ribosomal protein L35(OSSR64) and Glycine tRNA ligase beta subunit (OSSR138) were some of the functions of polymorphic EST-SSRs (Table 4).

Genetic diversity analysis

Phylogenetic relationships between the cultivars were assessed by constructing a UPGMA dendrogram using similarity coefficients. The dendrogram divided the accessions into 2 main clades (I and II). Clade I is divided into 2 subclades where subclade I contain 7 accessions, of which 6 are breeding lines or released varieties like DOV-69,

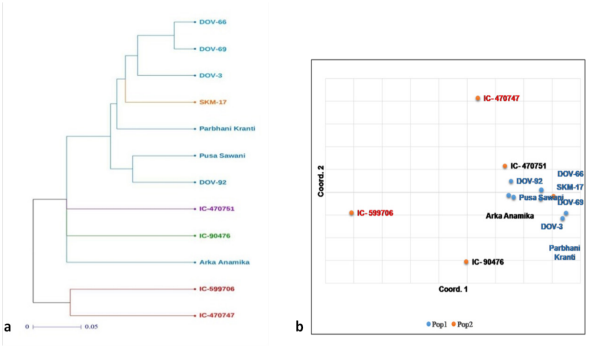


Fig. 3. (a) Hierarchical cluster dendrogram based on share allele distance showing genetic relationships among 12 okra accessions (listed in Table:1) using 23 polymorphic SSR primers. (b) Principal Coordinate Analysis (PCoA) of 12 okra accessions using Genealex v.6.5 into 2 groups

Table 3. Primers sequence information and characteristics of the 23 polymorphic EST-SSR markers used in the diversity analysis of 12 okra genotypes

S. No	SSR Name	SSR Repeat	SSR Motif	Forward Primer (5′-3′)	Tm(°C)	Reverse Primer (5′-3′)	Tm(°C)	Product Size (bp)	Na	Ne	I	Ho	He	PIC
1	OSSR23	p3	(CTT)25	GGTGAAGAGGCACTGCAGAT	60.03	CGAACCGACGATCAGTAGCA	59.90	239	2.000	1.385	0.451	0.333	0.278	0.239
2	OSSR67	p2	(TA)15	CAATTCGGCCTAACACGCTG	59.90	GGGAGGGGGCTTTCATTTCA	59.95	128	2.000	1.385	0.451	0.333	0.278	0.239
3	OSSR69	p2	(TA)19	CGGAGAAAGGAGGGGGAAAC	60.03	GGCTTCGTCGTCTATGGAGG	59.96	173	2.000	1.385	0.451	0.333	0.278	0.239
4	OSSR60	p2	(AG)18	GGATATGCTGGCAGGCTCAT	59.96	CATGCTGTTGGCGTATCTGC	59.97	158	2.000	1.087	0.173	0.083	0.080	0.077
5	OSSR64	p2	(TA)16	GGACGAAGATCAAGGGGCAA	60.03	CAAAGTTACCGGTGGTGCAC	59.69	153	2.000	1.280	0.377	0.250	0.219	0.195
6	OSSR103	p2	(TC)16	ATAAACGGAGAGTCGTCGCC	59.9	GCTTGAAGCAGCTTCCATGG	59.82	176	3.000	3.000	1.099	0.667	0.667	0.593
7	OSSR132	p3	(AGA)19	AGCTTCAATTGCTCGGTCCA	59.96	GCGGCACAAGAGTCCTTACT	60.03	264	3.000	1.767	0.723	0.583	0.434	0.369
8	OSSR116	p2	(TA)17	AGATGACCATCGGCAACAGG	60.10	AAGACTGGGCAAGGACGAAG	59.96	218	3.000	2.072	0.860	0.583	0.517	0.444
9	OSSR121	p2	(TC)16	CCCCCAATAGTTTCCCCCAC	60.03	GGGGGAAAATGCTCAAAGCC	59.75	202	3.000	1.291	0.456	0.250	0.226	0.212
10	OSSR138	p3	(TAT)17	CTTGCCGGACTGGGATTTCT	60.03	CCACCACCTGTAAAGGCCAA	60.17	175	4.000	3.388	1.281	0.750	0.705	0.649
11	OSSR139	p3	(CTT)16	TGTTTCTTTCACCTGCTGCA	58.23	GAAGCAAAGCTCACGTCAGC	60.11	216	3.000	2.667	1.028	0.833	0.625	0.545
12	OSSR143	p3	(GAG)15	GGAGGAGGAGGAGGAGGAAA	59.66	CAAGCAACTGTGGAGCGTTC	60.041	199	3.000	1.185	0.345	0.167	0.156	0.150
13	OSSR150	p2	(CT)13	GCCAAGCACCTCTTCCTTCT	59.963	AGTTTGGGTTCGTGAGTGGA	59.165	266	2.000	1.704	0.604	0.583	0.413	0.328
14	OSSR151	p2	(TG)12	CTCAGCCCACAAGAAGCTCA	59.964	TTGCACACACACACACACAC	59.76	246	3.000	1.811	0.778	0.417	0.448	0.397
15	OSSR153	p2	(AG)11	CTCCTTCCCTCTGCTCTCCT	60.032	CGTGGCAAATGCTTTGTGCT	60.598	215	2.000	1.087	0.173	0.083	0.080	0.077
16	OSSR169	p2	(GA)12	TTTCAGCACCTGTAGCCACT	59.235	CGACACGGCTGATGATTTGT	58.921	262	2.000	1.180	0.287	0.167	0.153	0.141
17	OSSR172	p2	(CT)11	GAGGAACACAGGTGGCTAGC	60.392	AGACCCACCATCGGAAGGTA	59.958	203	3.000	1.405	0.544	0.333	0.288	0.264
18	OSSR175	p2	(CT)12	AGTTCGTGCATGACCCATGT	59.963	AGAGAGGCGCCTTAATGAGC	59.894	280	2.000	1.180	0.287	0.167	0.153	0.141
19	OSSR178	p2	(GA)13	TCGGATCCAGCATCGATGTG	59.967	GGGCCGAATCCATTCCTGAA	60.107	217	3.000	1.291	0.456	0.083	0.226	0.212
20	OSSR145	p3	(GTT)19	CCATGGGCATGAGAGAGACC	59.892	GCACAAAGTTGGTCAACCCC	59.896	184	2.000	2.000	0.693	1.000	0.500	0.375
21	OSSR146	p3	(TAT)18	ACAATCGAGCTGCTTCTGCT	60.037	GGCATACGGTTAGGGGTACG	59.967	226	3.000	2.571	1.011	0.833	0.611	0.535
22	OSSR163	p2	(GA)12	CCCTTTCAGACACCCACCAT	59.594	CCGCTTAACCGGTCTCCATT	60.108	257	3.000	2.165	0.837	0.833	0.538	0.432
23	OSSR164	p2	(GA)13	TGTTTCAGCACCAGGTCCAG	60.179	GCAGTGTGTCGACCAAACAC	59.972	243	3.000	2.380	0.958	0.917	0.580	0.502
	Mean								2.609	1.768	0.623	0.460	0.367	0.319
	SE								0.122	0.137	0.065	0.062	0.041	

p^2 = Dinucleotide, p^3 = Trinucleotide, bp = Base pair, Tm = Primer melting temperature, Na = Number of alleles, Ne = Number of effective alleles, *I* Shannon index, *Ho*Observed heterozygosity, He = Expected heterozygosity and PIC = Polymorphic information content.

Table 4. List of polymorphic EST-SSR primers of okra with predicted function based on sequence homology

S. No	Nucleotide Type	Okra SSR ID	Gene id and function
1	Tri	OSSR23	XP_012439703.1 phosphoglycerate mutase-like protein AT74H
2	Di	OSSR64	XP_012447873.1 60S ribosomal protein L35
3	Di	OSSR67	KHG19352.1PHD finger 3
4	Di	OSSR116	XP_012446281.1 probable F-actin-capping protein subunit beta
5	Di	OSSR121	PPD75836.1hypothetical protein GOBAR_DD27238
6	Tri	OSSR138	KHG05639.1Glycine tRNA ligase beta subunit
7	Tri	OSSR145	XP_022716353.1uncharacterized membrane protein At1g16860-like
8	Tri	OSSR146	XP_017626125.1 dynamin-2A-like
9	Di	OSSR153	XP_022752529.1transcription factor E2FB-like isoform X2
10	Di	OSSR67	XP_022758327.1transcription factor UNE12-like isoform X1
11	Di	OSSR69	XP_016727940.1PREDICTED: uncharacterized protein LOC107939168 isoform X3
12	Di	OSSR60	EOY17057.1Nitroreductase family protein isoform 1

Table 5. Analysis of Molecular Variance (AMOVA) using 23 EST-SSR markers in 12 okra genotypes

Source	df	SS	MS	Est. Var.	% var.
Among Pops	1	11.833	11.833	0.906	12%
Within Pops	10	64.000	6.400	6.400	88%
Total	11	75.833		7.306	100%

**df* Degrees of freedom, *SS* Sum of square, *MS* Mean sum of square, *%Var.* percentage variance

DOV-66, DOV-3, Parbhani Kranti, Pusa Sawani, DOV-92 and one is *A. caillei* accession (SKM-17) (Fig. 3a). Subclade II consists of accessions like *A. angulosus* (IC-470751), *A. manihot* var. *tetraphyllus* (IC-90476) and Arka Anamika (Fig. 3a). CladeII contains accession of *A.manihot* var. *tetraphyllus* (IC- 470747)and another accession of *A. moschatus* (IC-599706). This result was confirmed by pairwise genetic distance calculated using GeneAlex software. The highest genetic distance of 25 was recorded between ParbhaniKranti and IC-599706 accessions (Supplementary Table S1). Delta K value reached its peak at 5, indicating a 5 structured population. Structure analysis showed an admixture in all the accessions studied (Supplementary Fig. S1). PCoA distributed all the accessions into 2 groups with a percentage of variation of 38.82, 24.45 and 16.22% in the first, second and third axis, respectively (Fig. 3b). Analysis of molecular variance (AMOVA) indicated 12 and 88% within and among populations, respectively (Table 5).

Discussion

This study's most common mononucleotide repeat pattern was A/T, identical to that found in *V. angularis* (Chen et al. 2015) and *A. thaliana* (Lawson and Zhang, 2006). Mononucleotides and trinucleotides repeat dominated in *Abelmoschus esculentus* in contrast to dinucleotide and trinucleotide repeats in many plants (Guo et al. 2017). Short motifs are abundant in species with long evolutionary history, implying that okra may have had a long evolutionary history. The percentage of trinucleotides in *Abelmoschus* was higher than in other plant species, possibly contributing to the okra's diversity. Furthermore, eukaryon trinucleotides may become abundant under mutation pressure and this diversity is beneficial for evolution (Metzgar et al. 2000).

A large percentage (80%) of newly developed 183 primer pairs successfully amplified the expected SSR fragments in a germplasm panel. More than 20 amplified EST-SSR primer pairs were polymorphic. The percentage of polymorphic primers was 14.1%, which is comparatively less may be due to fewer accessions studied. Low polymorphism of EST-derived SSR markers compared to anonymous SSR markers has been documented in grape (Strickler et al. 2012) and the conserved nature of coding areas of the genome explains it. Further, 20 of the 183 primer pairs failed to amplify and produced no band at all. Missassembly of the unigene with the SSR motif, or the presence of large introns in the genomic sequence of the gene carrying the SSR motif, could result in complete amplification failure (Ellis et al. 2006). When compared to randomly picked SSRs from other plant species, the polymorphism rate of the studied EST- SSRs was relatively high, which is likely owing to the amphipolyploid character of the plant, which allows for more gene sequence modifications than a homozygote diploid plant (Eujayl et al. 2002).

The majority of the amplified EST-SSRs were dinucleotides followed by trinucleotide repeating motifs, further, dinucleotide repeat SSRs have the greatest polymorphism rate (60%) among EST-SSRs. Polymorphic di-, tetra- and pentanucleotide repeat motifs would be more probable than tri- or hexanucleotide repeat motifs to produce frameshift mutations in coding areas (Schafleitner et al.

2013). Further, variations in EST allele sequences are more closely linked to protein function since they are responsible for significant modifications such as reading frame shifts, unexpected stop codons, and variable protein lengths and structures, as shown in other species (Wei et al. 2014). Moreover, the inclusion of coding region variants increases the functional use of these novel EST markers, primarily engaged in essential biological activities such as defence and response to abiotic and biotic stresses (Emebiri 2010). However, there is currently no clear understanding of how EST-SSRs in plant genomes might affect gene function or expression rate (Asadi and Monfared 2014).

We successfully developed 183 genic SSR markers. Most EST-SSRs were easily amplifiable and observable across all accessions, with just a few exhibiting no amplification, multiple bands, or scoring issues, most likely due to insertions/deletions or base mutations in primer binding regions or ploidy levels. On the other hand, the bulk of the EST-SSRs markers proved to be highly suited to discriminating across related *Abelmoschus* taxa, clearly identifying okra variants. The mean PIC (0.319) values fell into an intermediate category indicating promising levels of variability, as evidenced by their heterozygosity. This PIC content was lesser compared to 0.5 in a similar study on the okra crop where genomic SSRs were used (Ravishankar et al. 2018). Because EST-derived SSRs are connected with transcribed areas, they indicate lesser genetic diversity than genomic, neutral, and randomly chosen SSRs markers (Leonarduzzi et al. 2016). The structure of the population studied showed a higher mixture due to lesser individuals. The dendrogram analysis classified okra accessions into two major groups (Fig. 3). The accessions DOV-66, DOV-69, DOV-3, Parbhani Kranti, Pusa Sawani and DOV-92 which belong to *A. esculentus* species are grouped along with *A. caillei* accession SKM-17 which indicated their close genetic relatedness. *A. caillei* is also crossable with cultivated okra and has a lot of similarities in fruit and other morphological traits. It has been developed with cultivated okra as one parent. Therefore, it showed similarity with cultivated okra accessions and grouped in the same cluster. The presence of *A. esculentus* and *A. caillei* accessions in the same clade is reported in other studies also (Kumar et al. 2017; Das et al. 2022; Puneeth et al. 2023). Subclade II consists of wild accessions like *A. angulosus* (IC-470751), *A. manihot* var. *tetraphyllus* (IC-90476) and Arka Anamika (Fig. 3). The presence of Arka Anamika along with wild-type accessions might be due to its interspecific origin from *A. manihot* var. *tetraphyllus*. Clade II contains an accession of *A. manihot* var. *tetraphyllus* (IC- 470747) and another accession of *A. moschatus* (IC-599706). Largely the studied accessions are classified into cultivated and wild types where *A. caillei* accession is clustered with cultivated accessions. The above results are in concurrence with our previous report (Das et al. 2022; Puneeth et al. 2023) where SSR markers have classified cultivated and wild types separately. This result was also verified with the analysis of PCoA which also formed two clear coordinates.

The unique EST-SSR markers developed in this study will be a robust molecular tool for germplasm identification, genetic diversity analysis, genetic relationship investigations and comparative mapping in okra. Furthermore, these markers may help to create genetic linkage maps, which are required for the establishment of marker-trait associations and marker-assisted crop improvement.

Supplementary material

Supplementary Table S1 and Supplementary Fig. S1 are provided which can be accessed online www.isgpb.org.

Authors contribution

Conceptualization of Research (RKY and SL); Designing of experiments (RKY and SL); Contribution of the experimental materials (RKY and SL); Execution of the field/lab experiments and data collection (PPV, AT); Analysis of the data and interpretation (PPV, MAI, AT); Preparation of the manuscript (PPV, SL, RKY, HC).

Acknowledgment

The financial support from the National Agricultural Higher Education Project (NAHEP) - Centre for Advanced Agricultural Science and Technology (CAAST) project is gratefully acknowledged. The authors are also thankful to the Indian Council of Agricultural Research (ICAR), Ministry of Agriculture and Farmers' Welfare, Government of India, for providing a computational facility with the Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility at ICAR-IASRI, New Delhi, India.

Data availability statement

The data that support the findings of this study are openly available in NCBI at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA819417, Accession: PRJNA819417, ID: 819417. EST-SSRs developed in this study are freely available in GOAT (Genetic Output Analysis Tool) database (http://backlin.cabgrid.res.in/oyvmvtdb/).

References

Ahmed N., Nawaz S., Iqbal A., Mubin M., Butt A., Lightfoot D. A. and Maekawa M. 2013. Extraction of high-quality intact dna from okra leaves despite their high mucilage. Biosci. Methods., **4**: 4. http://dx.doi.org/10.5376/bm.2013.04.0004

Asadi A. A. and Monfared S. R. 2014. Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. Mol Genet Genom., **289**(4): 625-640. http://dx.doi.org/10.1007/s00438-014-0839-z

Barman P., Choudhary, A. K and Geeta, R. 2017. A modified protocol yields high-quality RNA from highly mucilaginous Dioscorea tubers. 3 Biotech, **7**(2): 1-5.

Chen H., Liu L., Wang L., Wang S.,Somta P. and Cheng X. 2015.

Development and Validation of EST-SSR Markers from the Transcriptome of Adzuki Bean (*Vignaangularis*). PLoS One., **10**(7): e0131939. doi:10.1371/journal.pone.0131939

Cho Y. G., Ishii T.,Temnykh S., Chen X.,Lipovich L.,Mccouch S. R. and Cartinhour S. 2000. Diversity of microsatellites derived from genomic libraries and gen bank sequences in rice (*Oryzasativa* L.). Theo. Appl. Genet., **100**(5): 713-722.

Das A., Yadav R. K., Choudhary H., Lata S., Singh S., Kumar C. and Talukdar A. 2022. Population structure, gene flow and genetic diversity analyses based on agro-morphological traits and microsatellite markers within cultivated and wild germplasms of okra (*Abelmoschusesculentus* (L.) Moench.). Genet. Res. Cr.Evol., **69**(2): 771-791.http://dx.doi.org/10.1007/s10722-021-01263-9

Ellis J. R., Pashley C. H., Burke J. M. and McCauley D. E. 2006. High genetic diversity in a rare and endangered sunflower as compared to a common congener. Mol. Eco., **15**(9): 2345-2355. http://dx.doi.org/10.1111/j.1365-294X.2006.02937.x

Emebiri L. C. 2010. An EST-SSR marker tightly linked to the barley male sterility gene (msg6) located on chromosome 6H. J. Her., **101**(6): 769-774.http://dx.doi.org/10.1093/jhered/esq083

Eujayl I.,Sorrells M. E., Baum M., Wolters P. and Powell W. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theo. Appl. Genet., **104**(2): 399-407.http://dx.doi.org/10.1007/s001220100738

Evanno G., Regnaut S. and Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Eco., **14**(8): 2611-2620 http://dx doi.org/10.1111/j.1365-294X.2005.02553.x

Guo R., Landis J. B., Moore M. J.,Meng A., Jian S., Yao X. and Wang H. 2017. Development and application of transcriptome-derived microsatellites in *Actinidiaeriantha*. Front. Pl. Sci., **8**: 1383. http://dx.doi.org/10.3389/fpls.2017.01383

Kalinowski S. T., Taper M. L. and Marshall T. C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. eco., **16**(5): 1099-1106.http://dx.doi.org/10.1111/j.1365-294X.2007.03089.x

Kumar S., Parekh M. J.,Fougat R. S., Patel S. K., Patel C. B., Kumar M. and Patel B. R. 2017. Assessment of genetic diversity among okra accessions using SSR markers. J. Plant. Biochem.Biot., **26**: 172–178

Lata S., Yadav R. K. and Tomar B. S. 2021. Genomic tools to accelerate improvement in okra (*Abelmoschus esculentus*). DOI: 10.5772/intechopen.97005

Lawson M. J. and Zhang L. 2006. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Geno.bio., **7**(2): 1-11.

Leonarduzzi C.,Spanu I.,Labriola M., González-Martínez S. C.,Piotti A. and Vendramin G. G. 2016. Development and characterization of three highly informative EST-SSR multiplexes for *Pinushalepensis* and their transferability to other Mediterranean pines. Pl. Mol. Bio. Rep., **34**(5): 993-1002. http://dx.doi.org/10.1007/s11105-016-0980-4

Metzgar D., Bytof J. and Wills C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Geno. Res., **10**(1): 72-80.

National Horticulture Board., data 2019-20 (second advance estimates). http://nhb.gov.in/StatisticsViewer.sWoUJibk35dW2g36TUJWAoZqESmAYFi7h2irlsmjlINTcFl1rG/kLbq8ZQbWUvuM

Peakall R. and Smouse P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Eco. Notes., **6**(1): 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

Perrier X. and Jacquemoud-Collet J. P. 2009. DARwin software: Dissimilarity Analysis and Representation for Windows (version 5.0. 157).

Puneeth P. V., Lata S., Yadav R. K., Wankhede D. P., Tomar B. S., Choudary H. and Talukdar A. 2023. Exploring the genetic diversity using CAAT box-derived polymorphism (CBDP) and start codon targeted (SCOT) markers in cultivated and wild species of okra (*Abelmoschus esculentus* (L.) Moench.). Genet. Res. Crop. Evol., **70**(3): 749-761.

Ravishankar K. V., Muthaiah G., Mottaiyan P. and Gundale S. K. 2018. Identification of novel microsatellite markers in okra (*Abelmoschus esculentus* (L.) Moench) through next-generation sequencing and their utilization in the analysis of genetic relatedness studies and cross-species transferability. J. gen., **97**(1): 39-47.

Schafleitner R., Kumar S., Lin C. Y.,Hegde SG., Ebert A. 2013. The okra (*Abelmoschusesculentus*) transcriptome as a source for gene sequence information and molecular markers for diversity analysis. Gene., **517**(1): 27-36.http://dx.doi.org/10.1016/j.gene.2012.12.098

Scott K. D., Eggler P., Seaton G., Rossetto M., Ablett E. M., Lee L. S. and Henry R. J. 2000. Analysis of SSRs derived from grape ESTs. Theo. Appl. Gen., **100**(5): 723-726.http://dx.doi.org/10.1007/s001220051344

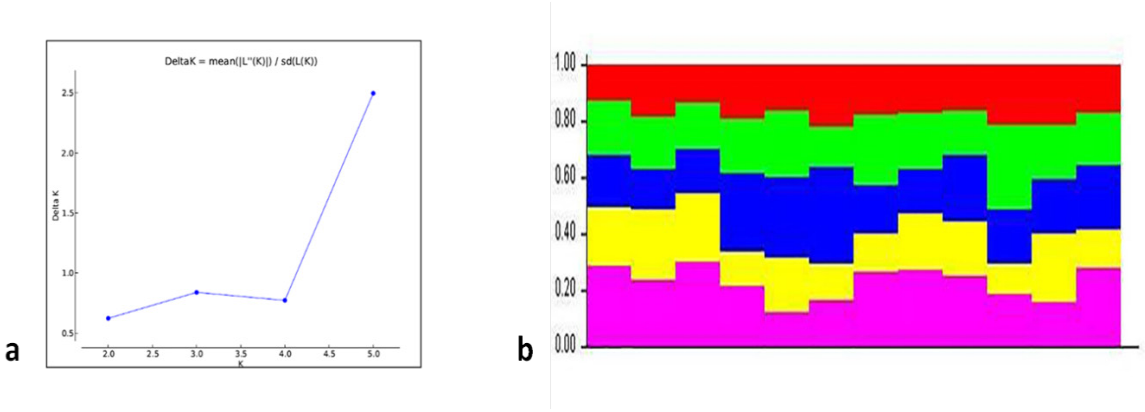
Strickler S. R., Bombarely A. and Mueller L. A. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. American J. Bot., **99**(2): 257-266.

Wei L., Li S., Liu S., He A., Wang D., Wang J. and Wu X. 2014. Transcriptome analysis of *Houttuyni acordata* by Illumina paired-end RNA sequencing and SSR marker discovery. PloS one, **9**(1): e84105.http://dx.doi.org/10.1371/journal.pone.0084105

You F. M., Huo N., Gu Y. Q., Luo M. C., Ma Y., Hane D. and Anderson O. D (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC bio. Info., **9**(1): 1-13.

Supplementary Table S1. Genetic distance among 12 okra genotypes for Principal Coordinate Analysis (PCoA)

DOV-92	DOV-66	DOV-69	Parbhani Kranti	Pusa Sawani	DOV-3	Arka Anamika	IC-599706	IC-470747	SKM-17	IC-90476	IC- 470751	
0												DOV-92
11	0											DOV-66
11	2	0										DOV-69
13	7	9	0									ParbhaniKranti
7	9	9	10	0								PusaSawani
12	6	5	9	7	0							H-3
18	12	12	15	17	15	0						ArkaAnamika
18	20	20	25	17	24	22	0					IC- 599706
16	13	15	20	14	20	19	22	0				IC- 470747
8	7	7	10	10	10	15	23	18	0			SKM-17
17	12	12	15	13	14	16	13	24	14	0		IC- 90476
12	9	11	16	16	16	14	20	14	8	15	0	IC- 470751



Supplementary Fig. S1. Population structure analysis using STRUCTURE HARVESTER v.0.6.94 based upon 23 EST SSR loci.a. ΔK of SSR primers calculated as ΔK = m|L0 0 (K)|/s [L(K)], reached peak at k = 5. b. Population structure bar plot of 12okra accessions based on 23 EST SSR markers