Indian J. Genet., 78(3): 348-356 (2018)
DOI: doi.org/10.31742/IJGPB.78.3.12

Genetic diversity computation in sesame genotypes using morphological traits and genic SSR markers

Adil Iqbal*, Udayan Bhattacharyya[1], Rumana Akhtar and Tapash Dasgupta[2]

Department of Genetics and Plant Breeding, Institute of Agricultural Science, University of Calcutta, Kolkata; [1]Bengal Institute of Technology, Tech Town, Dhapa Manpur, Hadia, Kolkata, West Bengal

(Received: December 2017; Revised: July 2018; Accepted: August 2018)

Abstract

Sesame (*Sesamum indicum* L.), the most ancient and important oilseed crop, with rich source of protein, high-quality seed oil and many antioxidant properties is extensively grown in India. In the present study, 235 genic markers developed through data mining of expressed sequence tags from the NCBI database were used to assess the molecular genetic diversity at molecular level of 70 genotypes, belonging to different eco-geographical regions of the world. Of the newly developed markers, 35 (15%) showed polymorphism with a total of 151 alleles identified across genotypes. The alleles detected by the markers varied from 3 to 6 with an average of 4.36 along with 69% of average genetic diversity. The polymorphic information content (PIC) ranged from 0.36 to 0.82, with an average of 0.61. Neighbor-Joining (NJ) analysis revealed the five major groups, and clustering was independent to geographic origin. By using UPGMA, 70 genotypes at 70% similarity coefficient among the morphological traits were classified into four distinct clusters and the clustering pattern was different than based on molecular markers.

Key words: Expressed sequence tags, simple sequence repeats, clustering, sesame

Introduction

Sesame is one of the most ancient and important oilseed crops grown and used by mankind. It was cultivated and domesticated in the Indian subcontinent during Harappan and Anatolian eras over 4,000 years ago (Bedigian and Van der Mesen 2003). Due to stability of its healthy oil, easiness of extraction and resistance to drought, Sesame was popular in the ancient world. Sesame is considered as a nutritious oilseed crop being rich source of protein (18-25%), carbohydrate (13.5%), minerals and healthy polyunsaturated fatty acid (Bedigian 1985). Sesame oil is favoured as a media of cooking by Indians and Africans. Presence of sesamol, a unique anti-oxidant and more poly-unsaturated fatty acid, have made it to queen of oilseed crop (Ashri 1998; Fukuda et al. 1994). In spite of being the first oilseed crop known to man and momentous history, sesame is typically a neglected crop. As a consequence of this, the use of molecular techniques for the improvement of sesame is very limited, and only a few reports (Ramprasad et al. 2017) are available on the use of molecular markers in breeding program. Simple-sequence repeats (SSRs) occur frequently in both protein-coding and non-coding regions of eukaryotic genome (Levinson and Gutman 1987). Mispairing during replication is primarily the reason for slipped-strand and variation in SSR length (Levinson and Gutman 1987). Although the utility of SSR markers has been well-established, their development can be more time consuming and costly than EST-SSRs development (Squirrell et al. 2003). In order to circumvent these problems, the publicly available genomic resources can be exploited for the development of gene- based SSR markers that are more likely to be transferable across taxonomic boundaries (Ellis and Burke 2007). The National Center for Biotechnology Information (NCBI) EST database (dbEST; Boguski et al. 1993) contains an ever-increasing number of these 'single-pass' cDNA sequences, meaning that the resources necessary for the efficient development of large numbers of so-called EST-SSRs already exist for a wide variety of taxa. In general, EST-SSRs have been found to be more

*Corresponding author's e-mail: adil_kol08@yahoo.co.in
[2]Present address: School of Agriculture and Rural Development, Ramakrishna Mission Vivekananda University, Narendrapur campus, Kolkata 700 103
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by indianjournals.com; www.isgpb.org

transferable across taxonomic boundaries than are SSRs (Chagne et al. 2004; Liewlaksaneeyanawin et al. 2004; Gutierrez 2005; Pashley et al. 2006), and reports of EST-SSR transferability have become increasingly common. Transferability is particularly prospective in plants, (Bandopadhyay et al. 2004). At present microsatellite markers especially expressed-sequence tag (EST) databases have become an increasingly valuable resource for SSR marker development (Pashley et al. 2006; Yi et al. 2006; Becher 2007; Heesacker et al. 2008). EST databases are now available for around 150 economically and biologically important plant species (Heesacker et al. 2008), and SSRs are abundant within these collections of sequence (Ellis and Burke 2007). A key advantage of these EST-SSRs is that they are often more transferable across species as compared to so-called 'anonymous' SSRs from non-coding sequences (Yu et al. 2004; Pashley et al. 2006), thereby facilitating comparative genetic analyses. This higher level of cross-taxon marker portability facilitates comparative genetic analyses.

Materials and methods

Sesame EST sequences were downloaded in FASTA format from the National Center for Biotechnology Information (NCBI) data base. (https://www.ncbi.nlm.nih.gov/) as of May 2015. VecScreen (https://www.ncbi.nlm.nih.gov/tools/vecscreen) and trimmest (http://embossgui.sourceforge.net/ demo/trimest.html) were used to remove vector contamination and poly A tail.

Microsatellites/SSRs were identified using the Simple Sequence Repeat Identification Tool (SSRIT) program (www.gramene.org/db/searches/ssrtool). SSR-containing EST sequences were clustered using Cap3 program to identify the non-redundant EST and were used for homology search using the Basic Local Alignment Search Tool (BLAST) (http://www.gramene.org/db/searches/blast). BLAST hits having the maximum score was selected from "top", "middle", "bottom", "region between top and middle" and "region between middle and bottom". A total of 235 SSR-containing non-redundant sesame EST sequences were selected and primer pairs were designed for selected 235 non-redundant SSR-containing sesame EST sequences using Primer3 program.

Seventy sesame genotypes were collected from diverse eco-geographical regions (Bangladesh, USA, Bulgaria and different parts of the India) of the world. Fifteen newly developed genotypes were also included in this study. These genotypes were planted in *kharif* 2015 season following Randomized Block Design (RBD) with 3 replications having 10cm spacing between plants and 40cm between rows. The experiment was carried out at Agricultural Experimental Farm, University of Calcutta, Baruipur during *rabi*-summer season, 2015. The farm is situated at an elevation of 10 meter above sea level, at approximately $22^0 51^/$ N latitude and $88^0 24^/$ E longitude. Statistical cluster analysis of 70 genotypes with respect to eleven morphological characters, namely, plant height (cm), days to 50% flowering (days), days to maturity, primary branches/plant, secondary branches/plant, inter node length (cm), no. of capsules/plant, capsule length, capsule breadth, 1000 seed weight (gm) and yield/plant (gm) was done by Average Linkage Between Groups (in other words UPGMA) Method. The divergence between accessions was evaluated using a Euclidean distance dissimilarity matrix. Euclidian distance between lines were calculated and UPGMA cluster analysis was performed with the help of IBM-SPSS software (Version: 16.0) for dendrogram formation.

EST-SSR analysis

DNA extraction was mainly carried out from young leaves of sesame plants around 8-12 days old (2-4 leaves stage). Extraction of DNA was done as described by Saghai-maroof et al. (1984). The purified DNA was quantified with Nanodrop Lite (Thermo Scientific, USA). A set of 235Est-SSR microsatellite markers were selected for the study. Each 25µl reaction mixture containing 0.2 mML^{-1} SSR primers, 0.2 mM of each dNTPs, 2 mML^{-1} $MgCl_2$, 1 X PCR buffer and 0.5 unit Taq polymerase and 50 ng sample DNA. The program for SSR and Est-SSR was carried out in a DNA thermo cycler. The program was set up as follows: denaturation at 95°C for 5 min, 40 cycles 2 min at 95°C, 45 s at on annealing temperature of the primer, 72°C for 30 second with final extension at 72°C for 5 min. Amplified PCR products were separated on 2% agarose gel stained with ethidium bromide. A 50 base pair ladder marker was used to estimate PCR fragment size.

Genetic diversity analysis using SSR profiles

The analysis was done with the help of 1-0 matrix. Those band were selected which was clear and visible for data analysis. Presence or absence of bands related to each primer was scored as '1' or '0' respectively. PIC was calculated by Nei's statistic (Nei 1973).

Effective allele per locus (Aep) was computed as suggested by Weir (1990). Expected heterozygosities (HE) calculated with the help of the formula [N(N+1)] / 2 Where n is number of alleles. All the data were analyzed for multivariate using the software Darwin Ver 6.0 (Sneath and Sokol, 1973). By using unweighted pair group method with arithmetic average (UPGMA) transformed data and the information was epitomized in dendrogram and Shan clustering program in Darwin Ver. 6.0. The un-weighted neighbour joining (NJ) method as implemented in DARwin 6.0 software was used to generate.

Results and discussion

The highest Euclidian distance was estimated between the genotypes Saheb and VR-1 (37.36). On the contrary, EC- 303432 (USA) and RT-348 (Rajasthan) were very close to each other with lowest Euclidean distance value of 4.56. By using UPGMA, 70 genotypes at 70% similarity coefficient were classified into four distinct clusters (CL) (Table 1). Most of the exotic collections were present in Cluster I, which was sub divided into Cluster IA consisting of 18 genotypes and Cluster IB having 20 genotypes. Overall, most of the CU developed genotypes were placed in Cluster-IV with 24 genotypes distributed across cluster IVA containing 3, IVB having 8 and IVC consisting of 13 genotypes. Cluster II and Cluster III were the smallest comprising of 6 and 2 genotypes, respectively. It was interesting to note that exotic collections collected from USA, Bangladesh and Bulgaria were mostly placed in CL I on the other hand, the CL IV contained the Indian genotypes developed from Calcutta university. Detailed study of the cluster composition revealed that each of the two clusters consisted of varieties belonging to different origin, i.e. from different states of India and also from different countries. This indicates that genetic divergence of genotypes is independent of geographic origin as has also been reported earlier by Banerjee and Kole (2009) and Ramprasad et al. (2017) in sesame and Goyal and Bisen (2017) in niger. The hybridization between the genotypes of different clusters is expected to get desirable segregants.

Of 44010 EST sequences available for sesame in the NCBI database, 9,106 EST sequences found to contain microsatellites. Among SSR containing EST sequences, only 7.9% belonged to Class I (motif length <20 nucleotides) and remaining were belonged to class II type (motif length < 20 nucleotides). Cap3 program identified 2184 as non redundant EST sequences containing microsatellites motifs. In total, only about 4.9% of sesame EST sequences were found with microsatellite motifs. The AG class of di-nucleotide repeats which includes CT, GA and TC repeats and the CCG class of tri-nucleotide repeats which includes CCG, GGC, CGC, GCC, GCG and CGG were the most abundant in sesame. The frequencies of tri-nucleotide repeats belonging to ATT (Isoleucine) and AAT (Asparagine) classes were very low (0.57% and 0.27%, respectively).

To identify candidate sequences from 2184 non-redundant sesame ESTs containing SSR sequences, about 323 EST sequences (14.76%) (selected based on GC content) were BLAST searched. Among the selected 300 candidate sequences, 241 sequences (80.3%) had putative annotations. The majority of these are classified as transcription factors or DNA binding proteins. Among the 300 selected sesame ESTs containing SSRs, 33 were di-nucleotide repeats (11%); 205 were tri-nucleotide repeats (68.3%); 38 were tetranucleotide repeats (12.66%) and 24 were penta-nucleotide repeats (8%). The most abundant repeats among the selected EST sequences were AG and CCG classes. A total of 39 of these EST-SSRs (13%) were Class I and 236 EST-SSRs were Class II, while the remaining 25 EST-SSRs were only 10 nucleotides (five di-nucleotide units) in length. The selected 300 candidate SSR-containing non-redundant sesame EST sequences were used to design primer pairs. Only 235 primers pairs were selected on the basis of GC content. These primers were named as CALCUTTA UNVIVERSITY EST SSR Primer (CU-ESSR) pairs with three digit serial number.

Seventy diverse genotypes collected from different part of the world were screened with all designed primer pairs. Out of 235, newly developed primer pairs, 157 (66.8%) amplified the template DNA and 126 (80.2%) of them produced simple and easy to score amplification products. Most of the primer pairs had amplicons sizes in the range of 150-200 bp. Of the 235 primer pairs, 35 primer pairs (15%) detected polymorphism in diverse collected genotypes of sesame.

The present investigation distinctly delineated that primers were efficient enough to distinguish the sesame genotypes (Table 2). A total of 151 alleles identified across genotypes with an average of 4.34 alleles per locus. The lowest band (68 bp) was observed in CU-ESSR-10 whereas highest band size (66 bp) was observed in the CU-ESSR-11. The number

Table 1. Clustering of genotypes based on morphological traits

Cluster	Sub cluster	No. of genotypes	Name of genotypes
I	I-A	18	RT348, IC-2621694, SAVITRI, EC-303432, EC-303435(4), IC-204063, CUMS 20, EC-182832(26), EC-334988-(3), EC-204704(44), EC-303442(32) EC-334973(38), TMV4, RT-54, EC-310448(36), CUMS-09, UMA
	I-B	20	TILLOTTAMA, OSC-593, NIC 8316, V-10, EC-303431, CUHY-13, IC-14053, EC-334971(23), EC-310448(36), EC-164966(52), RAMA, EC-335004, EC-335004(34), B9, EC-41923(49), EC-303431, EC-164966(52), EC-303433(17), HUMRA, TKG-22
II	II-A	05	GT-10, NIC-16434, CUMS-17, AMRIT, CUHY-24
	II-B	1	VR1
III		2	GT-2, CR-11A
IV	IV-A	3	CUMS-19, Gujrat local, ASSAM local
	IV-B	8	IC-152485, IC-131490, IC-96230, CUMS-11, CR-11A, CUMS-9, EC-334962(28)
	IV-C	13	TKG-352, CUHY-23, SAHEB, TKG-355, NIRMALA, IC-14331, CUHY-27, IC-43033, CUHY-36, CUHY-45, IC-204159, V-12, CUHY-57

of alleles ranged from to 6 (Table 2) These numbers were much higher than the average 2.0-5.5 alleles per locus for various classes of microsatellites reported by Cho et al. (2000) and Yu et al. (2003), using different parental lines for an international molecular breeding programme. The mean allele number found in this study, 4.34, was slightly lower than 7.6 obtained by Cho et al. (2011), higher than 3.11 reported by Badri et al. (2014) and 3.37 reported by Pandey et al. (2015) but very similar to 4.7 obtained by Dixit et al. (2005) and 4.15 recorded by Dossa et al. (2016) in sesame. This variation could be the result of inclusion of wild genotypes by Cho et al. (2011) resulting in more alleles, some of which novel, whereas only domesticated genotypes of different eco-geographical genotypes was used in the present study. In the present study medium to low magnitude of diversity was observed this is due to the fact that relatively narrow genetic basis exists in sesame from different sesame centers because ecological and geographical factor did not play an important roles in sesame evolution (Laurentin and Karlovsky 2006).

Although the genotypes from different ecological and geographical regions were evaluated but wide genetic variability as expected was not found. In fact expected variability within the working germplasm is less. It is not so much different within the collection from different countries. Normal expectation is the increase of variability due to geographic isolation. It can be concluded that narrow genetic variability can occur when a common gene pool is shared between the lines. Here in this study, shared gene pool might be the reason of displaying less diversity.

It was revealed from the molecular data (Table 2) that the EST-SSR primers CU-ESSR-14 (Fig. 1) followed by CU-ESSR- 12 and CU-ESSR- 27 and CU-ESSR-28 respectively were more polymorphic than others McCouch et al. (2002) concluded that microsatellite markers demonstrate high genetic diversity per locus because of their multi-allelism. The polymorphic information content (PIC) ranged from 0.36 (CU-ESSR-18) to 0.82 (CU-ESSR-26 (Table 1) with a low average of 0.607, which can be partly explained by the fact that all microsatellites were developed from ESTs, which are generally less variable than unexpressed regions of the genome. This value of PIC is slightly lower than 0.72 which was reported by Yepuri et al. (2013) but higher than reported by Anandan et al. (2017) and Dar et al. (2017). Based on PIC values, CU-ESSR-26 (0.82) appeared to be the most informative marker. These markers are expected to provide a valuable resource for diversity analysis of sesame. Comparison of molecular marker information indicates that the genetic basis was narrowed down and the genetic diversity was declining during domestication and selection of mordent cultivars from landraces.

NJ analysis revealed the clustering to be independent of geographic origin of the genotypes (Table 3). However, for better understanding, clusters were divided into five major groups. These major groups were dived into sub groups. Cluster I contain 22 genotypes, of which about 50% are exotic. Cluster

Table 2. Allele size, number of allele and genetic diversity among 70 SSR loci

Name	Primer sequence	No. of allele	Size ranges of allele(bp)	PIC value	H_E	Effective allel (Aep)
CUE-SSR -01	F:AGGAAGTGGTGTAACTGTTGA R:TCTTCTCTTCCTCAGTCTCCT	5	282-374	0.46	0.77	4.35
CU-ESSR -02	F:AAGAAAGCTAAGAAGGCAGAG R:GCTTGATAGAGAAGTTACGACA	6	154-205	0.71	0.80	5
CU-ESSR -03	F:TTTATTTATTACCCTCTCCTCT R:CTGAACAACAAACAAAGAAGG	4	299-357	0.54	0.66	2.94
CU-ESSR -04	F:AGCCAGATAAGTTTAGCATGA R:ATTCATTCATTCACTGCTGTT	4	186-222	0.62	0.70	3.33
CU-ESSR -05	F:CCATACACATCCGACGTATTA R:CAACTCCAGCATCTACAACTC	4	131-189	0.69	0.70	3.33
CU-ESSR -06	F:TGTTATACTCAGCCAGTCACC R:TGGTTGGGTTGATATAGTAGG	6	152-281	0.71	0.81	5.26
CU-ESSR -07	F:AATTACCCACAAAAAGAATC R:AATTACCCACAAAAAGAATCC	5	184-232	0.60	0.72	3.57
CUE-SSR -08	F:ACTCTCCTCTTCAACCTTCAC R:GAAGAGGTGGAGGAATTACG	5	103-250	0.63	0.79	4.76
CU-ESSR -09	F:GATTGTTGAAGAAGAAGGTGA R:TGGCTGAATCTTGAAAATCTA	3	377-555	0.67	0.63	2.70
CU-ESSR -10	F:ACGAGAAAAATGGTTGTGTAA R:TTTACTGGTGTGTGTGTGTGT	5	68-122	0.56	0.79	4.76
CU-ESSR -11	F:ATGCAAAAATACACACACACA R:CGCCACATTTTATGCTTATT	5	211-666	0.57	0.72	3.57
CU-ESSR -12	F:TTGTCAAAGTCAAGAGTTCGT R:TCTTATCCTTGCTAACAGCAG	4	143-194	0.66	0.71	3.45
CU-ESSR -13	F:GCAAAGGTAGAATTGAACAAG R:ATTAGCTTTCTTCAACCCTCT	4	86-215	0.6	0.72	3.57
CU-ESSR -14	F:CCCGGCTTTTCTTCTACTACT R:GGTTGTAGGTGTTGTTGTAGG	3	147-184	0.61	0.79	2.56
CU-ESSR -15	F:CGGTAAGATGACAAACAAGTC R:ATAGTCACCTGTTTTGAATGC	3	122-255	0.57	0.67	3.03
CU-ESSR -16	F:GAAGCCGTTGAATAGAAGAAT R:TACAGCAATGATGAAACAACA	5	171-232	0.43	0.72	3.57
CU-ESSR -17	F:CTACTGCACATTTCCTCATTC R:AGAAATATGATCCCCACTAGC	5	81-204	0.53	0.73	3.70
CU-ESSR -18	F:TGGAAAATTAACTCACAAAGG R:GAATGAGGAAATGTGCAGTAG	5	90-211	0.36	0.38	1.64
CU-ESSR -19	F:GGGGCTCATAATTCTCTTTT: R:CTGTGGGCAATAACAGTAAGA	4	202-255	0.66	0.72	3.45
CU-ESSR -20	F:TCCTCTTTCACTCTCTTTTCC R:CATAAGAGCAAACAGGATGAG	3	161-204	0.72	0.77	3.82
CU-ESSR -21	F:GGGGACTCTTCTTCTTCTTCT R:GAGGACCGTTGTAGTATCCTT	4	169-196	0.71	0.72	3.57
CU-ESSR -22	F:ACCAAGAACCACAACAAATC R:CCTAAAACCACAATCTGAGAA	5	106-198	0.74	0.79	3.70
CU-ESSR -23	F:GGGCTAATGTATCAGAGCTAA R:GCTCTAAATTGGTGATTTTCA	4	165-281	0.66	0.73	3.12

CU-ESSR -24	F:GATTGGCAGATACCTCATACA R:CAGGTTCAATCAAACATCAAT	5	121-191	0.75	0.80	4.20
CU-ESSR -25	F:AGACAGACATCCTCCTTTCTC R:GAGACAGAGAAGCAAGTTGAA	3	144-181	0.69	0.75	2.83
CU-ESSR -26	F:CTCTCTGACCTTTTCTTTTCC R:AGATAGGCTGTTGTTCCTTTC	4	207-261	0.82	0.86	3.03
CU-ESSR -27	F:CATTCAGTTCTTCTCTTCTGC R:GGAGAGATCGAGAACCAGTAT	4	158–172	0.80	0.847	3.12
CU-ESSR -28	F:TTCGTCATTTCTATCATTTCC R:AGGACTTCCATTGTTCATCTT	6	217–231	0.80	0.858	4.03
CU-ESSR -29	F:ACCTCGAGTCTTTCACTCTTC R:GAGGACCGTTGTAGTATCCTT	5	263–275	0.740	0.779	3.57
CU-ESSR -30	F:ATCTTCGCTCCTTCTCTGTT R:GGTGATGAGAGCTGAGTAGTG	3	289–307	0.340	0.437	4.76
CU-ESSR -31	F:GATGATGACGATGAAGAAGAG R:GGAGCTAAAGGATTGTCATCT	4	204–218	0.617	0.654	2.70
CU-ESSR -32	F:AATTACCCACAAAAAGAATCC R:ACTTCCTCATTCCTGGTAAAT	3	218–245	0.460	0.484	3.76
CU-ESSR -33	F:TAATTCGCAAGGATTAAGAGA R:GGTCCATGTGATATTCGTGTA	6	221–259	0.700	0.734	4.57
CU-ESSR -34	F: CTATGACGGGCAAGATTTAC R:AGAGAAAGGCTGAGAGAAAAA	3	181-228	0.52	0.55	4.18
CU-ESSR -35	F: TGTTATACTCAGCCAGTCACC R:TGGTTGGGTTGATATAGTAGG	4	119-217	0.62	0.67	3.71
MEAN		4.31		0.61	0.71	3.63

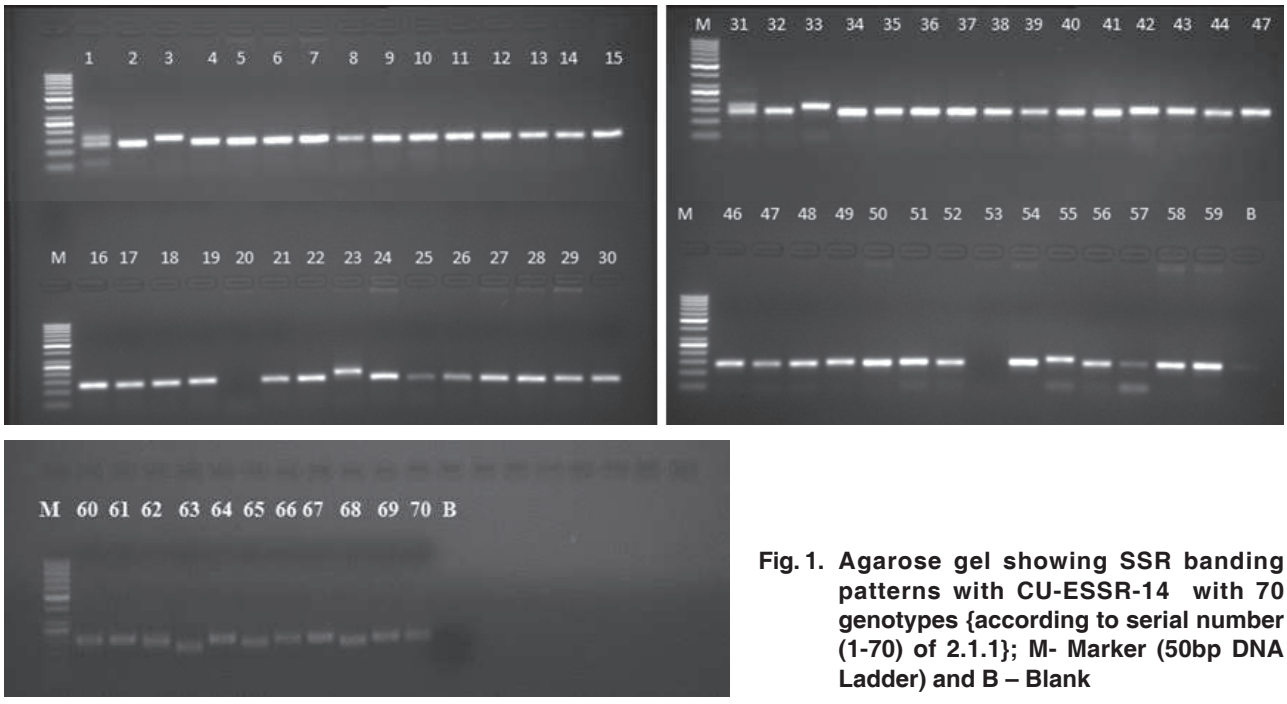

Fig. 1. Agarose gel showing SSR banding patterns with CU-ESSR-14 with 70 genotypes {according to serial number (1-70) of 2.1.1}; M- Marker (50bp DNA Ladder) and B – Blank

Table 3. Clustering of genotypes based on the Neighbour joining methods

Cluster	Sub cluster	No. of genotypes	Name of genotypes
I	A	5	EC-303433(17), EC-164966(50), CR-11A, IC-152485, NIC-16434
	B	6	UMA, SAHEB, EC-334973(38), CUHY-27, IC-204159, SAVITRI
	C	11	EC-334971(23), EC-182832(26), EC-303442(32), EC-41923-(49), EC-100043-A, EC-204704(44), CUMS 04, CUMS 20, EC-310448(39), EC-334962(28), EC-334988-(3)
II	A	8	Gujarat Local, ASSAM Local, GT-2, CUHY-57, RAMA, EC-303439, TKG 22, RT348
	B	4	CUMS-11,CUHY-13, EC-335004(34),B9
III	A	9	VR 1, RT-54, HUMRA, OSC-593, CUMS 9, TMV 4, CUHY-45, EC-303432, IC-14331
	B	6	GT-10, V-10, EC-303435(4), TKG-352, TILLOTTAMA, NIRMALA
IV	A	4	EC-335004, V-12, CUMS-9, CUHY-23
V	A	7	IC-14053, IC-2621694, IC-131490, EC-303431, CUHY-36, IC-96230, TKG 355
	B	3	IC-2621694, IC-43033, AMRIT
	C	7	CR-11, EC-310448(36), CUMS-19, CUMS-17, IC-204063, CUHY-24, IC-2621694

II contain 12 genotypes. High yielding varieties (HYV) are dominant in this group. Cluster III contains 15 genotypes. Accessions originating from India were grouped into 2 separate clusters. One cluster (IIIA) exclusively contained the accessions belonging to Indian origin HYV. In (IIIB) Indian origin HYV along with wild genotypes were present. Cluster IV, contained only four genotypes. Accessions originated from India mainly land races were present in the cluster V.

Wu et al. (2014) suggested the reason of genetic narrowness in cultivated sesame to be the domestication of landraces along with advanced plant breeding lines. The genetic variation in sesame was consequently reduced by selection and genetic drift. Characterization of genetic diversity of available landraces especially from different ecological and geographical collection by molecular markers is of great value to support parental line selection and breeding strategy design (Wu et al. 2014). The understanding of these landraces can provide a better foundation for further conservation and utilization of these resources (Pandey et al. 2005). Therefore, the combination of morphological and molecular based analysis in genetic diversity assessment of the sesame appears to be significant in developing any breeding program. The genic markers under study can be utilized as a useful tool for association mapping, genetic linkage map construction, genetic diversity detection, and marker-assisted selective breeding in future sesame breeding program.

Authors' contribution

Conceptualization of research (TD); Designing of the experiments (TD); Contribution of experimental materials (TD); Execution of field/lab experiments and data collection (AI, UB, RA); Analysis of data and interpretation (AI, TD); Preparation of manuscript (AI, RA).

Declaration

The authors declare no conflict of interest.

Acknowledgement

We acknowledge the university grants commission for providing the financial support to carry out the study under UGC Major Project on sesame titled "Genetic diversity among genotypes and molecular linkage map construction in sesame (*Sesamum indicum* L.)" UGC Reference No. F. 42-721/2013(SR).

References

Akagi H., Yokozeki Y., Inagaki A., Nakamura A. and Fujimura T. 1996. A codominant DNA marker closely linked to the rice nuclear restorer gene. Rf-1. identified with inter-SSR fingerprinting. Genome, **39**(6): 1205-1209.

Anandabn R., Deenathayalan T., Prakash M., Sunil kumar B. and Narayanan G. S. 2017. Assessment of genetic diversity among sesame (*Sesamum indicum* L.) germplasm as revealed by RAPD and SSR markers. Indian J. Biochem. Biophys., **55:** 143-150.

Ashri A. 1998. Sesame breeding. Plant Breed. Rev., **16**: 179-228.

Badri J., Yepuri V., Ghanta A., Siva S. and Siddiq E. A. 2014. Development of microsatellite markers in sesame (*Sesamum indicum* L.). Turkish J. Agric. Forestry, **38**(5): 603-614.

Bandopadhyay R., Sharma S., Rustgi S., Singh R., Kumar A., Balyan H. S. and Gupta P. K. 2004. DNA polymorphism among 18 species of *Triticum–Aegilops* complex using wheat EST–SSRs. Plant Sci., **166**(2): 349-356.

Banerjee P. P. and Kole P. C. 2009. "Analysis of Genotypic Diversity in Sesame (*Sesamum indicum* L.) Based on Some Physiological Characters". Czech J. Genet. Pl. Breed., **45:** 72-78.

Becher R. 2007. EST derived microsatellites as a rich source of molecular markers for oats. Plant Breed., **126**(3): 274-278.

Bedigian D. and Van der Maesen L. J. G. 2003. Slimy leaves and oily seeds: Distribution and use of *Sesamum* spp. and *Ceratotheca sesamoides* (Pedaliaceae) in Africa. In: Plant Resources of Tropical Africa: Proc. 1st PROTA Intern. Workshop. Nairobi., Kenya. 23-25 september 2002/GH Schmelzer and EA Omino (pp. 271-273).

Bedigian D. 1985. Is še-giš-i sesame or flax. Bull. Sumerian Agric., **2**: 159-178.

Boguski M. S., Lowe T. M. J. and Tolstoshev C. M. 1993. dbEST — database for "expressed sequence tags". Nature Genetics, **4**: 332. Nature Publishing Group http://dx.doi.org/10.1038/ng0893-332.10.1038/ng0893-332.

Chagné D., Chaumeil P., Ramboer A., Collada C., Guevara A., Cervera M. T., Vendramin G. G., Garcia V., Frigerio J-M., Echt C., Richardson T. and Plomion C. 2004. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor. Appl. Genet., **109**(6): 1204-1214.

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 GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet., **100**(5): 713-722.

Cho Y. I., Park J. H., Lee C. W., Ra W. H., Chung J. W., Lee J. R. and Park Y. J. 2011. Evaluation of the genetic diversity and population structure of sesame (*Sesamum indicum* L.) using microsatellite markers. Genes and Genomics, **33**(2): 187-195.

Dar A. A., Mudigunda S., Mittal P. K. and Arunmugam N. 2017. Comparative assessment of genetic diversity in *Sesamum indicum* L. using RAPD and SSR markers. Biotech., **3:** 7-10.

Dixit A., Jin M. H., Chung J. W., Yu J. W., Chung H. K., Ma K. H. and Cho E. G. 2005. Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). Mol. Ecol. Res., **5**(4): 736-738.

Dossa K., Wei X., Zhang Y., Fonceka D., Yang W., Diouf D., Liao B., Cissé N. and Zhang X. 2016. Analysis of genetic diversity and population structure of sesame accessions from Africa and Asia as major centers of its cultivation. Genes, **7**: 14. doi: 10.3390/genes 7040014.

Ellis J. R. and Burke J. M. 2007. EST-SSRs as a resource for population genetic analyses. Heredity, **99**(2): 125-132.

Fukuda Y., Osawa T., Kawakishi S. and Namiki M. 1994. Chemistry of lignan antioxidants in sesame seed and oil in food phytochemicals for cancer prevention, American Chemical Society, USA: 264-274.

Goyal V. K. and Bisen R. 2017. "Assessment of genetic divergence in Niger germplasm" Intnl. J. Chem. Studies, **5**(4): 1482-1485.

Gutierrez M. V. 2005. Cross-species amplification of Medicago truncatula microsatellites across three major pulse crops. Theor. Appl. Genet., **110**(7): 1210-1217.

Heesacker A., Kishore V. K., Gao W., Tang S., Kolkman J. M., Gingle A. and Rieseberg L. H. 2008. SSRs and INDELs mined from the sunflower EST database: abundance. polymorphisms and cross-taxa utility. Theor. Appl. Genet., **117**(7): 1021-1029.

Laurentin H. E. and Karlovsky P. 2006. Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genetics, **10-7**(1): 1471-2156. https://doi.org/10.1186/1471-2156-7-10.

Levinson G. and Gutman G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol., **4**(3): 203-221. https://doi.org/10.1093/oxfordjournals.molbev.a040442.

Li-Bin W. E. I., Zhang H. Y., Zheng Y. Z., Wang-Zhen G. U. O. and Zhang T. Z. 2008. Developing EST-derived microsatellites in sesame (*Sesamum indicum* L.). Acta Agronomica Sinica, **34**(12): 2077-2084.

Liewlaksaneeyanawin C., Ritland C. E., El-Kassaby Y. A. and Ritland K. 2004. Single-copy species-transferable microsatellite markers developed from loblolly pine ESTs. TAG Theor. Appl. Genet., **109**(2): 361-369.

McCouch S. R., Teytelman L., Xu Y., Lobos K. B., Clare K., Walton M. and Zhang Q. 2002. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res., **9**(6): 199-207.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA, **70:** 3321-3323.

Pandey S. K., Das A., Rai P. and Dasgupta T. 2015. Morphological and genetic diversity assessment of sesame (*Sesamum indicum* L.) accessions differing in origin. Physiol. Mol. Biol. Plants, **21**(4): 519-529. http://doi.org/10.1007/s12298-015-0322-2.

Pashley C. H., Ellis J. R., McCauley D. E. and Burke J. M. 2006. EST databases as a source for molecular markers: lessons from *Helianthus*. J. Hered., **97**: 381-388.

Ramprasad E., Senthilvel S., Jatoth J. L., Yamini K. N., Dangi K. S., Ranganatha A. R. G. and Varaprasad K. S. 2017. An insight into morphological and molecular diversity in Indian sesame cultivars. Indian J. Genet., **77**(2): 271-277. doi: 10.5958/0975-6906.2017.00036.0.

Saghai-Maroof M. A., Soliman K. M. Jorgensen R. A. and Allard R. W. 1984. Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance. Chromosomal location and population dynamics. Proc. Natl. Acad. Sci., USA. **81**: 8014-8019.

Sneath H. A. and Sokol R. R. 1973. Numerical Taxonomy; The Principles and Practice of Numerical Classification. W.H. Freeman, San Francisco CA., USA: 573.

Squirrell J., Hollingsworth P. M., Woodhead M., Russell J., Lowe A. J., Gibby M. and Powell W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol., **12**: 1339-1348.

Weir B. S. 1990. Genetic Data Analysis. Methods for discrete population genetic data. Sinauer Associates Inc., Sunderland, Massachusetts USA: 377.

Wu K., Yang M., Liu H., Tao Y., Mei J. and Zhao Y. 2014. Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. BMC Genetics, **15**: 35 http://doi.org/10.1186/1471-2156-15-35.

Yepuri V., Surapaneni M., Kola V. S. R., Vemireddy L. R., Jyothi B. Dineshkumar V. and Siddiq E. A. 2013. Assessment of genetic diversity in sesame (*Sesamum indicum* L.) genotypes using EST-derived SSR markers. J. Crop Sci. Biotechnol., **2**(16): 93-103.

Yi G., Lee J. M., Lee S., Choi D. and Kim B. D. 2006. Exploitation of pepper EST–SSRs and an SSR-based linkage map. TAG Theor. Appl. Genet., **114**(1): 113-130.

Yu J. K., La Rota M., Kantety R. V. and Sorrells M. E. 2004. EST derived SSR markers for comparative mapping in wheat and rice. Mol. Genet. Genom., **271**(6): 742-751.