Bangladesh J. Bot. **47**(2): 247-255, 2018 (June)

IN SILICO DEVELOPMENT AND ANNOTATION OF TOMATO EST-SSRS AND THEIR APPLICATION IN GENETIC DIVERSITY STUDY OF TOMATO GERMPLASM

SHILPA* AND RAJINDER KAUR

Department of Biotechnology, Dr. Y S Parmar University of Horticulture and Forestry, Nauni, Solan-173230, Himachal Pradesh, India

Keywords: SSRs, Genetic Variability, Genotypes, Primers, Polymorphism

Abstract

Present study aimed at finding out genetic diversity amongst a new set of germplasm representing various varieties and lines of tomato using *in silico* designed simple sequence repeats (SSRs) with extraction and assembly of 4200 expressed sequence tag (EST) sequences, finding out SSR containing EST sequences and designing of 55 primers, of which 20 were used to conduct diversity studies. On PCR amplification 62 bands were generated, representing 60 polymorphic (96.77% polymorphism) bands. Similarity values of data analysis ranged from 0.10 to 0.87 indicating a broad genetic base of germplasm. Dendrogram and principle component analysis (PCA) of clustering analysis divided the germplasm in two main clusters. Various fruit characters were also assessed morphologically. This study demonstrated the potential of EST-SSRs in genetic diversity analysis further facilitating the use of studied germplasm in breeding programs. These markers can also be used to successfully conduct QTL and marker assisted selection studies.

Introduction

Tomato (*Solanum lycopersicum* L.), is economically one of the most important and widely grown vegetable crops of the Solanaceae. Wild tomato species, native to South America exhibit great morphological and ecological diversity, and are very useful in breeding programs. Molecular markers offer efficient selection of desired agronomic traits in breeding studies because they are based on the plant genotypes and are also independent of environmental variations (Huq *et al*. 2009). SSRs, considered as one of the most powerful Mendelian markers, have been widely used in germplasm identification, population genetics, evolutionary studies and crop improvement (Mohan *et al*. 2013). On the basis of origin, SSR markers are of two types: genomic SSRs which are developed from enriched DNA libraries, and genic or EST-SSRs, derived from EST sequences originating from the expressed region of the genome which are submitted, in public domains, as cDNA clones (Chagne *et al*. 2004). The development of genomic SSRs is a time consuming job while genic-SSRs are easier to search *in silico* for a particular organism. However, the large EST datasets are highly redundant and contain a large number of repetitions in the sequences. A few attempts have been made to reduce this redundancy and develop non-repetitive EST datasets for tomato. Keeping in view the above information, the present study was conducted to fulfill the objectives of *in silico* EST-SSR primer designing and evaluation of genetic diversity among tomato genotypes. Non-redundant EST sets assembled from public datasets were mined for SSR motifs which were used to develop novel primers. The utility of the primers were checked on a broad germplasm of tomato.

*Author for correspondence: <shilpachaudhary05@gmail.com>.

Materials and Methods

The plant material comprised of different lines/accessions and variety of tomato is shown in Table 1. Seeds procured from Department of Vegetable Science, Dr. Y S Parmar University of Horticulture and Forestry, Nauni, Solan, India, were sown to obtain seedlings. The fresh and healthy leaves of each plant representing different lines and varieties were collected and stored at –80ºC until further use.

Retrieval of EST sequences: A total of 4,200 EST sequences of *S. lycopersicum* were downloaded in FASTA format from the EST database of NCBI platform (www.ncbi.nlm.nih.gov/nucest) and compiled in a single text file.

EST assembly: Non-redundant dataset from the EST sequences were obtained by using EGassembler-GenomeNet web server (Masoudi-Nejad *et al*. 2006). Data input was given in one-click assembly through file upload option in FASTA format. Parameters and algorithms used by software are as follows: sequence cloning process by minimum per cent identity of 96 for an alignment with a contaminant, repeat masking process by using RepBase repeats library of *Arabidopsis*, vector masking process by using core NCBI's vector library, organelle masking process by using plastids library of *Arabidopsis* and sequence assembly process by overlap per cent identity cutoff N > 65 value of 80. The software removed various contaminants (low quality and complexity regions) and assembled the EST sequences by creating contigs and singletons using CAP3 (Huang and Madan 1999, Vaidya *et al*. 2015).

SSR motif detection: Simple sequence repeat identification tool (SSRIT)-Gramene (Temnykh *et al.* 2001, Kaur *et al*. 2015) was used to find out SSR sequences among the contigs and singleton sequences. EST sequences were pasted in FASTA format in the text area. The SSRs were screened out by setting the search parameters to identify at least 5 repeats motifs with a maximum of 10 base pairs. These parameters strengthen the results. The software detects SSR sequences using following information: Sequence Id, SSR motif, No. of repeats, SSR start position and SSR end position.

Frequency of different repeat motifs with variability in their length was calculated by using the formula:

$$\text{Frequency of repeat} = \frac{\text{Number of times particular repeat occurred}}{\text{Total number of repeats of the same length}} \times 100$$

Primer designing: Primers were designed using Primer3 v.0.4.0 software (www.frodo.wimit.edu/primer3/) (Rozen and Skaletsky 2000) which were then custom synthesized from Genaxy Scientific (New Delhi, India). The parameters used for primer designing were as follows: Primer size- 18 to 27 bps, Primer Tm-57 to 63ºC and GC% - 20 to 80%.

Functional annotation: BLAST × tool from NCBI (https://blast.ncbi.nlm.nih.gov/Blast) was used for putative annotation of designed EST-SSRs by comparative analysis and assignment of functions on the basis of EST sequence homology. E value of <1E-5 was taken as a significant criterion of homology. Maximum target sequence of 100 along with BLOSUM62 as scoring matrix parameter, existence: 11 and extension : 1 as gap cost and conditional compositional score matrix adjustment were used as algorithm parameters.

PCR amplification: To conduct genetic marker analysis, DNA from leaves of seedlings was isolated using CTAB method (Doyle and Doyle 1987). A reaction mixture of 20 µl for PCR analysis was prepared using 1 X PCR buffer, 2 mM $MgCl_2$, 1 mM dNTPs, 20 pm of each primer (forward and reverse), 1 U Taq DNA polymerase, 50 ng template DNA following a thermal profile as : 5 min of initial denaturation at 95°C followed by 40 cycles of 1 min denaturation at

94°C, where annealing varied with Tm of each primer for 1 min and extension of 2 min at 72°C, further followed by final extension of 5 min at 72°C. The amplified DNA was mixed thoroughly with 6 X loading dye and then electrophoresed in 3.5% agarose gel supplemented 0.5 µg/ml ethidium bromide followed by a run at a constant voltage at the rate of 5 V/cm for about 3 hrs. Polymorphism study was carried out to find out the variability of the germplasm using the following formula:

$$\text{Per cent polymorphism} = \frac{\text{Total number of polymorphic bands}}{\text{Total number of bands}} \times 100$$

The Polymorphism information content (PIC) (Anderson *et al.* 1993) values for alleles at each SSR locus was calculated using the formula:

$$PIC = 1 - \sum pi^2$$

where 'pi' is the frequency of the ^{i}th allele.

Fingerprints of different genotypes were recorded based on the occurrence of unique bands. Effective multiplex ratio measures the number of polymorphic loci in the germplasm, analyzed per experiment (Varshney *et al*. 2007). It was calculated by the following formula: $E = n\beta$; where, "β" is the fraction of polymorphic markers and is estimated as β = np/(np + nnp), indicating "np" as polymorphic loci and "nnp" as non-polymorphic loci, "*n*" is the multiplex ratio, measured as the average number of DNA fragments amplified/detected per genotype using a marker system. The utility of a given marker system is a balance between the level of polymorphism detected and the extent to which an assay can identify multiple polymorphisms. Marker index (MI) provides a convenient estimate of the utility of any marker system (Varshney *et al*. 2007). It was estimated as follows: $MI = PIC \times E$.

Here, E and MI values were also calculated for each primer.

Dendrogram and PCA plot: To carry out data analysis binary matrix was prepared using banding pattern generated by all markers by indicating the presence of PCR product band by 1 and absence by 0. NTsys pc Version 2.02 (Rohlf 2000) was used to create dendrogram (UPGMA clustering) and PCA plots. Similarity matrix was prepared using SimQual function with Jaccard's coefficient of similarity module and dendrogram was constructed using SAHN module of clustering with UPGMA clustering method. Nine morphological descriptors of fruit viz., size, weight, shape, colour, intensity of colour, presence/absence of dark green stripes on the (shoulders) fruit, fruit shoulder shape, fruit cross-sectional shape and shape of pistil scar were used to analyze diversity present at the morphological level. H-index was calculated for these descriptors to find out morphological homogeneity by finding out the ratio of each category of descriptor followed by the sum of their square root.

Results and Discussion

EST-SSRs represent exonic regions of the genome which will transcribe into proteins and they remain conserved increasing their importance as markers for various molecular studies. For their development first and foremost requirement is the availability of EST sequences. In the present study, a total of 4200 initial EST sequences of *S. lycopersicum,* out of total dataset of 300290 was downloaded from NCBI EST database by choosing EST from database toolbar and specifying the name of crop as *S. lycopersicum* for further search. The dataset was assembled into 352 contigs and 1,918 singletons leading to a non-redundant dataset of 2,270 sequences with a 45.95% reduction in redundancy.

The EG assembler was software used to deduct overlapping sequences and also provided the unique sequences without any duplication in order to screen out and discard unwanted sequences; further increasing the precision of primer designing. Data output was in the form of ZIP format .Cap_all_zip. Repetitive elements including small RNA pseudo genes, LINEs, SINEs, LTR elements, vector sequences, organelle and other interspersed repeat were masked by the software through automatic screening and cleaning for various contaminants in the EST sequences. The server assembled the sequences into contigs and singletons using CAP3 with the overlap identity of 80% between one end of a default read to another end (Huang and Madan 1999). Kaur *et al*. (2015) and Vaidya *et al*. (2015) also reported a significant reduction in redundancy of 22.2% and 80.33%, respectively, by using EG assembler. A total of 94 SSRs was found, out of which 19 were obtained from contigs and 75 from singletons. The SSRs obtained were of di-nucleotide and tri-nucleotide repeats. A total of 55 di- nucleotide (58.51%) and 39 tri-nucleotide (41.48%) repeats was obtained. Five different types of di-nucleotide motifs were observed, of which AT/TA (36.36%) repeat motif was most common and GT/TG (7.27%) was least repeated motif. In case of tri-nucleotide repeats, 13 different motifs were found, among which TCT/CTT/TTC (17.94%) was most common, while GGC and ATC (2.56%) contributed the least. These results are supported by previous findings of Kaur *et al*. (2015), Parmar *et al*. (2010), Vaidya *et al*. (2012). A total of 55 primer pairs were designed from 94 SSR containing EST sequences, as remaining 39 SSR containing EST sequences were not found to contain any EST-SSR primer pair by Primer 3. Out of these, 20 primer pairs were custom synthesized and were further used to carry out diversity studies. Homology analysis with E value of <1E-5 using BLASTx tool revealed high level of sequence similarity, based on identity percentage, with four plants species: 14 sequences were homologous with *S. lycopersicum*, 4 showed homology with *S. tuberosum*, while 2 sequences were found to be homologous with *Nicotiana tabacum* and *Coffea canephora* (Table 2).

Out of the 20 EST-SSR primers pairs used for PCR amplification, of which 10 were of di- and other 10 were of tri-nucleotide repeats, 16 produced polymorphism and 4 did not produce any amplification. In total, 62 bands were produced, of which 60 revealed polymorphism (96.77%). A total of 440 fragments were amplified by all primers among the tomato germplasm varieties used which counted for 25.88 fragments per amplified primer and 13.75 fragments per accession. PIC values calculated for the amplified primers ranged from 0.15 to 0.92 with an average of 0.47 (Table 2). A total of 16 unique bands were produced, of which maximum number of 4 unique bands were produced by Contig 7 primer in accessions namely, EC-10304 (at 500 bp, 250 bp) and EC-528373 (at 150 bp, 400 bp), followed by Contig 340 in EC-2798 (300 bp), EC-251649 (200 bp) and EC-114375 (100 bp), gi|4387244| in EC-521054 (300 bp), EC-528373 (200 bp) and EC-168283 (100 bp), gi|4386782| in EC-521 (500 bp, 400 bp, 300 bp), which generated three unique bands each. Primer gi|4386508| produced two unique bands in EC-114375 (1000 bp), EC-6486 (200 bp) and the remaining single unique band was generated by gi|4386388| in EC-29914 (100 bp). The average E value and the MI value for the EST-SSR primers studied was 1.85 and 0.86, respectively. Maximum values of E and MI were found to be 0.46 and 0.42 with primer gi|4386508|.

Jaccard's similarity matrix coefficient was obtained through NTSYSpc. The similarity coefficient values ranged from 0.10 to 0.87 indicating a broad variety of the germplasm studied. In the dendrogram, two main clusters were found with further formation of sub-clusters. Cluster I consisted of 25 genotypes, while rest 7 were grouped in cluster II (Fig. 1). The maximum similarity of 87% was found between EC-521059 and EC-501074, while minimum similarity of 10% was observed between EC-12699 and EC-6486. It was found that 'Solan Vajr' which is a commercial variety, was present on the top of cluster, separating it from rest of the germplasm (Fig. 1). PCA was also done, which divided the germplasm into same groups as depicted by the

Table 1. List of tomato germplasm varieties used for evaluation of genetic diversity.

Sl. No.	Name/accession number	Source	Fruit size	Fruit wt .(g)	Fruit shape	Fruit colour	Intensity of colour	Presence /absence*	Fruit shoulder shape	Fruit cross-sectional shape	Shape of pistil scar
1.	Solan Vajr	Dr Yashwant Singh Parmar University of Horticulture and Forestry, Solan, Himachal Pradesh, India	Large	80-90	Round	Red	Intermediate	Absent	Flat	Round	Stellate
2.	Hawai 7998	Hawai, US	Medium	60-70	Round	Red	Intermediate	Absent	Flat	Round	Stellate
3.	EC-520075	Indian Institute of Vegetable Research, Varanasi, Uttar Pradesh, India	Small	10	Round	Red	Dark	Absent	Flat	Round	Dot
4.	EC-251649	IIVR, Varanasi, UP, India	Small	10	Round	Red	Dark	Absent	Flat	Round	Dot
5.	EC-528373	IIVR, Varanasi, UP, India	Small	10	Round	Red	Intermediate	Absent	Flat	Round	Dot
6.	EC-126902	IIVR, Varanasi, UP, India	Medium	50	Round	Dark red	Dark	Absent	Flat	Round	Stellate
7.	EC-6486	IIVR, Varanasi, UP, India	Small	30	Round	Red	Intermediate	Absent	Flat	Round	Stellate
8.	EC-2517	IIVR, Varanasi, UP, India	Medium	40	Round	Light red	Light	Absent	Flat	Round	Stellate
9.	EC-13904	IIVR, Varanasi, UP, India	Small	30	Ellipsoid	Red	Intermediate	Absent	Flat	Round	Stellate
10.	EC-36883	IIVR, Varanasi, UP, India	Medium	50	Ellipsoid	Light red	Light	Present	Flat	Round	Stellate
11.	EC-168283	IIVR, Varanasi, UP, India	Small	30	Round	Red	Intermediate	Absent	Slightly depressed	Round	Stellate
12.	EC-10662	IIVR, Varanasi, UP, India	Small	30	Round	Red	Intermediate	Absent	Flat	Round	Stellate
13.	EC-114375	IIVR, Varanasi, UP, India	Medium	40	Round	Red	Intermediate	Absent	Flat	Round	Stellate
14.	EC-2798	IIVR, Varanasi, UP, India	Small	30	Round	Light red	Light	Absent	Flat	Round	Stellate
15.	EC-25265	IIVR, Varanasi, UP, India	Medium	40	Round	Red	Intermediate	Absent	Flat	Round	Stellate
16.	EC-521059	IIVR, Varanasi, UP, India	Medium	50	Round	Light red	Light	Absent	Flat	Irregular	Stellate
17.	EC-10304	IIVR, Varanasi, UP, India	Small	20	Round	Red	Intermediate	Absent	Flat	Round	Stellate
18.	EC-501074	IIVR, Varanasi, UP, India	Small	10	Round	Red	Dark	Absent	Flat	Round	Dot
19.	EC-29914	IIVR, Varanasi, UP, India	Small	20	Round	Red	Intermediate	Absent	Slightly depressed	Round	Stellate
20.	EC-521041	IIVR, Varanasi, UP, India	Medium	50	Flat	Light red	Light	Present	Moderately depressed	Round	Stellate
21.	EC-362949	IIVR, Varanasi, UP, India	Medium	50	Round	Light red	Light	Absent	Flat	Round	Stellate
22.	EC-528367	IIVR, Varanasi, UP, India	Medium	40	Round	Light red	Light	Absent	Slightly depressed	Round	Stellate
23.	EC-251649	IIVR, Varanasi, UP, India	Small	10	Round	Orange	Light	Absent	Flat	Round	Dot
24.	EC-521054	IIVR, Varanasi, UP, India	Medium	40	Round	Red	Intermediate	Absent	Flat	Round	Stellate
25.	EC-35322	IIVR, Varanasi, UP, India	Small	30	Round	Red	Intermediate	Absent	Flat	Round	Stellate
26.	EC-251646	IIVR, Varanasi, UP, India	Medium	40	Flat	Orange	Intermediate	Present	Slightly depressed	Irregular	Stellate
27.	EC-528374	IIVR, Varanasi, UP, India	Small	30	Round	Red	Intermediate	Absent	Flat	Round	Stellate
28.	EC-521051	IIVR, Varanasi, UP, India	Medium	50	Ellipsoid	Light red	Light	Present	Moderately Depressed	Angular	Stellate
29.	EC-521	IIVR, Varanasi, UP, India	Medium	40	Ellipsoid	Red	Intermediate	Absent	Slightly Depressed	Round	Stellate
30.	EC-8591	IIVR, Varanasi, UP, India	Medium	30	Round	Red	Intermediate	Absent	Flat	Round	Stellate
31.	EC-12699	IIVR, Varanasi, UP, India	Medium	50	Round	Dark red	Dark	Absent	Flat	Round	Stellate
32.	EC-2791	IIVR, Varanasi, UP, India	Small	20	Round	Red	Intermediate	Absent	Flat	Round	Stellate

*Presence /absence of dark green stripes on the (shoulders) fruit.

Table 2. BLASTx derived putative annotation of EST sequences selected for primer design.

Sequence ID	Annotation	Sequence of primer (5' ⟶ 3')	Tm (ºC)	PIC value	E value	MI value	Polymorphic/ monomorphic
Contig7	Uncharacterized protein	F:GGGGAGATAGCACGGATTTT	57.36	0.77	0.15	0.11	Polymorphic
		R:GTGGGCGGCAAATTAAGGAT	58.89				
Contig143	H/ACA ribonucleoprotein complex subunit 2-like protein	F:GTCATCGACGAAACAAAGCA	57.06	0	0.03	0	Polymorphic
		R:CTTGCCTCTTGATCTTCGCC	58.99				
Contig162	Alpha-tubulin	F:TAATGCTGGACCTGGAACCA	58.63	0.69	0.12	0.08	Polymorphic
		R:GTCAGCAATAACCACAGGCT	58.17				
Contig265	Hop-interacting protein THI041	F:GGTGGAGGTGGGCATAATGA	59.45	0	0.03	0	Polymorphic
		R:GTAATTGATCCACCGGCGTC	59.06				
Contig340	Hydrogen peroxide-induced 1	F:GCCCTCTTGAGGACTTGGAG	59.75	0.71	0.15	0.10	Polymorphic
		R:ACCTTCAAAAGCAGTGCAGC	59.61				
Contig352	Actin-depolymerizing factor 2	F:GAACGCTCCCTCTACCTTTT	57.24	-	-	-	No amplification
		R:TAACTTTCAGCTGGCTCACC	57.81				
gi|116644211|	*Coffea canephora* DH200=94 genomic scaffold, scaffold_26	F:AAAAGATGCAACGCTGGAAC	57.58	0.50	0.06	0.03	Polymorphic
		R:ACTGAAACGCGCACATGTAA	58.78				
gi|4387244|	Protein EXECUTER 2, chloroplastic isoform X2	F:CGCAACTCCTTCTGCTGATG	59.27	0.77	0.15	0.11	Polymorphic
		R:GCAACTGAGTCCTTCGATGT	57.91				
gi|4386975|	Glutamate-1-semialdehyde 2,1-aminomutase, chloroplastic	F:TCTCCATCTCACCGTCGATG	58.97	0.66	0.12	0.07	Polymorphic
		R:AGCACGGACAGGGGAATTTA	59.01				
gi|4386907|	Polyadenylation and cleavage factor homolog 4	F:TTGTGGACTATCGGACCCTG	58.81	0.15	0.05	0.004	Polymorphic
		R:ACCATGGTTCCTGCAGATGA	59.01				
gi|4386813|	Uncharacterized protein	F:TGGGGTTTTGTTGTGAGGAA	57.47	-	-	-	No amplification
		R:ATATCCGGTGGCCTCGAAAT	58.94				
gi|4386782|	Polyadenylate-binding protein 1	F:ATCAATTTCACTCCACCGCC	58.54	0.26	0.15	0.03	Polymorphic
		R:TCGGCATCCATCTCTCCTTC	58.96				
gi|4386589|	Transcription factor JERF1	F:GGATTTCTCGCCGGTTAACC	58.99	0.50	0.06	0.03	Polymorphic
		R:TGGAGGATCTGTCAGCTTCG	59.18				
gi|4386576|	Transcription factor JERF1	F:GGTCACGTGTCAACCATCAT	58.19	0.41	0.06	0.02	Polymorphic
		R:AGGGGAAAGGGAAAGAGTCG	59.02				
gi|4386543|	Chloroplast-specific ribosomal protein	F:ACTCCTGAGATGTCGTGCAA	59.03	-	-	-	No amplification
		R:TGCCCCACAAAACTCAAACA	58.44				
gi|4386508|	Histone H2B.1	F:GCCAGCTGAGAAGAAACCAG	58.84	0.92	0.46	0.42	Polymorphic
		R:ACCCATAGACTTGCTGGAGA	57.75				
gi|4386388|	Carbonic anhydrase	F:CCTTACTCTCTCCCCTGCTC	58.59	0.55	0.09	0.04	Polymorphic
		R:TTTCTCGAGTGCAGCAATGG	58.84				
gi|4386332|	Ketol-acid reductoisomerase	F:GGCACGAGTCGATTTGTCTG	59.28	-	-	-	No amplification
		R:TTGGTTGATGATGCGGAGGA	59.38				
gi|4386229|	Elongation factor Tu	F:ACGAGCTTCCTTCTACCACA	58.37	0.24	0.045	0.01	Polymorphic
		R:GCGGGTGAATGGAGGAAAAG	59.19				
gi|76572221|	Phospholipase D	F:ATCACCATCTTCCTCTGCCT	58.11	0.53	0.09	0.04	Polymorphic
		R:CCCGTTGAAGTTGATCGCAA	59.13				

dendrogram (Fig. 2). Further, the authentication of the genetic relationship was also found congruent with an H-index value of fruit base morphological descriptors. H-index value ranged from 0.19 to 0.83 for nine descriptors with values of 0.47, 0.94, 0.19, 0.68, 0.46, 0.42, 0.78, 0.64, 0.83 and 0.74 for size, weight, shape, colour, intensity of colour, presence/absence of dark green stripes on the (shoulders) fruit, fruit shoulder shape, fruit cross-sectional shape and shape of pistil scar, respectively.

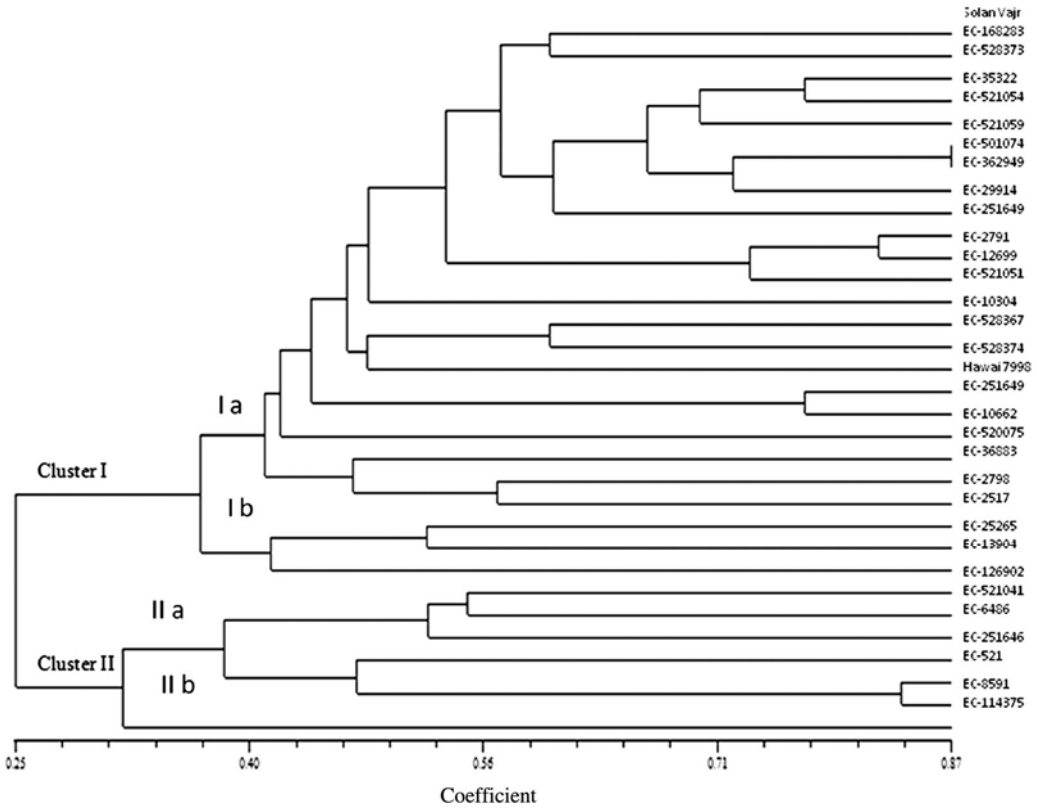

Fig. 1. Dendrogram of tomato germplasm genetic variance based on EST-SSR analysis.

EST-SSRs have evolved as an important class of molecular markers because of their easy availability, hyper variability, suitability for high throughput analysis and polymorphism in comparison to other available markers (Vaidya *et al*. 2012). In this study, more abundance of di-nucleotide repeat motifs were detected followed by tri-nucleotide repeat motifs. These results are in agreement with those found by Jung *et al*. (2005) in apricot and peach, Aggarwal *et al*. (2007) in coffee, Zhang *et al*. (2010) in walnut and Vaidya *et al*. (2012) in cauliflower. But some other findings in maize, rice, sorghum and wheat (Kantety *et al*. 2002), in barley (Thiel *et al*. 2003) and in oat (Becher 2007) showed that tri-nucleotide motifs were the leading SSR repeat motif type. While tetra-nucleotide repeat motifs were found to be the most abundant in sugarcane (Pinto *et al*. 2006). This variation in abundance level of EST-SSR may be attributed because of changes in SSR search criteria, SSR mining tools and the size of the database searched (Varshney *et al*. 2005). In present study broad genetic base of germplasm taken was revealed, which is congruent with the morphological findings as the germplasm consisted of small, medium and large varieties

and lines of different colours along with other distinct morphological characters. Grouping pattern obtained upon PCA was found congruent with dendrogram clustering, further confirming authentication of results. Also, high level of polymorphism (96.77%), high PIC (0.15 to 0.92), E (0.46) and MI (0.42) values and a sufficient number of unique bands obtained using EST-SSRs

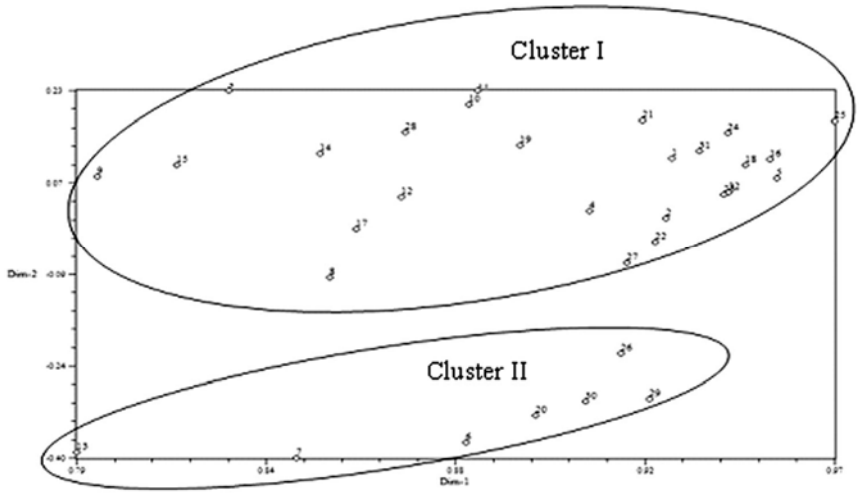

Fig. 2. Two dimensional PCA plot of tomato germplasm genetic variance based on EST-SSR analysis (1-32: genotype names are provided in Table 1).

reflects their usefulness for conducting diversity and fingerprinting studies. From these observation it may be concluded that these EST-SSRs are very promising to conduct varietal identification, genetic relationship and diversity studies with different germplasm set as well as for other marker studies like gen tagging and QTL identification. Vaidya *et al*. (2015) also revealed high usefulness of markers with more polymorphism, PIC, E and MI values towards markers studies. Use of different computational approaches for EST-SSR mining has also increased the repository of publicly available EST-SSRs which will prove a much valuable tool for further genetic and genomic studies. The use of these markers would reduce the cost and will also facilitate cultivar identification, genetic distance assessments, gene mapping and marker-assisted selection (MAS) (Sankhyan *et al*. 2017).

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V and Singh L 2007. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor. Appl. Genet. **114**: 359-372.

Anderson JA, Churchill GA, Autrique JE, Tanksley SD and Sorrells ME 1993. Optimizing parental selection for genetic linkage maps. Genome. **36**: 181-186.

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

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

Doyle JJ and Doyle JJ 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem. Bull. **19**: 11-15.

Huq S, Islam MS, Sajib AA, Ashraf N, Haque S and Khan H 2009. Genetic diversity and relationships in jute (*Corchorus* spp.) revealed by SSR markers. Bangladesh J. Bot. **38**(2): 153-161.

Huang X and Madan A 1999. CAP3: A DNA sequence assembly program. Genome Res. **9**: 868-877.

Jung S, Abbott A, Jesudurai C, Tomkins J and Main D 2005. Frequency, type, distribution and annotation of simple sequence repeats in *Rosaceae* ESTs. Funct. Integr. Geno. **5**: 136-143.

Kantety RV, Rota ML, Matthews DE and Sorrells ME 2002. Data mining for simple sequence repeats in expressed sequence tags from barely, maize, rice, sorghum and wheat. Plant Mol. Biol. **48**: 501-510.

Kaur R, Shilpa, Vaidya E and Kumar K 2015. Development, characterization and transferability of peach genic SSRs to some Rosaceae species. Adv. Res. **3**: 165-180.

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T and Goto S. 2006. EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res. **34**: 459-462.

Mohan C, Shanmugasundaram P and Senthil N 2013. Identification of true hybrid progenies in cassava using simple sequence repeat (SSR) markers. Bangladesh J. Bot. **42**(1): 155-159.

Parmar P, Oza VP, Chauhan V, Patel AD, Kathiria KB and Subramanian RB 2010. Genetic diversity and DNA fingerprint study of tomato discerned by SSR markers. Int. J. Biotechnol. Biochem. **6**(5): 657-666.

Pinto LR, Oliveira KM, Marconi T, Garcia AF, Ulian EC and De Souza AP 2006. Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed. **125**(4): 378-384.

Rohlf FJ 2000. NTSYS-pc numerical taxonomy and multivariate analysis version 2.0. New York: Applied Biostatics Inc. p. 25.

Rozen S and Skaletsky HJ 2000. Primer3 on the www for general users and for biologist programmers. *In*: Krawetz S and Misener S (Eds), p. 365-386. Bioinformatics methods and protocols: Methods in molecular biology. Humana Press, Totowa, NJ.

Sankhyan S, Kaur R, Shilpa and Kumar K 2017. Exploitation of walnut (*Juglans regia* L.) expressed sequence tags for development of SSR markers after *in silico* analysis. Proc. Natl. Acad. Sci., India, Sect. B Biol. Sci. doi.org/10.1007/s40011-017-0941-1.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch S 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Research **11**: 1441-1452.

Thiel T, Michalek W, Varshney RK and Graner A 2003. Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106**: 411-422.

Vaidya E, Kaur R and Bhardwaj SV 2012. Datamining of ESTs to develop dbEST-SSR for use in polymorphism study of cauliflower (*Brassica oleracea* var. *botrytis*). J. Hortic. Sci. Biotech. **87**: 57-63.

Vaidya E, Kaur R, Kumar K and Sharma N 2015. Exploitation of *Malus* ESTs for development of SSR markers after *in silico* analysis. J. Appl. Bot. Food. Qual. **88**: 164-169.

Varshney RK, Chabane K, Hendre PS, Aggarwal RK and Graner A 2007. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci. **173**: 638-649.

Varshney RK, Graner A and Sorrells ME 2005. Genic microsatellite markers in plants: Features and applications. Trends Biotech. **23**: 48-55.

Zhang R, Zhu AD, Wang XJ, Yu J, Zhang HR, Gao JS 2010. Development of *Juglans regia* SSR markers by data mining of the EST database. Plant Mol. Biol. Rep. **28**: 646-653.

(Manuscript received on 17 June, 2017; revised on 19 August, 2017)