Dr. Y.S. Parmar University of Horticulture and Forestry, Nauni – Solan, (H.P.) India

Exploitation of *Malus* ESTs for development of SSR markers after *in silico* analysis

Era Vaidya, Rajinder Kaur*, Krishan Kumar, Neha Sharma

Summary

The domestic apple (*Malus × domestica*) has been extensively characterized for EST sequences and currently 330181 ESTs are available in NCBI. Although this is a valuable resource for marker development, the redundancy in sequences makes the mining out of unique candidates for designing markers cumbersome. Keeping this in view, the present study was undertaken with the aim to remove the data redundancy in apple ESTs and then to develop SSR markers. The EST sequences were assembled into a non-redundant set of 5619 contigs and 59343 singletons (total sequences 64962), indicating 80 % reduction in data redundancy. These 64962 sequences were then used to mine out SSR motifs. 4575 EST-SSRs were detected, with dinucleotide repeats being predominant (72 %), followed by trinucleotide repeats (25 %). These markers were validated by designing primer pairs. 25 of these designed primers were tested on a group of 48 apple accessions. All the markers gave robust amplification, generating 93 % polymorphism. These findings indicate the usefulness of EST-SSRs in genome analysis. This study further emphasizes the importance of assembly of the vast amounts of data submitted in public databases. Our results have generated a set of non redundant apple ESTs which is of prime importance for development of marker systems without any repetition or overlapping.

Introduction

Although a vast array of genetic marker systems are available in literature for crop genome analysis, their efficiency and density across the genomes varies considerably. Markers based on SSRs are the markers of choice in genetics and breeding studies due to their multi-allelic nature, transferability over genotypes and extensive genome coverage.

In apple a number of microsatellite markers have been developed and deployed for the study of genetic polymorphism (POTTS et al., 2012; ZHANG et al., 2012; KITAHARA et al., 2005; LIEBHARD et al., 2002). However, most of these are genomic SSRs whose development is highly laborious, cumbersome and cost-intensive. Advances in genomic technologies have generated a large number of ESTs, which are available in public databases, thereby offering an opportunity to develop EST derived SSR markers by data mining. EST-SSR markers are expected to possess high transferability across boundaries as they belong to relatively conserved genic regions of the genome.

The large apple (*Malus × domestica*) EST database (comprising over 300000 sequences) provides a valuable resource for developing well characterized molecular markers (NEWCOMB et al., 2006; GASIC et al., 2008; IGARASHI et al., 2008). However, it has been observed that large datasets are highly redundant and contain a large number of repetitions in sequences. Little attempt has been made in reducing the redundancy present in apple EST dataset, so as to develop a set of non-repetitive EST sequences.

Keeping the above point of view, the present study was taken up with the aim to remove the data redundancy in apple ESTs by assembly

into singletons and contigs. This non-redundant set was used to mine out SSRs, which were further analyzed for abundance and distribution of different types of repeats. These SSRs were then used to develop novel SSR markers for apple, which were tested on a set of 48 apple accessions to check for their usefulness in genetic analysis and related applications in genetics and plant breeding.

Material and methods

Retrieval of EST sequences

EST sequences belonging to apple (*Malus × domestica*) were obtained in their FASTA format from the EST database on the NCBI website. A single text file containing all the sequences was compiled.

EST sequence assembly

A non-redundant data set was derived from the redundant set of 330181 apple ESTs for clustering and assembly analysis. An automated trimming and screening for various contaminants and low quality and complexity sequences was done through EGassembler web server (MASOUDI-NEJAD et al., 2006). The server clustered and assembled the sequences into contigs and singletons using CAP3 (HUANG and MADAN, 1999) with the criterion of 80 % overlap identity between one end of a read to another end. The results were obtained in two different output files, one comprising of contigs and the other of singletons. For SSR identification, these two files were combined to form non-redundant sequence data set.

Mining of SSRs

For detection of SSRs present in the EST sequences, a modified version of a Perl script SSR Identification Tool (SSRIT) (TEMNYKH et al., 2001; http://www.gramene.org/db/markers/SSRtool) was used. The SSRs were screened out by setting the search parameters to identify at least five repeats of SSR motifs with a maximum of ten base pairs. This software program takes a FASTA formatted sequence file as an input and produces an output file with sequence name, number of SSRs in the sequence, SSR type, SSR motif, repeat number, sequence coordinates for SSR and the length of the sequence.

Primer designing and marker validation

The SSRs detected were used for primer design. Primer pairs were designed to flank the target SSR repeats using PRIMER3 software (ROZEN and SKALETSKY, 2000) (www.frodo.wi/mit.edu/primer3/). The parameters for designing primers were as follows: primer length between 18 – 27 bp, GC content between 20 – 60 %, and optimum primer Tm between 57 ºC – 63 ºC. The designed primer pairs were synthesised by Eurofins mwg/operon (Eurofins Genomics, Bangalore, India). 25 primer pairs were custom synthesized and were validated for their ability for amplification on a set of 48 apple accessions.

---

* Corresponding author

SSR-EST similarity searches

The sequences used for primer designing were also analyzed for putative functions by pairwise comparison against the Swissprot/Uniprot protein database using BLASTx program at the Uniprot server (http://www.uniprot.org/), assuming an E value of < 1E-5 as a significant criterion of homology. The most significant matches for each sequence were recorded.

Gene ontology descriptions were assigned to SSR-ESTs on the basis of Swiss-Prot protein sequence matches.

SSR protein domain analysis

The sequences containing SSRs were also searched for functional protein domains using PROSITE (http://www. http://prosite.expasy.org/prosite.html). It has been observed that some regions, important for the function of a protein and/or for the maintenance of its three-dimensional structure have been better conserved than others during evolution. The properties of such groups of similar sequences are analyzed to derive a signature for a protein family or domain, which distinguishes its members from all other unrelated proteins.

Marker validation by PCR amplification and gel electrophoresis

A PCR protocol was standardized to amplify EST-SSR markers. The target sequences were amplified in 10 μl PCR reactions each containing 1×PCR buffer (Genie, Bangalore, India), 2.2 mM MgCl$_2$, 0.3 mM dNTPs, 20 μM of each primer, 1 Unit Taq DNA polymerase and 30 ng leaf DNA. SSR markers were amplified in a thermal cycler (Applied Biosystems), with the following programme: 5 min initial denaturation at 95 °C, followed by 40 cycles each of 1 min denaturation at 94 °C, 1 min annealing at the Tm for each primer and 2 min extension at 72 °C, with a final extension step of 5 min at 72 °C. The PCR products were then analysed on 2.5% (w/v) agarose gels. Electrophoresis was performed at 80V in 1×TAE buffer (40 mM Trisacetate, 1 mM EDTA; pH 8.0). The gels were stained with ethidium bromide (0.5 μg μl^{-1}) and visualized under UV light.

Data analysis

Genetic diversity, defined as Polymorphism Information Content (PIC) (ANDERSON et al., 1993), was used to measure allelic diversity at each SSR locus. PIC values were calculated as follows:

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

where pi was the frequency of the ith allele in the set of genotypes analysed.

The presence or absence of bands were encoded in binary mode by 0 and 1, respectively and used to derive a similarity matrix based on Jaccard's coefficients. A dendrogram was generated based on this matrix based on UPGMA clustering using NTSYSpc ver. 2.02h software.

Results

EST sequence retrieval and assembly

330181 EST sequences belonging to apple were retrieved from the EST database. The obtained sequences were found to be related to different plant tissues such as leaves, stem and root.

After sequence assembly, 5619 sequences were successfully assembled into contigs. The remaining 59343 showed no overlap with any ESTs, called as 'singletons'. The whole dataset was reduced to 64962 sequences, prior to the SSR analysis. Reduction in data redundancy in the assembled ESTs set of contigs and singletons was to be found 80.33% (Tab. 1).

Frequency of SSRs

SSRs in the ESTs were detected using the online interface of SSRIT. 849 (15%) contigs and 3724 (6%) singletons were found to contain SSRs. Analysis of SSR-ESTs revealed that dinucleotide SSRs were the most common SSRs at 72%. These were followed by trinucleotide SSRs at 25%. Tetranucleotide, pentanucleotide and hexanucleotide SSRs were less frequent, i.e. 1.24% SSRs were tetranucleotide, 0.37% were pentanucleotide and 0.61% were hexanucleotide repeats, respectively.

Frequency of occurrence of different SSR motifs was also calculated. Among the dinucleotide repeats, TC/CT motif was found to be the most frequent (37.76%) and CG/GC (0.09%) was least frequent (Tab. 2). Among trinucleotide repeats, the most frequent motif was CTC/TCT/CTT/TTC/CCT/TCC at 20.49% and CGG/CCG/CGC/GCG/GCC and TCA/CAT/ATC/ACT were the least frequent motifs at 4.46% (Tab. 2).

Tab. 2: Distribution of Repeat Motifs

S. No.	Repeat Motif	Number	Frequency
	DINUCLEOTIDE REPEATS		
1.	AG/GA	1159	34.84%
2.	AC/CA	185	5.56%
3.	TC/CT	1256	37.76%
4.	GT/TG	148	4.45%
5.	AT/TA	575	17.28%
6.	GC/CG	3	0.09%
	TOTAL	3326	
	TRINUCLEOTIDE REPEATS		
7.	CTC/TCT/CTT/TTC/CCT/TCC	234	20.49%
8.	AAC/CAC/ACC/CCA/CAA/ACA	169	14.79%
9.	CAG/CGA/AGC/ACG/GCA/GAC	143	12.52%
10.	TGT/GTG/TTG/TGG/GTT/GGT	63	5.51%
11.	CGG/CCG/CGC/GCG/GCC/GC	51	4.46%
12.	TGC/GCT/CGT/GTC/CTG/TCG	109	9.54%
13.	TCA/CAT/ATC/ACT	51	4.46%
14.	GGA/GAA/AAG/AGA/GAG	215	18.82%
15.	ATG/TGA/GAT/GTA/AGT/TGA	55	4.81%
16.	TAA/TTA/AAT/ATT/TAT/ATA	52	4.55%
	TOTAL	1142	

Tab. 1: Results of EST sequence assembly

Total Number of EST Sequences	Number of Singletons	Number of Contigs	Number of assembled sequences	% Reduction in Redundancy
330181	59343	5619	64962	80.33%

Tab. 3: List of EST-SSR markers designed

Sr.No.	Sequence ID	Sequence	Base pairs
1.	gi226857985	**F**: CGGTGCATGTACAACACGTA **R**: TCCAGCAGAAGAGGTGATGA	20 20
2.	gi226857903	**F**: GTATTGGCTCCACCACAACC **R**: GAAACCCGGTCGTGTAAAAA	20 20
3.	gi226857895	**F**: AAAGTGGCGTAAGGTTGTGG **R**: CACGCTGGACACATACATCC	20 20
4.	gi226857649	**F**: TTTCTTCTCTTCTCCAATCCTCTG **R**: TTGCAGAGAAAGATCAAGGTGA	24 22
5.	gi226857752	**F**: GGCCCAAATGTCATCGAATA **R**: TATTCCAAATTCCCGGACTG	20 20
6.	gi226814813	**F**: AACATATGGGGCAGCGTATT **R**: TTGTGATGCTTGTGGGGATA	20 20
7.	gi226811802	**F**: TCCTCTGGAGTAGCGAGCAC **R**: ACCGTCAACGAGGCTCAG	20 18
8.	gi226805503	**F**: CAACAAAGCCTTTTCCCAGT **R**: TTATTCGGCCTTTGTTTTGG	20 20
9.	gi226798745	**F**: CACGTTTCCAAGAGCCTTACA **R**: TTTGCCCTTAAAAGTTGGGTTA	21 22
10.	gi226796317	**F**: TCATCGCCACCTTGATGATA **R**: AGCTCAAAGAGGCCGTCATA	20 20
11.	gi226794476	**F**: CACTTGGGCCAATTTCTGAC **R**: CCAAGGAGTGAGTGGAGGAG	20 20
12.	gil226814826	**F**: TTCCAACGACACCAACCTTT **R**: GCGATGATGTATGGCACAGA	20 20
13.	gil226812665l	**F**: TCCGACAAATCAACATTGGA **R**: CTGACGGACGCTCATAACAA	20 20
14.	gil226810461	**F**: TGATGATGACGATGACGATG **R**: GAGCAAAAGTTGAAACCCTCA	20 20
15.	gil226810487l	**F**: GGGCTGCTTGATTTGAACTT **R**: GAAGCTAAAACTTCACCCTTGA	20 22
16.	gil226810743l	**F**: TGTTCATGTCGGTGCTCAAT **R**: GACGATGATCAGGCCATTCT	20 20
17.	gil226811551l	**F**: ACTGAGGGCTTGTGTTGGTC **R**: GCTTCTGAACGAGGAAGACG	20 20
18.	Contig 48-1	**F**: TGGTGTGCCCTGAAGTGTTA **R**: TATGGGAGCAGCAAATACCG	20 20
19.	Contig 51-1	**F**: AACAAGGAACCTATTTTCCGAAC **R**: CAGGGAGAAACGAGAGGTCA	23 20
20.	Contig 175-1	**F**: CCAGTAGAGGGCCAAAAACA **R**: TCTGTCACCGGATAGTGCTG	20 20
21.	Contig 208-1	**F**: AGCCCAAAAGGCAGTCTACA **R**: ACCCCCATTTTCAGCCTATC	20 20
22.	Contig 139-2	**F**: CGCTCCAGAAGTCAACATCA **R**: GGAGCTCCAATTCCATTTCA	20 20
23.	Contig 10-1	**F**: TGTTAGATCCGGACCCACTC **R**: AGATGGAAGGGGAGAAGAGC	20 20
24.	Contig 1-2	**F**: TGAGGTGACCACAAGCAAAG **R**: CAATTCTTTTGTGCGCGTAT	20 20
25.	Contig 27-1	**F**: TTTTCCCCCTCAAGAACAAA **R**: TGAAGCCGAAGGTTCATCAT	20 20

Primer designing

SSR primers were successfully designed for 25 sequences using PRIMER 3, for the purpose of polymorphism study. 17 singletons and 8 contigs were selected from the assembled data set and used for primer designing (Tab. 3).

BLASTX analysis for sequence identification

BLASTX analysis was carried out for the EST sequences selected for primer designing. Based on this analysis, a putative function could be assigned to 22 of the potential EST-SSR markers (15 singletons and 7 contigs) assuming a threshold value of <1E-5. The annotation results indicated that 13 EST-SSR sequences showed high homology with *Prunus persica* proteins, while four were observed to reveal high similarity with well characterized Malus proteins.

Analysis of SSR-protein domains

The 25 SSR containing sequences used for primer deigning were also analyzed for protein families and domains through PROSITE. A functional protein domain could be assigned to 23 sequences. The sequences were recorded to possess functional domains responsible for protein signatures like 2Fe-2S ferredoxin binding sites, anaphylatoxin domain signature, VWFC domain and thiolases active site. The SSR-FDMs provide information regarding transcribed genetic markers having putative functions.

SSR-ESTs of apple were gene ontologically assigned with their major molecular function, biological process and cellular component categories. Gene ontology for the corresponding SSRs was determined on the basis of sequence, domain and motif similarity. This indicates an indirect role of above SSR-ESTs in development of genetic markers for the protein and enzymes involved in various cellular processes.

Marker validation by PCR amplification

The 25 primer pairs were used to study genetic polymorphism in a set of 48 apple accessions (including cultivated varieties, rootstocks and wild relatives). All the primers amplified discrete DNA fragments, 24 being polymorphic (96%) and detected 94 alleles. The total level of polymorphism among the studied accessions of apple found to be 93%. PIC values were calculated for all the polymorphic primers. The PIC was found to range from 0.07 to 0.81. Fig. 1 shows amplification pattern obtained with primer E19.

Jaccard's similarity matrix coefficient was obtained through NTSYS-pc. The similarity coefficient values ranged from 0.303 to 0.903.

The generated dendrogram revealed a wide genetic base of the apple germplasm collection studied. The dendrogram generated for EST-SSR markers divided the genotypes into three main clusters, A, B and C (Fig. 2). Group A could be further classified into subgroup A1 and A2. Subgroup A1 comprised of 'Early Red One', 'Red Fuji', 'Compact Winter Banana', 'Black Ben Davis', 'Granny Smith', 'Spartan', 'Top Red', 'Early McIntosh', 'Well Spur', 'Anna', '*Malus orientalis*', 'Chaubattia Ambroise', 'Ambred', '*Malus baccata* Kashmir', '*Malus sargentii*', 'M7', 'Red Chief', 'Hardeman' and 'Vance Delicious'. Subgroup A2 contained '*Malus simcoe*', '*Malus baccata* Kinnaur', 'B9', 'Gale Gala', 'Scarlet Spur', 'M9' and 'M793'. Group B formed the other major cluster. It could be subgrouped into B1 and B2. B2 was found to contain only two accessions, i.e. 'Gloster' and 'Red Spur'. ' B1 comprised of accessions, 'Ambroyal', 'Bright-N-Early', 'Royal', '*Malus baccata* Shillong', 'Golden Spur', 'M26', 'MM106', 'Oregon Spur', 'Mutsu', '*Malus eseltine*', 'Skyline Supreme', 'Silver Spur', 'Hardispur', 'Golden Delicious', 'Scarlet Gala', 'MM111', 'Miller Sturdy Spur' and 'Criterion'. Group C contained 'Red Delicious' and 'Tydeman's Early Worcester'.

Discussion

EST databases represent a potentially valuable resource for developing molecular markers for evolutionary studies. Since EST-derived markers come from transcribed regions of the genome they are likely to be conserved across a range of genomes than other types of markers (PASHLEY et al., 2006). This is promising, considering the increasing number of EST-derived SSR markers in various crops.
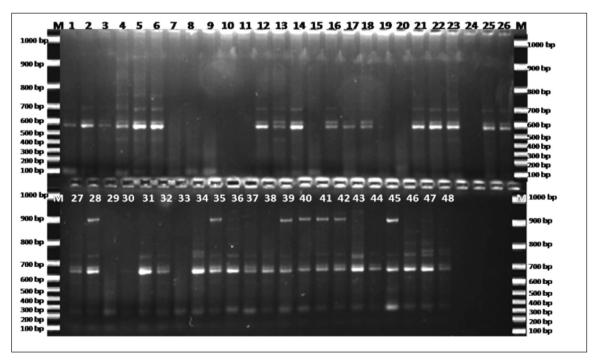

Fig. 1: Banding pattern with EST-SSR primer E19; 1-48: Apple accessions; M- Molecular weight ladder
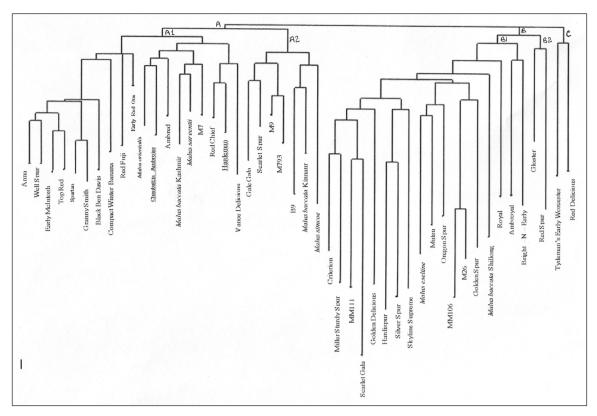
Fig. 2: Dendrogram showing genetic relationship among 48 apple accessions

Theoretically, the frequency of a repeat should decrease as its length increases. Accordingly, in the present study dinucleotide repeats are reported to be more abundant among the SSR types (72%). Similar reports have been obtained in other plant species (LI et al., 2002). The most dominant dinucleotide repeat motif was CT/TC followed by GA/AG (37% and 34%, respectively). A high abundance of (GA)n microsatellites has also been reported in some other plant genomes (GUYOMARC et al., 2002; GUPTA and VARSHNEY, 2000; SUWABE et al., 2002). Preference was given to trinucleotide repeats in SSRs, based on the hypothesis that di-, tetra-, or pentanucleotide variation would produce shifts in the reading frame that would result in negative selection and a lesser degree of polymorphism (VARSHNEY et al., 2005)

Marker Validation and Genetic diversity
ESTs have been effectively utilized for the development of SSR markers which have been subsequently used in the genetic diversity analysis (KUMAR et al., 2011; VAIDYA et al., 2012). EST-derived markers have been well documented in marker assisted selection that involves mainly the identification of QTL for important traits, genetic mapping, genetic diversity, and comparative genomics (CLOUTIER et al., 2009; DURAND et al., 2010; KUMAR et al., 2011; VAIDYA et al., 2012). EST-SSRs also have the potential for direct linkage association to functional genes encoding important horticultural traits. The level of polymorphism in the apple accessions revealed using the developed SSR markers, was relatively high, though a limited number of primer pairs had been used. A total of 25 EST-SSRs were used to amplify DNA from 48 apple accessions. All the primers yielded scorable bands, out of which 24 were found to be polymorphic. It was not possible to reach 100% success rate to design and amplify SSR primers, as was also reported by SZEWC-MCFADDEN et al. (1996). The PIC values, which provide an estimate of the discriminatory power of a locus or loci, recorded for all the informative

primers were found to range between 0.07 to 0.81. The success rate for EST-SSR primers (percentage of SSR primers producing discrete amplification products) in various reports was between 60 and 90% (VARSHNEY et al., 2005).

In this study, the relatively high level of polymorphism observed with apple EST-SSRs at the cultivar level might suggest that the apple cultivars used in this study were derived from cross-pollination and propagated vegetatively, and thus they might have a high level of heterozygosity, which can increase the polymorphism level.

Conclusion

The present study attempts to ascertain the frequency and distribution of SSRs in the apple EST database and develop those EST-SSRs for use in genetic studies. We demonstrated the utility of computational approaches for mining SSRs from ever increasing repertoire of publicly available plant EST sequences present in different databases. The resulting EST-SSR set is a valuable tool for further genetic and genomic applications. The developed 25 EST-SSR markers have a high rate of PCR amplification and can be used in apple breeding and genetic studies. The use of these markers would reduce the cost and therefore facilitate cultivar identification, genetic distance assessments, gene mapping and possible marker-assisted selection (MAS). The functional categorization of these markers corresponded to many genes with biological, cellular and molecular functions, thus providing an opportunity to investigate the consequences of SSR polymorphisms on gene function.

References

ANDERSON, J.A., CHURCHILL, G.A., AUTRIQUE, J.E., TANKSLEY, S.D., SORRELLS, M.E., 1993: Optimizing parental selection for genetic linkage maps. Genome 36, 181-186.

CLOUTIER, S., NIU, Z.X., DATLA, R., DUGUID, S., 2009: Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor. Appl. Genet. 119, 53-63.

DURAND, J., BODENES, C., CHANCEREL, E., FRIGERIO, J.M., 2010: A fast and cost-effective approach to develop and map EST- SSR markers: oak as a case study. BMC Genomics 11, 570.

GASIC, K., HAN, Y., KERTBUNDIT, S., SHULAEV, V., IEZZONI, A.F., STOVER, E.W., BELL, R.L., WISNIEWSKI, M.E., KORBAN, S.S., 2008: Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. Mol. Breeding 23, 397-411.

GUPTA, P.K., VARSHNEY, R.K., 2000: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113, 163-185.

GUYOMARC, H., SOURDILLE, P., CHARMET, G., EDWARDS, K.J., BERNARD, M., 2002: Characterization of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. Theor. Appl. Genet. 104,1164-1172.

HUANG, X., MADAN, A., 1999: CAP3: A DNA sequence assembly program. Genome Res. 9, 868-877.

IGARASHI, M., ABE, Y., HATSUYAMA, Y., UEDA, T., FUKASAWA-AKADA, T., KON, T., KUDO, T., SATO, T., SUZUKI, M., 2008: Linkage maps of the apple (*Malus × domestica* Borkh.) cvs 'Ralls Janet' and 'Delicious' include newly development EST markers. Mol. Breed 22, 95-118.

KITAHARA, K., MATSUMOTO, S., YAMAMOTO, T., SOEJIMA, J., KIMURA, T., KOMATSU, H., 2005: Molecular characterization of apple cultivars in Japan by SRNase analysis and SSR markers. J. Amer. Soc. Hort. Sci. 130, 885-892.

KUMAR, Y.H., RANJAN, A., ASIF, M.H., MANTRI, S., 2011: EST-derived SSR markers in *Jatropha curcas* L. development, characterization, polymorphism, and transferability across the species/genera. Tree Genet. Genomes 7, 207-219.

LI, Y.C., KOROL, A.B., FAHIMA, T., BEILES, A., NEVO, E., 2002: Microsatellites: genomic distribution, putative functions, and mutational mechanisms (a review). Mol. Ecol. 11, 2543-2565.

LIEBHARD, R., GIANFRANCESCHI, L., KOLLER, B., RYDER, C., TARCHINI, R., VAN DE WEG, E., 2002: Development and characterization of 140 new microsatellites in apple (*Malus × domestica* Borkh.). Mol. Breed. 10, 217-241.

MASOUDI-NEJAD, A., KOICHIRO, T., SHUICHI, K., YUKI, M., MASANORI, S., MASUMI, I., MINORU, K., TAKASHI, E., SUSUMU, G., 2006: EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res. 34, 459-462.

NEWCOMB, R.D., CROWHURST, R.N., GLEAVE, A.P., RIKKERINK, E.H.A., ALLAN, A.C., BEUNING, L.L., BOWEN, J.H., GERA, E., JAMIE-SON, K.R., JANSSEN, B.J., LAING, W.A., MCARTNEY, S., NAIN, B., ROSS, G.S., SNOWDEN, K.C., SOULEYRE, E.J.F., WALTON, E.F., YAUK, Y.K., 2006: Analyses of Expressed Sequence Tags from Apple. Plant Physiology 141(1), 147-166.

PASHLEY, C.H., ELLIS, J.R., MCCAULEY, D.E., BURKE, J.M., 2006: EST databases as a source for molecular markers: lessons from *Helianthus*. J. Heredity 97, 381-388.

POTTS, S.M., HAN, Y., KHAN, M.A., KUSHAD, M.M., RAYBURN, A.L., KORBAN, S.S., 2012: Genetic diversity and characterization of a core collection of *Malus* germplasm using simple sequence repeats (SSRs). Plant Mol. Biol. Rep. 30, 827-837.

ROZEN, S., SKALETSKY, H.J., 2000: Primer 3 on the www for general users and for biologist programmers, In: Krawetz, S., Misener, S. (eds.), Bioinformatics methods and protocols: methods in molecular biology, Humana Press, Totowa NJ.

SUWABE, K., IKETANI, H., NUMONE, T., KAGE, T., HIRAI, M., 2002: Isolation and characterization of microsatellites in *Bassica rapa* L. Theor. Appl. Genet. 104, 1092-1098.

SZEWC-MCFADDEN, A.K., KRESOVICH, S., BLIEK, S.M., MITCHELL, S.E., MCFERSON, J.R., 1996: Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated Brassica species. Theor. Appl. Genet. 93, 534-538.

TEMNYKH, S., DECLERK, G., LUKASHOVA, A., LIPOVICH, L., CARTINHOUR, S., MCCOUCH, S.R., 2001: Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11, 1441-1452.

VAIDYA, E., KAUR, R., BHARDWAJ, S.V., 2012: Data mining of ESTs to develop dbEST-SSRs for use in a polymorphism study of cauliflower (*Brassica oleracea* var. *botrytis*). J. Hortic. Sci. Biotech. 87, 57-63.

VARSHNEY, R.K., GRANER, A., SORRELS, M.E., 2005: Genic microsatellite markers in plants: features and applications. Trend Biotechnol. 23, 48-55.

ZHANG, Q., LI, L., ZHAO, Y., KORBAN, S., HAN, Y., 2012: Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. Plant Mol. Biol. Rep. 30(3), 1-8.

Address of the authors:

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