

Mining of simple sequence repeats in the Genome of *Gentianaceae*

R. Sathishkumar¹, P. T. V. Lakshmi^{1,2}, A. Annamalai³, V. Arunachalam⁴

¹Phytomatics Laboratory, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, ²Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry, ³Plant Cell and Molecular Biology Laboratory, Department of Biotechnology, Karunya University, Coimbatore, Tamil Nadu, ⁴Molecular Biology and Bioinformatics Laboratory, Central Plantation Crops Research Institute, Kasaragod, Kerala, India.

Submitted: 29-07-2010

Revised: 25-09-2010

Published: 07-04-2011

ABSTRACT

Simple sequence repeats (SSRs) or short tandem repeats are short repeat motifs that show high level of length polymorphism due to insertion or deletion mutations of one or more repeat types. Here, we present the detection and abundance of microsatellites or SSRs in nucleotide sequences of *Gentianaceae* family. A total of 545 SSRs were mined in 4698 nucleotide sequences downloaded from the National Center for Biotechnology Information (NCBI). Among the SSR sequences, the frequency of repeat type was about 429 -mono repeats, 99 -di repeats, 15 -tri repeats, and 2 --hexa repeats. Mononucleotide repeats were found to be abundant repeat types, about 78%, followed by dinucleotide repeats (18.16%) among the SSR sequences. An attempt was made to design primer pairs for 545 identified SSRs but these were found only for 169 sequences.

Key words: *Gentianaceae*, nucleotide, simple sequence repeats

Access this article online

Website:
www.phcogres.com

DOI:
10.4103/0974-8490.79111

Quick Response Code:



INTRODUCTION

Gentianaceae, or the Gentian family, is a family of flowering plants of 87 genera and over 1650 species.^[1] Plants are usually rhizomatous. These are annuals or perennials, mostly upright though a few species lie on the ground and have upright branch tips. Leaves are opposite or whorled with entire edges and bases connately attached to the stem, mostly without a petiole. Flowers have four to five sepals, petals, and stamens, but only one pistil. Sepals and petals are fused at the base, with four to five free lobes above. Stamens alternate with the corolla lobes. Ovary is superior; fruit is a capsule. Stipules is absent. Plants usually accumulate bitter iridoid substances; bicollateral bundles are present. The fruits are dehiscent septical capsules splitting into two halves. The *Gentianaceae* contains many species with interesting phytochemical properties. They have been widely used in traditional medicine and also as constituents in bitters and similar concoctions. The family

consists of trees, shrubs, and herbs showing a wide range of colors and floral patterns.

Simple sequence repeats (SSRs),^[2] or microsatellites^[3] or short tandem repeats,^[4] are short (1–6 bp) repeat motifs that show a high level of length polymorphism due to insertion or deletion mutations of one or more repeat types.^[5] Studies suggest that both protein coding and noncoding regions of DNA sequences contain SSRs.^[6] SSRs present in coding sequences are less polymorphic than those in the genomic sequences. Moreover, different taxon varies in abundance of different types of SSRs and these are present in greater abundance in noncoding regions than coding SSRs.^[7] The SSRs are either developed conventionally^[8] or from sequence databases.^[9] PCR-based techniques such as AFLP and microsatellites or SSRs have also played important roles in plant DNA profiling. Primers are essential components of PCR-based systems as well as modern microarray systems which utilize appropriate probes for PCR amplification.^[10]

In genetics, a sequence motif is a nucleotide or amino acid sequence pattern that is widespread and is believed to have, a biological significance. When a sequence motif appears

Address for correspondence:

Dr. PTV. Lakshmi, Centre for Bioinformatics
School of Life Sciences, Pondicherry University,
Puducherry - 605 014, India. E-mail: lakanna@bicpu.edu.in

in the exon of a gene, it may encode the "structural motif" of a protein, that is, a stereotypical element of the overall structure of the protein. "Noncoding" sequences are not translated into proteins. Outside of gene exons, there exist regulatory sequence motifs and motifs within the "junk," such as satellite DNA.^[11] Robinson *et al.*^[12] developed a computer program to identify and design PCR primers for amplification of SSR loci based on available DNA sequence information. SSR primers have been designed using publicly available expressed sequence tags (ESTs) in barley,^[13] almond (*Prunus communis* Fritsch.), and peach (*P. persica* (L.) Batsch.),^[14] *T. aestivum*, and *O. Sativa*.^[15] These SSRs are useful as molecular markers because their development is inexpensive, they represent transcribed genes, and their putative function can often be deduced by a homology search.^[16] SSRs have been the backbone to creating molecular maps for a number of years.

The increasing number of genomic and expressed sequences in public databases provides a valuable source for bioinformatical data mining. However, there are a number of exciting application of these sequence data; used in comparative genome analysis – to trace the evolution among the related species, to study the genome structure and their gene functions. Comparative genome analysis requires the same sets of genes (i.e., cross-reference genes) to be mapped to chromosomes in the species compared. Thus, comparative maps with sets of EST-derived markers (i.e., cross-species markers) are essential for comparative genome analysis. Several studies have utilized publicly available ESTs to mine SSRs or microsatellites markers for plants,^[17-20] catfish,^[21] insects,^[22] animals,^[23] and human.^[24] The EST-derived SSR markers (EST-SSRs) have proved very useful for the construction of genetic and comparative maps.^[25] The software used here is MISA, a microsatellite identifying tool which has the advantage of detecting the mono- to decamer repeats and also compound repeats. But it has the disadvantage of inability to detect above decanucleotide repeats. Riju and Arunachalam,^[26] mined the SSRs in oil palm ESTs with five different software and have reported that MISA program has given maximum coverage of SSRs in both oil palm ESTs and Contigs.

PCR primer design in general

Understanding of primer properties is very important for primer design. The major aspects of primer properties include specificity, melting temperature (T_m), and intraprimer or interprimer homology. Primer specificity is mostly determined by the 3'-end sequences. It was reported that single internal mismatches had no significant effect on PCR product yield while the 3'-terminal mismatches, especially the A:A, A:G, G:A, and C:C mismatches, markedly reduced overall PCR product yield.^[27] Khabar *et al.*^[28] assessed the annealing specificity of primers in PCR

reactions under different annealing temperatures (35°C, 40°C, and 45°C) and found perfect matches between at least eight bases at the 3'-end of the 5'-primers and the target region, whereas mispriming occurred only toward the 5'-end. Therefore it is critical to include 8–10 unique bases at the 3'-end of the primer.

Ideally the primer has a T_m in the range of 50–65°C, random nucleotide composition with a 40–60% GC-content, and 18–30 bases long. The intraprimer or interprimer homology is kept as low as possible to avoid formation of hairpin structures or primer dimmers (>3 bp complementarities between primers) which otherwise will interfere with annealing of primer to the DNA template.^[29]

ESTs, which represent the expressed part of genome, also serve as a source of SSRs.^[9] Detection of SSRs facilitates the development of SSR markers that are useful in the study of genetic variation, gene tagging, and linkage mapping,^[30] and are also useful across a number of related species.^[13] Microsatellites can be amplified for identification by the polymerase chain reaction (PCR) process, using the unique sequences of flanking regions as primers. Once the potentially useful microsatellites are determined (removing nonuseful ones such as those with random inserts within the repeat region), the flanking sequences can be used to design oligonucleotide primers which will amplify the specific microsatellite repeat in a PCR. Microsatellite loci are widely distributed throughout the genome and can be isolated from semidegraded DNA of older specimens, as all that is needed is a suitable substrate for amplification through PCR. Hence, the present study was to find out the distribution and abundance of SSRs for the development of markers and to annotate SSR-containing sequence in *Gentianaceae* family. Nucleotide database, which contains sequences of well-characterized genes as well as hundreds of thousands novel EST sequences, was retrieved to perform the analysis.

MATERIALS AND METHODS

Retrieval of nucleotide sequences and detection of SSRs

A total of 647 nucleotide sequences of *Gentianaceae* were downloaded from the NCBI (<http://www.ncbi.nlm.nih.gov/Nucleotide/?term=Gentianaceae>) and harvested for SSRs using a perl script. The minimum length of SSR was fixed at 14 bp according to the criteria used by Gupta *et al.*^[31] The SSRs were defined as 14-bp mononucleotide or dinucleotide repeats; 15-bp trinucleotide repeats; 16 tetranucleotide repeats; 20 pentanucleotide repeats; 18 hexanucleotide repeats. The poly A and poly T repeats were removed by using an inhouse developed perl script,

as these are not considered as SSRs due to their presence at 3'-end of mRNA/cDNA sequences.

Primer designing for SSRs

A pair of primer flanking each SSR was designed using FastPCR software available at www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi, which takes input according to user-defined conditions and pick primers according to these specified parameters. Default parameters of the FastPCR, viz, the optimum primer size of 20.0 (the range was 18–28), the optimum annealing temperature of 60.0 (the range was 57.0–63.0), and the range of % GC content of 44–60, were selected for primer designing.

Detection of SSR positions with respect to open reading frames

Open reading frames (ORFs) are predicted for all the SSR-containing sequences using ORF finder available at NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) using standard genetic code. Sequence fragments with maximum length uninterrupted by stop codon were taken as the primary encoding segment (ORF) of the query sequences. In all the predicted ORFs, the relative positions of SSRs were detected, that is, whether the SSR was present within the ORF, in the 5' UTR untranslated region (UTR) or in the 3' UTR.

RESULTS

Screening of *Gentianaceae* sequences for SSRs

In the present study, 4698 nucleotide sequences of *Gentianaceae* available at NCBI (<http://www.ncbi.nlm.nih.gov>) were searched for SSRs with a minimum length of 18 bp. A total of 545 SSRs were detected from 2889 kb of data mined, excluding poly A and poly T. Depending upon the length of the repeat unit itself (1–6 bp), the lengths of the identified SSRs varied from 14 to 48 bp, respectively.

Frequencies of classified repeat types of *Gentianaceae*

From a number of 4698 sequences screened, only a subset of 461 sequences contained 545 SSRs, suggesting that merely 9.83% of sequences contained SSRs. The frequencies of SSRs with mono-, di-, tri-, tetra-, and hexanucleotide repeat units showed the frequent repeat type within the nucleotide sequences of *Gentiana* family that were found to be in mononucleotide (84.58%) followed by dinucleotide repeats (18.16%), trinucleotide (2.75%), and hexanucleotide (0.65%), respectively [Figure 1]. Whereas, no tetranucleotide and pentanucleotide repeat was detected during the analysis.

The observed frequency of different repeat types comprising the SSRs is presented in Figure 2a–d and summarized in Table 1. SSRs were comprised of four

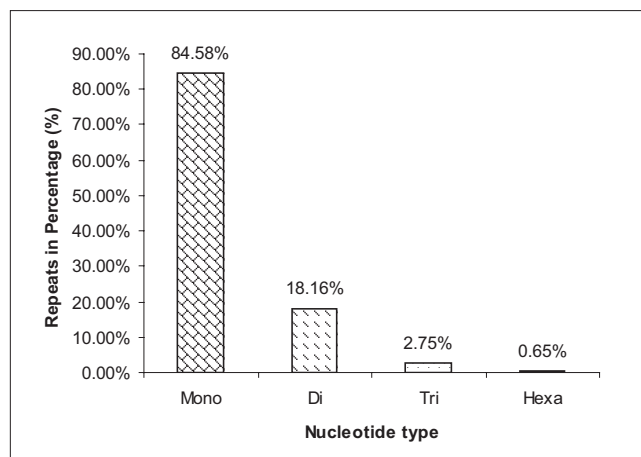


Figure 1: Frequency distribution of different repeat types identified in nucleotide sequences of *Gentianaceae*

different types of mononucleotide (A, T, C, and G), nine different types of dinucleotide (CA)_n, (TG)_n, (AC)_n, (GA)_n, (CT)_n, (TA)_n, (AT)_n, (GC)_n, (TC)_n, (AG)_n, (GT)_n repeats, seven different types of trinucleotide (GAG)_n, (ATG)_n, (CTT)_n, (TTA)_n, (CAA)_n, (AAC)_n, (ACA)_n repeats, and two types of hexanucleotide (CCACAC)_n, (GGTCAA)_n repeats.

Designing of primers for SSRs

Out of 545 SSRs detected, the primers could be designed only for 169 (31%) SSRs and the rest 376 (69%) sequences did not produce any acceptable primers. These 169 SSRs for which primers were designed include 133 mono-, 29 di-, 7 tri-, and no hexanucleotide repeats. The details of the accession numbers of nucleotide sequences of *Gentiana*, repeat motif of SSRs for which primer were designed, primer sequences, GC%, product size, and annealing temperature are given in Table 2.

Prediction of ORF in SSR-containing sequences

An attempt was made to predict the ORFs in SSR-containing sequences using ORF finder. Out of the 545 SSRs identified, the positions of 359 SSRs with respect to ORF were determined, while for the remaining 186 SSR-containing sequences, no ORF were predicted. Of these 359 SSRs, a large number of 161 (44.84%) were present in the 5' untranslated region, 129 (35.93%) SSRs occurred within ORF, and the remaining 69 (19.22%) occurred in the 3' untranslated region.

DISCUSSION

In the present study, a large number of nucleotide sequences (4698) of *Gentiana* retrieved from NCBI were mined for SSRs. In the sequences that were mined the SSRs were characterized, and a subset of these SSRs was

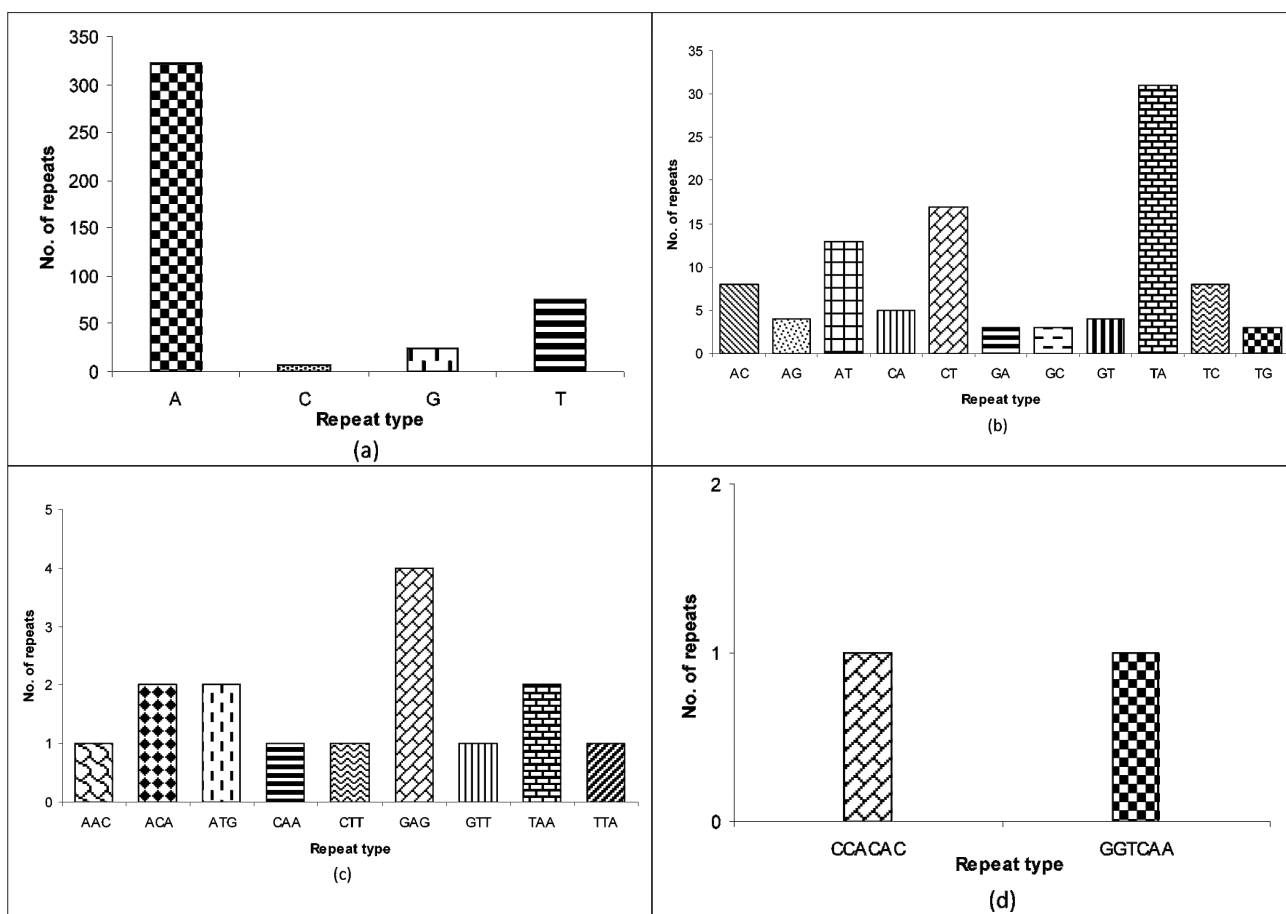


Figure 2: Frequency distribution of (a) mono-, (b) di-, (c) tri-, and (d) hexanucleotide repeat motifs in the genome of *Gentianaceae*

Table 1: Summary of in silico mining of Nucleotide sequences of *Gentianaceae*

Parameters	Values
Total number of sequences searched	4698
Total number of SSRs after removing poly A and poly T	545
Total size of examined sequences (bp)	2289303
Total number of sequences containing single SSRs	429
Number of sequences containing two SSRs	99
Number of sequences containing three SSRs	15
Number of sequences containing six SSRs	2
Number of sequences containing more than one SSR	57
Number of SSRs present in compound formation	47
Repeat type	
Mononucleotide	429 (84.58)
Dinucleotide	99 (18.16)
Trinucleotide	15(2.75)
Hexanucleotide	2 (0.65)

Data in parentheses is the percentage value of the repeat

used for designing the markers. A total of 545 SSRs was detected and this was in accordance to the findings of^[32] who reported that the abundance of different repeats varied broadly depending upon the species.

Microsatellites or SSRs are stretches of DNA containing tandem repeats of di-, tri-, tetra-, and above nucleotide

units ubiquitously distributed throughout the eukaryotic genome. They are found to be abundant in plant genomes and are thought to be the major sources of genetic variation in quantitative traits. The abundance of the different repeat motifs (1–6 bp) in the SSRs as detected in *Gentiana* family during the present study was variable so that the SSRs with different repeat motifs were not evenly distributed. The SSRs with dinucleotide repeats (18.16%) were abundant. This is in agreement with the results of earlier studies on *Arabidopsis* in which the dinucleotide repeats were also found to be abundant,^[33] perhaps because the genomic sequences of this species may include SSRs in noncoding regions too. The smaller repeat motifs were found to be predominant among SSRs identified and as the length of repeat unit increases, their occurrence decreases. We excluded poly A and poly T repeats due to which their number is under-represented. The abundance of trinucleotide SSRs may be attributed to the absence of frame shift mutations due to variation in trinucleotide repeats.^[34]

Molecular genetic markers can be used to examine a group of individuals or populations to estimate various diversity measures and genetic distances, intergenetic structure and

Table 2: Synthesis of primer

Accession number	Motif	Forward Primer		Reverse primer		Annealing temperature (C)	Product size (bp)
		Primer	GC %	Primer	GC %		
gi 261865195 gb GQ864096.1	(TA)7(AT)7	GGCAGAACTCAACGGGAAAGC	59.1	GAGGCTAAAGGGTCTCGGA	58.8	57.9	709
gi 169798749 gb EU370947.1	(A)13	CGAGGCTGTAAGTGCCTGC	60	GACGGTGAGAGGCTGTGTATG	57.1	56.85	415
gi 24494572 gb EF371451.1	(TG)17	GAGGCTGTAAGTGCCTGC	50	GACGGTTAGAGGCTGTGTAGG	54.5	54.25	414
gi 124494568 gb ef371447.1	(AG)7(TG)8 (GA)12	GTCTGAGTGAGGGAGCCATC	60	GTGATGCTGTGTCTCCCAAGAG	54.5	56.8	80
gi 112293600 gb dq822583.1	(CA)29(GA)21(AG)6	GGGATGGCACTCACCTACAGC	61.9	GGCATAAGTGTAGTCCAC	55	54.45	123
gi 112293594 gb dq822577.1	(CT)21	GCGTTTCACTCCGACTCCCGAG	65	GCAGAGCATAGTGGACCG	61.1	62	656
gi 156787489 gb ef569230.2	(A)21	CTGTGTGATGCTGAGGTTGC	55	GGCTGAACCCAGGGACAACC	65	59	461
gi 89199712 gb dq398766.1	(A)16	GATAGATTTCCAGAGGACG	50	GGTTGACACTTACAGCACG	50	57	760
gi 89199711 gb dq398765.1	(A)10	GCGATGTGTTCAAGACCG	55	GTGAACGCCTATCCAGTG	55.6	59	676
gi 89199709 gb dq398763.1	(A)16	CGCTGTTCCGCTCACGCTTC	65	GTTGACACTTACAGCCCCT	50	63	409
gi 89199708 gb dq398762.1	(A)16	GACGCTGTAAGTGTATAG	44.4	CGCTTAGGCTGACTTCGTG	57.9	56	349
gi 89199707 gb dq398761.1	(A)12	GTCGCTGAGTAGTGCCAAAG	57.9	GGCCGTTAAGGGCTGACACG	70	64	345
gi 89199706 gb dq398760.1	(A)12	GCTGCTGTAAGTGCCAAAG	55	GAGGCACGATTAGGGGCTGAC	61.9	65	414
gi 89199705 gb dq398759.1	(A)12	GAGCCTGTAAGTGCCAAAC	52.6	GTGCTGTGTAGAGCCTGAACCG	59.1	64	806
gi 89199704 gb dq398758.1	(A)11	GACGCTGTAAGTGCCAAAC	52.6	GTAGGCGTAACCTGTGACAC	52.6	58	228
gi 89199703 gb dq398757.1	(A)11	GAGGCTGTAGAGTGCCATC	57.1	CGCGGGATAGAGATAGAG	56.6	61	539
gi 89199702 gb dq398756.1	(A)11	GAGTCTGTAGAGTGCCAAAG	60	CGGAGATTAGAGCGTGAACAG	52.4	64	313
gi 89199701 gb dq398755.1	(A)11	GTCGCTGTAAGTGCTAAC	47.4	GTCATCAAGTCCGACAGCG	57.9	58	346
gi 89199700 gb dq398754.1	(A)12	CTGTTCAAAGGACCCGTGC	57.9	GAGCCTTTTGTCTGCTGGTG	60	56.35	415
gi 89199699 gb dq398753.1	(A)12	GGCGGCTGTAAGTCCATCC	63.2	GCTTGCTAACGGATTCTGCG	55	58	409
gi 89199698 gb dq398752.1	(A)11	GGAGAGTAAAGTGCCAAAC	55.6	GAGCATTGTCGGAACGAGCG	60	56.55	409
gi 89199697 gb dq398751.1	(A)11	GGAGGCTGGCTAAAGTGCTAC	57.1	GAGCCATTGCGAACAGTG	55.6	56.55	349
gi 89199696 gb dq398750.1	(A)15	CAATGTTCAAGCGATGCCGTC	52.4	CTATGGTCAGAAACGGTGG	57.9	57.05	345
gi 89199695 gb dq398749.1	(A)13	CATCAGAGTTCAAAGACCCG	52.6	GATACCTTGTCTACGAACCG	50	57.35	419
gi 89199694 gb dq398748.1	(A)21	GGAGTCTGCTAAGTGCCAAAC	57.1	GCACCTTGTCTGGAACGGCGG	70	58	349
gi 89199693 gb dq398747.1	(A)15	CATCGTGTCAAGGACCCG	55.6	GAACCTCTGTCGGAACCG	61.1	57.05	349
gi 89199692 gb dq398746.1	(A)12	GTGCTGTATCAAGGACCCG	57.9	GAGCCTTGTGAACGGTGG	61.1	57.05	141
gi 89199690 gb dq398744.1	(A)12	GGAGGCTAAGTGCCAAAC	55.6	GGTCTGGGTTAGGTTGG	66.7	59.75	143
gi 89199689 gb dq398743.1	(A)14	GGACGGTGTAAAGTGCCAAAC	57.9	GAAGGTCACCGTTAGGT	55.6	59.75	417
gi 89199688 gb dq398742.1	(A)15	GGAGGAGTAAAGTGCCAAAC	52.4	GGCTAATGTCGTCGGGAGG	63.2	58	417
gi 89199687 gb dq398741.1	(A)15	GGAGGAGTAAAGTGCCAAAC	55	GCCATTGTCGGAACGGAGG	63.2	58	415
gi 89199686 gb dq398740.1	(A)12	GGAGGAGTAAAGTGCCAAAC	55	GCCTTGTCTGGAACGGTGA	63.2	58	415
gi 89199685 gb dq398739.1	(A)12	GGAGGTCGTAAGTGCCAAAC	57.9	GCCTTGTCTGGAACGGTGG	66.7	58	417
gi 89199684 gb dq398738.1	(A)14	GGAGGTCGTAAGTGCCAAAC	57.9	GCCATTGTCGGAACGGAGG	63.2	58	411
gi 89199683 gb dq398737.1	(A)14	GGAGGTCGTAAGTGCCAAAC	57.1	GAGCCTATGTCGGAACCG	61.1	58	413
gi 89199682 gb dq398736.1	(A)10	GGGTCTTAAGTGCCAAAC	55.6	GCCAGTCTCGGAACAGTG	63.2	56.55	413
gi 89199681 gb dq398735.1	(A)12	GGAGGTCGTAAGTGCCAAAC	55	GCGGATTGTGGAACGGCG	66.7	56.55	413
gi 89199680 gb dq398734.1	(A)12	GGAGGTCGTAAGTGCCAAAC	57.9	GGCTATGCTCAGTCAGGG	61.1	56.55	413
gi 89199679 gb dq398733.1	(A)12	GGAGGTCGTAAGTGCCAAAC	57.9	GACATTGCTCGGAACAGGG	57.9	56.55	418
gi 89199678 gb dq398732.1	(A)15	GGAGGTCGTAAGTGCCAAAC	57.9	GAGCCTTCCGAAACGGTGG	66.7	58	462
gi 89199624 gb dq398678.1	(T)10	GTGTAGTCGGTCCATCG	55.6	GTTGGTCGGAGGAGTCCG	66.7	58.02	462
gi 89199623 gb dq398677.1	(T)10	GTGGTAGTGGTATCATCG	50	GGATAAGTCGGAAGAGGC	55.6	58.02	462

Table 2 (contd....)

Accession number	Motif	Forward Primer		Reverse primer		Annealing temperature (C)	Product size (bp)
		Primer	GC %	Primer	GC %		
gi 89199624 gb dq398678.1	(A)21	CGTAGAGTGCCATCCG	61.1	GATGACTACGAGGATGGCG	57.9	58.05	462
gi 89199623 gb dq398677.1	(T)10	CGTAGTCCGGTCCATCG	61.1	GTGATAGACAGAGGAGCGG	55	58.05	462
gi 89199612 gb dq398666.1	(T)10	GGATGACGACAGGAGAGCC	63.2	GATAGAGTCAGAGGAGGGC	57.9	58.05	461
gi 45738090 gb ay563392.1	(TA)8	GGTCCGATAGACTCAACGG	57.9	GTATCGCTATCGCACAGTC	52.6	58.15	676
gi 45738088 gb av563390.1	(TA)7(AT)7	GAGTCACAGTCGTCAGCG	61.1	GCGTGAGTATCGTAGCAGTC	55	58.15	639
gi 9994240 gb lat102469.2	(A)14	GCTGCTATCGAGACAC	61.1	GCACGGTATTTTCAGTCTCCGC	56.5	58.1	648
gi 9994232 gb lat102419.2	(A)10	CATGTCCTACGAGTGGC	57.9	GCACGAGTCAGTCTCCG	66.7	57.5	786
gi 9994224 gb lat102376.2	(A)10	CACGCCAATCTGACGCAC	63.2	GCACGGTTTTCAGTCTCCG	65	58.85	779
gi 224985956 gb fj014139.1	(A)10(A)14	CTGGATGGAACCCCTGAGTG	57.9	GCTTGACGCAGAACGGTG	61.1	55.45	356
gi 46403206 gb av596976.1	(GAG)5	GGAGACGATTGGAGTTGGTG	55	CTCGCTTAGATACTCGCC	55.6	56	669
gi 7578882 gb lat240764.1	(A)20	GGTGAGGGCATAGAGGC	66.7	GGTGGAGGGCATAGAGGC	57.9	55.85	691
gi 6685069 gb lat205859.1	(A)19	GAAGCCACAGGAAGCAG	61.1	GTCAAATCACTCCGCCAG	52.6	55.85	555
gi 6110321 gb ah008318.1	(A)12	GTATCGCGTATGTGGC	61.1	GCCAAACCCATTGTAAGTCC	55	57.8	550
gi 260079916 gb gq245007.1	(A)10	GGACACACAGCGGACAGC	66.7	GCGACGGTATTCACCTCCAC	55	57.95	320
gi 1644387 gb u72654.1	(A)23	GGCTGTTGGTAGATGGCTG	57.9	GGCGAATCTCTATGAACGG	57.9	57.8	690
gi 209483591 gb fj232569.1	(GC)7	CACCTGCGATAGCGGACGAC	63.2	GCAGCATCTCTCGGTGGAC	61.9	61.35	602
gi 94317216 gb dq449916.1	(C)10	CGAATCCAGGGCAAGCAGAG	61.9	CGCTTACAGTTCGGAGTCTC	61.9	62.75	608
gi 57634567 gb ay858677.1	(A)10	GGGAAACCCAGCGAGCGATG	63.2	GTGGCTTCGGGCAACTG	63.2	60.1	307
gi 259435649 emb hb880950.1	(A)18	GAGAAACCCATAGGAGGTC	55.6	CGCAATACCTCTCGTAGCTCG	55	63.1	822
gi 205289952 gb fj010824.1	(TA)7	GCGCAGAACTCAACGAC	61.1	GCGAGGCTATCCACCAGTC	65	58.15	671
gi 205289938 gb fj010810.1	(TA)6	GTGGCAGGACTCAACGGC	66.7	GCACGAGATCACGCCAGTC	65	58.15	669
gi 257693471 emb hb769727.1	(GAG)5	CTCGCTGTTGGCGTGAATC	60	CAGCCATAACCTCACGGATAG	52.4	55.6	549
gi 241661579 dbj ab453155.1	(T)13(T)10	GGGAGTATCTTATCGGAGCG	52.4	GCTGCTATTGATGCCCCGTC	55	54.95	312
gi 241661523 dbj ab453127.1	(T)10(T)10	GGGTTACTTATCGGGAATCG	50	CGATAGGCATTTTGGAGCGGC	57.1	58.55	331
gi 241661499 dbj ab453115.1	(T)11	GGGAGTTTATCGGGAATCG	52.6	CGTTAGGCGTTTGGGCTG	57.9	58.3	331
gi 89511875 dbj ab222605.1	(T)10	CACGAGACTTGGTTACGC	57.9	GAATCCCCCAACCCGAGG	63.2	56.25	337
gi 89511875 dbj ab222605.1	(T)11	GAATAAGAGGACGCCACG	55.6	CGCAGAGCAAGCCCAATG	66.1	56.4	714
gi 62183686 gb ay879942.1	(A)13	GAATAAGAGGACGCCACG	55.6	GCTTGACGGCAGAACGGC	66.7	56.25	299
gi 166407456 gb eu326062.1	(T)12(GT)7	GAGCAGCAGACGAGTAGC	61.1	GACGACGCACATCTCCAC	61.1	57.25	320
gi 219929423 emb fb699668.1	(GAG)5	GGAGACGATGGAGTTGGTG	57.9	CATAGGTGACATACGCCG	55.6	56	669
gi 218478034 dbj ab459662.1	(T)12(GT)7	GGCAGTATGGCTGCTGCTGC	65	CATAGGTGACATACGCCG	55.6	57.5	703
gi 218478034 dbj ab459662.1	(T)17	GCACCCGAAAGCCAGCACCT	65	GAACATACTTGCCACCCG	57.9	58.6	513
gi 164454772 dbj ab289445.1	(ATG)5	CTTCTTCTACTCCGCAGC	55.6	GCAGAAGATGACTCCTCCAG	55	54.9	559
gi 164454770 dbj ab289444.1	(CTT)7	CCAGAAAGTGAGGAAAGCG	55.6	GCAGTGACCCAGAGACCCC	66.7	56.95	549
gi 193795409 gb ef203258.1	(TA)7 (T)15 (A)29	CGTGAGGATTGGCTGTCGGC	65	CTATGCGACCAGCGATTCCAC	55	58.45	703
gi 193795407 gb ef203257.1	(G)14 (A)86	CAGACAAAGGAAACCCACCG	57.9	GGATGACGGCACCACCAAC	63.2	55.3	513
gi 113735151 dbj ab271691.1	(AT)6	GTAGCAGCAGTGTGGTCCGC	65	GATTTACAGCAACCCAGGTC	50	55	677
gi 170673145 gb eu541812.1	(G)10	CGACCTTTGTAGCAGCCG	61.1	CGCTTTGTTGTGCTCCG	55	55.15	499
gi 170673141 gb eu541808.1	(A)10	CACGACTCTCCAGCAGCG	68.4	GTCTCTGCTGTGCGTATCG	57.1	55.4	614
gi 170673139 gb eu541806.1	(T)10	CACGAATCATCTCAGTCTCTC	52.4	GTCTTTGTTGCTCTTCG	55.6	55.4	611
gi 170672874 gb eu528047.1	(TAA)6 (T)10	GGGTAATCTGAGCCAAATCC	50	GCAGGCTATCCCGACCAC	68.4	57.9	735
gi 4092183 gb lat102463.1	(T)10 (A)10	CGGGTCGCAATCTGAGCC	68.4	GCTTGACAGGCGAAGCCGGG	63.2	56.7	290

Table 2 (contd....)

Accession number	Motif	Forward Primer		Reverse primer		Annealing temperature (C)		Product size (bp)
		Primer	GC %	Primer	GC %	temperature (C)	GC %	
gi 4092141 gb af102421.1	(A)11	GGGTCCGCATCCTGAGCC	72.2	GCTTGACGGCAGAACCGG	66.7	56.7	290	
gi 4092124 gb af102404.1	(A)11	GGGGTCCGGATGTGAGCC	72.2	GCTTGACAGTGCAGAACCGG	63.2	56.25	290	
gi 133874211 dbj ab190181.1	(A)24	CGGTCCCTTGGATGCCCTGGG	71.4	CGCTCCTTCTCCACTGCC	65	58.3	592	
gi 124388815 gb ef069436.1	(A)21	GGCTCAATCGCTGTAAAC	57.9	GCCAGTCCAGTGAAGTCCG	63.2	57.7	137	
gi83758482 dbj AB19627.1	(TA)6	GGTTACGGTGAAGAGTGACAGG	54.5	GGATGGGAAGTGAAGACAGG	57.9	58.25	374	
gi 83758480 dbj AB219625.1	(TA)7	GGTTACGGGAAGAGTGACAGG	55	CCATACCAAGGCTCAATCC	52.6	56.2	228	
gi 156787487 gb ef569229.2	(AT)6 (A)20	CTTCTCCACGGTCGCCCTTAC	60	GTGACTGAAGCATCTTACC	52.6	60.15	539	
gi 156787485 gb ef569228.2	(A)18	GGATAGAGGCTGTGGATGC	60	CACCAGTCTCAACACCTC	55.6	56.75	313	
gi 146272406 dbj ab281494.1	(CAA)7	CGAGGATCAAGTTCACGGC	55	GAGTTCAGGGACCCGATAGC	60	58.15	371	
gi 147743640 gb ef571643.1	(CT)10	CCGTAGTGTGGTCAGAAAGCAGG	56.5	GCACCTGTAGAACGGATGATG	54.5	57.85	350	
gi 147743624 gb ef571635.1	(TC)6	GCAAGGGGAGCACCCCAAG	66.7	GTTAGCCAGGATGCGAAGC	57.9	57.75	712	
gi 147743612 gb ef571629.1	(TC)7	GCATTCCGGTCAGCCAAGC	61.1	GGTGGTATTCATCTTCCGCCG	57.1	58.55	825	
gi 147743609 gb ef571627.1	(CT)8	GCAACACAGCGGGACTAAC	57.9	GCACGACAGAACGAGCGG	66.7	60.25	617	
gi 147743597 gb ef571621.1	(CT)11	CCTCGGACCACCAATCAGC	63.2	GTGTGAACGACTCCGCTTG	57.9	56.35	486	
gi 147743595 gb ef571620.1	(CT)6	CAGTTTCGGTCAGCAAGC	55.6	GTGAGAACAACCCCGCTG	63.2	57.35	440	
gi 147743553 gb ef571599.1	(TC)7	GCAAAAGAGGAGCACCAAG	55.6	GCAGGTATTCATCTCCGCCG	60	59.05	705	
gi 44829182 gb av466118.1	(A)28	GCCTACCCAAAACGCTGACC	60	CACCTGTCTTCTGTAGCGG	55	59.2	645	
gi 124295133 gb ef203260.1	(A)18	GTTGTGGAGGTGGCTTCG	61.1	CTGCGTAACACTCATCAAGCC	52.4	56.02	603	
gi 118145133 dbj dd357421.1	(A)18	CCAAATGACGGACGCTACCC	60	CTATGCCACCAATGTCC	55.6	57.35	791	
gi 117935906 gb ef062505.1	(A)18	GGAGCCTATCGGAACGGG	66.7	GCGGTTTCACTTACACGAGC	52.4	58.85	700	
gi 95118583 gb dq497593.1	(A)11	CGTGATCAAGGACCCGCC	61.1	GGAGCCTATGCGAACCGG	66.7	56.35	345	
gi 95118580 gb dq497590.1	(A)11	CTGTGTTCAAAGACCCCGTGC	57.1	GAGCCTATGCGAACCGG	61.1	56.35	345	
gi 13018492 dbj e31220.1	(A)18	CTCTGGTCTTGTGATGCTG	50	GTGATTTGTAGCGGACG	52.6	54.8	650	
gi 92242664 dbj ba289870.1	(G)14(A)86	GGCTGGTACTACTGCTG	61.1	CCGCATCCACGACGGAAGC	68.4	58.95	561	
gi 128002 dbj ba289870.1	(A)15	CCTTGGTTGTGCTGCTCTGGT	59.1	GAAAGCAACAGAGGTGACCG	60	58.6	447	
gi 90959871 dbj ab027191.1	(T)14	GAGGATACGACACGCGACG	63.2	CAGAGCGTAGAACCCATCAG	55	55.7	704	
gi 90959867 dbj ab027189.1	(A)23	GATACCCACAGGAGAAC	57.9	GACAACTCCACGACGAGAC	57.9	55.45	680	
gi 90959865 dbj ab027188.1	(A)27	GCGAACTGTACTCAAACCCACC	54.5	CATTGAGCGACATCCCTGC	57.9	55.35	693	
gi 90959863 dbj ab027187.1	(A)18	CGACATCAACCCCTACGCTG	60	GCCATAAGGGACGCCATTG	57.9	57.1	730	
gi 6681691 dbj ab017370.1	(TA)7(T)15	CCTGTAACCTCCACCTCAACC	57.1	GCTGTCAATGTAGCTGCTTCC	52.2	57.9	679	
gi 6681687 dbj ab017368.1	(TA)8	GGTAGTAAGTCCAGGTGCTCG	57.1	CGCTTCTTCACTCCCTTTGC	55	56.95	590	
gi 89276224 gb dq402068.1	(T)14	CGCACCGAATAAAGCCACAG	55	GGTGAGAAGGAGACGGTGG	63.2	58.15	320	
gi 86451166 gb dq358898.1	(T)10	GCCCTCAAATCTTCTCTCTCG	54.5	CGGGTCTTGTGGCTGGCTATG	56.5	60.1	658	
gi 76443559 dbj ab011014.1	(A)18	GTTGTGGAGGTGGCTCTCG	63.2	GGTGAATGTTGAGAGACG	55	56.6	446	
gi 75812155 dbj d38043.1	(T)12	GCGAGGATAAAGACTGGCTG	60	CAACCCGCAACAGGACACGC	63.2	57.2	681	
gi 74267409 dbj ab208689.1	(A)10	GAATGGCAAGTGGTGGCTGG	60	GTAATGGCTGGCAAGGTTTC	55	54.8	460	
gi 18146806 dbj ab028666.1	(G)13	GCAAGCAAGGGAATCAGG	55.6	CCGCACTCCAGCACGGAAGC	68.4	58.6	629	
gi 6687482 emb aj236195.1	(A)10	GTGACTTCGTGAGACAGC	57.9	GTCGCAATGTCGTCGACGG	63.2	56	453	
gi 62433122 dbj d38168.1	(CT)8	CAGTCTCCTCCGTACCCGAAG	61.9	GAATGGAATCACGGGAG	52.9	56.75	450	
gi 62086548 dbj ab193314.1	(A)28	GAATCATCGCCATCTCCACC	55	GTCCATAACCCCTTTCAGCC	52.6	55.5	362	
gi 62086546 dbj ab193313.1	(A)12(A)12	GGTGTAGCGGGCAATCC	61.1	CAGACGCTCCACCTTCC	66.7	57.25	524	
gi 62086540 dbj ab193310.1	(A)19	GAATGGCAAGTGGTGGCTGG	60	GTAATGGCTGGCAAGGTTTC	55	54.8	620	
gi 38568130 emb aj489906.1	(C)10	GATGATTGCCGCCCTGCTGG	63.2	GGACTCGCATTTAGCCAGC	57.9	59.8	349	

Table 2 (contd....)

Accession number	Motif	Forward Primer		Reverse primer		Annealing temperature (C)		Product size (bp)
		Primer	GC %	Primer	GC %	temperature (C)	GC %	
gi 4455871 emb aj010509.1	(T)12	TGGTAAAAAGATGCTCCGTC	50	CCATAGGTAAGTCCGAAAAGTCC	50	55.15	317	
gi 37954874 gb ev251732.1	(A)19	GGTGCCAAATCCTGAGCCG	66.7	GCTTGACGGCAGAACCGGG	66.7	56.4	307	
gi 37954873 gb ev251731.1	(A)16	GGATTGAGCCTGGTATGG	55.6	GCCCTGACAGCAGAACCGGG	63.2	56.5	375	
gi 37991673 dbj jab124878.1	(A)16	GCAGTAAGGGAGCGTGAC	61.1	GTCGTCAAGTCAGCCAAC	55.6	54.6	460	
gi 22773822 dbj jab080739.1	(A)13	CATAGGAAGGCAAGAAAGC	52.6	GATTCAGGTAGCAACGGAGTGG	54.5	58.9	528	
gi 33945370 emb aj490190.1	(A)10	GGTGGCAGGACTCAACGG	66.7	GCGATTGACAGCAGAACCGGG	60	56.95	204	
gi 15149940 emb aj315324.1	(A)11	GTCGCTCCGCCGATTGGG	63.2	CTACACAGGTTCCGGAGGG	61.1	62.4	75	
gi 27530874 dbj jab076697.1	(A)18	GATTACAGCAGCCGAGGAG	61.1	GTTATGCCTCCGCTCAGTG	57.9	55.35	458	
gi 22796301 emb aj430909.1	(AT)7	GATGGCAGTCTCCGCTTC	63.2	CCGTTGATGTCTCTGTACC	55	55.95	597	
gi 18146804 dbj jab028665.1	(G)14(A)86	CAGTTCATCAAGCACACAGGGG	59.1	CTTCCGTCCGATGTACGG	61.1	56.7	622	
gi 1785485 dbj d14589.1	(A)15	GGCTGTTGAAATGGCGG	61.1	GAGCGTTTATCTGGGGC	55.6	56	672	
gi 3808127 emb aj011983.1	(A)12	CCGTAGGGTGGGCTTTC	61.1	GCCAACCAATTCGTAAGTCC	52.6	56.05	527	
gi 10189942 emb jax028840.1	(A)17	GGTCATTACTCGGGGTGTG	57.9	GCACTCTACATCACATTGGGC	54.5	56.6	797	
gi 7415596 dbj jab026494.1	(TA)6(A)21	CTTGAGAAATGCCGTGTTGC	50	CTACAAATGCCGCTCCAC	55	54.8	410	
gi 221150824 gb gh694656.1	(C)12	GGAGGAGTGAATCGGAACCC	60	GTGAGGTGGAGGGGACTGG	66.7	58.55	181	
gi 62086548 dbj jab193314.1	(A)28	GAATCATCGCCATCTCCACC	55	GTCCATAACCCCTTTCAGCC	52.6	55.5	362	
gi 62086546 dbj jab193313.1	(A)12(A)12	GGGTAGCGGGCATCCG	61.1	CAGACGCTCCACCTTCCC	66.7	57.25	424	
gi 62086540 dbj jab193310.1	(A)19	GAATGGCAAGTGGTCGCTGG	60	GATGGCTCGGCAAGGTTTC	55	54.8	620	
gi 38568130 emb aj489906.1	(C)10	GATGATTGCCGCTGCTGG	63.2	GGACTCGCATTTAGCCAGC	57.9	59.8	349	
gi 33945346 emb aj489882.1	(C)10	CACAAACGGCAGCAGAAAG	57.9	CAGCACAGGATTGAGGC	61.1	61.25	445	
gi 4455871 emb aj010509.1	(T)12	TGGTAAAAAGATGCTCCGTC	50	CCATAGGTAAGTCCGAAAAGTCC	50	55.15	317	
gi 37954874 gb ev251732.1	(A)19	GGTGCCAAATCCTGAGCCG	66.7	GCTTGACGGCAGAACCGGG	66.7	56.4	307	
gi 37954873 gb ev251731.1	(A)16	GGATTGAGCCTGGTATGG	55.6	GCCCTGACAGCAGAACCGGG	63.2	56.5	375	
gi 37991673 dbj jab124878.1	(A)16	GCAGTAAGGGAGCGTGAC	61.1	GTCGTCAAGTCAGCCAAC	55.6	54.6	460	
gi 22773822 dbj jab080739.1	(A)13	CATAGGAAGGCAAGAAAGC	52.6	GATTCAGGTAGCAACGGAGTGG	54.5	58.9	728	
gi 33945370 emb aj490190.1	(A)10	GGTGGCAGGACTCAACGG	66.7	GCGATTGACAGCAGAACCGGG	60	56.95	204	
gi 15149940 emb aj315324.1	(A)11	GTCGCTCCGCCGATTGGG	63.2	CTACACAGGTTCCGGAGGG	61.1	62.4	75	
gi 27530874 dbj jab076697.1	(A)18	GATTACAGCAGCCGAGGAG	61.1	GTTATGCCTCCGCTCAGTG	57.9	55.35	458	
gi 10189940 emb jax028838.1	(A)17	CGGGCTAAGAGACACCGGC	66.7	GGTGTGGCTCGTTGAATGCG	60	59.3	285	
gi 22796301 emb aj430909.1	(AT)7	GATGGCAGTCTCCGCTTC	63.2	CCGTTGATGTCTCTGTACC	55	55.95	597	
gi 18146804 dbj jab028665.1	(G)14(A)86	CAGTTCATCAAGCACACAGGGG	59.1	CTTCCGTCCGATGTACGG	61.1	56.7	522	
gi 1785485 dbj d14589.1	(A)15	GGCTGTTGAAATGGCGG	61.1	GAGCGTTTATCTGGGGC	55.6	56	672	
gi 3808127 emb aj011983.1	(A)12	CCGTAGGGTGGGCTTTC	61.1	GCCAACCAATTCGTAAGTCC	52.6	56.05	427	
gi 10189942 emb jax028840.1	(A)17	GGTCATTACTCGGGGTGTG	57.9	GCACTCTACATCACATTGGGC	54.5	56.6	797	
gi 7415596 dbj jab026494.1	(TA)6(A)21	CTTGAGAAATGCCGTGTTGC	50	CTACAAATGCCGCTCCAC	55	54.8	410	
gi 221150824 gb gh694656.1	(C)12	GGAGGAGTGAATCGGAACCC	60	GTGAGGTGGAGGGGACTGG	66.7	58.55	181	

clustering patterns, test for Hardy-Weinberg equilibrium and multilocus equilibrium, and to test polymorphic loci for the evidence of selective neutrality. This can be useful to plant breeders, germplasm managers, or others who are interested in population genetic properties of materials that they are working with. The three most common types of markers used today are RFLP, RAPD, and microsatellites. A wide variety of methods for the construction of libraries enriched for microsatellite sequences have been reported, the most popular among those being the ones based on vectorette PCR using anchored primers. But this method is highly time-consuming and expensive, and the alternative is to use bioinformatics, that is, computational tools to screen the public database and find SSR. EST-derived molecular markers, especially SSR and SNP, are highly useful in developing linkage maps and markers assisted breeding programs. These markers are also transferable to related genera.

Molecular marker techniques are advantageous as they directly reflect variations in the DNA sequences and therefore of independence of environment. Among many molecular marker techniques currently available, microsatellites and SSRs^[35] provide an improved technology in assessing genetic diversity and genetic relationships in plants as they are highly polymorphic, codominant, very informative, and PCR based. EST-SSRs offer the following advantages over other genome DNA-based markers: (1) they should detect variation in the expressed portion of the genome so that gene tagging should give “perfect” marker-trait associations; (2) they can be developed at no cost from the EST databases; and (3) once developed, these markers, unlike genomic SSRs, may be used across a number of related species. With the growth of sequence databases, several authors have reported an abundance of SSRs in different genomes. The Distribution of SSRs in the rice genome has also been studied on the basis of the two whole genome draft sequences released, respectively, by Syngenta and by the Beijing Genome Institute (BGI). In the draft sequence released by Syngenta, for instance, 48,351 SSRs (including di-, tri-, and tetranucleotide repeats) were available, giving a density of 8 kb per SSR in the whole genome; SSRs represented by di-, tri-, and tetranucleotide repeats accounted respectively for 24%, 59%, and 17% of the total SSRs.

SSRs are very polymorphic due to the high mutation rate affecting the number of repeat units. Such length-polymorphisms can be easily detected on high-resolution gels (e.g., sequencing gels), by running PCR-amplified fragments obtained using a unique pair of primers flanking the repeat.^[36] Chung and Staub^[37] developed a set of consensus chloroplast primer pairs for ccSSRs from *N. tabacum* chloroplast sequences. All primer pairs produced

amplicons after PCR employing chloroplast DNA from members of the *Cucurbitaceae* (six species) and *Solanaceae* (four species). Sixteen, 22, and 19 of the initial 23 primer pairs were successively amplified by PCR using template DNA from species of the *Apiaceae* (two species), *Brassicaceae* (one species), and *Fabaceae* (two species), respectively. Twenty of the 23 primer pairs were also functional in three monocot species of the *Liliaceae* (onion and garlic), and the *Poaceae* (oat). ccSSR primers were strategically “recombined” and referred to correctly as recombined consensus chloroplast primers (RCCP) for PCR analysis of cucumber DNA such that the primers designed for the SSR-containing genus of Gentiana family would be utilized for the production of amplicons from different members of family.

Kijas *et al.*^[38] tested two primer sets in 10 different *Citrus* species and two related genera and found conservation of the sequences. Cross-species amplification has also been reported between cultivated rice and related wild species^[39] and between *Vitis* species.^[40] Provan *et al.*^[41] could show successful amplification of two tomato SSR primer pairs tested on potato cultivars. Weising *et al.*^[42] reported conservation of SSR flanking sites in different species of kiwifruit (*Actinidia chinensis*). Usually, a low percentage of markers also amplified fragments from species belonging to other genera from the same family. Within the *Poaceae* family, primers worked even across different genera,^[43] but only 50% of microsatellite loci identified in wheat were also polymorphic in rye and barley cultivars. Whitton *et al.*^[44] tested 13 SSR loci in 25 representatives of the *Asteraceae*, where it was demonstrated that the regions flanking in the repeats are not highly conserved, neither in the nucleotide sequence nor in the relative position.

Indeed, in general, transferability of polymorphic markers in plants is likely to be successful mainly within genera (success rate close to 60% in eudicots and close to 40% in the reviewed monocots) rather than between genera (transfer rates are approximately 10% for eudicots) within the same family.^[45] This transferability of polymorphic markers nature in plant generally enhances the utilization of the primers in random way. Comparative genome analysis facilitates high-throughput comparative mapping with the assistance of cross-species markers, and further facilitates gene cloning by identifying cross-reference genes. Seventeen SSR primer sets developed for *Quercus petraea* were tested on eight different members of the *Fagaceae* family.^[46] In total 66% resulted in interpretable amplification products and most of them were really homologous to the originally cloned SSR fragment from *Q. petraea*. The primers could be designed successfully for a very large number (169, 31%) of SSRs [Table 2]. However, it was not possible to design the primers for remaining SSRs (376, 69%) because

the sequence flanking at both ends of the SSRs was inadequate in size to design the primers. The large number of primer pairs for the SSRs that have been designed during the present study may be utilized for a variety of purposes, for example, gene tagging, genetic mapping, population studies, etc. Due to a high level of potential for length polymorphisms, SSRs have become a valuable source of genetic markers and have been broadly applied to various areas of genetic research including studies of genome variation, establishment of genetic maps, integration of physical and genetic maps, determination of evolutionary relationships, and comparative genome analyses.

CONCLUSIONS

Nucleotide sequences of *Gentiana* family were systematically searched for SSRs using the “*ssr_finder.pl*” perl program for the development of SSR markers. This is a valuable approach for both costs and time, given a sufficient amount of available *Gentiana* family sequences. The use of SSRs in genetic diversity studies is a novel tool that reveals variation in genomes.

REFERENCES

1. Struwe L, Kadereit JW, Klackenberg J, Nilsson S, Thiv M, von Hagen KB, et al. Systematics, character evolution, and biogeography of *Gentianaceae*, including a new tribal and subtribal classification. In: Struwe, L. and V. A. Albert, editors. Cambridge: *Gentianaceae—Systematics and Natural History* Cambridge University Press; 2002. p. 21-309.
2. Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao YP, et al. Genetic mapping of a gene causing hypertensive rat. *Cell* 1991;67:213-24.
3. Litt M, Luty JA. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 1989;44:397-401.
4. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 1991;49:746-56.
5. Tautz D, Renz M. Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res* 1984;12:4127-38.
6. Katti MV, Ranjekar PK, Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 2001;18:1161-7.
7. Hancock JM. The contribution of slippage-like processes to genome evolution. *J Mol Evol* 1995;41:1038-47.
8. Ahmad R, Struss D, Southwick SM. Development and characterization of microsatellite markers in citrus. *J Am Soc Hortic Sci* 2003;128:584-90.
9. Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr. Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet* 2006;112:1248-57.
10. Yang X, Scheffler BE, Weston LA. Recent developments in primer design for DNA polymorphism and mRNA profiling in higher plants. *Plant Met* 2006;2:4.
11. Witzany G. Noncoding RNAs: Persistent viral agents as modular tools for cellular needs. *Ann NY Acad Sci* 2009;1178:244-67.
12. Robinson AJ, Love CG, Batley J, Barker G, Edwards D. Simple sequence repeat marker loci discovery using SSR primer. *Bioinformatics* 2004;20:1475-6.
13. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 2003;106:411-22.
14. Xu Y, Ma RC, Xie H, Liu JT, Cao MQ. Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. *Genome* 2004;47:1091-104.
15. Yu JK, La Rota M, Kantety RV, Sorrells ME. EST derived SSR markers for comparative mapping in wheat and rice. *Mol Genet Genomics* 2004;271:742-51.
16. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol* 2005;23:48-55.
17. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 2001;11:1441-52.
18. Kantety RV, La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 2002;48:501-10.
19. Pinto LR, Oliveira KM, Ulian EC, Garcia AA, de Souza AP. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome* 2004;47:795-804.
20. Peng JH, Lapitan NL. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Funct Integr Genomics* 2005;5:80-96.
21. Serapion J, Kucuktas H, Feng J, Liu Z. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar Biotechnol (NY)* 2004;6:364-77.
22. Prasad MD, Muthulakshmi M, Madhu M, Archak S, Mita K, Nagaraju J. Survey and analysis of microsatellites in the silkworm, *bombyx mori*: Frequency, distribution, mutations, marker potential and their conservation in heterologous species. *Genetics* 2005;169:197-214.
23. Dranchak PK, Chaves LD, Rowe JA, Reed KM. Turkey microsatellite loci from an embryonic cDNA library. *Poult Sci* 2003;82:526-31.
24. Haddad LA, Parra FC, Pena SD. Characterization and mapping of four novel human expressed polymorphic trinucleotide microsatellites. *Gene* 1998;223:369-74.
25. Fraser LG, Harvey CF, Crowhurst RN, De Silva HN. EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor Appl Genet* 2003;108:1010-6.
26. Riju A, Arunachalam V. Data mining for simple sequence repeats in oil palm expressed sequence tags. *Nature Proceedings* 2009.
27. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, et al. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* 1990;18:999-1005.
28. Khabar KS, Dhalla M, Bakheet T, Sy C, al-Haj L. An integrated computational and laboratory approach for selective amplification of mRNAs containing the adenylate uridylylate-rich element consensus sequence. *Genome Res* 2002;12:985-95.
29. Sharrocks AD. The design of primers for PCR. In: Griffin HG, Griffin AM, editors. *PCR technology: Current innovations*. London: CRC Press; 1994. p. 5-11.

30. Ramsay L, Macaulay M, degli Ivannissevich S, MacLean K, Cardle L, Fuller J, et al. Simple sequence repeat-based linkage map of barley. *Genetics* 2000;156:1997-2005.
31. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. EST-SSRs for transferability, polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics* 2003;270:315-23.
32. Toth G, Gaspari Z, Jurka J. Microsatellites in different eukaryotic genome, survey and analysis. *Genome Res* 2000;10:1967-81.
33. Cardle L, Ramsay L, Milborne D, Macaulay M, Marshall D, Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 2000;156:847-54.
34. Metzgar D, Bytof J, Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 2000;10:72-80.
35. Powell W, Machray GC, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1996;1:215-22.
36. Weber, May. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989;44:388-96.
37. Chung SM, Staub JE. The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theor Appl Genet* 2003;107:757-67.
38. Kijas JM, Fowler JC, Thomas MR, Scott NS. An evaluation of sequence tagged microsatellite site markers for genetic analysis within citrus and related species. *Genome* 1995;38:349-55.
39. Wu K, Tanksley SD. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mole Gen Genet* 1993;241:225-35.
40. Thomas MR, Scott NS. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites. *Theor Appl Genet* 1993;86:985-90.
41. Provan J, Waugh R, Powell W. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). *Theoretical and Applied Genetics* 1996;92:1076-84.
42. Weising K, Fung RW, Keeling DJ, Atkinson RG, Gardner RC. Characterisation of microsatellites from *Actinidia chinensis*. *Mole Breed* 1997;3:159-60.
43. Röder MS, Plaschke JS, König U, Börner A, Sorrells ME. Abundance, variability and chromosomal location of microsatellites in wheat. *Mole Gen Genet* 1995;246:327-33.
44. Whitton J, Rieseberg LH, Ungerer MC. Microsatellite loci are not conserved across Asteraceae. *Mole Biol Evol* 1997;14:204-9.
45. Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Mole Ecol* 2007;16:3759-67.
46. Steinkellner H, Lexer C, Turetschek E, Glössl J. Conservation of (GA)_n microsatellite loci between *Quercus* species. *Mole Ecol* 1997;6:1189-94.

Cite this article as: Sathishkumar R, Lakshmi PT, Annamalai A, Arunachalam V. Mining of simple sequence repeats in the Genome of *Gentianaceae*. *Phcog Res* 2011;3:19-29.

Source of Support: Nil, **Conflict of Interest:** None declared.

Author Help: Online Submission of the Manuscripts

Articles can be submitted online from <http://www.journalonweb.com>. For online submission articles should be prepared in two files (first page file and article file). Images should be submitted separately.

1) **First Page File:**

Prepare the title page, covering letter, acknowledgement, etc., using a word processor program. All information which can reveal your identity should be here. Use text/rtf/doc/pdf files. Do not zip the files.

2) **Article file:**

The main text of the article, beginning from Abstract till References (including tables) should be in this file. Do not include any information (such as acknowledgement, your names in page headers, etc.) in this file. Use text/rtf/doc/pdf files. Do not zip the files. Limit the file size to 1 MB. Do not incorporate images in the file. If file size is large, graphs can be submitted as images separately without incorporating them in the article file to reduce the size of the file.

3) **Images:**

Submit good quality color images. Each image should be less than **4096 kb (4 MB)** in size. Size of the image can be reduced by decreasing the actual height and width of the images (keep up to about 6 inches and up to about 1200 pixels) or by reducing the quality of image. JPEG is the most suitable file format. The image quality should be good enough to judge the scientific value of the image. Always retain a good quality, high resolution image for print purpose. This high resolution image should be sent to the editorial office at the time of sending a revised article.

4) **Legends:**

Legends for the figures/images should be included at the end of the article file.