Mining of microsatellites using next generation sequencing of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome

Ankit Jain · Saurabh Chaudhary ·
Prakash Chand Sharma

Received: 23 June 2013 / Revised: 31 August 2013 / Accepted: 9 October 2013 / Published online: 8 December 2013
© Prof. H.S. Srivastava Foundation for Science and Society 2013

Abstract Gene based microsatellite markers are becoming more popular as compared to traditional random genomic microsatellite markers due to rapid and inexpensive method of isolation and their cross species portability. The present study documents occurrence of microsatellites in the transcriptome of seabuckthorn, a plant with immense medicinal, nutritional and ecological value. De novo assembly of over 80 million high quality short reads generated by high throughput next generation sequencing yielded 88297 putative unigenes. Of these, 7.69 % unigenes harbored microsatellite repeats with an average of one microsatellite per 6.704 Kb transcriptome. Dinucleotide repeats were most abundant followed by trinucleotide repeats. Microsatellites were densely populated in coding regions followed by 3' and 5' untranslated regions. AG and AAG type repeats were most frequently represented. Of the microsatellite positive unigenes, 48.81 % could be assigned gene ontology (GO) terms in order to assess associations between microsatellite containing unigenes and biological role of known genes. Utility of unigene specific microsatellites was assessed on the basis of polymorphism(s) detected in 18 seabuckthorn collections from Leh (India) using a set of randomly selected 25 unigene specific microsatellites. The findings presented here are likely to find immense use in future breeding and molecular biology research projects in seabuckthorn aiming at its overall development as a crop.

Keywords Seabuckthorn · Transcriptome sequencing · Microsatellites markers · Gene ontology · Codon usage

Introduction

Microsatellites (Simple Sequence Repeats) represent tandem repetition of 1–6 nucleotide long motifs occurring throughout the prokaryotic as well as eukaryotic genomes (Tóth et al. 2000). Their abundance and hypervariability, co-dominant inheritance, reproducibility, multi-allelic nature and wide genomic distribution make them markers of choice for a variety of applications ranging from linkage mapping to diversity analysis (Varshney et al. 2005).

Often tedious and costly cloning and enrichment procedures are involved in the development of microsatellite markers (Zane et al. 2002). Recent revolutionary advancements in DNA sequencing technology have led to the accumulation of vast amount of sequence data in public domain. As the content available in public databases is increasing rapidly, there has been a great inclination towards development of molecular markers using computational tools (Sharma et al. 2007; Grover et al. 2012). This approach not only reduces the cost of marker development but also permits generation of markers at a large scale in a short span of time. However, a great variation has been observed in the output depending on the tools used and the criteria adopted for mining microsatellites (Grover et al. 2012).

Gene based microsatellite markers have provided a most valuable and popular tool for comparative genomics as the markers developed from one species could also be exploited in related species hitherto unexplored (Ellis and Burke 2007; Jain et al. 2010; Ma et al. 2012; Singh et al. 2012). Limited marker resources are available for seabuckthorn (*Hippophae* sp.) family Elaeagnaceae. It is a hardy, deciduous shrub, with significant medicinal and ecological importance. All the

Electronic supplementary material The online version of this article (doi:10.1007/s12298-013-0210-6) contains supplementary material, which is available to authorized users.

A. Jain · S. Chaudhary · P. C. Sharma (✉)
University School of Biotechnology, Guru Gobind Singh
Indraprastha University, Sector-16C, Dwarka, Delhi 110078, India
e-mail: prof.pcsharma@gmail.com

species of *Hippophae* are diploid with $2n=24$ (Elena et al. 2011). Seabuckthorn berries are rich in vitamins, fatty acids, free amino acids, flavanoids, tocopherol and other secondary metabolites (Dhyani et al. 2007). Preparations from seabuckthorn have frequently been utilized for the treatment of cardiovascular diseases, cancers, skin problems, burns, digestive tract disorders and immune system disorders. The plant parts contain many anti-inflammation and anti-bacterial substances (Chauhan et al. 2007; Dhyani et al. 2007).

Seabuckthorn has a strong ability to fix atmospheric nitrogen due to the association of *Frankia*, a soil actinomycete, with its root nodules thereby improving soil quality. Moreover, its extensive root system aids to soil conservation (Trajkovski and Jeppsson 1999). Cascades of genes imparting adaptive nature to seabuckthorn to withstand various abiotic stresses and genes related to biosynthesis of various bioactive compounds, have generated substantial interest among researchers all over the world (Yang et al. 2005; Guo et al. 2007; Xu et al. 2009; Ghangal et al. 2012; Fatima et al. 2012; Ghangal et al. 2013).

Attempts have been made to review different marker systems used to access phylogenetic relationship and genetic diversity in seabuckthorn (Ghangal et al. 2008; Kalia et al. 2011; Ruan et al. 2013). Till date, very limited information is available on microsatellite markers in seabuckthorn (Wang et al. 2008; Jain et al. 2010). It has been known for long that microsatellite variation sometimes leads to alteration in gene expression leading to phenotypic variation. Therefore, microsatellite variations could be implicated as a key player in the process of adaptation and evolution (Grover and Sharma 2011). The present study was undertaken to understand distribution of microsatellites in seabuckthorn transcriptome and assess their functional significance in developing Unigene Specific Microsatellite Markers (USMM).

Materials and methods

Sequence resource

Short read sequences were generated using Illumina HiSeq2000 platform under a project on seabuckthorn transcriptome profiling undergoing in our laboratory. The short reads have been submitted to NCBI's Short Read Archive (SRA) with study accession number SRP011938, containing seabuckthorn leaf and root sample data under the accession numbers SRS304528 and SRS304529, respectively. Assembly of short reads involved the use of short read assembler (ABYSS), followed by TGICL, yielded 88297 putative unigenes (Ghangal et al. 2013). Primary data thus generated was used for mining microsatellite repeats.

Microsatellite mining in seabuckthorn transcriptome

Microsatellite searching tool, MISA (<http://pgrc.ipk-gatersleben.de/misa/>; Thiel et al. 2003) was used to identify and classify microsatellites in putative unigenes of seabuckthorn, considering complementarily and occurrence in different frames (e.g. dinucleotide repeat GA class will include: GA/AG/CT/TC repeats). Mononucleotide repeats were not included in the current analysis because of the abundance of poly A/T repeats, arising due to the presence of PolyA/T tails in cDNA molecules derived from eukaryotic transcripts and sequencing errors prevailing in Illumina base sequencing platform.

Functional assessment of microsatellite carrying unigenes

Unigene sequences harboring microsatellite repeats (Supplementary file: Suppl.fa) were used for similarity search against existing non-redundant protein sequences in the public database to identify their putative functions. Similarity search was performed at a cut off value of $\leq 1E-10$ to keep stringency level high. In order to assess their complete annotation, these microsatellite carrying unigenes were annotated based on the Gene Ontology (GO) terms associated with the blast results using Blast2GO software (Conesa et al. 2005) which initially performs BLASTX analysis followed by mapping and annotation step.

Determination of the position of microsatellites within genes

OrfPredictor (Min et al. 2005) was used to predict open reading frames (ORF) within the putative unigenes. Microsatellites lying in the region upstream to ORF were considered as 5' UTR SSRs while others falling in the region downstream to ORF were taken as 3' UTR SSRs. Comparative abundance of different motifs was also determined manually in MS excel. Unigenes with no ORF were not included in further analysis.

Codon usage of tri-nucleotide repeats

Only tri-nucleotide repeats falling in coding regions were taken into consideration and examined for the possible encoded amino acid motif. These sequences were further submitted to CodonO server (Angellotti et al. 2007) to confirm the preferential codon usage within seabuckthorn transcriptome.

Designing and validation of primers

Class I microsatellites (≥ 20 nucleotide long) carrying unigenes were selected at random and analyzed for their 'Var' score by using SERV, a nonlinear model that generates

Fig. 1 Length and abundance relationship of class I and class II microsatellites in seabuckthorn unigenes

a numeric score on the basis of number of repeat units, unit length, and purity that correlates with repeat variability (Legendre et al. 2007). Further, these unigene sequences were used to design primers from microsatellite flanking regions using Primer 3 (Rozen and Skaletsky 2000). Parameters taken into consideration for primer design included 50–60 °C annealing temperature with a maximum difference of 5 °C between any individual pair of forward and reverse primers, 40–60 % GC content, 18–25 bp primer length, and desired amplicon size of 150–400 bp.

Genomic DNA isolated from 18 random collections of *Hippophae rhamnoides* (natural population collected from Leh (India), 34°08'43.43°N77°34'03.41°E) and 5 collections of *Hippophae salicifolia* (collected from the premises of DIHAR, Leh), were used to validate the putative unigene based microsatellite (SSR) markers. A 25 µl PCR reaction contained 5 pmol of each primer, 100 µM of each of dNTPs, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.5 U of Taq polymerase (GoTaq, Promega), and 50 ng of template DNA. Amplifications were carried out in a Mastercycler (Eppendorf, Germany) with thermal profile being 5 min of pre-amplification denaturation at 95 °C followed by 35 cycles of 30 s at 95 °C, 30 s at a primer-specific annealing

temperature, and 1 min at 72 °C. As a final step, products were given a final extension for 5 min at 72 °C and stored at 4 °C until further use. PCR products were resolved using electrophoresis in 4 % metaphor agarose (Lonza, Switzerland) in 1X TAE buffer. The amplicons were visualized after ethidium bromide staining and recorded using gel documentation system (Alpha Imaging).

Results and discussion

Microsatellite dynamics in seabuckthorn transcriptome

Out of total 88297 putative unigenes with an average size of 610 bp, 7.69 % unigenes showed the presence of microsatellite repeats. Of these microsatellite harboring unigenes, 14.58 % contained more than one microsatellite repeat. Further, 713 microsatellite repeats were found in compound formation. In the present study, overall microsatellite density was found to be one microsatellite spanning every 6.704 Kb, which is in correspondence with the earlier studies (Varshney et al. 2005; Poncet et al. 2006; Wohrmann and Weising 2011) where the frequency of microsatellites in model plants was

Fig. 2 Comparative abundance of microsatellites in different genic regions

Fig. 3 Predicted amino acid content encoded by the trinucleotide repeats

Fig. 4 Correlation between synonymous codon usage bias and GC composition in seabuckthorn microsatellite carrying unigene sequences

found to fall between the range of one microsatellite per 2 to 10 Kb. The variations in frequencies observed among earlier studies by different groups could be due to the different tools used to screen microsatellites (Varshney et al. 2005; Sharma et al. 2007) and variation in the size of sequence dataset.

Among microsatellites, dinucleotides were most frequent repeats (63.46 %) followed by trinucleotide repeats (31.35 %) while a very small share of 5.19 % was altogether contributed by tetra-, penta- and hexa-nucleotide repeats (Supplementary table 1). Similar abundance of dinucleotide repeats over trinucleotide repeats, was reported in other plant species (Tang et al. 2009; Roorkiwal and Sharma 2011). One of the possible reasons behind such abundance could be that shorter repeating units allow more number of possible slippage events per unit length of DNA. Considering the length of microsatellite loci, 75.56 % microsatellite were grouped as class II microsatellite repeats (< 20 nucleotide long) and 24.44 % microsatellite as class I microsatellite repeats ($\geq$ 20 nucleotide long) (Fig. 1). Similar preponderance of class II repeats has been reported in other studies also (Roorkiwal and Sharma 2011). The inverse relationship between the length of microsatellites and their abundance holds true in the present study (Fig. 1). This could be due to the fact that longer repeats have downward mutation bias (Katti et al. 2001). However, a slight increase in abundance of microsatellites with lengths 30–35 bp was observed in the present dataset, which could be attributed to further inclusion of hexanucleotide repeats in the interval class.

In terms of individual SSR motif, the most common dinucleotide repeat motif class was (AG) $_n$ (comprising of AG, GA, CT, TC) followed by AT repeat class, representing 63.30 % and 25.75 % of total dinucleotide repeats, respectively. Similar abundance of GA dinucleotide repeats, has been observed in many other plant species in the recent past (Tang et al. 2009; Roorkiwal and Sharma 2011; Victoria et al. 2011). This repeat class is known to play an active role in gene regulation in both animal and plant species (Trivedi 2004). Scarcity of GC dinucleotide repeats observed in this study could possibly be due to some strong negative selection by suppression of GC repeats, which reduces the stacking energy of DNA thus facilitating replication and transcription (Trivedi 2004).

Among trinucleotide motifs, (AAG) $_n$ repeat class was the most predominant as also observed in earlier studies in other plant species (Tang et al. 2009; Roorkiwal and Sharma 2011; Victoria et al. 2011). Grover and Sharma (2004) suggested tendency of making alternative DNA structures by repeat motif as the reason underlying such abundance.

Microsatellites in coding regions

Respective SSR positions were designated in unigene on the basis of prediction of ORF using OrfPredictor. In the present study, we found intense density of microsatellites in coding regions of the genes as compared to UTRs (Fig. 2). It could be

explained on the basis of direct relationship that exist between the rate of recombination and microsatellite density, as coding regions are known to have high recombination frequency (Trivedi 2004). Other possible reason could be the under representation of UTRs in unigene data. Location of 54 microsatellite loci (0.7 %) could not be assigned, possibly because of smaller size of unigene or the unigene might be representing a non-coding RNA.

Dinucleotide repeats were found to be richly distributed in all genic regions. The abundance of dinucleotide repeats in coding regions was followed by that in 3' and 5' untranslated regions contradicting some previous reports (Morgante et al. 2002). Nevertheless, such deviation could be due to the differences in definition parameters used during microsatellite screening or it could also be a species specific characteristic but more research is required to reach any concrete conclusion.

Majority of trinucleotide repeats (62.1 %) and hexanucleotide repeats (56.3 %) were found distributed in

Table 1 Top BLAST results for seabuckthorn unigenes containing microsatellites

Species	Percentage BLAST hit
<i>Vitis vinifera</i>	36.73407788
<i>Populus trichocarpa</i>	22.54753999
<i>Ricinus communis</i>	22.06459402
<i>Glycine max</i>	4.648354965
<i>Malus x domestica</i>	1.931783882
<i>Arabidopsis thaliana</i>	1.116812557
<i>Medicago truncatula</i>	0.784787202
<i>Arabidopsis lyrata</i>	0.784787202
<i>Gossypium hirsutum</i>	0.633866586
<i>Prunus persica</i>	0.543314217
<i>Nectria haematococca</i>	0.48294597
<i>Jatropha curcas</i>	0.301841232
<i>Cucumis melo</i>	0.301841232
<i>Nicotiana tabacum</i>	0.241472985
<i>Pyrenophora tritici-repentis</i>	0.181104739
<i>Pisum sativum</i>	0.241472985
<i>Oryza sativa</i>	0.181104739
<i>Pyrus pyrifolia</i>	0.120736493
<i>Hevea brasiliensis</i>	0.150920616
<i>Populus tremula</i>	0.150920616
<i>Solanum lycopersicum</i>	0.120736493
<i>Fagus sylvatica</i>	0.120736493
<i>Pyrenophora teres</i>	0.120736493
<i>Lotus japonicas</i>	0.120736493
<i>Citrus sinensis</i>	0.120736493
<i>Cicer arietinum</i>	0.120736493
<i>Betula pendula</i>	0.120736493
<i>Picea sitchensis</i>	0.120736493
<i>Cucumis sativus</i>	0.120736493
Others	4.769091458

coding regions. This could be due to the fact that reading frame would not be disturbed by slippage of a tri- or hexa-nucleotide repeat (Metzgar et al. 2000). There were 5.2 % microsatellites located between the boundaries of UTR – CDS, and role of such microsatellites have not been studied yet.

The predicted amino acid content for the trinucleotide repeats falling in coding regions is shown in Fig. 3. The amino acids serine, leucine, glutamine, and phenylalanine were predominant. Similar predominance of these amino acids has also been reported in other plant species (Jung et al. 2005; Lawson et al. 2006; Parida et al. 2006; Roorkiwal and Sharma 2011).

There exist a negative relationship between synonymous codons and GC composition as GC at 3rd position (GC3) plays a major role in this relationship (Fig. 4). In our data, there is a higher preference for GC3 and similar preference has been reported in case of other plants (Victoria et al. 2011).

Gene ontology annotation of microsatellite positive unigenes

Out of a total of 6787 putative unigenes harboring microsatellite loci, 3313 (48.81 %) were assigned gene ontology categories following BLAST matches to known proteins using BLAST2GO. *Vitis vinifera*, *Populus trichocarpa*, *Ricinus communis* were among the top BLAST- Hit species (Table 1). These SSR containing putative unigenes were assigned Gene Ontology (GO) terms (level 3) for biological process, cellular component and molecular function. In case of biological process category, unigenes were categorized into 31 different processes (Fig. 5a). Majority of the unigenes were involved in metabolic processes (16.4 %), followed by biosynthetic processes (8.065 %), while 5.57 % unigenes were found to be involved in responses to various stresses. However, other important processes included transcription

Fig. 5 Gene Ontology annotation of microsatellite carrying seabuckthorn unigenes **a**) Biological process; **b** Cellular component, and **c** Molecular function

Fig. 6 Polymorphism displayed by microsatellite locus USMM3. L: Molecular size marker (100 bp O'GeneRuler, Fermentas, Lithuania) 1–18: Accessions of *H. rhamnoides* collected from diverse locations of Leh

(7.033 %), transport (5.252 %), etc. For the cellular components category, majority of the unigenes were found to be related to plastid (13.93 %) followed by protein complex (12.97 %) and cytoplasm (9.41 %). The remaining sequences were assigned to other important cellular components like plasma membrane (8.63 %), mitochondria (7.95 %) and nucleus (6.96 %), etc. (Fig. 5b). When concentrating on molecular function category, majority of unigenes were assigned to protein binding (13.52 %) followed by nucleotide binding (10.89 %) (Fig. 5c).

Validation of unigene based microsatellite markers and application of variability score

Out of the 25 primer pairs designed to amplify microsatellite sequence loci, 21 pairs successfully resulted in amplification

of target loci in all the accessions of *H. rhamnoides*, and also cross-amplified in *H. salicifolia*. On resolving amplicons on 4 % metaphor agarose, eleven USMM were found polymorphic in *H. rhamnoides* (Fig. 6). We detected 2 to 7 alleles for different loci included in the present study. However, the number of polymorphic markers obtained in small marker set is quite high as compared to our own previous study (Jain et al. 2010) in which nine polymorphic marker were obtained out of 30 putative microsatellite markers. One possible reason behind such an observation could be biased selection of class I microsatellites carrying sequences from the complete data set for development of markers. Class I microsatellites have been found more polymorphic as compared to class II microsatellites in earlier studies also (Singh et al. 2010). Analysis of selected sequences for their 'Var' score further improved the proportion of polymorphic

Table 2 Amplification profile of polymorphic unigene specific microsatellite markers in seabuckthorn

Marker Code	Primer Sequences		Annealing Temperature (°C)	Microsatellite (sequence and length)	Location	Number of alleles	Var Score
	Forward 5'-3'	Reverse 5'-3'					
USMM 1 (Unigene 65)	GGC GAA ACT TGA CTT GTT GC	ACC GAT CAA TAC CGT TCT GC	55	(TAC) ₁₆	3'	4	1.1
USMM 3 (Unigene 997)	AAG GAT GTG GTC GAT CCA AG	GTT TGC AGG CAT TCC TTT GT	54	(TTC) ₁₀	CDS	7	0.9
USMM 5 (Unigene 1217)	TTC GAT CGG ATA AGG TCA TTG	GCA GTC GAG GAG GTT TGA AG	56	(AC) ₉ (AT) ₆	exon	5	NC
USMM 7 (Unigene 3650)	TCG CCG TCT GTT TCA GAT AA	GCT GAT CCA ACG GTC TCA TT	50	(AG) ₁₈	3'	3	1
USMM 8 (Unigene 6478)	GGT CGG AGA TGA TTC AAA GC	TAC GCT CTC AAC GCT TGC TA	50	(TCTTCA) ₇	CDS	5	0.4
USMM 12 (Unigene 2022)	CGT CAC CGT CAG TTC AGA GA	AAG GAG ACG CAG ACG AAG AG	50	(AAG) ₁₀	5'	4	1
USMM 13 (Unigene 2807)	GAT ATG CCG GCG GTA AAC TA	TGT ACC ACT TCC TGC TGC TG	57	(CAG) ₅ (CAC) ₆	CDS	3	NC
USMM 16 (Unigene 2900)	CCA ACC AAC CTC ATC GTT TC	ATC GGA GGG ATC GTT GAA AG	55	(TC) ₁₄	5'	2	1.1
USMM 24 (Unigene 8141)	TAG CAT TGC AGG CTC AGA GA	ATC CGT GGT TAA GGT TGC AC	55	(AG) ₁₁	3'	3	1
USMM 25 (Unigene 11683)	CGA GGT CCG AGT AGG AAG A	CAT TGG CCT TCA ATC TCC TC	55	(AAG) ₁₀	CDS-3'	3	0.9
USMM 26 (Unigene 11918)	TTC CCT CTC CGT GAC AAG TT	GCT TGA GGT GAA GAG GAA AGT C	55	(CT) ₁₆	3'	4	1.

markers. In the current study, out of total 21 amplified putative markers, eight microsatellites markers were found polymorphic out of 11 selected sequences with var cut off value of ≥ 0.9 . While one marker was found polymorphic out of 8 selected sequences with score less than 0.9 and two markers were found to be polymorphic where Var score could not be calculated due to the presence of compound repeat (Table 2). Thus overall selection of sequences with class I microsatellites and appropriate Var score, can allow higher return of polymorphic markers. Since the sample size is not very large and no statistical analysis conducted, chance factor can also be not ruled out.

High number of alleles observed in a small population in the present study may not alone be attributed to structural features of microsatellites and outcrossing nature of seabuckthorn, especially when markers have been designed from genic regions. Apart from the choice of the technique for the assessment (in terms of resolution and accuracy) of amplicons, such level of polymorphism could also be attributed to the nature of the genes, which need further investigations.

Conclusion

As unigene specific microsatellite markers are transferable from one taxon to another, microsatellites documented in the present study provide a useful resource for marker based applications in seabuckthorn and related taxa. These marker sites can be further exploited for recording allelic variance across diverse collection, and for the development of gene based functional markers. Further study of microsatellite positive genes involved in various pathways may help in understanding the regulatory role(s) of different microsatellite loci in plants.

Acknowledgments We are thankful to DIHAR, Leh, India for providing seabuckthorn samples. This work was supported by grants from DRDO (Defence Research and Development Organization), India and DBT (Department of Biotechnology), India.

References

- Angellotti MC, Bhuiyan SB, Chen G, Wan XF (2007) CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res* 35:W132–W136
- Chauhan AS, Negi PS, Ranterke RS (2007) Antioxidant and antibacterial activities of aqueous extract of Seabuckthorn (*Hippophae rhamnoides*) seeds. *Fitoterapia* 78:590–592
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Dhyani D, Maikhuri RK, Rao KS, Kumar K, Purohit VK, Sundriyal M, Saxena KG (2007) Basic nutritional attributes of *Hippophae rhamnoides* (Seabuckthorn) populations from Central Himalaya, India. *Curr Sci* 92:1148–1152
- Elena T, Capraru G, Rosu CM, Zamfirache MM, Olteanu Z, Manzu C (2011) Morphometric pattern of somatic chromosomes in three Romanian seabuckthorn genotypes. *Caryologia* 64:189–196
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* 99:125–132
- Fatima T, Snyder CL, Schroeder WR, Cram D, Datla R, Wishart D, Randall JW, Krishna P (2012) Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. *PLoS One* 7:e34099
- Ghangal R, Raghuvanshi S, Sharma PC (2008) Exploring seabuckthorn genome using molecular tools. In: Dwivedi SK, Perimalazagan T, Singh SB, Ahmad Z (eds) Seabuckthorn- the wonder plant. SSPH Publishers, Delhi, pp 347–371
- Ghangal R, Raghuvanshi S, Sharma PC (2012) Expressed sequence tag based identification and expression analysis of some cold inducible elements in seabuckthorn (*Hippophae rhamnoides* L.). *Plant Physiol Biochem* 51:123–128
- Ghangal R, Chaudhary S, Jain M, Purty R, Sharma PC (2013) Optimization of de novo short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. *Plos One* 8:e72516
- Grover A, Sharma PC (2004) Occurrence of simple sequence repeats in potato ESTs is not random: an in silico study on distribution and length of single sequence repeats. *Potato J* 31:98–105
- Grover A, Sharma PC (2011) Is spatial occurrence of microsatellites in the genome a determinant of their function and dynamics contributing to genome evolution? *Curr Sci* 100:859–869
- Grover A, Aishwarya V, Sharma PC (2012) Searching microsatellites in DNA sequences: approaches used and tools developed. *Physiol Mol Biol Plants* 18:11–19
- Guo W, Li B, Zhang X, Wang R (2007) Architectural plasticity and growth responses of *Hippophae rhamnoides* and *Caragana intermedia* seedlings to simulated water stress. *J Arid Environ* 69:385–399
- Jain A, Ghangal R, Grover A, Raghuvanshi S, Sharma PC (2010) Development of new EST based SSR markers in seabuckthorn. *Physiol Mol Biol Plants* 16:375–378
- Jung S, Abbott A, Jesudurai C, Tomkins J, Main D (2005) Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. *Funct Integr Genomics* 5:136–143
- Kalia RK, Singh R, Rai MK, Mishra GP, Singh SR, Dhawan AK (2011) Biotechnological interventions in sea buckthorn (*Hippophae* L.): current status and future. *Trees* 25:559–575
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18:1161–1167
- Lawson MJ, Zhang L, Tech V (2006) Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes. *Genome Biol* 7:R14
- Legendre M, Pochet N, Pak T, Verstrepen KJ (2007) Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res* 17:1787–1796
- Ma JQ, Ma CL, Yao MZ, Jin JQ, Wang ZL, Wang XC, Chen L (2012) Microsatellite markers from tea plant expressed sequence tags (ESTs) and their applicability for cross-species/genera amplification and genetic mapping. *Sci Hort* 134:167–175
- Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10:72–80
- Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33:W677–W680
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet* 3:194–200
- Parida SK, Kumar ARK, Dalal V, Singh NK, Mohapatra T (2006) Unigene derived microsatellite markers for the cereal genomes. *Theor Appl Genet* 112:808–817

- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, de Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Gen Genomics* 276:436–449
- Roorkiwal M, Sharma PC (2011) Mining functional microsatellite in legume unigenes. *Bioinformation* 7:264–270
- Rozen S, Skaletsky H (2000) Primer3 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, Totowa, pp 365–386
- Ruan CJ, Rumpunen K, Nybom H (2013) Advances in improvement of quality and resistance in a multipurpose crop: sea buckthorn. *Crit Rev Biotechnol* 33:126–144
- Sharma PC, Grover A, Kahl G (2007) Mining microsatellites in eukaryotic genomes. *Trends Biotechnol* 25:490–498
- Singh H, Deshmukh RK, Singh A, Singh AK, Gaikwad K, Sharma TR, Mohapatra T, Singh NK (2010) Highly variable SSR markers suitable for rice genotyping using agarose gels. *Mol Breeding* 25:359–364
- Singh BK, Thakur AK, Tiwari SK, Siddiqui SA, Singh VV, Rai PK (2012) Transferability of Brassica-derived microsatellites to related genera and their implications for phylogenetic analysis. *Nat Acad Sci Lett* 35:37–44
- Tang S, Okashah RD, Cordonnier-Pratt MM, Pratt LH, Johnson VE, Taylor CA, Arnold ML, Knapp SJ (2009) EST and EST-SSR marker resources for *Iris*. *BMC Plant Biol* 9:72
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106: 411–422
- Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10:967–981
- Trajkovski V, Jeppsson N (1999) Domestication of sea buckthorn. *Botanica Lithuanica Suppl* 2:37–46
- Trivedi S (2004) Microsatellites (SSRs): puzzles within puzzle. *Indian J Biotechnol* 3:331–347
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23: 48–55
- Victoria FC, da Maia LC, de Oliveira AC (2011) In silico comparative analysis of SSR markers in plants. *BMC Plant Biol* 11:15
- Wang A, Zhang Q, Wan D, Yang Y, Liu J (2008) Nine microsatellite DNA primers for *Hippophae rhamnoides* ssp. *sinensis* (Elaeagnaceae). *Conserv Genet* 9:969–971
- Wohrmann T, Weising K (2011) In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theor Appl Genet* 123:635–647
- Xu G, Li C, Yao Y (2009) Proteomics analysis of drought stress-responsive proteins in *Hippophae rhamnoides* L. *Plant Mol Biol Rep* 27:153–161
- Yang Y, Yao Y, Xu G, Li C (2005) Growth and physiological responses to drought and elevated ultraviolet-B in two contrasting populations of *Hippophae rhamnoides*. *Physiol Plant* 124:431–440
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation; a review. *Mol Ecol* 11:1–16