Biotechnology Research and Innovation (2018) 2, 9–18

SBBiotec
SOCIEDADE BRASILEIRA DE BIOTECNOLOGIA

Biotechnology
Research & Innovation

http://www.journals.elsevier.com/biotechnology-research-and-innovation/

RESEARCH PAPER

Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*

Nidhi Shukla[a,c], **Himani Kuntal**[a], **Asheesh Shanker**[b,*], **Satyendra Nath Sharma**[c]

[a] *Department of Bioscience and Biotechnology, Banasthali Vidyapith, Banasthali 304022, Rajasthan, India*
[b] *Department of Bioinformatics, Central University of South Bihar, Gaya, Bihar, India*
[c] *Seed Technology Research, Rajasthan Agricultural Research Institute, Jaipur 302018, Rajasthan, India*

Received 13 October 2017; accepted 17 August 2018
Available online 30 August 2018

KEYWORDS
Vigna;
Simple sequence repeats;
Chloroplasts;
Transferability

Abstract The genus *Vigna* is distributed throughout the world and it is phylogenetically closely related to agriculturally important crops. Simple sequence repeats (SSRs) are defined as sequence repeat units that occur between 1 and 6 bp and found in both coding and non-coding regions of genome. The availability of organelle genome sequence helps in understanding the organization of SSRs in them. In this study, chloroplast genome sequences of the genus *Vigna* were screened for the identification of chloroplast simple sequence repeats (cpSSRs). A total of 21, 24 and 25 cpSSRs in *Vigna angularis*, *Vigna radiata* and *Vigna unguiculata*, respectively, were mined computationally. The density of ~1SSR/6.1–7.2 kb was observed. Depending on the different repeat units, the length of SSRs ranged from 12 to 24 bp. Mononucleotides were the most abundant repeats followed by dinucleotide repeats. Hexanucleotide repeats were completely absent in the chloroplast genomes of genus *Vigna*. Seven cpSSRs were selected on the basis of repeats to study transferability. The transferable markers can be employed in genome mapping and comparative mapping among *Vigna* species in future studies.

Introduction

The genus *Vigna* comprises of several species including *V. angularis* (Willd.) Ohwi & H.Ohashi, *V. unguiculata* (L.) Walp., and *V. radiata* (L.) R.Wilczek. The *Vigna* species may be used as model legume plants in genetic research because of its short life cycle and small genome size. Molecular markers play important role to study genetic diversity of plants and Simple sequence repeats (SSRs) are one of them.

SSRs or microsatellites are a class of molecular markers based on tandem repeats of short (1–6 nucleotide) DNA sequences, and are present in large quantities in both coding and non-coding region of genome (Zane, Bargelloni, & Patarnello, 2002). SSRs are locus-specific, co-dominant, highly polymorphic and uniformly distributed over the genome (Powell, Machray, & Provan, 1996). A high level of genetic variation is observed due to differences in number of tandem repeating units of SSRs at a locus which produces

* Corresponding author
E-mail: ashomics@gmail.com (A. Shanker).

https://doi.org/10.1016/j.biori.2018.08.001

a highly polymorphic banding pattern detected by the Polymerase Chain Reaction (PCR) using locus specific flanking primers. Due to these characteristics, SSRs are widely used for genetic marker development and genome application (Cheng et al., 2016; Morgante & Olivieri, 1993; Oliveira, Padua, Zucchi, Vencovsky, & Vieira, 2006; Srivastava & Shanker, 2015; Tautz, 1989).

Potential coding regions were analyzed by BLAST2GO (Conesa et al., 2005) for functional annotation of the genes. It leads to consistent gene annotations, gene names, Gene Ontology (GO) numbers, and gene pathways. Moreover, it provides description of gene products relate to molecular function, biological process and cellular component. Gene annotation process brings novel experimental results in relationship with the current scientific knowledge stored in the ontology. GO development paradigm is one of the strength that has been performed by expert biologist-curators and as a consequence the GO is continually being updated with the new information. GO curators often request to add new terms and information of the GO structure and it also accepts requests from external users. In addition, the GO community works with scientific experts for the corresponding parts of the gene ontology process to evaluate and update (Diehl, Lee, Scheuermann, & Blake, 2007).

SSRs can be transferred from different genotypes within or between species (Wang, Walla, Zhong, Huang, & Dai, 2012). Earlier, cross-species transferability of SSRs was detected via PCR amplification in different related species using primer pairs designed from one species (Decroocq, Favé, Hagen, Bordenave, & Decroocq, 2003; Gasic et al., 2009). In addition, SSRs from adzuki bean (*V. angularis*) (Han et al., 2005; Wang, Kaga, Tomooka, & Vaughan, 2004) and cowpea (*V. unguiculata*) (Li, Fatokun, Ubi, Singh, & Scoles, 2001) have been transferred to mung bean (Chankaew, Somta, Sorajjapinun, & Srinives, 2011; Sangiri, Kaga, Tomooka, Vaughan, & Srinives, 2007) and relatively high polymorphism SSR rates were reported between *Vigna* species, while low variation level was found within species. Most of the previous studies on SSRs are based on sequence databases in which coding regions are mainly characterized and are generally nuclear-genome based (Zapiola, Cronn, & Mallory-Smith, 2010). However, SSRs have also been developed for organelle genomes which are maternally inherited (Jansen et al., 2005; Provan, Powell, & Hollingsworth, 2001) and play important role in deducing phylogenetic relationship in plants (Olmstead & Palmer, 1994; Shanker, 2013a, 2013b, 2013c, 2013d; Shanker, Sharma, & Daniell, 2011).

The conventional methods of generating SSRs from genomic libraries was replaced by *in silico* mining of SSRs from DNA sequences available in biological databases (Kabra, Kapil, Attarwala, Rai, & Shanker, 2016; Kapil, Rai, & Shanker, 2014; Kumar, Kapil, & Shanker, 2014). The availability of chloroplast genomes of *Vigna* at National Centre of Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov) provides opportunity to mine them and analyze for the presence of SSRs. Therefore, the present work was designed to computationally identify SSRs in the chloroplast genomes of genus *Vigna*, and to detect cross transferability of SSRs among genus *Vigna*. Moreover, wet lab experiments were performed to confirm the findings.

Table 1 The information of chloroplast genomes used in SSR mining.

S. No.	Organism	Accession number	Genome size (kb)
1	*Vigna angularis*	NC_021091	151.683
2	*Vigna radiata*	NC_013843	151.271
3	*Vigna unguiculata*	NC_018051	152.415

Materials and methods

Sequence retrieval, mining of SSRs and primer designing

The complete chloroplast genome sequences of different species of *Vigna* were retrieved in fasta and GenBank format from NCBI (Table 1). MISA (http://pgrc.ipk-gatersleben.de/misa), a perl script was used for the detection of both perfect (where each repeat follows the next without interruptions) and compound (where two or more repeat units are adjacent to each other) SSRs. The minimum repeat size was set to be >=12-mono, >=6-di, >=4-tri, >=3-tetra, penta, and hexanucleotides. The maximum difference of 0 was taken between two SSRs. Primer pairs for flanking sequences of each SSR were designed using Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm) with default parameters of GC content, melting temperature, primer and PCR product size.

Gene ontology

Functional annotation of the coding sequences of *Vigna* species was done using Blast2GO (Conesa et al., 2005). A fasta file with all coding cpSSRs was subjected to Blast2GO software against the gene ontology (GO) annotated sequences, and the obtained hits were compiled in order to assess associations between SSR loci and biological processes, cellular components and molecular function of known genes. By entering Pfam ID, InterProScan ID, ProSiteProfiles ID, SUPERFAMILY ID and Panther ID of interest, user will have the list of transcription factors (TFs) with annotation information. By clicking on a particular accession id, it will give the detail information about its family, contig position, length, TF family, physicochemical properties and sequence information (Singh, Sharma, Singh, & Sharma, 2017).

In-silico PCR

The primer pairs designed for cpSSRs of *Vigna* species were used to perform *in-silico* PCR and gel simulation using SPCR/Sim Gel (Cao et al., 2005) with default parameters to check product amplification and study transferability. The whole chloroplast genome sequences of *Vigna* species were used as templates and virtual amplicons are obtained for each primer pair. A total of seven primer pairs were selected for validation of transferability on the basis of different repeats (mono, di, tri and tetra). Out of these, 3 primer pairs

were of *V. angularis* and rest of *V. radiata* were chosen. Basic local alignment search tool (BLAST; Altschul et al., 1997) was used for using primer sequences as query against nuclear genome database of *V. angularis* and *V. radiata*. None of the primer sequence shows significant similarity with the nuclear genome. It suggests that the designed primers are chloroplast specific.

Plant materials

The seed material for experiment includes, 3 genotypes of *V. radiata* (RMG 344, RMG 492 and IPM-02-3), 1 genotype of *V. mungo* (IPU-94-1), 4 genotypes of *V. unguiculata* (RC 101, Pusa Komal, C-152-UAS and Gomati) and 2 genotypes of *V. unguiculata* subsp. sesquipedalis (Red and black yard long bean). The genotypes of genus *Vigna* were collected from pulse breeder All India Coordinated Research Project (AICRP), Rajasthan Agricultural Research Institute, Jaipur, India under standard cultivated conditions.

DNA extraction and wet lab PCR amplification

Two experimental methods are often used to collect cpDNA in plants. The first method is the whole chloroplast genome amplification from total DNA using PCR, and the second is isolating cpDNA based on sucrose gradient which is highly expensive method (Shi et al., 2012). Therefore, the former method was used to isolate cpDNA from fresh leaves of different species/genotypes of *Vigna* according to the Cetyl Trimethyl Ammonium Bromide (CTAB) method by Sharma, Kumar, Singh, and Sharma (2011) which is a completely modified version of Doyle and Doyle (1990). The quality of the extracted DNA was examined by using 0.8% (W/V) agarose gel and DNA concentration was quantified using a Nano-Drop spectrophotometer (Nano-Drop 1000 Version 3.1.1, USA). Extracted DNA was diluted to 100 ng/μl using nuclease free water. The PCR reactions were performed in a final volume of 25 μl containing 1X Taq assay buffer, 0.5 units of Taq DNA polymerase, 200 μM of each dNTPs (Bangalore Genei Pvt. Ltd., India), 0.2 μM primers and 50 ng of template DNA. The PCR amplification was performed using a thermal cycler (Model CGI-96, Corbett Research, Australia) and repeated thrice for each primer pair to ensure the reproducibility. The PCR was programmed as an initial step of melting of 4 min at 94 °C followed by 35 cycles of denaturation at 94 °C for 30 s, primer annealing at 60 °C for 30 s and elongation at 72 °C for 60 s and last cycle was followed by a final extension step for 7 min at 72 °C. The PCR products were analyzed on 1.2% agarose gel containing 0.5 μg/μl ethidium bromide. After separation, gels were documented using Biovis Image Plus software (Expert Vision Pvt. Ltd. Mumbai).

Results

Distribution of SSRs in chloroplast genomes of genus *Vigna*

A total of 21, 24 and 25 perfect cpSSRs were identified in *V. angularis*, *V. radiata* and *V. unguiculata*, respectively.

Table 2 Information of SSRs identified in chloroplast genomes of genus *Vigna*.

Parameters	Organisms		
	V. angularis	*V. radiata*	*V. unguiculata*
Chloroplast genome size	151683 bp	151271 bp	152415 bp
Total SSRs identified	21	24	25
Density of SSR	1 SSR/7.2 kb	1 SSR/6.3 kb	1 SSR/6.1 kb
Coding region	4 (19.05%)	5 (20.8%)	3 (12%)
Average length of SSR	13.76 bp	13.21 bp	13.04 bp
Repeat type			
Mononucleotides	8 (38.1%)	8 (33.3%)	7 (28%)
Dinucleotides	7 (33.3%)	5 (20.8%)	7 (28%)
Trinucleotides	1 (4.8%)	5 (20.8%)	6 (24%)
Tetranucleotides	5 (23.8%)	5 (20.8%)	5 (20%)
Pentanucleotides	–	1 (4.2%)	–

The details of all cpSSRs identified are represented in online supplementary file. The density of cpSSRs observed in *V. angularis*, *V. radiata* and *V. unguiculata* were 7.2, 6.3 and 6.1 cpSSR per kb, respectively. The average length of cpSSRs ranged from 13 to 14 nucleotides in these chloroplast genomes (Table 2).

Mononucleotides were the most abundant repeats identified (Fig. 1A and B) with T as most abundant repeat followed by A which is in support with previous results in *Oryza sativa* (Rajendrakumar, Biswal, Balachandran, Srinivasarao, & Sundaram, 2007), *Sesamum indicum* (Yi & Kim, 2012), Camellia species (Huang, Shi, Liu, Mao, & Gao, 2014) and *Cajanus cajan* (Kaila et al., 2016). Majority of SSRs in the chloroplast genome are mononucleotide A/T repeats (Wheeler, Dorman, Buchanan, Challagundla, & Wallace, 2014). Among dinucleotides, AT was followed by TA in *Vigna* species. It was reported that higher frequency of AT/TA motifs were found in olive species (Mariotti, Cultrera, Diez, Baldoni, & Rubini, 2010), *Sesamum indicum* (Yi & Kim, 2012) and *Glycine* species (Ozyigit, Dogan, & Filiz, 2015). TAT was the only trinucleotide repeat found in *V. angularis*. AAT was the most abundant repeat followed by TAA/TAT/TTA/ATA in *V. radiata* and *V. unguiculata*. Among tetranucleotide repeats, AATA was most abundant repeat followed by TTGA/CTAT in *Vigna* species. Only one pentanucleotide repeat AATAA was found in *V. radiata*. Hexanucleotides were completely absent in the chloroplast genomes of genus *Vigna*.

Gene ontology

Based on gene ontology analysis, the molecular functions of cpSSRs identified were classified as transporter activity, catalytic activity and binding. The majority involved in biological processes was related to cellular component organization or biogenesis, localization, metabolic process,

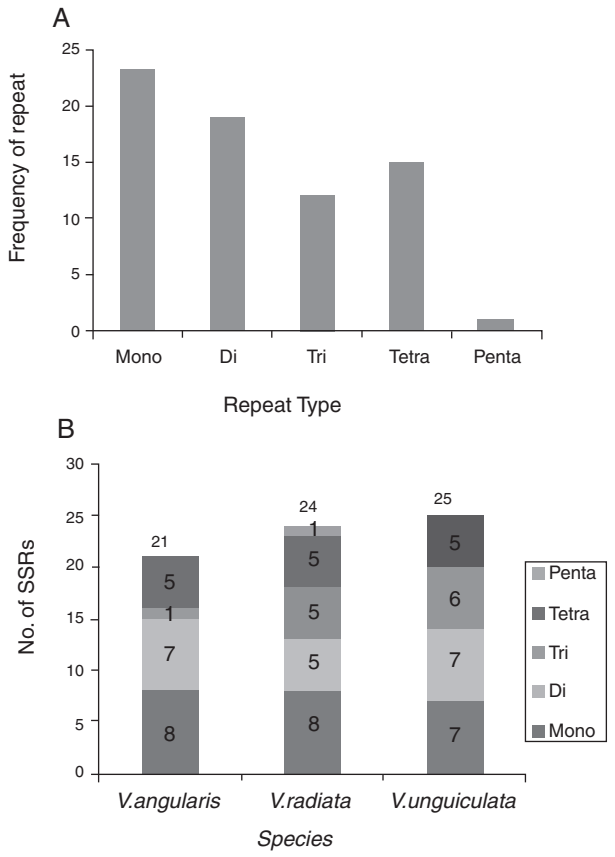


Fig. 1 (A) Distribution of cpSSRs in genus *Vigna* (B) Types of SSRs detected in three chloroplast genomes of genus *Vigna*.

single-organism process and cellular process. Moreover, the cellular component contained number of groups having macromolecular complex, membrane part, membrane, organelle part, cell part, organelle and cell (Fig. 2).

The transcriptome analysis, using Pfam (Protein families, https://pfam.xfam.org/) and Panther (Protein Analysis Through Evolutionary Relationships, http://pantherdb.org), of chloroplast genome of *Vigna* species indicated complete information about the function of protein-coding genes and their evolutionary relationship among the species.

One of the coding sequences of *V. angularis*, with a panther match score of 134.4, was characterized as NADH-Ubiquinone/plastoquinone oxido-reductase family. It had 3 GO ids: (i) GO: 0042773 for ATP synthesis coupled electron transport (ii) GO: 0055114 for oxidation-reduction processes and (iii) GO: 0016651 for oxido-reductase activity. The first two GO ids are under the category of biological processes and the third one, acting on NAD(P)H, comes under the category of molecular function.

One of the coding sequences of *V. radiata*, with panther match score of 79.4, was characterized by Acetyl-coenzyme-A carboxylase carbonyl transferase subunit beta, chloroplastic. It showed several pathways such as Aflatoxin biosynthesis, Candicidin biosynthesis, 3-hydroxypropanoate/4-hydroxybutanate, Jadomycin biosynthesis, Pyruvate metabolism, Biotin transport and metabolism, 3-hydroxypropanoate cycle, Propanoate metabolism, Glyoxylate assimilation, Octanoyl-[acyl-carrier protein] biosynthesis, Carbon fixation pathways in prokaryotes, Fatty acid biosynthesis, Fatty acid biosynthesis initiation-I. The second coding sequence, with panther match score of 58.9, was from Cytochrome C assembly protein family. It had 2 GO ids (i) GO: 0017004 for cytochrome complex assembly in the category of biological process and (ii) GO: 0020037 for heme-binding in the category of molecular function. The third coding sequence of *V. radiata*, with panther match score of 98.1, was NADH: ubiquinone oxidoreductase. It had 3 GO ids: (i) GO: 0055114 for oxidation-reduction process (ii) GO: 0042773 for ATP synthesis coupled electron transport, both are under the category of biological processes and (iii) GO: 0008137 for NADH dehydrogenase (ubiquinone) activity falls within the category of molecular function. The fourth coding sequence of *V. radiata*, with panther match score of 134.4, was NADH-Ubiquinone oxidoreductase subunit ND4L. It had 3 GO ids (i) GO: 0042773 for ATP synthesis coupled electron transport and (ii) GO: 0055114 for oxidation-reduction processes, both fall under the category of biological process and (iii) GO: 0016651 for oxido-reductase activity acting on NAD(P)H comes under the category of molecular function. One of the coding sequences of *V. unguiculata*, with panther match score of 50.6, was characterized as cytochrome-C assembly protein. It had 2 GO ids (i) GO: 0017004 for cytochrome complex assembly under the category of biological processes and (ii) GO: 0020037 for heme-binding comes under the category of molecular function. The second coding sequence, with panther match score of 97.8, was NAD(P)H Quninone oxidoreductase chain 4, chloroplastic. The GO ids were (i) GO: 0055114 for oxidation-reduction process in the category of biological processes and (ii) GO: 0016655 for oxidoreductase activity, acting on NAD(P)H, quinine, falls in the category of molecular function. The third coding sequence, with panther match score of 134.4, was NADH-ubiquinone oxidoreductase. It had 3 GO ids: (i) GO: 0016651 for oxidreductase, acting on NAD(P)H, falls in the category of molecular function (ii) GO: 0042773 for ATP synthesis coupled electron transport and (iii) GO: 0055114 for oxidation-reduction process under the category of biological process.

Cross-transferability of SSRs

Seven cpSSRs were selected to detect cross-transferability (Table 3) using virtual SPCR and wet lab technique for the chloroplast genomes of *Vigna* species (*V. radiata*, *V. mungo* and *V. unguiculata*). All the cpSSRs revealed 100% transferability within genus. The monomorphic amplicons were observed in most of the cpSSRs (Figs. 3 and 4).

Discussion

SSRs are distributed across plant genomes although their frequency and distribution patterns vary (Sonah et al., 2011). In mung bean the dominant repeat was mononucleotides (Chen et al., 2015), in citrus trinucleotides (Liu, Li, Long, Hu, & Zhang, 2013), in cucumber tetranucleotides (Cavagnaro et al., 2010), in cotton pentanucleotides (Zou, Lu, Zhang, & Song, 2012). The frequency of repeats also varied at the different genomic and transcriptional level (Moe et al., 2011).

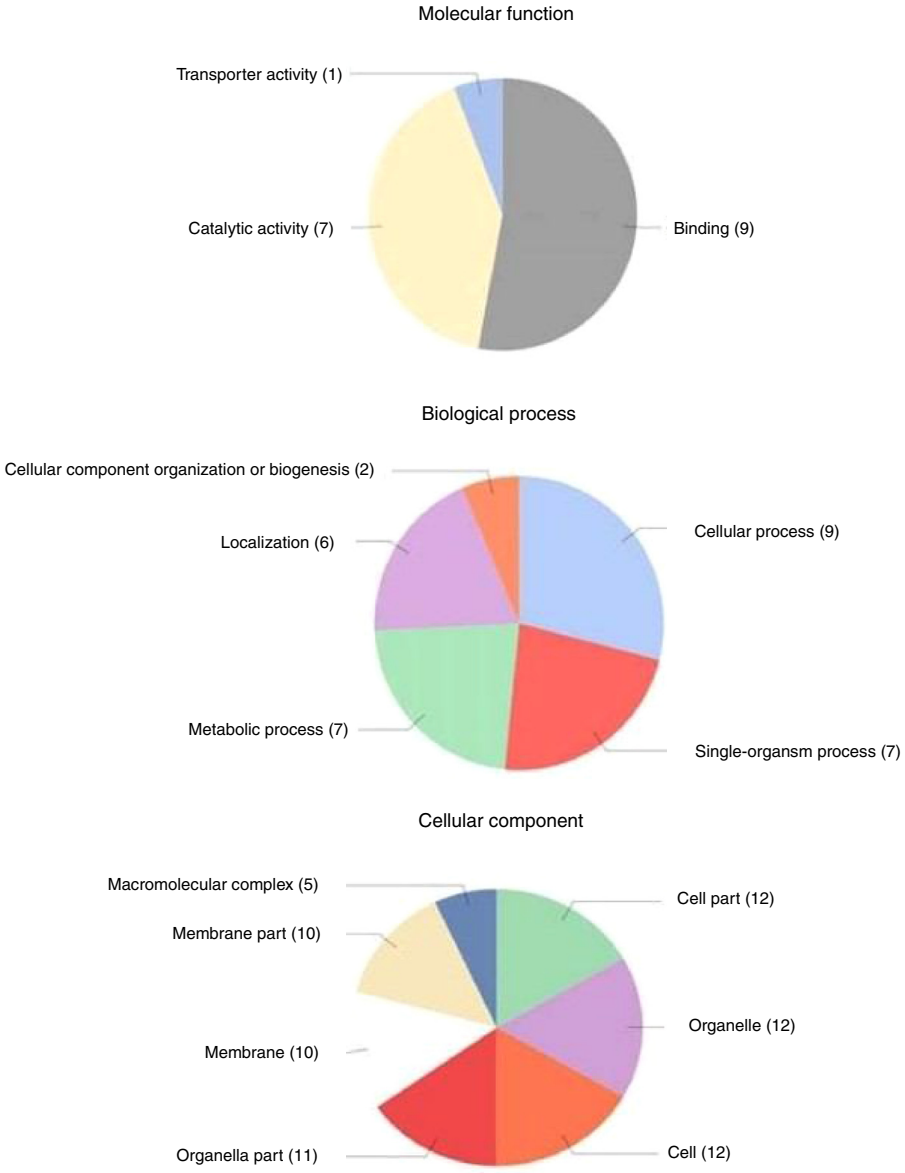


Fig. 2 Gene ontology distribution into molecular functions, biological processess and cellular components of cpSSRs for *Vigna* species.

SSRs were less in the coding regions compared to non-coding regions. Our findings were in agreement with the majority of SSRs that were embedded in non-coding DNA (Ellegren, 2004; Shanker, 2013d).

In this study, majority of SSRs identified in the chloroplast genomes are mononucleotide T/A repeats. These findings are in support with previous observations (Chen et al., 2015; Lawson & Zhang, 2006; Toth, Gaspari, & Jurka, 2000). The AT/TA repeats was the most predominant dinucleotide present in *Vigna* species which is in agreement with other studies (Powell et al., 1996). Whereas, the most frequent dinucleotide in monocot plant rice was AG/CT (Cheng et al., 2016). Among the dinucleotide repeats, the CG/CG repeat was extremely rare in genic and intergenic regions of the organellar genomes (Rajendrakumar et al., 2007) and in the *Vigna* species studied this repeat does not occur. In dinucleotide SSRs, the AT/TA percentage increased along with the evolution of plants from algae, fern and monocotyledon to dicotyledon (Qin et al., 2015). AAT was the most prominent trinucleotide repeat in *Vigna* species and this data was consistent with *Glycine* species (Ozyigit et al., 2015).

The density of mined cpSSRs is 1 SSR/6.1 kb, 1 SSR/6.3 kb and 1 SSR/7.2 kb in *V. unguiculata*, *V. radiata* and *V. angularis*, respectively. The average density of cpSSRs in *Vigna* species was found to be lower than the density of cpSSRs in family Solanaceae (1SSR/1.26 kb; Tambarussi et al., 2009), *Anthoceros formosae* (1 SSR/2.4 kb; Shanker, 2013d), rice (1 SSR/6.5 kb; Rajendrakumar et al., 2007), EST-SSRs in barley, maize, wheat, rye, sorghum and rice (1 SSR/6.0 kb, Varshney, Thiel, Stein, Langridge, & Graner, 2002). Whereas, the density of *Vigna* species was found higher as compared to EST-SSRs of cotton and poplar (1 SSR/20 kb and 1SSR/14 kb respectively; Cardle et al., 2000) and Unigene sequences of Citrus (1 SSR/12.9 kb; Shanker et al., 2007a).

Table 3 Description of primer pairs of cpSSRs in *Vigna* species.

S. No.	Motif	Start	End	Left/Right primer	Primer length	Tm	GC%
1	(AT)9	34674	34691	TCGTACAATCAGCGTAATCCAGA	23	59.621	43.478
				TGAACTGACGCTTAACCTGGT	21	59.585	47.619
2	(AT)7	64295	64308	CGAAATGAATTTTTAGTTCCGCGT	24	59.151	37.5
				TGGAATCGGACTCTAGGAAAGG	22	58.968	50
3	(AT)8	64961	64976	TGAAATGATATCCTTCCGAATGGGA	25	59.927	40
				GAGCTATGACACAATCAAACCCA	23	58.99	43.478
4	(AATA)3	112376	112387	AATTCTCGTGGCCCCGAATC	20	60.464	55
				TCTTGCATTACCGGGCATGA	20	59.746	50
5	(A)12	57485	57496	AAAGGGACCCTGCCTTTCAA	20	59.439	50
				CGTGAATAGTCATTGATTCATTCGC	25	59.157	40
6	(AAAG)3	116935	116946	CCCACATCATTTTCGATTCCGA	22	59.058	45.455
				CCCCACAATTCCGATCTAGAGT	22	59.299	50
7	(AAT)5	9917	9931	TGTATGGCGCAACCCAATCT	20	60.035	50
				TTTCCGGGGATAAAGCTGCC	20	60.395	55

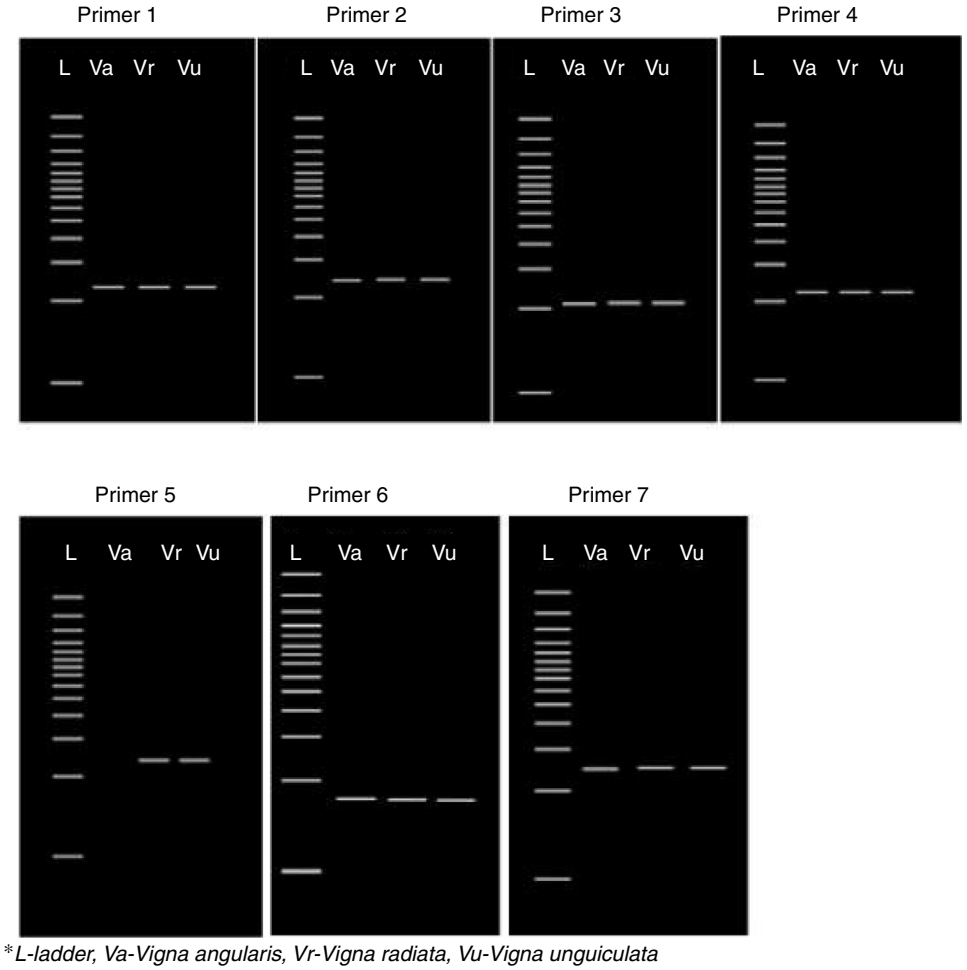
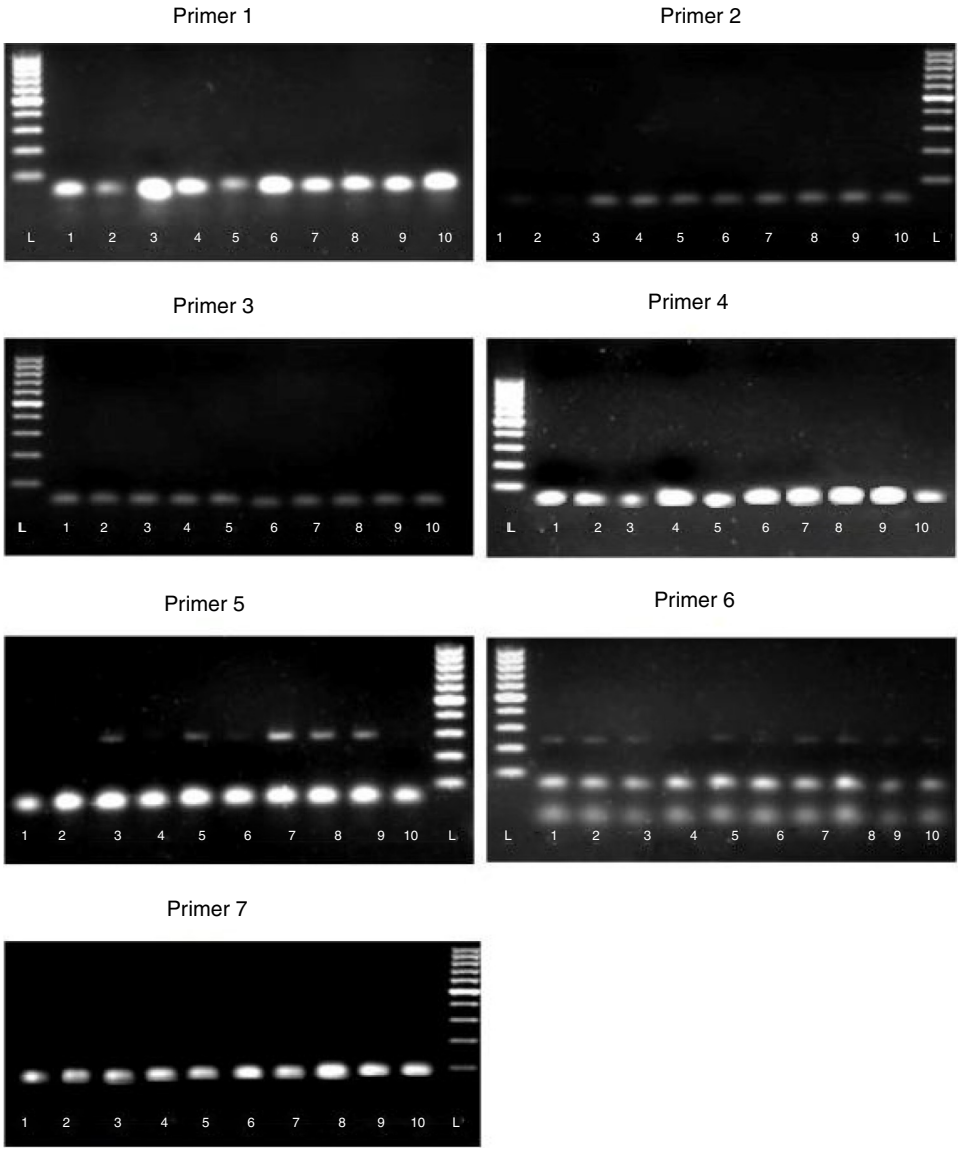


Fig. 3 *In-silico* transferability of cpSSRs for *Vigna* species.

Moreover, the frequencies of repeats having different region in genome also vary greatly (Mun et al., 2006; Qian et al., 2013) indicating that different repeats play different roles in gene ontologies. The gene anntotation of coding cpSSRs in different *Vigna* species (*V. angularis*, *V. radiata*, *V. unguiculata*) were categorized in many functions. It was observed that the distribution of sequences shared between *V. angularis*, *V. radiata* and *V. unguiculata* arose from a common ancestor. Therefore, this combined mapping would enhance the use of coding sequences in developing genetic



*L-ladder, 1-RMG 492, 2-IPM-02-3,3-IPU-94-1, 4-RC 101, 5-PUSA KOMAL, 6-C152-UAS, 7-GOMATI, 8-YARD LONG, 9-YARD LONG, 10-RMG 344 (genotypes 1,2,10 - *V-radiata*; 3-*V-mungo*; 4 to 9 -*V. unguiculata)*

Fig. 4 Polymorphism and transferability of cpSSRs validated through wet-lab experiments in *Vigna* species.

relationship and evolutionary research among species. A gene product might be associated or located with one or more cellular components such that it is active in one or more biological processes and performs one or more molecular functions. Therefore, some transcripts were annotated simultaneously with the three categories of gene ontology.

Approximately, 1850 genes encoding transcription factors (TFs) were identified by Pfam annotation in the *V. radiata* genome (Kang et al., 2014). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was used to analyze the gene annotation in rice bean (*V. umbellata* L.). Approximately, 9301 unigenes were annotated by the KEGG pathway database (Chen et al., 2016).

The transferability of seven cpSSRs generated from *Vigna* species was analyzed through *in-silico* and wet-lab technique. The 100% transferability observed in this study is in contrast with other studies where some SSRs could be transferable between species (Chagne et al., 2004; Mnejja, Garcia-Mas, Howad, & Arus, 2005). It has been previously reported that the cross transferability decrease with the increasing phylogenetic/evolutionary distances (Brondani, Rangel, Borba, & Brondani, 2003; Castillo et al., 2008; Chapman et al., 2009; Yodav, Mitchell, Fulton, & Kresovich, 2008).

Conclusion

This study revealed the abundance and distribution of cpSSRs in the chloroplast genomes of genus *Vigna* retrieved from public database. These SSRs can be used as a novel tool in genetic diversity and also helps in identifying genes

possessing SSRs in organelle genome and their relationship. It is an easy and efficient way to predict the repetitive region in genome and also minimize the cost and time of experiment. The transferability and polymorphism analysis of cpSSRs indicated the value of developed markers.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors are thankful to the Indian Council of Agricultural Research (ICAR), India for providing financial support. The authors are grateful to the Centre for Bioinformatics facility at Department of Biosciences and Biotechnology, Banasthali University, India. We are also thankful to the Professor Aditya Shastri, Vice Chancellor and Professor Vinay Sharma for providing their full support during the course of this study. The authors are also thankful to the concerned crop breeder, All India Co-ordinated Research Project, Indian Council of Agricultural Research for providing high quality *Vigna* seeds.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped blast and PSI-blast: A new generation of protein database search programs. *Nucleic Acids Research*, *25*, 3389–3402.

Brondani, C., Rangel, P. H. N., Borba, T. C. O., & Brondani, R. P. V. (2003). Transferability of microsatellite and sequence tagged site markers in *Oryza* species. *Hereditas*, *138*, 187–192.

Cao, Y., Wang, L., Xu, K., Kou, C., Zhang, Y., Wei, G., et al. (2005). Information theory-based algorithm for *in silico* prediction of PCR products with whole genomic sequences as templates. *BMC Bioinformatics*, *6*, 190.

Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D., & Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, *156*, 847–854.

Castillo, A., Budak, H., Varshney, R. K., Dorado, G., Graner, A., & Hernandez, P. (2008). Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. *BMC Plant Biology*, *8*, 97.

Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*, *11*, 1–18.

Chagne, D., Chaumeil, P., Ramboer, A., Collada, C., Guevara, M. A., Cervera, M. T., et al. (2004). Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics*, *109*, 1204–1214.

Chankaew, S., Somta, P., Sorajjapinun, W., & Srinives, P. (2011). Quantitative trait loci mapping of cercospora leaf spot resistance in mung bean, *Vigna radiata* (L.) Wilczek. *Molecular Breeding*, *28*, 255–264.

Chapman, M. A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Michelmore, R. W., et al. (2009). Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theoretical and Applied Genetics*, *120*, 85–91.

Chen, H., Wang, L., Wang, S., Liu, C., Blair, M. W., & Cheng, X. (2015). Transcriptome sequencing of mung bean (*Vigna radiata* L.) genes and the identification of EST-SSR markers. *PLoS ONE*, *10*, 1–15.

Chen, H., Chen, X., Tian, J., Yang, Y., Liu, Z., Hao, X., et al. (2016). Development of gene-based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data. *PLoS ONE*, *11*, 1–13. http://dx.doi.org/10.1371/journal.pone.0151040

Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., Trejo-Saavedra, D. L., et al. (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in capsicum. *Scientific Reports*, *6*, 18919.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676.

Decroocq, V., Favé, M. G., Hagen, L., Bordenave, L., & Decroocq, S. (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theoretical and Applied Genetics*, *106*, 912–922.

Diehl, A. D., Lee, J. A., Scheuermann, R. H., & Blake, J. A. (2007). Ontology development for biological systems: Immunology. *Bioinformatics*, *23*, 913–915.

Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13–15.

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*, 435–445.

Gasic, K., Han, Y., Kerbundit, S., Shulaev, V., Lezzoni, A. F., Stover, E. W., et al. (2009). Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. *Molecular Breeding*, *23*, 397–411.

Han, O. K., Kaga, A., Isemura, T., Wang, X. W., Tomooka, N., & Vaughan, D. A. (2005). A genetic linkage map for azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theoretical and Applied Genetics*, *111*, 1278–1287.

Huang, H., Shi, C., Liu, Y., Mao, S., & Gao, L. (2014). Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutioary Biology*, *14*, 1–17.

Jansen, R. K., Raubeson, L. A., Boore, J. L., dePamphilis, C. W., Chumley, T. W., Haberle, R. C., et al. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology*, *395*, 348–384.

Kabra, R., Kapil, A., Attarwala, K., Rai, P. K., & Shanker, A. (2016). Identification of common, unique and polymorphic microsatellites among 73 cyanobacterial genomes. *World Journal of Microbiology and Biotechnology*, *32*, 71.

Kaila, T., Chaduvla, P. K., Saxena, S., Bahadur, K., Gahukar, S. J., Chaudhury, A., et al. (2016). Chloroplast genome sequence of pigeon pea (*Cajanus cajan* (L.) Millspaugh) and *Cajanus scarabaeoides* (L.) Thouars: Genome organization and comparison with other legumes. *Frontiers in Plant Science*, *7*, 1–16.

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. *Nature Communications*, *5*, 1–9. http://dx.doi.org/10.1038/ncomms6443

Kapil, A., Rai, P. K., & Shanker, A. (2014). ChloroSSRdb: A repository of perfect and imperfect chloroplastic simple sequence repeats (cpSSRs) of green plants. *Database (Oxford)*, http://dx.doi.org/10.1093/database/bau107

Kumar, M., Kapil, A., & Shanker, A. (2014). MitoSatPlant: Mitochondrial microsatellites database of viridiplantae. *Mitochondrion*, *19*, 334–337.

Lawson, M. J., & Zhang, L. (2006). Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology*, *7*. R14.

Li, C. D., Fatokun, C. A., Ubi, B., Singh, B. B., & Scoles, G. J. (2001). Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. *Crop Science*, *41*, 189–197.

Liu, S. R., Li, W. Y., Long, D., Hu, C. G., & Zhang, J. Z. (2013). Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis. *PLoS ONE*, *8*, 1–10.

Mariotti, R., Cultrera, N. G. M., Díez, C. M., Baldoni, L., & Rubini, A. (2010). Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biology*, *10*, 211.

Mnejja, M., Garcia-Mas, J., Howad, W., & Arus, P. (2005). Development and transportability across *Prunus* species of 42 polymorphic almond microsatellites. *Molecular Ecology Notes*, *5*, 531–535.

Moe, K. T., Chung, J. W., Cho, Y. I., Moon, J. K., Ku, J. H., Jung, J. K., et al. (2011). Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mung bean. *Journal of Integrative Plant Biology*, *53*, 63–73.

Morgante, M., & Olivieri, A. M. (1993). PCR-Amplified microsatellites as markers in plant genetics. *The Plant Journal*, *3*, 175–182.

Mun, J. H., Kim, D. J., Choi, H. K., Gish, J., Debelle, F., Mudge, J., et al. (2006). Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics*, *172*, 2541–2555.

Oliveira, E. J., Padua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, *29*, 294–307.

Olmstead, R., & Palmer, J. (1994). Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany*, *81*, 1205–1224.

Ozyigit, I. I., Dogan, I., & Filiz, E. (2015). *In-silico* analysis of simple sequence repeats (SSRs) in chloroplast genomes of *Glycine* species. *Plant Omics Journal*, *8*, 24–29.

Powell, W., Machray, G., & Provan, J. (1996). Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, *1*, 215–222.

Provan, J., Powell, W., & Hollingsworth, P. M. (2001). Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution*, *16*, 142–147.

Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS ONE*, *8*, 1–12.

Qin, Z., Wang, Y., Wang, Q., Li, A., Hou, F., & Zhang, L. (2015). Evolution analysis of simple sequence repeats in plant genome. *PLoS ONE*, *10*, 1–18.

Rajendrakumar, P., Biswal, A. K., Balachandran, S. M., Srinivasarao, K., & Sundaram, R. M. (2007). Simple sequence repeats in organellar genomes of rice: Frequency and distribution in genic and intergenic regions. *Bioinformatics*, *23*, 1–4.

Sangiri, C., Kaga, A., Tomooka, N., Vaughan, D., & Srinives, P. (2007). Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) gene pool on the basis of microsatellite analysis. *Australian Journal of Botany*, *55*, 837–847.

Shanker, A., Bhargava, A., Bajpai, R., Singh, S., Srivastava, S., & Sharma, V. (2007). Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis*. *Scientia Horticulturae*, *113*, 353–361.

Shanker, A., Sharma, V., & Daniell, H. (2011). Phylogenomic evidence of bryophytes' monophyly using complete and incomplete data sets from chloroplast proteomes. *Journal of Plant Biochemistry and Biotechnology*, *20*, 288–292.

Shanker, A. (2013). Paraphyly of bryophytes inferred using chloroplast sequences. *Archive for Bryology*, *163*, 1–5.

Shanker, A. (2013). Inference of bryophytes paraphyly using mitochondrial genomes. *Archive for Bryology*, *165*, 1–5.

Shanker, A. (2013). Combined data from chloroplast and mitochondrial genome sequences showed paraphyly of bryophytes. *Archive for Bryology*, *171*, 1–9.

Shanker, A. (2013). Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. *Archive for Bryology*, *191*, 1–6.

Sharma, S. N., Kumar, V., Singh, G., & Sharma, R. (2011). A modified DNA extraction protocol and its utility in seed genetic purity assessment. *Seed Science and Technology*, *39*, 236–242.

Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y. J., & Gao, L. Z. (2012). An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS ONE*, *7*, 1–7.

Singh, A., Sharma, A. K., Singh, N. K., & Sharma, T. R. (2017). PpTFDB: A pigeonpea transcription factor database for exploring functional genomics in legumes. *PLoS ONE*, *12*, 1–8. http://dx.doi.org/10.1371/journal.pone.0179736

Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., et al. (2011). Genome-Wide distribution and organization of microsatellites in plants: An insight into marker development in Brachypodium. *PLoS ONE*, *6*, 1–9.

Srivastava, D., & Shanker, A. (2015). Mining of simple sequence repeats in chloroplast genome sequence of *Cocos nucifera*. *Applied Research Journal*, *1*, 51–54.

Tambarussi, E. V., Melotto-Passarin, D. M., Gonzalez, S. G., Brigati, J. B., de Jesus, F. A., Barbosa, A. L., et al. (2009). *In-silico* analysis of simple sequence repeats from chloroplast genomes of Solanaceae species. *Crop Breeding and Applied Biotechnology*, *9*, 344–352.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, *17*, 6463–6471.

Toth, G., Gaspari, Z., & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research*, *10*, 967–981.

Varshney, R. K., Thiel, T., Stein, N., Langridge, P., & Graner, A. (2002). *In-silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters*, *7*, 537–546.

Wang, X. W., Kaga, A., Tomooka, N., & Vaughan, D. A. (2004). The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi and Ohashi]. *Theoretical and Applied Genetics*, *109*, 352–360.

Wang, H., Walla, J. A., Zhong, S., Huang, D., & Dai, W. (2012). Development and cross-species/genera transferability of microsatellite markers discovered using 454 genome sequencing in choke cherry (*Prunus virginiana* L.). *Plant Cell Reports*, *31*, 2047–2055.

Wheeler, G. L., Dorman, H. E., Buchanan, A., Challagundla, L., & Wallace, L. E. (2014). A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Applications in Plant Sciences*, *2*, 1400059.

Yi, D.-K., & Kim, K.-J. (2012). Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. *PLoS ONE*, *7*, e35872.

Yodav, O. P., Mitchell, S. E., Fulton, T. M., & Kresovich, S. (2008). Transferring molecular markers from sorghum, rice and other cereals to pearl millet and identifying polymorphic markers. *Journal of Semi-Arid Tropical Agricultural Research*, *6*, 1–4.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: A review. *Molecular Ecology*, *11*, 1–16.

Zapiola, M. L., Cronn, R. C., & Mallory-Smith, C. A. (2010). Development of novel chloroplast microsatellite markers to identify species in the Agrostis complex (Poaceae) and related genera. *Molecular Ecology Resources*, *10*, 738–740.

Zou, C., Lu, C., Zhang, Y., & Song, G. (2012). Distribution and characterization of simple sequence repeats in *Gossypium raimondii* genome. *Bioinformation*, *8*, 801–806.