



ISSN: 2348-1900

Plant Science Today

<http://horizonepublishing.com/journals/index.php/PST>



Research Communication

Detection of simple sequence repeats in the chloroplast genome of *Tetraphis pellucida* Hedw.

Asheesh Shanker

Bioinformatics Programme, Center for Biological Sciences, Central University of South Bihar, India

Article history

Received: 08 March 2016
Accepted: 27 March 2016
Published: 13 June 2016

© Shanker (2016)

Special Section: New Frontiers
in Cryptogamic Botany

Section Editor

Afroz Alam

Publisher

Horizon e-Publishing Group

Corresponding Author

Asheesh Shanker
✉ ashomics@gmail.com

Abstract

Simple sequence repeats (SSRs) consist of short repeat motifs of 1-6 nucleotides and are found in DNA sequences. The present study was conducted to detect SSRs in chloroplast genome of *Tetraphis pellucida* (Accession number: NC_024291), downloaded from the National Center for Biotechnology Information (NCBI). The sequence was mined with the help of MISA, a Perl script, to detect SSRs. The length of SSRs defined as ≥ 12 for mono, di, tri and tetranucleotide, ≥ 15 for pentanucleotide and ≥ 18 for hexanucleotide repeats. In total, 41 perfect microsatellites were identified in 127.489 kb sequence mined. An average length of 13.56 bp was calculated for mined SSRs with a density of 1 SSR/3.04 kb. Depending on the repeat units, the length of SSRs ranged from 12 to 20 nt. Dinucleotides (14, 34.15%) were the most frequent repeat type, followed by tetranucleotides (10, 24.39%), trinucleotides (7, 17.07%), mononucleotides (6, 14.63%) and pentanucleotide (4, 9.76%) repeats. Hexanucleotide repeats were completely absent in chloroplast genome of *Tetraphis pellucida*. The mined SSRs can be used to develop molecular markers and genetic diversity studies in *Tetraphis* species.

Keywords

Bryophytes; chloroplast genome; microsatellites; moss; *Tetraphis*

Shanker, A. 2016. Detection of simple sequence repeats in the chloroplast genome of *Tetraphis pellucida* Hedw. *Plant Science Today* 3(2): 207-210. <http://dx.doi.org/10.14719/pst.2016.3.2.206>

Introduction

Bryophytes are the earliest and the simplest land plants with extant lineages of hornworts, liverworts, and mosses. Paraphyletic origin of bryophytes has been suggested by chloroplast and mitochondrial genome sequences based phylogenetic analysis (Shanker, 2013a; 2013b; 2013c). Chloroplasts are present in green plants as an intracellular organelles which contain their own autonomously replicating genome. In the recent past organelle genome sequences have been used for various purposes, including the

mining of simple sequence repeats (Kapil *et al.*, 2014; Kumar *et al.*, 2014).

Simple sequence repeats (SSRs) also known as microsatellites consist of short repeat motifs (1-6 nucleotides) and are found in both coding and non-coding regions of DNA sequences (Shanker *et al.*, 2007). These repeats act as molecular markers in many plant genomes and also identified in bryophytes (Shanker, 2014a; 2014b; 2014c; Pandey *et al.*, 2016). However, the distribution of SSRs in chloroplast genome of *Tetraphis pellucida* Hedw. (Bryophyta), a four-toothed moss is not well studied.

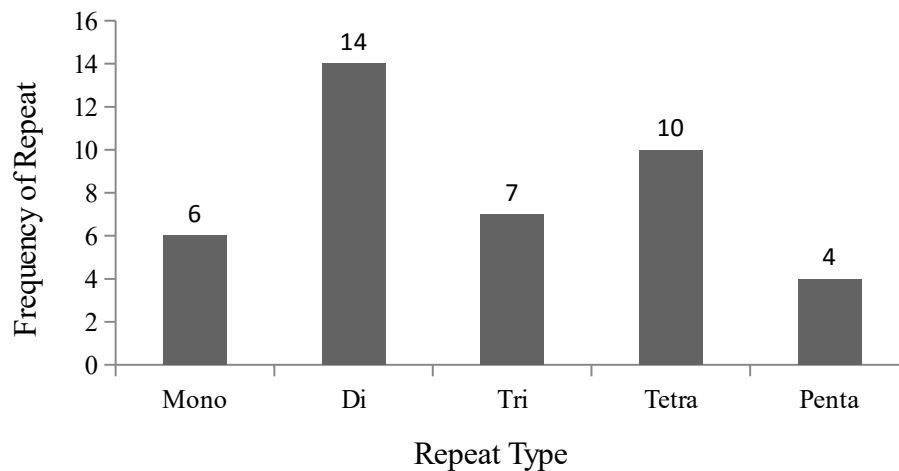


Figure 1. Frequency distribution of various repeat types

Materials and Methods

Tetraphis pellucida: Chloroplast genome sequence

The organelle genome resources at National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) contain only a handful of chloroplast genome sequences of bryophytes (Shanker, 2012; Shanker, 2015). Recently, Bell *et al.* (2014) published the chloroplast genome sequence of *Tetraphis pellucida*. It was downloaded from NCBI in FASTA and GenBank format (NC_024291, 127489bp).

Mining of simple sequence repeats

The chloroplast genome sequence of *Tetraphis pellucida* was mined using MISA, a Perl script (available at <http://pgrc.ipk-gatersleben.de/misa/misa>). SSRs length was considered as ≥ 12 for mono, di, tri and tetranucleotide, ≥ 15 for pentanucleotide and ≥ 18 for hexanucleotide repeats. The mined SSRs were classified into coding and non-coding SSRs using the information available in GenBank file of the chloroplast genome.

Results and Discussion

In this study perfect chloroplastic microsatellites or SSRs (cpSSRs) with a minimum length of 12 bp were identified in chloroplast genome sequence of *T. pellucida*. The length of the identified SSRs ranged from 12 to 20nt. Hexanucleotide repeats were not detected in chloroplast genome sequence of *T. pellucida*. The distribution of mined SSRs is presented in Fig. 1.

The chloroplast genome sequence of *T. pellucida* contains only 41 SSRs showing density of 1 SSR/3.04 kb in 127.489 kb sequence mined. Dinucleotides (14, 34.15%) were the most frequent repeat type, followed by tetranucleotides (10, 24.39%), trinucleotides (7, 17.07%),

mononucleotides (6, 14.63%), and pentanucleotide (4, 9.76%) repeats. Out of all mined SSRs only 12 (29.27%) lie in coding and 29 (70.73%) lie in non-coding region of the genome. Mined SSRs motif, their start-end position, the region in which they lie, and their length is presented in Table 1.

The density of cpSSRs in *T. pellucida* (1 SSR/3.04 kb) found to be lower than the density of cpSSRs in *Anthoceros formosae* (1 SSR/2.4 kb; Shanker, 2013d) and *Marchantia polymorpha* (1 SSR/1.83 kb; Shanker, 2014c), however, higher than the density of cpSSRs in *Aneura mirabilis* (1 SSR/5.68 kb; Shanker 2013e), *Pellia endiviifolia* (1 SSR/7.09 kb; Shanker, 2014b), and *Ptilidium pulcherrimum* (1 SSR/5.17 kb; Shanker, 2014a). Moreover, it is higher than the density of EST-SSRs in barley, maize, wheat, rye, sorghum and rice (1 SSR/6.0 kb; Varshney *et al.* 2002), cpSSRs of rice (1SSR/6.5 kb; Rajendrakumar *et al.*, 2007), Unigene sequences of *Citrus* (1 SSR/12.9 kb; Shanker *et al.*, 2007a), average density of chloroplast genomes of Rosales (1 SSR/5.12 kb; Srivastava and Shanker, 2015a) and Magnoliids (1 SSR/6.91 kb; Srivastava and Shanker, 2015b). The variation in SSR density is due to different mining parameters including minimum length of SSRs taken, the amount and type of data analyzed, and genomic composition of the sequence.

Similar to the cpSSRs in chloroplast genomes of other bryophytes (Shanker 2013d; 2013e; Shanker 2014b; 2014c) most of the cpSSRs identified in *T. pellucida* also lie in non-coding region. These cpSSRs can be useful to develop SSR markers and for other purposes. The exclusive genomic and morphological characteristics of *Tetraphis* authenticate its significance for deciding one of the key issues in the phylogeny of land plant and for understanding the evolution of the some special structures. Hence, this study provides a scientific base for phylogenetics, evolutionary genetics and diversity studies on different *Tetraphis* species.

Table 1: Information of mined SSRs in chloroplast genome sequence of *Tetraphis pellucida*

S. No.	Motif	Length	Start	End	Region
1.	(A)12	12	600	611	Coding
2.	(AT)6	12	8950	8961	Non-Coding
3.	(ATAAA)3	15	10430	10444	Non-Coding
4.	(TA)8	16	11397	11412	Non-Coding
5.	(AT)7	14	13254	13267	Non-Coding
6.	(ATA)4	12	16654	16665	Non-Coding
7.	(ATA)4	12	17241	17252	Non-Coding
8.	(AT)7	14	17688	17701	Non-Coding
9.	(TTTA)3	12	19034	19045	Non-Coding
10.	(A)12	12	20787	20798	Non-Coding
11.	(AT)6	12	22483	22494	Non-Coding
12.	(AT)6	12	24843	24854	Non-Coding
13.	(AT)7	14	28580	28593	Non-Coding
14.	(AT)8	16	28743	28758	Non-Coding
15.	(ATT)4	12	33607	33618	Non-Coding
16.	(TAT)4	12	35122	35133	Non-Coding
17.	(A)14	14	37037	37050	Non-Coding
18.	(TA)6	12	43992	44003	Non-Coding
19.	(ATA)4	12	46896	46907	Coding
20.	(CTTTT)3	15	47417	47431	Non-Coding
21.	(AATTT)3	15	47561	47575	Non-Coding
22.	(TTTA)3	12	49244	49255	Non-Coding
23.	(TAAA)3	12	51939	51950	Coding
24.	(T)12	12	52980	52991	Coding
25.	(TTA)4	12	55298	55309	Coding
26.	(TA)6	12	57964	57975	Non-Coding
27.	(AT)6	12	64213	64224	Non-Coding
28.	(TTTA)3	12	64467	64478	Non-Coding
29.	(TTAA)3	12	70006	70017	Non-Coding
30.	(AT)10	20	78884	78903	Non-Coding
31.	(ATAA)3	12	81351	81362	Coding
32.	(T)12	12	82128	82139	Coding
33.	(AAATA)3	15	87438	87452	Coding
34.	(AT)8	16	88439	88454	Non-Coding
35.	(TAAA)3	12	89252	89263	Non-Coding
36.	(AGGT)3	12	95994	96005	Coding
37.	(A)12	12	100547	100558	Coding
38.	(AT)6	12	111943	111954	Non-Coding
39.	(ATA)4	12	112000	112011	Non-Coding
40.	(TATT)3	12	113704	113715	Coding
41.	(CTAC)3	12	120917	120928	Coding

Acknowledgements

I am thankful to Dr. Afroz Alam for providing the opportunity to contribute in the special section on "New Frontiers in Cryptogamic Botany" of Plant Science Today.

References

- Bell N. E., J. L. Boore, B. D. Mishler, and Hyvonen J. 2014. Organellar genomes of the four-toothed moss, *Tetraphis pellucida*. *BMC Genomics* 15: 383. doi: 10.1186/1471-2164-15-383
- Kapil A., P. K. Rai and A. Shanker. 2014. ChloroSSRdb: a repository of perfect and imperfect chloroplastic

- simple sequence repeats (cpSSRs) of green plants. Database (Oxford) 2014 doi: 10.1093/database/bau107.
- Kumar M., A. Kapil and A. Shanker. 2014. MitoSatPlant: mitochondrial microsatellites database of viridi plantae. *Mitochondrion* 19: 334-337. doi: 10.1016/j.mito.2014.02.002
- Rajendrakumar P., A. K. Biswal, S. M. Balachandran Srinivasarao K, and R. M. Sundaram. 2007. Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* 23: 1-4.
- Pandey S., V. Sharma and A. Alam. 2016. Potential of microsatellites markers for the genetic analysis of bryophytes. *Not Sci Biol* 8(1): 37-46. doi: 10.15835/nsb.8.1.9748
- Shanker A. 2012. Chloroplast genomes of bryophytes: A review. *Arch Bryol* 143: 1-5.
- Shanker A. 2013a. Paraphyly of bryophytes inferred using chloroplast sequences. *Arch Bryol* 163: 1-5.
- Shanker A. 2013b. Inference of bryophytes paraphyly using mitochondrial genomes. *Arch Bryol* 165: 1-5.
- Shanker A. 2013c. Combined data from chloroplast and mitochondrial genome sequences showed paraphyly of bryophytes. *Arch Bryol* 171: 1-9.
- Shanker A. 2013d. Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. *Arch Bryol* 191: 1-6.
- Shanker A. 2013e. Mining of simple sequence repeats in chloroplast genome of a parasitic liverwort: *Aneura mirabilis*. *Arch Bryol* 196: 1-4.
- Shanker A. 2014a. Computational mining of microsatellites in the chloroplast genome of *Ptilidium pulcherrimum*, a liverwort. *Intl Jour Environ* 3: 50-58. doi: 10.3126/ije.v3i3.11063
- Shanker A. 2014b. Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. *Arch Bryol* 199: 1-5.
- Shanker A. 2014c. Simple sequence repeats mining using computational approach in chloroplast genome of *Marchantia polymorpha*. *Arctoa* 23, 145-149.
- Shanker A. 2015. An update on sequenced chloroplast genomes of Bryophytes. *Plant Sci Today* 2, 102-104. doi: 10.14719/pst.2015.2.4.143
- Shanker A., A. Bhargava, R. Bajpai, S. Singh, S. Srivastava, and V. Sharma. 2007a. Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis*. *Sci Hort* 113, 353-361. doi: 10.1016/j.scienta.2007.04.011
- Shanker A., A. Singh and V. Sharma. 2007. *In silico* mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. *Microbiol Res* 162, 250- 256.
- Srivastava D. and A. Shanker. 2015a. *In silico* mining and analysis of simple sequence repeats in chloroplast genomes of order Rosales. *Res Jour Bioinformatics* 2: 1-9.
- Srivastava D. and A. Shanker. 2015b. Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach. *Interdiscip Sci Comput Life Sci*. doi 10.1007/s12539-015-0129-4.
- Varshney R. K., T. Thiel, N. Stein, P. Langridge and A. Graner. 2002. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell & Mol Biol Lett* 7: 537-546.

