

In Silico Mining of Simple Sequence Repeats in Mitochondrial Genomes of Genus *Orthotrichum*

Sonu Kumar, Swati Kumari, and Asheesh Shanker*

Department of Bioinformatics, Central University of South Bihar, Gaya-824236, India.
ashomics@gmail.com*

Abstract: Simple sequence repeats (SSRs) or microsatellites play significant role in the study of genome evolution due to their high mutation rate. Bryophytes hold basal most position in the phylogeny of land plants and play significant role in diverse terrestrial ecosystem. In the present study, 7 mitochondrial genome sequences of genus *Orthotrichum* (*O. stellatum*, *O. macrocephalum*, *O. diaphanum*, *O. bicolor*, *O. callistomum*, *O. gymnostomum*, and *O. obtusifolium*) were computationally mined to detect SSRs. A total of 290 perfect SSRs were detected with an average density of 1SSR/2.53 kb sequence. Di-nucleotide (138; 47.59%) repeats were most abundant followed by tetra- (70; 24.14%), mono- (46; 15.86%), penta- (19; 6.55%), and tri-nucleotide (17; 5.86%) repeats. Hexa-nucleotide repeats were totally absent in studied mitochondrial genomes of genus *Orthotrichum*. Additionally, length variation of identified SSRs was also observed. These mined SSRs may be utilized for a variety of purposes including gene tagging, genetic mapping, and population genetics studies.

Index Terms: Bryophytes, Length variation, Mitochondria, Simple Sequence Repeats, *Orthotrichum*.

I. INTRODUCTION

Bryophytes have special characteristics to provide the information on the classification, ecology, and evolutionary relationship of plants. These are classified into liverworts, mosses, and hornworts, hold basal most position in the phylogeny of land plants, and play significant role in diverse terrestrial ecosystem (Shanker, 2013a-c). Complete chloroplast and mitochondrial genome (cp/mt-genome) sequences of bryophytes are available at National Center for Biotechnology Information (NCBI) database (Shanker, 2012a-e). These sequences can be used for various purposes and *in silico*

identification of simple sequence repeats is one of them (Shanker, 2014a-c & 2015; Kumar & Shanker, 2020a).

Simple sequence repeats (SSRs) also known as microsatellites or short tandem repeats (STRs) are tandem repetitions of short nucleotide motifs of length 1-6. SSRs play significant role in the study of genome evolution due to their high mutation rate. Characterization of SSRs is the main component of today's molecular marker development research (Shanker et al., 2007; Kapil et al., 2018) and polymorphic SSRs help to understand the genetic diversity within species or different species (Grassi et al., 2002). Considering their importance, SSR specific databases were developed (Kapil et al., 2014; Kumar et al., 2014; Kabra et al., 2016). Apart from this, a methodology was created to detect length variation in SSRs using computational approach (Kabra et al., 2016). Recently, Kumar & Shanker (2020b) demonstrated the significance of computational resources in the study of bryophytes.

Earlier, Zhao et al. (2014) detected mitochondrial SSRs (mtSSRs) in two liverworts (*Marchantia polymorpha* and *Pleurozia purpurea*), two mosses (*Physcomitrella patens* and *Anomodon rugelii*), and two hornworts (*Phaeoceros laevis* and *Nothoceros aenigmaticus*). Similarly, mtSSRs were detected in 4 species of order Hypnales (Bryopsida; Khushbu et al., 2019), *Aneura pinguis* (Kumar & Shanker, 2020a), and 6 species of genus *Nymphaea* (Kumar & Shanker, 2020c). Moreover, perfect SSRs were also identified in chloroplast genome of *Anthoceros formosae* (Shanker, 2013b), *Aneura mirabilis* (Shanker, 2013d), *Ptilidium pulcherrimum* (Shanker, 2014a), *Pellia endiviifolia* (Shanker, 2014c), and *Nothoceros aenigmaticus* (Shanker, 2015).

Despite these efforts, information of mtSSRs and their length variation in genus *Orthotrichum* has not been explored. Therefore, the present study was designed to mine mtSSRs and

* Corresponding Author

detect length polymorphism in complete mitochondrial genomes sequence of genus *Orthotrichum*.

II. MATERIAL AND METHODS

A. Mitochondrial genome sequences of genus *Orthotrichum*

Complete mitochondrial genome sequences of genus *Orthotrichum* were retrieved in FASTA and GenBank file format from National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov). A list of retrieved mitochondrial genome sequences is presented in table 1.

B. Mining of mtSSRs among genus *Orthotrichum*

To mine perfect mtSSRs among retrieved genome sequences of genus *Orthotrichum*, microsatellite identification tool (MISA; <http://pgrc.ipk-gatersleben.de/misa/misa>) which is based on PERL programming language was used. MISA takes nucleotide sequence data in FASTA format as an input and generates information of mined SSRs, if detected, along with statistical data in two separate files. The minimum length to mine mtSSRs was considered as ≥ 12 for mono to tetra, ≥ 15 for penta, and ≥ 18 for hexa-nucleotide repeats. Zero interruption was considered between two SSRs.

C. Detection of common, polymorphic and unique mtSSRs among genus *Orthotrichum*

Identical repeat motif with equal and varying length, having significant similarity with flanking regions (200 bases from both upstream and downstream of SSRs) across the genomes were considered as common and polymorphic SSRs, respectively. Moreover, identical repeat motif with no significant match of

flanking regions across the species and uniquely identified repeating units were considered as unique SSRs (Kumar & Shanker, 2018). Basic Local Alignment Search Tool (BLAST; Altschul et al., 1997) was used for similarity search.

III. RESULT AND DISCUSSION

A. Frequency and distribution of mined mtSSRs

In this study, 290 perfect mtSSRs were detected in 7 mitochondrial genome sequences of genus *Orthotrichum*, with an average density of 1SSR/2.53 kb sequence mined. The length of the identified mtSSRs ranged from 12 to 20 nucleotides. The average density of mtSSRs identified in the present study is higher than the average density of mtSSRs detected in order *Hypnales* (1SSR/4.21 kb; Khushbu et al., 2019) and genus *Aneura pinguis* (1SSR/6.36 kb; Kumar & Shanker, 2020a), while lower than *Physcomitrella patens* (1 SSR/2.06 kb; Kuntal & Sharma, 2011).

Apart from this, the average density of mtSSRs identified in this study is higher than SSRs identified in chloroplast genome of *Aneura mirabilis* (1SSR/5.68 kb; Shanker, 2013d), *Pellia endiviifolia* (1SSR/7.09 kb; Shanker, 2014c), *Ptilidium pulcherrimum* (1SSR/5.17 kb; Shanker, 2014a), and *Tetraphis pellucida* (1SSR/3.04 kb; Shanker, 2016), whereas lower than *Anthoceros formosae* (1SSR/2.4 kb; Shanker, 2013b) and *Marchantia polymorpha* (1SSR/1.83 kb; Shanker, 2014b). The variation in the density of identified mtSSRs might be due to nucleotide composition of sequences mined.

Table 1. List of mitochondrial genomes of genus *Orthotrichum* retrieved from NCBI

S. No.	Organism	Abbreviation	Accession	Size (Kb)	References
1.	<i>O. stellatum</i>	<i>Ost</i>	NC_024522.1	104.131	Liu et al., 2014
2.	<i>O. macrocephalum</i>	<i>Oma</i>	NC_029355.1	104.624	Vigalondo et al., 2016
3.	<i>O. diaphanum</i>	<i>Odi</i>	NC_029356.1	104.744	Vigalondo et al., 2016
4.	<i>O. bicolor</i>	<i>Obi</i>	NC_031389.1	104.736	Unpublished
5.	<i>O. callistomum</i>	<i>Oca</i>	NC_031390.1	104.785	Unpublished
6.	<i>O. gymnostomum</i>	<i>Ogy</i>	NC_031391.1	104.638	Unpublished
7.	<i>O. obtusifolium</i>	<i>Oob</i>	NC_031767.1	104.603	Unpublished

Table 2. Frequency of mono-hexa repeats mined

S. No.	Organism	Mono	Di	Tri	Tetra	Penta	Total	Density
1.	<i>O. stellatum</i>	7	19	2	10	1	39	2.67
2.	<i>O. macrocephalum</i>	6	20	2	10	3	41	2.55
3.	<i>O. diaphanum</i>	7	20	2	10	3	42	2.49
4.	<i>O. bicolor</i>	5	19	2	10	3	39	2.69
5.	<i>O. callistomum</i>	6	20	3	10	3	42	2.49
6.	<i>O. gymnostomum</i>	7	20	3	10	3	43	2.43
7.	<i>O. obtusifolium</i>	8	20	3	10	3	44	2.38
Total		46	138	17	70	19	290	

The present study showed that, di-nucleotide (138; 47.59%) motifs were most abundant followed by tetra- (70; 24.14%), mono- (46; 15.86%), penta- (19; 6.55%), and tri- (17; 5.86%) nucleotides. Whereas, hexa-nucleotide repeats were totally absent in mined mitochondrial genomes of genus *Orthotrichum*. Frequency and distribution of mtSSRs identified in genus *Orthotrichum* are presented in table 2.

Additionally, distribution of SSRs was also observed in coding, non-coding, and coding-non-coding regions of the mitochondrial genome. SSRs in non-coding (235; 81.03%) regions were most frequent followed by coding (48; 16.55%) and coding-non-coding (7; 2.41%) regions in mitochondrial genomes of genus *Orthotrichum* (Figure 1).

B. Length variation of mtSSRs among genus *Orthotrichum*

SSRs were categorized as common, polymorphic, and unique based on length variation of mined mtSSRs. The total

number of common and polymorphic mtSSRs ranged from 36 to 47 and 4 to 10, respectively, between each pair of species. Additionally, 4 unique mtSSRs were also detected among genus *Orthotrichum*. Some identical motif shows significant similarity with more than one loci over target genome sequence. Motif T, AT, CATA, ATATA, and TATAT frequently showed length polymorphism among genus *Orthotrichum*. Total numbers of common, polymorphic, and unique mtSSRs identified in genus *Orthotrichum* are presented in table 3. Previously, length variation of SSRs were observed in organelle genome sequences of 12 species of genus *Arabidopsis* (Kumar & Shanker, 2018), 4 species of order Hypnales (Bryopsida; Khushbu et al. 2019), and 6 species of genus *Nymphaea* (Kumar & Shanker, 2020c).

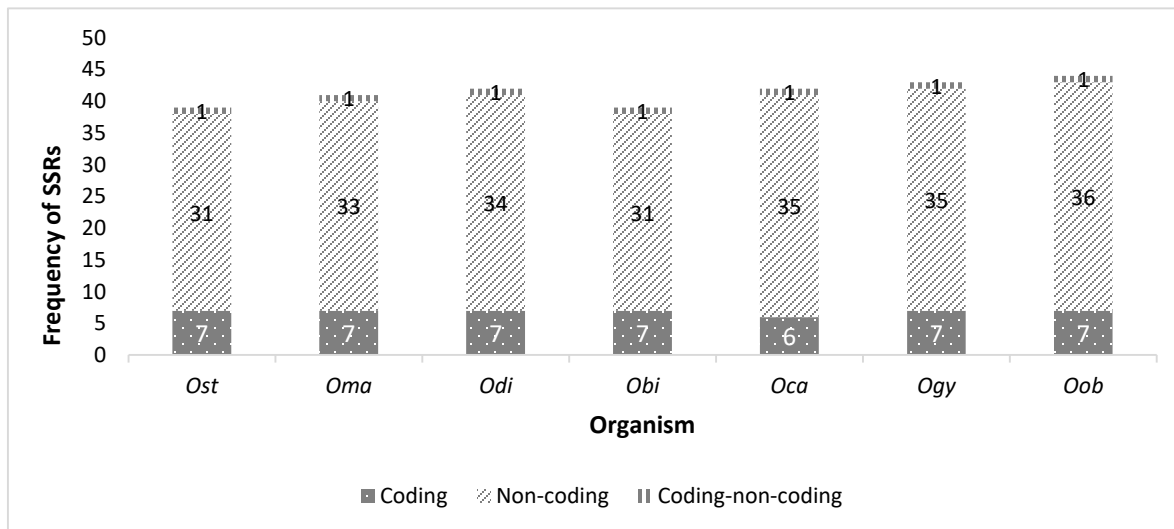


Figure 1. Distribution of mtSSRs in genus *Orthotrichum*. Organism abbreviations are given in table 1

Table 3. Common, **polymorphic** (bold), and unique (*) mtSSRs identified among mitochondrial genomes of genus *Orthotrichum*. Abbreviations of the organisms name are given in table 1

	Unique*	Ost	Oma	Odi	Obi	Oca	Ogy	Oob
Ost	-	-	40	40	37	40	38	36
Oma	-	7	-	47	43	45	38	37
Odi	1	7	4	-	43	45	38	37
Obi	-	8	5	5	-	42	38	36
Oca	2	7	4	4	5	-	38	37
Ogy	-	9	10	10	9	10	-	45
Oob	1	10	10	10	9	10	4	-

CONCLUSION

The mtSSRs were successfully mined which showed length variation in repeats among genus *Orthotrichum*. The identified mtSSRs will be useful for genetic diversity studies, development

of molecular markers, and evolutionary studies to distinguish the genotypes based on mitochondrial genome.

ACKNOWLEDGEMENT

S. K. is thankful to the University Grants Commission (UGC), New Delhi, India for financial support in the form of Rajiv Gandhi National Fellowship.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

REFERENCES

- Altschul, S. F., Madden. T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-3402.
- Anand, K., Kumar, S., Alam, A., & Shankar, A. (2019). Mining of microsatellites in mitochondrial genomes of order Hypnales (Bryopsida). *Plant Science Today*, 6(sp1), 635-638.
- Grassi, F., Labra, M., Scienza, A., & Imazio, S. (2002). Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Research Note*, 41, 157-158.
- 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(4), 71.
- Kapil, A., Jha, C. K., & Shanker, A. (2018). Data mining to detect common, unique, and polymorphic simple sequence repeats. In *Bioinformatics: Sequences, Structures, Phylogeny*, A. Shanker (ed.), Springer Singapore, 141-154.
- Kapil, A., Rai, P. K., & Shanker, A. (2014). ChloroSSRdb: a repository of perfect and imperfect chloroplastic simple sequence repeats (cpSSRs) of green plants. *Database*, 2014.
- Kumar, M., Kapil, A., & Shanker, A. (2014). MitoSatPlant: mitochondrial microsatellites database of viridiplantae. *Mitochondrion*, 19, 334-337.
- Kumar, S., & Shanker, A. (2018). Common, unique and polymorphic simple sequence repeats in chloroplast genomes of genus *Arabidopsis*. *Vegetos-An International Journal of Plant Research*, 31(special), 125-131.
- Kumar, S., & Shanker, A. (2020a). Analysis of Microsatellites in Mitochondrial Genome of *Aneura pinguis* (L.) Dumort. *Contemporary Research on Bryophytes*, 1, 87-94.
- Kumar, S., & Shanker, A. (2020b). Computational Resources for Bryology. *Contemporary Research on Bryophytes*, 1, 20-37.
- Kumar, S., & Shanker, A. (2020c). *In Silico* Comparative Analysis of Simple Sequence Repeats in Chloroplast Genomes of Genus *Nymphaea*. *Journal of Scientific Research*, 64(1).
- Kuntal, H., & Sharma, V. (2011). *In silico* analysis of SSRs in mitochondrial genomes of plants. *Omic: a journal of integrative biology*, 15(11), 783-789.
- Liu, Y., Medina, R., & Goffinet, B. (2014). 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. *Molecular Biology and Evolution*, 31(10), 2586-2591.
- Sablok, G., Padma Raju, G.V., Mudunuri, S.B., Prabha, R., Singh, D.P., Baev, V., Yahubyan, G., Ralph, P.J. and Porta, N.L. (2015). ChloroMitoSSRDB 2.00: more genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database*, 2015.
- Shanker, A. (2012a). Chloroplast genomes of bryophytes: a review. *Archive for Bryology*, 143, 2012.
- Shanker, A. (2012b). Comparison of mitochondrial genomes of bryophytes. *Archive for Bryology*, 142, 2012.
- Shanker, A. (2012c). Sequenced mitochondrial genomes of bryophytes. *Archive for Bryology*, 146, 2012.
- Shanker, A. (2013a). Combined data from chloroplast and mitochondrial genome sequences showed paraphyly of bryophytes. *Archive for Bryology*, 171, 2013.
- Shanker, A. (2013b). Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. *Archive for Bryology*, 191, 2013.
- Shanker, A. (2013c). Inference of bryophytes paraphyly using mitochondrial genomes. *Archive for Bryology*, 165, 2013.
- Shanker, A. (2013d). Mining of simple sequence repeats in chloroplast genome of a parasitic liverwort: *Aneura mirabilis*. *Archive for Bryology*, 196, 2013.
- Shanker, A. (2013e). Paraphyly of bryophytes inferred using chloroplast sequences. *Archive for Bryology*, 163, 2013.
- Shanker, A. (2014a). Computational mining of microsatellites in the chloroplast genome of *Ptilidium pulcherrimum*, a liverwort. *International Journal of Environment*, 3(3), 50-58.
- Shanker, A. (2014b). Simple sequence repeats mining using computational approach in chloroplast genome of *Marchantia polymorpha*. *Arctoa*, 23, 145-149.
- Shanker, A. (2014c). Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. *Archive for Bryology*, 199, 2014.
- Shanker, A. (2015). *In silico* mining of simple sequence repeats in chloroplast genome of *Nothoceros aenigmaticus*, *Researcher*, 7(3), 24-27.
- Shanker, A. (2016). Detection of simple sequence repeats in the chloroplast genome of *Tetraphis pellucida* Hedw. *Plant Science Today*, 3(2), 207-210.
- 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(4), 353-361.
- Vigalondo, B., Liu, Y., Draper, I., Lara, F., Garilleti, R., Mazimpaka, V., & Goffinet, B. (2016). Comparing three complete mitochondrial genomes of the moss genus *Orthotrichum* Hedw. *Mitochondrial DNA Part B*, 1(1), 168-170.
- Zhao, C. X., Zhu, R. L., & Liu, Y. (2016). Simple sequence repeats in bryophyte mitochondrial genomes. *Mitochondrial DNA Part A*, 27(1), 191-197.
