Volume 64, Issue 1, 2020
Journal of Scientific Research
Institute of Science,
Banaras Hindu University, Varanasi, India.

In Silico Comparative Analysis of Simple Sequence Repeats in Chloroplast Genomes of Genus *Nymphaea*

Sonu Kumar[1] and Asheesh Shanker[1*]

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

Abstract: **Simple sequence repeats (SSRs) also known as microsatellites, found in almost all organisms, are one of the widely used genetic markers. Despite the availability of a number of chloroplast genome of genus *Nymphaea* the information about its chloroplast SSR (cpSSR) is not well understood. In the present study, a total of 96 cpSSRs were mined in 6 chloroplast genome of genus *Nymphaea* (*N. alba, N. ampla, N. capensis, N. jamesoniana, N. lotus,* and *N. mexicana*). Mononucleotides (33, 34.38%) were most abundant followed by tri-nucleotides (24, 25%), tetra-nucleotides (19, 19.79%), di-nucleotides (14, 14.58%), whereas penta- and hexa-nucleotides (3, 3.13%) were found with equal frequency among chloroplast genomes of genus *Nymphaea*. Moreover, common, polymorphic, and unique SSRs were also searched between each chloroplast genomes. The identified common, polymorphic, and unique cpSSRs may play an important role in analysis of genetic diversity of genus *Nymphaea*.**

Keywords: **Chloroplast, Microsatellites, *Nymphaea*, Polymorphic, Simple sequence repeat**

I. INTRODUCTION

Simple sequence repeats (SSRs) or microsatellite is a tract of repetitive nucleotide sequence; consist of 1-6 base pair (bp). SSRs are widely used genetic/ molecular marker among the different molecular markers used in molecular genetics. These repeats found in coding and non-coding regions of both prokaryotic and eukaryotic genomes. Moreover, SSRs are also present in organellar genomes including chloroplast and amitochondria (Kapil et al., 2014; Kumar et al., 2014). SSRs have a higher mutation rate than other areas of genomes (Brinkmann et al., 1998) leading to high genetic diversity. These repeats are considered to be produced because of slippage mechanism during Deoxyribonucleic acid (DNA) replication (Tautz & Schlötterer, 1994). SSRs are commonly categorized into perfect, imperfect, and compound SSRs (Bachmann et al., 2004).

Traditional methods to develop SSRs are very expensive, laborious, and time consuming. Briefly, various experiments including DNA extraction, preparation of genomic libraries for targeted SSR, DNA fragmentation, cloning of DNA fragments, sequencing of clones carrying SSRs, PCR for the validation of SSR are involved to develop SSRs in laboratories (Zane et al., 2002; Squirrell et al., 2003; Zalapa et al., 2012; Zhu et al., 2012; Taheri et al., 2018). Resent development in computational resources play important role in many areas of scientific studies (Kumar & Shanker, 2018a, & b). Computational approaches have become a relatively less time consuming and inexpensive to develop SSRs. Previously, using computational approaches SSRs were mined in large amount of sequence data retrieved from National Centre for Biotechnology Information (NCBI) and online databases were developed for obtained information (Kapil et al., 2014; Kumar et al., 2014; Kabra et al., 2016). The availability of nucleotide/genome sequence of various organisms in biological databases proved to be a fast and inexpensive way for computational mining of SSRs (Shanker, 2013, 2014; Shanker, 2016; Kumar & Shanker, 2018c).

Moreover, variation in the chloroplast genome has been an important source of information in plant biology for several decades. Despite the availability of chloroplast genome of genus *Nymphaea* in public database the information about its chloroplast SSRs (cpSSRs) is not well studied. *Nymphaea* is a widespread genus of aquatic plants which belongs to the family Nymphaeaceae. Some species of genus *Nymphaea* has also been used as medicinal plant (Singh & Jain, 2017).

Earlier, SSRs were identified in various plants including *Citrus sinensis* (Shanker et al., 2007), *Gracilaria tenuistipitata* (Song et al., 2014), *Cocos nucifera* (Srivastava & Shanker, 2015), *Glycine species* (Ozyigit et al., 2015), Magnoliids (Srivastava & Shanker, 2016), *Euphorbia esula* (Sen et al., 2017), *Arabidopsis* (Kumar & Shanker, 2018), and *Triticum* (Kapil et al., 2018). However, the trend to detect common, polymorphic, and unique cpSSRs is relatively new (Kabra et al.,

2016; Kapil et al., 2018) and no such information is available for genus *Nymphaea*. Therefore, the present analysis was conducted to mine SSRs in chloroplast genomes of genus *Nymphaea* and detect common, polymorphic, and unique cpSSRs in them.

II. MATERIALS AND METHODS

A. Mining of cpSSRs

A total of 6 available chloroplast genome sequences of genus *Nymphaea* (Table 1) were retrieved from NCBI (https://www.ncbi.nlm.nih.gov/) in FASTA and GenBank format. FASTA is a text-based file format that contains sequence data of nucleotide or amino acids. The FASTA file starts with a ">" (greater-than) symbol which contains unique description in single line followed by lines of sequence data using their single-letter codes. Whereas, GenBank file format includes annotation section followed by sequence data. It includes various data like organism information, publications, accession ID, coding sequence (CDS) followed by sequence data. The end of GenBank file is marked with "//" symbol.

Table 1: Chloroplast genomes of genus *Nymphaea* retrieved from NCBI (Abbreviation- Abr*).

S. No.	Organism	Abr*	Accession No.	Size (Kb)	Reference
1.	*N. Alba*	*Nal*	NC_006050.1	159.930	Goremykin et al., 2004
2.	*N. Ampla*	*Nam*	NC_035680.2	159.861	Gruenstaeudl et al., 2017
3.	*N. Capensis*	*Nca*	NC_040167.1	159.998	Kim et al., 2019
4.	*N. Jamesoniana*	*Nja*	NC_031826.2	158.963	Gruenstaeudl et al., 2017
5.	*N. lotus*	*Nlo*	NC_041238.1	159.311	Kim et al., 2019
6.	*N. mexicana*	*Nme*	NC_024542.1	159.962	Yang et al., 2014

SSRs were mined in retrieved chloroplast genome sequences using Microsatellite identification tool (MISA; http://pgrc.ipk-gatersleben.de/misa/misa.html). MISA is based on Perl programming language and detect perfect and compound SSR in nucleotide sequence data. To mine the SSRs, the minimum repeat size of ≥12 for mono-, ≥6 for di-, ≥4 for tri-, ≥3 for tetra-, penta- and hexa-nucleotide was considered. Interruption between two SSRs was taken as 0. Moreover, information about coding, non-coding, and coding-non-coding region of mined cpSSRs was retrieved from GenBank file of respective organism.

B. Categorization of Common, Polymorphic, and Unique SSRs

Common, polymorphic, and unique cpSSRs were categorized in mined SSRs with the help of standalone Basic Local Alignment Search Tool (BLAST; Altschul et al., 1997). BLAST is a widely used tool to find similarity between nucleotide or protein sequences. In the present study SSRs of identical repeating units with equal and varying length, showing significant similarity of flanking regions (including 200 bases from both upstream and downstream of SSRs) across the species were considered as common and polymorphic SSRs, respectively. Whereas, SSRs of identical repeating units with no significant match of flanking regions across the species and uniquely identified repeating units were considered as unique SSRs (Kumar & Shanker, 2018). Apart from this, common and polymorphic cpSSRs were also illustrated using MapChart tool (Voorrips, 2002).

III. RESULTS AND DISCUSSION

A. Frequency of cpSSRs Identified in Genus Nymphaea

In the present study, a total of 96 cpSSRs were mined in 6 chloroplast genome of genus *Nymphaea*, with an average density of 1SSR/0.1 kb sequence. The average density observed in this study is higher than the average density of cpSSRs identified in Solanaceae family (1SSR/1.26 kb; Tambarussi et al. 2009), 17 genomes belonging to clade Magnoliids (1SSR/6.91 kb; Srivastava & Shanker 2016), genus *Vigna* (1SSR/6.1 kb; Shukla et al. 2018) and genus *Arabidopsis* (1SSR/0.33 kb; Kumar & Shanker 2018c). The variation in density of SSRs might be due to different parameters taken to mine SSRs and nucleotide composition of sequences mined. Distribution of cpSSRs identified among genus *Nymphaea* are presented in Table 2.

Table 2: Frequency of repeats identified in six chloroplast genomes of genus *Nymphaea.*

S. No.	Organism	Mono	Di	Tri	Tetra	Penta	Hexa	Total	Density
1	*N. alba*	7	-	4	2	1	-	14	0.09
2	*N. ampla*	6	2	4	3	-	1	16	0.1
3	*N. capensis*	7	5	4	3	-	-	19	0.12
4	*N. jamesoniana*	1	3	4	4	1	2	15	0.09
5	*N. lotus*	7	4	3	5	-	-	19	0.12
6	*N. mexicana*	5	-	5	2	1	-	13	0.08
	Total	**33**	**14**	**24**	**19**	**3**	**3**	**96**	

Mono-nucleotide (33, 34.38%) motifs were most abundant in this study followed by tri-nucleotides (24, 25%), tetra-nucleotides (19, 19.79%), and di-nucleotides (14, 14.58%); whereas penta- and hexa-nucleotides (3, 3.13%) were found with equal frequency among chloroplast genomes of genus *Nymphaea* (Fig 1). Moreover, cpSSRs identified in non-coding (72, 75%) regions were most frequent followed by coding (21, 21.88%), and coding-non-coding (3, 3.13%) regions among genus *Nymphaea*. The distribution of cpSSRs identified in coding, non-coding, and coding-non-coding regions among chloroplast genomes of genus *Nymphaea* is presented in Fig 2.

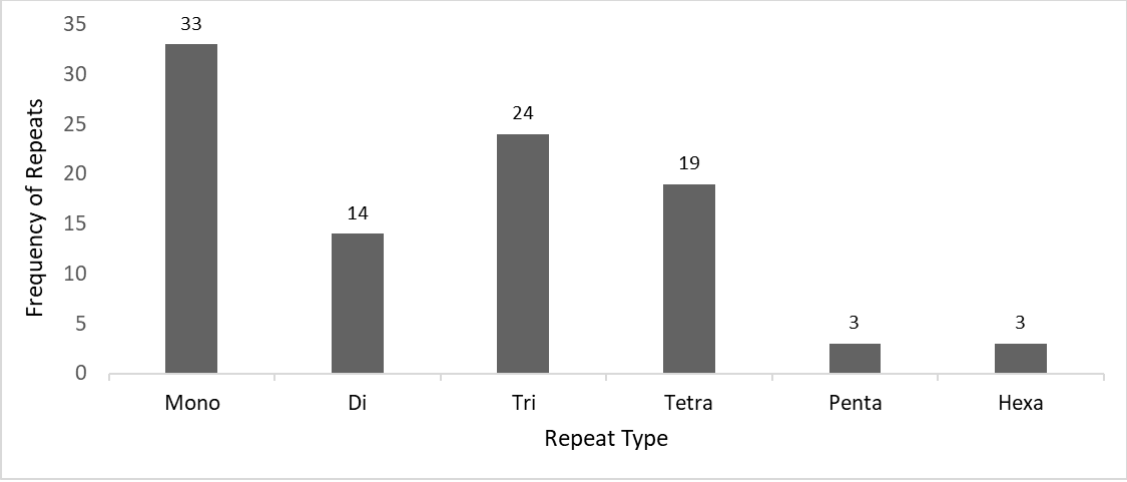

Fig. 1. Frequency of mono-hexa repeats identified in chloroplast genomes of genus *Nymphaea*.

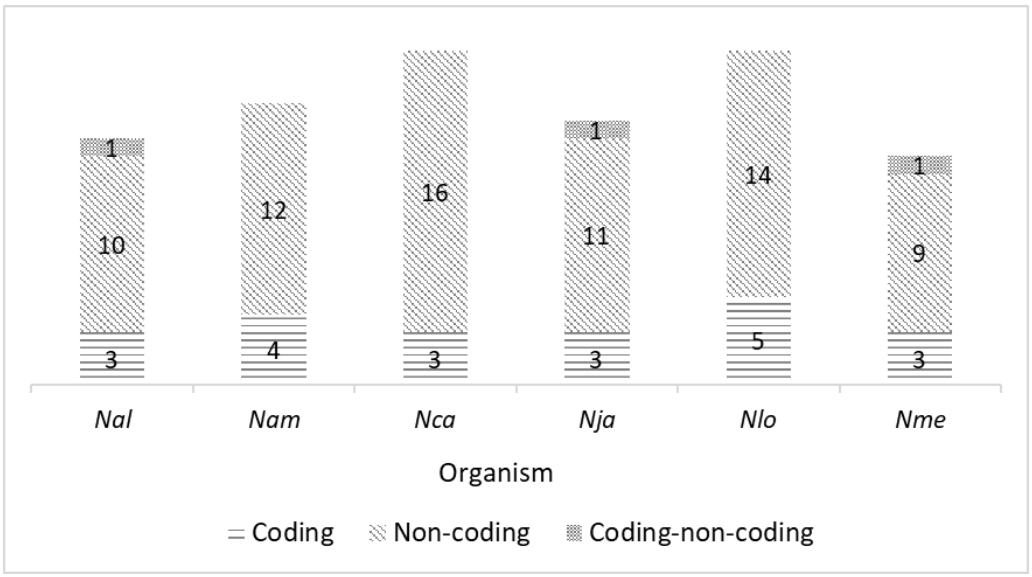

Figure 2: Frequency of cpSSRs identified in coding, non-coding, and coding-non-coding regions among genus *Nymphaea*. Abbreviations of the organisms name are given in Table 1.

In chloroplast genome of *N. alba*, a total of 14 cpSSRs were mined in 159.930 kb sequence with a density of 1SSR/0.09 kb. Among these mono-nucleotides (7, 50%) was most frequent followed by tri-nucleotides (4, 28.57%), tetra-nucleotides (2, 14.28%) and penta-nucleotides (1, 7.14%), whereas di- and hexa-nucleotides motif was completely absent in this chloroplast genome. Moreover, among mono-nucleotides, motif A (4, 57.14%) was most frequent followed by C (2, 28.57%) and T (1, 14.28%). Among tri-nucleotides, motif TAT (2, 50%) was most frequent followed by ATA and TTC (1, 25% of each). Motif ATTT and TTAG (1, of each) were found with equal frequency in tetra-nucleotides. ATACC (1, 100%) was the only motif found in penta-nucleotides in this chloroplast genome.

A total of 16 cpSSRs with a density of 1SSR/0.10 kb were identified in 159.861 kb sequence in *N.* ampla. Among these mono-nucleotides (6, 37.5%) was most frequent followed by tri-nucleotides (4, 25%), tetra-nucleotides (3, 18.75%), di-nucleotides (2, 12.5%), and hexa-nucleotides (1, 6.25%). Penta-nucleotides was completely absent in this genome. Moreover, among mono-nucleotides motif A (3, 50%) was most frequent followed by T (2, 33.33%) and C (1, 16.67%). TA (100%) was the only motif found in di-nucleotides. Motif ATA, TAT, TCC, and TTC (1, 25% of each) found with equal frequency among tri-nucleotides. Similarly, motif AAAG and TTGA (1, 50% of each) found with equal frequency in tetra-nucleotides. Motif AGTTAT (1, 100%) was the only motif identified in hexa-nucleotides in *N.* ampla.

In *N. capensis*, a total of 19 cpSSRs were found with a density of 1SSR/0.12 kb in 159.998 kb sequence. Among these mono-nucleotides (7, 36.84%) was most frequent followed by di-nucleotides (5, 26.32%), tri-nucleotides (4, 21.05), and tetra-nucleotides (3, 15.75%). Penta- and hexa-nucleotides were completely absent in *N. capensis*. Moreover, among mono-nucleotides motif A and C (3, 42.86% of each) was most frequent followed by T (1, 14.28%). TA (5, 100%) was the only motif found in di-nucleotides. Motif ATA, TAT, TCC, and TTC (1, 25% of each) found with equal frequency among tri-nucleotides. Similarly, AAAG, ATGT, and TTGA (1, 33.33% of each) found with equal frequency among tetra-nucleotides.

A total of 15 cpSSRs were mined with a density of 1SSR/0.09 kb in 158.963 kb sequence in *N. jamesoniana*. Among these, tri- and tetra-nucleotides (4, 26.66% of each) were most frequent with equal frequency followed by di-nucleotides (3, 20%) and hexa-nucleotides (2, 13.33%). Mono- and penta-nucleotides were found with equal frequency (1, 6.66% of each). Moreover, T (1, 100%) was the only motif found among mono-nucleotides. TA (2, 66.67%) was most frequent followed by AT (1, 33.33%) among di-nucleotides. ATA, TAT, TCC, and TTC (1, 25% of each) were found with equal frequency among tri-nucleotides. Motif TTGA (2, 50%) was most frequent followed by AGAA and ATTT (1, 25% of each) among tetra-nucleotides. ATCAA was the only motif found in penta-nucleotide. CTAATA and TAACTA (1, 50% of each) were found with equal frequency among hexa-nucleotides.

In *N.* lotus, a total of 19 cpSSRs with a density of 1SSR/0.12 kb were identified in 159.311 kb sequence. Among these, mono-nucleotides (7, 36.84%) were most frequent followed by tetra- (5, 26.31%), di- (4, 21.05%), and tri- (3, 15.78%) nucleotides. Penta- and hexa-nucleotides were completely absent in *N.* lotus. Among mono-nucleotides, motif A (3, 42.85%) was most frequent followed by C and T (2, 28.57% of each). Motif TA (3, 75%) was most frequent followed by AT (1, 25%) among di-nucleotides. Motif ATA, TAT, and TCC (1, 33.33% of each) were found with equal frequency among tri-nucleotides. Among tetra-nucleotides motif TTGA (2, 40%) was most frequent followed by ACAT, AGAA, and CTTT (1, 20% of each).

A total of 13 cpSSRs with a density of 1SSR/0.08 kb were found in 159.962 kb sequence of *N.* mexicana. Among these, mono- and tri-nucleotides (5, 38.46% of each) were most frequent followed by tetra-nucleotides (2, 15.38%) and penta-nucleotides (1, 6.69%). Di- and hexa-nucleotide repeats were completely absent in *N.* Mexicana. Among mono-nucleotide, motif A and T (2 of each) were most frequent followed by C (1). Motif TAT (3) was most frequent followed by ATA and TTC (1 of each) among tri-nucleotides. Motif ATTT and TTGA (1, 50% of each) found with equal frequency among tetra-nucleotides. ATACC (1, 100%) was the only motif found among penta-nucleotide.

B. Common, Polymorphic, and Unique cpSSRs Identified in Genus Nymphaea

The total number of common and polymorphic cpSSRs was observed from 3 to 10 and 1 to 5, respectively between each pair of species. Additionally, unique cpSSRs were also observed from 2 to 7 among genus *Nymphaea*. In the present study mono-nucleotide motifs (A, C, and T) frequently showed length polymorphism followed by di-nucleotides (TA) and tri-nucleotides (TCC) among genus *Nymphaea*. Total numbers of common, polymorphic, and unique cpSSRs identified in genus *Nymphaea* are presented in Table 3. Moreover, common and polymorphic cpSSRs between *N. alba* and rest of the chloroplast genomes among genus *Nymphaea* were illustrated (Fig 3).

Recently, length variation in 20 cpSSRs was detected and 3 were identified as putative polymorphic cpSSRs in three species of genus *Triticum* (*T. aestivum*, *T. monococcum*, and *T. urartu;* Kapil et al., 2018). In another study, common, polymorphic, and unique cpSSRs between each pair of species were identified in 12 chloroplast genome sequences of genus *Arabidopsis* and variation in frequency of cpSSRs were observed (Kumar & Shanker, 2018).

Table 3: Frequency of common, **polymorphic (bold)**, and unique (***U****) cpSSRs among genus *Nymphaea*. Abbreviations of the organisms name are given in Table 1.

	U*	***Nal***	***Nme***	***Nja***	***Nam***	***Nca***	***Nlo***
Nal	***2***	-	7	-	4	3	3
Nme	***3***	**3**	-	5	-	4	-
Nja	***4***	-	**1**	-	5	4	5
Nam	***3***	**3**	-	**3**	-	10	-
Nca	***6***	**3**	**1**	**4**	**1**	-	8
Nlo	***7***	**3**	-	**5**	-	**2**	-

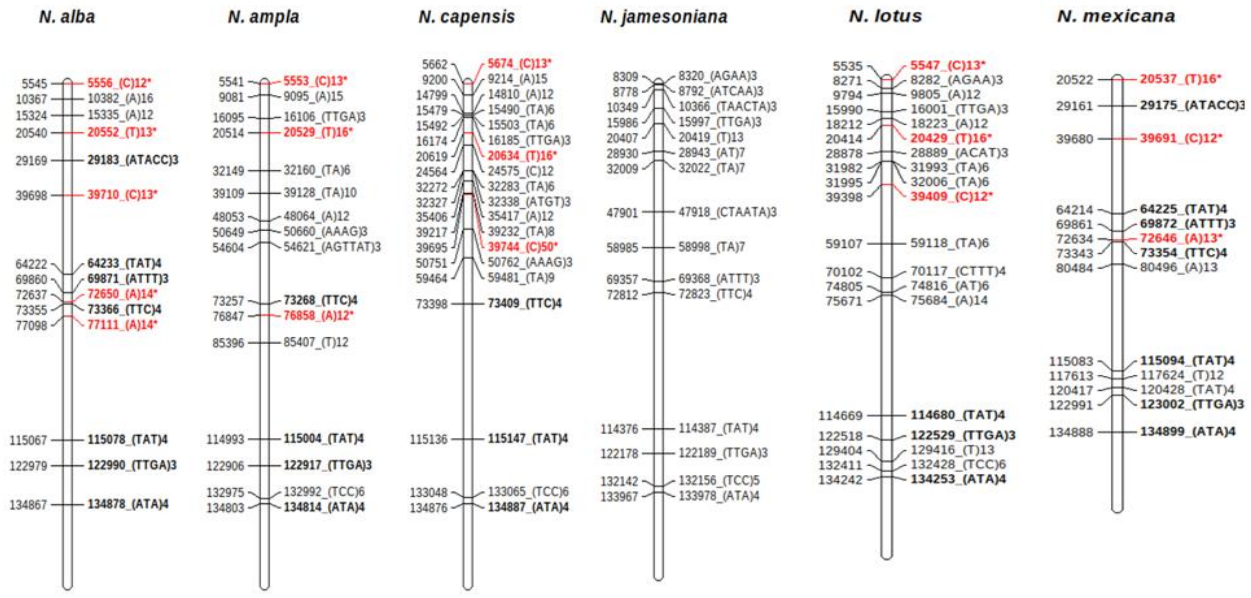

Figure 3: Common (bold) and polymorphic (*) cpSSRs between *N. alba* and rest of the chloroplast genomes.

CONCLUSION

Common, unique, and polymorphic simple sequence repeats identified in this study using a computational approach will play important role in studies including species identification, genetic mapping, and diversity analysis in genus *Nymphaea*.

ACKNOWLEDGMENT

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

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.

Bachmann, L., Bareiss, P., & Tomiuk, J. (2004). Allelic variation, fragment length analyses and population genetic models: a case study on *Drosophila* microsatellites. *Journal of Zoological Systematics and Evolutionary Research*, 42(3), 215-223.

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *The American Journal of Human Genetics*, 62(6), 1408-1415.

Goremykin, V. V., Hirsch-Ernst, K. I., Wölfl, S., & Hellwig, F. H. (2004). The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm. *Molecular biology and evolution*, 21(7), 1445-1454.

Gruenstaeudl, M., Nauheimer, L., & Borsch, T. (2017). Plastid genome structure and phylogenomics of Nymphaeales: Conserved gene order and new insights into relationships. *Plant systematics and evolution*, 303(9), 1251-1270.

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* (pp. 141-154). Springer, Singapore.

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*.

Kim, Y., Kwon, W., Song, M. J., Nam, S., & Park, J. (2019). The Complete chloroplast genome sequence of the *Nymphaea lotus* L. (Nymphaeaceae). *Mitochondrial DNA Part B*, 4(1), 389-390.

Kim, Y., Min, J., Kwon, W., Song, M. J., Nam, S., & Park, J. (2019). The Complete chloroplast genome sequence of the *Nymphaea capensis* Thumb. (Nymphaeaceae). *Mitochondrial DNA Part B*, 4(1), 401-402.

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

Kumar, S., & Shanker, A. (2018a). Bioinformatics Resources for the Stress Biology of Plants. In Biotic and Abiotic Stress Tolerance in Plants (pp. 367-386). Springer, Singapore.

Kumar, S., & Shanker, A. (2018b). Biological Databases for Medicinal Plant Research. In Biotechnological Approaches for Medicinal and Aromatic Plants (pp. 655-665). Springer, Singapore.

Kumar, S., & Shanker, A. (2018c). Common, Unique and Polymorphic Simple Sequence Repeats in Chloroplast Genomes of genus *Arabidopsis*. *Vegetos-An International Journal of Plant Research*, 31(special), 125-131.

Özyiğit, İ. İ., Doğan, İ., & Filiz, E. (2015). *In silico* analysis of simple sequence repeats (SSRs) in chloroplast genomes of *Glycine* species, 8(1), 24-29.

Sen, S., Dehury, B., Sahu, J., Rathi, S., & Yadav, R. N. S. (2017). 'In silico' mining and characterization of simple sequence repeats (SSRs) from 'Euphorbia esula' expressed sequence tags (ESTs): A potential crop for biofuel. *Plant Omics*, 10(2), 53.

Shanker, A. (2013). Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. Archive for Bryology 199, 1-6.

Shanker, A. (2014). Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. *Archive for Bryology*, 199, 1-5.

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.

Shukla, N., Kuntal, H., Shanker, A., & Sharma, S. N. (2018). Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*. *Biotechnology Research and Innovation*, 2(1), 9-18.

Singh, M., & Jain, A. P. (2017). A review on genus *Nymphaea*: multi-potential medicinal plant. *Asian Journal of Pharmaceutical Education and Research*, 6(4), 1-9.

Song, S. L., Lim, P. E., Phang, S. M., Lee, W. W., Hong, D. D., & Prathep, A. (2014). Development of chloroplast simple sequence repeats (cpSSRs) for the intraspecific study of *Gracilaria tenuistipitata* (Gracilariales, Rhodophyta) from different populations. *BMC research notes*, 7(1), 77.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M., & Powell, W. (2003). How much

effort is required to isolate nuclear microsatellites from plants? *Molecular ecology*, 12(6), 1339-1348.

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

Srivastava, D., & Shanker, A. (2016). Identification of simple sequence repeats in chloroplast genomes of Magnoliids through bioinformatics approach. *Interdisciplinary Sciences: Computational Life Sciences*, 8(4), 327-336.

Taheri, S., Lee Abdullah, T., Yusop, M., Hanafi, M., Sahebi, M., Azizi, P., & Shamshiri, R. (2018). Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules*, 23(2), 399.

Tambarussi, E. V., Melotto-Passarin, D. M., Gonzalez, S. G., Brigati, J. B., de Jesus, F. A., Barbosa, A. L., Dressano K.,& Carrer, H. (2009). In silico analysis of Simple Sequence Repeats from chloroplast genomes of *Solanaceae* species. *Crop Breeding & Applied Biotechnology*, 9(4).

Tautz, D., & Schlötterer, C. (1994). Simple sequences. *Current opinion in genetics & development*, 4(6), 832-837.

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity*, 93(1), 77-78.

Yang, J. B., Li, D. Z., & Li, H. T. (2014). Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources*, 14(5), 1024-1031.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., & Simon, P. (2012). Using next- generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American journal of botany*, 99(2), 93-208.

Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular ecology*, 11(1), 1-16.

Zhu, H., Senalik, D., McCown, B. H., Zeldin, E. L., Speers, J., Hyman, J., Bassil, N., Hummer, K., Simon, P., & Zalapa, J. E. (2012). Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theoretical and applied genetics*, 124(1), 87-96.

***