In silico mining of simple sequence repeats in chloroplast genome of *Nothoceros aenigmaticus*

Asheesh Shanker

Department of Bioscience and Biotechnology, Banasthali University, Banasthali-304022, Rajasthan, India
ashomics@gmail.com

Abstract: Simple sequence repeats (SSRs), also known as microsatellites, are found in DNA sequences and consist of short repeating motifs of 1-6 nucleotides. These repeats are ubiquitous and play an important role in the development of molecular markers. Therefore, the present analysis was conducted to identify SSRs in chloroplast genome of *Nothoceros aenigmaticus*. A total of 40 perfect SSRs were identified in 153.208 kb sequence mined with an average length of 19.35 bp. The identified SSRs showed a density of 1 SSR/3.65 kb. Depending on the repeat units, the length of SSRs ranged from 12 to 54 bp. Tetranucleotides (13, 32.5%) were found to be the most abundant repeat, followed by trinucleotide (12, 30%), dinucleotide (8, 20%), mononucleotide (4, 10%) and hexanucleotide (3, 7.5%) repeats. The pentanucleotide repeats were not detected in chloroplast genome of *Nothoceros aenigmaticus*.
[Shanker A. ***In silico* mining of simple sequence repeats in chloroplast genome of *Nothoceros aenigmaticus*.** *Researcher* 2015;7(3):24-27]. (ISSN: 1553-9865). http://www.sciencepub.net/researcher. 4

Key words: Chloroplast; Data Mining; Microsatellites; Simple sequence repeats

1. Introduction

Bryophytes have been broadly classified into liverworts, mosses & hornworts, and are considered as earliest land plants. In the recent past, paraphyly of bryophytes was inferred in maximum likelihood based phylogenetic analysis using chloroplast and mitochondrial genome sequences (Shanker, 2013; Shanker, 2013a; Shanker, 2013b). Moreover, chloroplast genome sequences of bryophytes have been used for the identification of microsatellites (Shanker, 2014; Shanker, 2014a; Shanker, 2014b).

Microsatellites, also known as simple sequence repeats (SSRs), have been known to present in DNA sequences and consist of short repeat motifs (1-6 nucleotides). These repeats are ubiquitous and found in both coding/non-coding regions of genome (Shanker et al., 2007). SSRs have been considered as molecular markers of choice in many plant genomes (Cardle et al., 2000; Gupta et al., 2003; Molla et al., 2011; Salem and Mattar, 2014). Recently SSRs have been detected in chloroplast and mitochondrial genomes of bryophytes (Shanker, 2013c; Shanker, 2013d; Villarreal et al., 2012). Moreover, considering the importance of these repeats SSR specific databases including MitoSatPlant (Kumar et al., 2014) and ChloroSSRdb (Kapil et al., 2014) have also been developed. Most of these studies have been done using computational approaches. In comparison to the biotechnological approaches, *in silico* analysis serve as a rapid and cost-effective method to identify SSRs in sequences deposited in public databases (Shanker et al., 2007a).

Despite all these efforts a detailed analysis of SSRs in chloroplast genome of *Nothoceros aenigmaticus* is not available. Therefore, in this analysis the chloroplast genome of *Nothoceros aenigmaticus* was screened to know the distribution of SSRs in coding and non-coding regions of genome.

2. Materials and Methods

2.1 Chloroplast genome sequence retrieval and data mining

The complete organellar genome sequences of bryophytes (Shanker, 2012; Shanker, 2012a) are available at National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov). The chloroplast genome sequence of *Nothoceros aenigmaticus* (Villarreal et al., 2013) was downloaded from NCBI in FASTA and GenBank format (Accession number: NC_020259, 153208 bp). The methods employed previously for SSRs mining in chloroplast genome of bryophytes (Shanker, 2013c; Shanker, 2013d; Shanker, 2014) were used in this study.

3. Results and Discussion

The present analysis deals with the identification of perfect chloroplastic simple sequence repeats (cpSSRs) in *Nothoceros aenigmaticus*. A minimum length of 12 nucleotides was considered to mine SSRs and the length of the identified SSRs ranged from 12 to 54 bp. Pentanucleotide repeats were not detected in chloroplast genome of *Nothoceros aenigmaticus*. The distribution of mined cpSSRs is presented in figure 1.
A total of 40 perfect SSRs were identified in 153.208 kb sequence mined with an average length of 19.35 bp. Depending on the repeat units, the length of SSRs ranged from 12 to 54 bp. Tetranucleotides (13, 32.5%)

were found to be the most abundant repeat, followed by trinucleotide (12, 30%), dinucleotide (8, 20%), mononucleotide (4, 10%) and hexanucleotide (3, 7.5%) repeats. The identified SSRs motif, their length, start-end position and the region in which they lie is presented in table 1. It is evident from this table that the majority of SSRs were found in non-coding region of the chloroplast genome.

Table 1. Identified SSRs motif, their length, start-end position in chloroplast genome of *Nothoceros aenigmaticus*.

S. No.	Motif	Length	Start	End	Region
1.	(TA)27	54	4563	4616	Non coding
2.	(AT)10	20	6887	6906	Non coding
3.	(TAA)8	24	18559	18582	Non coding
4.	(TATC)3	12	18896	18907	Non coding
5.	(ATAG)3	12	18909	18920	Non coding
6.	(TATGGG)3	18	26025	26042	Non coding
7.	(ACA)4	12	26510	26521	Non coding
8.	(AAGA)3	12	42130	42141	Non coding
9.	(TAAT)3	12	44715	44726	Non coding
10.	(AT)16	32	44962	44993	Non coding
11.	(AT)25	50	46645	46694	Non coding
12.	(ATT)4	12	47763	47774	Non coding
13.	(TTTC)3	12	48574	48585	Non coding
14.	(TTC)4	12	53861	53872	Non coding
15.	(TTA)9	27	54255	54281	Non coding
16.	(AGAAGT)3	18	62902	62919	Non coding
17.	(AT)7	14	63290	63303	Non coding
18.	(ACT)15	45	66535	66579	Non coding
19.	(AT)6	12	68118	68129	Non coding
20.	(T)12	12	70645	70656	Non coding
21.	(CCT)5	15	77985	77999	Non coding
22.	(TTTG)3	12	83848	83859	Non coding
23.	(ATT)4	12	85853	85864	Non coding
24.	(AT)9	18	89224	89241	Non coding
25.	(TAT)4	12	89315	89326	Non coding
26.	(AATT)3	12	91525	91536	Non coding
27.	(GGTT)3	12	93817	93828	Non coding
28.	(AT)6	12	94867	94878	Non coding
29.	(AAAT)3	12	96044	96055	Non coding
30.	(ATT)5	15	96889	96903	Non coding
31.	(TAT)4	12	96912	96923	Non coding
32.	(TTTA)3	12	102071	102082	Non coding
33.	(TAAT)3	12	103631	103642	Non coding
34.	(T)13	13	106665	106677	Non coding
35.	(T)12	12	110830	110841	Non coding
36.	(AGGT)3	12	114429	114440	Coding
37.	(TTA)6	18	126916	126933	Non coding
38.	(TTCTTT)3	18	136501	136518	Non coding
39.	(CTAC)3	12	146115	146126	Coding
40.	(A)12	12	149716	149727	Non coding

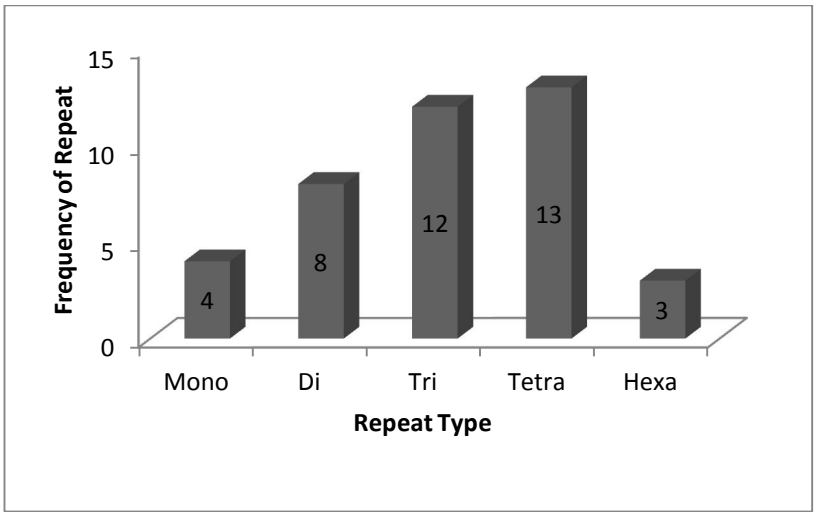

Figure 1. Frequency distribution of various repeat types.

The chloroplast genome of *Nothoceros aenigmaticus* contains 1 SSR/3.65 kb. The density of cpSSRs in this study found to be higher than the cpSSRs of *Pellia endiviifolia* (1 SSR/7.09 kb; Shanker, 2014), *Ptilidium pulcherrimum* (1 SSR/5.17 kb; Shanker, 2014a), *Aneura mirabilis* (1 SSR/5.68 kb; Shanker, 2013c), cpSSRs of rice (1 SSR/6.5 kb; Rajendrakumar et al., 2007) and Unigenes sequences of *Citrus* (1 SSR/12.9 kb; Shanker et al., 2007a). However the density of cpSSRs in *Nothoceros aenigmaticus* found to be lower than *Marchantia polymorpha* (1 SSR/1.83 kb; Shanker, 2014b), *Anthoceros formosae* (1 SSR/2.4 kb; Shanker, 2013d) and cpSSRs density in family Solanaceae (1 SSR/1.26kb; Tambarussi et al., 2009). The variation in SSR density might be due to different parameters including minimum length of SSRs taken, the amount of data analyzed and genomic composition of the sequence mined. As found in earlier studies of bryophytes (Shanker, 2013c; Shanker, 2013d; Shanker, 2014; Shanker, 2014a; Shanker, 2014b) most of the cpSSRs identified in the *Nothoceros aenigmaticus* also lie in the non-coding region.

4. Conclusion

SSRs were successfully identified using computational approach in the chloroplast genome of *Nothoceros aenigmaticus*. The identified cpSSRs can be useful to develop SSR markers. Moreover they can be further used in phylogenetic and diversity studies of *Nothoceros* species.

Acknowledgment:

I am thankful to University Grants Commission, India for providing financial support in the form of Major Research Project (F. No. 42-138/2013).

Correspondence to:
Asheesh Shanker,
Deptt. of Bioscience and Biotechnology,
Banasthali University, Banasthali-304022,
Rajasthan (INDIA).
Email: ashomics@gmail.com

References

1. Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D and Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 2000; 156:847-854.
2. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N and Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 2003; 270:315-323.
3. Kapil A, Rai PK, Shanker A. ChloroSSRdb: a repository of perfect and imperfect chloroplastic simple sequence repeats (cpSSRs) of green plants. Database (Oxford) 2014; doi:10.1093/database/bau107.
4. Kumar M, Kapil A, Shanker A. MitoSatPlant: Mitochondrial microsatellites database of viridiplantae. Mitochondrion 2014; 19:334-337.
5. Molla MR, Islam MN, Rohman MM, Ahmed I, and Rahman L. DNA fingerprinting of Rape seed (*Brassica rapa* L.) varieties of Bangladesh using SSR markers. Nature and Science, 2011;9: 222-228.
6. Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K and Sundaram RM. Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. Bioinformatics 2007; 23:1-4.

7. Salem KFM and Mattar MZ. Identification of microsatellite alleles for salt tolerance at seedling stage in wheat (*Triticum aestivum* L.). Life Science Journal 2014;11: 1064-1073.
8. Shanker A. Chloroplast genomes of bryophytes: a review. Arch Bryol 2012; 143:1-5.
9. Shanker A. Sequenced mitochondrial genomes of bryophytes. Arch Bryol 2012a; 146:1-6.
10. Shanker A. Paraphyly of bryophytes inferred using chloroplast sequences. Arch Bryol 2013; 163:1-5.
11. Shanker A. Inference of bryophytes paraphyly using mitochondrial genomes. Arch Bryol 2013a; 165:1-5.
12. Shanker A. Combined data from chloroplast and mitochondrial genome sequences showed paraphyly of bryophytes. Arch Bryol 2013b; 171:1-9.
13. Shanker A. Mining of simple sequence repeats in chloroplast genome of a parasitic liverwort: *Aneura mirabilis*. Arch Bryol 2013c; 196:1-4.
14. Shanker A. Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. Arch Bryol 2013d; 191:1-6.
15. Shanker A. Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. Arch Bryol 2014; 199:1-5.
16. Shanker A. Computational mining of microsatellites in the chloroplast genome of *Ptilidium pulcherrimum*, a liverwort. International Journal of Environment 2014a; 3:50-58.
17. Shanker A. Simple sequence repeats mining using computational approach in chloroplast genome of *Marchantia polymorpha*. Arctoa 2014b; 23:145-149.
18. Shanker A, Bhargava A, Bajpai R, Singh S, Srivastava S and Sharma V. Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis*. Sci Hortic 2007a; 113:353-361.
19. Shanker A, Singh A and Sharma V. *In silico* mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. Microbiol Res 2007; 162:250-256.
20. Tambarussi EV, Melotto-Passarin DM, Gonzalez SG, Brigati JB, de Jesus FA, Barbosa AL, Dressano K and Carrer H. *In silico* analysis of simple sequence repeats from chloroplast genomes of Solanaceae species. Crop Breed Appl Biotech 2009; 9:344-352.
21. Villarreal JC, Forrest LL, McFarland K, Goffinet B. Chloroplast, mitochondrial, and nuclear microsatellites from the southern Appalachian hornwort, *Nothoceros aenigmaticus* (Dendrocerotaceae). Am J Bot 2012; 99:e88-90.
22. Villarreal JC, Forrest LL, Wickett N, Goffinet B. The plastid genome of the hornwort *Nothoceros aenigmaticus* (Dendrocerotaceae): phylogenetic signal in inverted repeat expansion, pseudogenization, and intron gain. Am J Bot 2013; 100:467-477.

3/2/2015