Mining of simple sequence repeats in chloroplast genome of a parasitic liverwort: *Aneura mirabilis*

Asheesh Shanker

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

Abstract: *Aneura mirabilis* is a parasitic liverwort with a chloroplast genome size of 108007 bp. In this study simple sequence repeats (SSRs) were detected using bioinformatics approch in plastid genome of *Aneura mirabilis*. Due to its small genome size only 19 repeats were detected showing a density of 1 SSR/5.68 kb. The length of SSRs ranged from 12 to 20 bp. Mononucleotide repeats were the most frequent repeat type (36.84%) followed by tetranucleotide repeats (31.58%). Moreover hexanucleotide repeats were absent in chloroplast genome sequence of *Aneura mirabilis*.

Introduction

Bryophytes are the earliest land plants and are broadly classified into liverworts, mosses and hornworts. Phylogenetic analysis based on molecular sequences showed liverworts as the earliest diverging lineage and hornworts as sister group to vascular plants (Shanker 2013; Shanker 2013a; Shanker 2013b).

Simple sequence repeats (SSRs) also known microsatellites are short repeat motifs (1-6 bp) containing sequences (Shanker et al. 2007). These repeats are present in DNA sequences, both in coding and non-coding regions (Shanker et al. 2007a). Due to presence of large number of SSRs in the genome and their ability to associate with many phenotypes, these repeats have been widely used as molecular markers in many plant genomes (Gupta et al. 2003). However, there is lack of information about SSRs in chloroplast genomes of bryophytes.

Long established molecular methods of SSR extraction are costly and consumes time. However computational approaches developed in the recent past offer rapid and inexpensive SSR extraction using sequences deposited in public databases (Shanker et al. 2007). Consequently the present analysis was designed to mine SSRs in chloroplast genome sequence of *Aneura mirabilis*. It will enhance our understanding about the organization and distribution of these repeats in coding and non-coding regions of *Aneura mirabilis* chloroplast genome.

Materials and Methods

Retrieval of chloroplast genome sequence

A small number of organelle genome sequences of bryophytes are available (Shanker 2012; Shanker 2012a). The complete chloroplast genome sequence of *Aneura mirabilis* (NC_010359, 108007 bp; Wickett et al. 2008) was downloaded from National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) in FASTA and GenBank format.

Simple sequence repeats mining

To identify SSRs in chloroplast genome sequence of *Aneura mirabilis*, a Perl script named MISA (available at http://pgrc.ipkgatersleben.de/misa/misa) was used. It takes FASTA formatted sequence file as an input and generates information of mined SSRs, if detected, along with statistical data in two separate files. The length of SSRs were defined as ≥12 bp for mono, di, tri and tetranucleotide, ≥15 bp for pentanucleotide and ≥18 bp for hexanucleotide repeats. Based on the presence of repeats in coding and non-coding regions of chloroplast genome, the mined SSRs were classified as coding and non-coding SSRs. Information of coding and non-coding regions of chloroplast genome was taken from GenBank file.

Results and Discussion

In this study perfect SSRs was identified in chloroplast genome sequence of *Aneura mirabilis*. The length of the identified SSRs ranges from 12 to 20 bp. Hexanucleotide repeats were completely absent in chloroplast genome sequence of *Aneura mirabilis*. Majority of the detected SSRs were found in non-coding region of the genome. The frequency of identified SSRs is presented in Fig. 1.

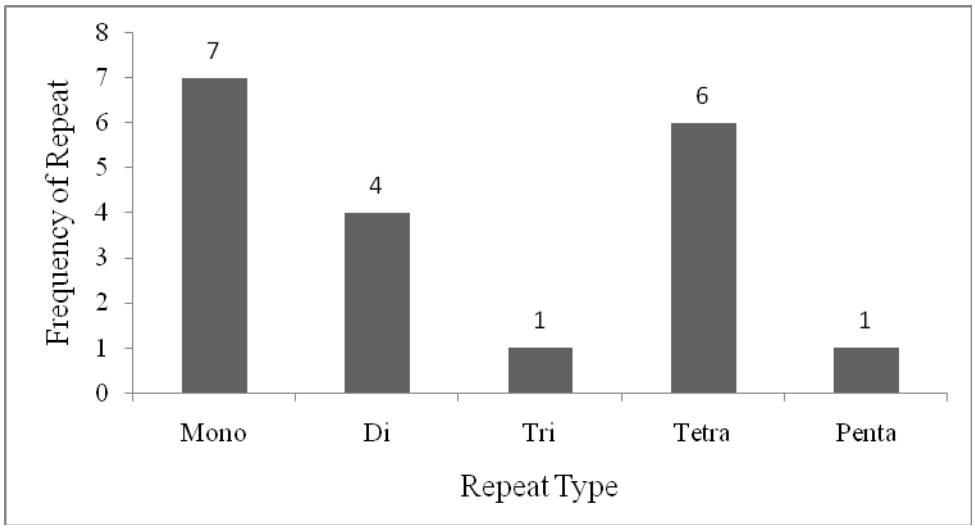

Figure 1. Frequency distribution of SSRs mined.

Only 19 SSRs were identified in chloroplast genome sequence of *Aneura mirabilis* representing density of 1 SSR/5.68 kb. Information of mined 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 mononucleotides were the most frequent repeat (7, 36.84%) followed by tetranucleotide (6, 31.58%) and dinucleotide (4, 21.05%). Tri and pentanucleotide repeats were found with equal frequencies (1, 5.26%). Hexanucleotide repeats were totally absent in chloroplast genome sequence of *Aneura mirabilis*. Out of all mined SSRs 4 (21.05%) lie in coding and 15 (78.95%) lie in non-coding region of the genome.

Table 1: Information of mined SSRs in chloroplast genome sequence of *Aneura mirabilis*.

S. No.	Motif	Length	Start	End	Region
1	(GGAGG)3	15	4763	4777	Coding
2	(GAAA)3	12	17148	17159	Non coding
3	(AT)6	12	20614	20625	Non coding
4	(C)12	12	20849	20860	Non coding
5	(AT)8	16	21527	21542	Non coding
6	(ATT)4	12	25121	25132	Non coding
7	(G)14	14	28778	28791	Non coding
8	(C)13	13	28921	28933	Non coding
9	(C)15	15	38762	38776	Non coding
10	(ATGT)3	12	41351	41362	Non coding
11	(TA)6	12	47090	47101	Non coding
12	(G)12	12	51986	51997	Non coding
13	(TA)10	20	57831	57850	Non coding
14	(G)12	12	69626	69637	Non coding
15	(T)12	12	74165	74176	Non coding
16	(AGGT)3	12	83455	83466	Coding
17	(AATC)3	12	87122	87133	Coding
18	(CTAA)3	12	88194	88205	Non coding
19	(CTAC)3	12	102093	102104	Coding

Earlier studies on chloroplastic SSRs of *Anthoceros formosae* (1 SSR/2.4 kb; Shanker 2013c) and in family Solanaceae (1 SSR/1.26 kb; Tambarussi et al. 2009) showed a higher density of SSRs than reported in this study for *Aneura mirabilis* chloroplast genome (1 SSR/5.68 kb). Contrary to this the density of chloroplastic SSRs in *A. mirabilis* was higher in comparison to the chloroplastic SSRs density in rice (1 SSR/6.5 kb; Rajendrakumar et al. 2007). In addition to this the density of chloroplastic SSRs in this analyis found to be higher than the density of EST-SSRs in barley, maize, wheat, rye, sorghum and rice (1 SSR/6.0 kb; Varshney et al. 2002), cotton and poplar (1

SSR/20 kb and 1 SSR/14 kb respectively; Cardle et al. 2000), Unigenes sequences of *Citrus* (1 SSR/12.9 kb; Shanker et al. 2007). The variation in SSR density may be due to the amount of data analyzed or might be due to adoption of different parameters (e.g. minimum length of SSRs taken), diversity of SSR identification tools and searching algorithms used during SSR detection. Most of the SSRs in chloroplast genome of *Aneura mirabilis* were found in non-coding region as detected earlier (Hancock 1995; Shanker 2013c). The chloroplastic SSRs identified in this study can be used to develop SSR markers.

References

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D and Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156: 847-54.

Filiz E and Koc I (2012) *In silico* chloroplast SSRs mining of *Olea* species. Biodiversitas 13: 114-117.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N and Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics 270: 315-323.

Hancock JM (1995) The contribution of slippage-like processes to genome evolution. J Mol Evol 41: 1038-1047.

Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K and Sundaram RM (2007) Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. Bioinformatics 23: 1-4.

Shanker A (2012) Chloroplast genomes of bryophytes: a review. Archive for Bryology 143: 1-5.

Shanker A (2012a) Sequenced mitochondrial genomes of bryophytes. Archive for Bryology 146: 1-6.

Shanker A (2013) Paraphyly of bryophytes inferred using chloroplast sequences. Archive for Bryology 163: 1-5.

Shanker A (2013a) Inference of bryophytes paraphyly using mitochondrial genomes. Archive for Bryology 165: 1-5.

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

Shanker A (2013c) Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. Archive for Bryology 191: 6 pp.

Shanker A, Bhargava A, Bajpai R, Singh S, Srivastava S and Sharma V (2007a) Bioinformatically mined simple sequence repeats in UniGene of *Citrus sinensis.* Sci Hort 113: 353-361.

Shanker A, Singh A and Sharma V (2007) *In silico* mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. Microbiol Res 162: 250- 256.

Tambarussi EV, Melotto-Passarin DM, Gonzalez SG, Brigati JB, de Jesus FA, Barbosa AL, Dressano K and Carrer H (2009) *In silico* analysis of simple sequence repeats from chloroplast genomes of Solanaceae species. Crop Breed Appl Biotech 9: 344-352.

Varshney RK, Thiel T, Stein N, Langridge P and Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell & Mol Biol Lett 7: 537-546.

Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, DePamphilis CW, Boore JL, Goffinet B (2008) Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol Biol Evol 25: 393-401.

ISSN 0945-3466