Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Computational Analysis of Microsatellite Repeats in Chloroplast Genomes By

G. V. Padma Raju¹

⁴ ¹ SRKR ENGINEERING COLLEGE AFFILIATED TO ANDHRA UNIVERSITY

Received: 13 December 2014 Accepted: 3 January 2015 Published: 15 January 2015

7 Abstract

Chloroplasts are the food producers of the cell. These organelles are found only in plant cells 8 and algae. Chloroplasts work to convert light energy of the Sun into sugars that can be used 9 by cells. Microsatellites are a special class of DNA repeats that are found to be helpful to 10 understand evolution, diseases and are widely used in various applications including, DNA 11 Fingerprinting, Paternity Studies, Linkage Analysis etc. These repeats are ubiquitously 12 present in all genomes including chloroplasts and very little is known about their presence in 13 organelle genomes. In this study, we have analyzed more than 370 chloroplast genomes and a 14 brief report on the distribution and frequency of these repeats in chloroplast genomes has 15 been presented. 16

17

3

5

18 Index terms— chloroplast; microsatellites; bioinformatics; genomes; repeats; distribution; computational 19 analysis;.

20 1 Introduction

²¹ hloroplasts, the organelles responsible for photosynthesis, are in many respects similar to mitochondria.

Both chloroplasts and mitochondria function to generate metabolic energy, evolved by endosymbiosis, contain their own genetic systems, and replicate by division. However, chloroplasts are larger and more complex than mitochondria, and they perform several critical tasks in addition to the generation of ATP. Most importantly, chloroplasts are responsible for the photosynthetic conversion of Carbon Di-oxide to carbohydrates. In addition, chloroplasts synthesize amino acids, fatty acids, and the lipid components of their own membranes. The reduction of nitrite to ammonia, an essential step in the incorporation of nitrogen into organic compounds, also occurs in chloroplasts. Moreover, chloroplasts are only one of several types of related organelles (plastids) that play a

29 variety of roles in plant cells [1][2][3][4][5][6][7].

Microsatellites (sometimes referred to as a variable number of tandem repeats or VNTRs) are short segments of DNA that have a repeated sequence, and they tend to occur in DNA. In some microsatellites, the repeated unit may occur four times, in others it may be seven, or two, or three [8]. These repeats are ubiquitous in nature and are responsible for causing several diseases and cancers [9] [10].

These are used in various applications like DNA Fingerprinting, DNA Forensics, Paternity Studies, and have been considered as potential markers for identifying species, for establishing phylogenetic relationships and also to study evolution [11]. Microsatellites are ubiquitously found in both coding and non-coding regions of all organisms and their distribution in coding regions (genes) is known to affect protein formation and gene regulation [12].

Next-generation sequencing enabled researchers to study biological systems at a level never before possible. Studying mutations in chloroplast microsatellite repeats can be very helpful to understand various biological questions and their usage in various other diverse applications. Few studies [13][14][15][16] earlier analyzed the distribution of microsatellites in chloroplast genomes but they are only confined to single or very low number of

42 genomes. This paper describes the study performed to analyze microsatellite repeats in more than 370 chloroplasts

43 genomes and details have been presented.

44 **2** II.

45 **3** Materials & Methods

46 Imperfect microsatellites have been extracted from Chloro Mito SSRDB [17] version 2.0, an opensource 47 microsatellite repository of sequenced organelle genomes. For this study, a total of 370 chloroplast genome 48 sequences have been used that belong to various classes as shown in Table 1.

⁴⁹ 4 Discussion a) Genome Size Analysis

We did a preliminary study to analyze the genome sizes of all chloroplasts. The chloroplast genome sizes vary from few kbs to a maximum of 1 Mb. The smallest chloroplast genome reported is of size 29529bp that belongs to plant named Plasmodium falciparum HB3 apicoplast (ID: NC_017928) belongs to Non-Viridiplantae category. The largest chloroplast genome spans about 1021616 bp of length that belongs to Paulinella chromatophora

54 chromatophore (ID: NC_011087) belongs to Rhizaria.

In Viridiplantae, the smallest chloroplast genome is Helicosporidium sp. ex Simulium jonesii plastid(ID: NC_008100) of length 37454 bp where as the largest chloroplast genome is Floydiella terrestris(ID: NC_014346) chloroplast of length 521168 bp.

In Non-Viridiplantae, the smallest chloroplast genome is found as Plasmodium falciparum HB3 apicoplast

- (ID: of length 29529 bp where as the largest chloroplast genome is Paulinella chromatophora chromatophore (ID:
 NC_011087) chloroplast of length 1021616 bp. It is observed that this non-Virdiplantae category genome size is
- 61 greater than the Viridiplantae genomes.

When the average genome sizes of chloroplast are considered category wise, it has been observed that the average lengths of Viridiplantae chloroplast genomes are little bit higher when compared to those of other non Virdiplantae(Refer Fig 1). 2 gives a summary of the total number of genomes categorized based on genome sizes of the two classes of chloroplast. It has been observed that majority of the genome sizes lie between 10kb to 500kb, only two genomes namely Floydiella terrestris chloroplast (NC_014346) and Paulinella chromatophora chromatophore (NC_011087) are found to be greater than 500kb. On the other hand, 311 plants of Viridiplantae

show genome sizes between 100kb and 500kb.

⁶⁹ 5 b) Distribution of Microsatellites

Microsatellites in or near genes (coding regions) are found to impact protein formation and gene regulation. When the distribution of microsatellites has been analyzed overall, it is found that around 57% of microsatellite repeats fall in coding regions of all chloroplast genomes. Out of the total 78536 chloroplast microsatellites, 45518 microsatellites fall in gene regions where as the rest 33018 repeats fall in non-coding regions. However, it is surprising to see that the distribution differs when the two classes have been compared separately (Refer Fig. 2). Genomes of Non-Viridiplantae are found to be having majority of its microsatellites in coding regions (64%). On the other hand, green plants (Viridiplantae) show that around 57% of their microsatellites to be distributed

⁷⁷ in coding regions. When two chloroplast categories are compared (Refer Fig. 3), these two categories exhibit a

similar distribution of its microsatellites in coding and non coding regions. It would be interesting to study the

reason behind the major number of microsatellite repeats in Viridiplantae.

⁸⁰ 6 c) Motif-size wise Analysis

We have further analyzed the distribution of chloroplast microsatellites based on their motif sizes. Table 3 lists 81 the proportionate distribution of chloroplast microsatellites motif-size wise. It has been observed that chloroplast 82 genomes are rich in tri and tetra nucleotide repeats which tohether account for more than 77% in Non-virdiplantae. 83 and around 62% in Virdiplantae. Mono, Penta and Hexa-nucleotide repeats are found to be very low in number. 84 When the microsatellite tract lengths have been analyzed, the genomes reported few interesting tract lengths for 85 almost all motif sizes. The average microsatellite tract lengths are usually observed to be not more than 19 bp. 86 But, it is surprising to note that some of the tetra and tri repeats have shown exceptional tract lengths as large 87 as 276bp have been observed. Based on the results in Table 4, we have further tried to find repeats in chloroplast 88 genomes that have exceptional tract lengths. Interestingly, we found 10 repeats in chloroplast with tract lengths 89 100bp or more; out of those, two repeats have tract lengths 200bp or more. Two significant tract lengths of 276 90 and 203 have been reported for genomes with IDs NC_020321, NC_008117 respectively. 91 IV. 92

93 7 Conclusion

In this paper, we have presented a brief description about the distribution of microsatellite repeats in all sequenced
chloroplast genomes of Plants. This study forms the first comprehensive analysis of microsatellite repeats in
chloroplast genomes and the statistics of this study can be a useful resource for biologists.

 $^{^{1}}$ © 2015 Global Journals Inc. (US) 1

 $^{^{2}}$ © 2015 Global Journals Inc. (US)



Figure 1: Figure 1 :



Figure 2: Figure 2 :

1



Figure 3: Figure 3 :

	Category	Total No.
	Alveolata	9
Cryptophyta		3
	Euglenozoa	5
Glaucocystophyceae		1
Haptophyceae		4
	Rhizaria	2
Rhodophyta		9
Stramenopiles		14
Viridiplantae		323
Total Genomes		370
Among the 370 genomes, 323 genomes belong		
to Viridiplantae (Green Plants), 47 genomes belongs to		
Non-Viridiplantae which include genomes of Alveolata,		
Cryptophyta,	Euglenozoa,	Glaucocystophyceae,
Haptophyceae,	Rhizaria,	Rhodophyta and

[Note: $C \otimes 2015$ Global Journals Inc. (US) Global Journal of C omp uter S cience and T echnology Volume XV Issue III Version I Year () C Stramenopiles (]

Figure 4: Table 1 :

 $\mathbf{2}$

	150000	Average genome sizes of chloroplast 148178.28	
Gene size	145000 omh#0000		136551.53
	$135000 \\ 130000$		
		Viridiplantae	Non- Viridiplantae
		Size Range	No. of plants
		>= 10 Kb and <50 Kb	
		Non-Viridiplantae	5
		Viridiplantae $>= 50 \text{ Kb}$ and $<100 \text{ Kb}$	2
		Non-Viridiplantae	10
		Viridiplantae $>= 100 \text{ Kb}$ and $<500 \text{ Kb}$	9
		Non-Viridiplantae	31

Figure 5: Table 2 :

3

Motif Size	Non-Viridiplantae	Viridiplantae
Mono	159(1.80%)	8602(12.33%)
Di	840(9.55%)	7909(11.34%)
Tri	3506(39.87%)	17055(24.45%)
Tetra	3300(37.52)	26796(38.42%)
Penta	623(7.08%)	5680(8.14%)
Hexa	365(4.15%)	3701 (5.31%)
Total	8793	69743

Figure 6: Table 3 :

 $\mathbf{4}$

	Non-Viridiplant	tae			Viridiplantae	
Motif Size	High	Low	Avg	High	Low	Avg
MONO	25	12	13.93	46	12	14.49
DI	54	11	12.90	83	11	13.24
TRI	51	11	12.19	276	11	12.38
TETRA	29	11	11.91	203	11	12.13
PENTA	65	14	15.27	100	14	15.41
HEXA	42	17	18.74	145	17	19.70

Figure 7: Table 4 :

7 CONCLUSION

- 97 [Sablok et al. ()] 'ChloroMitoSSRDB: open source repository of perfect and imperfect repeats in organelle 98 genomes for evolutionary genomics'. G Sablok , S B Mudunuri , S Patnana , M Popova , M A Fares ,
- ⁹⁹ La Porta , N . *DNA research* 2013. 20 (2) p. .
- [Palmer and Thompson ()] 'Chloroplast DNA rearrangements are more frequent when a large inverted repeat
 sequence is lost'. Jeffrey D Palmer , William F Thompson . Cell 1982. 29 (2) p. .
- [Shaw et al. ()] 'Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III'. Joey Shaw , Edgar B Lickey , Edward E Schilling ,
 Randall L Small . American journal of botany 2007. 94 (3) p. .
- [Reith and Munholland ()] 'Complete nucleotide sequence of the Porphyra purpurea chloroplast genome'. Michael
 Reith , Janet Munholland . *Plant Molecular Biology Reporter13* 1995. (4) p. .
- [Kassai-Jáger et al. ()] 'Distribution and evolution of short tandem repeats in closely related bacterial genomes'.
 E Kassai-Jáger , C Ortutay , G Tóth , T Vellai , Z Gáspári . Gene 2008. 410 (1) p. .
- 109 [Schlotterer ()] 'Evolutionary dynamics of microsatellite DNA'. C Schlotterer . Chromosoma 2000. 109 p. .
- [Rajendrakumar et al. ()] 'In silico analysis of microsatellites in organellar genomes of major cereals for
 understanding their phylogenetic relationships'. P Rajendrakumar , A K Biswal , S M Balachandran , R
- M Sundaram . In Silico Biol 2008. 8 p. .
 [Kuntal et al. ()] 'Microsatellite analysis in organelle genomes of Chlorophyta'. H Kuntal , V Sharma , H Daniell
- *Bioinformation* 2012. 8 p. .
- [Thibodeau et al. ()] 'Microsatellite instability in cancer of the proximal colon'. S N Thibodeau , G Bren , D
 Schaid . Science 1993. 260 (5109) p. .
- 117 [Li et al. ()] 'Microsatellites within genes: structure, function, and evolution'. Y C Li , A B Korol , T Fahima , 118 E Nevo . *Molecular biology and evolution* 2004. 21 (6) p. .
- [Goldstein and Schlotterer ()] Microsatellites: evolution and applications, D B Goldstein , C Schlotterer . 2001.
 Oxford: Oxford University Press.
- 121 [Cronn et al. ()] 'Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis tech-
- nology'. Richard Cronn , Aaron Liston , Matthew Parks , David S Gernandt , Rongkun Shen , Todd Mockler
 Nucleic acids research 2008. 36 (19) p. .
- [Bryan et al. ()] 'Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants'. G
 J Bryan , J Mcnicoll , G Ramsay , R C Meyer , W S Jong . *Theoretical and Applied Genetics* 1999. 99 (5) p. .
- [Powell et al. ()] 'Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the
 population genetics of pines'. W Powell , M Morgante , R Mcdevitt , G G Vendramin , J A Rafalski .
 Proceedings of the National Academy of Sciences 1995. 92 (17) p. .
- [Rajendrakumar et al. ()] 'Simple sequence repeats in organellar genomes of rice: frequency and distribution in
 genic and intergenic regions'. P Rajendrakumar , A K Biswal , S M Balachandran , K Srinivasarao , R M
 Sundaram . *Bioinformatics* 2007. 23 p. .
- 132 [Tautz and Schlotterer ()] 'Simple sequences'. D Tautz, C Schlotterer. Curr. Opin. Genet. Dev 1994. 4 p. .
- [Cosner et al. ()] 'The highly rearranged chloroplast genome of Trachelium caeruleum (Campanulaceae): mul tiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several
 repeat families'. Mary E Cosner, K Robert, Jeffrey D Jansen, Stephen R Palmer, Downie. Current genetics
- repeat families'. Mary E Cosner, K Robert, Jeffrey D Jansen, Stephen R Palmer, Downie. Current genetics
 1997. 31 (5) p. .