



## Analysis of the *Cryptophyta* Chloroplast Genome Reveals Presence of Additional Genes and Absence of Introns in their Genome

*B.C. Sharma and Dipen Ghimiray*

*Department of Botany, Darjeeling Government College, Darjeeling, (W.B.)*

*(Received 23 June, 2012, Accepted 25 July, 2012)*

**ABSTRACT :** We have studied the sequenced chloroplast genomes from the three species of *Cryptophyta* (*Cryptomonas*, *Guillardia* and *Rhodomonas*). In our study we find that the typical introns are absent in their genome; though it has been observed in other chloroplast genomes. Considerable number of genes of photosystem II and I and also genes for ribosomal proteins and other protein genes were observed, which may be due to their transfer into the nuclear genome was not completed in due course of evolution.

**Keywords :** Cryptophyta, chloroplast, genome, GenBank.

### INTRODUCTION

Besides nucleus, DNA is also present in the mitochondria and chloroplast. The pioneering sequencing work of the chloroplast genome was done by Shinozaki *et al.*, in 1986 on tobacco plant, followed by work of Ohyama *et al.*, on *Marchantia* in 1986, and by Hiratsuka *et al.*, on rice in 1989. After that, till today nearly 270 chloroplast genomes have been sequenced (NCBI GenBank 2012; Benson *et al.*, 2012; Sayers, 2012). The rapid increase of sequence information of different genomes is due to development of sophisticated sequencing methods and technologies (Cronn *et al.*, 2008). The average length of the chloroplast genome (hence forth mentioned as cpGenome) is about 200 kb. They consists of set of genes for photosynthetic system, protein synthesis system, chloroplast division, some conserved proteins and miscellaneous protein and some potential Open Reading Frames (ORFs) of unidentified functions.

The genes of photosystem I are designated as *psa* followed by respective alphabetical designation, that of photosystem II are designated by *psb*, genes of Cytochrome b6/F is designated by *pet*, *atp* for ATP Synthase, *chl* for chlorophyll biosynthesis, *rbc* for rubisco enzyme, *ndh* for NADH Oxidoreductase system, *rpo* for RNA polymerases, *rpl* for ribosomal proteins of larger sub-unit and *rps* for smaller sub-unit, *fts* and *min* for divisional genes, *ycf* for conserved genes. Usually the cpGenome have two sets of 23S, 16S, 5S and 4.5S rRNA genes and about 30-35 tRNA genes (Shimada and Sigura 1991; Robbens *et al.*, 2007). As we know, the initial activation of photosystem II by light leads to the transfer of electrons to photosystem I via mobile electron carriers, from which further movement of electrons takes place to cytochrome b6/F complex and NADH oxidoreductase system, all these systems functions in a coordinated manner for the entrapment of the solar energy to generate the ATP molecules which are used for photosynthetic reactions, thus these systems are very vital for the sustenance of life on this planet, as the green plants

are the only food producers in the ecosystem. As any reaction inside the living organisms are under the control of either enzyme system or triggered by some signaling molecules, both of them are under the direct or indirect control of the either single or set of genes. Thus, in this communication we are presenting the character and analysis of the genes of the chloroplast genome which are, in part, responsible for the maintenance of this vital process of photosynthesis. After the publication on comparative genomics of chloroplast genome on three species - tobacco, rice and *Marchantia* by Shimada and Shigura (1991) here we have tried to continue the work in this field by comparing cpGenome of three *Cryptophyta* species.

### MATERIAL AND METHODS

For this study, we have selected chloroplast genome of three species of *Cryptophyta* - *Cryptomonas paramecium* (NCBI Ref\_Seq NC\_013703; Donaher *et al.*, 2009), *Guillardia theta* (NCBI Ref\_Seq NC\_000926; Douglas and Penny 1999) and *Rhodomonas salina* (NCBI Ref\_Seq NC\_09573; Khan *et al.* 2007). Their cpGenome was obtained from the NCBI webpage and was used for the analysis of the cpGenome of *Cryptophyta*.

**Gene grouping.** Gene grouping was done first by manual inspection of the genome sequence and later classified them according to the set (gene group) to which they belong. The set parameters were followed as described in introduction section.

**Determination of Length.** As the positions of the genes were described in their sequence, we simply calculated their length by subtracting the initial position from final position and adding 1 to the result, *i.e.*, by the formula:

$$\text{Gene length} = (\text{final position} - \text{initial position}) + 1$$

The simple reason of adding the value of 1 is because we were following the translation from the first Open Reading Frame (ORF).

For this, it is assumed that the initial position is 11 and final position is 20 for a given gene. Simple subtraction of 11 from 20 leads to the result of 9. But manual counting of numbers between 11 and 20 will give us 10. It is because the translation machinery will follow the 3-letter codon pattern of the mRNA and here it has been assumed (in the sequence and gene annotation), by convention, that, translation is following the first open reading frame.

Gene Inspection. This was done by manual observation and comparative analysis of the three cpGenomes.

## RESULTS AND DISCUSSION

### Genes of Photosystem I [*psa*]

There are 11 genes belonging to Photosystem I in *Guillardia* and *Rhodomonas* but interestingly, no genes of photosystem I was found in the cpGenome of *Cryptomonas*. This is in excess to those found in the cpGenomes of higher plants like tobacco (Shinozaki *et al.*, 1986) and rice (Hiratsuka *et al.*, 1989). In these angiosperms, only *psaA*, *psaB*, *psaC* and *psaI* are found in the cpGenome; whereas in *Guillardia* and *Rhodomonas*, there are extra genes like *psaD*, *psaE*, *psaF*, *psaJ*, *psaK*, *psaL* and *psaM*. The longest among these genes is *psaA* with the length of 2259 bp, followed by *psaB* with length of 2205 bp. Though, most of the genes in *Guillardia* and *Rhodomonas* were of similar length, variations were observed in the genome of *Rhodomonas* in genes like *psaF*, *psaL* and *psaM* (Table 1).

**Table 1: Genes of Cryptophyta Chloroplast Genome.**

Genes	Respective gene length in		
	<i>Cryptomonas</i>	<i>Guillardia</i>	<i>Rhodomonas</i>
<b>Genes of Photosystem I (<i>psa</i>)</b>			
<i>psaA</i>		2259	2259
<i>psaB</i>		2205	2205
<i>psaC</i>		246	246
<i>psaD</i>		426	426
<i>psaE</i>		195	195
<i>psaF</i>	Not	567	552
<i>psaI</i>	Present	111	111
<i>psaJ</i>		129	129
<i>psaK</i>		264	264
<i>psaL</i>		462	459
<i>psaM</i>		93	90
<b>Genes of Photosystem II (<i>psb</i>)</b>			
<i>psbA</i>		1083	1082
<i>psbB</i>		1530	1530
<i>psbC</i>		1422	1386
<i>psbD</i>		1051	1056
<i>psbE</i>		255	255
<i>psbF</i>		129	129

<i>psbH</i>		204	204
<i>psbI</i>		117	117
<i>psbJ</i>		120	120
<i>psbK</i>	Not	138	138
<i>psbL</i>	Present	117	117
<i>psbN</i>		132	132#
<i>psbT</i>		99	99
<i>psbV</i>		489	486
<i>psbW</i>		351	351
<i>psbX</i>		120	120
<i>psbY</i>		114	114
<i>psbZ</i>		189	189

### Genes for Cytochrome b6/F (*pet*)

<i>petA</i>	-	966	957
<i>petB</i>	-	648	648
<i>petD</i>	-	483	483
<i>petF</i>	294	294	294
<i>petG</i>	-	114	114
<i>petL</i>	-	96	96
<i>petM</i>	-	-	99
<i>petN</i>	-	90	90

### Genes for ATP Synthase (*atp*)

<i>atpA</i>	1521	1509	1509
<i>atpB</i>	1418	1427	1416
<i>atpD</i>	555	558	543
<i>atpE</i>	290	396	396
<i>atpF</i>	-	549	547
<i>atpG</i>	510	492	468
<i>atpH</i>	249	249	249
<i>atpI</i>	747	747	741

### Genes for Chlorophyll biosynthesis (*chl*)

<i>chlI</i>	1002	1062	1059
-------------	------	------	------

### Genes for RUBISCO (*rbc*)

<i>rbcL</i>	1467	1467	1467
<i>rbcS</i>	420	420	420

### Genes for RNA Polymerase (*rpo*)

<i>rpoA</i>	933	957	945
<i>rpoB</i>	3216	3291	3294
<i>rpoC1</i>	1875	1872	1872
<i>rpoC2</i>	3693	3861	3912

### Genes for NADH Oxidoreductase (*ndh*)

<i>ndh</i>	Absent in all three species		
------------	-----------------------------	--	--

### Genes for rRNA

-	-	121	121
5S	121	-	-
	-	120	121
	-	1487	1491

16S	1493	-	-	<i>rps8</i>	401	398	398
	-	1487	1989	<i>rps9</i>	395	404	401
	-	2888	2895	<i>rps10</i>	314	308	308
23S	289	-	-	<i>rps11</i>	392	392	389
	-	2875	2894	<i>rps12</i>	371	371	374
<b>Genes for tRNA (5'-anticodon-3')</b>				<i>rps13</i>	377	368	371
<i>F - gaa</i>	73	73	73	<i>rps14</i>	302	302	302
<i>L - taa</i>	83	83	84	<i>rps16</i>	236	239	268
<i>caa</i>	82	81	81	<i>rps17</i>	272	272	237
<i>tag</i>	82	83	83	<i>rps18</i>	209	221	221
<i>I - gat</i>	74	74	74	<i>rps19</i>	278	278	278
<i>gat</i>	-	74	74	<i>rps20</i>	296	287	278
<i>M - cat</i>	86	86	84	<b>Genes for Ribosomal Proteins - Larger Sub-unit (<i>rpl</i>)</b>			
<i>cat</i>	74	74	74	<i>rpl1</i>	710	704	704
<i>cat</i>	88	89	87	<i>rpl2</i>	819	827	827
<i>V - tac</i>	72	71	72	<i>rpl3</i>	632	614	626
<i>S - tga</i>	89	88	88	<i>rpl4</i>	644	671	650
<i>gga</i>	87	-	87	<i>rpl5</i>	545	545	545
<i>gct</i>	88	89	87	<i>rpl6</i>	539	539	536
<i>P - tgg</i>	73	74	74	<i>rpl11</i>	425	425	425
<i>T - tgt</i>	73	73	73	<i>rpl12</i>	389	389	389
<i>A - tgc</i>	73	73	73	<i>rpl13</i>	431	431	437
<i>tgc</i>	-	73	73	<i>rpl14</i>	365	365	365
<i>Y - gta</i>	81	82	82	<i>rpl16</i>	407	404	413
<i>H - gtg</i>	73	73	73	<i>rpl18</i>	365	323	362
<i>Q - ttg</i>	72	72	72	<i>rpl19</i>	350	392	368
<i>N - gtt</i>	72	72	72	<i>rpl20</i>	347	344	350
<i>K - ttt</i>	72	72	72	<i>rpl21</i>	326	314	317
<i>D - gtc</i>	74	74	74	<i>rpl22</i>	335	365	356
<i>E - ttc</i>	73	73	73	<i>rpl23</i>	293	287	293
<i>C - gca</i>	73	71	71	<i>rpl24</i>	242	305	308
<i>W - cca</i>	73	73	73	<i>rpl27</i>	266	260	257
<i>R - ccg</i>	73	73	73	<i>rpl29</i>	203	197	206
<i>acg</i>	74	74	74	<i>rpl31</i>	233	218	221
<i>tct</i>	73	73	73	<i>rpl32</i>	-	161	178
<i>G - tcc</i>	70	71	71	<i>rpl33</i>	370	170	170
<i>gcc</i>	72	72	72	<i>rpl34</i>	140	140	134
<b>Genes for Translation Factors</b>				<i>rpl35</i>	197	200	197
<i>tsf</i>	620	660	660	<i>rpl36</i>	146	146	146
<i>infB</i>	1965	2208	2256	<b>Genes for Chloroplast Division</b>			
<i>tufA</i>	1227	1227	1230	<i>minE</i>	-	267	273
<b>Genes for Ribosomal Proteins - Smaller Sub-unit (<i>rps</i>)</b>				<i>minD</i>	Not	810	810
<i>rps2</i>	695	707	731	<i>ftsB</i>	Clear	309	339
<i>rps3</i>	653	650	656	<i>ftsH</i>	-	1896	1887
<i>rps4</i>	608	602	608	<b>Genes for Conserved Proteins (<i>ycf</i>)</b>			
<i>rps5</i>	509	509	506	<i>ycf3</i>	-	519	522
<i>rps6</i>	-	289	299	<i>ycf4</i>	-	546	540
<i>rps7</i>	479	470	470	<i>ycf12</i>	-	105	105

<i>ycf16</i>	765	762	762
<i>ycf17</i>	-	162	-
<i>ycf19</i>	279	276	279
<i>ycf20</i>	318	195	-
<i>ycf24</i>	1479	1452	1479
<i>ycf26</i>	-	-	1092
<i>ycf29</i>	696	627	705
<i>ycf31</i>	-	99	-
<i>ycf33</i>	-	198	210
<i>ycf35</i>	-	387	387
<i>ycf36</i>	-	468	498
<i>ycf37</i>	-	37	522
<i>ycf39</i>	-	945	939#
<i>ycf46</i>	-	1521	1479
<i>ycf61</i>	-	231	231
<i>ycf65</i>	-	300	300

#### Genes for Miscellaneous Proteins

<i>clpC</i>	2454	2460	2460
<i>dnaB</i>	-	1800	1842
<i>groEL</i>	1584	1590	1590
<i>hlp</i>	-	-	159
<i>rbcR</i>	-	933	927
<i>ccsI</i>	-	1245	1326
<i>secA</i>	2637	2634	2634
<i>ccsA</i>	-	906	963
<i>cpcB</i>	-	534	237
<i>cbbX</i>	870	882	870
<i>ilvH</i>	498	510	504
<i>rne</i>	-	1290	1290
<i>cemA</i>	837	837	837
<i>ilvB</i>	1725	1728	1725
<i>tatC</i>	771	873	753
<i>pbsA</i>	702	714	717
<i>secG</i>	-	-	210
<i>acpP</i>	246	246	246
<i>hlpA</i>	-	282	279
<i>dnaK</i>	1881	1884	1890
<i>secY</i>	1236	1263	1236

#### Unassigned Open Reading Frames (ORFs)

<i>ORF335</i>	1008	-	-
<i>ORF91</i>	276	-	-
<i>ORF555</i>	1668	-	-
<i>ORF147</i>	444	-	-
<i>ORF164</i>	495	-	-
<i>ORF125</i>	-	378	-
<i>ORF65</i>	-	198	-
<i>ORF27</i>	-	84	-
<i>ORF252</i>	-	759	-

<i>ORF282</i>	-	849	-
<i>ORF53</i>	-	162	-
<i>ORF62</i>	-	189	-
<i>ORF142</i>	-	-	429
<i>ORF146</i>	-	-	441
<i>ORF27</i>	-	-	84
<i>ORF403</i>	-	-	1212

(#) = Genes with introns

(-) = Absent in the Genome

#### Genes of Photosystem II [*psb*]

Like genes of photosystem I, those of photosystem II were not found to occur in the cpGenome of *Cryptomonas* whereas the cpGenome of *Guillardia* and *Rhodomonas* contain 18 genes for photosystem II. The longest of the group is *psbB* with the length of 1530 bp, second longest is *psbC* with the length of 1422 bp in *Guillardia* and 1386 bp in *Rhodomonas*. Shortest among the group is *psbT* with just 99 bp length. Like genes of photosystem I, there are certain extra genes in *Cryptophyta* in comparison to higher plants - these extra genes include *psbV*, *psbW*, *psbX* and *psbY*. In *Rhodomonas*, one gene *psbN* contains intron of 2235 bp length. Similarly, gene *psbV* has two intervening elements of very short length of 3 bp each (discussed below).

#### Genes of cytochrome b6/F [*pet*]

These genes are designated by *pet*. There is only 1 gene *petF* (294 bp long) in *Cryptomonas*. In *Guillardia* and *Rhodomonas*, there are 7 and 8 genes, respectively - *petA*, *petB*, *petD*, *petF*, *petG*, *petL*, *petM* and *petN*; *petM* is absent in *Guillardia*. In higher plants, *petF* and *petM* are absent. Longest gene among the group is *petA* of 957 bp in *Rhodomonas* and 966 bp in *Guillardia*, shortest genes is *petN* with length of 90 bp in both species. Lengths of remaining genes are equal in both the species.

#### Genes of ATP-Synthase [*atp*]

Designated as *atp*, there are 7 genes in *Cryptomonas*, where as, one additional gene is present in the cpGenome of *Guillardia* and *Rhodomonas* - *atpF* (549 bp). Interestingly, in higher plants *atpF* is also present, but *atpD* and *atpG* are absent in former. Longest among the group is *atpA* with the length of 1521 bp in *Cryptomonas*, 1509 bp in *Guillardia* and *Rhodomonas*. This is followed by *atpB* whose length is quite variable in three species - 1418 bp in *Cryptomonas*, 1427 bp in *Guillardia* and 1416 bp in *Rhodomonas*. Shortest gene among the group is *atpH* with the length of 249 bp in all three species.

#### Genes for Chlorophyll Biosynthesis [*chl*]

These are designated as *chl*, no genes were found to be present in the *Cryptomonas* cpGenome whereas in *Guillardia* and *Rhodomonas*, 1 gene *chlI* was found to be present, which was 1062 bp in *Guillardia* and 1059 bp in

*Rhodomonas*. Other genes such as *chlB*, *chlL* and *chlN* were absent.

#### Genes of RUBISCO [*rbc*]

In the *Cryptophyta*, it was observed in our analysis that both the larger and smaller sub-units of the enzyme RUBISCO are synthesized by their cpGenome. These are designated as *rbcL* for gene of larger sub-unit and *rbcS* for gene of smaller sub-unit. In higher plants, as well as, in green and red algae, *bryophytes*, *pteridophytes*, only larger sub-unit is synthesized by chloroplast DNA, smaller gene being under the control of nuclear genome. *rbcL* is of 1467 bp length (which is quite long in comparison to that of higher plants in which it is around 1434 bp) whereas, *rbcS* is 420 bp long.

#### Genes of RNA Polymerase [*rpo*]

Four genes for RNA Polymerase holoenzyme are synthesized by cpGenome - *rpoA*, *rpoB*, *rpoC1* and *rpoC2* respectively for  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$ . Their length is variable in all the three species. Longest of the group is *rpoC2* and shortest is *rpoB*.

#### Genes of NADH Oxido-reductase [*ndh*]

All the genes of NADH Oxido-reductase are absent in these three species of *Cryptophyta*.

#### Genes of rRNAs

Three genes of rRNAs - 5S, 16S and 23S, are encoded by the cpGenome of *Cryptophyta*, of which, these genes are present in single copy in *Cryptomonas*, and double copies in *Guillardia* and *Rhodomonas*. Gene of 5S is of constant length with 121 bp, 16S is little variable with 1493 bp in *Cryptomonas*, 1487 bp in *Guillardia* and 1491 bp in *Rhodomonas*. Similarly 23S is of 2879, 2888 and 2895 bps in *Cryptomonas*, *Guillardia* and *Rhodomonas*. Interestingly, the second set of genes of rRNAs in *Guillardia* and *Rhodomonas* are slightly variable than the first set, the reason for which is not known.

#### Genes of tRNA

29 genes of tRNAs were found in *Cryptomonas*, whereas in *Guillardia* and *Rhodomonas*, there were 30 and 31 genes. Generally there was one gene for one tRNA in all species, except for 6 (*Cryptomonas*) or 7 (*Guillardia* and *Rhodomonas*) tRNAs. These include Leucine, Isoleucine (not in *Cryptomonas*), Methionine, Serine, Alanine, Arginine and Glycine.

#### Genes of Translation Factors

cpGenome encodes some of the genes for translation factors such as for initiation and elongation. In *Cryptomonas* following genes were found - *tsf*, *infB*, and *tufA*. Similar genes were observed in *Guillardia* and *Rhodomonas*. Longest of these genes is *infB* and shortest is *tsf*.

#### Genes of Ribosomal Proteins (Larger sub-unit) [*rpl*]

26 genes for ribosomal proteins of larger sub-unit are found in the cpGenome of *Cryptophyta*, of which one, *rpl32* is absent in *Cryptomonas*. Longest of this group is *rpl2* followed by *rpl11*, whereas shortest is *rpl34*. Lack of consistency in the length of these genes among the three species was observed during the study. Additional genes present in the cpGenome in comparison to higher plants are of series 1, 3, 4, 5, 6, 11, 12, 13, 18, 19, 21, 24, 27, 29, 31, 34 and 35.

#### Genes of Ribosomal Proteins (Smaller sub-unit) [*rps*]

18 genes for ribosomal proteins of smaller sub-unit are found in the cpGenome of *Cryptophyta*, of which one, *rps6* is, again, absent in *Cryptomonas*. Longest of this group is *rps2* followed by *rps3*, whereas shortest is *rps16*. Here also gene length in the three species was observed as variable in this study. In compare to higher plants, some additional genes of ribosomal smaller subunit protein were found to be present, these are - *rps5*, 6, 9, 10, 13, 17 and 20. In *Rhodomonas*, *rps17* contains a very short intervening element of 3 bp, 5'-GTG-3'.

#### Chloroplast Division Genes

Like other chloroplast genome, genes of chloroplast division were also present in cpGenome of *Cryptophyta*, but these are not distinct in *Cryptomonas*. In *Guillardia*: *minE*, *minD*, *frbB* and *ftsH* were found. Longest of these is *ftsH* (1896 bp). Similar genes were also observed in *Rhodomonas*.

#### Genes of Conserved Proteins [*ycf*]

Genes of conserved proteins are designated by *ycf5*. Such genes in *Cryptomonas*, 18 in *Guillardia* and 16 in *Rhodomonas* were observed. In former, *ycf16*, *ycf19*, *ycf20*, *ycf24* and *ycf29* were only present; whereas in *Guillardia* *ycf3*, 4, 12, 17, 31, 33, 35, 36, 37, 39, 46, 61 and 65 were also found to be present; though, *ycf17*, 20 and 31 were absent in *Rhodomonas*, however, it contains extra *ycf26* gene. *ycf39* of *Rhodomonas* contains two intron-like element (discussed below).

#### Introns in cpGenome

No introns or intron-like elements were found in the genome of *Cryptomonas* and *Guillardia*. However, in *Rhodomonas* one true intron of group II, in the gene *psbN* was found to be present which starts from 127083 and ends in 129317, *i.e.*, of length 2235 bp. This intron has one potential ORF inside it which begins from 127475 and ends in 128752, *i.e.*, of length 1278 bp, and codes for a hypothetical protein. Whereas, in gene *rps17*, one intron like element was observed which is 3 bp long having the sequence 5'-GTG-3'. But *pet39* has two such intron-like elements, both of 3 bp length, with the sequence AGG and ATA (both reads from 5' to 3'). Also in the gene *psbV*, two such intron-like element were present, their sequence were

CTT and CTC (from 5' to 3').

The comparative analysis of plastid genome from the above three species of *Cryptophyta* shows the presence of large number of extra genes in their cpGenome. This may be due to incomplete transfer of their genes into the main DNA genome. But remarkably, it was found that, genes of photosystem I and II are absent in the *Cryptomonas*, whereas, in *Guillardia* and *Rhodomonas*, 12 genes were found to be present. This is in excess to the number of genes found in higher plants like tobacco, rice, etc. where only 4 genes are found to be present. Among this two species of *Cryptophyta*, gene length is almost equal except for 3 genes - in which they are longer in *Rhodomonas* (Table 1). Similarly, 4 extra genes of photosystem was found to occur in these two species compared to higher plants which includes *psbV*, *psbW*, *psbX* and *psbY*. One of the gene *psbC* of *Guillardia* is longer than that of *Rhodomonas* by 36 bp.

Regarding genes of Cytochrome b6/F, only one was found in *Cryptomonas*, 7 in *Guillardia* and 8 in *Rhodomonas*. In higher plants there are 6 genes of this set. The extra ones are *petF* and *petM*, later is also absent in *Guillardia* and present only in *Rhodomonas*. Except for *petA* which is longer in *Guillardia*, other genes of this set are identical in length, indicating their precise conservance in the evolutionary time scale. Though *petB* and *petD* genes of tobacco plant contain introns, no such features were found to be present in the members of *Cryptophyta*.

Among genes of ATP Synthase, 8 members were present in *Guillardia* and *Rhodomonas*, where as in *Cryptomonas*, *atpF* was absent. In contrast to this, tobacco contains only 6 genes - *atpD* and *atpG* being absent in their cpGenome. One gene of chlorophyll synthesis was found to be present, i.e., *chlI*, which was longest in *Guillardia* (1062 bp) followed by *Rhodomonas* (1059 bp) and was shortest in *Cryptomonas* where only 1002 bp was present. 4 genes of chloroplast division were also noted in their cpGenome except for *Cryptomonas* in which their presence was not remarkably noted. This includes *minE*, *minD*, *frbB* and *ftsH*. Among these 4 genes *ftsH* was longest with more than 1800 bp, where as, *minE* was around 270 bp and *frbB* was little more than 300 bp, *minD* was within 800 bp. Remarkably, both the genes for RUBISCO biosynthesis were present in the cpGenome of all three species. Normally in higher plants, gene for smaller sub-unit of this enzyme is encoded by nuclear gene, as it is not present in plastid genome. In the entire 3 species, gene for larger sub-unit of RUBISCO *rbcL* was 1467 bp and smaller sub-unit's gene *rbcS* was 420 bp, indicating the strong conservation in the sequence of these genes and their importance in translation machinery.

Interestingly, no genes of NADH Oxidoreductase system were found to be present in the cpGenomes of all

three species of *Cryptophyta*.

All the three species contains genes for translation machinery which includes those of r- and t-RNAs, translation factors and ribosomal proteins. Except *Cryptomonas*, other two species had 2 copies of each of 5S, 16S and 23S rRNA genes. In *Guillardia* one of the 5S genes was longer than the other due to one extra base, and in case of gene for 23S, there were extra 13 bases in one of the gene. Also, in case of *Rhodomonas*, genes of 5S rRNAs were of equal length, one of the genes of 16S was longer than other by 2 bp, and by 1 bp in case of 23S.

Among 3 genes of translation factors, none of them were found to be conserved in length among the three species. *tsf* was 620 bp in *Cryptomonas* but 660 bp in *Guillardia* and *Rhodomonas*. Similarly, *infB* was 1965 bp in *Cryptomonas*, whereas, it was 2208 and 2256 in *Guillardia* and *Rhodomonas*. Though *tufA* was 1230 bp in *Rhodomonas*, it was 1227 bp in both *Guillardia* and *Cryptomonas*.

Eighteen genes of ribosomal proteins of smaller subunit were found to be present except in *Cryptomonas* where *rps6* was absent and making total gene number of this group to be 17 only. In comparison to higher plants, additional 7 genes were found to be present - *rps5*, *rps6*, *rps9*, *rps10*, *rps13*, *rps17* and *rps20*. *rps17* in *Rhodomonas* is interrupted by 3 bp intervening sequence GTG, which is not present in other two species. Similarly, 26 genes of ribosomal proteins of larger subunit were found, though *rpl32* was absent in *Cryptomonas*. In higher plants, only 9 genes of this group are present in the cpGenome, the additional genes in *Cryptophyta*, here also, suggest the transfer of plastidal genes to nuclear genome was yet to accomplish. In both group of genes, sizes are variable in all three species, suggesting the frail conservation among this orthologs.

Regarding the genes of tRNAs, they were of more or less conserved in their base sequence as well as in their length among the three species. Gene of Serine tRNA was longest with 89, 88 and 87 base pairs where as that of Gly, Lys, Asn, Gln and Val were shortest among the group with length of 72 bp. Regarding their numbers in individual species, Leucine had 3, Methionine had 3 (2 in *Guillardia*), Serine had 3, Arginine had 3 and Glycine had 2 tRNA genes. All other amino acids had single tRNA genes, except, Isoleucine having 3 in *Guillardia* and 2 in *Rhodomonas*.

Only 5 conserved genes were present in all the three species suggesting their strong requirement for the function of photosynthesis. Other 12 genes were absent in *Cryptomonas* but present in *Guillardia* and 11 in *Rhodomonas*, suggesting their late appearance in the photosynthetic system, as well as, start of the complexity of genomes and inclusion of addition functions in the metabolism of evolving organisms. Interestingly, in

*Rhodomonas*, *ycf39* had two intron like elements of 3 bp each, suggesting the appearance of introns in the genes starts from short sequences like these. Beside these genes, as with other plastid genome, there were some ORFs with unassigned function in all the three species of *Cryptophyta*.

*Cyanobacteria* are regarded as the progenitors of higher plant chloroplasts (Gray 1992; Chatterjee *et al.*, 2004). May be due to this reason most of the genes of the chloroplast are devoid of any intronic sequences as opposed to the nuclear genes. The presence of the intronic genes may be due to the promiscuous DNA. Since, the genome analysis and their in-depth study is inseparable content of the evolutionary biology and genetics, such comparisons and findings may further expand our knowledge of DNA functions and expression of genes. Chloroplast, though, is a small organelle found in the plants, encodes its own genes - some of which are independent of the nuclear control and some are under the control of the latter. The presence of its own DNA may be because of the invasion of the non-photosynthetic cells by the independent cyanobacteria and modification of the same to the present chloroplast. Whatever may be the reason of its presence in the plant cell; it has made plants the sole authority as the "producers" in the ecosystem.

## REFERENCES

- Chatterjee A., Majee M., Ghosh, S. and Majumder, A.L. (2004). sIII722, an unassigned open reading frame of *Synechocystis* PCC 6803, codes for L-myo-inositol-1-phosphate synthase. *Planta*, **218**: 989-998.
- Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R. and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research*. **36**(19): E122.
- Benson, D.A., Karsch-Mizrachi, I., Clark K., Lipman, D. J., Ostell, J. and Sayers, E.W. (2012). GenBank. *Nucleic Acids Research* **40**: D48-D53, Database issue (2012).
- Donaher, N., Tanifuji, G., Onodera, N.T., Malfatti, S.A., Chain, P.S., Hara, Y. and Archibald, J.M. (2009). The complete plastid genome sequence of the secondarily nonphotosynthetic alga *Cryptomonas paramecium*: reduction, compaction, and accelerated evolutionary rate. *Genome Biol. Evol.*, **1**: 439-448.
- Douglas, S.E. and Penny, S.L. (1999). The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J. Mol. Evol.*, **48**(2): 236-244.
- Gray, M.W. (1992). The endosymbiotic hypothesis revisited. *International Review on Cytology*, **141**: 233-357.
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.*, **217**(2-3): 185-194.
- Khan, H., Parks, N., Kozera, C., Curtis, B.A., Parsons, B.J., Bowman, S. and Archibald, J.M. (2007). Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol. Biol. Evol.*, **24**(8): 1832-1842.
- Morton B.R. and Clegg M.T. (1993). A Chloroplast DNA mutational hotspot and gene conversion in a non-coding region near *rbcL* in the grass family (Poaceae). *Current Genetics*, **24**(4): 357-365.
- NCBI GenBank Home page. <http://www.ncbi.nlm.nih.gov/genomes>
- Ohyama, K., Fukazawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, **322**: 572-574.
- Robbens, S., Derelle, E., Ferraz, C., Wuyts, J., Moreau, H. & Deepier Y.V. (2007). The Complete Chloroplast and Mitochondrial DNA Sequence of *Ostreococcus tauri*: Organelle Genomes of the Smallest Eukaryote are Example of Compaction. *Mol. Biol. Evol.*, **24**(4): 956-968.
- Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Federhen, S., Feolo, M., Fingerhman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Krasnov, S., Landsman, D., Lipman, D.J., Lu, Zhiyong, Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Karsh-Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D. (2012). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **40**: D13-D25.
- Shimada, H. and Sugiura, M. (1991). Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic Acids Research*, **19**(5): 983-995.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matshubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal*, **5**: 2043-2049.