Journal of Applied Biology & Biotechnology Vol. 2 (03), pp. 021-026, May-June, 2014
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2014.2305

Comparative and Bioinformatics Analyses of the Solanaceae Chloroplast Genomes: Plastome Organization is more or Less Conserved at Family Level

Dipen Ghimiray* and Binod Chandra Sharma

Post Graduate Department of Botany, Darjeeling Government College, Lebong Cart Road, Darjeeling – 734101, West Bengal, India.

ARTICLE INFO

Article history:
Received on: 09/04/2014
Revised on: 01/05/2014
Accepted on: 19/05/2014
Available online: 27/06/2014

***Key words*:**
Solanaceae, *Nicotiana*, *Solanum*, Chloroplast genome, GenBank, Comparative analysis

ABSTRACT

Chloroplast genome of the seven members of the family solanaceae – *Nicotiana tabacum*, *N. sylvestris*, *N. tomentosiformis*, *N. undulate*, *Solanum tuberosum*, *S. lycopersicum*, and *S. bulbocastanum* were extracted from NCBI Gene Bank and were subjected to the bioinformatics study. In this study we found that the genes and amino acid sequences in the proteins were almost identical and conserved. Even the structure and sequences of introns were more or less conserved. Certain variations were observed in the nature and occurrence of open reading frames, which may be due to species specific genome. Some variations were also observed when compared these dicot genomes with the monocot genome of rice. But in contrast to some primitive species such as members of cryptophyta, there were fewer numbers of genes which may be due to the transfer of some chloroplast genes into the nuclear genome.

1. INTRODUCTION

Solanaceae is an important family of dicots, which includes important vegetable crops like potato, tomato, brinjal, chillies; medicinal plants such as datura, atropa; and garden plants like *Cestrum*, *Brunfelsia*, *Browallia*, tobacco, wood tobacco, etc. among many others. Chloroplast genomes of eight species from this family have been sequenced[1,2,3,4] and are available in the public databank such as NCBI, EMBL and DDBJ. Chloroplast genome on an average consists of 150-160 kb DNA in which nearly there are 90 protein encoding genes and 30-40 tRNA genes [1,5,6,7,8], therefore it is a short, double stranded and circular molecule. These genes are expressed inside the chloroplast itself using the available mechanisms, although the necessary amino acids for protein synthesis are imported from cytoplasm only because the chloroplast genomes lack the enzyme coding genes of amino acid biosynthesis pathways. Till today, over 360 sequenced chloroplast genomes of different species are available (NCBI GenBank Data, August 2013) which has provided us the opportunity for undertaking various bioinformatics study like comparative genome analysis [5,7,8,9,10], evolutionary history [11,12,13,14,15,16], promoter gene study [17], etc. In this study, gene bank available chloroplast genomes of seven species of solanaceae were analyzed considering three gene parameters - sequences, length and arrangement among the plastomes of this family.

2. MATERIALS AND METHODS

Sequenced chloroplast genome of seven species of Solanaceae – *Nicotiana tabacum* (NC_007602) [1], *N. sylvestris* (NC_007500)[2], *N. tomentosiformis* (NC_007602) [2], *N. undulate* (NC_016068) [3], *Solanum tuberosum* (NC_008096) [18], *S. lycopersicum* (NC_007898) [4], and *S. bulbocastanum* (NC_007943) [4] were extracted from the NCBI webpage [19].

2.1 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, as reported earlier.

2.2 Determination of length

As the positions of the genes were described in their sequence, gene lengths were derived by subtracting the initial position from final position and adding 1 to the result, *i.e.*, by the formula: Gene length = (final position - initial position) + 1, the reason of adding the value of 1 is for considering the translation from the first Open Reading Frame (ORF).

** Corresponding Author*
Post Graduate Department of Botany, Darjeeling Government College, Lebong Cart Road, Darjeeling, India. Phone & Fax : +91 354 2254019, Email: dipen23ray@gmail.com;

2.3 Gene Inspection

This was done by manual observation and comparative analysis of the seven cpGenomes.

3.RESULTS

Comparative study of the individual gene group from these species showed the following results.

3.1 Genes of photosystem I [psa]

In all the seven species of solanaceae, a total four genes of photosystem I was found to be present. Among all species the length of each gene was constant. The longest gene from the group is *psaA* (2253bp) followed by *psaB* (2205 bp). Shortest gene among this group is *psaI* with 111 bp length.

3.2 Genes of photosystem II [psb]

There are fifteen genes of photosystem II in solanaceae. Their number is constant in all species under study, except in *Solanum bulbocastanum* which has a bit longer *psbC* gene (1422 bp). Longest of the group is *psbB* with the length of 1527 bp, followed by *psbC* with 1386 bp. Shortest gene from this paralog is *psbM* and *psbT*, both with the length of 105 bp. Among the genes of this group *psbL* undergoes the RNA editing after the transcription and in gene *psbI* start codon is AUC instead of conventional AUG. Again, there are three clustered groups – first, cluster of *psbD* and *C*; second, cluster of *psbJ, L, F* and *E*; and the third cluster consists of *psbB, psbT, psbN* and *psbH*. All these clustered genes may have their common promoter for transcription. Multiple sequence alignment of the upstream regions of these three clusters shows the conserved base composition in all seven species, advocating the polycistronicity among these genes.

3.3 Genes of cytochrome B6/F [pet]

Total of six cytochrome B6/F genes were found to be present in the plastomes of solanaceae. Longest cytochromeB6/F gene is *petB* (including intron length) with the length of 1401 bp and shortest is the *petN* (90 bp only) and *petL* (96 bp). Among these six genes, *petB* and *petD* have single intron in them. For the *petB* it is 754 bp and in *petD* it is 743 bp. These lengths are more or less conserved in both the genus *Nicotiana* and *Solanum*. Excluding the intron length, longest gene would be *petA* (963 bp).

3.4 Genes of ATP-Synthase [atp]

Lengths of the genes of this group among the members of family solanaceae were found to be constant. There are total six genes encoded by chloroplast genome for ATP-Synthase. Including the intron length, the longest gene is the *atpF* (1249 bp) but excluding the intron, it would be *atpA* with 1524 bp as the longest gene. Shortest gene among the group is the *atpE* with the of length 402 bp.

3.5 Genes of RUBISCO [rbc]

As with the most chloroplast genome, in this family also only one gene *rbcL* was found to be present for the enzyme RUBISCO. This is the larger subunit encoding gene, whereas, smaller subunit of RUBISCO is encoded by nuclear genome. Length of the *rbcL* was found to be 1434 bp. In all the seven species under study, this length was conserved.

3.6 Genes of RNA polymerase [rpo]

As usual, four different genes encoding the subunits of RNA polymerase are found to be present in the plastomes of all seven species of solanaceae. Longest gene is the *rpoC2* (4167 bp), followed by *rpoB*, then *rpoC1* with the length of 2804 (including intron length), whilst, the shortest of the group is the *rpoA* (1014 bp).

3.7 Genes of NADH Oxido-reductase [ndh]

Genes of NADH-Oxidoreductase are designated as "**ndh**" followed by the respective alphabetical designation. There are total **twelve** (12) genes encoded as – *ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndfJ* and *ndhK*. Of these *ndhB* is present in **two** copies, one in the inverted repeat A and another copy is in inverted repeat B region. Again, among these, *ndhA* and *ndhB* contains intron. The total number and size of these genes are constant among the seven species of solanaceae under investigation.

There are two clusters in the location position of the ndh genes. One is the cluster of *ndhJ, ndhK* and *ndhC*. Second cluster consists of *ndhE, ndhG, ndhI, ndhA* and *ndhH*. Three genes *ndhF, ndhB* and *ndhD* do not belong to any cluster.

Out of 12 genes under this group, the longest gene is *ndhA* which is 2240 bp, followed by *ndhE* with the length of 2223 bp, *ndhB* with length of 2212 bp and *ndhD* with the length of 1503 bp. Amongst the shortest gene in this category *ndhE* which is just 306 bp followed by *ndhD* (363 bp).

3.8 Genes of rRNAs

There are four genes which encode rRNAs in the chloroplast genome of the solanaceae which are located in the inverted repeat regions, thus, there are total eight genes - two copies of each gene in the two repeat regions. Out of these genes, those encoding 23S, 4.5S and 5S are clustered together and 16S is isolated.

The presence of cluster may indicate that they are operating under the single promoter. The length of these genes is more or less constant in all the seven species under study, of which, the length of 4.5S gene is 103 bp and that of 5S gene is 121 bp in all the seven species. Minor deviation is observed at genus level, 16S rRNA gene is 1500 bp in members of *Solanum*, while in the members of *Nicotiana* the length of the same gene is 1491 bp. Again, gene length of 23S rRNA of *Solanum tuberosum* is 2809 bp in place of 2810 bp in other members of the family.

3.9 Genes of tRNA

The number of tRNA genes varied among the seven species of solanaceae. In *Nicotiana tabacum* there are total 37 tRNA genes. Among these genes, seven genes possess single intron in each. These intronic genes are – trnK-*UUU, trnG-UCC, trnL-UAA, trnV-UAC, trnI-GAU* (inverted repeat B), *trnA-UGC* and *trnI-GAU* (inverted repeat A). All other genes do not have introns. It was also observed that, genes of *tRNA-D, tRNA-Y, tRNA-E* and *tRNA-T* are present in one cluster.

Regarding the total number of tRNA genes, highest number is found in the *Solanum tuberosum* and *S. bulbocastanum*, 53 each; while lowest number (45) were observed in *Nicotiana tabacum* and *N. undulate* (Table 1).

Table. 1: Total number of genes in the seven species of Solanaceae.

Member	Protein Encoding Genes	Rna Encoding Genes
Nicotiana sylvestris	101	46
N. tabacum	98	45
N. tomentosiformis	102	46
N. undulate	110	45
Solanum bulbocastanum	85	53
S. lycopersicum	84	46
S. tuberosum	84	53

3.10 Genes of Translation Initiation Factor

There is one pseudo gene for the translational initiation factor in chloroplast genomes of all the studied members of solanaceae. It is designated as ***inf***, normally it is 308 bp long encoding amino acid of 105 bp. It's AT composition is 65% and GC composition is 35%, giving the high AT/GC ratio of 1.8.

3.11 Other genes

In other genes category, three genes have been placed – *matK, accD* and *clpP* which are present one copy per genome. Out of these three, *clpP* contains two introns thus making this gene as the one of the only two genes, which has three exons. Another three-exonic gene is *ycf3* described later (genes encoding conserved proteins). The length of these genes were found to be constant in all the studied members – *matK* (1530 bp), *accD* (1539 bp) and *clpP* (2035 bp) including the intron length. Of these, the presence of two introns in *clpP* is unique among the plastome genes; the only another gene with two introns is *ycf3*.

3.12 Genes encoding larger subunit of ribosome:

The genes encoding the larger subunit of ribosome are designated as "**rpl**". There are total eleven (11) genes in this group. These genes are designated as – *rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33* and *rpl36*. Among these, *rpl2* and *rpl23* is present twice in the genome. Again, *rpl2* and *rpl16* contains one intron each. Longest gene among this group is *rps2* (825 bp) while shortest gene is the rps36 (125 bp).

3.13 Genes encoding smaller subunit of ribosome:

The genes encoding the smaller subunit of ribosome are designated as "**rps**". There are thirteen genes designated as – *rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18* and *rps19*. Among these, *rps7* is present in two copies; whilst, *rps2* is a *trans*-splicing gene. In fact, *rps2* is the only *trans*-splicing gene in the chloroplast genome of solanaceae. Longest of this group is *rps16* (1117 bp including the intron length), without intron the longest is *rps2* (711 bp), and shortest gene among rps is *rps15* (264 bp).

Among this group of genes, rps12 is a unique *trans*-splicing gene. It has two parts – first, two exons separated by one intron; and, second, single exon which joins in *trans* with the first exons. Therefore, this gene in all the seven species of solanaceae consists of three exons each 26bp, 232 bp and 114 bp long. The *cis*-intron is 536 bp long and this size is more or less constant in this family.

3.14 Genes encoding conserved proteins

These genes are designated by "**ycf**". There are total nine (9) genes in this category. These includes – *ycf1, ycf2, ycf3, ycf4, ycf5 (=ccsA)* and *ycf10 (=cemA)*. Now, *ycf15* is also added in this list, which is located just after *ycf2*. This, *ycf2* and *ycf15* are present in two copies in the plastome. Among these genes, *ycf3* possesses three exons and two introns which are common in all the solanaceae members under the study.

3.15 Open Reading Frames

The number of potential open reading frames (also referred as ORF) varies among the seven members of solanaceae (**Table 2**). *Nicotiana tabacum* possess 14 ORFs designated as – ORF105, ORF74, ORF70A, ORF99, ORF103, ORF79, ORF131, ORF70B, ORF75 and ORF350 out of which four genes ORF75, ORF131, ORF70B and ORF79 are present in two copies. In the ORF75, cryptic gene for *trn-N* is also present, whereas in ORF70, start codon is GUG instead of AUG.

Table. 2: Different ORFs in the cpGenomes of Solanaceae.

Sl No.	*N. tabacum*	*N. sylvestris*	*N. tomentosiformis*	*N. undulate*	Location of ORFs
1	---	ORF63	ORF90	ORF90	Between *psbI* and *trnS*
2	---	ORF70C	ORF70C	ORF70C	Between *trnC* and *petN*
3	ORF105	ORF105	ORF74	ORF56	
4	---	ORF71A	ORF71A	ORF71A	Between *accD* and *psaI*
5	---	ORF75	---	---	Between *trnR* and *trnN*
6	---	---	ORF73	---	Before *rps4*
7	---	---	ORF71B	71B	Between *trnI* and *trnA*
8	---	---	ORF75	---	Before *trnN*
9	ORF350	---	ORF338	ORF339	
10	---	---	---	ORF79B	Cryptic within *rpl16*
11	---	---	---	ORF115	
12	---	---	---	ORF92	
13	---	---	---	ORF71	Cryptic in *trnI*
14	---	---	---	ORF75	Cryptic within *trnR* and *trnN*

In *Nicotiana sylvestris* there are four additional open reading frames which are ORFs 63, 70C, 71A and 75. Extra ORF63 is present in between *psbI* and *trnS*, ORF70C is located between *trnC* and *petN*. Similarly ORF71A is present between *accD* and *psaI*.

Again, in *Nicotiana tomentosiformis* and *Nicotiana undulate* ORF90 is present instead of ORF63. ORF105 of *Nicotiana tabacum* remains same in *N. sylvestris* but in *N. tomentosiformis* it is replaced by ORF74 and in *N. undulate* it becomes ORF57. ORF71A which is located between *accD* and *psaI*, is present in *N. sylvestris, N. tomentosiformis* and *N. undulate* but it is absent in the plastome of *N. tabacum.*

In this way, some of the unique ORFs present in *N. tomentosiformis* are – ORF73, ORF71B, ORF75 and ORF338; similar unique ORFs of *N. undulate* are – ORF339, ORF79B (which is cryptic within *rpl16*), ORF115, ORF92, ORF71B (cryptic within *trnI-GAU*) and ORF75 (cryptic in *trnR* and *trnN-GUU*).

Open reading frames in the species of *Solanum* are not mentioned, may be they are absent.

3.16 Introns in Protein encoding genes of cpGenome

There are total sixteen introns which are – *rps16, rpl2 (I), rpl2 (II), rpl16, ndhB (I), ndhB (II), ndhA, petB, petD, atpF, rpoC1, ycf3* (*1^{st}*), *ycf3* (*2^{nd}*), *clpP* (*1^{st}*), *clpP* (*2^{nd}*) and *rps12* (Table 3). Among these; since *ndhB* and *rpl2* are present in two copies, two different introns are present for these genes; and also for there are two introns in between the three exons in *ycf3* and *clpP*, these also have two different introns. Analysis of introns in tRNA genes was not done in this study. The exon-intron boundary between these introns and their respective genes are presented in Table 3. The sequences present are the consensus sequences which were obtained after comparing each gene and its intron of the 7 species under study. The length of the different introns present are as follows – intron in *rps12* is 536 nucleotide; *rps16* is 860; *rpl2* (I) is 394; *rpl2* (II) is 666; *rpoC1* is 737; *atpF* is 695; *ycf3* (1st intron) is 783; *ycf3* (2nd intron) is 738; *clpP* (1st intron) is 637 and *clpP* (2nd intron) is 807; *petB* is 753; *petD* is 742; *ndhB* (I) is 678; *ndhB* (II) is 679 and intron of *ndhA* is 1148 nucleotides long. Therefore, the longest intron among protein encoding genes is in *ndhA* with 1148 nucleotides followed by intron of *rpl16* (1020 nucleotides). Shortest intron is that of first copy of *rpl2* which is 394 nucleotides long. Genes which are present in two copies, their introns' length also vary.

Table. 3: Introns of protein encoding gene in solanaceae cpGenomes.

Gene	9 bp Border sequences of each side of Introns (From 5' to 3' : AUG-exon1-INTRON-exon2-STOP)
rps16	5'- GTGCGACTT ..... TATCCCAAT – 3'
rpl2 (I)	GCGGTTTGA ..... ACTTCAACC
rpl2 (II)	GTGCGGTTT ..... CTACTTCAA
rpl16	GTGTGACTC ..... AACTATAAC
ndhB (I)	GTGCGGTTC ..... CGACTCTGA
ndhB (II)	GTGCGGTTC ..... GACTCTGAC
ndhA	CATTTGGGT ..... GACTATGAT
petB	GTGTGACTT ..... TATCTCAAT
petD	GTGTGACTT ..... TATCCCAAT
atpF	GTGCGAGTT ..... TACTTTCAT
rpoC1	GTGTGATTT ..... TATCCCAAT
ycf3 (1st intron)	GTGTGATTT ..... GACGCTAAC
ycf3 (2nd intron)	AAATAAAAA ..... TATTCCGGC
clpP (1st intron)	TCAGATATA ..... TATATCATC
clpP (2nd intron)	TGGCGCCAA ..... TACCCTAAT
rps12 (*cis*-intron)	GTGCGTTGT ..... CCACCCTAC

4.DISCUSSIONS

Comparative and bioinformatics analysis of these seven members of solanaceae family shows the variable number of genes in their chloroplast genome; but their composition in terms of base sequence, AT-GC ratio, amino acid sequences in proteins, etc. show conserved pattern in their organelle genome with very low diversion in the protein and RNA encoding genes. Intergenic distances and sequences also show some conserved pattern but there are also some variations among these seven species. There are total 101 protein encoding genes in *Nicotiana sylvestris*, 98 in *N. tabacum*, 102 in *N. tomentosiformis* and 110 genes are present in *N. undulate*. From the genus Solanaceae, *S. bulbocastanum* has 84 protein encoding genes, whereas, *S. tuberosum* and *S. lycopersicum* has 84 protein encoding genes each. Considering total number of RNA encoding genes, highest number was observed in *Solanum bulbocastanum* and *S. tuberosum* with 53 genes in each species. *S. lycopersicum* has only 46 RNA genes. Similar numbers of genes were also found to be present in the *N. sylvestris* and *N. tomentosiformis* (46 each). *N. tabacum* and *N. undulate* has 45 RNA genes.

The lengths of two exons in petB in all the seven species were found to be same – 6 and 642 bp. The presence of such short exon of 6 bp is encountered in the organelle genome only and so far such short exons are not reported from bacterial, fungal, plant and animal genomes. But, length of the smaller exon in case of *petD* is variable, for example, in *S. bulbocastanum* it is 6 bp whereas in rest six species under study, its length is 8 bp. To compensate this, in the former there is 2 bp long second exon with 477 bp, whereas, in other six species, its length is only 475 bp.

16S rRNA genes in *Solanum* are little longer than *Nicotiana*. In the former it is 1500 bp, whereas, in latter its length is only 1491 bp. Similarly 23S gene in *S. tuberosum* is 1 bp shorter than the other members (2809 instead of 2810). Such variation is also obtained in the *rpoC2* gene, which in the species of *Solanum* shows length of 4179 bp whereas species of *Nicotiana* length of this is 4167 bp, thus there is difference of 12 bp. Regarding the hypothetical conserved gene *ycf1*, in all the three species of *Solanum*, there are additional and truncated copies of this gene but all of these are pseudo genes. Same trend is also observed in the three species of *Nicotiana* except *N. tabacum*. But the genes of photosystem I, cytochrome b6/F, ATP synthase, RUBISCO, NADH oxidoreductase and ribosomal proteins are strongly conserved in the species of solanaceae.

In the chloroplast genomes of three species of *Solanum*, traditional open reading frames are not mentioned, *i.e.*, they are absent in the genome, which also accounts for the lesser number of protein encoding genes in their genomes. However, the presence of higher number of RNA genes is due to the presence of two copies of some genes such those of *K-UUU*, *G-GGC*, *L-UAA*, *M-CAU*, *A-UGC* (in one IR region) and *I-GAU* (in one IR region); and also due to presence of some additional open reading frames.

The presence of some of the genes in clusters such as (i) cluster of tRNA genes of *D-GUC*, *Y-GUA* and *E-UUC* (since *T-GGU* is one A-strand, we considered this as a solo gene out of this cluster); (ii) *psbJ*, *psbL*, *psbF* and *psbE*; (iii) *psbB*, *psbT* and *psbH* (here also since *psbN* is on another strand, this particular gene is kept out of this cluster); (iv) *ndhD*, *ndhE*, *ndhG*, *ndhI*, *ndhA* and *ndhH*, along with the *psaC* and *rp15*; (v) there is also cluster of unrelated paralogous genes such as *rpoA*, *rps11*, *rpl36*, pseudo gene *infA*, *rps8*, *rpl14*, *rpl16*, *ORF79B*, *rps3*, *rpl22*, *rps19*, *rpl2*, *rpl23* and *trnI-CAU*; suggests the presence of polycistronic promoters in the chloroplast genomes. However, occurrence of many genes in isolated pattern suggests that monocistronic promoters are also present in the cpGenomes.

On comparing the chloroplast genomes of these dicot species with the monocot plants [we chose rice chloroplast genome (20) for this purpose], length variation was found to be quite common. Some of the genes were longer in dicot than those of monocots. Such gene includes – *psbI, psbK, psbL, psbN, ndhA, ndhB, ndhF, ndhI, ndhJ, rpl20, rpl22* and gene encoding 4.5S RNA. Even, one of the most conserved gene *rbcL* was longer in dicots. On the other hand, some groups of genes were shorter in dicots than their counterparts of monocots such as – *psbC, psbT, petB, petD, atpE, atpF, ndhJ, rpl2, rpl4, rpl16, rpl32*, and almost genes encoding smaller subunit of ribosomes, *matK*, *clpP* and gene encoding 23S RNA. Even genes for subunits of RNA polymerase such as *rpoB* and *rpoC2* were quite shorter than the monocot counterparts. Such differentiation in length might have appeared previously when monocots and dicots were separated on two different phylogenic lines. The exact reason for the occurrence of longer or shorter version of same gene in these two different groups is not precisely known, though it can be assumed that difference in anatomical structure and embryological differentiation pattern demands the specific size of proteins for these two groups. In this regard, some additional genes were found to be present in the genome of dicots which includes *psaJ, rps4, rps8* and *accD*. But in any case, proteins of photosystem I and genes encoding 16S and 5S are very much same in both monocots and dicots. Though 23 ORFs are present in monocot plants such as rice, species of *Nicotiana* shows total 14 ORFs altogether, whereas, *Solanum* does not have these ORFs as they are not mentioned in their sequence.

However, in comparison to plastid gene content in lower plant divisions such as Cryptophyta [21, 22, 23] we observed that orthologous gene content of higher plant (especially this family) is quite lower. Many genes of chloroplast genomes have been transferred into the nuclear genome in solanaceae and other higher group of plants. For example – genes of photosystem I *psaD, psaE, psaF, psaK, psaL* and *psaM*; genes of photosystem II *psbW, psbX* and *psbY*; genes of cytochromes like *petF and petM*; genes of ATP synthase *atpG*; *rbcS*; genes encoding larger and smaller subunits like *rpl1, rpl3, rpl5, rpl6, rpl12, rpl13, rpl18, rpl21, rpl23, rpl29, rpl34* and *rps5, rps6, rps9, rps13, rps20*; chloroplast division genes like *minE, minD, ftrB* and *ftsH*; conserved genes like *ycf12, ycf16, ycf17, ycf20, ycf24, ycf29, ycf34, ycf33, ycf35, ycf36, ycf37, ycf39, ycf46, ycf61* and *ycf65*; miscellaneous genes such as *clpC, dnaB, groEL, rbcR, ccs1, tsf, infB, tatC, dnaK, pbsA, acpP, secY, ilvB, chlI, cpcB, rne, cbbX, hlp, cpeB,* etc; which are present in the organelle genome of species of Cryptophyta [8] are not present in the chloroplast genome of solanaceae may be because they are transferred into nucleus or they have been deleted from the genome as they are not essential.

On the other hand, large numbers of introns are present among the genes of higher plants. Intron like element of only 3 bp long occurs in Cryptophyta genes like *rps17, pet39* and *psbV*; though *psbN* contains 2235 bp long intron including a cryptic potential ORF within it of length 1278 bp. But, genes of higher plants like members of solanaceae and other families show many introns especially in the genes such as – *petB, petD, atpF, ndhA, ndhB, ycf3, rpoC1, rpl2, tpl16, rps12* and *clpP.*

The comparative analysis of chloroplast genome from the seven species of the family solanaceae reveals that genes and amino acid sequence in the proteins are conserved and they have not undergone major diversion. This may be due to the fact that they are related closely as members of same genus or species under the same family (Solanaceae) or otherwise they may be withstanding the evolution and mutation pressure successfully.

5.ACKNOWLEDGEMENTS

The authors are grateful to the Department of Science and Technology, Govt. of India for providing infrastructural facilities under DST-FIST programme.

6. REFERENCES

1. 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 BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H and Sugiura M. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO Journal. 1986; 5: 2043-2049.

2 Yukawa M, Tsudzuki T and Sugiura M. The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum.* Molecular Genetics and Genomics. 2006; 275 (4):367-373.

3 Thyssen G, Svab Z and Maliga P. NCBI Genome Project, NIH, Bethesda, USA, 2011.

4 Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J and Jansen RK. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. Theoretical and Applied Genetics. 2006; 112 (8): 1503-1518.

5 Shimada H and Sugiura M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Research. 1991; 19 (5): 983-995.

6 Sato S, Nakamura Y, Kaneko T, Asamizu E and Tabata S. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Research. 1999; 6 (5): 283-290.

7 Robbens S, Derelle E, Ferraz C, Wuyts J, Moreau H & Deeper YV. The Complete Chloroplast and Mitochondrial DNA Sequence of Ostreococcus tauri: Organelle Genomes of the Smallest Eukaryote are Example of Compaction. Molecular Biology and Evolution. 2007; 24(4): 956-968.

8 Sharma BC and Ghimiray D. Analysis of the Cryptophyta chloroplast genome reveals presence of additional genes and absence of introns in their genome. Biological Forum 2012; 4(2): 1-7.

9 Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, Tong W, Wang J, Wang J, Yu J, Yang H and Zhu L. A comparison of rice chloroplast genomes. Plant Physiology. 2004; 135 (1):412-420.

10 Shanker A. Chloroplast genomes of bryophytes: A review. Archive for Bryology. 2012; 143:1-5.

11 Turmel M, Otis C and Lemieux C. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. Proc. Natl. Acad. Sci. U.S.A. 1999; 96 (18): 10248-10253.

12 Turmel M, Otis C and Lemieux C. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc. Natl. Acad. Sci. U.S.A. 2002; 99 (17): 11275-11280.

13 Turmel M, Otis C and Lemieux C. The complete chloroplast DNA sequences of the charophycean green algae *Staurastrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. BMC Biology. 2005, **3**:22 doi:10.1186/1741-7007-3-22.

14 Turmel M, Otis C and Lemieux C. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. Molecular Biology and Evolution. 2006; 23 (6):1324-1338.

15 Turmel M, Gagnon MC, O'Kelly CJ, Otis C and Lemieux C. The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix* and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. Molecular Biology and Evolution. 2009; 26 (3): 631-648.

16 Turmel M, Otis C and Lemieux C. The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri,* and *Oocystis solitaria* reveal a shared ancestry between the Pedinomonadales and Chlorellales. Molecular Biology and Evolution. 2009; 26 (10): 2317-2331.

17 Lyubetsky VA, Rubanov LI and Seliverstov AV. Lack of conservation of bacterial type promoters in plastids of Streptophyta. Biology Direct. 2010; 5:34 doi:10.1186/1745-6150-5-34.

18 Gargano D, Vezzi A, Scotti N, Gray JC, Valle G, Grillo S and Cardi T. The complete nucleotide sequence of potato (*Solanum tuberosum* cv. Desiree) chloroplast DNA. *In* Abstracts: 2nd Solanaceae Genome Workshop 2005: 107

19 NCBI GenBank, USA. http://www.ncbi.nlm.nih.gov/genomes

20 Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K & Sugiura M. 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. Molecular and General Genetics. 1989; 217 (2-3): 185-194.

21 Douglas SE and Penny SL. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. Journal of Molecular Evolution. 1999; 48 (2) :236-244.

22 Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S and Archibald JM. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Molecular Biology and Evolution. 2007; 24 (8): 1832- 1842.

23 Donaher N, Tanifuji G, Onodera NT, Malfatti SA, Chain PS, Hara Y and Archibald JM. The complete plastid genome sequence of the secondarily nonphotosynthetic alga *Cryptomonas paramecium*: reduction, compaction, and accelerated evolutionary rate. Genome Biology and Evolution. 2009; 1: 439-448.

How to cite this article:

Dipen Ghimiray and Binod Chandra Sharma., Comparative and Bioinformatics Analyses of the Solanaceae Chloroplast Genomes: Plastome Organization is more or Less Conserved At Family Level. J App Biol Biotech. 2014; 2 (03): 021-026.