J. Cotton Res. Dev. **32** (2) 167-176 (July, 2018) **ISSN No. 0972-8619**

A comparative plastomes analysis in *Gossypium* species : Gene loss, intron loss and inverted SSC region

P.I. JADEJA, P.P. BHATT AND V.S. THAKER*

Department of Biosciences, Saurashtra University, Rajkot- 360 005

**E-mail : thakervs@gmail.com*

ABSTRACT : Genus *Gossypium* is an important for its commercially valued fibers. The plastome engineering offers incorporation of desired gene(s) in mother plant. In this study, total 31 plastomes available in public database are studied for its structural and functional properties. The complete physical map of plastomes revealed that it is about 159 -160 kb long, circular molecule. It contains typical quadripartite structure of large single copy region (LSC) small single copy (SSC) region separated by a pair of inverted repeat regions (IRs). The comparison of plastomes in *Gossypium* species revealed notable distinctions. An inversion in SSC region of *G. anomalum*, *G. barbadense, G. bickii*, *G. herbaceum*, *G. longicalyx*, *G. stocksii*, *G. sturtianum* and *G. thurberi* was observed. In addition, gene and intron loss was also observed in few *Gossypium* species which suggest that it might have evolved before divergent. We report the key important differences in these plastomes which can be helpful to explore the possibilities at biotechnological level, phylogenic analysis and in the evolutionary studies.

Key Words : *Gossypium* genus, inter repeat regions, plastomes, single copy regions

Cotton is an important cash crop of India and worldwide for its spin able fiber and lipo-protein rich seeds. The estimated global business revenue from cotton production is estimated 500 million USD/year (Abdurakhmonov *et al.,* 2012). Today nearly 80 countries in 32-34 million ha land grow cotton for lint and oil seed production (Anonymous, 2013), even though the world demand and uses more cotton fiber than ever before. Among the 45 diploid species of genus *Gossypium* (2n=26) and six tetraploid (2n=52) species (Wang *et al.,* 2012) only two diploid *i.e. G. herbaceum* and *G. arboreum* and two tetraploid *i.e. G. hirsutum* and *G. barbadence* produce fibers of textile importance.

Due to the advent of next generation sequencing, remarkable progress in sequencing and assembly of whole cotton genomes of two diploid species, two tetraploid species and several chloroplast and mitochondrial genomes are worked out. These extensive data have helped the breeders and researchers to (i) identify the signal pathways responding to biotic and abiotic stresses (ii) identify the fiber specific genes and gene families which regulates fiber quality, and (iii) study comparative and functional genomics and evolution (Guo *et al.,* 2015; Yu *et al.,* 2015; Malik *et al.,* 2014). Thus the continuous progress in cotton genome sequencing and editing technologies, use of marker assisted selection as well as transgenic technologies, we can innovate cotton germplasms and effective breeding and agronomic practices.

The plastome (cp genome) and mitochondrial genome (mt genome) are maternally inherited and low rate of nucleotide variation in these genomes is compelling reason for evolutionary studies. Further, cytoplasmic–nuclear incompatibility is responsible for cytoplasmic male sterility (CMS) in plants (Luo *et al.,* 2013). In addition, plastome genetic engineering extend a new strategy to confer plant resistance to herbicides, insect and abiotic stress (Jin and Daniell, 2015). The plastome sequences of *G. hirsutum* and *G. barbadence* (Chen *et al.,* 2017) is reported and compared with other species of angiosperms. Subsequently, as on today, 31 plastome sequences of genus *Gossypium* is available in NCBI database, probably the largest dataset than any other genus of higher plants; however, little is known about its comparative analysis.

Considering the importance of plastome data analysis, we aimed in this study to evaluate structural and functional analysis of 31 plastomes of *Gossypium* species.

MATERIALS AND METHODS

Total 31 chloroplast genomes of *Gossypium* are available in organelle genome database of NCBI. These genome sequences (both FASTA and Genbank formats) and other parameters like genome size, GC content, number of genes, proteins, rRNAs, tRNAs and pseudo genes were obtained from NCBI and used for this study. For discovery of four chloroplast genomic regions including LSC, SSC, IRa and IRb their re annotation was performed through CpGAVAS (Liu *et al.,* 2012). Genome wide comparison, alignment, percent pairwise distance and graphical view were generated using Mauve plugin in geneious. Further annotation, intergenic region extraction and gene comparative study was performed by DOGMA. Chloroplast genome maps were obtained from CpGAVAS server. Graphical comparative view of SSC regions was obtained through Geneious.

RESULTS AND DISCUSSION

Analysis of 31 plastomes of *Gossypium* revealed following distinction

Genome organization : The general characteristic of the *Gossypium* plastomes showed a high level of similarity in overall composition of 31 platomes (Table 1). A typical quadripartite structure was observed for all genomes. The size of the genomes ranged from 159-160kb comprising of large single copy region (LSC), two identical inter repeat sequences (IRs) and small single copy region (SSC). LSC ranged from 71-72.5 kb, IRs about 25.5kb (± 200bp) each and SSC about 20-21kb. The SSC was slightly higher in *G. longicalyx* (22.478kb) and was exceptionally higher in *G. anomalum* (51,826bp) and *G. turneri* (55210 bp) due to overlap with IR regions. Thus the differences in *Gossypium* species are experienced in these regions (Table 1, Fig. 1).

GC content : GC content is a significant parameter in genomic studies. It is correlated with many cellular processes at molecular level in number of studies but not limited to; (i) mutation events (ii) numbers of miRNA binding sites (Zhang *7* 2014) (iii) recombination rate and

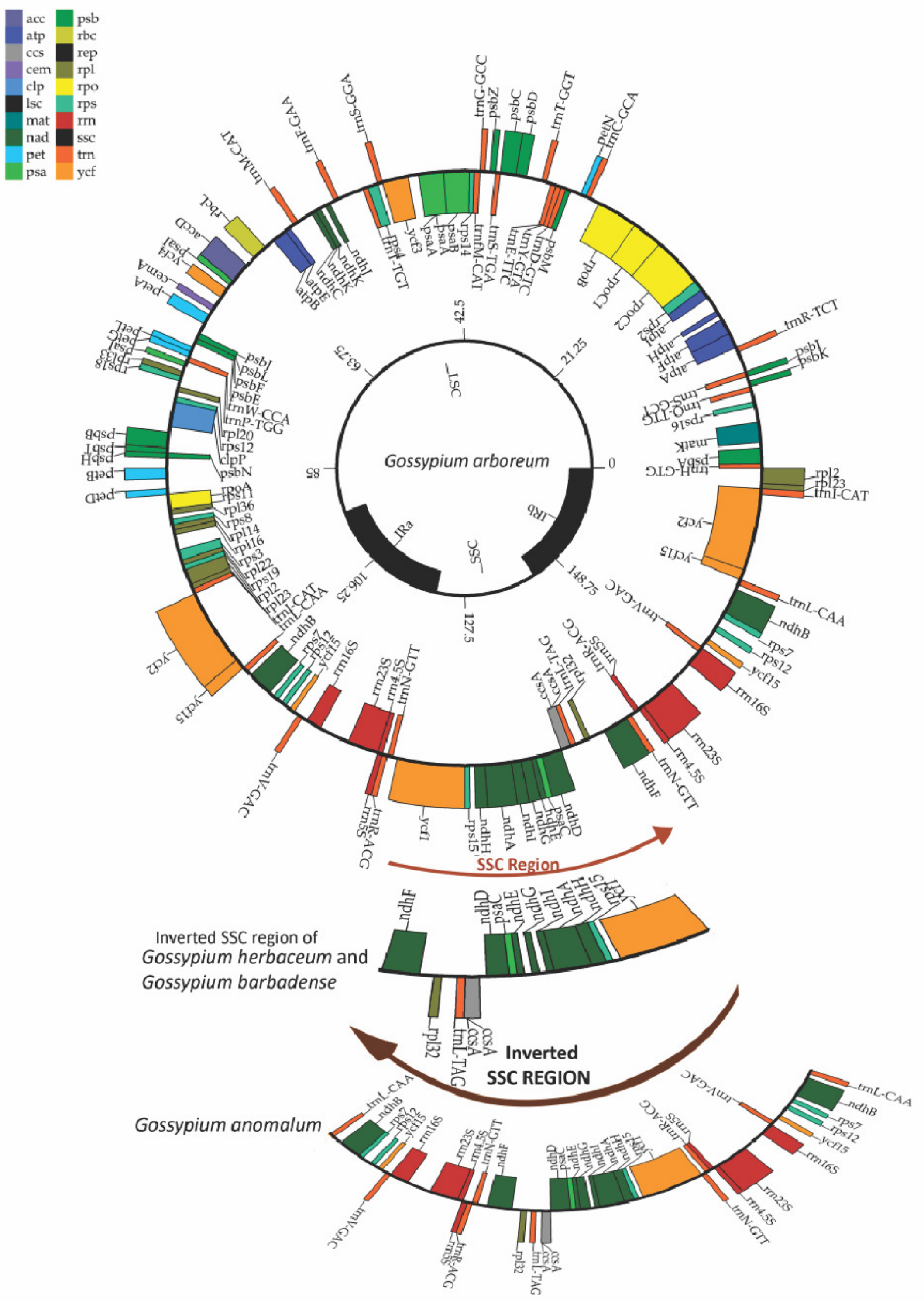

Fig. 1. A comparative diagram of the inverted SSC region in 31 plastomes; the complete circle indicates common features in genus *Gossypium*. Inverted SSC in *G. herbaceum* and *G. barbadense (*fiber producing species*)* and in *G. anomalum* with inverted and extended SSC region.

Table 1. General features of chloroplast genomes of all *Gossypium* species

No	Name	Size (bp)	GC content (%)	GC intergenic (%)	GC coding region (%)	Proteins	rRNAs	tRNAs	Genes	Pseudo gene	Intergenic region length	NCBI accession number
1	*Gossypium anomalum*	159507	37.33	33.1	38.6	86	8	37	133	2	67739	NC_023213.1
2	*Gossypium arboreum*	160230	37.23	32.7	38.5	83	8	37	128	-	68003	NC_016712.1
3	*Gossypium areysianum*	159572	37.37	33.2	38.6	83	8	37	128	-	67809	NC_018112.1
4	*Gossypium aridum*	160257	37.26	33	38.6	84	8	37	129	-	68416	NC_033396.1
5	*Gossypium armourianum*	160068	37.29	33	38.6	84	8	37	129	-	68267	NC_033400.1
6	*Gossypium australe*	159578	37.16	32.8	38.6	84	8	37	129	-	67852	NC_033401.1
7	*Gossypium barbadense*	160317	37.23	32.8	38.5	84	8	37	129	-	68257	NC_008641.1
8	*Gossypium bickii*	159422	37.2	32.8	38.3	87	8	36	132	1	67681	NC_023214.1
9	*Gossypium capitis-viridis*	159467	37.32	33.1	38.7	83	8	37	128	-	67699	NC_018111.1
10	*Gossypium darwinii*	160378	37.23	32.7	38.7	83	8	37	128	-	68155	NC_016670.1
11	*Gossypium davidsonii*	160072	37.31	33	38.7	84	8	37	129	-	68318	NC_033395.1
12	*Gossypium gossypioides*	159959	37.31	33.1	38.8	83	8	37	128	-	68180	NC_017894.1
13	*Gossypium harknessii*	160129	37.3	33	38.7	84	8	37	129	-	68342	NC_033333.1
14	*Gossypium herbaceum*	160140	37.31	33.1	38.4	86	8	36	132	2	68496	NC_023215.1
15	*Gossypium herbaceum subsp. africanum*	160315	37.22	32.7	38.7	83	8	37	128	-	68393	NC_016692.1
16	*Gossypium hirsutum*	160301	37.24	32.8	38.5	83	8	37	129	1	68066	NC_007944.1
17	*Gossypium incanum*	159205	37.39	33.2	38.7	83	8	37	128	-	67526	NC_018109.1
18	*Gossypium klotzschianum*	160097	37.3	33	38.7	84	8	37	129	-	68343	NC_033394.1
19	*Gossypium longicalyx*	160241	37.24	32.8	38.5	86	8	37	133	2	68003	NC_023216.1
20	*Gossypium mustelinum*	160313	37.22	32.7	38.7	83	8	37	128	-	68071	NC_016711.1
21	*Gossypium nelsonii*	162316	36.8	32.2	38.1	84	8	37	129	-	69006	NC_033399.1
22	*Gossypium populifolium*	159444	37.2	32.2	38.4	84	8	37	129	-	67730	NC_033398.1
23	*Gossypium raimondii*	160161	37.31	33.1	38.6	83	8	37	128	-	68320	NC_016668.1
23 a	*Gossypium raimondii*	160161	37.31	33.1	38.6	83	8	37	128	-	68320	NC_016668.1
24	*Gossypium robinsonii*	159849	37.16	32.7	38.4	83	8	37	128	-	68134	NC_018113.1
25	*Gossypium somalense*	159539	37.37	33.2	38.7	83	8	37	128	-	67776	NC_018110.1
26	*Gossypium stocksii*	159039	37.41	33.2	38.6	86	8	37	132	1	67401	NC_023217.1
27	*Gossypium sturtianum*	159627	37.13	32.7	38.3	86	8	36	131	1	67687	NC_023218.1
28	*Gossypium thurberi*	160264	37.22	32.7	38.5	85	8	40	134	1	68038	NC_015204.1
29	*Gossypium tomentosum*	160433	37.2	32.7	38.7	83	8	37	128	-	68198	NC_016690.1
30	*Gossypium trilobum*	160109	37.29	33	38.6	84	8	37	129	-	68381	NC_033397.1
31	*Gossypium turneri*	159927	37.26	33	38.7	83	8	37	128	-	68391	NC_026835.1

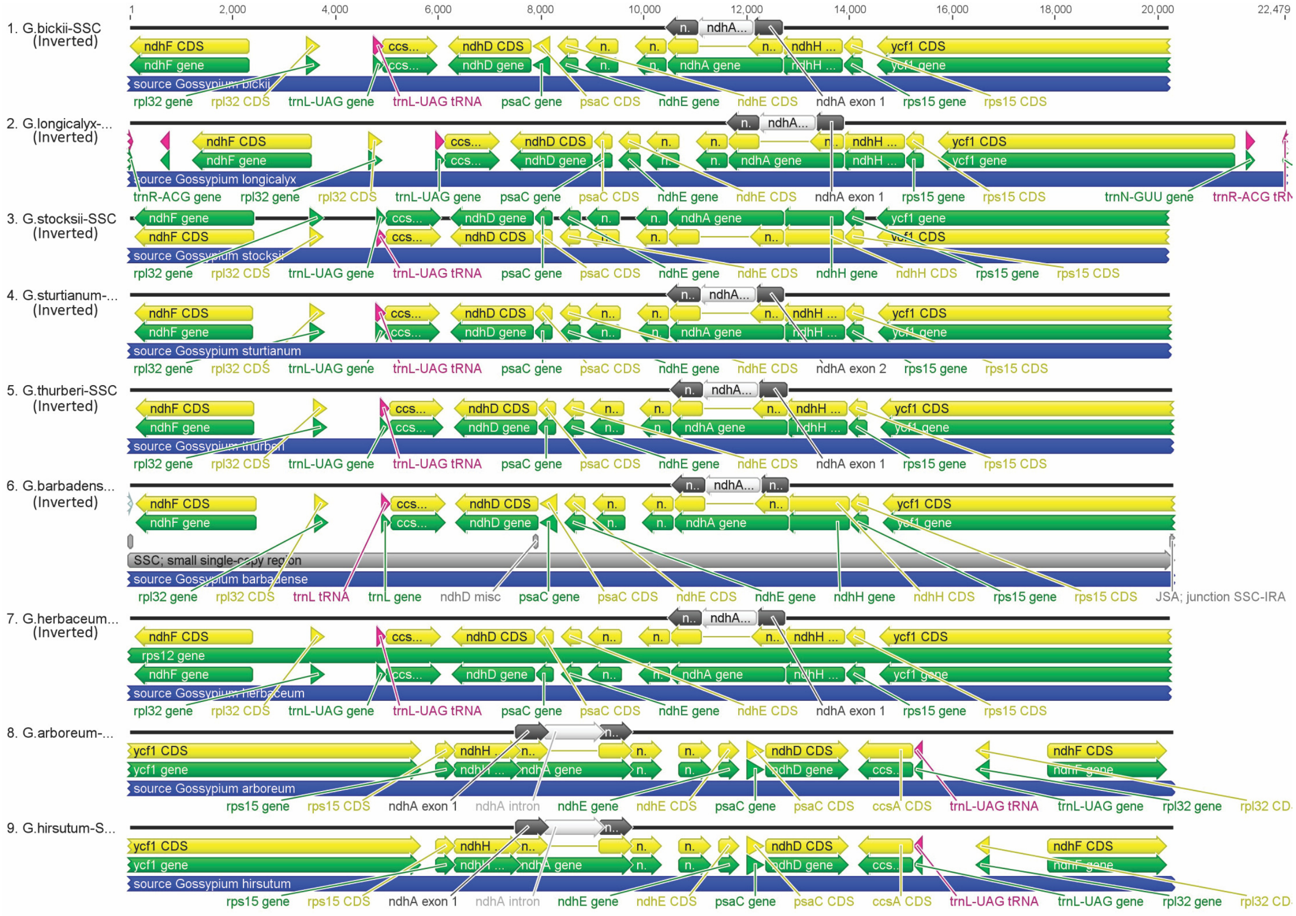

Fig.2. Details of normal and inverted SSC regions in fiber producing and other *Gossypium* species

gene distribution (Glemin *et al.,* 2014) (iv) 5'UTR and 3'UTR sequences. Low GC content is a characteristic of plastomes which might have resulted after endosymbiosis by DNA replication and repair. Comparative studies revealed about 37% (± 0.4) GC content in genus *Gossypium*, similar to the other plants reported elsewhere. In general, amongst the 4 regions of the plastome, the IRs record high GC content (apx.51-55%) mainly because of ribosomal genes located on this region. Further, variations in the GC contents between genes of different functions in plastomes are reported (Yurina *et al.,* 2017). In this study, GC content in non coding (intergenic) region was less than coding region (Table 1)

Numbers of the genes, rRNAS and tRNAS : The plastome harbors approximately 120-130 genes primarily involved in various functions of photosystem I and II, transcription, and translation (Jensen and Leister, 2014). In different *Gossypium* species, the protein coding genes were observed between 83-87 and total genes including tRNAs, and rRNAs ranged from 128-133 (Table 2). Eight Gossypium species *G. anomalum*, *G. bickii*, *G. herbaceum*, *G. hirsutum*, *G. longicalyx*, *G. stocksii*, *G. sturtianum* and *G.*

Table 2. Regions size (Highlighted regions having bigger SSC region)

Cp No.	Name	IRa	IRb	LSC	SSC
1	*Gossypium anomalum*	25,286	25,286	71,404	**51,826**
2	*Gossypium arboreum*	25,617	25,617	71,508	20,274
3	*Gossypium areysianum*	25,569	25,569	71,390	20,252
4	*Gossypium aridum*	25,647	25,647	71,537	20,243
5	*Gossypium armourianum*	25,593	25,593	71,410	20,224
6	*Gossypium australe*	25,567	25,567	71,355	20,221
7	*Gossypium barbadense*	25,591	25,591	71,476	20,294
8	*Gossypium bickii*	25,583	25,583	71,349	20,183
9	*Gossypium capitis-viridis*	25,602	25,602	71,402	20,198
10	*Gossypium darwinii*	25,603	25,603	71,472	20,266
11	*Gossypium davidsonii*	25,602	25,602	71,444	20,240
12	*Gossypium gossypioides*	25,576	25,576	71,156	20,004
13	*Gossypium harknessii*	25,599	25,599	71,419	20,221
14	*Gossypium herbaceum*	25,605	25,605	71,431	20,221
15	*Gossypium herbaceum subsp. africanum*	25,620	25,620	71,525	20,285
16	*Gossypium hirsutum*	25,602	25,602	71,484	20,280
17	*Gossypium incanum*	25,565	25,565	71,326	20,196
18	*Gossypium klotzschianum*	25,604	25,604	71,443	20,235
19	*Gossypium longicalyx*	25,647	25,647	71,574	22,478
20	*Gossypium mustelinum*	25,610	25,610	71,489	20,269
21	*Gossypium nelsonii*	26,596	26,596	72,556	19,364
22	*Gossypium populifolium*	25,577	25,577	71,247	20,093
23	*Gossypium raimondii*	25,651	25,651	71,507	20,205
23a	*Gossypium raimondii*				
24	*Gossypium robinsonii*	25,580	25,580	71,363	20,203
25	*Gossypium somalense*	25,569	25,569	71,389	20,251
26	*Gossypium stocksii*	25,487	25,487	71,153	20,179
27	*Gossypium sturtianum*	25,580	25,580	71,376	20,216
28	*Gossypium thurberi*	25,628	25,628	71,527	20,271
29	*Gossypium tomentosum*	25,615	25,615	71,501	20,271
30	*Gossypium trilobum*	25,604	25,604	71,441	20,233
31	*Gossypium turneri*	25,413	25,413	71,038	55,210

thurberi possess one or two pseudogenes. In all *Gossypium* species rRNA genes were found in duplicate for *rrna4.5*, *rrra5*, *rrna16* and *rrna23* on inverted repeat regions. In most *Gossypium* sp., tRNAs numbers are 37 except 36 in *G.bickii*, *G.herbaceum*, *G.sturtianum* and 40 in *G. turberi* . The intergeneic region length between 67,401 bp (*G. stockii*) to 69,006bp (*G. nelsonii*) (Table 2).

Introns : Introns are ubiquitously present in eukaryotic genomes. Most genes are interrupted by one or more introns which removed upon transcription to generate intact open reading fames of mRNA. Location and numbers of the introns in protein coding genes seems to be conserved. The comparative studies revealed that total 12 genes have one or two intron and it is observed in all *Gossypium* species (Table 3). While in 7 tRNAs genes, the distinction in presence and absence of intron was quite prominent (Table 3). Ever since its discovery, intron has remained the topic of great interest for molecular biologists especially for gene expression and evolutionary studies. The size of the intron varies between the species and mutational preference for deletions controlled by the size of the intron. Natural selection favors shorter and less number of introns in highly expressing genes and thus has evolutionary significance (Francis and Worheide, 2017).

Table 3. Inverted SSC region in 8 *Gossypium* genome with two normal one as a outgroup

Cp genome No.	GENOME	SSC
1	*Gossypium anomalum*(abnormal)	97892-149718
8	*Gossypium bickii*(abnormal)	113656-133839
19	*Gossypium longicalyx*(abnormal)	113215-135693
26	*Gossypium stocksii* (abnormal)	113373-133552
27	*Gossypium sturtianum*(abnormal)	113831-134047
28	*Gossypium thurberi*(abnormal)	114365-134636
7	*Gossypium barbadense*(abnormal)	114432-134726
14	*Gossypium herbaceum*(abnormal)	114314-134535
2	*Gossypium arboreum*(normal)	114328-134602
16	*Gossypium hirsutum*(normal)	114408-134688

Gene loss and inversion in SSC region : Gene loss in LSC and IR region of *Gossypium* genus is experienced in many species (Table-4). In LSC region, *rpl22* loss observed in 14 plastomes*, trnT-GGU* and *trnM-CAU* loss observed in 21 plastomes, *trnH-GUG* was observed in one plastome. In IR region, *ycf15* loss is observed in 22 plastomes, *rpl19* in one and *ycf1* loss is two plastomes (Table 4). With evolution of chloroplast genome loss and gain of the genes have been reported. For example, *infA, trnP-GGG* is not found in any *Gossypium* species. The *infA* as a psuedogene is reported in many chloroplast genomes (Raman and Park, 2015, Ni *et al.,* 2017) and considered as a mobile gene between nucleus and chloroplast (Lin et al., 2015, Huang *et al.,* 2017) while in some plastome the complete CDS is reported (Yang *et al.,* 2013). *trnP-GGG* is observed in *Cryptomeria japonica* (Hsu *et al.,* 2016). *rpl22* codes for ribosomal protein22 large subunit which reported as absent in 14 plastomes of genus *Gossypium*, and other genes for tRNAs loss in large number of plastomes advocate importance of its analysis which may shed some light on evolution of *Gossypium.*

The SSC region of all *Gossypium* species contained *ndhF, rpl32*, *trnL-UAG*, *ccsA*, *ndhD*, *psaC ndhE*, *ndhG*, *ndhI*, *ndhA*, *ndhH*, *rps15* and *ycf1* genes. However, genome alignment showed inverted SSC region in 8 *Gossypium* sp (Table-5, Fig. 2). A marked distinction in 8 species of the *Gossypium* showed inversion in SSC region (Table 5, Fig. 1). They are *G. anomalum*, *G. barbadense, G. bickii*, *G. herbaceum*, *G. longicalyx*, *G. stocksii*, *G. sturtianum* and *G. thurberi.*

Further, in *G. anomalum*, SSC region contain two copies of *trnL-CAA*, *ndhB*, *rps7*,*trnV-GAC*,*trnI-GAU*,*trnA-UGC*, *rrn23*,*rrn4.5*,*trnN-GUU* and *trnR-ACG*; due to the polymorphism of one base pair (C replaced by T) in IR region (Fig.2), it was a shift in SCC region or represented as a overlap. While in *G. longicalyx*, extra genes *trnN-GUU* in two copies was observed.

The noteworthy point in SSC region observed in fiber producing *Gossypium* sp. *i.e. G. herbaceum* (2n) and *G. barbidences* (4n) showed inverted SSC regions as compared to *G. arboretum* (2n) and *G. hirsutum* (4n). The important parameter for investigation in

Table 4. Intron content among all *Gossypium* chloroplast genomes

Genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
atpF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
rpoC1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
psaA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ycf3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
clpP	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
rpl2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ycf15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ndhB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ndhA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ndhB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ycf15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
rpl2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
trnL-TAA	-	-	-	-	1	-	-	-	-	-	1	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-
trnI-TAT	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
trnR-TCT	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
trnL-TAG	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
trnL-TCT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
trnI-AAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-
trnR-TCG	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-

Table 5. Region wise Gene loss or gain in *Gossypium* genomes

Regions	Gene	Gene loss in cp genome
LSC	rpl22	2, 3, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25, 28, 29
	trnT-GGU	1, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 17, 18, 21, 22, 23, 24, 25,26, 27, 30, 31
	trnM-CAU	1, 3, 4, 5, 6, 8, 9, 11, 12, 13, 14, 17, 18, 21, 22, 23, 24, 25,26, 27, 30, 31
	trnH-GUG	31
IR	ycf15	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 29, 30
	rps19	28
	ycf1	2, 7

comparative genomics is the identification and characterization of homologous regions in closely related genomes; which can uncover large-scale events in the evolutionary aspect of a genome, such as segmental duplications or deletion. However, these regions in different species can differ significantly. On the other hand, comparative analysis of small loci can produce detailed evolutionary histories of groups of neighboring genes and provide examples of the types of changes possible in a genome thus the loss of genes in many *Gossypium* plastomes and rearrangement of the genes (by inverted SSC region) in 8 *Gossypium* plastomes may suggest an important measure for evolutionary studies. We report the key important differences in these plastomes which can be helpful to explore the possibilities at biotechnological level, phylogenic analysis and in the evolutionary studies.

ACKNOWLEDGEMENT

Authors are grateful to UGC-CAS Department of Biosciences, Saurashtra University Rajkot and Gujarat state Biotechnology Mission, Gandhinagar for the support to this work.

REFERENCES

Abdurakhmonov, I.Y., Buriev, Z.T., Shermatov, S.E., Abdullaev, A.A., Urmonov, K., Kushanov, F., Egamberdiev, S.S., Shapulatov, U., Abdukarimov, A., Saha, S., Jenkins, J.N., Kohel, R.J., Yu, J.Z., Pepper, A.E., Kumpatla, S.P. and Ulloa, M . 2012. “Genetic diversity in *Gossypium* genus” in Genetic Diversity in Plants, ed.M.Galiskan (Rijeka:InTechPress), 331.

Chen, Z., Grover, C.E., Li, P., Wang, Y., Nie, H., Zhao, Y., Wang, M., Liu, F., Zhou, Z., Wang, X. and Cai, X. 2017. Molecular evolution of the plastid genome during diversification of the cotton genus. *Mol. Phylogenet. and evol***. 112**:268**.**

Francis, W.R. and Worheide, G. 2017. Similar Ratios of Introns to Intergenic Sequence across Animal Genomes. *Genome Biol. Evol.* **9**:1582.

Glemin, S., Clement, Y., David, J. and Ressayre A. 2014. GC content evolution in coding regions of angiosperm genomes: a unifying hypothesis. *Trends Genet.* **30**:263.

Guo, S., Wang, Y., Sun, G., Jin, S., Zhou, T., Meng, Z. and Zhang, R. 2015. Twenty years of research and application of transgenic cotton in China. *Sci. Agi. Sin.* ***48***:3372.

Huang, Y., Wang, J., Yang, Y., Fan, C. and Chen, J., 2017. Phylogenomic Analysis and Dynamic Evolution of Chloroplast Genomes in Salicaceae. *Front. Plant. Sci.* **8**:1050.

Hsu, C.Y., Wu, C.S. and Chaw, S.M., 2016. Birth of four chimeric plastid gene clusters in Japanese umbrella pine. *Genome Biol. Evol.* **8**:1776.

Jensen, P.E. and Leister, D., 2014. Chloroplast evolution, structure and functions. *F1000prime reports, 6.*

Jin, S. and Daniell, H. 2015. The engineered chloroplast genome just got smarter. *Trends Plant Sci.* **20**:622.

Lin, C.S., Chen, J.J., Huang, Y.T., Chan, M.T., Daniell, H., Chang, W.J., Hsu, C.T., Liao, D.C., Wu, F.H., Lin, S.Y. and Liao, C.F., 2015. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* **5**:9040.

Liu, C., Shi , L., Zhu, Y., Chen, H., Zhang, J., Lin, X. and Guan, X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* **13**:715.

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., Fang, C., Zhang, Q., Bai, M., Yao, N. and Wu, H., 2013. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice**.** *Nature Genet.* **45**:573.

Malik, W., Ashraf, J., Iqbal, M.Z., Khan, A.A., Qayyum, A., Abid, M.A., Noor, E., Ahmad, M.Q. and Abbasi, G.H. 2014. Molecular markers and cotton genetic improvement: current status and future prospects.*The Sci World J.* **2014**:15.

Ni, L., Zhao, Z., Xu, H., Chen, S. and Dorje, G., 2017. Chloroplast genome structures in *Gentiana* (Gentianaceae), based on three medicinal alpine plants used in Tibetan herbal medicine. *Curr. Genet.* **63**:241.

Raman, G. and Park, S., 2015. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. *PLoS One*, **10**:0141329.

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S. and Zou, C. 2012. The draft genome of a diploid cotton *Gossypium raimondii. Nature Genet. 44*:1098.

Yang, J.B., Tang, M., Li, H.T., Zhang, Z.R. and Li, D.Z. 2013. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Biol. 13*:84.

Yu, L.H., Wu, S.J., Peng, Y.S., Liu, R.N., Chen, X., Zhao, P., Xu, P., Zhu, J.B., Jiao, G.L., Pei, Y. and Xiang C.B. 2015. Arabidopsis EDT1/ HDG11improves drought and salt tolerance in cotton and poplar and increases cotton yield in the field. *Plant Biotechnol. J.* **14**:72.

Yurina, N.P., Sharapova, L.S. and Odintsova, M.S., 2017. Structure of plastid genomes of photosynthetic eukaryotes. *Biochemistry (Moscow)*, **82**:678.

Zhang, Y., Yan, C. and Kuang, H.2014. GC content fluctuation around plant small RNA-generating sites. *FEBS Letters.* **588** : 764.

Received for publication : March 4, 2018
Accepted for publication : May 18, 2018