Journal of Plant Sciences
2018; 6(4): 134-143
http://www.sciencepublishinggroup.com/j/jps
doi: 10.11648/j.jps.20180604.13
ISSN: 2331-0723 (Print); ISSN: 2331-0731 (Online)

Comparative Chloroplast Genome Analysis of Single-Cell C_4 Bienertia Sinuspersici with Other Amaranthaceae Genomes

Lorrenne Caburatan, Jin Gyu Kim, Joonho Park*

Department of Fine Chemistry, Seoul National University of Science and Technology, Seoul, South Korea

Email address:
jhpark21@seoultech.ac.kr (J. Park)

*Corresponding author

To cite this article:
Lorrenne Caburatan, Jin Gyu Kim, Joonho Park. Comparative Chloroplast Genome Analysis of Single-Cell C_4 Bienertia Sinuspersici with Other Amaranthaceae Genomes. *Journal of Plant Sciences*. Vol. 6, No. 4, 2018, pp. 134-143. doi: 10.11648/j.jps.20180604.13

Received: August 15, 2018; **Accepted**: August 31, 2018; **Published**: September 26, 2018

Abstract: Bienertia sinuspersici is a single-cell C_4 (SCC_4) plant species whose photosynthetic mechanisms occur in two cytoplasmic compartments containing central and peripheral chloroplasts. The efficiency of the C_4 photosynthetic pathway to suppress photorespiration and enhance carbon gain has led to a growing interest in its research. A comparative analysis of B. sinuspersici chloroplast genome with other genomes of Amaranthaceae was conducted. Results from a 70% cut off sequence identity showed that B. sinuspersici is closely related to Beta vulgaris with slight variations in the arrangement of few genes such ycf1 and ycf15; and, the absence of psbB in Beta vulgaris. B. sinuspersici has the largest 153, 472 bp while Spinacea oleracea has the largest protein-coding sequence 6,754 bp larger than B. sinuspersici. The GC contents of each of the species ranges from 36.3 to 36.9% with B. sinuspersici having the same GC percentage as Haloxylon persicum and H. ammodendron (36.6%). The IR size also varies yet in all six species, the Ira/LSC border is generally located upstream of the trnH-GUG gene. A total of 107 tandem repeats were found in each of the species, most of which are situated in the intergenic space. These results provide basic information that may be valuable for future related studies.

Keywords: Bienertia Sinuspersici, Single-Cell C_4, Central and Peripheral Chloroplasts, Comparative Anaylysis, Tandem Repeats

1. Introduction

Chloroplasts are double membrane organelles containing their own genomic machineries. These plant plastids encode genes not only essential for photosynthetic and genetic functions, but also participate in other functional mechanisms such as biosynthesis of nucleotides, lipids, fatty acids, and starch [1-3], as well as the accumulation of pigments, photoreceptors, and hormones [4]. The majority of chloroplast genome sequences of higher plants, such as angiosperms, are highly conserved [3]. These circular genomes are about 120–160 kb in size and are present in 1,000–10,000 copies per cell [5-6].

Plants can be classified as either C_3 or C_4, based on the employed carbon fixation mechanisms. In C_4 plants, two kinds of chloroplasts are found on either mesophyll cells (MC) or bundle sheath cells (BSC) of the so called Kranz anatomy [7-8]. In these Kranz C_4 species, there is a separation of primary and secondary carbon fixation reactions between the MC and BSC.

However, there is another type of C_4 plant in which the carbon-concentrating mechanism does not depend on cooperation between MC and BSC. Such is the case of the single-cell C_4 (SCC_4) plant species.

Bienertia sinuspersici is an SCC_4 plant belonging to the family Amaranthaceae. Chloroplasts of B. sinuspersici are situated in two intracellular cytoplasmic compartments: central and peripheral. The former is packed with chloroplasts and mitochondria, while the latter was previously thought to lack mitochondria [9-11], although recent studies have tried to prove otherwise [12-13]. Nonetheless, both compartments contain peroxisomes, which

are involved in photorespiration.

Moreover, high rates of photosynthesis and efficient use of water and nitrogen resources characterize C_4 plants. In such plants as B. sinuspersici, C_4 photosynthesis is an adaptation to suppress photorespiration [7, 14-15] and increase carbon gain [16].

The efficiency of C_4 photosynthesis led to the growing interest to engineer the C_4 photosynthetic pathway into C_3 crops [17-23]. It is envisioned that such a breakthrough would benefit the global economy by improving crop production and increasing yield to meet the ever-growing challenges of a surging population [23-27].

Therefore in this study, a comparative genome analysis between B. sinuspersici and other species in Amaranthaceae was conducted to further understand its genomic status and relation to species in the same family and provide basic information that may help future studies for chloroplast engineering.

2. Materials and Methods

2.1. Chloroplast Isolation and DNA Sequencing

Fresh leaves (~2 cm) of B. sinuspersici were collected. Chloroplast DNA was extracted following the protocol of WizPrep™ Plant DNA mini kit (Wizbiosolutions, South Korea). Total DNA concentration was measured using Optizen POP UV/Vis spectrophotometer (Mecasys, South Korea) at 260 nm. The genome of B. sinuspersici chloroplast was analyzed using a combined approach with 454 GS FLX Titanium system (Roche Diagnostics, Brandford, CT) with an 8-kb paired-end library and the Illumina GAIIx (San Diego, CA). The 454 GS FLX sequencing achieved about 3.5-fold coverage, while 290.2-fold read coverage was achieved by Illumina paired-end sequencing. The reads generated by the Illumina GAIIx and the 454 GS FLX Titanium were assembled using Celera Assembler 7.0.

2.2. Genome Assembly and Annotation

De novo assembly was conducted using Celera Assembler 7.0 [28]. Gene prediction and annotation were carried out using Glimmer3, the RAST annotation server, and the NCBI COG database. Geneious version 8.1.6 was used to annotate the chloroplast genome. The annotation results were manually checked. Codon positions were compared with homologue genes from previously known chloroplast DNA, with slight adjustments on codon position. The circular DNA map was drawn through OGDRAW software [29]. The final chloroplast genome of B. sinuspersici was deposited to GenBank with accession number: KU726550 [30].

2.3. Genome Comparison and Sequence Analysis

B. sinuspersici genome was obtained from a previous study of Kim et al. [30] with a little modification using the software Geneious version 8.1.6 (Biomatters, NZ). Genomes of other Amaranthaceae species were obtained from GenBank with the following accession number: Beta vulgaris, KJ081864 [31]; H. persicum, KF534479 and H. ammodendron, KF534478 [32]; S. oleracea, NC_002202 [33]; and, Dianthus longicalyx, KM668208 [3].

B. sinuspersici chloroplast genome was compared with the mentioned species. Sequence analysis of genes was performed using the mVISTA program in Shuffle-LAGAN mode with 70% cut-off identity. The genome sequence of B. sinuspersici was used as the reference, while the outgroup was that of D. longicalyx. Tandem repeats were also determined using Tandem Repeats Finder version 4.09 [34].

3. Results and Discussion

3.1. Chloroplast Genome Features of B. Sinuspersici

The complete chloroplast genome of B. sinuspersici is 153,472 bp in length. It is divided into a quadripartite structure with a long single-copy (LSC) bp length of 84,560 (55.10%) and a short single-copy (SSC) bp length of 19,016 (12.39%), separated by two IR regions (IRa and IRb) with a length of 24,948 bp each (32.51%). The total GC content of the chloroplast genome sequence is 36.6%. It is highest in the IR regions (42.94%) while in the LSC and SSC, the GC content is much lower (34.47% and 29.42% respectively).

Figure 1 shows the positions of all the genes identified in the chloroplast genome of B. sinuspersici and the corresponding functional categorization. The complete chloroplast genome encodes for 130 predicted functional genes, 113 of which are unique with 80, 29, and 4 unique protein-coding, tRNA, and rRNA genes, respectively.

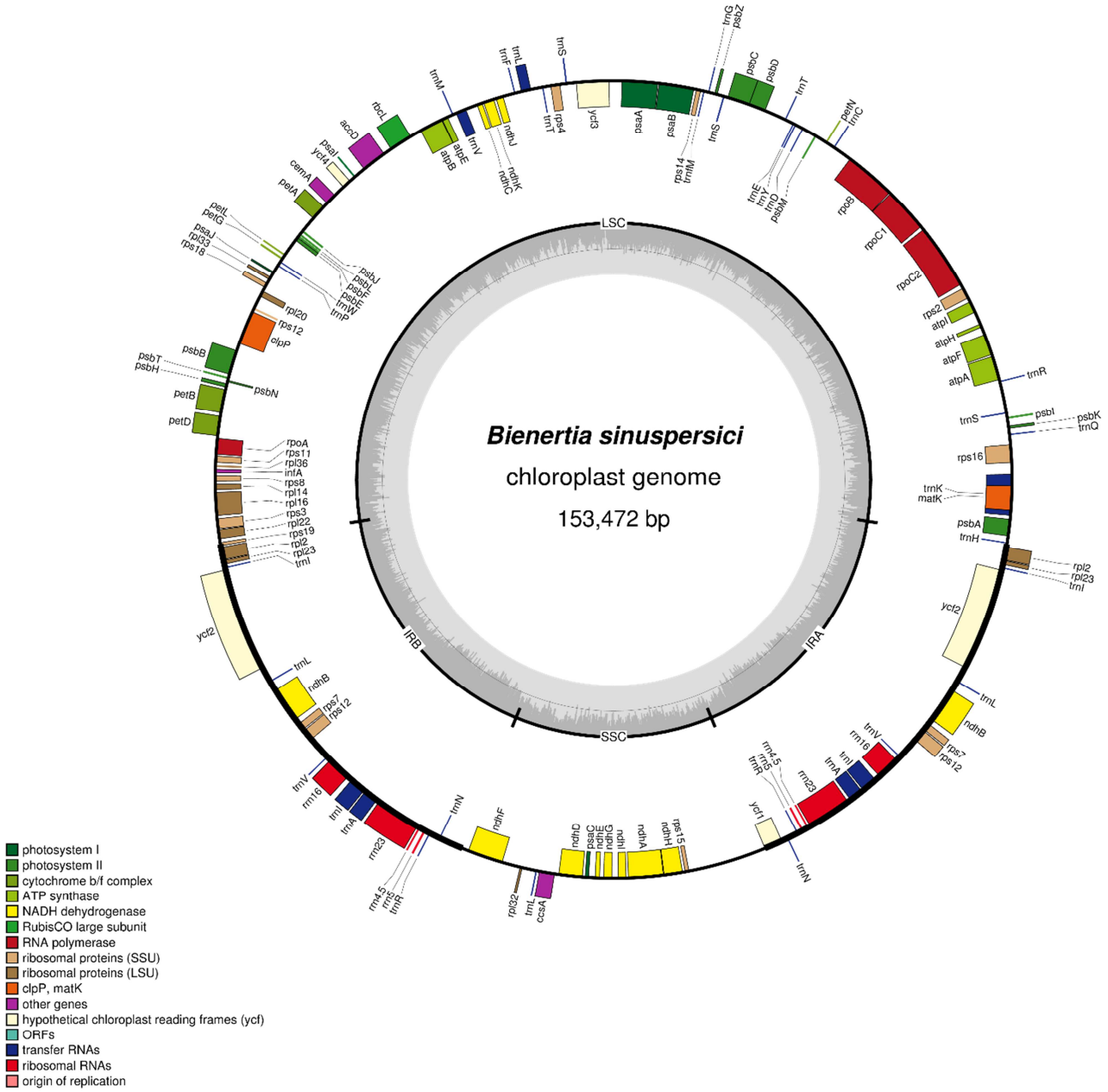


Figure 1. *Gene map of Bienertia sinuspersici chloroplast genome. Exons are annotated by coloured boxes while introns are annotated by black boxes. Genes drawn inside the circle are transcribed clockwise while those outside are transcribed counterclockwise.*

The LSC region contains 83 genes while the SSC region contains 30 genes. A total of 17 genes are duplicated in the IR regions. This is composed of six protein-coding, seven tRNA, and four rRNA genes. The ribosomal RNA genes have the most abundant transcripts in plastids, the biosynthesis of which is highly regulated during development at both the transcriptional and post-transcriptional levels [35]. Gene duplication has been asserted [36] to be one of the factors to have caused the evolution or synthesis of new genes arising from pre-existing ones either through a complete or partial duplication of genes [37]. These duplicated genes are said to contribute to genomic instability, leading to gene rearrangement and thence, speciation [38].

Moreover, there are a total of 16 intron-containing genes, 14 containing one intron (nine protein-coding and five tRNA genes) and two (clpP, and ycf3) containing two introns. The longest intron is found in the ndhA gene, accounting for 1,105 bp.

The protein-coding genes are comprised of 24,871 codons. A total of 10.6% of all codons (2,645) encodes leucine as the most prevalent amino acid, and 1.2% (298) encodes cysteine as the least prevalent amino acid. The protein-coding genes accounted for 48.62% (74,613 bp) of the whole genome sequence, while the tRNA and rRNA genes accounted for 1.78% (2,733 bp) and 5.89% (9,045 bp), respectively. The remaining regions were comprised of non-coding sequences (CNS), including introns, intergenic spacers, and a hypothetical protein.

3.2. Comparison with Other Chloroplast Genomes in the Family Amaranthaceae

B. sinuspersici genome was compared to five other species. B. sinuspersici and the other four species belong to the family Amaranthaceae, except for D. longicalyx (family Caryophyllaceae), which serves as the outgroup. The average GC content of all six species is 36.57%. S. oleracea has the largest GC content (36.9%), while D. longicalyx has the smallest (36.3%). As with B. sinuspersici, the Haloxylon species (H. persicum and H. ammodendron) also have 36.6% GC content.

Notable differences observed among these species also include genome size, gene losses and intron losses, pseudogenization of protein-coding genes, sequence divergence and tandem repeats, and IR expansion and contraction.

3.2.1. Genome Size

B. sinuspersici has the largest genome size of all the species compared, followed by Haloxylon persicum. D. longicalyx has the smallest genome size, which is 97 bp smaller than Beta vulgaris. Table 1 presents a detailed summary of the general features between these species. The observed variation in genome size is mainly due to the variation in the length of the IR regions and the length of the protein-coding region. S. oleracea has the largest protein-coding region among the analyzed species. It is 6,754 bp larger than B. sinuspersici. The latter appeared to contain the least length of protein-coding genes (74,612 bp). Therefore, B. sinuspersici has more prominent conserved noncoding sequence (CNS) than the others.

CNS implications and functional significance in the genome has captured interest [39] and it has been observed that CNS may be correlated to transcription regulation as grass regulatory genes are rich in orthologous CNS in a study presented by Guo and Moose [40] and as has been found out by Thomas and colleagues [41] that Arabidopsis genes encoding for transcription factors are rich in homologous CNS.

3.2.2. Gene Loss

The gene psbB is absent in Beta vulgaris. It is one of the fifteen genes which encodes for subunits of photosystem II (PS II) which is a membrane protein complex that catalyzes the light-driven oxidation of water [42]. In a study presented by Delannoy and colleagues [43], psbB has also been claimed to be absent in the underground orchid, Rhizanthella gardneri and a pseudogene in a parasitic plant, Epifagus virginiana. In B. sinuspersici, ycf15 and the tRNA gene trnG-GCC which codes for glycine is absent. This is the reason why B. sinuspersici only has 29 tRNA genes.

The retention of some genes within the organelle genome and their absence may be attributed to several mechanisms working together to affect a particular phenomenon inside the cell. There are a few hypotheses presented in previous studies to explain gene retention in organelles such as the chloroplast and mitochondria. One well-accepted hypothesis is the 'Co-location of genes and gene products for Redox Regulation of gene expression' (CoRR). This hypothesis is based on ten principles and is concentrated on redox-dependent genes: rbcL; rps2, 3, 4, 7, 8, 11, 12, 14, 19; and rpl2, 14, 16, 20, 36) [44]. Retention of genomes in organelles was thought to be caused by regulatory coupling between electron transfer and gene expression [45]. While the CoRR hypothesis concerns the significance of redox-dependent genes for genome retention, another hypothesis focused on the retention of ribosomal assembly genes in the organelle. Maier et al. [46] suggested that ribosomal assembly imposes functional constraints that are subordinate to redox regulation for electron transport chain components, which govern the retention of protein-coding genes in the organelles. Finally, the hypothesis proposed by Keller et al. [44] refers to the loss of the chloroplast target peptide of the nuclear-encoded ribosomal protein, or the rps16 gene in their study.

Table 1. *General chloroplast genome features of four Amaranthaceae and an outgroup (D. longicalyx).*

Features	B. sinuspersici	Beta vulgaris	H. persicum	H. ammodendron	S. oleracea	D. longicalyx
Genome size	153,472	149,637	151,586	151,570	150,725	149,539
LSC length	84,560	83,057	84,217	82,719	82,719	82,805
SSC length	19,016	17,701	19,015	19,014	17,860	17,172
IR length	49,896	48,879	48,354	48,342	50,146	49,606
Coding size	74,612	76,998	77,130	77,145	81,366	78,588
Spacer size	62,508	62,242	57,444	57,411	50,814	55,084
Intron size	16,352	10,397	17,012	17,014	18,545	15,867
GC content (%)	36.6	36.4	36.6	36.6	36.9	36.3
Total gene number	130	130	131	131	129	129
Protein-coding genes	80	79	78	78	78	78
Duplicated genes	17	17	19	19	17	17
tRNA genes	29	30	30	30	30	30
rRNA genes	4	4	4	4	4	4
Genes with intron	16	17	17	17	17	17
Pseudogenes	0	1	2	2	2	2

3.2.3. Intron loss

The trnG-GCC has no introns in D. longicalyx, S. oleracea, and Beta vulgaris while trnG-UCC has no introns in B. sinuspersici, S. oleracea, and D. longicalyx. The trnG-UCC gene may contain an intron, which would be unique to chloroplasts [1, 47]. Hence, the absence of introns in this gene in B. sinuspersici and the other two mentioned species is interesting. It was also observed that the rpl2 gene has no introns in all of the studied species. As claimed by Downie and colleagues [48], this gene has lost its introns in all the Caryophyllales families to which Amaranthaceae belongs. Thus, this is considered a distinguishing feature of core members of Caryophyllales [3, 49].

There also are no introns in petB and petD genes in Beta vulgaris. These genes encode subunits of two functionally distinct photosynthetic complexes (PSII and cytochrome b6/f) whose accumulation is affected differently by light. Proteins of cytochrome b6/f is the one which accumulate in the dark [50]. As spliced and unspliced RNAs in petB and petD encode alternative petB and petD open reading frames in maize [50], the absence of introns in Beta vulgaris petB and petD genes may have also shown a quite diverse set of sequences in both genes.

The presence of introns is a possible advantage for the regulation of gene expression as introns are deemed essential for such regulating mechanism [51-54]. Intron sequences are also said to regulate alternative splicing [51, 55-57] and are claimed to control mRNA transport or chromatin assembly [51, 58-60].

3.2.4. Gene Pseudogenization

As with most angiosperms, the species in this study characterize pseudogenization of some genes. These pseudogenes are copies of genes that have coding-sequence deficiencies, like frameshifts and premature stop codons, but resemble functional genes [61]. There are two known pseudogenes for most of the Amaranthaceae species in this study. The infA and rpl23 genes are pseudogenes in D. longicalyx. The infA gene functions as a translation initiation factor which assists in the assembly of the translation initiation complex [62] and was found to be either a pseudogene or missing gene in some species under the order Caryophyllales, as well as that of Brassicales, Fabales, Cucurbitales, Malvales, and Myrtales [3]. Loss of this gene in some higher plants is claimed to be due to its multiple independent transfers to the nucleus in the course of time [63].

Moreover, rpl23 is also a pseudogene in S. oleracea and in both H. persicum and H. ammodendron. In a study of Raman and Park [3] with 32 angiosperms, it was also reported that rpl23 is either a pseudogene or lost gene exclusively in members of the Caryophyllales family such as Dianthus, Lychnis and Spinacia. Thus, in this study rpl23 is only functional in B. sinuspersici and Beta vulgaris. Chloroplast ribosomal protein genes, such as rpl23, are a class of genes expressed constitutively [64], whose protein products are required for plastid metabolism, regardless of developmental stage [65]. Lastly, ycf15 is found to be a pseudogene in Beta vulgaris, S. oleracea and in both H. persicum and H. ammodendron. It is a hypothetical chloroplast protein which is not considered a protein-coding gene due to a number of premature stop codons [33, 66].

3.2.5. Sequence Divergence and Tandem Repeats

The overall sequence identity to evaluate sequence divergence was plotted in a cut-off of 70% identity, using B. sinuspersici as the reference (Figure 2). It can be noted that the IR regions are more conserved, in comparison to the SC region, and the coding region is less divergent than the non-coding region. However, among the most divergent coding regions were matK, rps16, rpoC1, ycf3, clpP, petB, rpl16, ycf2, ndhF, ndhD, and ndhA. However, further observation showed that B. sinuspersici plastid genome is more similar to Beta vulgaris and is most different from the outgroup, D. longicalyx.

Sequential arrangement of genes also showed remarkable differences. In the forward (F) direction, psbI is followed by trnR-UCU in B. sinuspersici while in Beta vulgaris, H. persicum and H. ammodendron, psbI (F) is followed by trnG-UCC (F). The trnG-UCC gene in B. sinuspersici is also in a forward direction and is followed by ycf3 in the same direction. Although all other species contain the same order and direction with regards to trnG-UCC as B. sinuspersici, the reason of an extra trnG-UCC in Beta vulgaris, H. persicum and H. ammodendron is due to the fact that in these three species, trnG-UCC is transpliced in the LSC. Another difference that can be observed is the presence of a hypothetical gene in B. sinuspersici situated in between trnR-ACG and rpl32 in the forward direction. In other species, except S. oleracea, a ycf1 gene oriented in the reverse direction (R) is present instead of a hypothetical gene. The psbB operon (psbB, psbT, psbH, petB, petD) in all of the species studied are all in the same orientation and arrangement except that a PS I hypothetical protein is present in Beta vulgaris in the place of psbB.

Moreover, tandem repeats among the six species were distributed along the intergenic space (71), coding region (30), and along regions of introns (60) (Table 2). There are a total of 107 tandem repeats. Repeat sequences in the intergenic space are high among B. sinuspersici (17), D. longicalyx (19), and among the Haloxylon species (12). That of the coding region is highest in Beta vulgaris (10). Most repeats ranged from 30 to 44 bp (Figure 3a). The longest repeat is 177 bp in B. sinuspersici. These repeats are more profound in the intergenic space (66%) than in the regions of introns (6%) (Figure 3b).

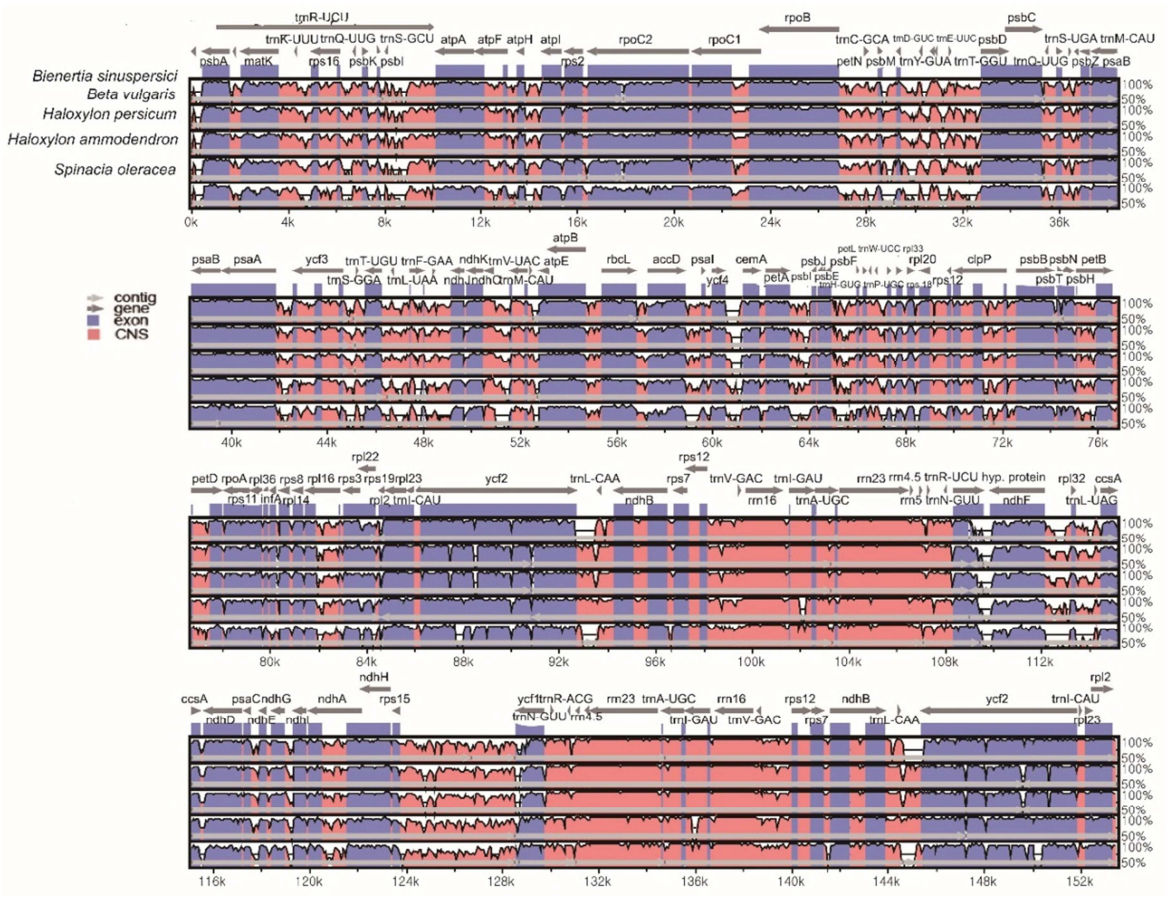
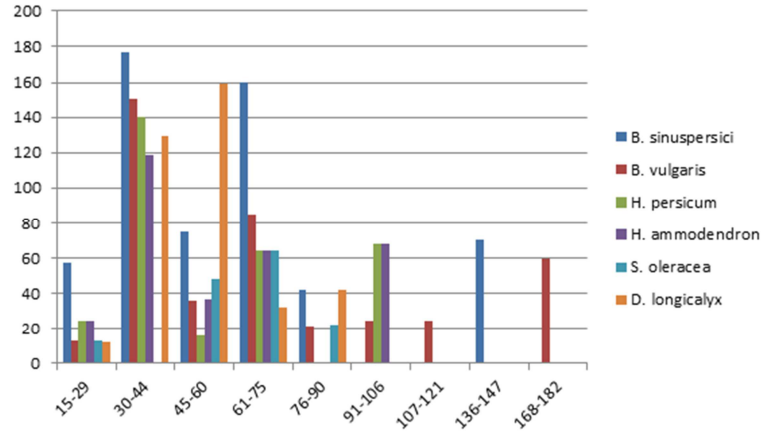


Figure 2. *Comparison of six Caryophyllales chloroplast genome using mVISTA. A cut-off of 70% identity was used for the plots. The Y-scale represents identity between 50-100%. Gray arrows represent gene position and orientation. Genome regions in blue and red represent protein-coding (exons) and non-coding sequences (CNS).*

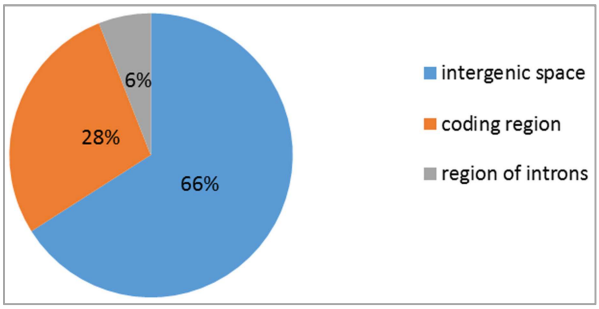
Table 2. *Tandem repeat distribution among four Amaranthaceae species and an outgroup (D. longicalyx).*

	Intergenic Space	Coding Sequence	Intron
B. sinuspercisi	17	11	0
Beta vulgaris	9	10	1
H. persicum	12	1	2
H. ammodendron	12	1	2
S. olerace	2	4	1
D. longicalyx	19	3	1
Total	71	30	6

Four genes occur in the coding regions where tandem repeats are found: rps19 (B. sinuspersici); ycf2 (B. sinuspersici, Beta vulgaris, S. oleracea, and D. longicalyx); ycf1 (Beta vulgaris, S. oleracea, and D. longicalyx); and ndhF (H. persicum and H. ammodendron). On the other hand, two intron-containing genes occur where tandem repeats are found: rpl16 (Beta vulgaris, H. persicum, H. ammodendron, and D. longicalyx); and clpP (H. persicum and H. ammodendron).



(A)



(B)

Figure 3. *Tandem repeat analysis in sequences of Caryophyllales species. (A) Frequency of tandem repeat sequences by length. The length of repeats is subdivided into 30 base pairs. (B) Location distribution of all tandem repeats in percent.*

3.2.6. IR Expansion and Contraction

The expansion and contraction of the IR regions, as well as those of the single-copy boundary regions, primarily contribute to size variation of the chloroplast genome [6, 32]. The SC and IR boundaries of the six studied chloroplast genome are presented in Figure 4.

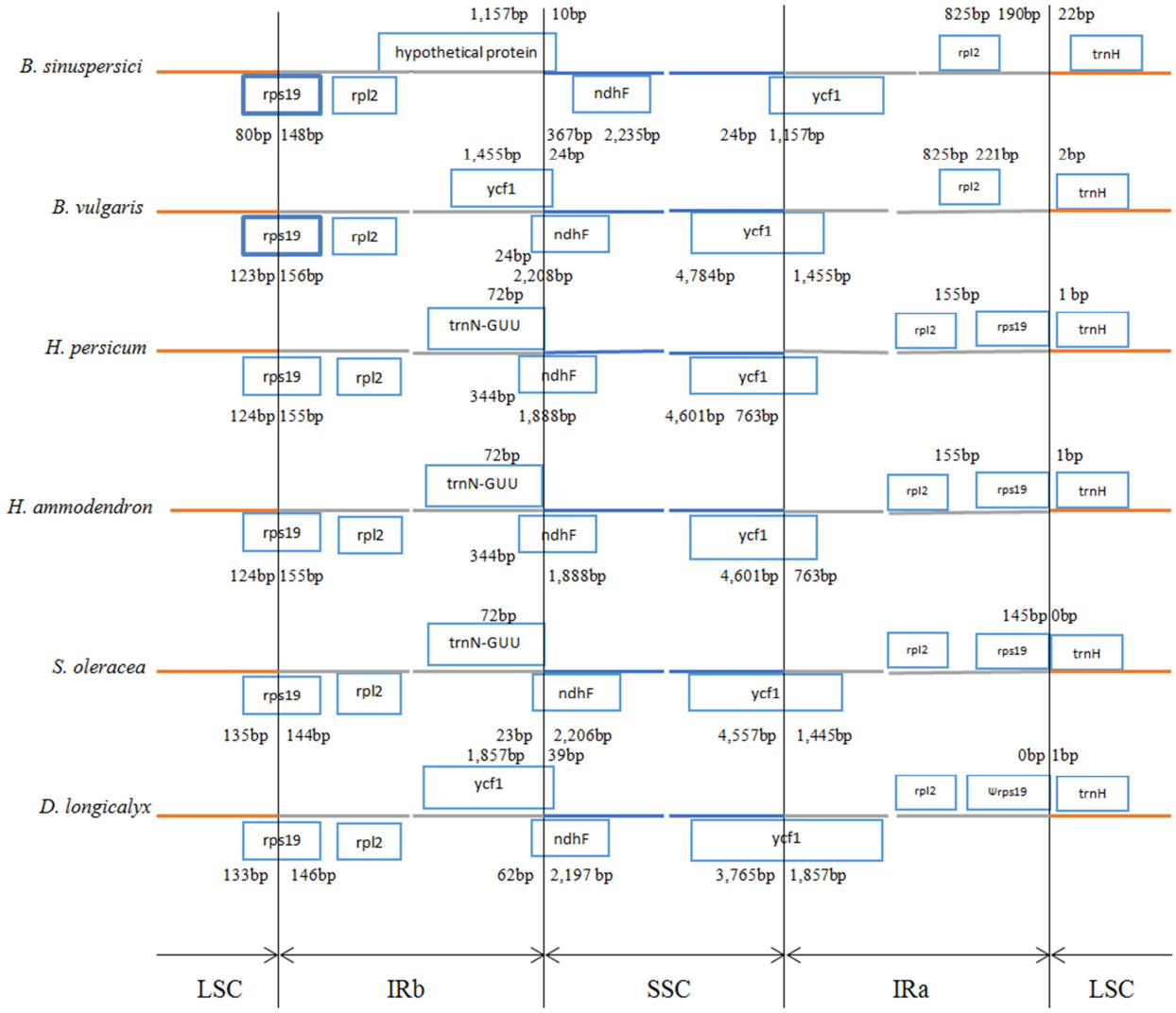


Figure 4. *Comparison of the junctions between the IR and single copy regions among five Caryophyllales chloroplast genomes. Annotated genes are represented by blue boxes. The symbol for Greek letter psi (Ψ) refers to the presence of a pseudogene.*

IR regions are said to have played a significant role in conserving genomic stability, as extensive sequence rearrangements are observed in genomes that have lost their inverted repeats [67-68]. In this study, S. oleracea has the

largest IR length, 250 bp larger than B. sinuspersici. D. longicalyx is also noted for its IR size, 1,264 bp larger than the smallest IR size of H. ammodendron, despite having the smallest genome size. This expansion of the IR regions in both S. oleracea and D. longicalyx is mainly due to differences in the individual sizes of the genes that are duplicated in the IR regions. Most of these Caryophyllales species contain only 17 genes duplicated in the IRs, except for the Haloxylon species (19), which manifest duplications of ycf15 and rps19 genes. Thus, the differences in the size of the IR regions can only be attributed to the differences in the length of nucleotide sequences in each of the individual genes, as the number of genes is nearly identical.

It can also be observed that, in all six species, the IRa/LSC border is generally located upstream of the trnH-GUG gene. This gene is the only tRNA gene that codes for the amino acid histidine. The distance of the trnH-GUG gene from the LSC border is 22 bp in B. sinuspersici, 2 bp for Beta vulgaris, and 3 bp for the remaining species except for S. oleracea, which characterizes no separation. The IRa region expanded by 1,157 bp in B. sinuspersici as it entered the 5' end of the ycf1 gene and it greatly expanded in D. longicalyx by 1,857 bp, while in Beta vulgaris and S. oleracea it expanded by 1,455 and 1,445 bp, respectively. That of the Haloxylon species were both shortly expanded by 763 bp.

Furthermore, at the IRb/SSC junction, the ycf1 gene extended by 10 bp in B. sinuspersici, 24 bp in Beta vulgaris, and 39 bp in D. longicalyx. The ycf1 gene is not duplicated in the IRb region of S. oleracea or in either the Haloxylon species. Instead, the trnN gene, which codes for asparagine, is 72 bp in length inside the IRb boundary of these three species. The ndhF gene of B. sinuspersici is the only gene at SSC that does not extend in reverse direction to the IRb region. It lies 367 bp away from the IRb border, while other species extended by 23 bp to 344 bp.

The rps19 gene, which lies on the border of the IRb/LSC junction in all five species, is 279 bp in length, except for B. sinuspersici, with a shorter span of 228 bp. This gene in reverse direction extends to the LSC border by 80 bp in B. sinuspersici, 133 and 135 bp in D. longicalyx and S. oleracea, respectively. That of Beta vulgaris (123 bp) extends 1 bp shorter than the Haloxylon (124 bp) species.

Moreover, the intron-containing rps12 gene in all of the species under study is a divided gene transpliced with the 5' end located in the LSC region and the duplicated 3' end in the IR regions. It lies closely to the rps7 gene in B. sinuspersici, Beta vulgaris, H. persicum, and H. ammodendron. On the other hand, it is 25,455 and 26,460 bp length away from rps7 in D. longicalyx and S. oleracea, respectively. The rps12 gene is thought to have a vital function in initiating translation, in Chlamydomonas reinhardtii [69] and is capable of controlling translation fidelity of E. coli ribosomes.

4. Conclusion

A comparative analysis of B. sinuspersici chloroplast genome was conducted along with four other species in the family Amaranthaceae and an outgroup from Caryophyllaceae (D. longicalyx). Differences among the chloroplast genomes of the studied species include genome size, gene losses and intron losses, pseudogenization of protein-coding genes, sequence divergence and tandem repeats, and IR expansion and contraction. Results further showed that B. sinuspersici is closely related to Beta vulgaris with slight differences in the arrangement of genes. Tandem repeats among the studied species were also variable and were mostly found in the intergenic space. Data obtained from this study may prove helpful for any further genomic studies.

Acknowledgements

We thank Dr. Offerman for providing plant material used in this study. This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01095306)" Rural Development Administration, Republic of Korea.

References

[1] Sugiura, M. (1992) The chloroplast genome. Plant molecular biology, 19(1), 149-168.

[2] Tiller, N. and Bock, R. (2014). The translational apparatus of plastids and its role in plant development. Molecular plant, 7(7), 1105-1120.

[3] 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(10), e0141329.

[4] Grabsztunowicz, M., Koskela, M. M. and Mulo, P. (2017) Post-translational modifications in regulation of chloroplast function: Recent advances. Frontiers in plant science, 8, 240.

[5] Bendich, A. J. (1987) Why do chloroplasts and mitochondria contain so many copies of their genome?. BioEssays, 6(6), 279-282.

[6] Yang, Y., Yuanye, D., Qing, L., Jinjian, L., Xiwen, L. and Yitao, W. (2014) Complete chloroplast genome sequence of poisonous and medicinal plant Datura stramonium: organizations and implications for genetic engineering. PLoS One, 9(11), e110656.

[7] Furbank, R. T. and Taylor, W. C. (1995). Regulation of photosynthesis in C3 and C4 plants: a molecular approach. The Plant Cell, 7(7), 797.

[8] Offermann, S., Okita, T. W. and Edwards, G. E. (2011) How do single cell C4 species form dimorphic chloroplasts?. Plant signaling & behavior, 6(5), 762-765.

[9] Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Artyusheva, E. G., Freitag, H. and Edwards, G. E. (2002) Proof of C4 photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae). The Plant Journal, 31(5), 649-662.

[10] Akhani, H., Barroca, J., Koteeva, N., Voznesenskaya, E., Franceschi, V., Edwards, G. … and Ziegler, H. (2005) Bienertia sinuspersici (Chenopodiaceae): a new species from Southwest Asia and discovery of a third terrestrial C4 plant without Kranz anatomy. Systematic Botany, 30(2), 290-301.

[11] Lara, M. V., Offermann, S., Smith, M., Okita, T. W., Andreo, C. S. and Edwards, G. E. (2008) Leaf development in the single-cell C4 system in Bienertia sinuspersici: expression of genes and peptide levels for C4 metabolism in relation to chlorenchyma structure under different light conditions. Plant physiology, 148(1), 593-610.

[12] Voznesenskaya, E. V., Koteyeva, N. K., Chuong, S. D., Akhani, H., Edwards, G. E. and Franceschi, V. R. (2005) Differentiation of cellular and biochemical features of the single-cell C4 syndrome during leaf development in Bienertia cycloptera (Chenopodiaceae). American Journal of Botany, 92(11), 1784-1795.

[13] Chuong, S. D., Franceschi, V. R. and Edwards, G. E. (2006) The cytoskeleton maintains organelle partitioning required for single-cell C4 photosynthesis in Chenopodiaceae species. The Plant Cell, 18(9), 2207-2223.

[14] Bräutigam, A. and Gowik, U. (2016) Photorespiration connects C3 and C4 photosynthesis. Journal of experimental botany, 67(10), 2953-2962.

[15] Gowik, U. and Westhoff, P. (2011) The path from C3 to C4 photosynthesis. Plant Physiology, 155(1), 56-63.

[16] Offermann, S., Okita, T. W. and Edwards, G. E. (2011) Resolving the compartmentation and function of C4 photosynthesis in the single-cell C4 species Bienertia sinuspersici. Plant physiology, 155(4), 1612-1628.

[17] Aubry, S., Brown, N. J. and Hibberd, J. M. (2011) The role of proteins in C3 plants prior to their recruitment into the C4 pathway. Journal of experimental botany, 62(9), 3049-3059.

[18] Furbank, R. T. (2011) Evolution of the C4 photosynthetic mechanism: are there really three C4 acid decarboxylation types?. Journal of experimental botany, 62(9), 3103-3108.

[19] Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B. J., Dionora, M. J. A., … and Quick, W. P. (2011) Strategies for engineering a two-celled C4 photosynthetic pathway into rice. Journal of experimental botany, 62(9), 3001-3010.

[20] Miyao, M., Masumoto, C., Miyazawa, S. I. and Fukayama, H. (2011) Lessons from engineering a single-cell C4 photosynthetic pathway into rice. Journal of experimental botany, 62(9), 3021-3029.

[21] Nelson, T. (2011) The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C4 leaves. Journal of experimental botany, 62(9), 3039-3048.

[22] Peterhansel, C. (2011) Best practice procedures for the establishment of a C4 cycle in transgenic C3 plants. Journal of experimental botany, 62(9), 3011-3019.

[23] Sage, R. F. and Zhu, X. G. (2011) Exploiting the engine of C4 photosynthesis. Journal of experimental botany, 62(9), 2989-3000.

[24] Mitchell, P. L. and Sheehy, J. E. (2006) Supercharging rice photosynthesis to increase yield. New Phytologist, 171(4), 688-693.

[25] Hibberd, J. M., Sheehy, J. E. and Langdale, J. A. (2008) Using C4 photosynthesis to increase the yield of rice—rationale and feasibility. Current opinion in plant biology, 11(2), 228-231.

[26] Dawe, D., Pandey, S. and Nelson, A. (2010) Emerging trends and spatial patterns of rice production. Rice in the Global Economy: Strategic Research and Policy Issues for Food Security. Los Baños, Philippines: International Rice Research Institute (IRRI).

[27] Zhu, X. G., Long, S. P. and Ort, D. R. (2010) Improving photosynthetic efficiency for greater yield. Annual review of plant biology, 61, 235-261.

[28] Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., ... and Anson, E. L. (2000) A whole-genome assembly of Drosophila. Science, 287(5461), 2196-2204.

[29] Lohse, M., Drechsel, O. and Bock, R. (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current genetics, 52(5-6), 267-274.

[30] Kim, B., Kim, J., Park, H. and Park, J. (2016) The complete chloroplast genome sequence of Bienertia sinuspersici. Mitochondrial DNA Part B, 1(1), 388-389.

[31] Li, H., Cao, H., Cai, Y. F., Wang, J. H., Qu, S. P. and Huang, X. Q. (2014) The complete chloroplast genome sequence of sugar beet (Beta vulgaris ssp. vulgaris).

[32] Dong, W., Xu, C., Li, D., Jin, X., Li, R., Lu, Q. and Suo, Z. (2016) Comparative analysis of the complete chloroplast genome sequences in psammophytic Haloxylon species (Amaranthaceae). PeerJ, 4, e2699.

[33] Schmitz-Linneweber, C., Maier, R. M., Alcaraz, J. P., Cottet, A., Herrmann, R. G. and Mache, R. (2001) The plastid chromosome of spinach (Spinacia oleracea): complete nucleotide sequence and gene organization. Plant Molecular Biology, 45(3), 307-315.

[34] Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research, 27(2), 573.

[35] Suzuki, J. Y., Sriraman, P., Svab, Z. and Maliga, P. (2003) Unique architecture of the plastid ribosomal RNA operon promoter recognized by the multisubunit RNA polymerase in tobacco and other higher plants. The Plant Cell, 15(1), 195-205.

[36] Zhang, J. (2003) Evolution by gene duplication: an update. Trends in ecology & evolution, 18(6), 292-298.

[37] Kondrashov, F. A. and Kondrashov, A. S. (2006) Role of selection in fixation of gene duplications. Journal of Theoretical Biology, 239(2), 141-151.

[38] Cotton, J. A. (2008) The impact of gene duplication on human genome evolution. eLS.

[39] Freeling, M. and Subramaniam, S. (2009) Conserved noncoding sequences (CNSs) in higher plants. Current opinion in plant biology, 12(2), 126-132.

[40] Guo, H. and Moose, S. P. (2003) Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. The Plant Cell, 15(5), 1143-1158.

[41] Thomas, B. C., Rapaka, L., Lyons, E., Pedersen, B., and Freeling, M. (2007) Arabidopsis intragenomic conserved noncoding sequence. Proceedings of the National Academy of Sciences, 104(9), 3348-3353.

[42] Bock, R. (2007) Structure, function, and inheritance of plastid genomes. In Cell and molecular biology of plastids (pp. 29-63). Springer Berlin Heidelberg.

[43] Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M. and Small, I. (2011) Rampant gene loss in the underground orchid Rhizanthella gardneri highlights evolutionary constraints on plastid genomes. Molecular Biology and Evolution, 28(7), 2077-2086.

[44] Keller, J., Rousseau-Gueutin, M., Martin, G. E., Morice, J., Boutte, J., Coissac, E., ... and Aïnouche, A. (2017) The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from Lupinus. DNA Research, 24(4), 343-358.

[45] Allen, J. F., Puthiyaveetil, S., Ström, J. and Allen, C. A. (2005) Energy transduction anchors genes in organelles. Bioessays, 27(4), 426-435.

[46] Maier, U. G., Zauner, S., Woehle, C., Bolte, K., Hempel, F., Allen, J. F. and Martin, W. F. (2013) Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. Genome biology and evolution, 5(12), 2318-2329.

[47] Deno, H. and Sugiura, M. (1984) Chloroplast tRNAGly gene contains a long intron in the D stem: Nucleotide sequences of tobacco chloroplast genes for tRNAGly (UCC) and tRNAArg (UCU). Proceedings of the National Academy of Sciences, 81(2), 405-408.

[48] Downie, S. R., Olmstead, R. G., Zurawski, G., Soltis, D. E., Soltis, P. S., Watson, J. C. and Palmer, J. D. (1991) Six independent losses of the chloroplast DNA rpl2 intron in dicotyledons: molecular and phylogenetic implications. Evolution, 45(5), 1245-1259.

[49] Logacheva, M. D., Samigullin, T. H., Dhingra, A. and Penin, A. A. (2008) Comparative chloroplast genomics and phylogenetics of Fagopyrum esculentum ssp. ancestrale–a wild ancestor of cultivated buckwheat. BMC Plant Biology, 8(1), 59.

[50] Rock, C. D., Barkan, A. and Taylor, W. C. (1987) The maize plastid psbB-psbF-petB-petD gene cluster: spliced and unspliced petB and petD RNAs encode alternative products. Current genetics, 12(1), 69-77.

[51] Jo, B. S. and Choi, S. S. (2015) Introns: the functional benefits of introns in genomes. Genomics & informatics, 13(4), 112-118.

[52] Gruss, P., Lai, C. J., Dhar, R. and Khoury, G. (1979) Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40. Proceedings of the National Academy of Sciences, 76(9), 4317-4321.

[53] Callis, J., Fromm, M. and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. Genes & development, 1(10), 1183-1200.

[54] Beaulieu, E., Green, L., Elsby, L., Alourfi, Z., Morand, E. F., Ray, D. W. and Donn, R. (2011) Identification of a novel cell type-specific intronic enhancer of macrophage migration inhibitory factor (MIF) and its regulation by mithramycin. Clinical & Experimental Immunology, 163(2), 178-188.

[55] Sorek, R. and Ast, G. (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Research, 13(7), 1631-1637.

[56] Pan, Q., Shai, O., Lee, L. J., Frey, B. J. and Blencowe, B. J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nature genetics, 40(12), 1413.

[57] Roy, M., Kim, N., Xing, Y. and Lee, C. (2008) The effect of intron length on exon creation ratios during the evolution of mammalian genomes. Rna, 14(11), 2261-2273.

[58] Valencia, P., Dias, A. P. and Reed, R. (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. Proceedings of the National Academy of Sciences, 105(9), 3386-3391.

[59] Schwartz, S., Meshorer, E., & Ast, G. (2009). Chromatin organization marks exon-intron structure. Nature Structural and Molecular Biology, 16(9), 990.

[60] Spies, N., Nielsen, C. B., Padgett, R. A. and Burge, C. B. (2009) Biased chromatin signatures around polyadenylation sites and exons. Molecular cell, 36(2), 245-254.

[61] Tutar, Y. (2012). Pseudogenes. Comparative and functional genomics, 2012.

[62] Dong, W., Xu, C., Cheng, T. and Zhou, S. (2013) Complete chloroplast genome of Sedum sarmentosum and chloroplast genome evolution in Saxifragales. PLoS One, 8(10), e77965.

[63] Millen, R. S., Olmstead, R. G., Adams, K. L., Palmer, J. D., Lao, N. T., Heggie, L., ... and Calie, P. J. (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. The Plant Cell, 13(3), 645-658.

[64] Christopher, D. A., Cushman, J. C., Price, C. A. and Hallick, R. B. (1988) Organization of ribosomal protein genes rp123, rp12, rps19, rp122 and rps3 on the Euglena gracilis chloroplast genome. Current genetics, 14(3), 275-286.

[65] Thomson, W. W. and Whatley, J. M. (1980) Development of nongreen plastids. Annual Review of Plant Physiology, 31(1), 375-394.

[66] Steane, D. A. (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, Eucalyptus globulus (Myrtaceae). DNA Research, 12(3), 215-220.

[67] Palmer, J. D. and Thompson, W. F. (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell, 29(2), 537-550.

[68] Perry, A. S. and Wolfe, K. H. (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. Journal of Molecular Evolution, 55(5), 501-508.

[69] Liu, X. Q., Gillham, N. W. and Boynton, J. E. (1989) Chloroplast ribosomal protein gene rps12 of Chlamydomonas reinhardtii. Wild-type sequence, mutation to streptomycin resistance and dependence, and function in Escherichia coli. Journal of Biological Chemistry, 264(27), 16100-16108.