

# Comprehensive analysis of chloroplast intron-containing genes and conserved splice sites in dicot and monocot plants

• **Nguyen Dinh Sy**

Tay Nguyen University; Chonnam National University

• **Hunseung Kang**

Chonnam National University

(Received on June 5<sup>th</sup> 2016, accepted on April 14<sup>th</sup> 2017)

## ABSTRACT

Despite the increasing knowledge on the importance of the intron splicing of chloroplast genes during plant growth and stress responses, identification of intron-containing chloroplast genes and determination of splice sites in chloroplast introns are still lacking. Here, we carried out a comprehensive analysis of the chloroplast genome sequences in important plants and crops, including four dicots (*Arabidopsis thaliana*, *Coffea arabica*, *Nicotiana tabacum*, and *Panax schinseng*) and four monocots (*Musa acuminata*, *Oryza sativa*, *Triticum aestivum*, and *Zea mays*). The results showed that both dicot and monocot chloroplast genomes harbor 6 intron-containing tRNAs (*trnA*, *trnG*, *trnI*, *trnK*, *trnL*, and *trnV*) and 10-12 intron-containing mRNAs (*atpF*, *rpl2*, *rpl16*, *rps16*, *ndhA*, *ndhB*, *petB*, *petD*, *rpoC1*, *rps12*, *ycf3*, and *clpP*). Notably, *rpoC1* and

*clpP* lacked introns in monocot plants, except *M. acuminata*. Analysis of the nucleotide sequences of chloroplast introns revealed that the 5'-splice sites, 3'-splice sites, and branch-point sites of the chloroplast introns were highly conserved among dicots and monocots. Notably, the 5'-splice sites and 3'-splice sites of the chloroplast introns were similar to those of the nuclear U12 introns, whereas the branch-point sites of the chloroplast introns were homologous to those of the nuclear U2 introns. Taken together, these results indicated that the chloroplast genomes contained strictly limited intron-containing genes with conserved splice sites, suggesting that splicing of chloroplast introns was important for chloroplast biogenesis and function in both dicot and monocot plants.

**Keywords:** chloroplast, dicot, intron, monocot, splicing

## INTRODUCTION

The chloroplast, the green plastid that is found only in plant and algal cells, is not only a major cellular organelle for photosynthesis but also plays important roles in many aspects of plant physiology and development, such as the biosynthesis of phytohormones, amino acids, fatty acids, and vitamins, the storage of a variety of products, the assimilation of sulfur and nitrogen, and function as a global sensor of abiotic stresses [1]. It was

hypothesized that plastids are derived from endosymbiotic cyanobacteria [2], and cyanobacteria-origin of chloroplasts was firmly illustrated by recent molecular phylogenetic analysis [3]. The chloroplast is a double membrane-bound organelle that contains the thylakoid system where the process of light reaction of photosynthesis occurs. The chloroplast is semi-autonomous and contains a single circular DNA as its own genome. Since the first chloroplast genome

was sequenced in tobacco (*Nicotiana tabacum*) which consists of 155,844 bp containing 4 rRNA genes, 30 tRNA genes, and 50 protein coding genes [4], the complete chloroplast genome sequences over 800 plants and algae have been determined and deposited in the NCBI database (<http://www.ncbi.nlm.nih.gov>), including some important crop plants and model plants in both monocot and dicot species such as *Arabidopsis thaliana* (154,478 bp containing 4 rRNA, 37 tRNA, and 87 protein coding genes) [5], *Coffea arabica* (155,189 bp containing 4 rRNA, 35 tRNA, and 91 protein coding genes) [6], *Panax schinseng* (156,318 bp containing 4 rRNA, 30 tRNA, and 70 protein coding genes) [7], *Musa acuminata* (169,972 bp containing 4 rRNA, 30 tRNA, and 79 protein coding genes) [8], *Oryza sativa* (134,525 bp containing 4 rRNA, 30 tRNA, and 76 protein coding genes) [9], *Triticum aestivum* (134,540 bp containing 4 rRNA, 30 tRNA, and 71 protein coding genes) [10], and *Zea mays* (140,387 bp containing 4 rRNA, 30 tRNA, and 70 protein coding genes) [11]. The chloroplast genome is transcribed as polycistronic units by plastid-encoded and nuclear-encoded RNA polymerases [12,13], which encodes approximately 120–140 genes that participate in photosynthesis, transcription, and translation of chloroplast genes. Importantly, it has been demonstrated that chloroplast gene expression is regulated mainly at the posttranscriptional level, such as RNA splicing, RNA processing, RNA editing, RNA degradation, and translation [14–16].

RNA splicing is the process of cutting introns out of precursor RNAs (pre-RNAs) and ligating the exons together to form mature RNA, which is one of the most important posttranscriptional regulations of the gene expression in the chloroplast as well as in the nucleus. For accurate splicing to occur, specific signals on RNA precursors must exist to identify where to “cut and paste”, and many previous studies have revealed the consensus sequences at 5'-splice site, 3'-splice site, and branch-point site found in the

introns of nuclear mRNAs: introns in mammals contain the conserved sequences 5'-AG/GUAAGU-intron-YNCURAC-Y<sub>n</sub>NYAG/G-3', introns in plants contain the conserved sequences 5'-AG/GUAAGU-intron-CRUAY-GCAG/G-3', and introns in yeast contain the conserved sequences 5'-AG/GUAAGU-intron-UACUAAAC-YAG/-G-3, where N is any bases, Y and R is either pyrimidine (U or C) or purine (A or G), respectively, and A is the conserved adenine nucleotide at branch-point site [17].

The introns found in approximately 20 chloroplast genes are classified as group II introns, except a single group I intron found in *trnL* gene, by virtue of the conserved features of primary sequences and predicted secondary structures [18–23]. Although chloroplast introns belong to group I or group II introns, the splicing of which occurs via a self-splicing mechanism, splicing of chloroplast introns is not self-splicing but is dependent on many nuclear-encoded protein factors [23, 24]. Because the splicing of chloroplast group II introns differs from that of self-splicing group II introns, it is interesting to determine whether chloroplast introns contain splicing signals similar to or different from the signals found in nuclear introns. To answer this question, we analyzed the nucleotide sequences in exon-intron junctions of chloroplast genes in selected dicot and monocot plants, and report that the 5'-splice sites, 3'-splice sites, and branch-point sites of the chloroplast introns are highly conserved among dicots and monocots.

## MATERIALS AND METHODS

### Retrieval of intron-containing genes from the chloroplast genome sequences

The maps of the chloroplast genomes in diverse plant species, including four dicot species (*Arabidopsis thaliana* [5], *Coffea arabica* [6], *Nicotiana tabacum* [4], and *Panax schinseng* [7]) and four monocot species (*Musa acuminata* [8], *Oryza sativa* [9], *Triticum aestivum* [10], and *Zea mays*

[11]), were obtained from the indicated references, and the nucleotide sequences of intron-containing tRNA and mRNA genes described in each reference were downloaded from the National Center for Biotechnology Information (NCBI) database (<http://ncbi.nlm.nih.gov>).

#### **Analysis of the 5'-splice site, 3'-splice site, and branch-point site of chloroplast introns**

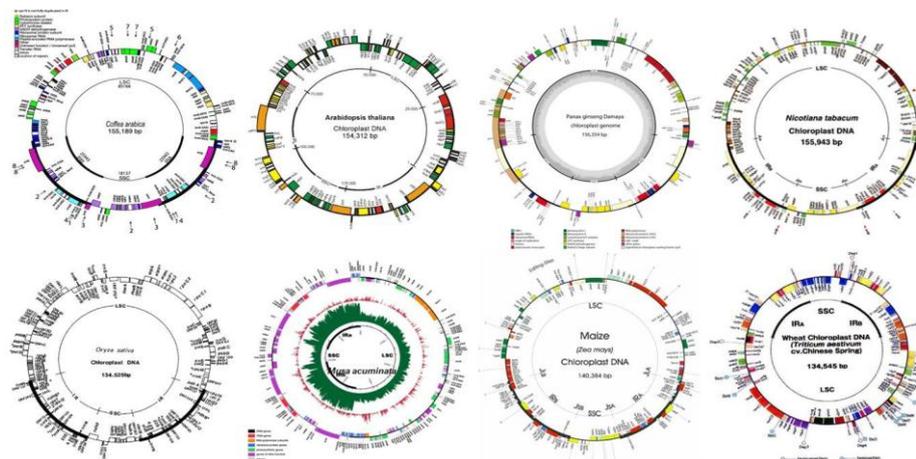
For the identification of conserved intron sequences at the 5'- and 3'-splice sites, the sequences at the 5'-end of introns spanning 5 nucleotides (nts) in the exon and 10 nts in the intron, and the sequences at the 3'-end of introns spanning 10 nts in the intron and 5 nts in the exon were extracted from the predicted exon-intron junction sequences, and the sequences were analyzed using the WEBLOGO program (<http://weblogo.threeplusone.com>). For the prediction of branch-point sites, approximately 100 nts upstream of the 3' splice sites were compared with the previously confirmed branch-point sequences [25-26], and the sequences showing a high similarity were extracted and analyzed using the WEBLOGO program.

### **RESULTS AND DISCUSSION**

#### **Identification and characterization of chloroplast intron-containing genes**

The genomes of plant chloroplasts are circular DNA consisting of approximately 130,000–160,000 base pairs and harbor approximately 140 genes,

among which 16 to 20 genes contain introns. To obtain information on which genes contain introns, the maps of chloroplast genomes in diverse plant species, including four dicot species (*Arabidopsis thaliana*, *Coffea arabica*, *Nicotiana tabacum*, and *Panax schinseng*) and four monocot species (*Musa acuminata*, *Oryza sativa*, *Triticum aestivum*, and *Zea mays*), were obtained from indicated references, and the structures of chloroplast genomes were analyzed (Fig. 1). The results showed that there are six intron-containing tRNA genes (*trnA*, *trnG*, *trnI*, *trnK*, *trnL*, and *trnV*) in the chloroplasts of both dicot and monocot plants. Monocot plants contain ten intron-containing mRNA genes (*atpF*, *rpl2*, *rpl16*, *rps16*, *ndhA*, *ndhB*, *petB*, *petD*, *rps12*, and *ycf3*), whereas dicot plants harbor twelve intron-containing genes with two additional intron-containing genes (*rpoC1* and *clpP*) in addition to the ten intron-containing mRNA genes found in monocot plants. However, *M. acuminata*, a monocot plant, retains introns in *rpoC1* and *clpP* genes (Table 1). Notably, three mRNA genes, *rps12*, *ycf3*, and *clpP*, contained two introns. Interestingly, the *rps12* is split into three separate parts on the chloroplast genome; the exon 2 and exon 3 are separated by a cis-intron and are transcribed with the downstream *rps7*, whereas the exon 1 is co-transcribed with the upstream *clpP* and the downstream *rpl20* genes, after which the two separate transcripts are joined together by a trans-splicing process to form mature *rps12* mRNA [27].



**Fig. 1.** Structure of chloroplast genomes. The maps of chloroplast genomes were obtained from different plant species; for dicots, *Arabidopsis thaliana* [5], *Coffea arabica* [6], *Nicotiana tabacum* [4], and *Panax schinseng* [7]), and for monocots, *Musa acuminata* [8], *Oryza sativa* [9], *Triticum aestivum* [10], and *Zea mays* [11]. Each genome map shows the length of genome and the number of genes.

**Table 1.** Names of the intron-containing chloroplast genes in different plant species

Gene names	Dicot species				Monocot species			
	<i>Arabidopsis thaliana</i>	<i>Coffea arabica</i>	<i>Nicotiana tabacum</i>	<i>Panax schinseng</i>	<i>Musa acuminata</i>	<i>Oryza sativa</i>	<i>Triticum aestivum</i>	<i>Zea mays</i>
<i>trnA</i>	X	X	X	X	X	X	X	X
<i>trnG</i>	X	X	X	X	X	X	X	X
<i>trnI</i>	X	X	X	X	X	X	X	X
<i>trnK</i>	X	X	X	X	X	X	X	X
<i>trnL</i>	X	X	X	X	X	X	X	X
<i>trnV</i>	X	X	X	X	X	X	X	X
<i>atpF</i>	X	X	X	X	X	X	X	X
<i>rpl2</i>	X	X	X	X	X	X	X	X
<i>rpl16</i>	X	X	X	X	X	X	X	X
<i>rps16</i>	X	X	X	X	X	X	X	X
<i>ndhA</i>	X	X	X	X	X	X	X	X
<i>ndhB</i>	X	X	X	X	X	X	X	X
<i>petB</i>	X	X	X	X	X	X	X	X
<i>petD</i>	X	X	X	X	X	X	X	X
<i>rpoC1</i>	X	X	X	X	X			
<i>rps12</i> *	X	X	X	X	X	X	X	X
<i>ycf3</i> *	X	X	X	X	X	X	X	X
<i>clpP</i> *	X	X	X	X	X			

<sup>x</sup> denotes the genes whose sequences were included in the present analysis.

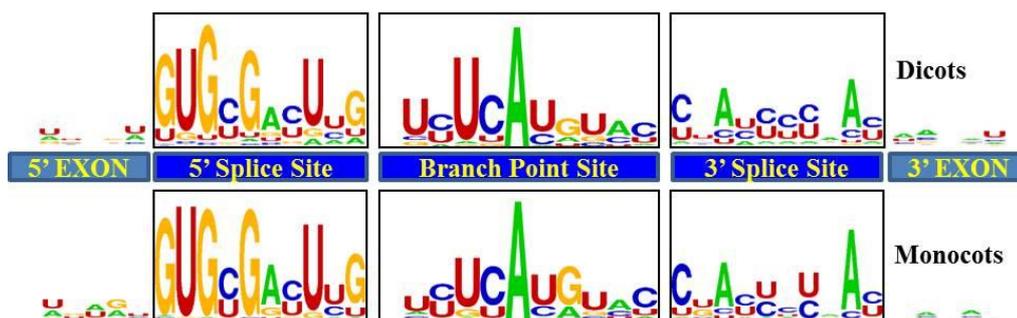
\* denotes the genes having two introns.

**Consensus sequences in the 5'-splice sites, 3'-splice sites, and branch-point sites of the chloroplast introns**

To obtain the information on the sequences in the 5'-splice sites, 3'-splice sites, and branch-point sites of the chloroplast introns, nucleotide sequences of 48 intron-containing chloroplast mRNAs in dicots and 42 intron-containing chloroplast mRNAs in monocots (Table 1) were downloaded from NCBI database. The sequences of 10 nucleotides at the 5'- and 3'-ends of each intron were analyzed using the WEBLOGO software. The results showed that GU at the 5'-end and (A/C)(C/U) at the 3'-end were highly conserved in both dicot and monocot plants (Fig. 2). These predicted 5'- and 3'-splice sites of chloroplast introns are slightly different from those found in the splice sites of nuclear introns (Fig. 3). The major U2-type introns contain the highly conserved GU at the 5'-splice sites and AG at the 3'-splice sites, whereas the minor U12-type introns contain the conserved

(G/A)U at the 5'-splice sites and A(G/C) at the 3'-splice sites [26, 28] (Fig. 3).

We then analyzed the intron sequences to identify putative branch-point sites. The sequences of approximately 100 nucleotides at the upstream of the 3'-splice sites were selected and aligned with the known branch-point sequences of humans and plants [25, 26]. The sequences of putative branch-point sites are showed in Fig. 2. Seventy two out of the 86 putative branch-point sites (83.7 %) were found at approximately 60–40 nucleotides upstream of the 3'-splice sites. Analysis of the nucleotide sequences using the WEBLOGO software revealed that (C/U)U(C/U)A(U/C) is conserved at the branch-point sites in the chloroplast introns of both dicot and monocot plants (Fig. 2). These predicted branch-point sites in chloroplast introns are not identical to those found in nuclear introns that contain the conserved (C/U)U(A/G)A(U/C) in the U2-type introns and the UU(A/G)A(U/C) in the U12-type introns (Fig. 3).



**Fig. 2.** Conserved sequences of splicing sites and branch point sites in the introns of chloroplast genes. Intron sequences in 48 mRNAs from 4 dicot species (*Arabidopsis thaliana*, *Coffea arabica*, *Nicotiana tabacum*, and *Panax schinseng*) and in 42 mRNAs from 4 monocot species (*Musa acuminata*, *Oryza sativa*, *Triticum aestivum*, and *Zea mays*) were analyzed using the WEBLOGO program (<http://weblogo.threeplusone.com>). The height of the letters at each nucleotide position indicates the degree of conserved sequence.

Species	U2-type introns			U12-type introns		
	5' splice site	3' splice site	Branch site	5' splice site	3' splice site	Branch site
<i>Arabidopsis thaliana</i>						
<i>Glycine max</i>						
<i>Vitis vinifera</i>						
<i>Oryza sativa</i>						
<i>Zea mays</i>						
<i>Physcomitrella patens</i>						
<i>Selaginella moellendorffii</i>						
<i>Chlamydomonas reinhardtii</i>						

**Fig. 3.** Consensus sequences in the 5'-splice, 3'-splice, and branch-point sites of the nuclear introns. The nucleotide sequences in the U2-type and U12-type introns in plants, moss, or alga were analyzed using the WEBLOGO program. The height of the letters at each nucleotide position indicates the degree of conserved sequence [26].

Correct splicing of introns in pre-RNAs is one of the most important steps for the regulation of gene expression in the chloroplast as well as in the nucleus. Nucleotide sequences in the 5'-splice sites, 3'-splice sites, and branch-point sites are highly conserved in the nuclear introns of plants [26, 28] (Fig. 3). By contrast, our current analysis revealed that the 5'- and 3'-splice sites in chloroplast introns are less conserved than those in nuclear introns (Fig. 2), and that the 5'- and 3'-splice sites in chloroplast introns are more similar to those in U12-type introns than to those in U2-type introns (Fig. 3). Notably, more variations were observed in the sequences and positions of branch-point sites between nuclear and chloroplast introns. The branch-point sites of the nuclear U12-type introns in plants harbor the conserved -UUnAn- sequences and are located approximately 10-16 nucleotides upstream of the 3'-end of the U12 introns, whereas the branch-point sites of the nuclear U2-type introns harbor the conserved -

nUnAn- sequences and are located approximately 20-40 nucleotides upstream of the 3'-end of the U2 introns [26, 29] (Fig. 3). Interestingly, the sequences in the branch-point sites of chloroplast introns in both dicots and monocots are similar to those of the U2 introns (Fig. 2), whereas the positions of branch-point sites are quite different in that approximately 84% of chloroplast introns have the branch-point sites approximately 40-60 nucleotides upstream of the 3'-end of introns. These conserved and divergent sequences in the 5'-splice, 3'-splice, and branch-point sites of the nuclear and chloroplast introns suggest that the nucleus and chloroplasts not only share common machineries for intron splicing, but also harbor specific components to mediate intron splicing in each organelle. It would be interesting to determine how the differences in intron sequences between the nucleus and chloroplasts are recognized, and how the introns are correctly spliced out by different splicing complexes.

## CONCLUSIONS

Analysis of the chloroplast genome sequences revealed that introns are found in identical genes in dicot and monocot plants, and that the 5'-splice sites, 3'-splice sites, and branch-point sites of chloroplast introns are highly conserved in dicot and monocot plants. Given that the chloroplast introns contain both identical and divergent sequences compared with nuclear introns, it would be interesting to

investigate how the intron sequences in chloroplasts are recognized by specific protein factors and to determine how the splicing of chloroplast introns is modulated by developmental and environmental cues.

*Acknowledgements: This work was supported by the grant from the Next-Generation BioGreen21 Program (PJ01103601), Rural Development Administration, Republic of Korea.*

# Phân tích gene mang trình tự intron trong bộ gene lục lạp và trình tự bảo tồn tại vị trí “cắt và dán” giữa exon-intron ở thực vật một lá mầm và hai lá mầm

• **Nguyễn Đình Sỹ**

Trường Đại học Tây Nguyên; Đại học Quốc gia Chonnam, Hàn Quốc

• **Hunseung Kang**

Đại học Quốc gia Chonnam, Hàn Quốc

## TÓM TẮT

Cơ chế điều hòa loại bỏ intron để tạo mRNA trưởng thành ảnh hưởng đến quá trình sinh trưởng và khả năng đáp ứng với môi trường bất lợi của thực vật được nghiên cứu ngày càng tăng, tuy vậy, việc xác định gene chứa intron và trình tự tại vị trí “cắt dán” của exon-intron còn hạn chế. Trong nghiên cứu này, bộ gene lục lạp của 4 loài thực vật 2 lá mầm (*Arabidopsis thaliana*, *Coffea arabica*, *Nicotiana tabacum*, and *Panax schinseng*) và 4 loài thực vật 1 lá mầm (*Musa acuminata*, *Oryza sativa*, *Triticum aestivum*, and *Zea mays*) được chọn lọc để phân tích, kết quả cho thấy có tổng số 6 tRNA gene (*trnA*, *trnG*, *trnI*, *trnK*, *trnL*, and *trnV*) và 10-12 mRNA gene mang

trình tự intron (*atpF*, *rpl2*, *rpl16*, *rps16*, *ndhA*, *ndhB*, *petB*, *petD*, *rpoC1*, *rps12*, *ycf3*, and *clpP*). Đặc biệt, *rpoC1* và *clpP* mang trình tự intron ở thực vật 2 lá mầm, trong khi thực vật 1 lá mầm, trình tự này được loại bỏ trong quá trình tiến hóa ngoại trừ loài *M. acuminata*. Phân tích trình tự “cắt-dán” của exon-intron tại vị trí 5', 3' và điểm nhánh (branch-point) cho thấy trình tự được bảo tồn giữa các gene và giữa thực vật 1 lá mầm và 2 lá mầm. Hơn nữa, trình tự cắt tại vị trí 5' và 3' giống với trình tự U12 intron, nhưng điểm nhánh lại giống với trình tự U2 intron trong bộ gene nhân tế bào. Từ những kết quả trên cho thấy chỉ có một số lượng nhỏ gene chứa intron trong lục lạp

và trình tự “cắt-dán” được bảo tồn, do đó quá trình điều hòa loại bỏ intron để hình thành RNA trưởng thành là rất quan trọng trong quá trình hoạt động và

thực hiện chức năng của lục lạp đối với cả thực vật 1 lá mầm và 2 lá mầm.

**Từ khóa:** Exon, hai lá mầm, intron, lục lạp, một lá mầm

## REFERENCES

- [1]. B. Biswal, P.N. Joshi, M.K. Raval, U.C. Biswal, Photosynthesis, a global sensor of environmental stress in green plants: stress signaling and adaptation. *Curr. Sci.*, 101, 47–56 (2011).
- [2]. W. Martin, K.V. Kowallik, Annotated English translation of Mereschkowsky's 1905 paper: Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Eur. J. Phycol.*, 34, 287–295 (1999).
- [3]. W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, D. Penny, Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA*, 99, 12246–12251 (2002).
- [4]. K. Shinozaki, M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K.Y. Shinozaki, C. Ohto, K. Torazawa, K.B. Meng, M. Sugital, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, M. Sugiura, The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.*, 5, 2043–2049 (1986).
- [5]. S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, S. Tabata, Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.*, 6, 283–290 (1999).
- [6]. N. Samson, M.G. Bausher, S.B. Lee, R.K. Jansen, H. Daniell, The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotech. J.*, 5, 339–353 (2007).
- [7]. K.J. Kim, H.L. Lee, Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among vascular. *Plants DNA Res.*, 11, 247–261 (2004).
- [8]. G. Martin, F.C. Baurens, C. Cardi, J.M. Aury, A.L. D'Hont, The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): Insight into plastid monocotyledon evolution. *PLoS ONE* 8, e67350 (2013).
- [9]. J. Hiratsuka, H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C.R. Sun, B.Y. Meng, Y.Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, M. Sugiura, The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.*, 217, 185–194 (1989).
- [10]. Y. Ogihara, K. Isono, T. Kojima, A. Endo, M. Hanaoka, T. Shiina, T. Terachi, S. Utsugi, M. Murata, N. Mori, S. Takumi, K. Ikeo, T. Gojobori, R. Murai, K. Murai, Y. Matsuoka, Y. Ohnishi, H. Tajiri, K. Tsunewaki, Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: complete sequence and contig clones. *Plant Mol. Biol. Rep.*, 18, 243–253 (2000).
- [11]. R.M. Maier, K. Neckermann, G.L. Igloi, and K. Hans, Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.*, 251, 614–628 (1995).
- [12]. T. Shiina, Y. Tsunoyama, Y. Nakahira, M.S. Khan, Plastid RNA polymerases, promoters, and transcription regulators in higher plants. *Int. Rev. Cytol.*, 244, 1–68 (2005).

- [13]. K. Liere, T. Börner, Transcription of plastid genes. In: K.D. Grasser, (eds.), *Regulation of transcription in plants*, Blackwell Publishing, Oxford, 184–224 (2007).
- [14]. J. Nickelsen, Chloroplast RNA-binding proteins. *Curr. Genet.*, 43 392–399 (2003).
- [15]. D.L. Herrin, J. Nickelsen, Chloroplast RNA processing and stability. *Photosynth. Res.*, 82, 301–314 (2004).
- [16]. D.B. Stern, M. Goldschmidt-Clermont, M.R. Hanson, Chloroplast RNA metabolism. *Ann. Rev. Plant Biol.*, 61, 125–155 (2010).
- [17]. R.F. Weaver, RNA processing I: Splicing. In: *Molecular Biology (4 eds.)*, McGraw-Hill companies, Inc., 394–430 (2012).
- [18]. F. Michel, B. Dujon, Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J.*, 2, 33–38 (1983).
- [19]. F. Michel, K. Umesono, H. Ozeki, Comparative and functional anatomy of group II catalytic introns-a review. *Gene*, 82, 5–30 (1989).
- [20]. D.W. Copertino, R.B. Hallick, Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem. Sci.*, 18, 467–471 (1993).
- [21]. A.M. Lambowitz, M.G. Caprara, S. Zimmerly, P.S. Perlman, Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: R.F. Gesteland, T.R. Cech, J.F. Atkins (2 eds), *The RNA World*, Cold Spring Harbor Laboratory Press, 451–485 (1999).
- [22]. L. Bonen, J. Vogel, The ins and outs of group II introns. *Trends Genet.*, 17, 322–331 (2001).
- [23]. N.A. Eckardt, Chloroplast intron splicing mechanisms. *Plant Cell*, 19, 3838 (2007).
- [24]. C. Schmitz-Linneweber, A. Barkan, RNA splicing and RNA editing in chloroplasts. In: R. Bock (ed.), *Cell and Molecular Biology of Plastids*, 213–248 (2007).
- [25]. K. Gao, A. Masuda, T. Matsuura, K. Ohno, Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.*, 36, 2257–2267 (2008).
- [26]. M.W. Szczesniak, M. Kabza, R. Pokrzywa, A. Gudys, I. Makalowska, ERISdb: A database of plant splice sites and splicing signals. *Plant Cell Physiol.*, 54, 1–8 (2013).
- [27]. M. Sugita, M. Sugiura, Regulation of gene expression in chloroplasts of higher plants. *Plant Molecular Biology*, 32, 315–326 (1996).
- [28]. N. Sheth, X. Roca, M.L. Hastings, T. Roeder, A.R. Krainer, R. Sachidanandam, Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.*, 34, 3955–3967 (2006).
- [29]. Z.J. Lorkovic, D.A.W. Kirk, M.H.L. Lambermon, W. Filipowicz, Pre-mRNA splicing in higher plants. *Trends Plant Sci.*, 5 160–167 (2000).