



**HAL**  
open science

# The Intertwined Chloroplast and Nuclear Genome Coevolution in Plants

Mathieu Rousseau-Gueutin, Jean Keller, Julie Ferreira de Carvalho,  
Abdelkader Ainoüche, Guillaume Martin

► **To cite this version:**

Mathieu Rousseau-Gueutin, Jean Keller, Julie Ferreira de Carvalho, Abdelkader Ainoüche, Guillaume Martin. The Intertwined Chloroplast and Nuclear Genome Coevolution in Plants. Diah Ratnadewi; Dr. Hamim. Plant Growth and Regulation - Alterations to Sustain Unfavorable Conditions, InTechOpen, 2018, 978-1-78984-285-2. 10.5772/intechopen.75673 . hal-02787510

**HAL Id: hal-02787510**

**<https://hal.inrae.fr/hal-02787510v1>**

Submitted on 20 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

7,200

Open access books available

190,000

International authors and editors

205M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# The Intertwined Chloroplast and Nuclear Genome Coevolution in Plants

---

Mathieu Rousseau-Gueutin, Jean Keller,  
Julie Ferreira de Carvalho,  
Abdelkader Ainouche and Guillaume Martin

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.75673>

---

## Abstract

Photosynthetic eukaryotic cells arose more than a billion years ago through the engulfment of a cyanobacterium that was then converted into a chloroplast, enabling plants to perform photosynthesis. Since this event, chloroplast DNA has been massively transferred to the nucleus, sometimes leading to the creation of novel genes, exons, and regulatory elements. In addition to these evolutionary novelties, most cyanobacterial genes have been relocated into the nucleus, highly reducing the size, gene content, and autonomy of the chloroplast genome. In this chapter, we will first present our current knowledge on the origin and evolution of the plant plastome in the different Archaeplastida lineages (Glaucophyta, Rhodophyta, and Viridiplantae), focusing on its gene content, genome size, and structural evolution. Second, we will present the factors influencing the rate of DNA transfer from the chloroplast to the nucleus, the evolutionary fates of the nuclear integrants of plastid DNA (*nupts*) in their new eukaryotic environment, and the drivers of chloroplast gene functional relocation to the nucleus. Finally, we will discuss how cytonuclear interactions led to the intertwined coevolution of nuclear and chloroplast genomes and the impact of hybridization and allopolyploidy on cytonuclear interactions.

**Keywords:** endosymbiosis, plastome evolution, functional gene transfer, nuclear integrant of plastid DNA (*nupt*), nucleo-cytoplasmic interactions

---

## 1. Introduction

Photosynthetic eukaryotic organisms harbor a chloroplast genome (also called 'plastome') within their cells. This genome derives from the endosymbiosis of a prokaryotic organism, which was then gradually converted into the chloroplast. With the increased number of sequences within

---

publicly available databases and the emergence of very sophisticated phylogenetic and phylogenomic analyses, we can infer much more precisely the origin of this primary endosymbiotic event. In addition, these comparative analyses allow for investigation of plastome evolutionary dynamics in the different plant lineages and the extent of nuclear influence over the chloroplast genome. Overall, plant plastomes harbor a very low gene content compared to their prokaryotic ancestor, which appears to result from either gene loss due to redundant functions in both chloroplast and nuclear genomes or functional transfer and relocation of chloroplast genes into the nucleus. The relocation of thousands of chloroplast genes from the chloroplast to the nucleus was rendered possible due to the massive transfer of DNA from the chloroplast to the nucleus. However, chloroplast genes that have been integrated into the nucleus are not immediately functional and have to adapt to their new eukaryotic environment by acquiring various regulatory elements (i.e., promoter, polyadenylation signal, and target peptide). Despite most of these functional transfers occurred soon after the endosymbiotic event, some clever real-time experiments (using a selectable marker) have allowed for understanding how easily and by which molecular mechanisms DNA is transferred from the chloroplast to the nucleus. Such experiments have also permitted the study of the subsequent evolution of chloroplast DNA in the nuclear genome, and how a chloroplast gene becomes functional in the nucleus.

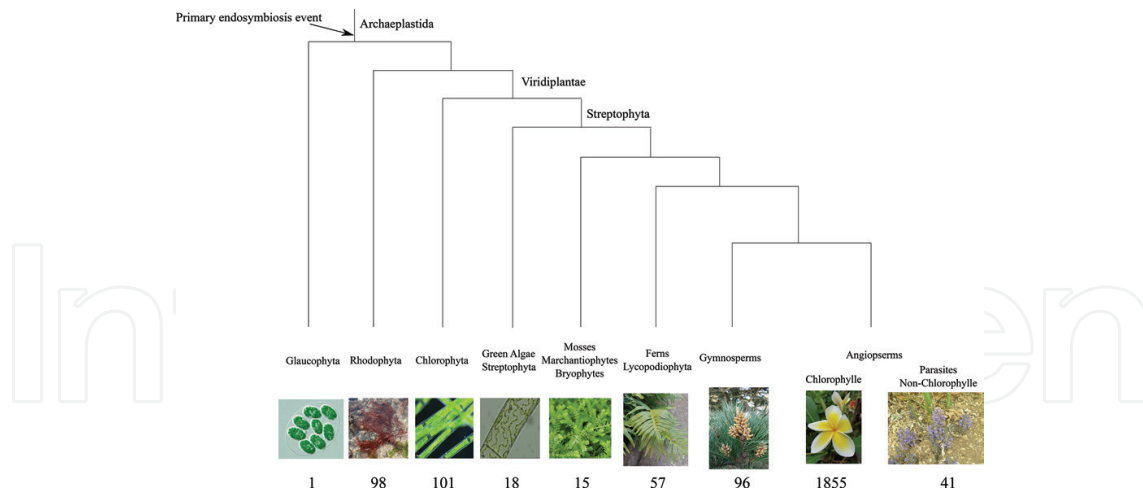
## 2. Chloroplast origin and evolution

Photosynthetic eukaryotic cells arose through the engulfment of a cyanobacterium that was then converted into the chloroplast, enabling plants to use sunlight to fix carbon. This major functional innovation allowed for eukaryotes to transition from heterotrophy to autotrophy. This primary endosymbiotic event is at the origin of the astonishing biodiversity visible today in plants, including the Glaucophyta, Rhodophyta, and Viridiplantae lineages (**Figure 1**). With the advent of next-generation sequencing technologies, the number of fully sequenced plastomes has hugely expanded, providing insight into chloroplast evolution in the different plant lineages. In this part, we will present our current knowledge on chloroplast origin and what has been unraveled on the chloroplast genome evolution, regarding genome size, gene content, structure, and mutation rate.

### 2.1. Primary endosymbiosis event and origin of chloroplasts

The first hypothesis of the endosymbiotic origin of chloroplasts is commonly credited to Russian botanist K. Mereschkowsky, who observed similarities between cyanobacteria and chloroplasts of plants and algae [1]. This hypothesis was then reaffirmed by Margulis in the 1970s. The origin of this primary endosymbiosis event is still debated. While fossil-based phylogeny estimated the origin of chloroplasts to be around 1.4–1.7 billion years ago [2], gene-based approaches dated it around 0.9 billion years ago [3]. Different phylogenetic analyses aimed at determining the cyanobacterial lineage from which the chloroplast was derived and revealed that chloroplasts were closely related to the nitrogen-fixing cyanobacteria Chroococcales, *Nostoc* sp., and *Anabaena variabilis* [4, 5].

It is now widely accepted that this primary endosymbiotic event has a single origin [6–8]; however, it is still unclear how long it took for the conversion of the bacterial endosymbiont into a fully integrated organelle. This transition from endosymbiont to organelle surely involved many



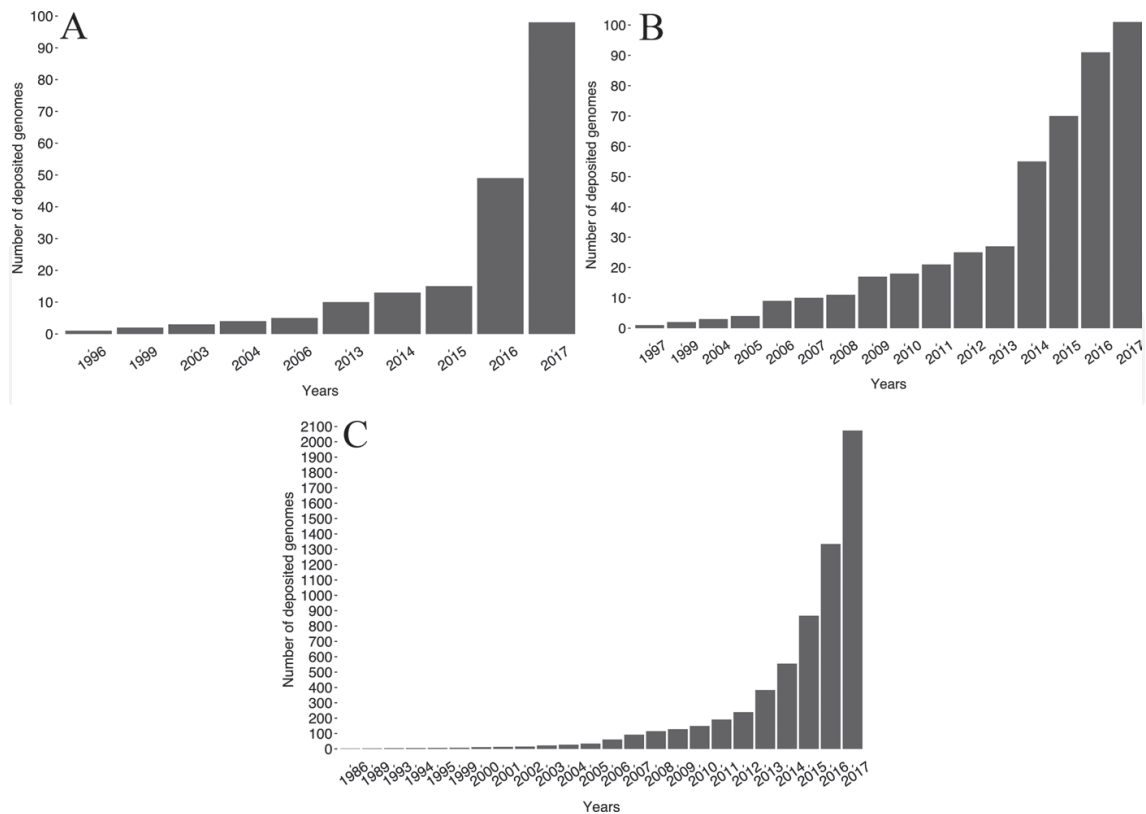
**Figure 1.** Phylogenetic relationships of the different plant lineages formed after the primary endosymbiosis of a cyanobacterium by an ancestor of the Archaeplastida. The number of available genomes on GenBank is indicated under the image. For simplicity, “Mosses, Marchantiophytes and Bryophytes” on one side, as well as “Ferns and Lycopodiophyta” on the other side, were grouped together in the tree. Pictures copyright to L. Brient, M.T. Misset, R. Delourme, and J. Keller.

steps. The first steps corresponded to the loss of the bacterial wall and the early acquisition by the endosymbiont of a transport system to transfer proteins and metabolites from the cytosol to the chloroplast. This latter step is constituted by two protein complexes: translocon of the outer (TOC) membranes of the chloroplast and translocon of the inner (TIC) membranes of the chloroplast [9–11]. The TIC/TOC complexes allow for transportation of the pre-proteins (proteins with a cleavable chloroplast target peptide) from the cytosol, where they are synthesized, to the chloroplast, where the target peptide is cleaved (reviewed in [11]). The presence of the same protein import apparatus in the different Archaeplastida lineages is the best evidence of the single origin of chloroplasts. Finally, the transition also necessitated the gradual functional transfer of endosymbiont genes to the nucleus [12], leading to the massive reduction of plastome size and gene content.

## 2.2. Evolution of chloroplast genomes

### 2.2.1. An unequal sequencing effort

Most of our current knowledge of the conversion from endosymbiont to organelle has been obtained by comparing contemporary Archaeplastida organelles with their closest bacterial relatives. During the last few years, advances in high-throughput sequencing and bioinformatic methods greatly facilitated the assembly, analysis, and publication of complete plastomes. To date, more than 2300 plastomes are fully assembled and deposited in the GenBank database. This number of plastomes actually doubled in the last 2 years. However, the number of sequenced plastomes varies greatly between the different Archaeplastida lineages. Indeed, almost 80% of them belong to Angiosperms. Thus, there is an important inequality in the sequencing effort. The poor level of plastome sequencing in plant lineages outside of the Angiosperms needs to be improved to fully understand chloroplast genome evolution in plants. Some efforts to fill this gap have been performed in the last 2–5 years, but they are still insufficient. In the Glaucophyta, only one chloroplast genome is available (NC\_001675), and another is sequenced but not yet published (Lang et al., unpublished). In



**Figure 2.** Cumulative numbers of full chloroplast genomes deposited in GenBank for (A) Rhodophyta, (B) Chlorophyta, and (C) Streptophyta.

contrast, the sequencing of Rhodophyta and Chlorophyta (green algae *sensu stricto*) species greatly improved since 2012: from less than 30 plastomes available in 2012 to around 100 in 2017 (**Figure 2A** and **B**).

### 2.2.2. Gene content evolution

As mentioned previously, the conversion of the cyanobacterial endosymbiont into a chloroplast necessitated the functional transfer or replacement of most cyanobacterial genes into the nucleus. Compared to the thousands of genes (at least 2000) thought to have been once present in the cyanobacterial genome, Archaeplastida plastomes encode a maximum of around 250 genes [13, 14]. This observation indicates that most genes (includes protein coding and structural RNAs) present in the cyanobacterial ancestor have been functionally transferred relatively soon after the endosymbiotic event. Despite gene content among modern chloroplast genomes being relatively well conserved, there are important variations. Thus, Rhodophyta have the highest number of genes (237 in average; minimum 207; up to 266 in *Grateloupia taiwanensis*) compared to the Glaucophyta (195), Chlorophyta (118 in average; minimum 68; maximum 210) or Streptophyta (129 in average; minimum 64; maximum 313), when excluding parasitic and non-chlorophyll species (**Table 1**).

These variations in gene content revealed the divergent evolution of plastomes in the different lineages. As an example, Rhodophyta gene content is characterized by the complete absence of the NADPH dehydrogenase complex [15]. Conversely, some genes are Rhodophyta-specific or rare in other Archaeplastida such as RNase P RNA, tmRNA, or signal recognition

	Glaucophyta	Rhodophyta	Chlorophyta	Green Algae Streptophyta	Mosses Marchantiophytes Bryophytes	Ferns Lycopodiophyta	Gymnosperms	Angiosperms	
								Chlorophyll	Parasites Non-Chlorophyll
Number of plastomes	1	98	101	18	15	57	96	1 855	41
Average genome size (bp)	135 599	183 209	161 975	156 310	129 090	15 004	130 398	152 782	71 736
<i>min/max genome size (bp)</i>		<i>90,243 / 610,063</i>	<i>37,454 / 521,168</i>	<i>107,236 / 207,850</i>	<i>108,007 / 161,162</i>	<i>127,840 / 157,260</i>	<i>107,122 / 166,341</i>	<i>113,064 / 242,575</i>	<i>11,348 / 128,921</i>
Average number of proteins	149	202	83	97	80	86	82	84	33
Average number of structural RNAs	43	35	35	40	45	42	39	45	30

**Table 1.** Plastome numbers and characteristics (average size, number of proteins, and structural RNAs) among the Archaeplastida. The minimum and maximum genome sizes are indicated in italic.

particle RNA [16–18]. Rhodophyta chloroplasts generally have a large genome size (see later) characterized by a high number of genes and other features such as the presence of bacteria-like operons, suggesting that Rhodophyta plastomes are phylogenetically closest to the ancestral cyanobacteria genome than any other algae [15]. Gene content variations are also well documented in the Angiosperm family in which multiple independent gene losses have been found such as *infA*, *ycf1*, *rps16*, and *accD* genes, which have been repeatedly lost in several lineages [19]. Within non-parasitic Angiosperms, a few families, such as the Fabaceae and the Campanulaceae, have recently lost various chloroplast genes [19]. In these lineages, recent gene losses from the plastome coincide with the transfer of those genes to the nucleus, providing insight into the underlying molecular mechanisms implicated in such events. For example, chloroplast gene loss may occur through a relaxed selective constraint on the chloroplast copy when a nuclear copy is already functional. This relaxation of selective constraint allows for non-sense mutations that may render the chloroplast copy non-functional [19, 20]. In addition, genes can become non-functional following the loss of their splicing capacity, as observed for *rps16* [21, 22]. The plastome gene content reduction is even more pronounced in non-chlorophyll organisms, such as parasites and obligate symbionts. Among the Angiosperms, 41 plastomes from parasitic plants have been sequenced and showed a great reduction in gene content (with only 63 genes) and size (around 70 kb in average), in line with the progressive loss of photosynthetic abilities. Similarly, plastome reduction is also observed among algae such as in the parasitic *Helicosporium* sp. (green algae) or *Choreocolax polysiphoniae* (red algae). On the contrary, increase in gene number may also be observed but in a lesser extent. In *Pelargonium*, which has among the highest number of chloroplast genes in Angiosperms (more than 180 in *P. transvaalense* and *P. hortorum*), there have been multiple duplication events in 39 genes [23]. Despite the number of coding sequences increased in the species belonging to this genus, this increased number of genes was due entirely to duplications and not to neo-functionalization processes.

### 2.2.3. Size variation

Among plants, chloroplast genomes range from less than 100 kb to more than 1 Gb, again excluding the non-chlorophyll species that exhibit significantly smaller chloroplast genomes (Table 1). The largest chloroplast genome ever sequenced has very recently been found in the red algae *Corynoplastis japonica*. Its genome size goes up to 1 Mb and contains 209 genes [24].

On average, the largest plastomes are found in the Rhodophyta with an average size of about 183 kb (minimum = 149,987 kb, maximum = 610,063 kb, excluding the small 90 kb genome of the parasite of *C. polysiphoniae*), whereas Glaucophyta and Streptophyta have an average chloroplast genome size of between 130 and 160 kb (minimum = 107,236 kb; maximum = 242,575 kb, excluding the parasitic and non-chlorophyll species), respectively (**Table 1**).

Several factors can explain the important size variations found among the Archaeplastida. In the case of the red algae *C. japonica* and *Bulboplastis apyrenoidosa* (more than 1 Mb and 600 kb long plastomes, respectively), the increase of plastome size is due to an expansion of the intron number with more than 200 introns found in these species [24]. In Angiosperms, plastome variations have been observed but in a lesser extent. For example, in *Pelargonium* that encompasses species with the largest chloroplast genomes found in Angiosperms (almost 243 kb), increased size is correlated to the expansion of the inverted repeats (IRs) that can be as long as 75 kb [23, 25]. This has also been observed in the Campanulaceae, *Lobelia thuliniana* [26], and *Musa acuminata* [27]. Expansion of plastomes has been linked to the presence of an increase number of repeats such as in *Trifolium* [28] or the Mimosoid *Acacia* and *Inga* [29]. This increase of plastome size by repeats is presumably the result of a less efficient chloroplast DNA repair mechanism [30, 31]. In contrast, plastome size reductions are also relatively common and can be due to loss of both coding and non-coding regions, especially in the non-chlorophyll species [32] that have an average plastome size of 71,736 bp in Angiosperms (**Figure 2**).

#### 2.2.4. Structural evolution

Among plants, most plastomes seem to exhibit a conserved quadripartite structure, with a large and small single copy separated by two inverted repeats (Palmer 1983). However, multiple rearrangements occurred in diverse lineages, which modified this conserved structure. One of the most striking examples is the loss of one IR that occurred multiple times in the different chloroplast-bearing lineages, such as in the Fabaceae and the Geraniaceae [30, 33, 34]. This has also been reported for different Gymnosperms species such as *Pseudotsuga menziesii*, *Pinus radiata*, *Cephalotaxus oliveri*, as well as in multiple lineages of Chlorophyta [35–37].

Chloroplast genome structure and gene order are also highly affected by inversions. Many inversions have been described in the literature, especially in legumes, with, for instance, fragments of 50 kb in the Papilionoideae [38], 36 kb in the Genistoids [39]; 29 kb in Sophoreae [40] or 7 kb in *Tylosema esculentum* [41]. Multiple inversions have also been found in Geraniaceae, Campanulaceae (more than 40 inversions detected), and other lineages [25, 42, 43]. Inversions can be caused through flip-flop recombinations between repeat sequences [39, 44].

#### 2.2.5. Evolution rates of plastomes

Chloroplast genomes are known to be highly conserved, with relatively low rates of mutations, especially when compared to the plant nuclear genome. Indeed, the chloroplast genome evolves on average 10 times slower than the nuclear genome [45], with about 1 or less mutation/kb/million years [46] compared with approximately 7 mutations/kb/million years for the nuclear genome [47]. However, there are some exceptions, especially in three Angiosperm families (i.e., Fabaceae, Campanulaceae, and Geraniaceae) that are known to have accelerated evolutionary rates of their plastomes along with multiple structural rearrangements and size



variations [19, 28, 30, 42, 44, 48, 49]. For example, the *ycf4* gene appears to be a hotspot of variation in *Lathyrus*, and this gene evolves 20 times faster than the rest of the chloroplast genome [19]. This localized hypermutable chloroplast region evolves even faster than the nuclear genome. Similarly, faster evolution has been observed in the *clpP* gene in Mimosoid [29]. In *Lupinus* (Fabaceae), two hypervariable regions have been identified (*ycf1* gene and *psaA-ycf4* region) and are characterized by high numbers of indels (with length usually superior to 20 bp) and mutations [22].

To sum up this first section on the origin and evolution of plant plastomes originating from the primary endosymbiosis event, the recent sequencing and bioinformatics progress significantly increased the number of chloroplast genomes available for the scientific community. These advances have greatly improved our knowledge about the evolutionary dynamics of plastomes. Despite the diversity of organisms that harbor chloroplasts, plastomes in general seem to be relatively well conserved among the Archaeplastida (in terms of structure, size, and gene content); however, multiple independent alterations of these features have been observed in the different lineages. In addition, a few plant families (or group of species) seem to present an atypical evolution of the chloroplast genome. It is certain that the continuous effort to sequence much more plastomes (especially in the Glaucophyta and Rhodophyta) will allow the identification of new examples of such atypical evolution and will permit a better understanding of what are the causes and the molecular mechanisms involved in limiting or increasing plastome evolution.

### **3. Impact of the cyanobacterial endosymbiosis on plant nuclear genome evolution and origin of chloroplast proteins**

Since the endosymbiotic event, the host genome (nuclear) has acquired most of the cyanobacterial genes, leading to the gradual loss of autonomy of the endosymbiont and the reduction of its genome. In this part, we will present our current knowledge on the mechanisms as well as the numerous cases of chloroplast DNA transfers to the nucleus and where it is now integrated in the nuclear genome. We will then detail the subsequent evolution and adaptation processes of the chloroplast genome that took place in its new eukaryotic environment. We will also discuss which factors can influence relocation of a chloroplast gene to the nucleus, and how a chloroplast gene transferred to the nucleus may become functional. Finally, we will discuss the important role that transfer of chloroplast DNA to the nucleus plays in the process of diversifying the plant nuclear gene content.

#### **3.1. DNA transfer from the chloroplast to the nucleus**

Much earlier than the complete sequencing and assembly of the first chloroplast genome (*Nicotiana tabacum*: [50]), Kawashim et al. [51] observed that the gene encoding the small subunit of the Rubisco chloroplast protein could be transferred by pollen and thus must be encoded in the nucleus. From this early observation arose the question of whether nuclear genes encoding chloroplast proteins were of eukaryotic origin or resulted from transfer of DNA from the chloroplast to the nucleus. The existence of DNA transfer from the chloroplast to the nucleus was discovered a decade later using Southern Blot, by observing the presence of sequences with high homology between spinach (Chenopodiaceae) chloroplast

and nuclear genomes [52, 53], as well as in other closely related Chenopodiaceae species [54]. With the advent of the polymerase chain reaction, Ayliffe & Timmis [54] amplified and sequenced a chloroplast DNA sequence from *N. tabacum* nuclear DNA. This nuclear integrant of plastid DNA (also called 'nupt') presented more than 99% homology with its homologous chloroplast sequence, indicating that this chloroplast DNA fragment had been transferred to the *Nicotiana* nucleus during the last million year. Using similar techniques, these authors also observed that the tobacco nuclear genome contained long tracts of chloroplast DNA at different locations. These different *nupts* may be as large as the whole chloroplast genome (about 150 kb) and the different *nupts* did not consist of the same sequence homology to the chloroplast homologous sequence, indicating that chloroplast DNA had been transferred at multiple times to the nucleus during plant evolution [54]. To decipher how frequently chloroplast DNA is transferred to the nucleus, experiments using an antibiotic resistance gene tailored for nuclear expression (i.e., nuclear promoter and terminator) were performed [55, 56]. After introducing this selectable marker (antibiotic resistance gene) into *N. tabacum* chloroplast genome and obtaining homoplastomic lines, it was demonstrated that DNA transfer occurred once in about 16,000 pollen grains [55] or once for every 5 million somatic cells [56], highlighting the high rate of DNA transfer from the chloroplast to the nucleus. This deluge of DNA transfer may be even higher in the presence of environmental stresses, such as mild heat [57] or cold stress [58]. It is important to note that in these experiments, the reported transfer rate may be underestimated as only the transfer of the selectable marker (about 2 kb) from the chloroplast genome could be identified. The higher rate of transfer observed in reproductive tissue (from pollen grains) compared to somatic cells may be explained by the higher degree of degradation of chloroplast DNA during pollen development (since chloroplast genomes are maternally inherited) than in somatic cells (more stable plastids). This hypothesis was supported by the observation of a much lower frequency of DNA transfer from female germlines (about 1 every 270,000 ovules) [59]. Some of these newly transferred chloroplast sequences were characterized and demonstrated that integration occurred by non-homologous end joining [60] and predominantly in open chromatin [61]. Surprisingly, it has also been demonstrated that DNA fragments from various plastome regions may insert simultaneously at the same nuclear location [60].

### 3.2. Short-term and long-term evolution of chloroplast DNA transferred to the nucleus

Some of the chloroplast DNA fragments that were experimentally shown to insert in the nuclear genome were characterized [55, 60] and were often large in size (usually greater than 10 kb in length). Considering the massive transfer of chloroplast DNA to the nucleus, one would expect that some of these *nupts* would be deleted to avoid a rapid increase of the nuclear genome size. This hypothesis was tested by studying the fate of these newly integrated chloroplast fragments [62]. Half of the lines presented an unstable inheritance of the *nupts*, after only one to two generations. Most lines presented a varying level of instability between the different areas of the same plant, and the loss of the *nupt* most often occurred during somatic cell division. However, it was also observed that some *nupt* loss occurred during meiosis [62]. Thus, even if constantly and massively integrated into the nucleus, at least some of these novel *nupts* may be rapidly removed and likely an even larger number

may be deleted over longer evolutionary time scales. Many *nupts* have been identified in various sequenced plant nuclear genomes [63–69], by fluorescent in situ hybridization using a chloroplast DNA probe [70] or using PCR-derived methods [71, 72]. Using the nuclear genome sequences of 17 plant species, the number, size, and genomic organization of *nupts* were studied [65]. They found a positive correlation between nuclear genome size, organelle numbers in cells, and cumulative lengths of *nupts*, as previously observed from a smaller number of plant nuclear genomes [64, 67, 73, 74]. To date, the largest identified *nupt* was found in the rice nuclear genome (131 kb) and corresponds to almost the entire chloroplast genome size (97.4%). A detailed analysis of the *nupts* presents in the rice genome revealed that *nupts* were mainly integrated within the pericentromeric regions [68]. Thereafter, they were rapidly fragmented, vigorously shuffled, and 80% of them were eliminated in the million years following their integration. Accordingly, the largest *nupts* were found to be the youngest, whereas the smallest *nupts* were found to be older. Most of the *nupts* identified in rice were less than 1 million years old (myo), whereas only a few were older than 5 myo. The recently integrated *nupts* were assumed to be decaying over evolutionary time into smaller fragments [64]. In rice, the half-lives of *nupts* were evaluated to be 0.5 myo for fragments whose length is superior to 1.6 kb and 2.2 myo for fragments with length inferior to 1.6 kb [68]. This result differs from those obtained experimentally in *N. tabacum*, where several old *nupts* (up to 6 myo) were larger than 2 kb [71]. The evolutionary fate of *nupt* sequences was scrutinized and revealed the prevalence of G:C → A:T transitions, which partly resulted from the deamination of methylcytosine [71]. However, over-representation of these transition types was similar to what was observed in the *Arabidopsis* nuclear genome, indicating that *nupts* evolved in a nuclear-specific manner. Similarly, the fate of potential protein-coding sequences and non-coding sequences presented within *nupts* was similar and evolved both neutrally, in accordance with the non-functionality of almost all *nupts*.

### **3.3. Functional replacement of hundreds to thousands chloroplast proteins in the nucleus**

Following endosymbiosis, the symbiont to organelle transition involved many steps. This includes the loss of the bacterial cell wall, the acquisition of a protein machinery that transfers nuclear-encoded proteins from the cytosol to the chloroplast (also known as the TIC and TOC complexes [75, 76]), and finally, the functional relocation of most chloroplast genes to the nucleus. As detailed below, a chloroplast gene may be replaced either only after its functional transfer to the nucleus, or directly substituted by a gene of a mitochondrial or eukaryotic origin.

Since the endosymbiosis event, thousands of genes have relocated within the nuclear genome. Indeed, cyanobacterial genomes encode a minimum of 2000 proteins, whereas current plant plastomes encode only 80–200 proteins, although 800 to more than 2000 proteins have been found in some algae and plant chloroplasts [77], respectively. Apart from some genes that presented redundant functions in both chloroplast and nuclear genomes, most chloroplast genes have been functionally relocated to the nucleus with their proteins targeted back to the organelle. Thus, the spectrum of proteins required for function and biogenesis of the cytoplasmic organelle did not greatly evolve since its creation.

### 3.3.1. Functional transfer and relocation of a chloroplast gene to the nucleus

The current plastome of most plants encodes a maximum of 200 proteins [78] whereas more than 2000 proteins in the chloroplast, suggesting the functional gene transfer and relocation of most chloroplast genes to the nucleus. As chloroplast genes are of prokaryote origin, they are not readily functional in the nuclear genome. To function in this novel environment, a chloroplast gene has to acquire or hijack nuclear gene regulatory elements (eukaryote promoter and terminator), as well as a transit peptide to target the protein back to the chloroplast [60, 79]. However, the acquisition of all these nuclear elements does not have to take place right after the transfer of the chloroplast gene to the nucleus, as they can retain their open reading frames for several million years [71]. In addition, some chloroplast genes can be relatively easily functional as a few chloroplast promoters (i.e., *psbA* and *16S rrn* [80, 81]) were shown to be functional in the nucleus. Similarly, some transit peptides may be of cyanobacterial origin [82] and the AT-richness of 3'UTR chloroplast gene regions may mimic a polyadenylation signal.

To date, the number of chloroplast-encoded proteins (about 80) is relatively well conserved among flowering plants. However, a few chloroplast genes have been independently lost in various plant lineages [19], allowing to understand how they became functional. Such chloroplast gene losses were most particularly observed in the Fabaceae, for which the plastome has been extensively reorganized and contains localized accelerated mutation rates [19]. Some of these genes, such as *rpl22* [83] and *accD* [19], have been shown experimentally to have been functionally transferred to the nucleus. Similarly, recent functional transfers of chloroplast genes, such as *rpl32* [84] or *infA* [85], have been demonstrated. In addition, the functional relocation of *infA* and *accD* genes to the nucleus occurred several times independently [19, 85, 86]. Indeed, after the functional transfer of a chloroplast gene to the nucleus, two genes present in two different cellular compartments will encode for the same chloroplast protein. On one hand, the retention of the chloroplast copy is favored as the chloroplast genome evolves slower than the nuclear genome. On the other hand, even if the nuclear copy loses its functionality, the whole process can be repeated again.

### 3.3.2. Functional replacement of a chloroplast gene by a gene of mitochondrial (prokaryotic) or eukaryote origin

The functional replacement of a chloroplast gene does not necessarily necessitate its functional transfer from the chloroplast to the nucleus. In the case of the chloroplast RPS16 protein, the chloroplast *rps16* gene has been replaced by a nuclear *rps16* gene of mitochondrial origin [22, 83]. This nuclear *rps16* of mitochondrial origin had been functionally transferred to the nucleus soon after the formation of the mitochondria [22], and it acquired a dual target peptide to transfer the RPS16 protein to both chloroplasts and mitochondria [20]. Such functional replacement is not so surprising and many more similar functional transfers may have occurred as the prokaryote ancestors of chloroplast and mitochondria may encode similar proteins.

Another evolutionary mechanism enabling the functional replacement of a chloroplast gene may occur *via* the acquisition of a chloroplast transit peptide by a eukaryotic gene presenting the same function. Such event was observed for the chloroplast *accD* and the eukaryote *aac* genes,

which both encode an acetyl-CoA carboxylase. In *Arabidopsis*, the nuclear *acc* gene has been duplicated in tandem, and one copy has acquired a chloroplast targeted protein and thus also encodes a chloroplast ACCD protein [87].

The continuous deluge of organellar DNA to the nucleus has facilitated the functional transfer of almost all chloroplast genes to the nucleus, reducing extensively the plastome size. Additionally, this organellar DNA was not only used to replace organellar genes but also enabled diversifying the plant nuclear gene content [77].

### **3.4. Importance of chloroplast DNA transferred to the nucleus in diversifying the plant nuclear gene content**

Chloroplast gene sequences transferred to the nucleus may present different fates. As presented in the two previous sections: (i) they may remain non-functional, decay, and ultimately be lost; (ii) they may acquire all the necessary elements to conserve the same function and have the protein targeted back to the chloroplast; or (iii) they may acquire new subcellular locations and functions. As mentioned earlier, Martin *et al.* [77] extrapolated that about 18% of *Arabidopsis thaliana* genes were acquired from the cyanobacterial ancestor of plastids and that more than half of these cyanobacterially derived proteins were not targeted to the chloroplasts, suggesting either that they conserved their function but in another cellular localization or that they acquire a new function. These proteins are involved in many different functional categories that are not typically cyanobacterial, such as disease resistance and intracellular protein routing, indicating that they served as a rich source of genetic raw material and led to functional novelties. Similar analyses were performed in the glaucophyte *Cyanophora paradoxa* [88–90] and the green alga model *Chlamydomonas reinhardtii* [91]. Compared to what was observed in the flowering plant *A. thaliana*, only 6–7% of genes were inferred to be of cyanobacterial origin. Of these genes of cyanobacterial origin, 90% were inferred to be targeted back to the chloroplast in *C. paradoxa* [88], indicating that the impact of *nupts* on creating novel genes (new function or new cellular location) varies between plant lineages. We can speculate that many factors could explain these differences, such as the nuclear genome size and its structural evolutionary dynamics. Another major evolutionary impact of *nupt* on plant proteome evolution was determined by observing that *nupts* can generate novel nuclear exons encoding proteins with a different function to the preexisting organellar coding sequence. Additionally, Noutsos *et al.* [92] found that the  $K_a/K_s$  ratios (non-synonymous substitutions/synonymous substitutions) were higher than 1, reflecting a non-neutral evolution of *nupts* and their involvement into innovative functions.

## **4. Cytonuclear interactions, coadaptation processes, and incompatibilities**

The conversion of the cyanobacterial endosymbiont into the chloroplast partly results from the gradual transfer of hundreds to thousands of endosymbiont genes to the nuclear host. Across all lineages, more than 90% of the plant chloroplast proteins are now encoded in the nucleus. Within the few chloroplast-encoded proteins, about 40% of them are involved in

chloroplast protein complexes that are made up of proteins encoded in both the chloroplast and the nucleus. These complexes exhibit important functions that are vital for the plant, such as photosystems I and II. One can only wonder how the stoichiometry between those two compartments is maintained. Indeed, one cell might contain hundreds to thousands of chloroplast copies compared to only one copy in the nucleus. Furthermore, chloroplast inheritance is often maternal, whereas nuclear bi-parental inheritance occurs in angiosperms during sexual reproduction. Therefore, coevolving interactions between cytoplasmic and nuclear genomes have been necessary and have resulted in significant coadaptation processes. When these fine-tuned coevolutionary interactions are disrupted, after intra-interspecific hybridization and/or genome doubling, for instance, incompatibilities and deleterious phenotypes can be observed. These evolutionary processes will be discussed in the light of previous work on synthetic and natural hybrids, as well as in polyploid species.

#### 4.1. Hybridization and cytonuclear intergenomic complexes

Several evolutionary scenarios can explain coadaptation between chloroplast and nuclear genomes after intraspecific hybridization. First, cytoplasmic genomes lack sexual reproduction and are more susceptible to fix and accumulate deleterious mutations by genetic drift [93]. Only positive selection for compensatory nuclear alleles will allow for regaining of optimal organelle function [94]. This mechanism of *compensatory coadaptation* has been shown in several plant species with photosynthesis dysfunction (reviewed in [95]). One of the best examples with detailed genetic studies comes from the genus *Oenothera* [96], where three basic haploid nuclear genomes can be associated with five different chloroplast haplotypes. Of the 30 possible combinations, only 12 produce a green viable phenotype, whereas the 18 remaining associations lead to various degrees of cytonuclear incompatibilities, from reduced phenotypic capacity to embryo lethality [97]. Subsection *Oenothera* has apparently separated into three distinct evolutionary lineages (represented by the three basic haploid genomes A, B, and C) that have coevolved with chloroplast haplotypes I, III, and V, respectively [97]. Recent molecular work suggests that the radiation within this subsection started approximately 1 million years ago [98]. Thus, these results suggest that, in *Oenothera*, cytonuclear incompatibilities and associated coadaptation mechanisms have rapidly lead to strong post-zygotic barriers after only 1 million years apart [99].

Second, some mutations in the organelles could also be adaptive in specific environments and fixed in the population by natural selection. Subsequently, coadaptation process may favor specific nuclear variants to preserve intergenomic interactions. This mechanism is called *adaptive divergence*. However, experimental studies in the genus *Helianthus* are giving some hints of the effects of extrinsic selection on cytonuclear interactions. Exchange of the common sunflower cytoplasm with closely related species' organelles leads, just as in *Oenothera*, to deleterious phenotypes (from altered biomass to reduced seed weight and pollen unviability), suggesting, again, a role of cytonuclear incompatibilities in establishing reproductive barriers between populations [100]. Additional study demonstrated the contrasting adaptive potential of two cytoplasmic genomes in two alternative ecological environments. Sambatti et al. [101] have performed reciprocal transplant experiments of *H. annuus* and *H. petiolaris* and all possible backcross combinations of nuclear and cytoplasmic

genomes into two contrasted ecological environments. The authors elegantly showed that each cytoplasm of *H. annuus* and *H. petiolaris* exhibits higher fitness in mesic and xeric habitats, respectively, and is therefore differentially adapted to these two contrasting habitats. More recently, authors have benefited from the model system *A. thaliana* to investigate the contribution of cytonuclear interactions into plant fitness variation [102]. In this study, a field experiment has been set with 56 different cytoplasmic lines (based on eight natural accessions of *A. thaliana*) combining the nuclear genome of one parent with the organelle genomes of another. Using 28 adaptive phenotypic traits (such as germination, phenology, and fecundity), authors showed that a large proportion of those traits are affected by inter-specific cytonuclear interactions. However, the genetic factors and molecular interactions underlying such phenomenon are still to be elucidated.

As mentioned above, the examples for intergenomic coadaptation and incompatibilities are scarce, and we are still very far from unraveling the molecular processes underlying such interactions. Applications of genome-wide studies in association with high-throughput sequencing would greatly improve our understanding of cytonuclear coevolution.

#### **4.2. Effects of whole genome doubling and interspecific hybridization on cytonuclear complex stability**

As shown above, cytonuclear interactions are extremely fine-tuned coevolved molecular processes that are still largely understudied. However, in recent years, efforts have been made, especially in neo-polyploid plant species (natural and resynthesized) to better apprehend the consequences of whole genome duplication (WGD) and interspecific hybridization on cytonuclear interactions and stability. In this last section, we will review our knowledge on such systems and elaborate on the many future issues to address.

Although completely overlooked, it is astonishing to envision the numerous and drastic consequences of a WGD event on copy number variation and stoichiometry on those cytonuclear complexes. Impacts of WGD on genomic structure and functional changes have been extensively studied in a large variety of plant systems. Genome redundancy can lead to changes in epigenetic patterns (including transposable element dynamics), altered gene expression (changes in global gene expression but also possible biased contribution of redundant copies), and fractionation processes (gene loss, homologous and non-homologous exchanges). However, to date, very few studies have investigated how the duplication of nuclear genes would affect the assembly dynamics of the multi-subunit cytonuclear complexes [103]. Different hypotheses predict the fate of nuclear and cytoplasmic genes implicated in cytonuclear complexes. They are based on the prediction that selection will favor compensatory mechanisms to maintain coordinated expression between cytoplasmic and nuclear genes leading *in fine* to a functional complex. Immediate impacts of WGD could therefore lead to downregulation of nuclear genes and/or upregulation of cytoplasmic genes. Additionally, another path to achieve the same outcome would be for the cell to enhance organelle biogenesis and produce a larger number of chloroplasts. This has been shown in cotton and alfalfa polyploids, which exhibit larger chloroplast size and higher chloroplast number per cell relative to their diploid progenitors [104, 105]. For instance, chloroplast number in guard cells

is increased by 25, 72, and 102% in triploid, tetraploid, and hexaploid cottons, respectively, compared to diploids [105]. Consequently, it is hypothesized that larger chloroplasts could carry more genome copies per organelle. In maize, only chloroplast number per cell (and not chloroplast size) is accentuated with ploidy [106]. However, it seems that chloroplast proliferation might be more correlated to cell size than nuclear ploidy [107]. Indeed, a positive relationship exists between nuclear genome size and cell size [108], but the direct impact of WGD and presence of redundant genomes have yet to be elucidated.

Only a handful of studies have looked at the consequences of WGD on a longer time scale, in that case, occurrences of subfunctionalization and pseudogenization of duplicated copies are to be expected. Coate et al. [109] stated that there might be a considerable influence of cytonuclear complex sensitivity to gene dosage imbalance and thus the need to return to single copy status or stay in duplicates. More specifically, Coate et al. [109] demonstrated that in *Glycine max*, *Medicago truncatula*, and *A. thaliana* photosystem gene families are preferentially retained as duplicates after WGD. This trend is likely explained by the high dosage sensitivity of these cytonuclear complexes. The authors hypothesized that if one of the duplicated gene copies implicated in the same cytonuclear complex is lost, it will cause gene dosage imbalance between genes, and the complex will not function properly. On the contrary, other complexes are apparently less affected by gene dosage imbalance and tolerate different copy numbers among genes (of the same complex).

All of these processes could be enhanced through allopolyploidy, where divergent parental species first hybridized before genome doubling. In that case, the nuclear genome is redundant and a mixture of two, more or less, divergent parental genomes, whereas the organelles have (usually) a uniparental origin. Therefore, as chloroplast inheritance is usually maternal, selection should favor maintenance of maternal nuclear copies over the paternally inherited homoeolog as to preserve pre-existing coadaptive cytonuclear interactions. In allopolyploids, different scenarios leading to pseudogenization of paternal copies can be envisioned and were tested in a limited set of genes and species. The first scenario involves downregulation and relaxed selection of the paternally inherited homoeolog. An alternative scenario involves preferential gene conversion to the maternal homoeolog resulting in the loss of the paternal-like copy. It is important to note that both scenarios are not exclusive but could be part of a dynamic and gradual process, with first overexpression of the maternal copies leading to paternal homoeolog pseudogenization and maternally biased gene conversion. These hypotheses have only been tested in the Rubisco nuclear-encoded gene *rbcS* in various allopolyploids. In cotton, an ancient allopolyploid formed 1–2 MYA (progenitors diverged 5–10 MYA), and it has been shown in five different allopolyploid species that putative events of gene conversion occur between subgenomes but not in synthetic hybrids [110]. Interestingly, maternal homoeologs are preferentially expressed in wild and cultivated allopolyploids as well as in the synthetic F1 hybrid (whereas no such bias is observed between the diploid progenitors) [110]. These patterns have been shown also in other polyploid models. Following the same methods, concerted evolution is reported between homoeologous genomes of *Arabidopsis suecica*, *Arachis hypogaea*, and *N. tabacum* [111]. Additionally, there is preferential



occurrence of maternal to paternal gene conversion in signaling and regulatory domains of the *rbcS* gene copies. In those polyploids, preferential expression of paternal homoeologs carrying the maternal-like gene conversions has also been described [111]. In contrast, the allotetraploid *Brassica napus* showed no sign of homoeologous exchanges or bias expression probably because of either recent (compared to the other models) divergence time between diploid parental species (only 4 MYA). In the same way, resynthesized reciprocal hybrids and allotetraploids formed between *Oryza sativa indica* and *japonica* (that diverged around 9000 yr. ago) did not exhibit biased expression of *rbcS* alleles or homoeologs and also no biased gene conversion toward maternal gene copies [112]. In *Tragopogon miscellus*, a very recent neo-allopolyploid formed only 80 years ago, homoeolog gene loss and biased expression were limited, occurring only in 12 and 16% of individuals coming from two naturally and repeatedly formed polyploid populations [113]. However, the bias was mainly toward maintaining the maternal nuclear copy of *rbcS* (in 7 of 10 cases of homoeolog loss). Therefore, although parental genomes of the neotetraploid *T. miscellus* polyploid are quite divergent [114], very little evidence for functionalization and homogenization of duplicated copies is visible in the polyploids. This might be due to the recent formation of such polyploids (less than 100 years ago) and the lack of time for such events to take place. Thus, in the cases of allopolyploid formation, divergence between parental species and age of polyploids seems to be important factors driving cytonuclear coevolution processes.

These few studies already highlight the complexity of the different model systems that can be highly influenced by various evolutionary processes such as pre-existing coadaptive mechanisms, natural selection, and divergence between parental individuals (different populations to different species). As all Angiosperms have experienced at least one round of genome duplication and most of them multiple WGDs (Triticum and Brassica), paleopolyploid species are perfect candidates to elucidate the long-term impact of diploidization and biased genome fractionation on rates of asymmetric gene loss and pseudogenization. Additionally, it seems essential to integrate plant families that have contrasted rate of chloroplastic evolution (such as in Geraniaceae, Campanulaceae, and Fabaceae) and paternally inherited chloroplast genomes (such as in Actinidia, Medicago, and most Conifers). Finally, life history features such as reproductive strategy (perennial vs. annual), mating system (selfer vs. outcrosser), population level dynamics, and effective population size will also impact fixation rate of mutations.

## Acknowledgements

We would like to thank the European Union Seventh Framework Program (FP7-CIG-2013-2017; Grant No. 333709 to Mathieu Rousseau-Gueutin) and an Agreenskills Plus fellowship to Julie Ferreira de Carvalho. We would also like to thank Dr. Christina Richards (Department of Integrative Biology, University of South Florida) for carefully and critical reading of the manuscript.

## Conflict of interest

No conflict of interest.

## Author details

Mathieu Rousseau-Gueutin<sup>1\*</sup>, Jean Keller<sup>2</sup>, Julie Ferreira de Carvalho<sup>1</sup>, Abdelkader Aïnouche<sup>2</sup> and Guillaume Martin<sup>3,4</sup>

\*Address all correspondence to: mathieu.rousseau-gueutin@inra.fr

1 IGEPP, INRA, Agrocampus Ouest, University of Rennes 1, Le Rheu Cedex, France

2 UMR CNRS 6553 Ecobio, OSUR University of Rennes 1, Rennes, France

3 CIRAD, UMR AGAP, F-34398, Montpellier, France

4 AGAP, University Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

## References

- [1] Mereschkowsky C. Über Natur and Ursprung der Chromatophoren im Pflanzenreiche. *Biologisches Centralblatt*. 1905;**25**:593-604
- [2] Parfrey LW, Lahr DJ, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*. 2011;**108**(33):13624-13629
- [3] Shih PM, Matzke NJ. Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proceedings of the National Academy of Sciences*. 2013;**110**(30):12355-12360
- [4] Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, Allen JF, et al. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Molecular Biology and Evolution*. 2008;**25**(4):748-761
- [5] Falcón LI, Magallón S, Castillo A. Dating the cyanobacterial ancestor of the chloroplast. *The ISME Journal*. 2010;**4**(6):777-783
- [6] McFadden GI, van Dooren GG. Evolution: Red algal genome affirms a common origin of all plastids. *Current Biology*. 2004;**14**(13):R514-R516
- [7] Keeling PJ. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology*. 2013;**64**(1):583-607
- [8] Zimorski V, Ku C, Martin WF, Gould SB. Endosymbiotic theory for organelle origins. *Current Opinion in Microbiology*. 2014;**22**:38-48

- [9] Andrès C, Agne B, Kessler F. The TOC complex: Preprotein gateway to the chloroplast. *Biochimica et Biophysica Acta*. 2010;**1803**(6):715-723
- [10] Nakai M. The TIC complex uncovered: The alternative view on the molecular mechanism of protein translocation across the inner envelope membrane of chloroplasts. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 2015;**1847**(9):957-967
- [11] Jarvis P, López-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews. Molecular Cell Biology*. 2013;**14**(12):787-802
- [12] Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Reviews. Genetics*. 2004;**5**(2):123-135
- [13] Raven JA, Allen JF. Genomics and chloroplast evolution: What did cyanobacteria do for plants? *Genome Biology*. 2003;**4**(3):209
- [14] Kaneko Y, Keegstra K. Plastid biogenesis in embryonic pea leaf cells during early germination. *Protoplasma*. 1996;**195**(1):59-67
- [15] Lang BF, Nedelcu AM. Plastid genomes of algae. In: Bock R, Knoop V, éditeurs. *Genomics of Chloroplasts and Mitochondria* [Internet]. Dordrecht: Springer Netherlands; 2012. p.59-87. Disponible sur: [http://www.springerlink.com/index/10.1007/978-94-007-2920-9\\_3](http://www.springerlink.com/index/10.1007/978-94-007-2920-9_3)
- [16] Shevelev EL, Bryant DA, Löffelhardt W, Bohnert HJ. Ribonuclease-P RNA gene of the plastid chromosome from *Cyanophora paradoxa*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*. 1995;**2**(5):231-234
- [17] Turmel M, Gagnon M-C, O'Kelly CJ, Otis C, 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
- [18] Andersen ES, Rosenblad MA, Larsen N, Westergaard JC, Burks J, Wower J, et al. The tmRDB and SRPDB resources. *Nucleic Acids Research*. 2006;**34**(90001):D163-D168
- [19] Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, et al. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*. 2010;**20**(12):1700-1710
- [20] Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S-I, Tsutsumi N, et al. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. *Molecular Biology and Evolution*. 2008;**25**(8):1566-1575
- [21] Roy S, Ueda M, Kadowaki K, Tsutsumi N. Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus *Arabidopsis* and closely related species. *Genes & Genetic Systems*. 2010;**85**(5):319-326
- [22] Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, et al. 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. 2017;**24**(4):343-358
- [23] Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of pelargonium x hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Molecular Biology and Evolution. 2006;**23**(11):2175-2190
- [24] Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM, et al. The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. Current Biology. 2017;**27**(11):1677-1684.e4
- [25] Palmer JD, Nugent JM, Herbon LA. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proceedings of the National Academy of Sciences of the United States of America. 1987;**84**(3):769-773
- [26] Knox EB, Palmer JD. The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): Expansion of the inverted repeat in an ancestor of the Campanulales. Plant Systematics and Evolution. 1999;**214**(1-4):49-64
- [27] Martin G, Baurens F-C, Cardi C, Aury J-M, D'Hont A. The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): Insight into plastid monocotyledon evolution. PLoS One. 2013;**8**(6):e67350
- [28] Cai Z, Guisinger M, Kim H-G, Ruck E, Blazier JC, McMurtry V, et al. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. Journal of Molecular Evolution. 2008;**67**(6):696-704
- [29] Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Scientific Reports. 2015;**5**:16958
- [30] Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the Angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. Molecular Biology and Evolution. 2011;**28**(1):583-600
- [31] Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proceedings of the National Academy of Sciences of the United States of America. 2008;**105**(47):18424-18429
- [32] Funk HT, Berg S, Krupinska K, Maier UG, Krause K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biology. 2007;**7**(1):45
- [33] Lavin M, Doyle JJ, Palmer JD. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. Evolution. 1990;**44**(2):390-402

- [34] Wojciechowski MF, Sanderson MJ, Steele KP, Liston A. Molecular phylogeny of the « temperate herbaceous tribes » of papilionoid legumes: A supertree approach. In: Herendeen PS, Bruneau A, éditeurs. *Advances in Legume Systematics*. Kew/Surrey: Royal Botanic Gardens, Kew; 2000. p. 277-298
- [35] Turmel M, Otis C, Lemieux C. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Scientific Reports*. 2017;7(1) Disponible sur: <http://www.nature.com/articles/s41598-017-01144-1>
- [36] Strauss SH, Palmer JD, Howe GT, Doerksen AH. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proceedings of the National Academy of Sciences*. 1988;85(11):3898-3902
- [37] Yi X, Gao L, Wang B, Su Y-J, Wang T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biology and Evolution*. 2013;5(4):688-698
- [38] Doyle JJ, Doyle JL, Ballenger JA, Palmer JD. The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Molecular Phylogenetics and Evolution*. 1996;5(2):429-438
- [39] Martin GE, Rousseau-Gueutin M, Cordonnier S, Lima O, Michon-Coudouel S, Naquin D, et al. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Annals of Botany*. 2014;113(7):1197-1210
- [40] Choi I-S, Choi B-H. The distinct plastid genome structure of *Maackia fauriei* (Fabaceae: Papilionoideae) and its systematic implications for genistoids and tribe Sophoreae. Vendramin GG, éditeur. *PLoS One*. 2017;12(4):e0173766
- [41] Kim Y, Cullis C. A novel inversion in the chloroplast genome of marama (*Tylosema esculentum*). *Journal of Experimental Botany*. 2017;68(8):2065-2072
- [42] Weng M-L, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution*. 2014;31(3):645-659
- [43] Cosner ME, Raubeson LA, Jansen RK. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology*. 2004;4(1):27
- [44] Haberle RC, Fourcade HM, Boore JL, Jansen RK. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution*. 2008;66(4):350-361
- [45] Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*. 1987;84(24):9054-9058

- [46] Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, et al. Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PLoS One*. 2014;**9**(3):e85761
- [47] Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*. 2010;**327**(5961):92-94
- [48] Doyle JJ, Doyle JL, Palmer JD. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Systematic Botany*. 1995;**20**(3):272
- [49] Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. *Genomics of Chloroplasts and Mitochondria* [Internet]. Dordrecht: Springer Netherlands; 2012. p. 103-126. Disponible sur: [http://www.springerlink.com/index/10.1007/978-94-007-2920-9\\_5](http://www.springerlink.com/index/10.1007/978-94-007-2920-9_5)
- [50] Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, et al. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *The EMBO Journal*. 1986;**5**(9):2043-2049
- [51] Kawashima N, Wildman SG. Studies on fraction I protein IV. Mode of inheritance of primary structure in relation to whether chloroplast or nuclear DNA contains the code for a chloroplast protein. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. 1972;**262**(1):42-49
- [52] Timmis JN, Scott NS. Sequence homology between spinach nuclear and chloroplast genomes. *Nature*. 1983;**305**(5929):65
- [53] Scott NS, Timmis JN. Homologies between nuclear and plastid DNA in spinach. *Theoretical and Applied Genetics*. 1984;**67**(2-3):279-288
- [54] Ayliffe MA, Timmis JN. Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. *Theoretical and Applied Genetics*. 1992;**85**(2-3):229-238
- [55] Huang CY, Ayliffe MA, Timmis JN. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature*. 2003;**422**(6927):72-76
- [56] Stegemann S, Hartmann S, Ruf S, Bock R. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;**100**(15):8828-8833
- [57] Wang D, Lloyd AH, Timmis JN. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;**109**(7):2444-2448
- [58] Ruf, Braune, Endries, Hasse, Stegemann, Bock. *Plastid transmission, gene transfer and the impact of the environment*. In Maynooth; 2010
- [59] Sheppard AE, Ayliffe MA, Blatch L, Day A, Delaney SK, Khairul-Fahmy N, et al. Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. *Plant Physiology*. 2008;**148**(1):328-336

- [60] Lloyd AH, Timmis JN. The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. *Molecular Biology and Evolution*. 2011;**28**(7):2019-2028
- [61] Wang D, Timmis JN. Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biology and Evolution*. 2013;**5**(6):1060-1064
- [62] Sheppard AE, Timmis JN. Instability of plastid DNA in the nuclear genome. *PLoS Genetics*. 2009;**5**(1):e1000323
- [63] Noutsos C, Richly E, Leister D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Research*. 2005;**15**(5):616-628
- [64] Richly E, Leister D. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Molecular Biology and Evolution*. 2004;**21**(10):1972-1980
- [65] Yoshida T, Furihata HY, Kawabe A. Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. *DNA Research*. 2014;**21**(2):127-140
- [66] Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Molecular Biology*. 2003;**52**(5):923-934
- [67] Michalovova M, Vyskot B, Kejnovsky E. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: Size, relative age and chromosomal localization. *Heredity*. 2013;**111**(4):314-320
- [68] Matsuo M, Ito Y, Yamauchi R, Obokata J. The Rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell Online*. 2005;**17**(3):665-675
- [69] Guo X, Ruan S, Hu W, Cai D, Fan L. Chloroplast DNA insertions into the nuclear genome of rice: The genes, sites and ages of insertion involved. *Functional & Integrative Genomics*. 2008;**8**(2):101-108
- [70] Roark LM, Hui AY, Donnelly L, Birchler JA, Newton KJ. Recent and frequent insertions of chloroplast DNA into maize nuclear chromosomes. *Cytogenetic and Genome Research*. 2010;**129**(1-3):17-23
- [71] Rousseau-Gueutin M, Ayliffe MA, Timmis JN. Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. *Plant Physiology*. 2011;**157**(4):2181-2193
- [72] Rousseau-Gueutin M, Ayliffe MA, Timmis JN. Plastid DNA in the nucleus. *Plant Signaling & Behavior*. 2012;**7**(2):269-272
- [73] Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: Mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genetics*. 2010;**6**(2). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2820518/>
- [74] Smith DR, Crosby K, Lee RW. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biology and Evolution*. 2011;**3**(0):365-371

- [75] Jarvis P, Soll J. Toc, tic, and chloroplast protein import. *Biochimica et Biophysica Acta*. 2001;**1541**(1-2):64-79
- [76] Gross J, Bhattacharya D. Revaluating the evolution of the Toc and tic protein translocators. *Trends in Plant Science*. 2009;**14**(1):13-20
- [77] Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;**99**(19):12246-12251
- [78] Lee J, Cho CH, Park SI, Choi JW, Song HS, West JA, et al. Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology*. 2016;**14**(1). Disponible sur: DOI: <http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-016-0299-5>
- [79] Stegemann S, Bock R. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell*. 2006;**18**(11):2869-2878
- [80] Cornelissen M, Vandewiele M. Nuclear transcriptional activity of the tobacco plastid psbA promoter. *Nucleic Acids Research*. 1989;**17**(1):19-29
- [81] Wang D, Qu Z, Adelson DL, Zhu J-K, Timmis JN. Transcription of nuclear Organellar DNA in a model plant system. *Genome Biology and Evolution*. 2014;**6**(6):1327-1334
- [82] Bruce BD. Chloroplast transit peptides: Structure, function and evolution. *Trends in Cell Biology*. 2000;**10**(10):440-447
- [83] Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *The EMBO Journal*. 1991;**10**(10):3073-3078
- [84] Cusack BP, Wolfe KH. When gene marriages don't work out: Divorce by subfunctionalization. *Trends in Genetics*. 2007;**23**(6):270-272
- [85] Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*. 2001;**13**(3):645-658
- [86] Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. Potential functional replacement of the Plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology*. 2013;**161**(4):1918-1929
- [87] Babiychuk E, Vandepoele K, Wissing J, Garcia-Diaz M, De Rycke R, Akbari H, et al. Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;**108**(16):6674-6679
- [88] Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Current Biology: CB*. 2006;**16**(23):2320-2325



- [89] Reyes-Prieto A, Weber APM, Bhattacharya D. The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics*. 2007;**41**(1):147-168
- [90] Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, et al. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science*. 2012;**335**(6070):843-847
- [91] Moustafa A, Reyes-Prieto A, Bhattacharya D. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS One*. 2008;**3**(5):e2205
- [92] Noutsos C, Kleine T, Armbruster U, DalCorso G, Leister D. Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. *Trends in Genetics*. 2007;**23**(12):597-601
- [93] Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren JH. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (hymenoptera: Pteromalidae). *Molecular Biology and Evolution*. 2008;**25**(10):2167-2180
- [94] Rand DM, Haney RA, Fry AJ. Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution*. 2004;**19**(12):645-653
- [95] Burton RS, Pereira RJ, Barreto FS. Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*. 2013;**44**(1):281-302
- [96] Stubbe W. *Oenothera*—An ideal system for studying the interactions of genome and plastome. *Plant Molecular Biology Reporter*. 1989;**7**(4):245-257
- [97] Greiner S, Rauwolf U, Meurer J, Herrmann RG. The role of plastids in plant speciation. *Molecular Ecology*. 2011;**20**(4):671-691
- [98] Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic Acids Research*. 2008;**36**(7):2366-2378
- [99] Dietrich W, Wagner WL, Raven PH. Systematics of *Oenothera* Sect. *Oenothera* Subsect. *Oenothera* (Onagraceae). 1997; Disponible sur: <http://repository.si.edu/handle/10088/9721>
- [100] Levin DA. The cytoplasmic factor in plant speciation. *Systematic Botany*. 2003;**28**(1):5-11
- [101] Sambatti JBM, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecology Letters*. 2008;**11**(10):1082-1091
- [102] Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, et al. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;**113**(13):3687-3692
- [103] Birchler JA, Veitia RA. Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences*. 2012;**109**(37):14746-14753

- [104] Bingham ET. Stomatal chloroplasts in alfalfa at four Ploidy levels. *Crop Science*. 1968;**8**(4):509-510
- [105] Krishnaswami R, Andal R. Stomatal chloroplast number in diploids and polyploids of *Gossypium*. *Proceedings of the Indian Academy of Science, Section B*. 1978;**87**(5):109-112
- [106] Rhoades MM, Dempsey E. Induction of chromosome doubling at meiosis by the elongate gene in maize. *Genetics*. 1966;**54**(2):505-522
- [107] Kawade K, Horiguchi G, Ishikawa N, Hirai MY, Tsukaya H. Promotion of chloroplast proliferation upon enhanced post-mitotic cell expansion in leaves. *BMC Plant Biology*. 2013;**13**:143
- [108] Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *The New Phytologist*. 2008;**179**(4):975-986
- [109] Coate JE, Schlueter JA, Whaley AM, Doyle JJ. Comparative evolution of photosynthetic genes in response to Polyploid and Nonpolyploid duplication. *Plant Physiology*. 2011;**155**(4):2081-2095
- [110] Gong L, Salmon A, Yoo M-J, Grupp KK, Wang Z, Paterson AH, et al. The Cytonuclear dimension of allopolyploid evolution: An example from cotton using Rubisco. *Molecular Biology and Evolution*. 2012;**29**(10):3023-3036
- [111] Gong L, Olson M, Wendel JF. Cytonuclear evolution of Rubisco in four allopolyploid lineages. *Molecular Biology and Evolution*. 2014;**31**(10):2624-2636
- [112] Wang X, Dong Q, Li X, Yuliang A, Yu Y, Li N, et al. Cytonuclear Variation of Rubisco in Synthesized Rice Hybrids and Allotetraploids. *Plant Genome* [Internet]. 2017. Disponible sur: <https://dl.sciencesocieties.org/publications/tpg/articles/0/0/plantgenome2017.05.0041>
- [113] Sehrish T, Symonds VV, Soltis DE, Soltis PS, Tate JA. Cytonuclear coordination is not immediate upon allopolyploid formation in *Tragopogon miscellus* (Asteraceae) allopolyploids. *PLoS One*. 2015;**10**(12):e0144339
- [114] Mavrodiev EV, Krahulec F, Soltis DE, Soltis PS. A cryptic taxon rather than a hybrid species of *Tragopogon* (Asteraceae) from the Czech Republic. *Kew Bulletin*. 2013;**68**(1):133-141