



# The chloroplast genome: a review

Jędrzej Dobrogojski<sup>2</sup> · Małgorzata Adamiec<sup>1</sup> · Robert Luciński<sup>1</sup>

Received: 17 September 2019 / Revised: 26 March 2020 / Accepted: 11 May 2020 / Published online: 18 May 2020  
© The Author(s) 2020

## Abstract

Chloroplasts are the metabolically active, semi-autonomous organelles found in plants, algae and cyanobacteria. Their main function is to carry out the photosynthesis process involving a conversion of light energy into the energy of chemical bonds used for the synthesis of organic compounds. The Chloroplasts' proteome consists of several thousand proteins that, besides photosynthesis, participate in the biosynthesis of fatty acids, amino acids, hormones, vitamins, nucleotides and secondary metabolites. Most of the chloroplast proteins are nuclear-encoded. During the course of evolution, many genes of the ancestral chloroplasts have been transferred from the chloroplast genome into the cell nucleus. However, these proteins which are essential for the photosynthesis have been retained in the chloroplast genome. This review aims to provide a relatively comprehensive summary of the knowledge in the field of the chloroplast genome arrangement and the chloroplast genes expression process based on a widely used model in plant genetic research, namely *Arabidopsis thaliana*.

**Keywords** *Arabidopsis thaliana* · Chloroplast · Chloroplast genome · Gene expression

## Abbreviations

|       |   |
|-------|---|
| cpDNA | Chloroplast genome                                |
| IRs   | Inverted repeats                                  |
| LSC   | Long single copy section                          |
| mtDNA | Mitochondrial genome                              |
| NCBI  | National Center for Biotechnology Information     |
| NEP   | T3/T7 phage-type RNA polymerases                  |
| PAPs  | Polymerase associated proteins                    |
| PEP   | Bacterial-type multi-subunit RNA polymerase       |
| SIG   | Sigma factors, transcriptional initiation factors |
| SSC   | Short single copy section                         |
| TAIR  | The Arabidopsis Information Resource              |

## Introduction

The first two plants in which the chloroplast genome (cpDNA) was successfully sequenced were tobacco (*Nicotiana tabacum*, Ohyama et al. 1986) and liverwort (*Marchantia polymorpha*, Shinozaki et al. 1986). Since then, the cpDNA of 3721 plant species including green algae and both terrestrial and aquatic plants have been described. They are available in the National Center for Biotechnology Information (NCBI) organelle genome database. Such a significant progress in the field of chloroplast genetics has been accomplished by the advancement of the high-throughput sequencing technologies over the last several years. A real breakthrough in the cpDNA study has come along with the fully sequenced genome of *Synechocystis* sp. PCC6803, a bacteria from the phylum of the cyanobacteria (Sato et al. 1999). It is one of the best known primitive photosynthetic organism widely used for research on photosynthesis, the assimilation of carbon and nitrogen and the evolution of plastids (Martin et al. 2002). *Synechocystis* quickly received status as a model microorganism due to its characteristic features, such as: a fully sequenced genome, both auto and heterotrophic mode of nutrition and a simple photosynthetic apparatus. Studies conducted on *Synechocystis* genome remarkably contributed to a better understanding of the plants cpDNA and also provided the key arguments confirming the endosymbiotic theory of the chloroplast origin (Timmis et al.

Communicated by P. Wojtaszek.

✉ Jędrzej Dobrogojski  
dobrogojski@gmail.com

<sup>1</sup> Department of Plant Physiology, Faculty of Biology, Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland

<sup>2</sup> Department of Biochemistry and Biotechnology, Faculty of Agronomy and Bioengineering, Poznań University of Life Sciences, Dojazd 11, 60-637 Poznań, Poland

2004). Moreover, these works were used to determine the degree of similarity between modern living plants and their ancestors. Thanks to this it was possible to trace the process of a mass gene transfer from the precursors of the present chloroplasts towards the cell nucleus. In the process of genes transfer not every gene ended in the host genome, some of them were lost. Eventually, the phylogenetic comparative studies that were conducted between *Arabidopsis thaliana* and various microorganisms' species have shown that over 4500 genes encoding plants proteins have cyanobacterial origin (Martin et al. 2002). At this point it is worth mentioning that the chloroplast proteome comprises approximately 3000 proteins from which the vast majority is encoded in the nucleus (Zoschke and Bock 2018).

The chloroplasts of the higher plants are the metabolic centers that simply sustain life on Earth by conducting the photosynthesis. In this process, plants use solar energy, water, mineral soils and carbon dioxide to synthesize organic compounds. The byproduct of this process is oxygen released into the atmosphere. Moreover, plastids are an essential environment for a number of biochemical processes such as the synthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins but also assimilation of sulfur and nitrogen (Daniell et al. 2016). Many of the metabolites synthesized in chloroplasts are necessary to maintain a proper communication between the different parts of the plant. These metabolites also play an important role in plant reaction to both the biotic and the abiotic stresses (Daniell et al. 2016).

Chloroplasts are the semi-autonomous organelles containing several thousand proteins that participate mainly in the photosynthesis process but also in a biosynthesis pathways of many relevant compounds, i.e. fatty acids (Brehelin et al. 2007; Vidi et al. 2006; Lippold et al. 2012) and jasmonate (Schaller et al. 2005). The majority of the chloroplast proteins are encoded by the nuclear genome. The translation of these genes takes place on eukaryotic 80 s ribosomes. Transport of these proteins from the cytoplasm to the light of the chloroplast is complex and it depends on both the transported protein's structure as well as the outer and inner envelope membranes protein translocons (Sun and Zerges 2015). Short amino acid sequence located at the N-terminal, called the transit peptide allows the recognition and transport of the protein of interest through the outer and inner envelope membranes via translocons called TOC (translocon on the outer chloroplast membrane) and TIC (translocon on the inner chloroplast membrane). Import of the nuclear-encoded chloroplast protein is modulated through the ubiquitin E3 ligase SP1. Depending on the developmental stage or environmental stimuli, SP1 promotes the degradation of TOC complexes (Ling et al. 2012). On the other hand, a more recent study indicates an essential role of the cpDNA encoded conserved hypothetical chloroplast YCF2 protein in

the nuclear-encoded proteins translocation across the inner membrane through the TIC (Kikuchi et al. 2018).

However, not all of the chloroplast proteins are encoded in the nuclear genome. Some of them are encoded by the chloroplast genome. The number of the genes encoded by the cpDNA varies between different plant species from 0 to 315 (NCBI). Importantly, proteins encoded by some of these genes play a key role in the processes occurring in chloroplasts, especially photosynthesis. A majority of the proteins encoded by the cpDNA are synthesized by their own internal expression system. Undoubtedly, one of the most important elements of the mentioned system are the bacterial-type 70 s ribosomes (Barkan and Goldschmidt-Clermont 2000). They reflect the evolutionary origin of the plastids, thus reinforcing the theory of endosymbiosis. In addition, the expression of proteins encoded by the cpDNA is regulated at various stages of this process, both by the transcription factors found in chloroplasts, as well as by those having a non-chloroplast origin (Sun and Zerges 2015). Moreover, the chloroplast proteome undergoes significant rearrangements while the plant is exposed to different types of abiotic stresses (Watson et al. 2018). Mostly, it concerns proteins involved in the photosynthesis. Goulas and coworkers observed a decrease in the abundance of ribulose-5-phosphate-3-epimerase and PsbO1 under cold stress while Wilson and coworkers showed that oxidative stress influence on the depletion of the chloroplast phosphatase SAL1 (Goulas et al. 2006; Wilson et al. 2009). Furthermore, heat may cause perturb in the import of the nuclear-encoded proteins (Heckathorn et al. 1998). Interestingly, most of the changes in the abundance of the chloroplast proteins caused by the abiotic stresses take place due to the post-transcriptional modification rather than by the alterations on the transcriptional level (Woodson and Chory 2008).

## Size and arrangement of the cpDNA

At the beginning of the plastom arrangement studies, cpDNA was considered to form only a circular molecule inside the living plant cells (Mower and Vickrey 2018). Recent analyses based mainly on microscopy have shown a multibranching linear structures of cpDNA in many species of angiosperms (Mower and Vickrey 2018). However, it is still uncertain which shape of the cpDNA is the most abundant. It differs between plant species and depends on many factors like cell development stage, tissue type and the experimental design (Mower and Vickrey 2018). Among plants the cpDNA size varies from 15,553 base pairs (bp) in *Asarum minus* to 521,168 bp in *Floydia terraestrans* (NCBI). Among terrestrial plants, the cpDNA is highly conserved, in terms of both structure and the content and order of the placement of particular genes (Shaw et al. 2007). In most

photosynthetic organisms, the cpDNA forms continuously packed structures called nucleoids (Morley et al. 2019a). On average, plant cells contain from 1000 to 1700 copies of chloroplast nucleoids. In turn, one chloroplast of the *A. thaliana* mesophilic cell can contain from 20 to 35 copies of the cpDNA (Zoschke et al. 2007). However, the cpDNA number is highly variable during the plant development (Oldenburg and Bendich 2015).

In the structure of the cpDNA, two identical fragments called inverted repeats (IR) can be distinguished. IRs are separated by one long single copy section (LSC) and one short single copy section (SSC) (Mower and Vickrey 2018). Among different plant species, the IR regions are highly conserved and have a length ranging from 20,000 to 25,000 bp (Morley et al. 2019a).

Similarly, to the IR regions the chloroplast intron sequences can also be considered as highly conserved (Daniell et al. 2016). Exceptions from this rule are a few plant species, such as barley (*Hordeum vulgare*), bamboo (*Bamboo* sp.), Cassava (*Manihot esculenta*) or chickpeas (*Cicer arietinum*) in which a loss of introns is observed in genes encoding some proteins. Like bacteria, many chloroplast genes are organized into operons whose transcription is carried out using one or more promoters (Börner et al. 2015).

## Gene transfer from chloroplasts to the cell nucleus

During the course of evolution, many genes of the ancestral chloroplasts have been transferred from the cpDNA into the cell nucleus. This process that also occurred in the second semi-autonomous organelles, which are mitochondria, is called endosymbiotic gene transfer (Ku et al. 2015). It has been estimated that around 18% of the genes from the *A. thaliana* nuclear genome have cyanobacterial origin (Ku et al. 2015). An example of the gene that has been "acquired" by the cell nucleus genome from the cpDNA is a gene coding the chloroplast translation initiation factor 1 (INFA). A plant's INFA protein is a homologue of the *Escherichia coli* *infA* gene product and has been identified in plants such as *A. thaliana*, soybean (*Glycine max*), tomato (*Solanum lycopersicum*) and *Mesembryanthemum crystallinum*. Studies carried out on *A. thaliana* and Soybean (*Glycine max*), showed that the *INFA* gene is translated in the cytosol. However, due to the chloroplast transit sequence, which is attached to the formulated protein, the whole complex goes to the appropriate compartment in the chloroplast (Daniell et al. 2016). A similar situation takes place in case of the *RPL22* and *RPL32* genes encoding ribosomal proteins L22 and L32, respectively (Daniell et al. 2016). A more complex problem concerns the family of 12 chloroplast genes that encode a subunit of the chloroplast NAD(P)

H dehydrogenase (*NDH*) complex involved in photosynthesis. The NDH complex is attached to the photosystem I (PS I) and mediates cyclic electron transport during the bright photosynthesis phase. The distribution of the *NDH* genes between the nuclear, mitochondrial and the chloroplast genome varies between plant species. Moreover, data generated by the full genome sequencing of plants such as black pine (*Pinus thunbergii*), saguaro (*Carnegieia gigantea*) or erodium cicutarium (*Erodium cicutarium*) shows a partial or complete deletion of genes from the *NDH* family. However, the lack of genes encoding the NDH dehydrogenase subunits does not disturb these plants from carrying out an efficient photosynthesis process (Daniell et al. 2016). The mass transfer of the genes from the cpDNA to the nucleus also took place in the case of a wide group of genes related to DNA repair and recombination process, called the *RAR* genes. Research in this field is of particular importance to medicine, due to high homology between the large number of genes involved in repair and recombination of DNA identified in the *A. thaliana* and human genes whose mutations are considered to be caused by diseases such as breast cancer, non-lipid colon cancer or Cockayne syndrome (The Arabidopsis Genome Initiative 2000).

## Organization of genes in the *A. thaliana* chloroplast genome

*Arabidopsis thaliana* is one of the best known model plants, for decades widely used in various studies. Its cpDNA complete nucleotide sequence has been demonstrated in 1999 (Sato et al. 1999). Depending on the different factors, the chloroplast genome of *A. thaliana* can take both the spatial form, namely circular and linear DNA molecule. (Oldenburg and Bendich 2015) It is composed of 154,478 base pairs and are more than 2 times smaller than the mitochondrial genome (mtDNA), which has about 367,000 bp (Table 1). The *A. thaliana* cpDNA contains two IRs of 26,264 bp each and separating them LSC and SSC of 84,197 and 17,780 bp, respectively

**Table 1** General comparison of the *A. thaliana* nuclear, plastid and mitochondrial genomes (based on the TAIR10.1)

|                                | Nucleus     | Plastid | Mitochondrion |
|--------------------------------|-------------|---------|---------------|
| Genome size (bp)               | 119,668,634 | 154,478 | 367,808       |
| Number of genes                | 38,311      | 129     | 284           |
| Number of protein coding genes | 48,265      | 85 (87) | 33            |
| GC content (%)                 | 36.05       | 36.3    | 44.8          |
| tRNA                           | 684         | 37      | 22            |

*GC content (%)* guanine-cytosine content





chloroplast proteins, three duplicated ribosomal RNA genes localized in IRs region can be distinguished. There is also one more RNA gene localized in LSC encoded 5S rRNA. The 3 mentioned ribosomal RNA duplicated genes create a cluster of 16S-23S-4.5S rRNA genes separated by two tRNA genes (*TRN1* and *TRNA*). As many as 37 genes encoding tRNA molecules are located within the regions of LSC and SSC (Sato et al. 1999).

A comparison of the general characteristics of the nuclear, mitochondrial and chloroplast genome in *A. thaliana* is shown in Table 1. Presented data indicate that the cpDNA is smaller than the nuclear and mtDNA. However, the number of the protein coding genes encoded by the cpDNA is larger than those encoded within mtDNA. In turn, Table 2 presents a list of 87 protein coding genes located in the *A. thaliana* cpDNA. One duplicated gene namely *ORF77 (YCF15)* is not included in all sources. 45 genes of the cpDNA encode proteins related to the photosynthesis process. Among them, 6 genes coding for protein subunits of the cytochrome b6f complex, 5 genes coding for proteins being a part of the photosystem I (PS I), 15 genes coding for proteins being a part of the photosystem II (PS II), 6 genes encoding the ATPase chloroplast subunits, 12 genes encoding the NADH dehydrogenase subunits and one gene encoding the RuBisCo large subunit. In addition, the *A. thaliana* cpDNA includes 29 genes related to the gene expression processes, including 4 genes encoding the chloroplast RNA polymerase subunits and 25 genes encoding protein components of the ribosomes (Sato et al. 1999). The list of chloroplast genes supplements 12 genes with other and also not known yet, physiological functions (The Arabidopsis Information Resource (TAIR), The UniProt Consortium (UniProt).

## Chloroplast transcription machinery

It is believed that chloroplasts emerged as a consequence of endosymbiosis between photosynthetic cyanobacteria and the ancestor of today's eukaryotic plant cells followed by various genome rearrangements (Timmis et al. 2004). In comparison to the cyanobacteria genome size, which is ca. 3 Mbp, the terrestrial plants' cpDNA is 20 times smaller. Despite such a difference in size between those two genomes the expression of chloroplast genes is regulated by much more complex systems than cyanobacteria genes (Yagi and Shiina 2014). Importantly, the expression of chloroplast genes is strongly dependent on post-transcriptional regulation, which includes the polycistronic mRNA processing, intron splicing and RNA editing (Yagi and Shiina 2014).

The transcription of chloroplast genes in angiosperm plants is carried out by two types of RNA polymerases, the bacterial-type multi-subunit RNA polymerase (PEP)

encoded by the cpDNA and the T3/T7 phage-type RNA polymerases (NEP) encoded by the nucleus genome.

PEP is an enzyme composed of many subunits. Although most of the genes for PEP subunits have been transferred to the nuclear genome, genes coding  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  subunits of the PEP's core are retained in the cpDNA. One of the fundamental differences between the PEP core subunits is their molecular weight. The  $\alpha$  subunit has a mass of 38 kDa,  $\beta$ —120 kDa,  $\beta'$ —85 kDa,  $\beta''$ —185 kDa (Liere and Börner 2007). Similarly, to bacteria, in most terrestrial plants, the *RPOA* gene encoding  $\alpha$  subunit of the PEP core, together with the genes of the ribosomal proteins are organized into one operon under the control of the same promoter. Contrarily, *RPOB*, *RPOC1* and *RPOC2* genes encoding  $\beta$ ,  $\beta'$  and  $\beta''$  subunits of the PEP core respectively, form a separate operon designated as *RPOBC* (Börner et al. 2015). Besides the core subunits, PEP is composed of additional protein factors encoded by the nuclear genome including sigma factors and the polymerase associated proteins (PAPs). The chloroplast sigma factors, which in fact stands for the bacterial transcription initiation factors, play a crucial role in the chloroplast-encoded gene transcription. Sigma factors regulate transcription on different developmental and fitness stages by recognition of the specific promoters, which allow a whole complex of the PEP to start polymerization. It seems likely that the second highlighted PEP's factor, the PAPs are involved in almost every single step of the transcription (Pfannschmidt et al. 2015).

NEP is a single-subunit enzyme. One protein performs the entire transcription from the recognition of the promoter to the end of the process, regardless of the DNA template structure (Börner et al. 2015). The NEP polymerase has evolved through the duplication of the nuclear gene encoding the mitochondrial RNA polymerase and shows a high similarity to the phage-type RNA polymerases, also made of one subunit (Börner et al. 2015). Three types of NEP polymerases are distinguishable. All three types of the NEP polymerase are encoded by the *RPOT* genes. The *RPOT* abbreviation comes from the full name, the RNA polymerase of the phage T3/T7 type. In monocotyledonous plants *RPOTp* polymerase occur, whereas in dicotyledonous plants *RPOTp* and *RPOTmp* have been identified (Yagi and Shiina 2014). In addition, the third enzyme form the NEP family, *RPOTm* occurred in the mitochondria.

Both PEP and NEP are essential for the transcription of chloroplast proteins. Even though NEP and PEP recognize different promoters, many chloroplast genes have promoters recognized both by PEP and by NEP. The promoters recognized by NEP can be divided into three types, designated as Ia, Ib and II (Morley et al. 2019b). All promoters belonging to the Ia type are characterized by a conserved core motif designated by the abbreviation YRTa. In the Ia type promoters several nucleotides are located above the transcription

**Table 2** Representation of all *A. thaliana* protein-coding genes encoded by cpDNA (Sato et al. 1999)

|    | Gene                       | Localization within the cpDNA | Biological function of a gene product | Name of a gene product                         | Intron presence | Number of aminoacids |
|----|----------------------------|-------------------------------|---------------------------------------|--|-----------------|----------------------|
| 1  | <i>RPOA</i> <sup>ab</sup>  | LSC                           | Transcription                         | RNA polymerase $\alpha$ subunit (PEP)          | No              | 329                  |
| 2  | <i>RPOB</i> <sup>ab</sup>  | LSC                           | Transcription                         | RNA polymerase $\beta$ subunit                 | No              | 1072                 |
| 3  | <i>RPOC</i> <sup>ab</sup>  | LSC                           | Transcription                         | Podjednostka $\beta'$ polimerazy RNA (PEP)     | Yes             | 680                  |
| 4  | <i>RPOC2</i> <sup>ab</sup> | LSC                           | transcription                         | RNA polymerase $\beta''$ -subunit              | No              | 1376                 |
| 5  | <i>RPL2</i> <sup>ab</sup>  | IR                            | Translation                           | Ribosomal protein L2                           | Yes             | 274                  |
| 6  | <i>RPL2</i> <sup>ab</sup>  | IR                            | Translation                           | Ribosomal protein L2                           | Yes             | 274                  |
| 7  | <i>RPL14</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L14                          | No              | 122                  |
| 8  | <i>RPL16</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L16                          | yes             | 135                  |
| 9  | <i>RPL20</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L20                          | No              | 117                  |
| 10 | <i>RPL22</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L22                          | No              | 160                  |
| 11 | <i>RPL23</i> <sup>ab</sup> | IR                            | Translation                           | Ribosomal protein L23                          | No              | 93                   |
| 12 | <i>RPL23</i> <sup>ab</sup> | IR                            | Translation                           | Ribosomal protein L23                          | No              | 93                   |
| 13 | <i>RPL32</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L32                          | No              | 52                   |
| 14 | <i>RPL33</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L33                          | No              | 66                   |
| 15 | <i>RPL36</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein L36                          | No              | 37                   |
| 16 | <i>RPS2</i> <sup>ab</sup>  | LSC                           | Translation                           | Ribosomal protein S2                           | No              | 236                  |
| 17 | <i>RPS3</i> <sup>ab</sup>  | LSC                           | Translation                           | Ribosomal protein S3                           | No              | 218                  |
| 18 | <i>RPS4</i> <sup>ab</sup>  | LSC                           | Translation                           | Ribosomal protein S4                           | No              | 201                  |
| 19 | <i>RPS7</i> <sup>ab</sup>  | IR                            | Translation                           | Ribosomal protein S7                           | No              | 155                  |
| 20 | <i>RPS7</i> <sup>ab</sup>  | IR                            | Translation                           | Ribosomal protein S7                           | No              | 155                  |
| 21 | <i>RPS8</i> <sup>ab</sup>  | LSC                           | Translation                           | Ribosomal protein S8                           | No              | 134                  |
| 22 | <i>RPS11</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein S11                          | No              | 138                  |
| 23 | <i>RPS12</i> <sup>ab</sup> | IR                            | Translation                           | Ribosomal protein S12                          | Yes             | 123                  |
| 24 | <i>RPS12</i> <sup>ab</sup> | IR                            | Translation                           | Ribosomal protein S12                          | Yes             | 123                  |
| 25 | <i>RPS14</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein S14                          | No              | 100                  |
| 26 | <i>RPS15</i> <sup>ab</sup> | SSC                           | Translation                           | Ribosomal protein S15                          | Yes             | 88                   |
| 27 | <i>RPS16</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein S16                          | Yes             | 79                   |
| 28 | <i>RPS18</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein S18                          | No              | 101                  |
| 29 | <i>RPS19</i> <sup>ab</sup> | LSC                           | Translation                           | Ribosomal protein S19                          | No              | 92                   |
| 30 | <i>PETA</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome f                                   | No              | 320                  |
| 31 | <i>PETB</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome b6                                  | Yes             | 215                  |
| 32 | <i>PETD</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome b6/f complex subunit 4              | Yes             | 160                  |
| 33 | <i>PETG</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome b6/f complex subunit 5              | No              | 37                   |
| 34 | <i>PETL</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome b6/f complex subunit 6              | No              | 31                   |
| 35 | <i>PETN</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Cytochrome b6/f complex subunit 8              | No              | 29                   |
| 36 | <i>PSAA</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Photosystem I P700 apoprotein A1               | No              | 750                  |
| 37 | <i>PSAB</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Photosystem I P700 apoprotein A2               | No              | 734                  |
| 38 | <i>PSAC</i> <sup>ab</sup>  | SSC                           | Photosynthetic apparatus              | Photosystem I subunit VII (iron-sulfur center) | No              | 81                   |
| 39 | <i>PSAI</i> <sup>ab</sup>  | LSC                           | Photosynthetic apparatus              | Photosystem I reaction center subunit VIII     | No              | 37                   |

**Table 2** (continued)

| Gene                         | Localization within the cpDNA | Biological function of a gene product | Name of a gene product                       | Intron presence | Number of aminoacids |
|------------------------------|-------------------------------|---------------------------------------|--|-----------------|----------------------|
| 40 <i>PSAJ</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem I reaction center subunit IX     | No              | 44                   |
| 41 <i>PSBA</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II reaction center protein D1    | No              | 353                  |
| 42 <i>PSBB</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II CP47 chlorophyll apoprotein   | No              | 508                  |
| 43 <i>PSBC</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II CP43 chlorophyll apoprotein   | No              | 473                  |
| 44 <i>PSBD</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II reaction center protein D2    | No              | 353                  |
| 45 <i>PSBE</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II cytochrome b559 alpha subunit | No              | 83                   |
| 46 <i>PSBF</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II cytochrome b559 beta subunit  | No              | 39                   |
| 47 <i>PSBH</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II 10 kDa phosphoprotein         | No              | 73                   |
| 48 <i>PSBJ</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein I                     | No              | 36                   |
| 49 <i>PSBJ</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein J                     | No              | 40                   |
| 50 <i>PSBK</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein K                     | No              | 61                   |
| 51 <i>PSBL</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein L                     | No              | 38                   |
| 52 <i>PSBM</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein M                     | No              | 34                   |
| 53 <i>PSBN</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein N                     | No              | 43                   |
| 54 <i>PSBT</i> <sup>ab</sup> | LSC                           | Photosynthetic apparatus              | Photosystem II protein T                     | No              | 33                   |
| 55 <i>PSBZ</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | Photosystem II subunit PsbZ                  | No              | 62                   |
| 56 <i>ATPA</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF1 subunit alpha               | No              | 507                  |
| 57 <i>ATPB</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF1 subunit beta                | No              | 498                  |
| 58 <i>ATPE</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF1 subunit epsilon             | No              | 132                  |
| 59 <i>ATPF</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF0 subunit I                   | Yes             | 184                  |
| 60 <i>ATPH</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF0 subunit III                 | No              | 81                   |
| 61 <i>ATPI</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | ATP synthase CF0 subunit IV                  | No              | 249                  |
| 62 <i>NDHA</i> <sup>ab</sup> | SSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 1  | Yes             | 360                  |
| 63 <i>NDHB</i> <sup>ab</sup> | IR                            | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 2  | Yes             | 389 or 512           |
| 64 <i>NDHB</i> <sup>ab</sup> | IR                            | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 2  | Yes             | 389 or 512           |
| 65 <i>NDHC</i> <sup>ab</sup> | LSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 3  | No              | 120                  |
| 66 <i>NDHD</i> <sup>ab</sup> | SSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 4  | No              | 506 or 500           |
| 67 <i>NDHE</i> <sup>ab</sup> | SSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 4L | No              | 101                  |
| 68 <i>NDHF</i> <sup>ab</sup> | SSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 5  | No              | 746                  |
| 69 <i>NDHG</i> <sup>ab</sup> | SSC                           | Photosynthetic metabolism             | NADH-plastoquinone oxidoreductase subunit 6  | No              | 176                  |

**Table 2** (continued)

| Gene  | Localization within the cpDNA | Biological function of a gene product   | Name of a gene product                                    | Intron presence | Number of aminoacids |
|---|-------------------------------|---|---|-----------------|----------------------|
| 70 <i>NDHH</i> <sup>ab</sup>                              | SSC                           | Photosynthetic metabolism               | NADH-plastoquinone oxidoreductase subunit 7               | No              | 393                  |
| 71 <i>NDHI</i> <sup>ab</sup>                              | SSC                           | Photosynthetic metabolism               | NADH-plastoquinone oxidoreductase subunit I               | No              | 172                  |
| 72 <i>NDHJ</i> <sup>ab</sup>                              | LSC                           | Photosynthetic metabolism               | NADH-plastoquinone oxidoreductase subunit J               | No              | 158                  |
| 73 <i>NDHK</i> <sup>a</sup> Or <i>PSBG</i> <sup>b</sup>   | LSC                           | Photosynthetic metabolism               | NADH-plastoquinone oxidoreductase subunit K               | No              | 225                  |
| 74 <i>RBCL</i> <sup>ab</sup>                              | LSC                           | Photosynthetic metabolism               | RuBisCo large subunit                                     | No              | 479                  |
| 75 <i>ACCD</i> <sup>ab</sup>                              | LSC                           | Photosynthetic metabolism               | acetyl-CoA carboxylase, carboxyl transferase subunit beta | No              | 488                  |
| 76 <i>CLPP</i> <sup>ab</sup>                              | LSC                           | Protelitic function                     | CLP protease ATP binding subunit                          | Yes             | 196                  |
| 77 <i>MATK</i> <sup>ab</sup>                              | LSC                           | Probable splicing factor                | intron maturase K   | No              | 526 or 504           |
| 78 <i>YCF1</i> <sup>ab</sup>                              | IR                            | Protein transport                       | Hypothetical protein ArthCp087                            | No              | 1786                 |
| 79 <i>YCF1</i> <sup>ab</sup>                              | IR                            | Protein transport                       | Hypothetical protein ArthCp070                            | No              | 343                  |
| 80 <i>YCF2</i> <sup>ab</sup>                              | IR                            | ATPase activity                         | Conserved hypothetical chloroplast protein YCF2           | No              | 2294                 |
| 81 <i>YCF2</i> <sup>ab</sup>                              | IR                            | ATPase activity                         | Conserved hypothetical chloroplast protein YCF2           | No              | 2294                 |
| 82 <i>YCF3</i> <sup>ab</sup>                              | LSC                           | Probable photosystem I complex assembly | photosystem I assembly protein YCF3                       | Yes             | 168                  |
| 83 <i>YCF4</i> <sup>ab</sup>                              | LSC                           | Probable photosystem I complex assembly | photosystem I assembly protein YCF4                       | No              | 184                  |
| 84 <i>YCF5</i> <sup>b</sup> or <i>CCSA</i> <sup>a</sup>   | SSC                           | cytochrome complex assembly             | Cytochrome c biogenesis protein CCSA                      | No              | 328                  |
| 85 <i>YCF10</i> <sup>b</sup> or <i>CEMA</i> <sup>a</sup>  | LSC                           | Probable proton transporter             | Envelope membrane protein                                 | No              | 229                  |
| 86 <i>ORF77</i> <sup>b</sup> or <i>YCF15</i> <sup>b</sup> | IR                            | Unknown function                        | Uncharacterized protein YCF15                             | No              | 77                   |
| 87 <i>ORF77</i> <sup>b</sup> or <i>YCF15</i> <sup>b</sup> | IR                            | Unknown function                        | Uncharacterized protein CYF15                             | No              | 77                   |

<sup>a</sup>Information about gene available in NCBI

<sup>b</sup>Information about gene available in TAIR

start sequence. NEP's type Ib promoters contain an additional conservative motif in their structure, called "GAA box", which is located between the 18th and 20th nucleotide above the YRTa motif. Assays carried out on mutated tobacco have indicated a crucial role of the GAA motif for a proper recognition of the promoter conducted by NEP (Liere and Börner 2007). Contrarily, type II promoters all consist of the NEP promoters lacking the YRTa motif.

Many promoters recognized by PEP polymerase resemble bacterial  $\sigma 70$  promoters, characterized by two motifs of canonical sequences spaced from the transcription start site by 10 and 35 nucleotides. The first motive, 10 nucleotides away from the transcription site, is the TATAAT sequence, while the second motif—remote from the transcription site

by 35 nucleotides—is the TTGACT sequence. Due to the huge diversity among plants, the location of the canonical sequences of certain PEP promoters may vary. For example, in barley's cpDNA, the TATAAT sequence is located 3–9 nucleotides upstream from the transcription start site, while TTGACT is located 15–21 nucleotides from the same transcription initiation site (Börner et al. 2015). PEP is mostly responsible for the transcription of chloroplast genes, whose protein products are in a various way associated with the photosynthesis. Although some of the genes coding proteins involved in photosynthesis are also transcribed by NEP (Yagi and Shiina 2014).

There is a small group of chloroplast genes unrelated to photosynthesis, which are transcribed exclusively by the



NEP. It includes the *ACCD* gene encoding the acetyl-CoA carboxylase subunit in dicotyledonous plants; the *RPL23* gene coding for the ribosomal protein L23; *CLPP* gene encoding the ATP-dependent proteolytic subunit of ClpP protease in monocotyledonous plants; and the *RPOB* gene encoding the  $\beta$ -subunit of the PEP<sup>+</sup> core in all higher plants (Liere and Börner 2007). Thus, chloroplast genes can be divided together with their promoters into 3 categories: genes transcribed only with the participation of NEP polymerase, genes transcribed only with the participation of PEP polymerase and genes that are transcribed both by NEP and by PEP polymerase (Hajdukiewicz 1997). In the case of dicotyledonous plants where two different NEP have been distinguished, the expanded division of the promoters can be proposed. The activity of RPOTp and RPOTmp vary in tissues at various stages of a plant development. In *A. thaliana*, the increased activity of RPOTmp is observed mainly in young dividing cells and photosynthetically inactive tissues, whereas the boosting activity of RPOTp is observed in green, photosynthetically active tissues. Differences in the structure of the promoters recognized by both types of NEP polymerase were also detected. Interestingly, both NEP and PEP are active at all stages of plant development in plastids of the unbleached tissues such as roots, fruits or seeds. Constant activity of these polymerases is related to their participation in the transcription of the housekeeping genes such as genes encoding tRNA (Börner et al. 2015).

### The regulation and control of the chloroplasts' proteins expression

The expression of the genes encoded by cpDNA is largely dependent on an extensive group of nuclear origin factors. In turn, these factors regulate expression of the plastid genes in response to various developmental and environmental signals (Barkan and Goldschmidt-Clermont 2000). Regulation factors are widely present at different stages of the plastid genes expression including transcription, RNA editing, RNA post-transcriptional modification, RNA splicing and translation.

### Nuclear control of chloroplast gene transcription

The transcription of the cpDNA genes is controlled by different factors of nuclear origin. The primary factors affecting the cpDNA genes transcription are NEP polymerase and additional or so-called non-core subunits of PEP polymerase. In the group of the additional subunits of PEP polymerase, the additional nuclear-encoded protein factors (PAPs) and

transcriptional initiation factors (sigma factors, SIG), are distinguishable.

It is well known that in overall, PAPs are essential in the transcription regulation. However, some of the PAPs's precise functions remain unestablished. On the contrary, the SIGs are crucial for specific binding of PEP polymerase to the promoters of the respective genes (Chi et al. 2015). As many as six different SIGs are involved in the transcription of *A. thaliana*'s genes. The plant's SIGs of bacterial origin show homology to their ancestors only through the conservative region located at the end of their molecule. The so-called non-conservative region appears to be decisive for the function of the particular SIG. The SIGs are regulated by the phosphorylation of the specific sequences within mentioned non-conservative region. Research on the SIGs's activation has been conducted almost exclusively on SIG1 and SIG6. The SIGs phosphorylation seems to be a complex process conducted by different enzymes. In 1996, Baginsky and his group suggested that the phosphorylation of SIG6 is conducted by a PEP-associated Ser/Thr protein kinase named plastid transcription kinase (PTK; Baginsky 1997). In addition, in literature PTK is commonly abbreviated as cpCK2 due to the similarities between catalytic components of PTK and the subunit of casein kinase 2 (CK2). Although, it has been demonstrated that cpCK2 uses SIG6 as a substrate for regulatory phosphorylation, SIG6's most sensitive phosphorylation site is not typical for cpCK2. Therefore, it entailed the hypothesis of other kinase(s) involved in this process (Chi et al. 2015). The phosphorylation of the SIGs can both initiate and block the transcription of the genes recognized by PEP polymerase complex. The type of promoter recognized by a given factor seems decisive in this process. For example, phosphorylation of the SIG6 factor is crucial for transcription of the *ATPB* gene but has no effect on *PSBA* gene transcription. In turn, the lack of phosphorylation of the SIG1 factor reduces the transcription of both the *DOGA* and *PSBA* genes (Chi et al. 2015). A brief description of the all 6 identified in *A. thaliana* SIGs is demonstrated in Table 3.

The transcription of the genes encoded by cpDNA can also be affected by smaller molecules like rare nucleotides. Recent studies show an arising interest in these molecules. They are considered as the novel signaling molecules evoking response of plant to different types of biotic and abiotic stresses. In this case they were called alarmones (Pietrowska-Borek 2014). In vitro studies shown that under various stress conditions, guanosine tetraphosphate (ppGpp) is produced in plastids and then binds to the  $\beta$ -subunit of PEP polymerase, inhibiting RNA synthesis (Liere et al. 2011).

**Table 3** Description of the *A. thaliana* sigma factors ( $\sigma$ ) family members (Yagi and Shiina 2014; Börner et al. 2015; Chi et al. 2015)

| Sigma factor ( $\sigma$ ) | Description and function   |
|---------------------------|--|
| SIG1                      | The sigma factor characterized with the highest abundance. It is expressed during the latest stage of the leaf development. It binds to the promoters of genes encoding proteins included in PS I and PS II. It plays an important role in PS I acclimatization to the light quality changes. It is involved in the WRKY3 dependent biotic stress response         |
| SIG2                      | Involved in the transcription of several tRNAs such as the <i>DOGJ</i> gene and the <i>ATP</i> operon. Participates in signaling between plastid and nucleus. Cooperates with SIG6 in the chloroplast development process. It is also believed that SIG2 is a key factor favoring PEP polymerase over NEP in transcription at early plant development stages       |
| SIG3                      | Involved in the transcription of <i>PSBN</i> , <i>ATP</i> and antisense RNA into <i>PSBT</i> . Mutants without SIG3 factor do not show phenotypic changes  |
| SIG4                      | It regulates the transcription of the <i>NDHF</i> gene. Factor-free mutants do not show phenotypic changes   |
| SIG5                      | SIG5 expression is induced by a wide range of abiotic stresses. SIG5 is involved in the transcription of the <i>PSBD</i> gene and in the developmental processes on the embryonic stage. It is believed that SIG5 plays a significant role in the control of circadian regulation of plastid gene transcription  |
| SIG6                      | Expressed during the initial stages of chloroplast biogenesis. It cooperates with SIG2 in the chloroplast development process, as evidenced by anatomical chlorophyll disorder in mutants lacking SIG6. SIG6 is involved in the transcription of <i>ATP</i> genes. It interacts with the DG1 protein and participates in the signaling between plastid and nucleus |

### The control of the chloroplast RNA editing and RNA post-transcriptional modification mediated by nucleus

RNA editing, as one of the post-transcriptional gene expression processes, alters the identity of nucleotides so that information about the protein encoded in mRNA differs from the prediction of genomic DNA. The editing events include the deletion, insertion and base substitution of nucleotides. The first demonstrated modification of the RNA molecule mediated by the RNA editing has been observed in mitochondrial mRNA of kinetoplastid protozoa (Benne et al. 1986; reviewed in Benne 1986). An example of the RNA edition occurring in the chloroplast might be the conversion of cytosine into uracil in the flowering plants. The whole process of the RNA edition depends on specific proteins. Around 40 editing spots at the chloroplast RNA of the flowering plants have been located, to which specific proteins bind. These proteins contain repeating elements that bind to the 5' end motifs of the related RNA sequence at a distance of 20–25 nucleotides from the edited nucleotide. Then, this RNA–protein complex reacts with other proteins. It is suggested that the latter proteins are specific linkages between the complex and the unknown enzyme molecule with deaminase activity (Takenaka et al. 2013). There is strong evidence that the process of chloroplast RNA editing involves factors encoded only by nuclear genes. The editing of chloroplast RNA was detected in a barley mutant lacking plastid ribosomes and in the presence of chloroplast translation inhibitors (Barkan and Goldschmidt-Clermont 2000).

One of the most important stages of post-transcriptional RNA processing is a process called RNA maturation. This process involves egzo- and endonucleolytic processing of mRNA. In case of the terrestrial plants, the 5' end of the mRNA is often changed. In this process, the most prominent

molecules are proteins belonging to the PPR protein family, which in the case of *A. thaliana* has more than 450 individuals. An example is PPR10 protein encoded by the *PPR10* gene. Mutation within *PPR10* gene causes changes in the accumulation of transcripts that are not subject to egzo-nucleotide processing. The only explanation for this situation could be the hypothesis that the PPR10 protein has the 5' and 3' protective function of the RNA end of the transcripts of some genes. Another protein belonging to the PPR protein family is PGR3 protein stabilizing the RNA of the *PETL* operon and thus responsible for the correct translation process of the *PETL* and *NDHA* gene (Shikanai and Fujii 2013).

### Splicing of the chloroplast genes

Only a few chloroplast genes contain a non-coding intron sequences in their structure. This group includes the *RPOCI* gene encoding the  $\beta'$  subunit of PEP, the *RPL2* gene encoding the CL2 protein of the large (50S) ribosomal subunit, or the *RPS16* gene encoding the CS16 protein of the small (30S) ribosomal subunit. All of the genes encoded by the cpDNA containing introns are described in Table 2. Introns of the chloroplast genes can be divided into two groups, type I and type II introns. Both types are often called self-trimming introns, but some studies showed that in case of group I introns this statement is true only under in vitro condition. There were no nuclear mutations directly affecting the chloroplast group I intron processing. However, genetic analysis of the genomes clearly indicated factors affecting the group II intron splicing. It has been demonstrated that mutations within the nuclear coded *CRS1* gene from maize negatively affects the splicing of the intron from the chloroplast-coded *ATPF* gene. The same study shows that the

*CRS2* gene affects the normal splicing of many introns from the second group (Barkan and Goldschmidt-Clermont 2000).

## Regulation of the chloroplast gene translation

The *A. thaliana* cpDNA contains a total of 129 protein coding and non-protein coding genes. Chloroplasts are equipped with a semi-autonomous expression apparatus that was inherited from their cyanobacterial ancestors. The last and at the same time one of the key stages of gene expression is the process of polypeptides chain synthesis on the mRNA matrix, i.e. translation. The complete translation process within the chloroplast stroma is possible due to the presence of the bacterial 70S ribosomes and a number of both nucleus and chloroplast encoded translational factors (Sun and Zerges 2015). Biochemical methods of protein identification were the main way to characterize plastid translation machinery. This method allowed the identification of the chloroplast ribosomal proteins encoded by the nuclear genome, translation initiation and elongation factors and tRNA synthetases (Barkan and Goldschmidt-Clermont 2000).

The contribution of the translational factors to the chloroplast genes expression regulation is closely dependent on the plant's ontogenesis stage. Studies in the initial phases of chloroplast differentiation revealed high activity of PEP polymerase and, consequently, a drastic increase in mRNA encoding proteins of the photosynthetic apparatus. However, little is known about the regulation of translation in this process. Well-known translation regulator in *A. thaliana* is the nuclear-encoded ATAB2 factor. This factor is responsible for the correct placement of the polycistronic mRNA of *PSAA-PSAB* and *PSBC-PSBD* operons on ribosomes (Sun and Zerges 2015). Another essential nuclear-encoded protein for proper expression of cpDNA encoded proteins is pentatricopeptide repeat protein (LPE1). Lack of LPE1 protein cause in a decrease in *psbN* translation and a complete loss of *psbJ* translation (Williams-Carrier et al. 2019). Recently, a CHLOROPLAST RIBOSOME ASSOCIATED (CRASS) protein has been identified. It is a nuclear-encoded protein interacting with proteins of the 30S subunit either directly or via another factor. CRASS is not essential for plant survival under controlled growth conditions. However, it supports ribosomal activity under stressful conditions, especially upon cold treatment, by playing a regulatory or stabilization role of the 30S subunit (Pulido et al. 2018). Studies focusing on the chloroplasts' translation in the mature tissues indicate that light is the main external factor regulating this process. The energy in the solar rays reaching the chloroplasts affects the multilevel regulation of the synthesis and redox states of some biochemical elements of the photosynthetic apparatus.

These changes ultimately lead to the modulation of the proton gradient ( $\Delta\text{pH}$ ) across the thylakoid membrane. In this way, among others, a pH-sensitive factor is activated and regulates translation of the *PSBA* gene encoding the photosystem II core D1 protein. Another gene whose regulation at the level of translation has been well understood is the *PSBB* gene encoding the protein component of the CP47 complex, constituting the internal PSII energy antenna. This gene is regulated by the translational factors NAC2 and RBP40. Through the light-mediated redox reactions, these factors bind to each other by covalent linkages to form a *PSBB* gene translation-promoting complex. This activation consists of changing the secondary structure preventing the initiation of the translation process from occurring on chloroplast ribosomes (Sun and Zerges 2015).

Numerous studies on *A. thaliana* under the light stress were conducted. Transcriptomic analyses showed that this type of abiotic stress, among others, cause high accumulation of mRNA of *PSBA* gene encoding D1 protein of photosystem II (Sakurai et al. 2012; Walter et al. 2015; Adamiec et al. 2018). D1 protein is highly sensitive to the high light intensity and in case of any damage rapidly replacement. One of the co-translational factors responsible for this process is the nuclear factor REP27. In opposite, translation of the *RBCL* gene encoding the RuBisCo is inhibited during the high light stress condition. The regulation of this process has not been described (Sun and Zerges 2015). The regulation of chloroplast gene expression by microRNA is still a very poorly understood process. Available literature data only indicate miR398 as a regulator of gene expression encoding both cytoplasmic and chloroplast superoxide dismutase (Han-Hua et al. 2008).

## Summary

During an event estimated to have happened over 1.2 billion years ago within the process of endosymbiosis, free-living cyanobacterium started to coexist inside a eukaryotic cell, which turned out to be essential for life as we know it. During the evolution of the plant cell the cpDNA undergone a substantial reduction. Many genes of the ancestral chloroplasts have been transferred from the cpDNA into the cell nucleus. Most of the chloroplast proteins are encoded by the nucleus, however the chloroplast genomes still encode between 50 and 200 essential proteins depending on the plant species. In comparison, cyanobacterial genomes encode several thousand proteins. Despite such a notable reduction of the encoded protein in the cpDNA, the expression of chloroplast genes is regulated by much more complex systems than cyanobacteria genes. Chloroplasts are considered to be half-independent organelle; however, they are in constant close relation with nucleus. The most pronounced

example of this relation is the presence of both PEP and NEP polymerases and the factors necessary for their proper functioning. The cpDNA differs between the plant species both in term of the size and the form they may take. Recent cpDNAs studies show that these molecules can take both a multi branched linear as well as circular structures. All this together gives an image of a diverse and quite complicated genetic system occurring inside chloroplasts. Understanding of chloroplast genetic machinery is one of the basic conditions for a full explanation of the functioning of chloroplasts and the processes occurring in them, including photosynthesis—the basic process that allows life on Earth in the form we know.

**Author contributions statement** JD: wrote the draft of the manuscript, prepared all tables, and was involved in writing of the final version of the manuscript. RL was involved in writing of the manuscript and literature screening. MA and RL formulated conception of the manuscript and supervised all stages of writing of the manuscript.

**Acknowledgements** We thank Professor Emmanuel Douzery (Université de Montpellier, FR) for letting us use the chloroplast genome map which is his original work.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adamiec M, Misztal L, Kosicka E, Paluch-Lubawa E, Luciński R (2018) *Arabidopsis thaliana* *egy2* mutants display altered expression level of genes encoding crucial photosystem II proteins. *J Plant Physiol* 231:155–167
- Baginsky S, Tiller K, Link G (1997) Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (*Sinapis alba*). *Plant Mol Biol* 34:181–189
- Barkan A, Goldschmidt-Clermont M (2000) Participation of nuclear genes in chloroplast gene expression. *Biochimie* 82:559–572
- Benne R, Van Den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC (1986) Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46:819–826
- Börner T, Aleynikova AY, Zuboa YO, Kusnetsov VV (2015) Chloroplast RNA polymerases: role in chloroplast biogenesis. *Biochim Biophys Acta* 9:761–769
- Brehelin C, Kessler F, van Wijk KJ (2007) Plastoglobules: Versatile lipoprotein particles in plastids. *Trends Plant Sci* 12:206–266
- Chi W, He B, Mao J, Jiang J, Zhang L (2015) Plastid sigma factors: their individual functions and regulation in transcription. *Biochim Biophys Acta* 9:770–778
- Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution and applications in genetic engineering. *Genome Biol* 17:2–29
- Goulas E, Schubert M, Kieselbach T, Kleczkowski LA, Gardeström P, Schröder W, Hurry V (2006) The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short- and long-term exposure to low temperature. *Plant J* 47:720–734
- Hajdukiewicz PTJ, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J* 16:4041–4048
- Han-Hua L, Xin T, Yan-Jie L, Chang-Ai W, Cheng-Chao Z (2008) Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA* 14:836–843
- Heckathorn SA, Downs CA, Coleman JS (1998) Nuclear-encoded chloroplast proteins accumulate in the cytosol during severe heat stress. *Int J Plant Sci* 159:39–45
- Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bedard J, Hirabayashi-Ishioka Y et al (2018) A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* 30:2677–2703
- Ku C, Nelson-Sathi S, Roettger M, Garg S, Hazkani-Covo E, Martin WF (2015) Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc Natl Acad Sci USA* 33:10139–10146
- Liere K, Börner T (2007) Transcription and transcriptional regulation in plastids. *Cell and molecular biology of plastids*. Springer-Verlag, Heidelberg, Berlin, pp 121–174
- Liere K, Weihe A, Börner T (2011) The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J Plant Physiol* 168:1345–1360
- Ling Q, Huang W, Baldwin A, Jarvis P (2012) Chloroplast biogenesis is regulated by direct action of the ubiquitin–proteasome system. *Science* 338:655–659
- Lippold F, vom Dorp K, Abraham M, Holz G, Wewer V, Yilmaz JL, Lager I, Montandon C, Besagni C, Kessler F, Stymne S, Dormann P (2012) Fatty acid phytyl ester synthesis in chloroplasts of *Arabidopsis*. *Plant Cell* 24:2001–2014
- Maier RM, Neckermann K, Igloi GL, Kössel H (1995) 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
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) 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
- Morley SA, Peralta-Castro A, Briebe LG, Miller J, Ong KL, Ridge PG, Nielsen BL (2019a) *Arabidopsis thaliana* organelles mimic the T7 phage DNA replisome with specific interactions between Twinkle protein and DNA polymerases Pol1A and Pol1B. *BMC Plant Biol* 19:241
- Morley SA, Niaz A, Brent LN (2019b) Plant organelle genome replication. *Plants* 8:358
- Mower JP, Vickrey TL (2018) Chapter nine—structural diversity among plastid genomes of land plants. *Adv Bot Res* 85:263–292
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete

- sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572–574
- Oldenburg DJ, Bendich AJ (2015) DNA maintenance in plastids and mitochondria of plants. *Front Plant Sci* 6:1–15
- Pfannschmidt T, Blanvillain R, Merendino L, Courtois F, Chevalier F, Liebers M, Grübler B, Hommel E, Lerbs-Mache S (2015) Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J Exp Bot* 66:6957–6973
- Pietrowska-Borek M, Czekaka Ł, Belchí-Navarro S, Pedreño MA, Guranowski A (2014) Diadenosine triphosphate is a novel factor which in combination with cyclodextrins synergistically enhances the biosynthesis of trans-resveratrol in *Vitis vinifera* cv. Monastrell suspension cultured cells. *Plant Physiol Biochem* 84:271–276
- Pulido P, Zagari N, Manavski N, Gawronski P, Metthes A, Schaff BL, Meurer J, Leister D (2018) Chloroplast ribosome associated supports translation under stress and interacts with the ribosomal 30S subunit. *Plant Physiol* 177:1539–1554
- Sakurai I, Stazic D, Eisenhut M, Vuorio E, Steglich C, Hess WR, Aro EM (2012) Positive regulation of psbA gene expression by cis-encoded antisense RNAs in *Synechocystis* sp. PCC 6803. *Plant Physiol* 160:1000–1010
- Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6:283–290
- Schaller F, Schaller A, Stinzi A (2005) Biosynthesis and metabolism of Jasmonates. *J Plant Growth Regul* 23:179–199
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* 94:275–288
- Shikanai T, Fujii S (2013) Function of PPR proteins in plastid gene expression. *RNA Biol* 9:1446–1456
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsumayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049
- Sun Y, Zerges W (2015) Translational regulation in chloroplasts for development and homeostasis. *Biochim Biophys Acta* 9:809–820
- Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A (2013) RNA editing in plants and its evolution. *Ann Rev Genet* 47:335–352
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Gen* 5:123–135
- Vidi PA, Kanwisher M, Baginsky S, Austin JR, Csucs G, Dormann P, Kessler F, Brehelin C (2006) Tocopherol cyclase (VTE 1) localization and vitamin E accumulation in chloroplast plastoglobule lipoprotein particles. *J Biol Chem* 281:11225–11234
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *PNAS USA* 91:9794–9798
- Walter B, Pieta T, Schünemann D (2015) *Arabidopsis thaliana* mutants lacking cpFtsY or cpSRP54 exhibit different defects in photosystem II repair. *Front Plant Sci* 6:250
- Watson SJ, Sowden RG, Jarvis P (2018) Abiotic stress-induced chloroplast proteome remodeling: a mechanistic overview. *J Exp Bot* 69:2773–2781
- Williams-Carrier R, Brewster C, Belcher ES, Rojas M, Chotewutmontri P, Ljungdahl S, Barkan A (2019) The *Arabidopsis* pentatricopeptide repeat protein LPE1 and its maize ortholog are required for translation of the chloroplast psbJ RNA. *The Plant J* 99:56–66
- Wilson PB, Estavillo GM, Field KJ, Pornisiwong W, Carrol AJ, Howel NS, Lake JA, Smith SM, Harvey SM, Harvey Millar A, von Caemmerer S, Pogson BJ (2009) The nucleotidase/phosphatase SAL1 is a negative regulator of drought tolerance in *Arabidopsis*. *Plant J* 58:299–317
- Woodson JD, Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Gen* 9:383–395
- Yagi Y, Shiina T (2014) Recent advances in the study of chloroplast gene expression and its evolution. *Front Plant Sci* 5:1–7
- Zoschke R, Bock R (2018) Chloroplast translation: structural and functional organization, operational control, and regulation. *Plant Cell* 30:45–770
- Zoschke R, Liere K, Börner T (2007) From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *Plant J* 50:710–722

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.