

The evolution of the plastid chromosome in land plants: gene content, gene order, gene function

Susann Wicke · Gerald M. Schneeweiss ·
Claude W. dePamphilis · Kai F. Müller ·
Dietmar Quandt

Received: 25 July 2010 / Accepted: 19 February 2011 / Published online: 22 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract This review bridges functional and evolutionary aspects of plastid chromosome architecture in land plants and their putative ancestors. We provide an overview on the structure and composition of the plastid genome of land plants as well as the functions of its genes in an explicit phylogenetic and evolutionary context. We will discuss the architecture of land plant plastid chromosomes, including gene content and synteny across land plants. Moreover, we will explore the functions and roles of plastid encoded genes in metabolism and their evolutionary importance regarding gene retention and conservation. We suggest that the slow mode at which the plastome typically evolves is likely to be influenced by a combination of different molecular mechanisms. These include the organization of plastid genes in operons, the usually uniparental

mode of plastid inheritance, the activity of highly effective repair mechanisms as well as the rarity of plastid fusion. Nevertheless, structurally rearranged plastomes can be found in several unrelated lineages (e.g. ferns, Pinaceae, multiple angiosperm families). Rearrangements and gene losses seem to correlate with an unusual mode of plastid transmission, abundance of repeats, or a heterotrophic lifestyle (parasites or myco-heterotrophs). While only a few functional gene gains and more frequent gene losses have been inferred for land plants, the plastid *Ndh* complex is one example of multiple independent gene losses and will be discussed in detail. Patterns of *ndh*-gene loss and functional analyses indicate that these losses are usually found in plant groups with a certain degree of heterotrophy, might rendering plastid encoded *Ndh1* subunits dispensable.

S. Wicke (✉) · G. M. Schneeweiss
Department of Biogeography and Botanical Garden, University
of Vienna, Rennweg 14, 1030 Vienna, Austria
e-mail: susann.wicke@uni-muenster.de

Present Address:

S. Wicke
Institute for Evolution and Biodiversity, University of Muenster,
Huefferstr. 1, 48149 Muenster, Germany

C. W. dePamphilis
Department of Biology and Institute of Molecular Evolutionary
Genetics, Pennsylvania State University, 403 Life Sciences
Building, University Park, Pennsylvania, PA 16802, USA

K. F. Müller
Institute for Evolution and Biodiversity, University of Muenster,
Huefferstr. 1, 48149 Muenster, Germany

D. Quandt
Nees Institute for Biodiversity of Plants, University of Bonn,
Meckenheimer Allee 170, 53115 Bonn, Germany

Keywords Plastid genome · Land plants · Genome
evolution · Plastid gene function · Gene retention

Abbreviations

gII(A)	Group II(A) intron
IR	Inverted repeat
LSC	Large single copy region
NEP	Nuclear encoded polymerase
ORF	Open reading frame
Ori	Origin of replication
PEP	Plastid encoded polymerase
PSI	Photosystem I
PSII	Photosystem II
PSRSP	Nuclear-encoded plastid ribosome specific proteins
SC	Single copy
SDR	Small dispersed repeat
SSC	Small single copy region

Introduction

Plastids are one of the main distinguishing characteristics of the plant cell. The central function of the plastid is to carry out photosynthesis, but other major cellular functions also take place in plastids, including synthesis of starch, fatty acids, pigments and amino acids (reviewed by Neuhaus and Emes 2010). As early as 1905, Konstantin S. Mereschkowski hypothesized that plant “chromatophores” are the result of the uptake of a cyanobacterium by a eukaryotic organism (English translation available by Martin and Kowallik 1999). It is now generally accepted that the plastid originated via incorporation of a free-living cyanobacterial-like prokaryote into a eukaryotic cell (primary endosymbiosis), thereby enabling the transition from heterotrophy to autotrophy by gaining the ability of utilizing photoenergy. Recent phylogenetic analyses of plastid genes from major plant lineages have converged on the hypothesis that plastids of the plant kingdom, i.e. the clade including Glaucophytes, Rhodophytes, Chlorophytes, and Streptophytes (Fig. 1; Keeling 2004), are derived from a single origin (Palmer 2000; McFadden and van Dooren 2004; Keeling 2010). This is also supported by several

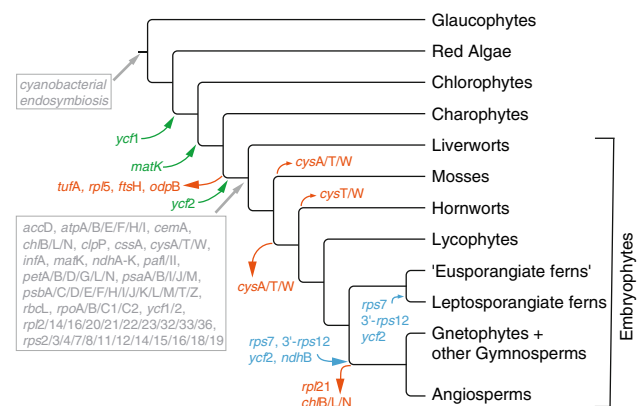


Fig. 1 Evolution of plastid gene content in land plants. Events of gene losses in Embryophytes, as well as gains and duplication of protein coding genes in green plant lineages are depicted along the branches/nodes of the Plant Tree of Life (Palmer et al. 2004; Qiu et al. 2006; Zhong et al. 2010). The putatively ancestral gene content, as reflected in *Marchantia* and derived from parsimony analysis after Maul et al. (2002), is given at the first land plant node. Gene losses during the evolution of land plants are indicated by red arrows (those occurring before the emergence of Embryophytes are not considered here); a green arrow indicates the evolution of a novel gene prior to the transition to land; blue arrows refer to gene duplications. Changes in the content of transfer RNAs are not considered here (refer to Gao et al. 2010 for review). A detailed summary of gene losses during the evolution of angiosperms is provided by Jansen et al. (2007) and Magee et al. (2010). Although *chl*-subunits are still present in some gymnosperm plastomes, multiple losses and pseudogenizations indicate a functional transfer to the nuclear genome. As *chl* genes have been lost entirely from angiosperm plastomes, functional *chl*-gene transfer might have already occurred in a common ancestor

biochemical features, such as the composition of light harvesting complexes and their components, structural RNAs, membrane structure, and the protein import/targeting machinery (Weeden 1981; Bölder et al. 1998; Keeling 2004; Yang and Cheng 2004; Koziol et al. 2007; Vesteg et al. 2009).

Over evolutionary time, genetic information was functionally or more often non-functionally transferred from the endosymbiont's genetic system to the host nuclear genome, genetically intertwining the two genomes. Except for genes involved in photometabolic processes, most other genes have been incorporated into the nuclear genome. This has resulted in a highly reduced plastid genome in Streptophytes (land plants plus their closest algal relatives), comprising less than 5–10% of the genes hypothesized for the ancestral cyanobacterial genome (ca. 2000 to 3000 genes; Martin et al. 2002). A corollary of this process is that the plastid genome (plastome) became subjected to nuclear regulation (Timmis et al. 2004), locking in their symbiotic relationship. The transfer of sequences and both functional and non-functional genes from the plastid genome to both the nuclear and the mitochondrial genome remains an ongoing process (Stern and Lonsdale 1982; Stern and Astwood 1986; Nakazono and Hira 1993; Albus et al. 2010, 1998; Shahmuradov et al. 2003; Matsuo et al. 2005; Guo et al. 2008; Sheppard and Timmis 2009). This intracellular gene transfer is considered “frequent and [to occur] in big chunks” (Martin 2003:1; Stegemann et al. 2003; Noutsos et al. 2005). The question of how many genes can eventually be transferred to the nuclear genome (and whether the plastome could eventually be lost) has been discussed for some time (Barbrook et al. 2006). Massive gene loss has been observed in several parasitic plants (e.g. *Orobanchaceae*: Wolfe et al. 1992; *Cuscuta*: Funk et al. 2007, McNeal et al. 2007). In these plants, gene loss is not restricted to genes that are primarily involved in photosynthesis and related pathways (Wolfe et al. 1992; Krause 2008); additional losses or pseudogenization is seen in genes encoding subunits of the genetic apparatus (e.g., plastid-encoded RNA polymerase, some tRNAs, some ribosomal proteins; dePamphilis and Palmer 1990; Wolfe et al. 1992; Lohan and Wolfe 1998).

Four decades of genetic, genomic and physiological research have contributed substantially to assign genes and gene functions to land plant plastid encoded proteins. Plastid genes have been grouped into functionally defined classes, including (i) those involved in primary and secondary photosynthesis pathways (photosynthetic light and dark reactions), (ii) genes not involved in photosynthetic pathways, such as sulfate transport and lipid acid synthesis, (iii) genes involved in transcription and translation, and (iv) a number of structural RNA genes (Palmer 1991; Sugiura 1992; Bock 2007). Subsequent studies have identified the

roles of additional genes not falling into any of these genes classes, including genes involved in post-transcriptional modification (*matK*, Liere and Link 1995), protein turnover or protein complex assemblies (Peltier et al. 2004). Currently, only two genes remain, *ycf1* and *ycf2*, whose metabolic or genetic roles have not yet been unambiguously defined (Bock 2007).

In this review, we will discuss functional and evolutionary insights from research on land plant plastid chromosomes, providing a synthesis of our knowledge of their evolution and conservation. Accordingly, particular emphasis will be placed on genetics of plastomes in the context of land plant diversification, with special attention to the roles of plastid-encoded proteins in photosynthesis and other principal genetic pathways.

Plastid genetics and synteny of land plant plastid chromosomes

Plastid inheritance

The transmission (inheritance) of plastids has been disputed for many years. For seed plants, mechanisms and occurrences of plastid inheritance have been studied in a great number of species (reviewed in Hagemann 2004; Bock 2007; Zhang and Sodmergen 2010). However, little is known about plastid transmission in earlier land plant lineages, probably due to methodological difficulties. Ultrastructural studies of functional sperm cells of bryophytes, lycophytes, horsetails and water ferns (heterosporous ferns) reported the presence of proplastids (reviewed in Sears 1980). In liverworts and mosses, the sperm cell's proplastids are "discarded" before fertilization (Sears 1980, and references therein). Maternal plastid transmission was subsequently demonstrated for the liverwort *Pellia* (Pacak and Szwejkowska-Kulińska 2002) and several moss representatives (*Rhizomnium*: Jankowiak et al. 2005; *Sphagnum*: Natcheva and Cronberg 2007; *Plagiomnium*: Jankowiak-Siuda et al. 2008). Maternal inheritance of plastids was shown for the horsetail *Equisetum variegatum* (Guillon and Raquin 2000), but nothing is known about the fate of the sperm cell's proplastid. Most, though probably not all, plastid-like structures are lost from the spermatozooids of lycophytes, and it seems as if there was a strong bias towards predominantly maternal plastid transmission caused by degradation prior or immediately after fertilization (Sears 1980). The absence of a plastid-like structure in sperm cells was shown in representatives of leptosporangiate ferns (*Pteridium*: Bell et al. 1966; *Thelypteris*: Sears 1980). This suggested maternal plastid transmission, which was later confirmed using molecular biological methods for *Cheilanthes* (Gastony and Yatskiyevych 1992)

and *Asplenium* (Vogel et al. 1998). In gymnosperms and angiosperms, uniparental inheritance is more frequent than biparental transmission (Hagemann 2004). Maternal inheritance is typical for angiosperms and the gymnosperm groups cycads and gnetophytes. In the majority of gymnosperms (conifers) paternal transmission is the dominant mode (Hagemann 2004; Zhang and Sodmergen 2010). However, biparental inheritance has evolved multiple times in seed plants, in particular in eudicot angiosperms such as Geraniaceae (e.g. Tilney-Bassett and Almouslem 1989), Campanulaceae (Corriveau and Coleman 1988) and Fabaceae (Corriveau and Coleman 1988). In gymnosperms, biparental inheritance is much less frequent (Hagemann 2004).

Architecture of plastid chromosomes

In vivo structure and molecular conformation of the plastid chromosome has long been thought to be exclusively circular. However, several studies employing in situ hybridization techniques demonstrated that often only a minor proportion of the molecules occur in a circular and covalently closed form. Instead, the majority of plastid chromosomes are arranged in concatemers of two or more molecules in either circularized or linear form (Deng et al. 1989; Bendich and Smith 1990; Bendich 1991, 2004; Harada et al. 1997; Lilly et al. 2001). It is still unknown how these concatemeric molecules are formed, and how linkage and breakage is carried out in vivo. It is speculated that the formation of these supermolecules might facilitate maintenance of gene organization and genome integrity (Day and Madesis 2007; Maréchal and Brisson 2010). However, the formation of supermolecules as a primary stabilizing factor needs to be evaluated carefully. Mitochondrial DNA forms concatemeric molecules as well, but exhibits a great variety of genome size and structure among land plants (Palmer and Herbon 1988; Bendich 2007).

The size of photosynthetic land plant plastid chromosomes ranges from 120 kb to 160 kb. The plastome in photosynthetic plants comprises 70 (gymnosperms) to 88 (liverworts) protein coding genes and 33 (most eudicots) to 35 (liverworts) structural RNA genes (Wakasugi et al. 1994; Ohyama 1996; Bock 2007), totaling 100–120 unique genes (Fig. 1). The vast majority of these genes are arranged in operons (or operon-like structures) and transcribed as polycistronic precursor molecules that are subjected to splicing and nucleolytic cleavage in order to produce mature and translatable mRNAs (Stern et al. 2010). Functional gene classes (translation/transcription, electron transfer, and photosystems) are often arranged in close vicinity to one another (Fig. 2; Cui et al. 2006). Using a parametric bootstrap-approach, Cui et al. (2006) showed that the genomic rearrangements of some

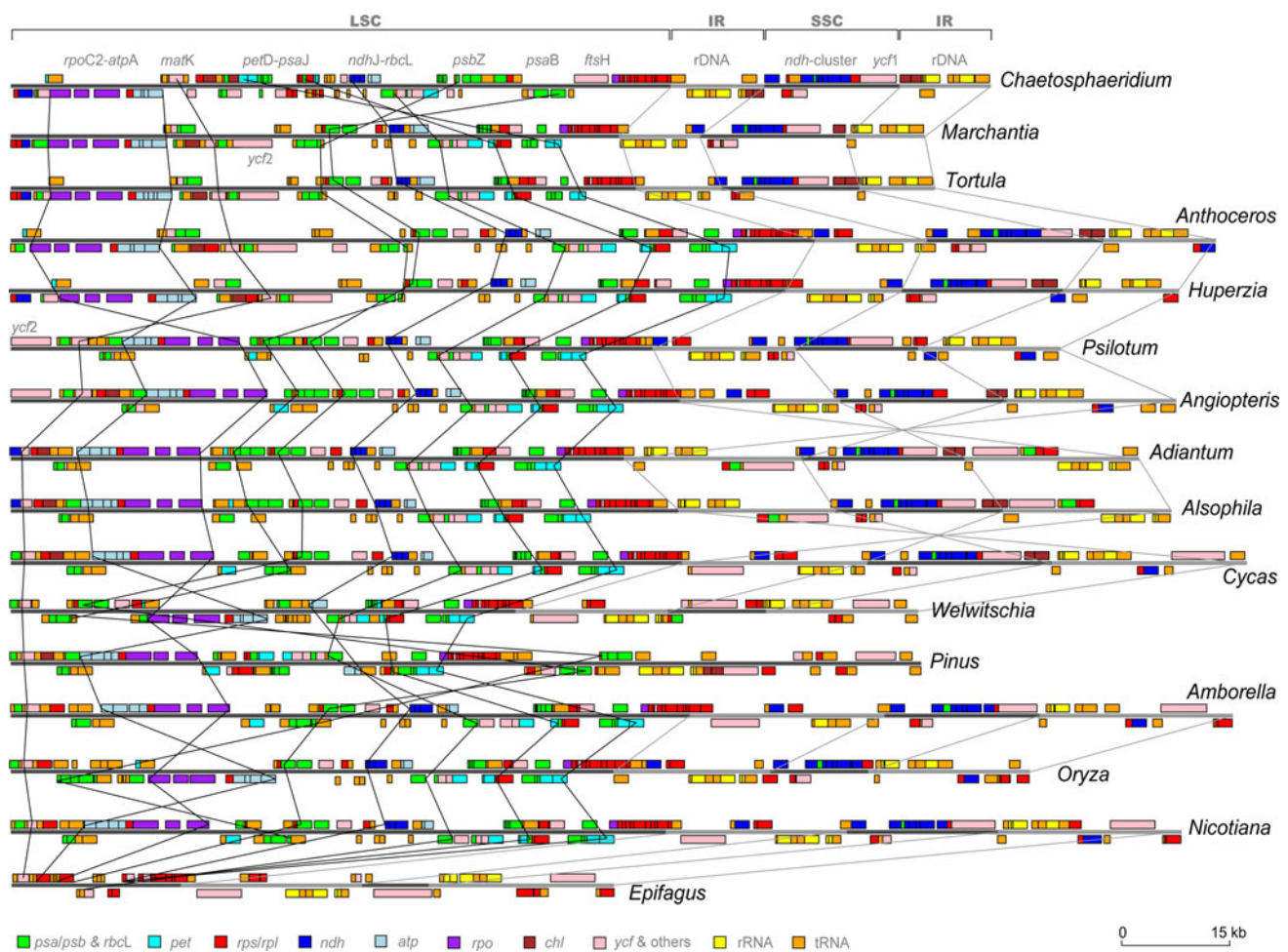


Fig. 2 Synteny of land plant plastid chromosomes. The plastid chromosomes are shown in linearized form illustrating relative gene synteny. Genes are depicted by boxes colored according to their relevant functional class (see legend). Genes encoded by the leading strand (+ strand) or by the lagging strand (- strand) are shown above or below the grey chromosome bar, respectively. Lengths of boxes do not reflect lengths of genes, but are artificially increased to aid legibility (consequently, overlapping genes on \pm strand do not indicate overlapping reading frames). Lines from selected gene/gene-regions mentioned above the first chromosome bar roughly

indicate genes clusters that have been reorganized during land plant evolution. Not all regions that underwent genomic relocations prior or during land plant evolution are depicted here. The chromosome bars are colored gray to highlight the positions of the two large Inverted Repeat regions (IR_A/IR_B) and are connected by gray lines between the different lineages. Gray lines are discontinued once to indicate loss of the large inverted repeat in *Pinus*. Drawn with *GenomePixelizer* (Kozik et al. 2002) using genome annotations deposited in public sequence databases. Refer to the text for genome references and original publications.]

chlorophytic algae (e.g. *Chlamydomonas*) relative to others are not random. Results indicated that the physical clustering of genes belonging to a similar functional class is positively selected. Furthermore, expression analysis indicated that some of these newly formed cluster are co-transcribed which led the authors to speculate that these could represent new regulons (Cui et al. 2006).

The plastid chromosome displays a quadripartite structure, i.e. it is divided into four major segments (Fig. 2). Two of those contain only single copy (SC) genes and are referred to as Single Copy regions. The Large Single Copy region (LSC) harbors the majority of plastid genes; its smaller counterpart is known as the Small Single Copy

region (SSC). The third segment is duplicated and exists in two nearly identical copies separating the SC regions (Kolodner and Tewari 1979). These copies are inverted and, therefore, termed large Inverted Repeats A and B (IR_A, IR_B). An IR is between 20 and 30 kb in size in angiosperms compared to only 10–15 kb in most non-seed plant lineages (Kolodner and Tewari 1979; Palmer 1991; Raubeson and Jansen 2005; Wu et al. 2009; Wolf et al. 2010a). However, several lineages deviate strongly from the average, such as *Cycas* (25 kb, Wu et al. 2007), the cypress *Cryptomeria* (114 kb, Hirao et al. 2008) or the eudicot Geraniaceae (*Monsonia*: 7 kb, Guisinger et al. 2010; *Pelargonium*: 76 kb, Chumley et al. 2006). As the

IRs are essentially identical, one might describe the plastid genome structure also as tripartite (as in Bock 2007), since the IRs share molecular evolutionary patterns that clearly differ from those observed in the SC regions. This quadripartite (or tripartite) architecture is already present in algal lineages including the closest relatives of land plants (e.g. *Chaetosphaeridium*, *Chara*; Turmel et al. 2002, 2006), implying a pre-land plant origin for this important conserved structural feature.

The plastid chromosomes of charophyte algae, the closest relatives of land plants (Qiu et al. 2006), are larger than those of land plants. They contain several genes that have either been lost or functionally transferred to the nuclear genome in Embryophytes (Turmel et al. 1999; 2006). Parsimony analyses reconstructing unambiguous changes in gene content among plants revealed that the gene *ycf1* was gained in a common ancestor of several green algae and land plants (Maul et al. 2002). The gain of an intron in the *trnK_{UUU}* coding regions, including an intact open reading frame (ORF; *matK*), is shared by Charophytes and Embryophytes (Maul et al. 2002; Lewis and McCourt 2004; McNeal et al. 2009). Comparative analysis revealed that the plastome structure and gene content in *Chaetosphaeridium*, a unicellular freshwater charophyte alga, is most similar to that of early land plants (Turmel et al. 2002): Large blocks of co-linear groups of genes are already present in this genus. Yet, in order to obtain the structural organization of early land plant plastomes, several functional gene transfers to the nuclear genome (e.g. *tufA*, *ftsH*, *odpB*, *rpl5*), one gene gain (*ycf2*), and a minimum of eight inversions are necessary (Turmel et al. 2006; Gao et al. 2010). One of those inversions involves a region of the LSC approximately 30 kb in length (Raubeson and Jansen 1992). A huge inversion of the complete *matK—atpA-I—rpoB-C1/2*-region is shared between ferns and seed plants (Fig. 2), whereas liverworts (Ohyama et al. 1988; Wickett et al. 2008a), mosses (Sugiura et al. 2003; Oliver et al. 2010), hornworts (Kugita et al. 2003), and lycophytes (Wolf et al. 2005; Tsuji et al. 2007; Karol et al. 2010) show a more ancestral organization similar to that of *Chaetosphaeridium* (Quandt et al. 2003; Turmel et al. 2002). Generally, the presence of such rearrangements implies that additional transitional forms probably existed and might still be observable in lineages that have remained unstudied so far.

Synteny and structural rearrangements

Plastome rearrangements

Hotspots for structural rearrangements within plastid genomes include the IRs, which are frequently subject to expansion, contraction or even complete loss. Such

changes occurred several times independently during the evolution of land plants and often are specific for single orders and families, sometimes even for just one or a few species within a genus (Downie and Bewley 1992; Goulding et al. 1996; Plunkett and Downie 2000; Daniell et al. 2006; Guisinger et al. 2010; Wolf et al. 2010a). Furthermore, extensive changes within the IRs appear to have an effect on the structural integrity of the entire plastid chromosome beyond the IRs and their immediate neighborhood. This is likely due to their role as putatively important players in the stabilization of the plastid chromosome via homologous recombination-induced repair mechanisms (Maréchal et al. 2009; Rowan et al. 2010; reviewed in detail by Maréchal and Brisson 2010).

Early branching gymnosperms (McCoy et al. 2008; Wu et al. 2009), angiosperms (Goremykin et al. 2003; Cai et al. 2006) and derived leptosporangiate ferns possess much larger IRs than the remaining land plant lineages (Wakasugi et al. 1998; Roper et al. 2007; Karol et al. 2010). Thus, large scale expansions of the IRs most likely occurred at least twice independently over the evolution of major land plant groups, including once in the common ancestor of seed plants. Additional large- (Guisinger et al. 2010) and small-scale (Goulding et al. 1996) expansions have occurred within angiosperms. As a result of the re-location into the IR, several previously SC genes became duplicated, including the largest plastid gene, *ycf2* (Wolf et al. 2010a). A duplication of the *ycf2* gene occurs independently in derived leptosporangiate ferns (tree and polypod ferns) and might be functionally relevant for plant development. In angiosperms, *ycf2* expression is highest in fruits (Drescher et al. 2000), but comparable data for leptosporangiate ferns (or other land plant lineages) are lacking so far. Interestingly, plastome re-structuring in ferns is correlated with an expansion of the IR (Thompson et al. 1986; Stein et al. 1992; Raubeson and Stein 1995; Wolf et al. 2010a).

Contraction of the large inverted repeats involves only few (tens to hundreds of) base pairs up to and including complete IR loss. The positions of the LSC-IR junctions vary slightly within groups, but usually this has only negligible effects on plastome size (Goulding et al. 1996; Daniell et al. 2006; Wang et al. 2008). It has been suggested that such positional changes of IR-junctions among species are the result of gene conversion (Goulding et al. 1996). In several groups, one of the IR-region has been completely lost, for instance in several legumes (Palmer et al. 1987b; Cai et al. 2008; Jansen et al. 2008; Tangphatsornruang et al. 2010), members of Geraniaceae (Guisinger et al. 2010), and some representatives of Orobanchaceae (Downie and Bewley 1992; S. Wicke, C. W. dePamphilis, D. Quandt and G. M. Schneeweiss, unpublished data). So far, no properties have been

identified that are shared between these rather distantly related angiosperms and might provide an explanation for these IR losses. In legumes, the loss apparently affects overall structural stability, leading to mutational hotspots (Palmer et al. 1987b; Milligan et al. 1989; Cai et al. 2008; Magee et al. 2010) and an overall increase of nucleotide substitution rates (Perry and Wolfe 2002). The changes in gene order of a *Vigna angularis* cultivar relative to other members of Fabaceae have been proposed to either be caused by a large inversion or mediated by a two-step model including IR expansion and contraction (Perry et al. 2002).

Small dispersed repeats

Reorganizations are in many cases associated with small dispersed repeats (SDR), which are hypothesized to contribute to the double-strand break induced repair mechanism (Milligan et al. 1989; Maul et al. 2002; Odom et al. 2008). SDRs often contribute significantly to repeat space in genomes with highly rearranged gene order and add to structural polymorphism in even closely related lineages (Maul et al. 2002). SDRs mainly occur in non-coding DNA fractions (spacer, introns; Raubeson et al. 2007), where they are often associated with small hairpin structures (Quandt et al. 2003; Kim and Lee 2005). The greatest concentrations of SDRs have so far been reported in green algal plastid genomes (ca. 20% of the *Chlamydomonas* plastome), although this seems to be highly lineage specific (Maul et al. 2002). Large repeats are assumed to be suppressed (or selectively eliminated) in plastid DNA because of their ability to cause recombination that may destabilize genome structure (Gray et al. 2009; Maréchal and Brisson 2010). Among angiosperms, the most abundant sizes of SDRs are on average smaller than 50 bp with direct repeats being more frequent than inverted repeats (Raubeson et al. 2007). A significant increase of repeats larger than the average has been reported in highly rearranged genomes such as Geraniaceae (Guisinger et al. 2010), Campanulaceae (Haberle et al. 2008), and Fabaceae (Cai et al. 2008), supporting the notion that repeats and genomic rearrangement are causally related. Possibly, tRNA genes might be recognized as repeated elements causing rearrangements by intramolecular or non-homologous recombination (Ogihara et al. 1988; Hiratsuka et al. 1989). In many cases, breakpoints of inversions are flanked by tRNA genes and short repetitive sequences (Hiratsuka et al. 1989; Haberle et al. 2008; Guisinger et al. 2010).

A unique switch in IR orientation (inversion) has occurred along the branch separating early diverging fern lineages (*Psilotum*, *Angiopteris*: Wakasugi et al. 1998; Roper et al. 2007; Karol et al. 2010) from derived leptosporangiate ferns (*Adiantum*, *Alsophila*: Wolf et al. 2003;

Gao et al. 2009). This might be an outcome of the flip-flop recombination process proposed by Palmer (1983). Two smaller rearrangements occur at the breakpoint of the large inversion that is synapomorphic to all vascular plants except lycophytes (Raubeson and Jansen 1992; Wolf et al. 2003). The inversions reported in derived leptosporangiates are likely to be caused by two overlapping inversions during the evolution of leptosporangiate ferns (Wolf et al. 2003, 2010).

Several small and large inversions that are not accompanied by expansion and contraction of an IR have been reported for diverse angiosperm lineages (Asteraceae: Jansen and Palmer 1987; Kim et al. 2005; *Spinacia*: Schmitz-Linneweber et al. 2001; some Oleaceae: Lee et al. 2007; Mariotti et al. 2010; grasses: Hiratsuka et al. 1989; Bortiri et al. 2008), but seem to be less frequent in early land plants lineages. However, one large inversion (71 kb), affecting nearly the entire LSC, is found in the model moss *Physcomitrella patens* (Sugiura et al. 2003). This inversion was shown to be autapomorphic to *Physcomitrella* and Funariales, but absent in other mosses (Goffinet et al. 2007). Due to the small number of plastid genomes sequenced from early land plant lineages, little is known about other structural rearrangements in bryophytes. As of this writing, no structural changes (inversions) have been identified in liverworts (L. L. Forrest and B. Goffinet, Ecology and Evolutionary Biology, University of Connecticut/USA, personal communication). Some of the largest inversions observed may be attributable to flip-flop recombination due to the existence of the large inverted repeats (Palmer 1983). In the flowering plants studied so far, it has been shown that flip-flop recombination and inversions predominantly occur around the origin of replication (*ori*). In some angiosperms, the *ori_B* maps to the rDNA-*ycf1* region within the IR, which is located more closely to the IR-SSC-boundary than to the IR-LSC junction (Thompson et al. 1986; Lu et al. 1996; Kunnimalaiyaan and Nielsen 1997; Eisen et al. 2000; Mackiewicz et al. 2001).

Genome size reduction, gene transfer, and gene gains

Genome size reduction is another major aspect of non-canonical structural evolution. The most dramatic changes in genome size and gene content have been reported for non-photosynthetic parasitic plants. The plastome of *Epifagus* (Wolfe et al. 1992) measures only about half the size of an average eudicot plastome (Bock 2007). This is mainly due to non-functionalization of most photosynthesis-related genes (dePamphilis and Palmer 1990) and some genes for transcription and translation (Morden et al. 1991). Although there is a general trend of (functional) plastid genome reduction in parasitic plants, the size and

gene content seem to vary widely among different lineages because some highly heterotrophic species retain photosynthetic ability (Revill et al. 2005; Funk et al. 2007; McNeal et al. 2007; Nickrent and García 2009). Independent of parasitism, genome reduction was observed in Pinaceae and Gnetophytes (McCoy et al. 2008; Wu et al. 2009), due in large part to the loss of *ndh* genes. The plastomes of *Gnetum* and *Welwitschia* are also more compact than in other seed plant lineages due to the reduction of intron and spacer regions (McCoy et al. 2008; Wu et al. 2009). This genome reduction is speculated to be the result of a low-cost strategy that could facilitate rapid genome replication under disadvantageous environmental conditions (McCoy et al. 2008; Wu et al. 2009).

Translocation of single genes is rare in plastid genomes, and this is likely a reflection of the overall rarity of inserted (vs. lost or rearranged) sequences in plastid genomes. Reports of foreign DNA being naturally inserted into the plastid DNA are rare (Maul et al. 2002; Haberle et al. 2008; Guisinger et al. 2010); perhaps in part because of the difficulty of detecting insertions in poorly conserved intergenic regions. Many of the repetitive elements found in highly rearranged genomes seem to be derived from plastid sequences (Cai et al. 2008; Haberle et al. 2008; Guisinger et al. 2010). However, some are unique which might suggest either rapid divergence or a non-plastid origin (Guisinger et al. 2010). As already mentioned by Park et al. (2007), the putatively horizontally acquired *rbcL* gene copies found in several *Phelipanche* species (Orobanchaceae) are most likely located in the nuclear or mitochondrial genome, and are not plastid encoded. *RbcL* appears to be generally absent from *Phelipanche* plastid genomes (S. Wicke, D. Quandt, C. W. dePamphilis, G. M. Schneeweiss, unpublished data).

Gene gains, too, are exceptional during plant evolution (e.g. *matK*, *ycf1/2*; Fig 1). The organization and regulation of genes in operons might be one stabilizing factor. Most often, localized changes of gene order are caused by the loss of single genes to the nuclear genome, or due to non-functionalization in parasitic or mycotrophic plants.

Functional transfer of genes and subsequent loss of the plastid gene copy has been reported for some rosids (Jansen et al. 2010), some monocots (e.g. Hiratsuka et al. 1989; Masood et al. 2004; Sasaki et al. 2007) and the spikemoss *Selaginella uncinata* (Tsuji et al. 2007).

Contrasting with the overall high degree of conservation of plastome structure and gene content in land plants, massive structural changes are occasionally found in several unrelated lineages. These include derived angiosperm families such as Geraniaceae (Palmer et al. 1987a; Chumley et al. 2006; Guisinger et al. 2010), Fabaceae (Palmer et al. 1987b); Milligan et al. 1989; Cai et al. 2008; Tangphatsornruang et al. 2010), members of Onagraceae

(*Oenothera*: Hupfer et al. 2000; Greiner et al. 2008), Campanulaceae (Knox and Palmer 1999; Cosner et al. 1997, Cosner et al. 2004; Haberle et al. 2008), but also leptosporangiate ferns (Wolf et al. 2003, 2010; Gao et al. 2009). Because some of the extensively re-shuffled angiosperm plastomes occur in lineages with biparental plastid inheritance (Corriveau and Coleman 1988), it is tempting to speculate that the nature of plastid inheritance may affect plastid genome stability. Biparental inheritance combined with fusion of paternal and maternal plastids (although rare; Wellburn and Wellburn 1979) would likely result in homologous recombination between putatively divergent plastome copies (experimentally shown by Fejes et al. 1990), eventually leading to alteration of the genome structure. In other plants, major rearrangements, in particular gene losses, are obviously connected to a change in lifestyle from autotrophy to parasitism or myco-heterotrophy (*Aneura*: Wickett et al. 2008a; Orobanchaceae: dePamphilis and Palmer 1990; Wolfe et al. 1992; Convolvulaceae: Funk et al. 2007; McNeal et al. 2007, 2009; Viscaceae: Nickrent and García 2009; and Lennoaceae: Y. Zhang and C.W. dePamphilis, unpublished data).

The precise mechanisms underlying structural changes are as yet unknown, but they are often associated with the presence of nearby repeat sequences, including small repeated sequences that are dispersed through the genome (Maul et al. 2002; Cui et al. 2006; Omar et al. 2008; Cai et al. 2008; Gray et al. 2009; Maréchal and Brisson 2010). Similarly to the plastid genome, in both the nuclear and mitochondrial genomes, structural reorganizations often are observed in proximity to structural RNA genes and short repetitive flanking sequence motifs (Grewe et al. 2009). In the nuclear genome, the latter is often associated with transposon activity (Woodhouse et al. 2010). In mitochondrial genomes, transposons are restricted to angiosperms (Knoop et al. 1996; Kubo et al. 2000; Notsu et al. 2002), but are absent in early land plant lineages (Ohya 1996; Knoop 2004; Grewe et al. 2009). No (retro-) transposons, or traces thereof, have ever been reported from land plant plastomes. Yet, the plastid chromosome of the model green algae *Chlamydomonas* harbors two copies of the non-functional transposable element *Wendy* (Fan et al. 1995, Maul et al. 2002). Consequently, mechanisms suggested for nuclear and mitochondrial genomes are less likely for plastid genomes given the current knowledge on their evolution (reviewed in Palmer 1991; Raubeson and Jansen 2005; Bock 2007).

Other possible candidates for causing restructuring of plastid genomes are relaxed repair mechanisms and/or recombination processes. Recently, several nuclear encoded genes and gene families have been identified that mediate stabilization, repair and maintenance of the plastid chromosome (Day and Madesis 2007; Maréchal and

Brisson 2010). It might be possible that mutations in these proteins could lead to impaired maintenance of the plastid genome structure (Guisinger et al. 2010).

Gene content and function of the plastid genome

The central function of the chloroplast is to carry out photosynthesis and carbon fixation. Besides genes encoding elements for the genetic apparatus, such as structural and transfer RNAs, the plastome encodes numerous proteins for photometabolic pathways (Palmer 1991; Sugiura 1992; Raubeson and Jansen 2005; Bock 2007). The following functional protein categories can be distinguished (Table 1): proteins for the genetic apparatus, for non-photosynthesis related metabolic pathways, for primary (light-dependent) photosynthetic reactions, and for secondary (light-independent) photosynthesis pathways. In most cases, fully functional protein complexes are assembled from plastid encoded gene products and nuclear encoded subunits that are imported into the plastid organelle.

Plastid encoded elements for the plastid genetic apparatus

Many genes that encode pathways for the plastid genetic apparatus have been transferred to the nucleus and are now imported into the plastid. However, genes for transcription and protein biosynthesis are retained in the plastome. These comprise structural RNAs (rRNA, tRNA), some ribosomal proteins, and genes for a DNA-dependent RNA polymerase as well as few genes coding for DNA and protein processing enzymes.

Genes for DNA/RNA processing enzymes

Plastid genetics is sometimes described as “chimeric” in that eukaryotic cytosolic (e.g. poly-A-binding proteins) and eubacterial components (e.g. Shine-Dalgarno interactions) are combined with novelties such as regulating stem loops in the 5'- and 3'- untranslated regions of plastid mRNAs (Zerges 2000). Transcription of plastid genes is carried out by a set of DNA-dependent RNA polymerases: nuclear encoded (phage-type) polymerase (NEP) and plastid-encoded (eubacterial-type) polymerase (PEP). Both transcribe distinct groups of genes (Hajdukiewicz et al. 1997; Cahoon and Stern 2001; Shiina et al. 2005) and require different transcription promoting signals (Weihe and Börner 1999). Promoter signals of PEP-transcribed genes are highly similar to those of eubacterial σ 70-promoters with AT-rich sequences in the -35 promoter element (consensus 5'-TTGACA-3') and the -10 TATA-box (consensus 5'-TATAAT-3') upstream of the transcription initiation site

(Briat et al. 1986). Promoter elements of NEP-transcribed genes are less conservative and share only short elements (Weihe and Börner 1999). Three different types are known. Two are characterized by a common core promoter YRT-element (i.e. purine-pyrimidine-thymidine stretch) that is highly conserved among flowering plants. This motif is localized in close proximity to the start codon (less than 10 bp away), where it can be preceded by a GAA-box. The different classes of promoters are recognized by two phage type polymerases. In *Arabidopsis*, the existence of at least two plastid targeted NEPs has been experimentally corroborated (Swiatecka-Hagenbruch et al. 2008), but evidence for differential usage or affinity to particular promoters is currently lacking. In eudicots, one of these NEPs is targeted to mitochondria and plastids (Kobayashi et al. 2001), which is reflected in partially shared promoter architectures between both organelles (Kühn et al. 2005). However, this dual-targeted phage type polymerase appears to be absent from other land plants including monocots and early diverging angiosperms (Yin et al. 2010).

PEP is lost or pseudogenized in some parasitic plants with minimal or no photosynthetic activity such as *Cuscuta* (Funk et al. 2007; McNeal et al. 2007) and Orobanchaceae (Wolfe et al. 1992; Delavault et al. 1996). The loss of PEP subunits renders its promoters dispensable, potentially allowing them to be lost from the plastome (Krause et al. 2003). However, NEP seems to be able to take over at least some of PEP's transcriptional functions as suggested by the frequent presence of both NEP and PEP promoters upstream of several plastid transcription units, for instance in the *rrn16-trnV* region (Krause et al. 2003). In both *Cuscuta* (Berg et al. 2004) and *Lathraea* (Lusson et al. 1998) expression of the *rbcL* gene is accomplished by NEP after the loss of PEP.

MatK—a general group IIA intron maturase?

Protein coding genes that are related to (post-) transcriptional activity include the *matK* gene. The *matK*-gene product is thought to act as a splicing factor for plastid group IIA (gIIA) introns (Liere and Link 1995). It is commonly referred to as a ‘general’ maturase associated with several different intron-containing plastid mRNAs (Zoschke et al. 2010). *MatK* is transcribed from the sole intact plastid gII intron ORF localized between the exons coding for the lysine-tRNA (*trnK_{UUU}*). In contrast to other gII ORFs, *MatK* has lost domains assigned to a reverse transcriptase and endonuclease function. Similarity to typical gII ORF maturases is only retained in the DNA-binding domain (Mohr et al. 1993; San Filippo and Lambowitz 2002; Mohr and Lambowitz 2003; Lambowitz and Zimmerly 2004; Pyle and Lambowitz 2006; Hausner

Table 1 Summary of plastid encoded genes in land plants. Genes are divided primarily according to their principal function (light-independent pathways, light-dependent pathways, genetic apparatus), and, secondarily according to the function of their respective subunits in a given protein

Function	Gene class	Subunits	Protein type	Functional/structural remarks	Losses and pseudogenizations
Light independent proteins related to photosynthetic dark reactions	<i>cemA</i> inner membrane protein	–	Integral	Mediates CO ₂ -uptake	Lost from <i>Epifagus</i>
	<i>chl</i> Protochloro-phyllide reductase	B, L, N	–	–	Absent from angiosperm plastomes Lost/pseudogenized from some gymnosperms, e.g. <i>Gnetum</i> , <i>Welwitschia</i>
	<i>ccsA</i> Cytochrome <i>c</i> biogenesis protein	–	–	Mediates heme attachment to <i>c</i> -type cytochromes	Lost from <i>Epifagus</i>
	<i>rbcL</i> large subunit of RuBisCO	–	Cytosolic	Presumably the most abundant protein on earth—primarily involved in photosynthetic carbon fixation Putative additional function in lipid acid metabolism, that is decoupled from photosynthesis	Pseudogenized in <i>Epifagus</i> , <i>Orobanchaceae</i> ^a , <i>Hyobanthe</i> sp. ^b Retained and expressed in <i>Cuscuta</i> ^c and non-photosynthetic and broomrape <i>Lathraea</i> ^d , some species of <i>Harveya</i> ^b , <i>O. corymbosa</i> ⁴ , <i>O. fasciculata</i> ^a ; putatively functional in myco-heterotrophic <i>Aneura mirabilis</i> ^e but expression analysis are currently unavailable
Proteins not related to photosynthesis	<i>accD</i>	–	Cytosolic	Involved in lipid acid synthesis and not related to photosynthesis	Lost from Poales, <i>Acorus</i> , and <i>Trachelium</i> , some Geraniaceae ^f , <i>Passiflora</i>
	<i>cys</i> putative ABC-containing sulfate transporter genes	A, T	– Extrinsic	<i>cysA</i> : contains ATP-binding cassette and nucleotide binding motifs (Walker sites A, B) Involved in sulfate metabolism and not related to photosynthesis	<i>cysA</i> present in the <i>Marchantia</i> ^g and <i>Anthoceros</i> ^h plastome, <i>cysT</i> pseudogenized in <i>Aneura mirabilis</i> ^e All subunits lost in mosses ^{i,j} and all vascular plants

Table 1 continued

Function	Gene class	Subunits	Protein type	Functional/structural remarks	Losses and pseudogenizations
Light-dependent proteins of the photosynthetic light reactions	<i>atp</i> F-type ATP Synthase	A, B, E F, I, H	Extrinsic Integral	Extrinsic domains form catalytic complex F ₁ Integral domain forms proton pumping complex F ₀	Some subunits are lost from <i>Epifagus</i> , some are retained as pseudogenes
	<i>ndh</i> NAD(P)H dehydrogenase complex (<i>Ndh1</i>)	A, B, C, D, E, F, G, H, I, J, K	Integral Extrinsic	Subunits A-D and H-K homologous to mitochondrial complex I Subunits E, F, G are unique to plastid/cyanobacterial <i>Ndh1</i> complex—no homologs to any subunit of complex I	<i>ndh</i> genes are either pseudogenized or lost from <i>Aneura mirabilis</i> ^e , Pinaceae, Gnetales, orchids, <i>Epifagus</i> , some <i>Erodium</i> species, some <i>Cuscuta</i> ^c ; some members of carnivorous Lentibulariaceae ^k
	<i>pcf</i> Photosystem I assembly factor	I (<i>ycf3</i>), II (<i>ycf4</i>)	–	<i>pcfI</i> is indispensable for PSI assembly, whereas <i>pcfII</i> might not be essential	<i>pcfII</i> absent from some legumes; <i>pcf I/II</i> absent from <i>Epifagus</i>
	<i>pet</i> Cytochrome b ₆ /f complex	A, B, D, G, L, N	Extrinsic Integral	A, B, D + ne PetC (Rieske protein) form core complex G,L,N + ne PetM arranged peripherally around the core	Lost from <i>Epifagus</i> Some subunits are pseudogenized in <i>Cuscuta</i> -species ^c , and <i>Aneura mirabilis</i> ^e
	<i>psa</i> Photosystem I	A, B, C, I, J	Integral	Extrinsic subunits are nuclear encoded	Some subunits are either lost from or pseudogenized in <i>Epifagus</i> , some are pseudogenized in <i>Cuscuta</i> -species ^c , and <i>Aneura mirabilis</i> ^e
	<i>psb</i> Photosystem II	A, B, C, D, E, F, H, I, J, K, L, M, N, T, Z	Integral	<i>psbB/C</i> form large extrinsic loops but are intrinsically bound to the membrane Extrinsic subunits <i>psbO</i> , P, Q are nuclear encoded	Some subunits are either lost from or pseudogenized in <i>Epifagus</i> Some subunits are pseudogenized in <i>Cuscuta</i> -species ^c and <i>Aneura mirabilis</i> ^e

Table 1 continued

Function	Gene class	Subunits	Protein type	Functional/structural remarks	Losses and pseudogenizations
Genetic apparatus Proteins for transcription and post-transcriptional modification	<i>mark</i>	–	Cytosolic	Fast evolving gene with nearly equal substitution rates at all codon positions Only plastid gII-ORF localized in the trnK _{UUU} -intron	Lost from some <i>Cuscuta</i> -species ^a along with seven gIIA-introns Loss of the surrounding trnK _{UUU} intron from most leptosporangiate ferns, some lycophytes ¹ <i>rpoA</i> highly diverged in Geraniaceae ^{f,m} , although remaining subunits are intact; Subunits are either lost from or pseudogenized in parasitic plants including <i>Epifagus</i> , and <i>Cuscuta</i> ^c Lost from <i>Passiflora</i> , <i>Scaevola</i>
	<i>rpo</i> DNA-dependent RNA polymerase	A, B, C1, C2	Cytosolic	–	
Proteins for translation and protein-modifying enzymes	<i>clpP</i> Proteolytic subunit of Clp-protease	–	Cytosolic	Contains two gII-introns, however, loss of introns occurs several times independently	
	<i>infA</i> Translation initiation factor A	–	Cytosolic		Pseudogenized in the lycophyte <i>Isoetes</i> ; Pseudogenized/lost in rosids, Solanales, <i>Lemna</i> , <i>Trachelium</i> , <i>Ranunculus</i> , <i>Cuscuta</i> ^c
Large ribosomal proteins	<i>rpl</i>	2, 14, 16, 20, 21, 22, 23, 32, 33, 36	Cytosolic	Primary 23S binding Subunits ⁿ : L2, L16, L20, L22, 23, plus NE. L1, L3, L4, L9, L10, L11, L12, L15, L17, L18, L24, L29) Secondary 23S binding subunits ⁿ : L14, L19, L21 plus NE L5, L6, L13, L27	<i>rpl21</i> lost from all seed plants <i>rpl22</i> lost from Fabaceae <i>rpl23</i> pseudogenized in several angiosperms taxa <i>rpl32</i> lost from <i>Yucca</i> , <i>Populus</i> , some parasitic plants <i>rpl33</i> lost from <i>Phaseolus</i>
	Small ribosomal proteins	2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19	Cytosolic	Binding features unknown: L32, L33, L36, plus ne L31, L34 and 50S-PSRPs Primary 16S binding subunits ⁿ : S4, S7, S8, S15, (ne S13, S20) Secondary 16S binding subunits ⁿ : S11, S12, S16, S18, S19, (ne S6, S9, S13, S5) Tertiary 16S subunits ⁿ : S14, S10, S3, S2, (ne S21) Binding unknown: 30S-PSRPs 3' <i>rps12</i> (exons 2/3) are trans-spliced to 5'- <i>rps12</i> (exon 1)	Some <i>rpl</i> genes are pseudogenized or lost from some parasitic plant plastomes <i>rps7</i> lost from <i>Passiflora</i> <i>rps11</i> lost from <i>Scaevola</i> <i>rps16</i> gene lost multiple times during land plant evolution, multiple intron losses Some <i>rps</i> genes are either pseudogenized or lost from some parasitic plant plastomes

Table 1 continued

Function	Gene class	Subunits	Protein type	Functional/structural remarks	Losses and pseudogenizations
Structural RNAs	<i>rrm</i>	4.5S, 5S, 16S, 23S	–	–	–
	Ribosomal RNAs	Ala ^(UGC) , His ^(GUG) , Lys ^(UUU) , Gln ^(UUG) , Cys ^(GCA) , Asp ^(GUC) , Glu ^(UUC) , Tyr ^(GUA) , fMet ^(CAU) , Phe ^(GAA) , Met ^(CAU) , Trp ^(CCA) , Asn ^(GUU) , Gly ^(UCC, GCC) , Thr ^(GGU, UGU) , Ile ^(GAU, CAU) , Val ^(GAC, UAC) , Pro ^(UGG, GGG) , Arg ^(UCU, ACG, CCG) , Leu ^(UAA, CAA, UAG) , Ser ^(GCU, GGA, UAG)	–	Anticodon of tRNA-Leu ^{UAA} mutated to CAA, RNA-edited to UAA → loss of tRNA-Leu ^{CAA} in some ferns	trn ^{P_{GCG}} : loss is synapomorphic in angiosperms tRNA-Arg ^{CCG} : pseudogenized in gymnosperms plus Gnetales, absent from angiosperm plastomes tRNA-SeC ^{UCA} gene in <i>Adiantum</i> has not yet been found in any other Embryophyte plastome ^o Some tRNAs have been lost from the plastomes of parasitic plants
Proteins of unknown function	<i>ycf1</i>	–	–	–	Lost from Poales, Acorales, <i>Passiflora</i> Highly diverged in some ferns and some members of carnivorous Lentibulariaceae ^k Lost from Poales
	<i>ycf2</i>	–	–	–	Highly diverged in some ferns and some members of carnivorous Lentibulariaceae ^k

Where more than one subunit exists for a given gene class, subunit specifications are in alphanumeric order. The type of encoded protein is indicated as cytosolic, integral or extrinsic. The term 'cytosolic' is used here to describe localization in either the plastid stroma or lumen without associations to membranes. Genomic or genetic features as well as incidents of gene losses are provided for each gene. Refer to the text for further information about the function of certain genes. [Abbreviations: ne—nuclear encoded. References for structural characteristics/gene losses/pseudogenization are (if not stated otherwise): Jansen et al. 2007 and/or Magee et al. 2010—photosynthetic angiosperms; Wolfe et al. (1992)—holoparasitic (non-photosynthetic) *Epifagus*; Wakasugi et al. 1994 and/or Wu et al. 2009—Pinaceae and Gnetales; Wolf et al. 2010a—early vascular plants and ferns.]

^a Wolfe and dePamphilis (1997), Leebens-Mack and dePamphilis (2002); ^bRandle and Wolfe (2005); ^cMcNeal et al. (2007), Funk et al. (2009); ^dDelavault et al. (1996); ^eWickett et al. (2008a, b); ^fGuisinger et al. (2010); ^gOhyama (1996); ^hKugita et al. (2003); ⁱOliver et al. (2010); ^kB. Schäferhoff, S. Wicke, C. W. dePamphilis and K. F. Müller, unpublished data.; ^lTsuji et al. (2007); ^mChumley et al. (2006); ⁿRefers to eubacterial-type ribosomes, see Grondek and Culver 2004 for assembly maps. No assembly maps are currently available for binding of plastid ribosomal proteins to 16S rRNA or 23S rRNA, respectively. PSRP do not harbour homologues in *E. coli* ribosomes. *E. coli* ribosomal proteins L7, L8, L30, L25, and L26 have not been detected in chloroplast ribosomes by Yamaguchi and Subramanian (2000); ^oList may be incomplete due to additional changes/mutations or RNA-editing of anticodon sequences as reported from *Adiantum* by Wolf et al. 2004. A more thorough overview of tRNA changes among land plants is provided by Gao et al. 2010. The set of tRNAs shown here refers to the reference plastome *Nicotiana tabacum* (Shinozaki et al. 1986)

et al. 2006). The molecular evolution of the *matK* coding region is unusual compared to other plastid genes in that all three codon positions evolve at nearly equal rates (Hilu and Liang 1997). This feature makes it particularly useful for phylogenetic reconstruction (Müller et al. 2006; Wicke and Quandt 2009). Equal substitution rates at all codon positions, however, are indicative of relaxed purifying selection (Müller et al. 2006; Duffy et al. 2009), which led several authors to question its function or functionality in land plants (Hausner et al. 2006). Substitution rate analysis, however, demonstrated purifying selection for *matK* in parasitic lineages including Orobanchaceae (Young and dePamphilis 2000) and some *Cuscuta* species (McNeal et al. 2009), providing evidence for sustained functionality. In *Cuscuta*, however, *matK* is absent from species (Funk et al. 2007; McNeal et al. 2007) that have lost all of the seven gIIA introns that likely depend upon the *matK* maturase for splicing (McNeal et al. 2009; Zoschke et al. 2010), which lends further support to the hypothesis of a more general demand for the *matK*-encoded maturase function.

Structural RNAs

Reflecting their localization within the IR region, two sets of structural ribosomal RNA species (*rrn23*, *rrn16*, *rrn5*, *rrn4.5*) are encoded in most plastid genomes of green plants studied so far. The few exceptions with only one set occur in lineages that have lost one copy of the IR. The ancient duplication of the plastid ribosomal DNA operon and its conservation throughout plant evolution might be attributed to generally high quantities of rRNA required for ribosome synthesis during early developmental stages (Bendich 1987). The large ribosomal subunit (*rrn23*, cpLSU) is arranged upstream of the smallest ribosomal subunits of 4.5S (*rrn4.5*) and 5S RNA (*rrn5*), which might facilitate expression and delivery of either subunit at equal ratios. Moreover, the existence of two copies facilitates the maintenance of these genes by, e.g., gene conversion (Lemieux and Lee 1987). The small ribosomal subunit (*rrn16*, cpSSU) is separated from the remainder rRNAs by two tRNA genes. Functional domains of either rRNA species are highly conserved and show 65–80% similarity to eubacterial (cyanobacterial) ribosomal RNAs (Palmer 1985; Harris et al. 1994; Stoebe and Kowallik 1999; Zerges 2000).

30 different tRNAs are encoded in a typical angiosperm plastid genome. Recognition of all 61 codons is possible by superwobbling (“two out of three”-mechanism; Lagerkvist 1978; Pfitzinger et al. 1990; Rogalski et al. 2008). Superwobbling allows reading of all possible codons even if there is only *one* tRNA encoded as in the case of alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, histidine, lysine, phenylalanine, proline, tryptophan, and tyrosine

(Palmer 1991; Sugiura 1992; Bock 2007). In addition to protein biosynthesis, glutamyl tRNA (encoded by the plastid *trnE* gene) plays a prominent role during activation of heme biosynthesis (Smith 1988; Howe and Smith 1991; Jahn et al. 1992). This and the low rates of tRNA import into cell organelles (Dietrich et al. 1992, Dietrich et al. 1996; Lohan and Wolfe 1998) led Barbrook et al. (2006) to suggest that a minimal plastid genome would at least contain the *trnE* gene. However, experimental data concerning the import machinery for small structural RNAs are rare and evidence for general tRNA import into plastids is lacking. Therefore, it remains speculative to what extent the plastid genome could eventually be reduced.

Nonphotosynthetic and minimally photosynthetic angiosperms typically retain only a fraction of tRNAs (Morden et al. 1991; Lohan and Wolfe 1998; Funk et al. 2007; McNeal et al. 2007, 2009; Nickrent and García 2009). In Orobanchaceae, the loss of some tRNA-genes, e.g. *trnC*, seems to be correlated with the loss of photosynthesis (Taylor et al. 1991). Because expression analyses of retained genes in the highly reduced plastomes of *Epifagus* (Wolfe et al. 1992) and *Conopholis* (Wimpee et al. 1991, Wimpee et al. 1992) suggest an intact translation apparatus, the loss of tRNAs from their genomes might be indicative of tRNA import into plastid organelles. Pseudogenization of tRNAs has been reported for the mistletoe *Arceuthobium* (Nickrent and García 2009) and for *Cuscuta* (Funk et al. 2007; McNeal et al. 2007). In non-parasitic plants, the loss of e.g. *trnK_{UUU}* has occurred independently multiple times (*Selaginella*: Tsuji et al. 2007; leptosporangiate ferns: Duffy et al. 2009; Wolf et al. 2010; Gao et al. 2010; Geraniaceae: Guisinger et al. 2010).

Plastid ribosomal proteins and ribosomes

Plastid protein biosynthesis is carried out at eubacterial-like 70S ribosomes (Zerges 2000). These are assembled from the small 30S ribosomal subunit and the large 50S subunit. The 16S ribosomal RNA builds the backbone of the 30S ribosome subunit, which additionally includes 25 ribosomal proteins (Yamaguchi et al. 2000). The remaining three plastid rRNA species together with 33 ribosomal proteins constitute the 50S ribosome subunit (Yamaguchi and Subramanian 2000). Most genes coding for ribosomal subunit proteins have been transferred to the nuclear genome. However, land plant plastomes commonly encode twelve proteins for the small ribosomal subunits (*rps* genes) and nine large ribosomal subunit proteins (*rpl* genes). Loss of *rps* and *rpl* genes from plastomes is rare, but has been detected in rosids (e.g. *rpl22*, *rpl23*, *rps16*; see Jansen et al. 2007; Jansen et al. 2010; Magee et al. 2010 for an overview) and a variety of non-photosynthetic or minimally photosynthetic angiosperms (*Epifagus*: dePamphilis

and Palmer 1990; *Conopholis*: Y. Zhang and C. W. dePamphilis, unpublished data; *Cuscuta*: Funk et al. 2007; McNeal et al. 2007; *Arceuthobium*: Nickrent and García 2009). Whether parasitic angiosperms are able to translate proteins with a reduced set of ribosomal proteins or import missing components is still unknown.

Other proteins associated with plastid ribosomes are a nuclear encoded ribosome recycling factor and several plastid ribosome specific proteins (PSRPs) that are unique to plants and show no similarities to bacterial homologs (Yamaguchi et al. 2000, Yamaguchi et al. 2003; Yamaguchi and Subramanian 2000; Sharma et al. 2010). The assembly of the eubacterial-type ribosomes has been studied intensively (reviewed in Moore 1998), but so far no such studies are available for plastid ribosomes. Given the high similarity of ribosomal RNA and most ribosomal proteins between eubacteria and plastids, it can be assumed that plastid ribosome assembly is similar to that of eubacteria. Most of the ribosomal proteins of the 30S ribosome subunit bind to the so-called S7-branch or are dependent on other (plastid encoded) proteins for binding (Grondek and Culver 2004). Thus, through analogy with eubacterial ribosomal proteins, plastid encoded ones might be divided into primary, secondary and tertiary binding components of the 30S and the 50S (Table 1) ribosome subunit according to their rRNA binding features.

Four proteins that are bound to the 30S ribosome subunits have no homologs in the eubacterial (i.e. *E. coli*-type) ribosome and are nuclear-encoded PSRPs. Two additional PSRP-proteins are bound to the 50S ribosome subunit (Yamaguchi et al. 2000; Yamaguchi and Subramanian 2000). It remains unknown how PSRPs are assembled into functional ribosome complexes. Recent analyses of PSRPs suggest that they play a role in light-dependent regulation of transcription/translation processes (Sharma et al. 2010).

One translation initiation factor assisting in the assembly of the translation initiation complex is encoded by the plastid gene *infA* (translation initiation factor; a total of three are known from eubacterial translation mechanisms). *InfA* has been lost multiple times independently during land plant evolution. Although present in all bryophyte and fern lineages, it is pseudogenized in the lycophyte *Isoetes* (Karol et al. 2010), but appears to be functional in other lycophytes (*Selaginella*: Tsuji et al. 2007; *Huperzia*: Wolf et al. 2005). In angiosperms, multiple losses have been reported (summarized in Jansen et al. 2007; Magee et al. 2010), accumulating in lineages known for their non-canonical plastid genome evolution (e.g. legumes; Millen et al. 2001).

clpP—a protein-modifying enzyme

High levels of photosynthetic gene expression coincide with an enormous protein turn-over in plastids. Both

maturation and protein degradation involve ATP-dependent synthase/protease complexes that act as molecular chaperones restoring or degrading damaged proteins according to the severity of protein denaturation (Wawrzynow et al. 1996; Adam et al. 2001; Adam and Clarke 2002). In plastids, three different protease complexes have been identified: *Fts* (filamentation temperature sensitive protease), *DegP/HtrA* (high temperature requirement protease A) and *Clp* (Caseinolytic protease). Whereas all subunits of the first two complexes are encoded by the nuclear genome, *ClpP* is plastid encoded.

Plastid genes coding for protein subunits involved in photosynthetic dark reactions and biogenesis

Genes for protochlorophyllide reductase subunits, proteins for CO₂ uptake and cytochrome C biogenesis

Bryophytes, lycophytes, ferns and most gymnosperms harbor genes for three subunits of a light-independent protochlorophyllide reductase (*chlB*, *chlL*, *chlN*) in their plastomes. This enzyme is involved in porphyrin and chlorophyll metabolism (Reinbothe and Reinbothe 1996; Karpinska et al. 1997). In gnetophytes, an aberrant gymnosperm group with still controversial phylogenetic position (e.g. Zhong et al. 2010), *chlB*, *chlL* and *chlN* are lost to different extents (McCoy et al. 2008; Wu et al. 2009). In *Ephedra*, sister group to the remaining Gnetales (Zhong et al. 2010), all three genes are present and intact, whereas *Gnetum* and *Welwitschia* possess pseudogenes of two subunits and have lost the third (McCoy et al. 2008; Wu et al. 2009). Different patterns in pseudogenization and *chl*-gene loss in both genera might indicate relaxation of evolutionary constraints to maintain functional copies, perhaps due to import of as yet unidentified nuclear substitutes.

The gene *ccsA* (*ycf5*) encodes a protein mediating the attachment of heme to c-type cytochromes during cytochrome biogenesis (Xie and Merchant 1996; Saint-Marcoux et al. 2009). The gene is localized in the plastid SSC region, and widely conserved among photosynthetic plants. However, *ccsA* is lost from *Epifagus* (Wolfe et al. 1992), and pseudogenized in *Aneura mirabilis* (Wickett et al. 2008a) The reading frame is, however, retained in all *Cuscuta* species sequenced so far (McNeal et al. 2007; Funk et al. 2007).

Land plant plastomes also encode a protein localized in the inner envelope membrane (inner-envelope protein, *cemA/ycf10*; Sasaki et al. 1993b). Knockouts of the gene *cemA* in *Chlamydomonas* severely affected the uptake of CO₂, while not affecting photosynthetic reactions (Rolland et al. 1997). *CemA* is lost from the plastid genome of *Epifagus* (Wolfe et al. 1992) and other Orobanchaceae (S. Wicke et al., unpublished data), but present in *Cuscuta*

(Funk et al. 2007; McNeal et al. 2007), and *Aneura* (Wickett et al. 2008a).

rbcL

The *rbcL* gene encodes the large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). RuBisCO is estimated to be the most abundant protein on earth (Ellis 1979). With the assistance of chaperones, it is assembled from eight large subunits (RbcL) and eight small subunits (RbcS). In contrast to red algae and Glaucophytes, Chlorophytes and Streptophytes do not possess a functional gene copy for the small RuBisCO subunit (*rbcS* gene) in the plastid genome. Instead, RbcS is encoded as a nuclear gene family and targeted to the plastid (Clegg et al. 1997). In contrast to many other photosynthesis related genes, *rbcL* is often retained in non-photosynthetic plants. Putatively functional copies of *rbcL* are retained in several representatives of Orobanchaceae, such as *Lathraea* (Delavault et al. 1996; Lusson et al. 1998), *Orobanche corymbosa*, *O. fasciculata* (Wolfe and dePamphilis 1997; Leebens-Mack and dePamphilis 2002), most species of *Harveya* (Leebens-Mack and dePamphilis 2002; Randle and Wolfe 2005), and the non-photosynthetic liverwort *Aneura mirabilis* (Wickett et al. 2008a). In other broomrape species, *rbcL* is only found as a pseudogene (as in *Epifagus*: Wolfe et al. 1992, *O. cernua*: Wolfe and dePamphilis 1997; *Hyobanche*, Randle and Wolfe 2005), or has been completely lost (S. Wicke et al., unpublished data). Retention, expression, and evidence for strong purifying selection in hemiparasitic and some holoparasitic plants have led to the speculation that *rbcL* is involved in another, yet photosynthesis unrelated pathway (Leebens-Mack and dePamphilis 2002; Randle and Wolfe 2005; McNeal et al. 2007; see section “Plastid encoded genes for photosynthesis unrelated pathways”).

Plastid genes for thylakoid complexes involved in photosynthetic light reactions

Oxygenic photosynthesis requires efficient light harvesting systems as well as an electron transport chain. The inner (thylakoid) membrane of the plastid contains at least five major protein complexes: photosystem I (PSI), photosystem II (PSII), cytochrome *b₆f* complex, ATP synthase and an NAD(P)H-plastoquinone oxidoreductase-complex (summarized in Table 1; Gounaris et al. 1986; Nixon et al. 1989).

Photosystem I and II (*psa* and *psb* genes)

In plants, light is harvested by two photosynthetic reaction centers, PSI and PSII. These are localized in the thylakoid

membrane and form supercomplexes, each with its own light harvesting complex that absorbs light via antenna molecules (chlorophyll a/b, and carotenoids). The exact number of proteins and cofactors associated with PSI and PSII supercomplexes is not known. PSII contains at least 17 subunits, 15 of which are encoded by the plastid genome (*psbA*, B, C, D, E, F, H, I, J, K, L, M, N, T, Z). These genes are scattered across the LSC region. All plastid *psb*-gene products form transmembrane helices (Nelson and Yocum 2006) and bind to the subunits *PsbA* (syn. D1), B, C, and D (syn. D2; Eckardt 2001). The gene products of *psbN* and *psbZ* (syn. *ycf9*) supposedly interact with the chlorophyll-bound subunit *PsbC* that reaches into the thylakoid lumen (Nelson and Yocum 2006). The structure of PSI is less complex than that of PSII, because it contains fewer polypeptides in its reaction center. The genes encoding for its plastid encoded subunits (*psa* genes) are found in the LSC region with the exception of *psaC*, which is embedded in an operon of *ndh*-genes in the plastome SSC region. Five subunits of plastid encoded PSI (A, B, C, I, J) are transmembrane proteins. The structurally highly similar apoproteins *PsaA* and *PsaB* bind to the iron-sulfur reaction center that mediates the transfer of excited electrons from plastoquinone to ferredoxin (Nelson and Yocum 2006). *PsaC* codes for a peripheral subunit on the stromal side of PSI, which is directly involved in ferredoxin reduction by binding the terminal electron acceptor molecules and linking them to the PSI iron-sulfur center (Fischer et al. 1998). Subunits I and J are not essential for PSI function (Bock 2007).

Photosystem assembly factors (*ycf3*, *ycf4*)

Both photosystems have been shown to be assembled with the help of chaperones (Nelson and Yocum 2006). The products of two plastid genes, *ycf3* (orf62) and *ycf4* (orf184), function as assembly factors for the photosystem I complex (Boudreau et al. 1997a; Ruf et al. 1997; Naver et al. 2001; Ozawa et al. 2009). Mutations in certain amino acid residues that mediate protein–protein interactions led to decreasing levels of PSI accumulation in the thylakoid membrane, as did gene disruption experiments (Boudreau et al. 1997a). Recently, it has been shown that *Ycf3* interacts with at least one nuclear encoded protein during the assembly of PSI (Albus et al. 2010). The naming of both genes is somewhat misleading as it implies that their function is still unknown. However, the transcripts of both ORFs are obviously translated and the resulting polypeptides assist during the assembly of the photosystem I. We therefore suggest renaming both genes to *PSI assembly factor I* (*pafl*, the former *ycf3*) and *II* (*paflII*, the former *ycf4*). The specifications I and II refer to the timing at which they are thought to interact with PSI following the model proposed by Ozawa et al. (2009).

Cytochrome *b₆f* complex (*pet*-genes) and ATP-Synthase complex (*atp*-genes)

PSII and PSI are electrochemically connected in series by the cytochrome *b₆f* complex. This is a functional complex composed of nine different subunits plus several inorganic cofactors, such as chlorophyll a, heme, β -carotene and an iron-sulfur cluster (Baniulis et al. 2008).

Six subunits are plastid-encoded (*petA*, B, D, G, L, N). These participate in electron transfer, generating a proton gradient across the thylakoid membrane (Stroebele et al. 2003). Together with the nuclear encoded Rieske protein (PetC), the gene products of *petA* (cytochrome *f*), *petB* (cytochrome *b₆*) and *petD* (subunit IV) form the core complex that acts in the linear electron transport (Kurisu et al. 2003). The remaining subunits (PetN, PetG, PetL plus nuclear encoded PetM, PetH) are hydrophobic molecules and are arranged peripherally around the core (Cramer et al. 2006).

Plastid ATP Synthase is a multi-subunit complex composed of nine different proteins generating ATP using the proton gradient. These constitute an integral membrane domain (F₀-domain) and an extrinsic catalytic domain (F₁-domain) reaching into the stroma (McCarty 1992). The F₁-subunit consists of five different polypeptides (α – ϵ), three of which are encoded by the plastome (*atpA*, B, E). The F₀-domain involved in proton translocation is built from three different polypeptides (*a*–*c*) that are exclusively plastid encoded (*atpF*, I, H; Vollmar et al. 2009).

All plastid-encoded genes for the photosynthetic apparatus are highly conserved in land plant plastomes (with the exception of *ndhA*–*K*, see below). Loss or pseudogenization have only been reported in non-photosynthetic parasitic (Krause 2008) or myco-heterotrophic (Wickett et al. 2008a, b) plants.

Plastid NAD(P)H-complex (*ndh*-genes)

Electrons are recycled around PSI in different pathways. One of which is carried out by a plastid NAD(P)H-dehydrogenase complex (*Ndh1*-complex) incorporated in the thylakoid membrane (Casano et al. 2000; Nixon 2000). This complex might also be involved in chlororespiration, i.e. the process of respiratory electron transport in addition to and/or in interaction with the photosynthetic electron transport. The plastid *Ndh1*-complex non-photochemically reduces and oxidizes plastoquinones. Furthermore, it may also mediate the transport of electrons from PSI-ferredoxins back to PSII (reverse electron transport; Peltier and Cournac 2002). Subunit composition appear to be highly divergent between cyanobacteria and derived land plants (reviewed in Suorsa et al. 2009). Together with several partly uncharacterized subunits, *Ndh1* consists of distinct

subcomplexes ranging from ca. 500 to over 1,000 kDa (Suorsa et al. 2009).

Eleven subunits of the *Ndh1*-complex are encoded by the plastid genome (*ndhA*, B, C, D, E, F, G, H, I, J, K). Plastid subunits A–D as well as H–K are homologous to the eubacterial (mitochondrial) proton pumping complex I (Friedrich et al. 1995). Experimental studies have shown that plastid encoded *Ndh1*-subunits might not be essential for plant survival in tobacco, although *ndh*-gene knockouts did cause phenotypic alterations (Peltier and Cournac 2002 and references therein).

The plastid encoded genes of the *Ndh1* are pseudogenized or entirely lost several times during land plant evolution. Given current data, these losses seem to be predominantly connected to a heterotrophic lifestyle in land plants (parasitism, some forms of myco-heterotrophy). This includes the myco-heterotrophic and non-photosynthetic liverwort *Aneura mirabilis* (Wickett et al. 2008a), the photosynthetic or partially non-photosynthetic parasitic *Cuscuta* (McNeal et al. 2007; Funk et al. 2007), the non-photosynthetic parasite *Epifagus* (dePamphilis and Palmer 1990), orchid species (Chang et al. 2006; Wu et al. 2010), and some gymnosperms (Wu et al. 2009) as well as some representatives of carnivorous Lentibulariaceae (B. Schäferhoff, S. Wicke, C. W. dePamphilis and K. Müller, unpublished data), and some species of Geraniaceae (Blazier et al. 2011). *Ndh* genes are also absent from several chlorophyte algae genomes (incl. *Chlamydomonas*), but they are present in plastomes of the closest relatives of land plants (Turmel et al. 2006; see also Martín and Sabater 2010).

The *Ndh1* complex may also be associated with other pathways, and might play an important role in adaptation to environmental stress (reviewed in Suorsa et al. 2009). Abiotic stress, such as nutrient starvation (in particular nitrogen starvation), affected and up-regulated *ndh*-gene expression indicating a putative regulating function of *Ndh1* for the photosynthetic electron flow (Peltier and Schmidt 1991). Due to the presence of nuclear genes of *Arabidopsis* with strong similarities to *ndh* complexes and plastid transit peptide sequences (Peltier and Cournac 2002), the existence of a second, nuclear encoded plastid *ndh* complex (*Nda2*) has long been suspected. Recently, an alternative form of an plastid localized *Ndh*-complex involved in non-photochemical plastoquinone reduction was identified (Sirpiö et al. 2009; Takabayashi et al. 2009; Ishida et al. 2009; Suorsa et al. 2009, 2010). The existence of a second form might explain the multiple losses of *Ndh1* genes from land plant plastomes. It may be that the function of an alternative *Ndh*-complex, or of fewer or incompletely assembled *Ndh1*-subcomplexes is sufficient for photosynthetic and related pathways in some, yet not all, plants—in particular, if they exhibit a certain degree of heterotrophy (e.g. myco-heterotrophy, parasitism,

carnivory). It might be that nutrient supplies could affect the activity of the *Ndh1* complex in a way that renders it dispensable. In the light of expression analyses under nitrogen starvation (Peltier and Schmidt 1991), the responsible factor may include the type of nitrogen source (nitrate vs. ammonium) or the excess of nitrogen (and/or other nutrients or even assimilates) obtained from a host plant. It is unclear whether this also accounts for the loss of *ndh* genes from the plastomes of Pinaceae, Gnetophytes and some Geraniaceae. As with many land plants, gymnosperms live in close association with mycorrhizae (Wang and Qiu 2006). Thus, it may be possible that fungal associations, or the fungal symbiont itself contributes to the fate of *ndh*-genes. On the other hand, throughout land plants, the presence of mycorrhizae and *ndh* loss appear to be only imperfectly correlated; evidently, more data is necessary before sound conclusions can be drawn, since other reasons such as multiple independent functional gene transfers must be considered as well (see also Blazier et al. 2011).

Plastid encoded genes for photosynthesis unrelated pathways

Plastid genes for metabolic pathways unrelated to photosynthesis include proteins for fatty acid synthesis, and sulfur metabolism.

AccD and the *RuBisCO* “shunt”

Acetyl-CoA carboxylase is another key enzyme in plastids mediating the irreversible conversion of acetyl-CoA to malonyl-CoA during the biosynthesis of fatty acids (Neuhaus and Emes 2010). The beta subunit of this multimeric enzyme (*accD*) is encoded in the LSC of the plastome in Streptophytes (Sasaki et al. 1993a) and is considered to be crucial for leaf development (Kode et al. 2005). The *accD* gene has been lost from the plastid genome several times in angiosperms (Jansen et al. 2007) where its function is fulfilled by nuclear copies (Nakkaew et al. 2008).

Recently, *RuBisCO* has been found to be involved in a previously unrecognized glycolysis bypassing reaction that converts carbohydrates to fatty acids at low carbon cost in oily seeds of white turnip (*Brassica rapa*, Schwender et al. 2004). This has been proposed as a likely reason for the retention of a photosynthetic pathway in parasitic species of *Cuscuta* that are fully heterotrophic, yet nonetheless would benefit from the *RuBisCO* “shunt” to enable rapid and efficient lipid synthesis (McNeal et al. 2009).

Genes related to sulfur metabolism

Liverworts contain at least two more protein coding genes absent from most other land plants, *cysA* and *cysT*. *CysA*

(designated *mbpX* in the *Marchantia polymorpha* plastome) shows functional domains similar to inner membrane sulfate ABC (*ATP binding cassette*) transporters. Although conservation of amino acid composition drops dramatically towards the N-terminus, similarity searches suggest that both genes might belong to sulfate related transport complexes or sulfate permeases and thus may have a function related to sulfate metabolism (Laudenbach and Grossman 1991). However, both subunits are lacking from most other land plant plastid genomes (mosses, ferns, seed plants) and the green algae *Chlamydomonas* (Sugiura 1992; Maul et al. 2002; Melis and Chen 2005; Lindberg and Melis 2008). In hornworts, a *cysA*-like gene is present in the plastid genome, but it appears to be non-functional (Kugita et al. 2003).

Plastid genes of unknown function

ycf1 and *ycf2*

Green algae, including the closest relatives of Embryophytes, possess an *ftsH* reading frame, which encodes a metalloprotease. Predominantly at the carboxyl-terminus, *ftsH* exhibits similarities to the largest, yet functionally uncharacterized ORF (*ycf2*) in land plants (Wolfe 1994). Nucleotide sequence similarity among land plant *ycf2* is extraordinarily low compared to other plastid-encoded genes, being less than 50% across bryophytes, ferns, and seed plants. *Ycf2* harbors nucleotide binding sites typical for green algal and eubacterial *ftsH* and CDC48 gene families, which are involved in cell division processes, proteolysis, and protein transport (Wolfe 1994). Furthermore, a “DPAL”-motif, shared by CDC48 and *ycf2*, is highly conserved. In several angiosperm plastomes, a smaller ORF, *ycf15*, is present directly downstream of the *ycf2* gene (Raubeson et al. 2007 and references therein). So far, an exact function has not been assigned to the *ycf15* gene product. Expression studies in spinach suggested that *ycf15* might act as a regulator for *Ycf2* on the RNA level, but might not function on protein level (Schmitz-Linneberger et al. 2001). Consistent with an RNA-level function, Raubeson et al. (2007) showed that *ycf15* is not under purifying selection as expected for most protein coding sequences. A non-protein function might also account for the conservation of the cryptic reading frame *ycf68* found in the IRs of several angiosperms (Raubeson et al. 2007) and *Aneura mirabilis* (Wickett et al. 2008a). The persistence of both *ycf15* and *ycf68* ORFs might be attributable to their localization in the slowly evolving IR region.

Ycf1, the second largest gene in plastid genomes, codes for a protein of approximately 1800 amino acids, yet its precise function remains to be determined. Experimental

data and comparisons of *Chlamydomonas* and angiosperm *ycf1* homologs revealed conserved nucleotide binding sites (Boudreau et al. 1997b). Based on these data, functions of *ycf1* and *ycf2* have been hypothesized to involve ATPase-related activities, chaperone-function, activity in cell divisions (depicted from similarities with *ftsH*) and structural remodeling and/or linkage of plastid chromosomes to protein and/or membrane structures (Wolfe 1994; Boudreau et al. 1997b). Available data on gene expression in tobacco show that, similar to *ycf2*, *ycf1* is expressed in fruits (Drescher et al. 2000). Products of both genes are essential for plant cell survival (Drescher et al. 2000; Boudreau et al. 1997b). In most land plant lineages, *ycf1* and *ycf2* genes have elevated substitution rates and may have undergone pseudogenization (Oliver et al. 2010; Wolf et al. 2010a). For the most part, however, the 5' end of both genes are relatively conserved, whereas other parts seem to evolve more freely. In the case of *ycf1*, this might be due to the co-localization of a replication origin (*ori_B*) in this region (Kunnimalaiyaan and Nielsen 1997). This implies that both genes seem to undergo at least weak selective constraints. Analyses regarding differences in d_n/d_s ratios and mutational hotspots within the genic region might corroborate the assignment of a function to both these genes. The losses observed in several photosynthetic lineages, however, raise the question whether they really carry out essential functions in *all* plants. Complete loss of both *ycf1* and *ycf2* from the plastomes of some (but not all) derived monocot lineages and putative pseudogenization in other plants (Downie et al. 1994) are in contrast to the high structural conservation in parasites (dePamphilis and Palmer 1990; Wolfe et al. 1992; McNeal et al. 2007). This might in fact point towards a function decoupled from photosynthesis. Nuclear encoded and plastid targeted proteins similar to Ycf1/Ycf2 were not found in lineages where both genes have been lost from the plastid genome, such as Poaceae (Downie et al. 1994).

Conclusions

In terms of structure, land plant plastid chromosomes evolve much more slowly than their mitochondrial or nuclear counterparts. This structural conservatism might be a result of the common organization of genes in operons that are conserved features between cyanobacteria, green algae and land plants. Other relevant factors include the mode of plastid transmission, the activity of highly effective repair mechanisms, as well as the rarity of plastid fusion and fission. The latter property is one of the major differences relative to mitochondrial genomes that have

been shown to frequently fuse, and in doing so, provide opportunities for exchanging divergent genome copies. Most plastome rearrangements appear to be restricted to lineages that show one or more of the following characteristics: (i) aberrant behavior of the inverted repeat region (expansion, contraction, loss), (ii) biparental plastid transmission; (iii) a high frequency of small dispersed repeat sequences, (iv) heterotrophic lifestyle (parasites, myco-heterotrophs). Among land plants, angiosperms show the greatest variation in plastome structure, although distortion of gene synteny by rearrangements and gene loss is still rare compared to the genomes of other cell compartments. Interestingly, plastid chromosome restructuring appears to occur most commonly in the more derived clades of a given lineage (leptosporangiate ferns, Funariales within mosses, Pinaceae and Gnetophytes within gymnosperms, eudicots and Poales within angiosperms). It will be interesting to see whether similar patterns occur in liverwort plastome evolution. The gene content of land plants does not appear to have dramatically changed, and only few gene losses or putative functional transfers (*chl*, *cys*) might have taken place in the course of land plant evolution. The retention of photosynthetically relevant genes might be attributable to several factors. On the one hand, functional gene transfer is a complex issue since it involves the transfer itself and the evolution of transit peptides; thus, it is expected to be rare. On the other hand, most protein subunits encoded by the plastome (in particular photosynthesis relevant proteins) harbor trans-membrane proteins, and might therefore be difficult to import (as known from mitochondria). Finally, many gene products are required at high expression levels and at early developmental stages (e.g. translation/transcription apparatus, photosynthesis genes) and their retention might be selected for.

Acknowledgments Thanks are especially due to Monika Ballmann (Bonn) and Ortrun Lepping (Münster) for assistance in retrieving genome annotations from public databases. We are very grateful to Norman Wickett (PSU), Paul Wolf (USU), Yan Zhang (PSU) and Josh Der (PSU) for fruitful and inspiring discussions on plastid genome evolution in land plants. We thank Laura Forrest and Bernard Goffinet (UConn) as well as Bastian Schäferhoff (IEB, Muenster) for sharing unpublished data with us. We appreciate helpful suggestions on an earlier version of this manuscript by two anonymous reviewers. Funding of our own research on parasitic plants and carnivores was obtained from the Austrian Science Fund (FWF grant 19404 to G.M.S), DFG grant MU2875/2, to K.F.M), and N.S.F. grants DEB-0120709 and DBI-0701748 to C.W.D.). Financial support to S.W. from the University of Vienna (Austria) and the Botanical Society of America is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Adam Z, Clarke AK (2002) Cutting edge of chloroplast proteolysis. *Trends Plant Sci* 7:451–456
- Adam Z, Adamska I, Nakabayashi K, Ostersetzer O, Haussuhl K, Manuell A, Zheng B, Vallon O, Rodermeil SR, Shinozaki K, Clarke AK (2001) Chloroplast and mitochondrial proteases in *Arabidopsis*. A proposed nomenclature. *Plant Physiol* 125:1912–1918
- Albus CA, Ruf S, Schottler MA, Lein W, Kehr J, Bock R (2010) Y3IP1, a nucleus-encoded thylakoid protein, cooperates with the plastid-encoded Ycf3 protein in photosystem I assembly of tobacco and *Arabidopsis*. *Plant Cell* 22:2838–2855
- Ayliffe MA, Scott N, Timmis JN (1998) Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Mol Biol Evol* 15:738–745
- Blazier JC, Guisinger ME, Jansen RK (2011) Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol*. doi:10.1007/s11103-011-9753-5
- Baniulis D, Yamashita E, Zhang H, Hasan SS, Cramer WA (2008) Structure-function of the cytochrome *b₆/f* complex. *Photochem Photobiol* 84:1349–1358
- Barbrook AC, Howe CJ, Purton S (2006) Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci* 11:101–108
- Bell P, Frey-Wyssling A, Mühlethaler K (1966) Evidence for the discontinuity of plastids in the sexual reproduction of a plant. *J Ultrastruct Res* 15:108–121
- Bendich AJ (1987) Why do chloroplasts and mitochondria contain so many copies of their genome? *BioEssays* 6:279–282
- Bendich AJ (1991) Moving pictures of DNA released upon lysis from bacteria, chloroplasts, and mitochondria. *Protoplasma* 160:121–130
- Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16:1661–1666
- Bendich AJ (2007) The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. *BioEssays* 29:474–483
- Bendich AJ, Smith SB (1990) Moving pictures and pulsed-field gel electrophoresis show linear DNA molecules from chloroplasts and mitochondria. *Curr Genet* 17:421–425
- Berg S, Krause K, Krupinska K (2004) The *rbcl* genes of two *Cuscuta* species, *C. gronovii* and *C. subinclusa*, are transcribed by the nuclear-encoded plastid RNA polymerase (NEP). *Planta* 219:541–546
- Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) *Cell and Molecular Biology of Plastids*. Springer, Berlin Heidelberg, pp 29–63
- Bortiri E, Coleman-Derr D, Lazo G, Anderson O, Gu Y (2008) The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res Notes* 1:61
- Boudreau E, Takahashi Y, Lemieux C, Turmel M, Rochaix J (1997a) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J* 16:6095–6104
- Boudreau E, Turmel M, Goldschmidt-Clermont M, Rochaix J, Sivan S, Michaels A, Leu S (1997b) A large open reading frame (orf1995) in the chloroplast DNA of *Chlamydomonas reinhardtii* encodes an essential protein. *Mol Gen Genet* 253:649–653
- Briat J, Lescure A, Mache R (1986) Transcription of the chloroplast DNA: a review. *Biochimie* 68:981–990
- Cahoon AB, Stern DB (2001) Plastid transcription a menage à trois? *Trends Plant Sci* 6:45–46
- Cai Z, Penaflor C, Kuehl JV, Leebens-Mack JH, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol Biol* 6:77
- Cai Z, Guisinger M, Kim H, Ruck E, Blazier J, McMurtry V, Kuehl J, Boore J, Jansen R (2008) Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol* 67:696–704
- Casano LM, Zapata JM, Martín M, Sabater B (2000) Chlororespiration and poisoning of cyclic electron transport. *J Biol Chem* 275:942–948
- Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, Chaw SM (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol* 23:279–291
- Chumley TW, Ferraris JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol* 23:2175–2190
- Clegg MT, Cummings MP, Durbin ML (1997) The evolution of plant nuclear genes. *Proc Natl Acad Sci USA* 94:7791–7798
- Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am J Bot* 75:1443–1458
- Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet* 31:419–429
- Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in *Campanulaceae*: phylogenetic utility of highly rearranged genomes. *BCM Evol Biol* 4:27
- Cramer WA, Zhang H, Yan J, Kurisu G, Smith JL (2006) Transmembrane traffic in the cytochrome *b₆/f* complex. *Annu Rev Biochem* 75:769–790
- Cui L, Leebens-Mack J, Wang LS, Tang J, Rymarquis L, Stern DB, dePamphilis CW (2006) Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. *BMC Evol Biol* 6:13
- Daniell H, Lee S, Grevich J, Sasaki C, Quesada-Vargas T, Guda C, Tomkins J, Jansen RK (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor Appl Genet* 112:1503–1518
- Day A, Madesis P (2007) DNA replication, recombination, and repair in plastids. In: Bock R (ed) *Cell and Molecular Biology of Plastids*. Springer, Berlin Heidelberg, pp 65–119
- Delavault PM, Russo NM, Lusson NA, Thalouarn P (1996) Organization of the reduced plastid genome of *Lathraea clandestina*, an achlorophyllous parasitic plant. *Physiol Plant* 96:674–682
- Deng X, Wing RA, Gruissem W (1989) The chloroplast genome exists in multimeric forms. *Proc Natl Acad Sci USA* 86:4156–4160
- dePamphilis CW, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348:337–339
- Dietrich A, Weil JH, Marechal-Drouard L (1992) Nuclear-encoded transfer RNAs in plant mitochondria. *Ann Rev Cell Biol* 8:115–131
- Dietrich A, Small I, Cosset A, Weil JH, Maréchal-Drouard L (1996) Editing and import: strategies for providing plant mitochondria

- with a complete set of functional transfer RNAs. *Biochimie* 78:518–529
- Downie SR, Bewley JD (1992) Restriction site mapping of the chloroplast DNA inverted repeat—a molecular phylogeny of the Asteridae. *Ann Mo Bot Gard* 79:266–283
- Downie SR, Katz-Downie DS, Wolfe KH, Calie PJ, Palmer JD (1994) Structure and evolution of the largest chloroplast gene (ORF2280): internal plasticity and multiple gene loss during angiosperm evolution. *Curr Genet* 25:367–378
- Drescher A, Ruf S, Calsa T, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J* 22:97–104
- Duffy AM, Kelchner SA, Wolf PG (2009) Conservation of selection on *matK* following an ancient loss of its flanking intron. *Gene* 438:17–25
- Eckardt NA (2001) A role for PsbZ in the core complex of photosystem II. *Plant Cell* 13:1245–1248
- Eisen J, Heidelberg J, White O, Salzberg S (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* 1:6 Research0011
- Ellis RJ (1979) The most abundant protein in the world. *Trends Biochem Sci* 4:241–244
- Fan WH, Woelfle MA, Mosig G (1995) Two copies of a DNA element, ‘Wendy’, in the chloroplast chromosome of *Chlamydomonas reinhardtii* between rearranged gene clusters. *Plant Mol Biol* 29:63–80
- Fejes E, Engler D, Maliga P (1990) Extensive homologous chloroplast DNA recombination in the *Nicotiana* somatic hybrid. *Theor Appl Genet* 79:28–32
- Fischer N, Hippler M, Setif P, Jacquot J, Rochaix J (1998) The PsaC subunit of photosystem I provides an essential lysine residue for fast electron transfer to ferredoxin. *EMBO J* 17:849–858
- Friedrich T, Steinmüller K, Weiss H (1995) The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett* 367:107–111
- Funk H, Berg S, Krupinska K, Maier U, Krause K (2007) Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol* 7:45
- Gao L, Yi X, Yang Y, Su Y, Wang T (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol Biol* 9:130
- Gao L, Su Y, Wang T (2010) Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *J Syst Evol* 48:77–93
- Gastony GJ, Yatskievych G (1992) Maternal inheritance of the chloroplast and mitochondrial genomes in cheilanthoid ferns. *Am J Bot* 79:716–722
- Goffinet B, Wickett NJ, Werner O, Ros RM, Shaw AJ, Cox CJ (2007) Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). *Ann Bot* 99:747–753
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 20:1499–1505
- Goulding SE, Olmstead R, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252:195–206
- Gounaris K, Barber J, Harwood JL (1986) The thylakoid membranes of higher plant chloroplasts. *Biochem J* 237:313–326
- Gray B, Ahner B, Hanson M (2009) Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. *Transgenic Res* 18:559–572
- Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG (2008) The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucl Acids Res* 36:2366–2378
- Grewe F, Viehoveer P, Weisshaar B, Knoop V (2009) A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucl Acids Res* 37:5093–5104
- Grondek JF, Culver GM (2004) Assembly of the 30S ribosomal subunit: positioning ribosomal protein S13 in the S7 assembly branch. *RNA* 10:1861–1866
- Guillon J, Raquin C (2000) Maternal inheritance of chloroplasts in the horsetail *Equisetum variegatum* (Schleich.). *Curr Genet* 37:53–56
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2010) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol*. doi:10.1093/molbev/msq229
- Guo XY, Ruan SL, Hu WM, Ca DG, Fan LJ (2008) Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites and ages of insertion involved. *Funct Integr Genom* 8:101–108
- Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol* 66:350–361
- Hagemann R (2004) The sexual inheritance of plant organelles. In: Daniell H, Chase C (eds) *Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and Mitochondria*. Springer, Dordrecht, The Netherlands, pp 104–105
- Hajdukiewicz PT, 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
- Harada T, Soma N, Ishikawa R, Niizeki M (1997) Topological resolution of rice plastid DNA by pulsed-field gel electrophoresis. *Bull Fac Agric Hirosaki Univ* 61:25–32
- Harris EH, Boynton JE, Gillham NW (1994) Chloroplast ribosomes and protein synthesis. *Microbiol Mol Biol Rev* 58:700–754
- Hausner G, Olsen R, Johnson I, Simone D, Sanders ER, Karol KG, McCourt RM, Zimmerly S (2006) Origin and evolution of the chloroplast *trnK* (*matK*) intron: a model for evolution of group II intron RNA structures. *Mol Biol Evol* 23:380–391
- Hilu KW, Liang HP (1997) The *matK* gene: sequence variation and application in plant systematics. *Am J Bot* 84:830–839
- Hirao T, Watanabe A, Kurita M, Kondo T, Takata K (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol* 8:70
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol Gen Genet* 217:185–194
- Howe CJ, Smith AG (1991) Plants without chlorophyll. *Nature* 349:109
- Hupfer H, Swiatek M, Hornung S, Herrmann RG, Maier RM, Chiu WL, Sears BB (2000) Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. *Mol Gen Genet* 263:581–585
- Jahn D, Verkamp E, Söll D (1992) Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends Biochem Sci* 17:215–218

- Jankowiak K, Rybarczyk A, Wyatt R, Odrzykoski I, Pacak A, Szweykowska-Kulińska Z (2005) Organellar inheritance in the allopolyploid moss *Rhizomnium pseudopunctatum*. *Taxon* 54:383–388
- Jankowiak-Siuda K, Pacak A, Odrzykoski I, Wyatt R, Szweykowska-Kulińska Z (2008) Organellar inheritance in the allopolyploid moss *Plagiomnium curvatulum*. *Taxon* 57:145–152
- Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc Natl Acad Sci USA* 84:5818–5822
- Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack JH, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104:19369–19374
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S, Daniell H (2008) Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol Phylogenet Evol* 48:1204–1217
- Jansen RK, Saski C, Lee S, Hansen AK, Daniell H (2010) Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol Biol Evol*. doi: 10.1093/molbev/msq261
- Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KDE, Hall JD, Hansen SK, Kuehl JV, Mandoli D, Mishler BD, Olmstead RG, Renzaglia K, Wolf PG (2010) Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evol Biol* 10:321
- Karpinska B, Karpinski S, Hällgren J (1997) The *chlB* gene encoding a subunit of light-independent protochlorophyllide reductase is edited in chloroplasts of conifers. *Curr Genet* 31:343–347
- Keeling PJ (2004) Diversity and evolutionary history of plastids and their hosts. *Am J Bot* 91:1481–1493
- Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Phil Trans R Soc B* 365:729–748
- Kim KJ, Lee HL (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. *Mol Cell* 19:104–113
- Kim K, Choi K, Jansen RK (2005) Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Mol Biol Evol* 22:1783–1792
- Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet* 46:123–139
- Knoop V, Unseld M, Marienfeld J, Brandt P, Sunkel S, Ullrich H, Brennicke A (1996) *Copia*-, *Gypsy*- and *Line*-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. *Genetics* 142:579–585
- Knox EB, Palmer JD (1999) The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): expansion of the inverted repeat in an ancestor of the Campanulales. *Plant Syst Evol* 214:49–64
- Kobayashi Y, Dokiya Y, Sugita M (2001) Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochem Biophys Res Commun* 289:1106–1113
- Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J* 44:237–244
- Kolodner R, Tewari KK (1979) Inverted repeats in chloroplast DNA from higher plants. *Proc Natl Acad Sci USA* 76:41–45
- Kozik A, Kochetkova E, Michelmore R (2002) GenomePixelizer—a visualization program for comparative genomics within and between species. *Bioinformatics* 18:335–336
- Krause K (2008) From chloroplasts to “cryptic” plastids: evolution of plastid genomes in parasitic plants. *Curr Genet* 54:111–121
- Krause K, Berg S, Krupinska K (2003) Plastid transcription in the holoparasitic plant genus *Cuscuta*: parallel loss of the *rrn16* PEP-promoter and of the *rpoA* and *rpoB* genes coding for the plastid-encoded RNA polymerase. *Planta* 216:815–823
- Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys(GCA). *Nucl Acids Res* 28:2571–2576
- Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K (2003) The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucl Acids Res* 31:716–721
- Kühn K, Weihe A, Börner T (2005) Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucl Acids Res* 33:337–346
- Kunnimalaiyaan M, Nielsen BL (1997) Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. *Nucl Acids Res* 25:3681–3686
- Kurisu G, Zhang H, Smith JL, Cramer WA (2003) Structure of the cytochrome *b₆f* complex of oxygenic photosynthesis: tuning the cavity. *Science* 302:1009–1014
- Lagerkvist U (1978) “Two out of three”: an alternative method for codon reading. *Proc Natl Acad Sci USA* 75:1759–1762
- Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annu Rev Genet* 3388:1–35
- Laudenbach DE, Grossman AR (1991) Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport. *J Bacteriol* 173:2739–2750
- Lee H, Jansen RK, Chumley TW, Kim K (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol* 24:1161–1180
- Leebens-Mack JH, dePamphilis CW (2002) Power analysis of tests for loss of selective constraint in cave crayfish and nonphotosynthetic plant lineages. *Mol Biol Evol* 19:1292–1302
- Lemieux C, Lee RW (1987) Nonreciprocal recombination between alleles of the chloroplast 23S rRNA gene in interspecific *Chlamydomonas* crosses. *Proc Natl Acad Sci USA* 8844:4166–4170
- Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. *Am J Bot* 91:1535–1556
- Liere K, Link G (1995) RNA binding activity of the *matK* protein encoded by the chloroplast *trnK* intron from mustard (*Sinapis alba*). *Nucl Acids Res* 23:917–921
- Lilly JW, Havey MJ, Jackson SA, Jiang J (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell* 13:245–254
- Lindberg P, Melis A (2008) The chloroplast sulfate transport system in the green alga *Chlamydomonas reinhardtii*. *Planta* 228:951–961
- Lohan AJ, Wolfe KH (1998) A subset of conserved tRNA genes in plastid DNA of nongreen plants. *Genetics* 150:425–433
- Lu Z, Kunnimalaiyaan M, Nielsen BL (1996) Characterization of replication origins flanking the 23S rRNA gene in tobacco chloroplast DNA. *Plant Mol Biol* 32:693–706
- Lusson NA, Delavault PM, Thalouarn P (1998) The *rbcL* gene from the non-photosynthetic parasite *Lathraea clandestina* is not transcribed by a plastid-encoded RNA polymerase. *Curr Genet* 34:212–215

- Mackiewicz P, Mackiewicz D, Kowalczyk M, Cebrat S (2001) Flip-flop around the origin and terminus of replication in prokaryotic genomes. *Genome Biol* 2:Interactions1004
- Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res* 20:1700–1710
- Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytol* 186:299–317
- Maréchal A, Parent J, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N (2009) Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci USA* 106:14693–14698
- Mariotti R, Cultrera N, Munoz Diez C, Baldoni L, Rubini A (2010) Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biology* 10:211
- Martín M, Sabater B (2010) Plastid *ndh* genes in plant evolution. *Plant Physiol Biochem* 48:636–645
- Martin W (2003) Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc Natl Acad Sci USA* 100:8612–8614
- Martin W, Kowallik KV (1999) Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *Eur J Phycol* 34:287–295
- 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
- Masood S, Nishikawa T, Fukuoka S, Njenga PK, Tsudzuki T, Kadowaki K (2004) The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. *Gene* 340:133–139
- Matsuo M, Ito Y, Yamauchi R, Obokata J (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* 17:665–675
- Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB (2002) The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *Plant Cell* 14:2659–2679
- Mccarty RE (1992) A plant biochemist's view of H⁺ ATPases and ATP synthases. *J Exp Biol* 172:431–441
- McCoy S, Kuehl J, Boore J, Raubeson L (2008) The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol Biol* 8:130
- McFadden GI, van Dooren GG (2004) Red algal genome affirms a common origin of all plastids. *Curr Biol* 14:R514–R516
- McNeal JR, Kuehl J, Boore J, dePamphilis C (2007) Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol* 7:57
- McNeal JR, Kuehl J, Boore J, Leebens-Mack JH, dePamphilis CW (2009) Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PLoS ONE* 4:e5982
- Melis A, Chen H (2005) Chloroplast sulfate transport in green algae—genes, proteins and effects. *Photosynth Res* 86:299–307
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermini LS, Wolfe KH (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658
- Milligan BG, Hampton JN, Palmer JD (1989) Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol Biol Evol* 6:355–368
- Mohr G, Lambowitz AM (2003) Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucl Acids Res* 31:647–652
- Mohr G, Perlman PS, Lambowitz AM (1993) Evolutionary relationships among group-II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucl Acids Res* 21:4991–4997
- Moore PB (1998) The three-dimensional structure of the ribosome and its components. *Ann Rev Biophys Biomol Struct* 27:35–58
- Morden CW, Wolfe KH, dePamphilis CW, Palmer JD (1991) Plastid translation and transcription genes in a nonphotosynthetic plant—Intact, missing and pseudogenes. *EMBO J* 10:3281–3288
- Müller K, Borsch T, Hilu KW (2006) Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. *Mol Phylogenet Evol* 41:99–117
- Nakazono M, Hira A (1993) Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. *Mol Gen Genet* 236:341–346
- Nakkaew A, Chotigeat W, Eksomtramage T, Phongdara A (2008) Cloning and expression of a plastid-encoded subunit, beta-carboxyltransferase gene (*accD*) and a nuclear-encoded subunit biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). *Plant Sci* 175:497–504
- Natcheva R, Cronberg N (2007) Maternal transmission of cytoplasmic DNA in interspecific hybrids of peat mosses, *Sphagnum* (Bryophyta). *J Evol Biol* 20:1613–1616
- Naver H, Boudreau E, Rochaix J (2001) Functional studies of Ycf3: Its role in assembly of photosystem I and interactions with some of its subunits. *Plant Cell* 13:2731–2745
- Nelson N, Yocum CF (2006) Structure and function of photosystems I and II. *Annu Rev Plant Biol* 57:521–565
- Neuhaus HE, Emes MJ (2010) Nonphotosynthetic metabolism in plastids. *Annu Rev Plant Physiol Plant Mol Biol* 51:111–140
- Nickrent D, García M (2009) On the brink of holoparasitism: plastome evolution in dwarf mistletoes (*Arceuthobium*, Viscaceae). *J Mol Evol* 68:603–615
- Nixon PJ (2000) Chlororespiration. *Philos Trans R Soc B* 355:1541–1547
- Nixon PJ, Gounaris K, Coomber SA, Hunter CN, Dyer TA, Barber J (1989) *psbG* is not a photosystem two gene but may be an *ndh* gene. *J Biol Chem* 264:14129–14135
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268:434–445
- Noutsos C, Richly E, Leister D (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res* 15:616–628
- Odom OW, Baek K, Dani RN, Herrin DL (2008) *Chlamydomonas* chloroplasts can use short dispersed repeats and multiple pathways to repair a double-strand break in the genome. *Plant J* 53:842–853
- Ogihara Y, Terachi T, Sasakuma T (1988) Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc Natl Acad Sci USA* 85:8573–8577
- Ohyama K (1996) Chloroplast and mitochondrial genomes from a liverwort, *Marchantia polymorpha*—gene organization and molecular evolution. *J Mol Evol* 60:16–24

- Oliver M, Murdock A, Mishler BD, Kuehl J, Boore J, Mandoli D, Everett K, Wolf PG, Duffy A, Karol KG (2010) Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC Genomics* 11:143
- Ozawa S, Nield J, Terao A, Stauber EJ, Hippler M, Koike H, Rochaix J, Takahashi Y (2009) Biochemical and structural studies of the large Ycf4-photosystem I assembly complex of the green alga *Chlamydomonas reinhardtii*. *Plant Cell* 21:2424–2442
- Pacak A, Szwejkowska-Kulińska Z (2002) Organellar inheritance in liverworts: an example of *Pellia borealis*. *J Mol Evol* 56:11–17
- Palmer JD (1983) Chloroplast DNA exists in two orientations. *Nature* 301:92–93
- Palmer JD (1985) Comparative organization of chloroplast genomes. *Annu Rev Genet* 119:325–354
- Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) *Cell Culture and Somatic Genetics of Plant*, vol 7A. *Molecular Biology of Plastids*, Academic Press, San Diego, pp 5–53
- Palmer JD (2000) Molecular evolution: a single birth of all plastids? *Nature* 405:32–33
- Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J Mol Evol* 28:87–97
- Palmer JD, Nugent JM, Herbon LA (1987a) Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc Natl Acad Sci USA* 84:769–773
- Palmer JD, Osorio B, Aldrich J, Thompson WF (1987b) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr Genet* 11:275–286
- Palmer JD, Soltis DE, Chase MW (2004) The plant tree of life: an overview and some points of view. *Am J Bot* 91:1437–1445
- Park J, Manen J, Schneeweiss GM (2007) Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). *Mol Phylogenet Evol* 43:974–985
- Peltier G, Cournac L (2002) Chlororespiration. *Annu Rev Plant Biol* 53:523–550
- Peltier G, Schmidt GW (1991) Chlororespiration: an adaptation to nitrogen deficiency in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA* 88:4791–4795
- Peltier J, Ripoll DR, Friso G, Rudella A, Cai Y, Ytterberg J, Giacomelli L, Pillardy J, van Wijk KJ (2004) Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. *J Biol Chem* 279:4768–4781
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol* 55:501–508
- Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH (2002) Evolutionary re-organisation of a large operon in Adzuki bean chloroplast DNA caused by inverted repeat movement. *DNA Res* 9:157–162
- Pfistering H, Weil JH, Pillay DTN, Guillemot P (1990) Codon recognition mechanisms in plant chloroplasts. *Plant Mol Biol* 14:805–814
- Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. *Syst Bot* 25:648–667
- Pyle AM, Lambowitz AM (2006) Group II introns: Ribozymes that splice RNA and invade DNA. In: Gesteland RF, Cech TR, Atkins JF (eds) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA, pp 449–505
- Qiu Y, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrowska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103:15511–15516
- Quandt D, Müller K, Huttunen S (2003) Characterisation of the chloroplast DNA *psbT*-H region and the influence of dyad symmetrical elements on phylogenetic reconstructions. *Plant Biol* 55:400–410
- Randle CP, Wolfe AD (2005) The evolution and expression of RBCL in holoparasitic sister-genera *Harveya* and *Hyobanche* (Orobanchaceae). *Am J Bot* 92:1575–1585
- Raubeson LA, Jansen RK (1992) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* 255:1697–1699
- Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry R (ed) *Diversity and Evolution of Plants - Genotypic and Phenotypic Variation in Higher Plants*. CABI Publishing, Wallingford, UK, pp 45–68
- Raubeson LA, Stein DB (1995) Insights into fern evolution from mapping chloroplast genomes. *Am Fern J* 85:193–204
- Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 88:174
- Reinbothe S, Reinbothe C (1996) The regulation of enzymes involved in chlorophyll biosynthesis. *Eur J Biochem* 237:323–343
- Revill MJW, Stanley S, Hibberd JM (2005) Plastid genome structure and loss of photosynthetic ability in the parasitic genus *Cuscuta*. *J Exp Bot* 56:2477–2486
- Rogalski M, Karcher D, Bock R (2008) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 15:192–198
- Rolland N, Dorne A, Amoroso G, Sultemeyer DF, Joyard J, Rochaix J (1997) Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. *EMBO J* 16:6713–6726
- Roper JM, Kellon Hansen S, Wolf PG, Karol KG, Mandoli DF, Everett KDE, Kuehl J, Boore JL (2007) The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). *Am Fern J* 97:95–106
- Rowan BA, Oldenburg DJ, Bendich AJ (2010) RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *J Exp Bot* 61:2575–2588
- Ruf S, Kössel H, Bock R (1997) Targeted inactivation of a tobacco intron-containing open reading frame reveals a novel chloroplast-encoded photosystem I-related gene. *J Cell Biol* 139:95–102
- Saint-Marcoux D, Wollman F, de Vitry C (2009) Biogenesis of cytochrome b_6 in photosynthetic membranes. *J Cell Biol* 185:1195–1207
- San Filippo J, Lambowitz AM (2002) Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J Mol Biol* 324:933–951
- Sasaki Y, Hakamada K, Suama Y, Nagano Y, Furusawa I, Matsuno R (1993a) Chloroplast-encoded protein as a subunit of acetyl-CoA carboxylase in pea plant. *J Biol Chem* 268:25118–25123
- Sasaki Y, Sekiguchi K, Nagano Y, Matsuno R (1993b) Chloroplast envelope protein encoded by the chloroplast genome. *FEBS Lett* 316:93–98
- Saski C, Lee S, Fjellheim S, Guda C, Jansen R, Luo H, Tomkins J, Rogli O, Daniell H, Clarke J (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet* 115:571–590

- Schmitz-Linneweber C, Maier RM, Alcaraz J, Cottet A, Herrmann RG, Mache R (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Mol Biol* 45:307–315
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y (2004) Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432:779–782
- Sears BB (1980) Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. *Plasmid* 44:233–255
- Shahmuradov IA, Akbarova YY, Solovyev VV, Aliyev JA (2003) Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Mol Biol* 52:923–934
- Sharma MR, Dönhöfer A, Barat C, Marquez V, Datta PP, Fucini P, Wilson DN, Agrawal RK (2010) PSRP1 is not a ribosomal protein, but a ribosome-binding factor that is recycled by the ribosome-recycling factor (RRF) and elongation factor G (EF-G). *J Biol Chem* 285:4006–4014
- Sheppard AE, Timmis JN (2009) Instability of plastid DNA in the nuclear genome. *PLoS Genet* 5:e1000323
- Shiina T, Tsunoyama Y, Nakahira Y, Khan MS (2005) Plastid RNA polymerases, promoters, and transcription regulators in higher plants. *Int Rev Cytol* 244:1–68
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi 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
- Smith AG (1988) Subcellular localization of two porphyrin-synthesis enzymes in *Pisum sativum* (pea) and *Arum* (cuckoo-pint) species. *Biochem J* 249:423–428
- Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci USA* 100:8828–8833
- Stein DB, Conant DS, Ahearn ME, Jordan ET, Kirch SA, Hasebe M, Iwatsuki K, Tan MK, Thomson JA (1992) Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. *Proc Natl Acad Sci USA* 89:1856–1860
- Stern DB, Astwood JD (1986) Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping, and locations of transposed chloroplast DNA sequences. *Nucl Acids Res* 14:5651–5666
- Stern DB, Lonsdale DM (1982) Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* 299:698–702
- Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. *Annu Rev Plant Biol* 61:125–155
- Stoebe B, Kowallik KV (1999) Gene-cluster analysis in chloroplast genomics. *Trends Genet* 15:344–347
- Stroebel D, Choquet Y, Popot J, Picot D (2003) An atypical haem in the cytochrome b_6/f complex. *Nature* 426:413–418
- Sugiura M (1992) The chloroplast genome. *Plant Mol Biol* 19:149–168
- Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucl Acids Res* 31:5324–5331
- Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Börner T (2008) Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucl Acids Res* 36:785–792
- Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthapaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: Structural organization and phylogenetic relationships. *DNA Res* 17:11–22
- Taylor GW, Wolfe KH, Morden CW, dePamphilis CW, Palmer JD (1991) Lack of a functional plastid tRNA^{Cys} gene is associated with loss of photosynthesis in a lineage of parasitic plants. *Curr Genet* 20:515–518
- Thompson WF, Stein DB, Palmer JD (1986) Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. *Curr Genet* 10:835–841
- Tilney-Bassett RAE, Almouslem AB (1989) Variation in plastid inheritance between pelargonium cultivars and their hybrids. *Heredity* 63:145–153
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5:123–135
- Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K (2007) The chloroplast genome from a lycophyte (microphylophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. *J Plant Res* 120:281–290
- Turmel M, Otis C, Lemieux C (2002) The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci USA* 99:11275–11280
- Turmel M, Otis C, Lemieux C (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol* 23:1324–1338
- Vogel JC, Russel SJ, Rumsey FJ, Barrett JA, Gibby M (1998) Evidence for maternal transmission of chloroplast DNA in the genus *Asplenium* (Aspleniaceae, Pteridophyta). *Bot Acta* 111:247–249
- Vollmar M, Schlieper D, Winn M, Büchner C, Groth G (2009) Structure of the c14 rotor ring of the proton translocating chloroplast ATP synthase. *J Biol Chem* 284:18228–18235
- 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*. *Proc Natl Acad Sci USA* 91:9794–9798
- Wakasugi T, Nishikawa A, Yamada K, Sugiura M (1998) Complete nucleotide sequence of the plastid genome from a fern, *Psilotum nudum*. *Endocytobiosis Cell Res* 13(Suppl):147
- Wang B, Qiu Y (2006) Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16:299–363
- Wang R, Cheng C, Chang C, Wu C, Su T, Chaw S (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 8:36
- Wawrzynow A, Banecki B, Zylicz M (1996) The Clp ATPases define a novel class of molecular chaperones. *Mol Microbiol* 21:895–899
- Weihe A, Börner T (1999) Transcription and the architecture of promoters in chloroplasts. *Trends Plant Sci* 4:169–170
- Wellburn FAM, Wellburn AR (1979) Conjoined mitochondria and plastids in the barley mutant ‘albostrians’. *Planta* 147:178–179
- Wicke S, Quandt D (2009) Universal primers for the amplification of the plastid *trnK/matK* region in land plants. *Anales Jard Bot Madrid* 66:285–288
- Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, dePamphilis CW, Boore JL, Goffinet B (2008a) Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Mol Biol Evol* 25:393–401
- Wickett NJ, Fan Y, Lewis P, Goffinet B (2008b) Distribution and evolution of pseudogenes, gene losses, and a gene rearrangement in the plastid genome of the nonphotosynthetic liverwort, *Aneura*

- mirabilis* (Metzgeriales, Jungermannioopsida). J Mol Evol 67:111–122
- Wimpee C, Wrobel R, Garvin D (1991) A divergent plastid genome in *Conopholis americana*, an achlorophyllous parasitic plant. Plant Mol Biol 17:161–166
- Wimpee CF, Morgan R, Wrobel RL (1992) Loss of transfer RNA genes from the plastid 16S–23S ribosomal RNA gene spacer in a parasitic plant. Curr Genet 21:417–422
- Wolf PG, Rowe CA, Sinclair RB, Hasebe M (2003) Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res 10:59–65
- Wolf PG, Rowe CA, Hasebe M (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. Gene 339:89–97
- Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL (2005) The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350:117–128
- Wolf PG, Der J, Duffy A, Davidson J, Grusz A, Pryer KM (2010) The evolution of chloroplast genes and genomes in ferns. Plant Mol Biol. doi:10.1007/s11103-010-9706-4
- Wolf PG, Roper JM, Duffy AM (2010b) The evolution of chloroplast genome structure in ferns. Genome 53:731–738
- Wolfe KH (1994) Similarity between putative ATP-binding sites in land plant plastid ORF2280 proteins and the FtsH/CDC48 family of ATPases. Curr Genet 25:379–383
- Wolfe AD, dePamphilis CW (1997) Alternate paths of evolution for the photosynthetic gene *rbcL* in four nonphotosynthetic species of *Orobanchaceae*. Plant Mol Biol 3:965–977
- Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 89:10648–10652
- Woodhouse MR, Pedersen B, Freeling M (2010) Transposed genes in *Arabidopsis* are often associated with flanking repeats. PLoS Genet 6:e1000949
- Wu CS, Wang YN, Liu SM, Chaw SM (2007) Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Mol Biol Evol 24:1366–1379
- Wu C, Lai Y, Lin C, Wang Y, Chaw S (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. Mol Phylogenet Evol 52:115–124
- Wu F, Chan M, Liao D, Hsu C, Lee Y, Daniell H, Duvall M, Lin C (2010) Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol 10:68
- Xie Z, Merchant S (1996) The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes. J Biol Chem 271:4632–4639
- Yamaguchi K, Subramanian AR (2000) The plastid ribosomal proteins—Identification of all the proteins in the 50 S subunit of organelle ribosome (chloroplast). J Biol Chem 275:28466–28482
- Yamaguchi K, von Knoblauch K, Subramanian AR (2000) The plastid ribosomal proteins—Identification of all the proteins in the 30 S subunit of organelle ribosome (chloroplast). J Biol Chem 275:28455–28465
- Yamaguchi K, Beligni MV, Prieto S, Haynes PA, McDonald WH, Yates JR, Mayfield SP (2003) Proteomic characterization of the *Chlamydomonas reinhardtii* chloroplast ribosome. J Biol Chem 278:33774–33785
- Yin C, Richter U, Börner T, Weihe A (2010) Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm *Nuphar advena* encodes two mitochondrial and one plastid phage-type RNA polymerases. BMC Evol Biol 10:379
- Young ND, dePamphilis CW (2000) Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. Mol Biol Evol 17:1933–1941
- Zerges W (2000) Translation in chloroplasts. Biochimie 82:583–601
- Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? J Plant Res 123:201–206
- Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2010) The position of gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. Mol Biol Evol 27:2855–2863
- Zoschke R, Nakamura M, Liere K, Sugiura M, Börner T, Schmitz-Linneweber C (2010) An organellar maturase associates with multiple group II introns. Proc Natl Acad Sci USA 107:3245–3250