Chapter 1

Plastid Genomes of Flowering Plants: Essential Principles

Tracey A. Ruhlman and Robert K. Jansen

Abstract

The plastid genome (plastome) has proved a valuable source of data for evaluating evolutionary relationships among angiosperms. Through basic and applied approaches, plastid transformation technology offers the potential to understand and improve plant productivity, providing food, fiber, energy, and medicines to meet the needs of a burgeoning global population. The growing genomic resources available to both phylogenetic and biotechnological investigations is allowing novel insights and expanding the scope of plastome research to encompass new species. In this chapter, we present an overview of some of the seminal and contemporary research that has contributed to our current understanding of plastome evolution and attempt to highlight the relationship between evolutionary mechanisms and the tools of plastid genetic engineering.

Key words Angiosperm, Plastome, Genome evolution, DNA recombination, Replication and repair, Intergenic region, Inverted repeat, Inheritance, Phylogeny

1 Introduction

The most notable defining feature of the plant cell is the presence of plastids, the bioenergetic organelles responsible for photosynthesis and myriad metabolic activities. Contemporary plastids carry the remnant genome (plastome) of their evolutionary ancestor, a photosynthetic bacterium believed to be related to extant cyanobacteria. Over eons, the coding capacity of the plastome has been greatly reduced relative to its progenitor such that a very small fraction of the ancestral gene complement remains [1]. The majority of plastid proteins are encoded in the nucleus and posttranslationally imported; the expression of plastome sequences is controlled by imported nuclear factors.

The evolutionary trajectory, from free-living organism to endosymbiont, to organelle, that has shaped the plastome is ongoing [2–4]. A survey of the plastome sequences in publicly available databases reveals that despite the prevalence of highly conserved sequence and organization in the majority of flowering plants sampled to date, a salient fraction show marked variation in their

Pal Maliga (ed.), *Chloroplast Biotechnology: Methods and Protocols*, Methods in Molecular Biology, vol. 2317,
https://doi.org/10.1007/978-1-0716-1472-3_1,

rates of sequence evolution and genomic architecture [5]. Although some examples of divergence are subtle, others are conspicuous and lead one to wonder; by what mechanisms do these changes arise and how are plants able to tolerate changes that appear disruptive?

In this chapter, we will outline some general features of the typical angiosperm plastome including its structure, organization and gene content. We will consider cases where genes otherwise found in the plastome are disrupted or missing and how these changes, along with genomic characters such as rearrangements, are used not only to infer phylogenetic relationships but also to extend our understanding of how organelle genomes change through evolutionary time. Further, we will discuss mechanisms that may be influencing genomic stability and consider how these same activities are inherently involved in the introduction of exogenous DNA sequences via plastid transformation.

2 Characteristics of the Angiosperm Plastome

Although not representative of the predominant physical state of plastomes, completed sequences are often illustrated schematically as circular maps (Fig. 1). Early studies concerned with the architectural features of plastomes used denaturation mapping and restriction enzyme digestion of DNA molecules isolated from purified plastids to characterize plastome size and structure. Prior to 1990 plastid DNA (ptDNA) from a diverse range of angiosperm species, representative gymnosperms and ferns, and a number of photosynthetic green algae, had already been scrutinized [6]. Among angiosperms the findings were largely in agreement and provided the framework for our current understanding of plastome structure. With the development of technology facilitating the direct sequencing of complete plastomes many of the seminal predictions have been confirmed providing a reasonably clear picture of the typical angiosperm plastome. The circular maps presented in the literature represent a single monomer, however numerous studies have identified more abundant alternative forms including concatemeric, linear and branched molecules (discussed below) along with some multimeric circles. The monomer is highly gene dense relative to nuclear or mitochondrial genomes, with 120–130 genes packed into 120–170 kb. Nonetheless, the gene space accounts for only ~50% of total nucleotide sequence with the remainder comprising introns, regulatory regions and intergenic spacers [7]. Plastomes are also highly AT rich; overall GC content is typically on the order of 30–40% and in some regions that do not encode proteins AT content exceeds 80% [8]. The proportion of GC, which is higher in protein coding regions, varies across plastomes by location, codon position and by functional group. For example, genes encoding photosynthetic functions have the highest GC content while the NAD(P)H genes have the lowest.

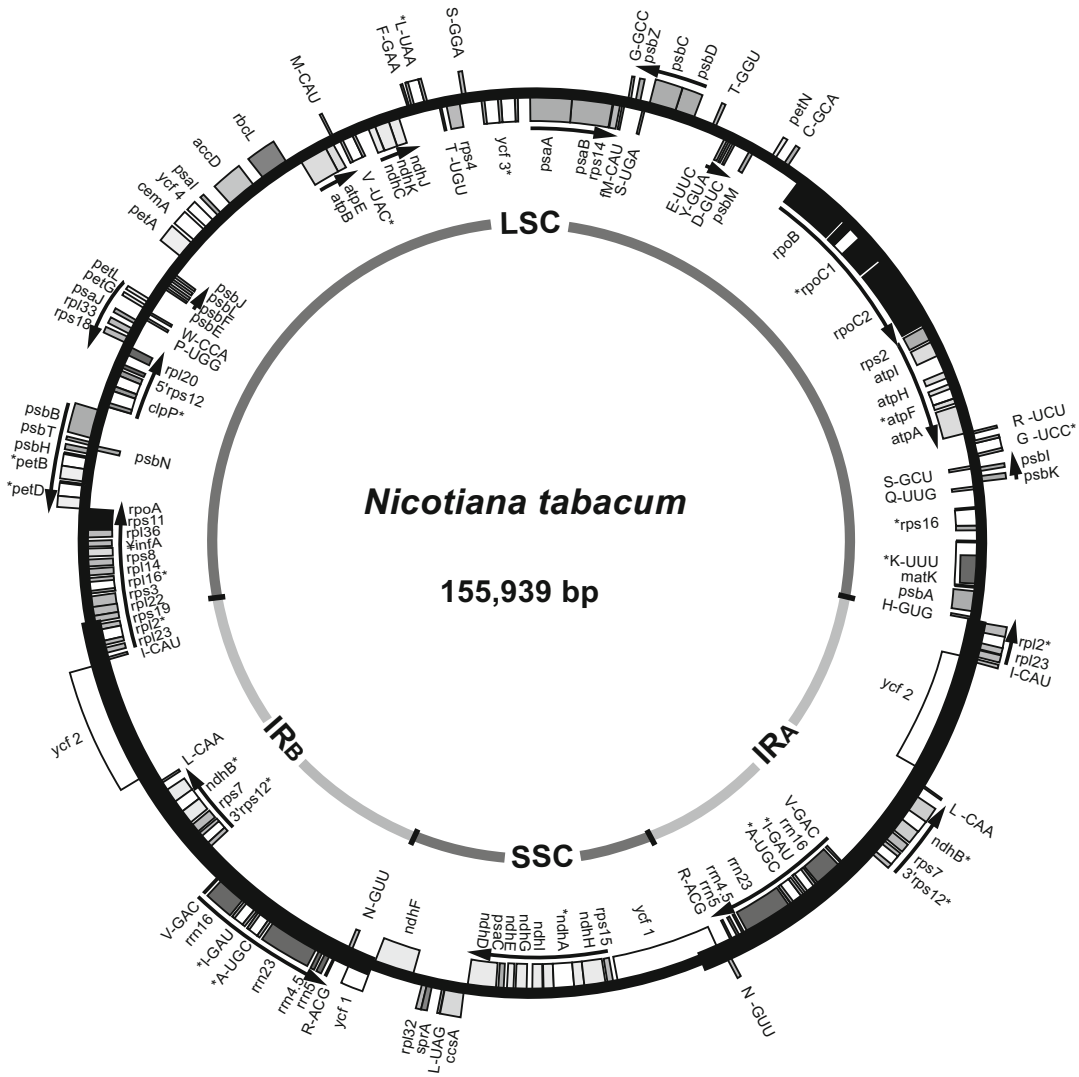


Fig. 1 Circular representation the physical map of the *Nicotiana tabacum* plastid genome (NC_001879). The *N. tabacum* plastid genome is representative of the ancestral gene order that is seen among unrearranged angiosperms. Genes annotated inside the circle are transcribed clockwise while those outside are transcribed counterclockwise. Arrows indicate polycistronic transcription units. Introns are annotated as open boxes and genes containing introns are marked with asterisks

Like those of its prokaryotic predecessor, a number of plastid genes are cotranscribed from operons-like units. Polycistronic precursors often contain transcripts encoding proteins involved in similar functions or subunits of higher order complexes, reflecting the need for concerted regulation of expression. Some plastid polycistronic RNAs, however, are multifunctional such as the ubiquitous ribosomal operon containing tRNA and rRNA sequences. Others, like the *rpl23* gene cluster (syn. S10 operon), encode polypeptides belonging to different functional complexes.

3 Gene Content

The gene complement of most angiosperm plastomes is highly conserved and the arrangement of sequences collinear (Fig. 2). Genes have been grouped into three general classes: those of the genetic system, photosynthetic components and "others." A comprehensive catalogue of the typical plastome gene content is presented by Bock [9]. Genes that do not fit well into either of the major functional categories include *clpP*, *accD*, *ccsA* and *cemA*. Encoding a subunit of plastid localized ATP-dependent caseinolytic protease (Clp) is *clpP*. Apart from this subunit, all other polypeptides that assemble in the plastid Clp holoenzyme are imported from the cytoplasm; as many as 20 plastid targeted Clp polypeptides are predicted for *Arabidopsis* [10, 11]. Among the metabolic activities found in plastids is fatty acid synthesis. The *accD* gene encodes the beta-carboxyl transferase subunit of acetyl-CoA carboxylase (ACCase), which assembles with three nuclear encoded subunits and is essential for leaf development in *Nicotiana tabacum* (tobacco; [12, 13]). The functional copy of *accD* was predicted to have been lost from angiosperm plastomes at least seven times ([5]; Fig. 3). More recent evaluations of *accD* sequences in atypical plastomes have suggested that several of those previously assessed as missing may in fact be divergent but functional genes. In *Passiflora* [14, 15], *Geranium* [16] and *Medicago* [17, 18] the *accD* sequence is interrupted by tandem repeats, however the reading frames are preserved and all five conserved functional motifs along with the putative catalytic site [19] were found intact. In grasses, the plastid encoded product has been substituted by a single subunit eukaryotic ACCase [20] and a nuclear copy of the prokaryotic *accD*, likely transferred from the plastid, has been identified in transcriptomes of *Trifolium* [21, 22], *Hypseocharis* and *Monsonia* [16]. Required for heme attachment to cytochrome, *ccsA*-encoded Cytochrome c biogenesis protein (CcsA) associates with a number of nuclear proteins forming the thylakoid-bound system II cytochrome assembly machinery [23, 24]. Another integral membrane protein, the product of *cemA* is thought to interact with heme molecules due to homology with characterized heme attachment domains [25, 26]. At present, no functional analyses have been conducted in higher plants but localization studies confirm the product of *cemA* as a polytopic protein of the inner envelope membrane. Inferences have been drawn from analyses in *Chlamydomonas* where the *cemA* product is two times larger than its angiosperm homolog [27]; these studies suggest a role in proton extrusion and promotion of efficient inorganic carbon uptake into plastids.

Plastid protein coding genes are usually named as an abbreviation of the encoded protein's function. Predicted open reading frames (ORF) receive the designation *ycf* (hypothetical chloroplast

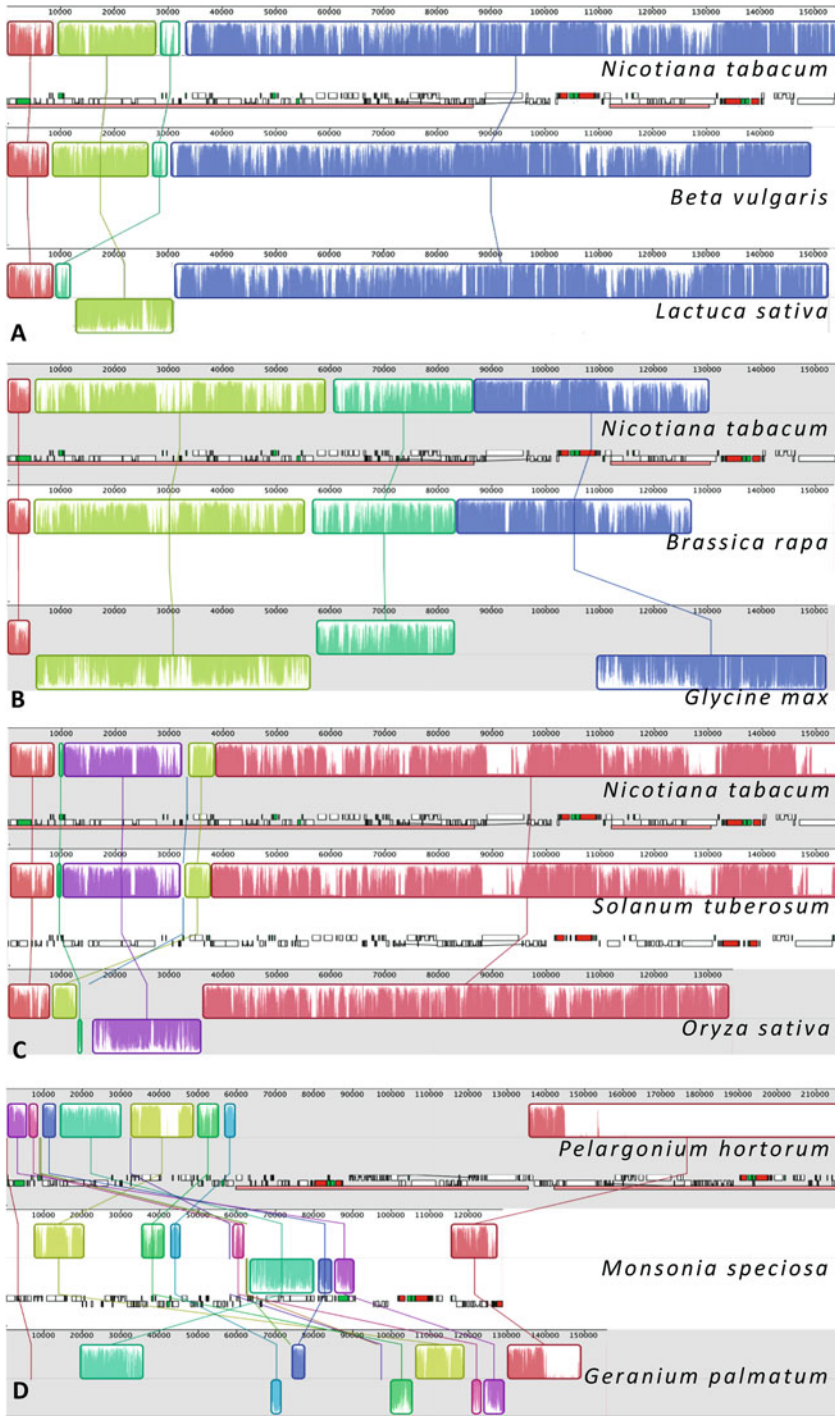
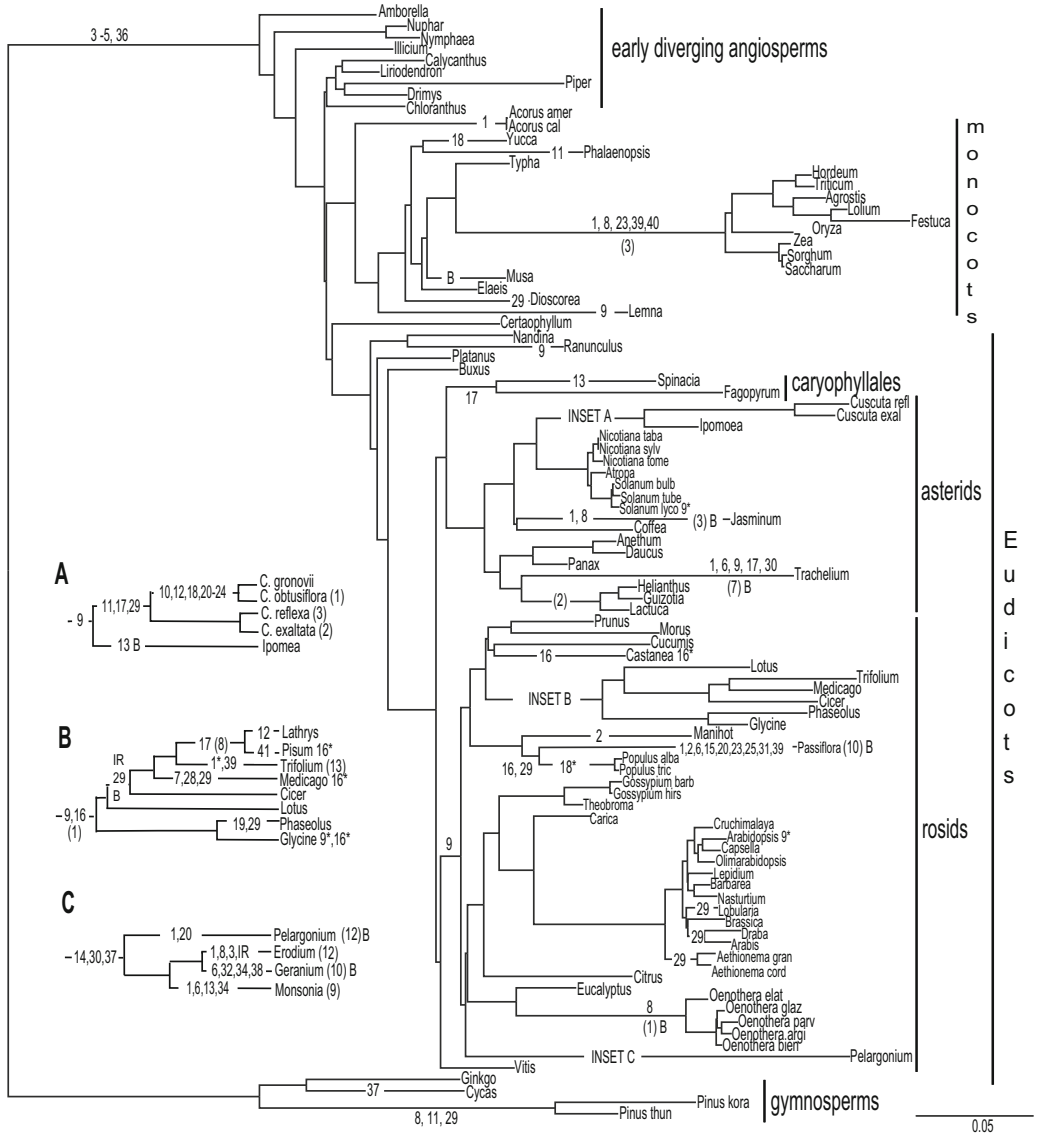


Fig. 2 Gene order comparison of highly conserved and rearranged plastid genomes. Whole plastid genome sequences were downloaded from GenBank for *N. tabacum* (NC_001879), *B. vulgaris* (EF534108.1), *L. sativa* (NC_007578), *Brassica rapa* (DQ231548.1), *Glycine max* (NC_007942), *Solanum tuberosum* (NC_008096), *O. sativa* (NC_001320), *P. hortorum* (NC_008454), *M. speciosa* (NC_014582) and *G. palmatum* (NC_014573). Alignments were performed in Geneious Pro [290] with the mauveAligner algorithm [291], which aligns syntenic blocks of genes and predicts inversions relative to a reference genome. *Nicotiana tabacum* was set as the reference genome in A, B and C. In D, *P. hortorum* was set as the reference



1	accD	10	matK	18	rpl32	26	rps11	34	trnG-UCC intron
2	atpF intron	11	ndhA-K	19	rpl33	27	rps12	35	trnK-UUU
3	chlB	12	psa I	20	rpoA	28	rps12 intron	36	trnP-GGG
4	chlL	13	rpl2 intron	21	rpoB	29	rps16	37	trnT-GGU
5	chLN	14	rpl16 intron	22	rpoC1	30	rps16 intron	38	trnV-UAC
6	clpP	15	rpl20	23	rpoC1 intron	31	rps18	39	ycf1
7	clp intron1	16	rpl22	24	rpoC2	32	trnA-UGC	40	ycf2
8	clp intron 1-2	17	rpl23	25	rps7	33	trnG-GCC	41	ycf4
9	infA								

IR: Inverted repeat loss B: potential for biparental inheritance (#) number of inversions

Fig. 3 Angiosperm phylogeny. The large maximum likelihood phylogram was constructed from 97 taxa based on 81 plastid gene sequences and is adapted from Jansen et al. [83]; Scale bar indicates the increment of 0.05 substitutions per site. Inset trees are cladograms, branches do not represent evolutionary distance. Gene and intron losses are represented by numbers (see key) plotted on branches and are based on sequence homology or divergence, relative to the ancestral angiosperm plastome, as reported in the literature (see references in main text) or examined by the authors. Asterisk indicates reported cases of nuclear transfer, that is, cDNA of nuclear origin detected or fusion product imported into plastids. Note: Not all changes described in the text are plotted on branches

reading frames) until some function can be ascribed to their product. Like those described above, most *ycf*s have been renamed as their role in plastid function has been elucidated. The genes for the conserved Photosystem I assembly proteins *ycf3* and *ycf4*, are yet to be renamed, and *ycf15* is no longer considered to be a protein coding gene [28–30]. Short, nonconserved ORFs are observed in all plastomes but are presumed to be nonfunctional due to the lack of conservation across closely related species and biochemical evidence for their expression [9, 29, 31]. Although many *ycf*s have been characterized and renamed, as yet no definitive function has been attributed to the two largest ORFs present in plastomes, *ycf1* and *ycf2*. Both *ycf1* and *ycf2* have been designated as essential genes because their targeted disruption results in unstable mutant phenotypes [32]. However, *ycf1* is highly divergent and may be lost in several lineages. In *Passiflora* [5, 15] and *Trifolium* [33] the *ycf1* gene appears to be pseudogenized and in all Geraniaceae genera except *Hypseocharis* [34–36] both *ycf1* and *ycf2* are lost as in Poaceae plastomes [37–39]. Other lineages that harbor highly divergent, degraded and pseudogenized copies of *ycf1* and/or *ycf2* include Campanulaceae [40, 41], Ericaceae [42, 43], *Silene* [44] and the saguaro cactus *Carnegiea gigantea* [45]. One suggestion for the product of *ycf1* was a role in plastid protein import as a Tic subunit [46] while others have speculated that its tendency to be lost or highly divergent concurrently with the plastid *accD* gene, as in grasses, Geraniaceae and *Passiflora*, could suggest a role in ACCase holoenzyme assembly [47]. However, the products of these two genes are not known to have any interaction.

Greater than one third of the typical plastome comprises protein coding genes involved in photosynthesis and related reactions. Apart from *ndhB* all of these sequences are contained within the single copy regions; an additional ten single-copy genes encode subunits of the plastid NAD(P)H dehydrogenase (NDH). In several lineages including saguaro cactus [45], Alismatales [48, 49], Orchidaceae [50–53] and *Erodium* [53, 54] the plastid NDH genes are degraded as to be nearly undetectable. Attempts to locate the missing plastome NDH sequences in the nuclear genomes have failed to uncover a functional transfer event. More than 30 nuclear genes encode subunit polypeptides or accessory proteins of the plastid NDH complex, which are translated in the cytoplasm and transported across the plastid membranes where they assemble with locally encoded constituents (*see* [53], Additional file 1; [55]). In all cases examined, including investigations in gymnosperm lineages missing the plastid NDH genes [53, 56, 57], transcriptome [53] and nuclear genome (*Halophila ovalis,* Hydrocharitaceae, Alismatales; [58]) studies revealed the concomitant loss of the nuclear complement suggesting that under favorable conditions NDH function may be dispensable.

There are 22 genes for subunits of photosystems I and II (*psa* and *psb*), six genes each encode subunits of the cytochrome b6f complex (*pet*) and the F_0F_1 ATP synthase (*atp*). With the exception of the RuBisCo large subunit (*rbcL*) all of the genes in this class encode subunits of membrane complexes that are assembled together with nuclear-encoded proteins. The gene for the small subunit of RuBisCo (*rbcS*) is likewise found in the nuclear genome but the holoenzyme assembles in the plastid stroma and does not associate with membranes [59].

The genetic system genes are the largest functional class including 30 tRNAs, four rRNAs, 21 proteins that associate with the large or small ribosomal subunits, four subuints of the plastid encoded RNA polymerase (PEP), intron-encoded maturase K (*matK*; group II intron splicing factor), and translation initiation factor 1 (*infA*).

4 The Large Inverted Repeat

4.1 Genes Are Duplicated in the Inverted Repeat

Most angiosperm plastomes retain a large (~25 kb) inverted repeat (IR); the two copies designated IR_A and IR_B. Genes contained within the IR are duplicated as IR_A and IR_B are thought to be perfectly identical in nucleotide sequence. Usually the genes of the plastid ribosomal operon (*rrn16*, *rrn23*, *rrn4.5* and *rrn5*) and seven tRNA genes are duplicated in the IR. In addition to three complete sequences for ribosomal proteins, the IR also contains *ycf2* and exons two and three of the trans-spliced *rps12* gene. The two IR copies are separated by unequally sized single copy domains, referred to as the large (~80 kb) and small (~20 kb) single copy regions, LSC and SSC, respectively. This arrangement is seen in *N. tabacum* (Fig. 1) and is thought to be the ancestral form for angiosperms [28]. The widespread occurrence of the IR across not only land plants but also in chlorophyte lineages suggests this is an ancient feature. The publication of the *Amborella* plastome [29] has confirmed the IR's presence in the most basal angiosperm. Speculation as to the origin of the IR, and its possible role in plastome stability has fostered decades of research and hypotheses. Kolodner and Tewari [60] detected the presence of the large palindromic sequences in *Lactuca sativa* (lettuce), *Spinacia oleracea* (spinach) and *Zea mays* (corn) using intramolecular homoduplex formation visualized by electron microscopy and reported the absence of this feature in *Pisum sativum* (pea). Subsequent analyses including complete plastome sequences have confirmed the loss of one IR copy in a lineage of papilionoid legumes [22, 61, 62].

4.2 The IR and Recombination

Long before sequencing of complete plastomes became routine, restriction endonuclease mapping was widely used to characterize and compare the structure of angiosperm plastomes [6]. Intramolecular recombination between the IR copies was suggested as a

mechanism to prevent the divergence of the two copies from each other and has the potential to reverse the polarity of the segments between the two copies (i.e., the phenomenon later described by Stein et al. as flip-flop recombination; [63]). The observed difference in conformation of circular dimers (head to tail in pea, head to head 70–80% in *L. sativa* and *S. oleracea*) led to the suggestion that the inverted sequences were also involved intermolecular recombination among the many DNA copies present in plastids.

4.3 Speculation Regarding the Origin of the IR

Based on restriction mapping in *Nicotiana* species, a dimeric circle comprising two complete plastome copies, joined together in a head-to-head (and tail-to-tail) fashion, was proposed by Tassopolo and Kung [64] as a possible origin of the IR. Repeated deletions in an identified "hot spot" mapped to a region down stream of JL_A (junction of LSC and IR_A) extending some distance into the LSC region eventually yielded the IR currently observed. More likely the widely observed variation at IR-single copy junctions results from both expansion and contraction of the repeats [65]. In the case of *N. acuminata*, the abundant plastome sequences now available suggest that a large insertion, not a deletion, in the "hot spot" region caused plastome expansion in this lineage.

In 2010 [66] phylogenetic analyses of the cyanobacterial ancestor of plastids suggested that the order Chroococcales includes the closest extant relative. This placement was refuted in a more recent analysis [67], which suggested a more ancient cyanobacterial origin of plastids. Cyanobacteria, including some of the previously suggested close relatives, have genomes that contain at least duplicated copies of a ribosomal operon (dispersed and/or in reverse orientation). As the rRNAs are the most commonly observed IR genes in plant plastids, a possible origin of the plant IR may lie in a very distant ancestor.

Whatever its origin, the presence of the IR is deemed ancient, and certainly predates the branch eventually leading to angiosperms. Across most angiosperms gene content and position of the IR are highly conserved. Usually around 25 kb, variation in the size of the IR, which can range from a little as 7 and up to ~88 kb, often accounts for plastome size variation overall [34]. The most dramatically wide variation is seen in a small but fascinating minority whose study may yield insights into the mechanisms involved in expansion and contraction of the repeats.

4.4 IR Persistence and Loss

Despite its persistence in the vast majority of angiosperm plastomes, the IR does not appear to be essential as one copy has been lost several times in eudicots. Within Fabaceae, this molecular synapomorphy defines the Inverted Repeat-lacking clade (IRLC; [68]) that includes *Trifolium*, *Pisum*, *Cicer* and *Medicago*, a designation that is strongly supported by independent phylogenetic analyses of molecular data. Furthermore IR loss within this clade

is unambiguously confirmed by assembled and annotated complete plastome sequences [21, 22, 33, 69]. Although not essential, IR loss may have had an impact on plastome evolution in the IRLC. Acceleration in the rate of synonymous nucleotide substitution in formerly IR genes and changes in gene order are observed in legume plastomes where the IR is lost [70–73].

At least three independent losses are reported within the Geraniaceae where the ancestral plastome was inferred to contain an IR [36]. Although no supporting data were available at the time, Palmer and Downie [74] inferred IR loss in *Monsonia* (formerly *Sarcocaulon*) and *Erodium* based on hybridization of small probes with sequences surrounding IR-single copy junctions. More recent publications based on complete plastome sequences have revealed several losses across three genera, *Erodium*, *Geranium* and *Monsonia* [34, 36, 75–77]. Plastome sequences have been generated for *M. vanderietieae*, the species reported earlier as lacking the IR [74] but as yet the plastome is not completed. A number of clades within Geraniaceae exhibit unprecedented reorganization of plastome sequences making assembly of finished genomes a challenging endeavor. Also reconfigurations in gene order can obscure data interpretation when hybridization approaches are employed for IR detection. Among the data from the *M. vanderietieae* draft plastid genome, two large repeats of at least 3 kb were detected [78] suggesting that the IR may not be completely lost, rather just hiding. *Monsonia speciosa* retains a modest IR ([34]; 7313 bp), encoding four protein genes and the ribosomal operon, but excluding *rrn16*, and two additional species, *M. emarginata* and *M. marlothii* lack the IR entirely [76, 77].

While most *Erodium* species sequenced thus far show loss of the IR, species of the long branch clade (LBC) appear to have developed novel IRs that include the ribosomal operon and other typical IR genes [75]. Whether there were two independent losses or a single loss and regrowth of the IR in LBC taxa remains unresolved for *Erodium*. Another instance of IR regrowth was detected in the IRLC where there is less ambiguity as to gain or loss. A novel IR was uncovered in *Medicago minima*, which is placed well within the IRLC, a monophyletic group defined by the single shared loss of the IR. Among the 19 *Medicago* species completed, just one contained a novel IR of >9 kb [18].

Two additional reports describe IR losses, one in the saguaro cactus (*Carnegiea gigantea*; [45]) and the other in *Tahina spectabilis* [79]. The loss in saguaro resulted in the smallest plastome among photosynthetic angiosperms detected to date (~113 kb; [45]). The loss in *Tahina*, however, remains ambiguous as technical rigor regarding identification and confirmation of the loss were lacking [79].

Although the synonymous substitution rate was accelerated in IRLC taxa for formerly IR genes, the situation was more complex in *Pelargonium*. While all sequenced *Pelargonium* contain an IR,

the boundaries vary greatly between clades within the genus (IR size range ~36–88 kb; [35]). An examination of evolutionary rates in 22 species where typical IR genes were excluded or typical single copy genes included in the IR suggested that rate heterogeneity in *Pelargonium* plastomes was a mixture of locus-specific, lineage-specific and IR-dependent effects [35].

Among the asterid nonphotosynthetic parasites there are two reported cases of IR loss. The Orobanchaceae holoparasite *Conopholis americana* was predicted by one group to contain [80], and by another to have lost the IR [74]. Recently assembled plastome sequences confirm the *C. americana* loss; however, a definitive loss cannot as yet be confirmed for *Striga asiatica*, also in the Orobanchaceae. Again assembling complete sequences for species that contain highly reorganized plastomes make assignment of particular characters challenging. As yet the loss in *S. asiatica* has not been confirmed however three other species of *Striga* were shown to all contain expanded IRs over 60 kb [81]. In addition to *C. americana*, IR loss has been confirmed for *Phelipanche ramosa* (>62 kb). The nearest relative examined in this study, *P. purpurea*, was found to have a small IR, perhaps accounting for the 1182 bp difference in plastome size between the congeners [82]. These findings suggest at least two independent losses within the Orobanchaceae. If data analysis ultimately yields resolution in favor of IR loss in *Striga*, this will represent a third, independent loss within the family.

IR loss is far from a universal feature among the parasites. *Epifagus virginiana*, which has the second smallest sequenced angiosperm plastome at ~70 kb, has experienced extreme reduction in the single copy regions due to the loss of the photosynthetic genes while retaining an IR of ~25 kb [83]. The reduced IR (9767 bp) found in the subterranean orchid *Rhizanthella gardneri*, plastome (~60 kb) uniquely lacks the rRNA genes and is the primary reason for the size difference between the two smallest angiosperm plastomes characterized to date [84]. Within the same family, in the *Neottia nidus-avis* plastome, the IR is thought to have undergone expansion relative to other heterotrophic orchids contributing to its larger plastome of ~92 kb [85].

4.5 The IR and Plastome Stability

The presence of the IR may play a role in structural stability of the plastome and is involved in maintaining the homogeneity of the sequences encoded in each copy. Differential rates of nucleotide substitution between genes of the single copy regions and those of the IR [38, 39, 73, 86, 87], along with the thinking that both copies display identical sequence, suggest an efficient gene conversion mechanism is at work. Such a mechanism has been invoked to explain copy correction in plastid transformants when point mutations or foreign sequence are introduced into the IR [88–90].

Following a mutation or transgenic integration event, although replication and partitioning may contribute, homoplasmy is likely to be driven predominantly by recombination processes. As

illustrated by Khakhlova and Bock [91], gene conversion proceeds very rapidly and early in regenerating callus and shoots under selection. While biased gene conversion may favor wild-type alleles where mutations are neutral [87], deleterious mutations may be removed or resistance alleles (i.e., transgenic resistance alleles) fixed under selective conditions yielding homoplasmy [91, 92].

For some groups differential rate of nucleotide substitution between the single copy regions and IR is well established [39, 73, 86, 87, 93]. In these studies, IR genes experienced significantly lower rates of substitution with some variability depending on the species and genes examined. Comparison of synonymous substitution rates (d_s) of six formerly IR-located genes in IRLC plastomes with those in IR-containing relatives revealed that d_s was higher in the IRLC genes. In the IR-containing plastomes single copy regions genes displayed d_s that is 2.3-fold higher than IR genes while in the IRLC, d_s was uniform across the entire plastome [73]. Although d_S was consistently higher across the plastome in IR-containing papilionoid taxa compared with IRLC and nonpapilionoids, the IRLC had the highest d_S across genes that were formerly in the IR. However, in agreement with Perry and Wolfe [73] the rate of formerly IR genes was not significantly different than other genes in IRLC taxa [93].

The homogenizing effect of gene conversion in the IR is demonstrated both theoretically and experimentally. It remains unknown if there is a significant difference in the frequency of IR recombination within or between plastome molecules, nor is there much information about intra- or intermolecular recombination between single copy regions. It has been long acknowledged that single copy inversion isomers, the products of the so-called flip-flop recombination [63], are present at roughly equimolar proportions [60, 62, 94–97]. Given that these isomers are not produced from replication initiated by IR recombination within a single plastome unit suggests a mechanism that requires the interaction of *different* copies of the unit genome. Small inversions mediated by intramolecular recombination events have played a major role in the evolution of pseudogene diversity and can be useful for increasing phylogenetic resolution at the species level [98–101]. One could speculate that simple proximity facilitates more rapid and frequent recombination between IR copies in the same molecule but given the various hypotheses regarding the physical form of ptDNA previous notions of what constitutes intra- versus intermolecular may have to be reconsidered.

We know that transplastomic plants have been generated by targeting insertion to the LSC, and that these lines reach homoplasmy [90]. The very few reports that describe gene conversion with regard to single copy genes in wild-type plants are based on examination of sequence data alone. Maintenance of an *rpl23* pseudogene situated in the LSC by gene conversion with the functional

allele present in the IR has been suggested for representatives of the grass family [102]. In this study, limited nonparallel sampling for the functional copies and pseudogenes makes the proposal of maintenance for this pseudogene, by any mechanism, highly speculative. The polyploid state of the plastome would provide ample template for intermolecular gene conversion to restore the wild-type sequence following a mutation and may be one of the ways plastomes have escaped Muller's ratchet [103]. In any case we can be certain that intermolecular recombination occurs in plastids as evidenced by the transplastomic plants that arise through site-specific integration of foreign sequences by homologous recombination with the flanking sequence on the transformation vector.

Presence of an IR does not ensure genome stability. The IR, while nearly omnipresent, undergoes seemingly constant expansion and contraction through evolutionary time [65]. This can be observed not only in numerous examinations of the junction sites [39, 85, 104–107] but also in the plastomes where the IR is massively expanded or drastically reduced [15, 34, 35, 84, 108–111]. Within the family Oleaceae in the Jasmineae tribe IR length is somewhat expanded relative to most angiosperms (29 kb), and moderate rearrangement is seen. Arrays of repeats, duplicated sequences inserted into coding and noncoding regions, and gene and intron loss are reported [112]. Two overlapping inversions resulted in the relocation of the *ycf4–psaI* region in Jasmineae plastomes while this feature and other peculiarities are not seen in the tribe that includes cultivated olives (Oleaceae; [113]). The *ycf4-psaI r*egion was recently described as hypermutable among IRLC members [21, 22]. An extreme locus-specific elevation in nucleotide substitution coupled with length mutations in the same region led the authors to speculate that repeated DNA breakage and, presumably error prone, repair was involved.

Within or between monomers, recombination plays an integral and ongoing role in plastome maintenance, with regard to both structural and sequence variation. The recombination-dependent replication (RDR) DNA repair pathway could account for inversions and drive repeat accumulation in plastomes with large non-IR repeats [95, 114]. *Monsonia emarginata* contains a number of large repeats, in fact ~22% of its plastome sequence comprises repeats ranging from ~1 to >3 kb making it a candidate in which to detect alternative plastome arrangements derived from RDR. PacBio SMRT long read sequencing identified alternative plastome arrangements in *M. emarginata* that appeared to be mediated by recombination between long repeats. PacBio reads that included one copy of a large repeat and adjacent sequences revealed variation of adjacencies indicative of alternative sequence arrangement around the repeats [76]. Examination of the novel IR in *Medicago minima* suggested a role for RDR in generation of the duplicated, inverted sequence identified there [18]. A shift in the thinking

about the physical from of plastome molecules away from monomeric circles to a collection of large and dynamic molecules containing several to many copies of the unit genome should foster realistic hypotheses about mechanisms shaping plastome structural evolution.

5 Synteny Across Plastomes

Plastid genes tend to be collinear across a broad range of angiosperms, and simple, phylogenetically informative inversions can be useful for the resolution of relationships, especially at deep nodes in the tree of life [115–117]. There are, however, several cases where numerous overlapping inversions, insertions and deletions (indels), and expansions and contractions of the IR result in such reconfiguration of gene order as to confound prediction of evolutionary relationships.

5.1 Genomic Rearrangements

In the Campanulaceae, Cosner et al. [118] used restriction site/gene mapping to evaluate the distribution of gene order changes among 18 members of this family relative to *N. tabacum*. Across the 18 taxa, so many gene order changes were observed that inference of the individual events within the family was highly problematic. However, despite this complexity, the resulting gene order phylogenies exhibited low levels of homoplasy and were congruent with independent trees generated from DNA sequences for the same taxa.

Nowhere among photosynthetic angiosperm plastomes are genomic rearrangement and substitution rate acceleration more dramatically illustrated than in the family Geraniaceae [36]. As mentioned, genera within the Geraniaceae contain both the largest (*Pelargonium*) and smallest (*Monsonia*) IRs of any reported to date as well as IR-lacking lineages (*Erodium*, *Geranium*, *Monsonia*). Massive accumulation of dispersed repeats, often associated with changes in gene order, is observed in across the family excluding the monotypic genus *California*. Also highly unusual is the disruption of two conserved transcription units within the family and the presence of rRNA genes outside the IR in *Monsonia*. The presence of large repeats (>100 bp) in *Pelargonium*, *Erodium*, *Geranium* and *Monsonia* is unprecedented. Sequence homology indicates that some of the large repeats have arisen from full or partial duplication of genes. Many of the duplicated sequences, along with sequences of protein and tRNA genes commonly present in plastomes, are predicted to be pseudogenes [34]. Members of the long branch clade (LBC) of *Erodium*, unlike sister species *E. texanum*, are further lacking functional genes encoding all subunits of NAD(P)H [119].

Estimating the minimum number of events required to arrive at the observed configuration, Chumley and colleagues [110] proposed eight IR boundary changes, 12 inversions and several insertions of duplicated sequence for the *Pelargonium* × *hortorum* plastome. Indels can arise readily in the face of IR expansion and contraction. As a change occurs at an IR boundary, that change will be reflected as an indel in the other IR copy through copy correction by gene conversion. However attractive it may be to attempt, reconstructing evolutionary history based on the number of rearrangement events required to convert one genome to another would be risky when considering plastomes like those in the Geraniaceae. In this case it may be impossible to accurately estimate the number of genomic changes or the order in which they occurred. Perhaps a more reliable approach to reconstructing the evolutionary history of genomic changes would be to compare genomes of very closely related species (i.e., [35, 120]) or even different populations of the same species in the lineages that exhibit such a high degree of rearrangement.

As it turns out when species that are thought to be closely related display substitution rate acceleration and structural changes in plastomes accurate phylogenetic tree topologies remain obscure and may suggest that such lineages have undergone recent and rapid divergence. Although not as extreme as seen in Geraniaceae, genomic rearrangement and elevated nonsynonymous substitution rates have been documented in *Silene* plastomes [107]. Four species were recently examined and while two, *S. latifolia* and *S. vulgaris,* appeared to retain ancestral plastome characteristics, two others, *S. noctiflora* and *S. conica*, were estimated to have experienced four and one repeat-mediated inversions, respectively. In all four species *infA* and *rpl23* are pseudogenized and in *S. noctiflora* and *S. conica matK*, *rpoA* and *accD* have diverged and may also represent pseudogenes. These two species have lost the introns in *rpoC1* and both of the *clpP* introns while *S. noctiflora* has lost the *rpl16* intron, *S. conica* has lost the intron of *atpF*. All of the intron losses noted in *Silene* plastomes have been documented in other angiosperm lineages as well (Fig. 3).

A positive correlation has been noted between nucleotide substitutions and genomic rearrangements ([5, 34, 93, 107]; Fig. 3). At first glance this seems fairly intuitive. For example in plastomes such as those described, a fully or partially duplicated gene sequence is inserted at some alternate locus (a genomic rearrangement). It may be separated physically from the sequences that regulate its expression, it may be truncated; for whatever reason it is nonfunctional. While the functional copy is maintained, over time the duplicated sequence experiences an accelerated rate of nucleotide substitution because there is reduced selective constraint on its divergence. This source of substitution rate acceleration is not usually examined, however. Comparisons are made based on

intact genes and duplicated sequences are not included in such analyses suggesting other explanations for the observed correlation. Aberrant DNA repair pathways, such as those involved in double strand break repair, and RDR are implicated in genomic rearrangement and could further contribute to divergence in nucleotide sequences, particularly in repeat rich plastomes (i.e., *M. emarginata*; [76]).

6 Plastome Gene Loss

6.1 Defining Gene Losses

The *functional* loss of plastome encoded sequences occurs by various means. Authors have often described this situation with little regard to the mechanics that yield a loss of function. There are numerous examples where plastome gene losses are reported in angiosperms, and in only a few cases, functional replacement by nuclear homolog is suggested or demonstrated (Fig. 3) [16, 21, 22, 121–125]. Gene loss may include pseudogenization due to nonsynonymous substitution resulting in amino acid changes that impair the functional capacity of a protein, introduction of stop codon or elimination of start or stop codons. Descriptions of gene loss based on substitution by sequence analysis typically do not consider changes in regulatory regions such as promoters and 5′ or 3′ untranslated regions (UTRs), or intron sequences, that could potentially disrupt expression of the gene product yielding a functional loss. Apart from pseudogenization of coding regions by nucleotide substitution there is gene loss due to indels. Insertions into coding and regulatory regions or deletions that compromise regulatory function, introduce frameshift mutations or, in the case of deletions, remove entire coding regions constitute another class of gene loss. Among the Campanulaceae, the Geraniaceae and in *Trifolium*, genome rearrangements have resulted in disruption of transcription units that are highly conserved in other angiosperms [33, 34, 40, 75, 76, 112]. By their definition the genes of a transcription unit are cotranscribed from a promoter situated upstream of the most distal gene at the 5′ end. While the coding regions of the "stranded" genes appear to be functional by sequence inspection, whether and how these genes are transcribed and regulated remains a mystery. Interestingly, a recent analysis of transcriptional start sites using differential RNA sequencing has identified numerous examples where plastid genes exhibited independence from their canonical operon promoters [126]. In any case, assignment of functional status to plastid encoded, or plastid targeted, genes must be made with a caveat; conclusions reached through sequence inspection should be supported by biochemical evidence.

6.2 Old and New Gene Loss

Notwithstanding the preceding discussion, recent gene losses reinforce the notion that the plastome is dynamic, still following its evolutionary trajectory while adhering to the constraints of its environment. Recent events include the loss of *ndh* genes in *Erodium*, more than a dozen independent ribosomal protein gene losses and, in one clade of the parasitic genus *Cuscuta*, the loss of *psaI*, *matK* and the entire gene suite for subunits of the plastid encoded RNA polymerase. In addition to protein coding genes, recent losses are also reported for tRNA genes. The loss of the *ndh* genes in Orchidaceae [52, 53, 127]; *accD*, *ycf1* and *ycf2* in Poaceae, and *infA* in all but one rosid represent older events, but still occurred long after the origin of angiosperms (Fig. 3). The early and rapid transfer of genes to the nucleus following endosymbiosis [1] reduced the coding capacity of the plastome to less than 5% of the ancestral genome by the time the angiosperm lineage originated. The analysis of nuclear genomes from evolutionarily distant *Arabidopsis* and *Oryza* reveals that large tracts of ptDNA have been integrated into these genomes. One insert in *Oryza sativa* (rice) represents a nearly complete copy of the plastome possessing very high identity (>99%) to the sequences of the plastid genome [128–130]. With the raw material present in the nucleus, is it possible for any transferred gene sequence to eventually gain the appropriate features and functionally to replace its plastid counterpart? Lloyd and Timmis [131] have recently observed three distinct mechanisms by which a gene of plastid origin may become activated as a nuclear integrant and suggest that such phenomena are more common than previously thought. Nonetheless, the early processes involved in acquisition of sequence elements conferring transcriptional activity and plastid localization are likely to be stochastic and therefore rare events.

6.3 Gene Retention

The persistence of organellar genomes has inspired a number of theoretical hypotheses (reviewed in [132–134]). Recent and/or repeated functional transfer events, that is replacement of a plastome encoded gene product with an imported, nuclear encoded gene product, seem to favor ribosomal proteins or proteins whose activity is not intimately linked to photosynthesis. According to the CORR (**co**-location for **r**edox **r**egulation) hypothesis proposed by Allen [135] the nonrandom sample of genes that are ubiquitously retained in photosynthetic plastomes link supply and demand at the level of regulation of gene expression within plastids. More precisely, this core set of genes provides plastids with a mechanism to respond rapidly and directly to the local redox environment. The CORR hypothesis does not address the need for import of many more nuclear encoded gene products to execute transcription, translation, photosynthesis, and related reactions. The absence of *rpoA* from the moss plastomes where the α-subunit of PEP is imported from the cytoplasm [136] weakens the so-called

autonomy of plastids in terms of gene expression, a central feature of the CORR theory. More recently Wright et al. [137] have proposed that organellar genomes could themselves serve as long term redox damage sensors by providing a tonic, or constitutive, retrograde signal. Sensing oxidative damage to ptDNA molecules as a means to activate control mechanisms could, theoretically, be perpetuated through any subset of genes as damage and repair pathways should not preferentially favor the specific subset retained in contemporary plastomes. Retention of the *rbcL* gene in plastids defies at least one assumption of hydrophobicity hypotheses; it is not a membrane protein with hydrophobic domains. Transgenic experiments have shown the large subunit can be imported by plastids from the cytoplasm following nuclear expression [138]. The concept of ordered assembly, outlined by Daley and Whelan [134], offers another plausible motivation for retention of photosynthetic genes in plastomes. The study of multisubunit protein complexes, such as photosystem II, has shown that assembly proceeds in an ordered manner, with the sequential addition of specific scaffolding and functional subunits [139]. Again the problem that arises is that plastomes do not encode the entire complement of genes required for photosystem assembly [140]. In this case perhaps retention of at least some of the subunit genes in the plastome assures plastid localized assembly.

It is quite possible that there are a variety of reasons, several of which have been proposed, why some genes as opposed to others presently remain plastid encoded. The reasons may be different for different genes or subsets of genes and the underlying mechanisms combinatorial. Given the small window we have opened into the evolution of plastomes it is likely that we have but scratched the surface and continued study will reveal further intricacies governing the subcellular localization of genes. With the availability of complete nuclear genome sequences come new opportunities to employ bioinformatic approaches to study the fate of plastid genes in the nucleus.

7 Plastid Genes in the Nucleus and Other Promiscuous Sequences

7.1 Transcriptional Activation in the Nucleus

Using data from the most current annotation of the 12 *O. sativa* chromosomes and the TIGR Database of Rice Transcript Assemblies, Akbarova et al. [141] identified cDNAs of nuclear origin that include the complete sequences of five plastid genes: *atpI*, *psbJ*, *psbL*, *rpl14* and *rps7*. This finding may be indicative of an intermediate stage in the evolution of the *O. sativa* plastome, which currently contains apparently transcriptionally active copies of these five genes. Both prokaryotic and the single subunit eukaryotic ACCase may be functioning in *Monsonia*. The *accD* gene appears to have been functionally transferred to the nucleus where it has

acquired the sequences for expression and plastid targeting. A duplication of the gene encoding the eukaryotic enzyme was detected in all examined Geraniaceae except *Hypseocharis*, and one copy is predicted to contain plastid targeting peptide [16]. This situation fits perfectly with the paradigm for plastid genes functionally transferred to the nucleus, as suggested by Timmis et al. [2]. Theory predicts that there would be a period during which both the nuclear and plastid encoded products were active in plastids, a period of functional redundancy. During this time selection would act to favor one copy over the other; perhaps one copy would be lost or potentially undergo mutation and adopt another functional role in the cell. If the plastid copy were lost this would constitute a functional nuclear transfer event. Such an intermediate state of functional redundancy is supported by the finding that while plastid encoded *rps16* continues to be transcribed and spliced in species representing at least four genera, a mitochondrially derived *rps16* gene now located in the nucleus possesses a dual targeting sequence capable of delivering its product into plastids [142].

Far outside of plants, in the photosynthetic protist *Paulinella*, examination of 3000 expressed sequence tags revealed that the *psaI* gene has been relocated to the nucleus in a recent transfer event. Here the chromatophore (analogous to the much older plastid) copy of *psaI* is easily detectable but has been interrupted by two stop codons within the former coding region [143], while ten other genes for photosystem I (PSI) subunits are encoded in the *Paulinella* plastome. Although unable to locate the nuclear copies, Magee et al. [21] report the loss of the PSI genes *ycf4* and *psaI* from *Pisum* and *Lathyrus*, respectively. Until now the only reported loss of photosynthesis related genes among angiosperms had been in members of the nonphotosynthetic heterotrophs of Orchidaceae and *Cuscuta* and *E. virginiana*.

7.2 Promiscuous ptDNA and Experimental Artifacts

From inspection of the available angiosperm nuclear genomes it is now clear that both short and long tracts of nuclear ptDNA, or NUPTs, are present and retain high identity to plastome sequences [2, 129]. One investigation revealed that intact open reading frames of plastid origin could be maintained in the nucleus within NUPTs for several million years [144].

In addition ptDNA insertions are found abundantly in mitochondrial DNA (mtDNA) [145–149]. These transferred sequences present a problem that must be recognized and managed when analyzing amplified ptDNA for phylogenetic reconstructions. Duplications such as those uncovered in *O. sativa*, where 53 plastid genes were present in multiple copies throughout the 12 chromosomes [141], could result in PCR amplification of nontarget sequences. This phenomenon has been evaluated with regard to spurious amplification of mtDNA from nuclear inserts (NUMTs)

and a recent comment by Arthofer et al. [130] offers suggestions about how artifactual phylogenetic inferences may be avoided when ptDNA sequences are considered. Careful consideration of the potential to amplify spurious sequences should inform primer design and novel approaches, such as the application of methylation sensitive PCR to distinguish nuclear sequences from organelle as described by Kumar and Bendich [150], may alleviate some of these issues.

The so-called promiscuous DNAs can confound molecular analysis in transplastomic experiments as well. The plastome, which persists in all plastid differentiation types, may be present in up to 100 identical copies per plastid in leaf cells [151]. Mesophyll cells of a mature leaf may carry up to 100 chloroplasts with the result that the plastome can comprise a significant portion (up to ~20%; [152, 153]*)* of cellular DNA content. For the establishment of stable transplastomic lines homoplasmy, the iteration of the transgene throughout all of the copies comprising the plastome, must be achieved. In the case of deletion studies, it is desirable to demonstrate homoplasmy of the transplastome as evidence that the mutation involved a nonessential gene [12, 32, 154, 155]. The mixed genotype is unstable and this condition, known as heteroplasmy, does not persist in the absence of selection, resolving into one state (wild type) or the other (transformed) over time [156]. Culturing on nonlethal selective media can facilitate the transformation of all copies. Plastids carrying a resistance marker, and in turn the cells that harbor these plastids, are preferentially maintained as plastome molecules are divided up between daughter plastids and subsequently as plastids are partitioned between daughter cells at mitosis [157, 158]. The same applies to the maintenance of a wild-type allele when the experimental design includes disruption of an essential gene.

The homoplasmic state of putative transformants is typically evaluated by Southern analysis of restriction fragments that differ between the wild type and transplastome, PCR amplification across the insertion/deletion site or by germination of seed from self-fertilized transplastomic plants. Given the high identity and length of NUPTs, hybridization experiments, and PCR in particular due to its sensitivity, could lead to the an incorrect assignment of heteroplasmy as the NUPTs will more likely produce a result corresponding to the wild-type plastome [159, 160]. Analysis of transplastomic *N. tabacum* carrying a mutation in the *psbZ* gene found persistent heteroplasmy even when cultured in sucrose media, thereby alleviating the demand for photoautotrophy. This led investigators to propose an essential, nonphotosynthetic function [161, 162]. By analyzing ptDNA purified by pulsed-field gel electrophoresis, Swiatek and coworkers [163] subsequently demonstrated homoplasmy of the transplastome for mutant line for *psbZ* and *cemA*, but not the essential *ycf1* mutant. In this study

total genomic samples and ptDNA prepared by gradient centrifugation were shown to contain promiscuous ptDNA sufficient to produce misleading PCR results. Furthermore, quantitative evaluation of template abundance responsible for promiscuous *cemA* signals suggest multiple copies may be present in the nucleus and/or mitochondria [163].

7.3 "Foreign" DNA in Wild-Type Plastomes

The transfer of DNA among the genome bearing compartments of individual cells has shaped the evolution of plastomes. Plastomes were thought to be recalcitrant to the incorporation of foreign DNA, either by horizontal or intracellular transfer. As more plastome sequences are obtained bona fide cases of foreign DNA in plastomes have been uncovered. The first, from Goremykin et al. [147], was uncovered during the sequencing of the *Vitis vinifera* (grape) mitogenome. Approximately 1.5 kb of nonplastid sequence was identified in an intergenic region of the *Daucus carota* IR (Apiaceae, carrot; [164]) that contained two regions showing high similarity to grape and other published mitogenomes. The integration of carrot mtDNA in the plastome was confirmed and the insert was identified across the Daucinae clade and in *Caucalis platycarpos* (Torilidinae; [165, 166]). Also in Apiaceae, the plastomes of *Petroselinum* and *Crithmum* contain DNA sequence with no identity to angiosperm plastome sequences situated between the boundary of IR_A and the 3′ end of *trnH-GUG* [167]. The suggestion that insertions of mtDNA into the plastome could involve the IR was reinforced by the detection of an insertion (2.7 kb) in the *trnI-CAU-trnL-CAA* spacer in the herbaceous bamboo genus *Pariana* (Poaceae, Olyreae). This insertion returned hits with high similarity to angiosperm mitochondrial sequences in blastn searches. The best scoring hit was to the mitogenome of another bamboo, *Ferrocalamus rimosivaginus* (Arundinarieae), and covered 97% of the insertion sequence [168]. The *trnI-CAU-trnL-CAA* spacer may be a hot spot for these insertions in Poaceae as mitochondrial inserts 976–11,393 bp were reported for the same region in the *Paspalum* lineage [169]. Although an insert in the *Paraneurachne trnI-CAU-trnL-CAA* spacer bore no similarity to the *Paspalum* insert, the 2808 bp sequence shared 92% identity to the *Sorghum bicolor* mitochondrial genome [169]. A large mtDNA insertion (6767 bp) in the completed plastome of *Anacardium occidentale* (cashew) was also identified in the IR. Primers designed to amplify the inserted DNA in other *Anacardium* species revealed that the insert is present in four taxa that occur in a strongly supported clade, suggesting a single shared insertion event [170].

Thus far a single instance of mtDNA insertion has been reported for the plastome single copy regions in a location well removed from the IR and its boundaries. A large insertion was identified in the *rps2-ycf4-psaI 2-rpoC2* IGS of *Asclepias syriaca*

that produced best blast hits to a region of the *Nicotiana tabacum* mitogenome [171]. Later the same year, Straub et al. [172] confirmed the plastome insertion of a 2.4 kb segment of the *Asclepias* mitogenome. Using complete plastome sequences and a PCR based survey, inserts ranging in size from ~2.4 to ~4.7 kb were identified across the tribe Asclepiadeae but not in other Apocynaceae.

8 Recombination and the Plastome

The generation of novel plastome genotypes by transformation initially relies on integration of foreign sequence by intermolecular homologous recombination (HR; [89, 173]). Mechanistically related to gene conversion, HR and repair pathways participate in the subsequent events that yield homoplasmic transplastomic cells and eventually stable transplastomic plants. Intra- or intermolecular recombination between repeated sequences, both in wild-type plastomes [12, 34, 40, 100, 174, 175] and as a result of transplastomic experiments [154, 176, 177] can generate inversions when repeats are palindromic, or deletions when direct depending on the loci and mechanisms involved. Furthermore the role of HR proteins in damage repair may be compromised when foreign DNA is introduced, and through associated tissue culture and selective pressure, as these manipulations can place additional stress on recombination machinery leading to unintended events [95, 178]. Numerous recombination products have been observed during the development of transplastomic lines [177, 179] and likely vastly more remain either unexplored or unreported as such recombination products are undesirable to biotechnologists.

8.1 DNA Replication, Repair, and Recombination

Among the DNA repair and recombination genes identified in the nuclear genomes of *Oryza* and *Arabidopsis* TargetP [180] prediction suggests 19% and 17%, respectively, are targeted to plastids [181]. As yet only a small number of such activities have been explored.

8.1.1 RecA Protein

Early work in *Pisum sativum* demonstrated DNA strand transfer activity for plastid localized RecA [182, 183]. Further study implicated RecA in recombination mediated repair of damaged ptDNA [184]. Using *Arabidopsis* mutant lines generated by T-DNA insertion Rowen et al. [185] showed that reduced RecA1 (AT1G79050) activity leads to a destabilization and reduction in ptDNA. The reduction in plastome copy number in mutant lines relative to wild type suggests that RecA1 participates in recombination-mediated replication.

8.2 DNA Replication, Physical Form and Copy Number

It is likely that ptDNA replicates by a number of mechanisms acting concomitantly and/or sequentially along a developmental continuum. Proposed models include displacement-loop and rolling circle (θ and σ; [186]) replication of circular templates, and recombination dependent mechanisms acting on circular or linear templates [94, 187, 188]. There is strong experimental evidence for multiple replication origins mapping to both the IR and single copy regions [189–196]. Several different groups using distinct approaches have demonstrated that ptDNA is present in vivo in many forms. These include monomeric and multimeric circles, linear molecules that can be highly branched forms and condensed, high molecular weight DNA complexes [94, 186, 197–200]. Given the range of potential templates available to the replication machinery it is plausible that different mechanisms predominate in replication of each of the different forms of ptDNA and at different developmental stages.

Replication is thought to be most active in plastids of meristematic tissues and leaf primordia, when cells and their organelles are growing and dividing most rapidly. At these early developmental stages, branched linear molecules were reported to comprise the most abundant topological form of ptDNA in some species, with reducing complexity as leaves develop [138, 150]. Plastome copy number reaches a maximum in young leaves, and appears to be reduced approximately three fold in terms of genomes per plastid in mature leaves well before senescence [94, 200, 201]. Given the abundance of plastids in the cells of mature leaves, this still constitutes a considerable amount of DNA on a per cell basis. Estimation of ptDNA copy number in plastids of maturing leaves of *Medicago*, *Pisum*, *Zea* and *Arabidopsis* [200, 202–204] suggested that a large proportion of mature leaf plastids were completely devoid of ptDNA, having lost earlier populations through degradation. In their examination of isolated plastids of three maize cultivars, Zheng et al. [204] report that during greening of dark grown seedlings, 6 h of exposure to light induced ptDNA instability resulting in the loss of half of the ptDNA that was present before transfer to the light. By 24 h of light exposure ptDNA abundance was equal to that of the mature plastids of light grown plants, which were found to contain tenfold fewer plastome copies than etioplasts. The authors propose that the high plastome copy numbers seen prior to the initiation of photosynthesis are needed to facilitate accumulation of plastid encoded proteins. Once photosynthesis is initiated, this demand is ablated and copy number declines. While this group observed disagreement in copy number estimates between DAPI staining and quantitative PCR, they suggest that the detection of NUPTs, both in PCR-based and hybridization approaches artificially inflates plastome copy number estimates.

At least three studies challenge the notion that plastome abundance declines as cells mature in *Arabidopsis* and report an absence of significant variation in plastome abundance during development [205–207]. Further analysis in *Beta vulgaris* (beet) supports the majority of *Arabidopsis* findings and demonstrates the constancy of plastome copy number as leaves mature [153]. Methodological differences may indeed be responsible for the inconsistencies; regardless, three independent labs using different methods report agreement of their results for *Arabidopsis*.

8.3 Maintaining Plastome Integrity

Recombination dependent replication is likely to play an integral role in the maintenance of plastome integrity. In the potentially genotoxic environment of the plastid, ptDNA damage is bound to arise. Collapsed replication forks, DNA lesions, double strand breaks (DSB), and slipped strand mispairing are among the phenomena that threaten plastome stability and may set up scenarios under which error prone DNA repair can occur. Aberrant replication and repair mechanisms can ultimately lead to reconfiguration of plastomes. Very little is known about the nuclear encoded proteins that function in plastids to ameliorate these phenomena, but recent studies are shedding some light on mechanisms involved in securing ptDNA.

8.3.1 Plastid Targeted Whirlies

Like RecA1, the plastid targeted Whirly proteins are characterized as single stranded DNA binding proteins and are proposed to contribute to plastome stability by prohibiting illegitimate recombination events [208]. *Arabidopsis* plants in which one or both known plastid targeted Whirlies are knocked out have been used to investigate the role and mechanism of these proteins in plastome maintenance. Outward and inward facing PCR amplification and sequencing identified 40 reorganized ptDNA loci in the double knock out lines including deletions and circularized and/or head to tail concatemers delimited by direct repeats of 10–18 bp. In the lines where only one of the two Whirlies was disrupted rearrangements were less frequent. Surprisingly, even analysis of wild-type plants revealed two rearrangement events. Extension of this approach to maize Whirly mutants confirmed that these proteins are similarly involved in plastome maintenance in this system. Analysis of the short direct repeats associated with illegitimate recombination events in *Arabidopsis* mutants identified an overrepresentation of A/T homopolymeric regions, known to cause slipped-strand mediated replication stalling and DSB [209]. In some lines repeats bordering the rearranged regions contained some mismatches suggesting that recombination events did not require perfect homology [208]. Continued study of Whirlies has revealed that a specific motif conserved across this protein family is integral to the assembly of higher order homomeric structures [210]. While previous experiments had

demonstrated that the phenotype of the *Arabidopsis why1 why3* mutant lines could be complemented by transformation with the wild-type *WHY1*, transformation with a *why1* construct carrying the mutation for a specific lysine residue shown to abolish higher order assembly in *why2* lines failed to recover Whirly-1 activity [210].

8.3.2 Double Strand Break Repair in Plastids

The frequency and resolution of DSBs in plastomes is likely influenced by several protein factors including RecA, Whirlies, DNA polymerase Ib [211] and single strand binding proteins such as the homolog of bacterial *MutS*, MSH1 [212]. To explore how plastids of wild-type *Arabidopsis* cope with a DSB scenario, Kwon et al. [213] created an inducible system in which the I-CreII endonuclease of *Chlamydomonas reinhardtii* is targeted to plastid where it cleaves the endogenous *psbA* gene. Amplicon sequencing of nested PCR products from across the I-CreII cleavage site showed deletions mapping to either side of the cleavage site. Evaluation of the junction sequences indicated that DSBs had occurred and were predominantly repaired by microhomology mediated end joining (MMEJ), but in a few cases data was suggestive of nonhomologous end joining (NHEJ). Perfectly homologous repeats of 6–12 bp, or imperfect 10–16 bp repeats, were sufficient to direct DSB repair by MMEJ. No rearrangements were detected in this study, likely due to the presence of factors (i.e., Whirlies) that minimize activation of a microhomology mediated break induced replication (MMBIR) repair pathway. Wild-type *Arabidopsis* plants were resistant to ptDNA rearrangements compared to the Whirly double knock out plants, which accrue a heterogeneous pool of ptDNA molecules when treated with DNA damaging agents. Because MMEJ can only account for deletions following a repair event [214] the investigators proposed that an MMBIR response yielded the varied and complex ptDNA molecules observed [208, 215]. This response, which relies on template switching after replication fork collapse, could utilize microhomologous regions within the same molecule or on two different molecules producing the diversity of rearranged products observed in these studies [216].

Double strand breaks repaired by pathways that utilize microhomologies were detected even in wild-type *Arabidopsis* plastomes, which have a low incidence of large repeats [95] and inverted microhomologies at stalled replication forks led to reinitiation on the wrong strand forming short range inversions [217]. Longer homeologous or nonhomologous sequences have been identified in plastomes, particularly those containing abundant repeats, and, along with microhomologies, can participate in DSB repair via RDR pathways. The series of large repeats in the *Monsonia emarginata* plastome likely facilitated replicative repair leading to the accumulation of alternate arrangements of the plastome monomer [76]. Although unexplored in *Erodium*, the return of the IR in

Medicago minima can be explained by DSB repair via synthesis dependent strand annealing or Holiday junction formation/resolution [18]. The resulting arrangement of syntenic blocks of sequence represents a significant structural change, however mechanisms that utilize homologies to template replicative repair are active in plastomes and typically considered conservative resolutions to DSB [95].

Repair pathways mediated by microhomologous regions in the plastome represent an error prone mechanism to overcome severe DNA damage. Nonhomologous end joining (NHEJ) can resolve a DSB with high fidelity or alter DNA, with the result being tightly linked to the nature of the break and the activity of attendant protein factors [218, 219]. A major DNA repair pathway in the eukaryotic nucleus and also active in bacteria and plant mitochondria, NHEJ was not thought to occur in angiosperm plastids and was suggested to exclude promiscuous DNA from plastome integration. Whether or not NHEJ is involved is uncertain, however plastomes in several lineages have been identified to contain insertions of mitochondrial DNA [165–171]. While it is possible that inefficient integration mechanisms could be limiting foreign DNA insertions in plastomes, more likely the lack of an active DNA uptake system [220] and the plastid double membrane have precluded widespread transfer of extraplastidal DNA.

Investigating transposition of the IS 150 element in transplastomic *N. tabacum*, Kohl and Bock [221] were unable to detect religated, IS 150 lacking plastomes, again suggesting the absence of NHEJ in plastids. Observing DSBs at high frequency at both ends of IS element, these authors proposed two possible pathways to explain the lack of detection of religated molecules: the damaged plastomes were degraded or repaired via homologous recombination with IS 150-containing plastomes. More recently two independent investigations have reported NHEJ products in *Arabidopsis*. In one case repair of DSBs by NHEJ following I CreII activity was detected at low frequency [213], in another NHEJ repair events represented 17% of the rearranged products in Whirly knock out lines [215]. The finding that NHEJ, although likely a quantitatively minor pathway, can occur in plastids makes elucidating the nature of linear ptDNA ends all the more compelling.

9 Plastome Evolution: Phylogenetic and Biotechnological Considerations

The processes of DNA replication, recombination and repair in plastids [95, 222] have shaped the evolution of a conservative genome in most angiosperms. Given the highly orchestrated nature of these activities, perturbances to one aspect of the system can be

ramified in unpredictable ways. Large and small scale rearrangements, indels and pseudogenization have diversified plastomes and consideration of such diversity is imperative to experimental design for phylogenetic reconstruction as well as for biotechnological applications. Although gene order and protein coding sequences are well conserved across most angiosperms, pertinent differences are observed across families and genera. More subtle differences can be detected at and below the species level facilitating resolution over short evolutionary time scales [101, 223].

9.1 Targeting Plastome Insertion

To achieve efficient foreign sequence integration by homologous recombination transplastomic approaches have used endogenous plastome sequences to target insertions. Studies demonstrate a positive correlation between the rate of recombination and both the length and degree of sequence homology in prokaryotes [224, 225]. Decreasing identity between transformation vectors and target plastome regions, that is, using *N. tabacum* flanking sequences to transform different species, yields much lower transformation efficiencies [226–230]. Both intra- and interfamilial variation in target sequences negatively impacted homologous recombination events and concomitantly transformation efficiency.

Between *L. sativa* and *N. tabacum* in which the ribosomal operon region has been targeted for transgene integration, gene order is conserved but large and small deletions within the introns of *trnI* and *trnA* unique to *Nicotiana*, restricts the length of nearly identical sequence to less than 600 bp across the integration site. Although the minimum flanking sequence length has not been defined empirically, as little as a 400 bp on either side of the expression cassette appears to be sufficient to obtain transformation at a reasonable frequency [231]. Targeting sequences generally extend from 1–1.5 kb on either side of the expression cassette [232, 233].

9.2 Phylogenetic Utility of "Noncoding" Sequences

The level of conservation observed for plastome protein coding sequences is so great as to limit their phylogenetic utility when sampling few genes or taxa, especially at lower taxonomic levels. Alternatively, concatenated sequences for all protein coding regions plus the four ribosomal RNAs have been employed to gain resolution among early diverging lineages and several major angiosperm clades ([5] and references therein; [234–236]). For evaluating recent divergence such as interspecific hybridization or cultivars/haplotypes within species, amplified noncoding sequences have been used. Introns and intergenic regions, including spacers and regulatory sequences, yield some phylogenetically informative sites, but also demonstrate the need to recognize the functional nature of regulatory elements when analyzing DNA sequence [98–101, 237, 238]. More recently investigators have explored the use of whole plastome sequences to evaluate relationships at and below the

species level. Studies such as those of wild *Oryza* species [239] and individuals of *Jacobaea vulgaris* [240] have found that most phylogenetically informative sites occurred in intergenic regions.

9.2.1 Small Inversions in Regulatory Elements

Indeed the intergenic regions used for phylogenetic studies include various elements associated with transcription and translation of plastid genes. The observation of small inversions predicted to arise from recombination between palindromes in both 5′ and 3′ UTR sequences is not surprising. The mRNA binding proteins involved in both activation and degradation of plastid transcripts are likely highly sensitive to cognate structures. The orientation of the loop sequence may be free from directional selection, and thus have no evolutionary significance (feature would be freely variable within groups), as long as the overall structure is preserved. Like the IR mediated interconversion of single copy regions shown by Palmer to exist in equimolar proportions within an individual [62], such small inversions could be present as a mixed population of isomers. A mutation that alters the thermodynamic stability of the structure such that one orientation was favored over another could lead to the fixation of that allele over time.

Oppositely, the very same kinds of nucleic acid-protein interactions should deter mutations that disrupt conserved stem and loop structures in UTRs, ribosome binding sites, conserved nucleotides involved in intron structure and processing or promoter elements. While the plastid gene expression system appears to be quite robust in its ability to utilize heterologous regulatory elements [90, 230, 241], specific stem and loop interactions are observed [242–246] and predicted structures for some 5′ regions, such as the *psbA* 5′ UTR, are distinct between angiosperm families [230]. Similar structures localized in the 3′ UTRs of plastid transcripts, unlike those of Eubacteria, do not act as terminators of transcription but rather participate in polycistron processing 3′ end maturation and transcript stability [244, 246–248]. The structural and functional constraints placed on sequence evolution in noncoding regions should be widely recognized and incorporated into phylogenetic approaches [98, 249].

9.2.2 Nucleotide Substitution in Noncoding Regions

In addition to rearrangements, the diversity of angiosperm plastomes is reflected in nucleotide substitution rates [39, 127, 250–252]. Rate comparisons may consider synonymous or nonsynonymous substitutions, or a ratio of the two, for protein coding genes. In regions that do not code for amino acid sequences, base substitutions have, at least for phylogenetic purposes, often been deemed neutral and compared to synonymous coding substitutions. An evaluation of 34 single copy introns and intergenic regions across a small but taxonomically disparate set of angiosperms found that nucleotide substitutions accounted for approximately 72% of the phylogenetically informative characters identified

[253]. However, where functional sequences are concerned base substitutions may not be neutral. On closer inspection regions of extreme sequence variability tend to localize in areas with the least structural and functional constraint [254, 255], an observation consistent with directional selection on regulatory sequences. Analysis of nucleotide substitution rates for protein coding genes allows for unambiguous assignment in terms of conservative and nonconservative change; nonetheless, this approach typically does not consider protein domain functions, such as participation in higher order intramolecular structure, interaction with other polypeptide subunits and enzymatic capacity. Thus some residues or groups of residues may be under localized selection pressures, a point that could be overlooked when evaluating rate variation at the genomic scale, across genes or taxa.

10 Plastome Inheritance

A cursory review of available plastome sequences could miss the diversity that becomes apparent with critical examination. Against the background of apparently restrictive conservation, the unusual, the unexplained, beg investigation. Plastids and their genomes are inherited from the seed parent in approximately 80% of angiosperms [256–259]. It is intriguing that a number of the most bizarre plastomes exhibit biparental inheritance (Fig. 3; *see* [36]).

10.1 Biparental Inheritance

Given the highly active recombination system functioning in plastids could there be a relationship between the presence of two plastome genotypes and genomic instability? Extensive recombination demonstrated in somatic hybrids and male sterile lines of *Nicotiana* [260–262] and through the incorporation of foreign sequences using plastid transformation strategies leaves little room for the notion that different plastome genotypes cannot recombine. Further evidence for ptDNA recombination has been documented in interspecific somatic hybrids of *Solanum* [263] and *Brassica* [264], and in wild populations of lodgepole pine [265]. Likely the greatest impediment to recombination is a physical one, the presence of the plastid envelopes presents a formidable barrier. The mechanical forces applied during plastid transformation overcome this barrier, but in the absence of such interventions how would different genotypes achieve the necessary proximity? Plastid fusion has been reported in *Hosta* mutants [266] suggesting the possibility for the exchange of genetic material. Also, the presence of stroma filled tubules, or stromules, connecting two or more plastids constitute a possible pathway for exchange [267–270]. Recent work is now uncovering the interconnectedness of plant cell organelles and suggests that membrane contact sites (MCS) establish foci for the transfer of molecules between cellular

compartments and even between compartments in adjacent cells via plasmodesmata [271]. Observed stress-inducible MCS include direct interaction between plastids and peroxisomes, the nucleus and mitochondria via stromules, peroxules, and mitochondrial protrusions, and direct mitochondrial-plastid MCS (reviewed in [272]).

Biparental inheritance is thought to be a derived trait in angiosperms and has arisen independently in several lineages [256, 259, 273, 274]. An intriguing hypothesis that addresses the possible correlation between plastome instability and the emergence of biparental inheritance of ptDNA has been proposed. Zhang and Sodmergen [274] suggest that, rather than contributing to disruptive change, the presence of paternal plastids and their plastomes provides resolution to plastid-genome incompatibility (PGI) resulting from deleterious mutations in the maternal plastids. The potential for biparental inheritance (PBPI) is identified using cytological evidence; the detection of plastid DNA signals in male gametic cells [256, 258]. While species known to exhibit biparental inheritance are always positive for PBPI, genetic inheritance is rarely demonstrated in the majority of PBPI species. This suggests that although the potential exists for paternal plastids and their DNA to be transmitted the occurrence is low relative to maternal transmission and would be easy to miss were there no defect to favor retention of the male parent plastid and its haplotype. The "paternal rescue" hypothesis suggests that PBPI can overcome defects that may render maternal plastids incompatible with the nuclear genome through stable transmission of paternal plastids to progeny but does not address a mechanism. Are maternal, defective plastids excluded in toto, or can we imagine a scenario where two plastid haplotypes, maternal and paternal, recombine (i.e., undergo gene conversion) to restore a wild-type phenotype?

10.2 Coevolved Complexes and Plastome–Genome Incompatibility

The plant genome comprises a coevolving, integrated genetic system housed in three subcellular compartments. This integration and coevolution is an effective postzygotic reproductive barrier and as such plays a role in speciation processes [275–278]. There is a strong correlation between biparental inheritance and higher rates of nucleotide substitutions, genomic rearrangements and plastid genome incompatibility (PGI), and this correlation has been used to suggest that biparental inheritance of organelles may rescue incompatibilities [76, 279–281]. Self-fertilized or outcrossed individuals within the same population carry very few plastome polymorphisms between them and those used to differentiate parental types in screens of progeny are usually synonymous changes with no adaptive effect. Transmission of two plastome types, maternal and paternal, to progeny (heteroplasmy) will typically resolve into homoplasmy over time in the absence of selection and may be

completely random [156, 282, 283]. However, if parental plastome differences impair interactions between coevolved polypeptides encoded in the plastid and nucleus PGI can result.

Previous studies of coevolution between nuclear and plastid genomes have primarily focused on examining subunits of genes that are encoded in the nucleus and the plastid that form multisubunit complexes in plastids [77, 284, 285]. In the case of PGI several plastid genes or regions have been implicated (*clpP*, *ycf1*, *ycf2* in *Campanulastrum* [280]; *accD* in *Pisum*, [286]; *rpo* genes in Geraniaceae, [284]; intergenic promoter region between *clpP* and *psbB* in *Oenothera*, [287]). Greiner and Bock [276] indicated that PGI could also be caused by disruption of regulatory pathways, making nuclear genes involved in plastid transcription, translation and mRNA maturation prime candidates.

Two studies utilized the Geraniaceae model system to evaluate correlation between rates of nucleotide substitutions (*dN*, *dS*) and *dN/dS*. Examining cytonuclear interactions in the PEP holoenzyme identified six major clades of nuclear encoded sigma factors across the Geraniales that correspond to the six sigma factors (sig1–6) present in *Arabidopsis* [284, 288]. Correlated rates were detected between interacting plastid (PEP) and nuclear (sigma factor) sequences across the Geraniales were likely caused by plastid-nuclear genome coevolution and could cause PGI in Geraniaceae hybrids. Involving a greater number of both plastid and nuclear constituents, examination of the protein subunits of the plastid ribosome relative to control sequences revealed strong signals of cytonuclear coevolution between plastid and nuclear encoded subunits, including nonsynonymous substitutions in plastid encoded and nuclear encoded plastid targeted subunits occurring along the same branches in the Geraniaceae phylogeny [285]. Increased *dN/dS* in plastid encoded ribosomal proteins was mainly due to intensified positive selection whereas increased *dN/dS* in nuclear encoded plastid targeted genes was facilitated by relaxed purifying selection.

Employing model systems like Geraniaceae or *Silene* (Caryophylaceae; [289]) is illuminating the evolutionary forces acting on plastomes and the influence of plastome–nuclear genome interactions is certainly a factor. The plastomes of these atypical lineages are useful for the diversity of changes they have experienced through evolutionary time. Applying the tools of plastome transformation provides the opportunity to precisely introduce specific nonsynonymous base changes in plastid genes that interact with nuclear counterparts and observe resulting phenotypes.

11 Conclusion

Like any endeavor involving empirical investigation, studies of plastome evolution, however enlightening, leave unresolved questions and newly envisioned directions for research. The dizzying advance of DNA sequencing technology promises databases replete with information. Integrating bioinformatic, genomic, and biochemical data will be paramount in the design and execution of experiments aimed at elucidating the role of nuclear encoded proteins in plastome maintenance. Of particular interest will be projects that target DNA replication, recombination and repair pathways in those families where plastome stability is, or once was, destabilized. Plastid biotechnology, aided by available genomic resources, should be extended to include important agronomic and model species. Plastome modification as a tool to address basic research questions has yielded previously elusive insights and we enthusiastically encourage novel applications of this technology to progress our understanding of plastome evolution in form and function.

Acknowledgments

Support for this study was provided by the National Science Foundation (DEB-0717372 to R.K.J. and IOS-1027259 and DEB-1853024 to R.K.J. and T.A.R.), the Sidney F. and Doris Blake Professorship in Systematic Botany to R.K.J., and the Fred C. Gloeckner Foundation (to T.A.R. and R.K.J.). The authors thank J. Chris Blazier for comments on an earlier version of the manuscript.

References

1. 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 U S A 99:12246–12251
2. Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5:123–135
3. Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci U S A 100:8828–8833
4. Sheppard AE, Madesis P, Lloyd AH, Day A, Ayliffe MA, Timmis JN (2011) Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. Plant Mol Biol 76:299–309
5. Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee S, 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 U S A 104:19369–19374

6. Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids Cell culture and somatic cell genetics of plants, vol 7a. Springer, Vienna, pp 5–53
7. 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
8. Cai Z, Penaflor C, Kuehl J, Leebens-Mack J, Carlson J, dePamphilis C, Boore J, Jansen R (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol Biol 6:77
9. Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids, vol 19. Springer-Verlag, Berlin, pp 29–63
10. Clarke AK, MacDonald TM, Sjogren LLE (2005) The ATP-dependent Clp protease in chloroplasts of higher plants. Physiol Plant 123:406–412
11. Kuroda H, Maliga P (2003) The plastid clpP1 protease gene is essential for plant development. Nature 425:86–89
12. 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
13. Cronan JE, Waldrop GL (2002) Multi-subunit acetyl-CoA carboxylases. Prog Lipid Res 41:407–435
14. Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, Dornelas MC, Varani AM, Oliveira GCX, Bergès H, Vieira ML (2017) The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. Front Plant Sci 8:334
15. Rabah SO, Shrestha B, Hajrah NH, Sabir JM, Alharby H, Sabir JM, Gilbert LE, Ruhlman TA, Jansen RK (2019) *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J Syst Evol 57:1–14
16. Park S, Ruhlman TA, Weng M-L, Hajrah NH, Sabir JSM, Jansen RK (2017) Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. Genome Biol Evol 9:1766–1780
17. Gurdon C, Maliga P (2014) Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *accD* coding region in *Medicago truncatula*. DNA Res 21:417–427
18. Choi I-S, Jansen RK, Ruhlman TA (2019) Lost and found: return of the inverted repeat in the legume clade defined by its absence. Genome Biol Evol. https://doi.org/10.1093/gbe/evz076
19. Lee SS, Jeong WJ, Bae JM, Bang JW, Liu JR, Harn CH (2004) Characterization of the plastid-encoded carboxyltransferase subunit (*accD*) gene of potato. Mol Cells 17:422–429
20. Konishi T, Shinohara K, Yamada K, Sasaki Y (1996) Acetyl-CoA carboxylase in higher plants: most plants other than Gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol 37:117–122
21. Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, Stefanovic 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
22. Sabir J, Schwarz E, Ellison N, Zhang J, Baeshen NA, Mutwakil M, Jansen RK, Ruhlman T (2014) Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol J 12:743–754
23. Page MLD, Hamel PP, Gabilly ST, Zegzouti H, Perea JV, Alonso JM, Ecker JR, Theg SM, Christensen SK, Merchant S (2004) A homolog of prokaryotic thiol disulfide transporter CcdA is required for the assembly of the cytochrome b6f complex in *Arabidopsis* chloroplasts. J Biol Chem 279:32474–32482
24. Hamel PP, Dreyfuss BW, Xie Z, Gabilly ST, Merchant S (2003) Essential histidine and tryptophan residues in CcsA, a system II polytopic cytochrome c biogenesis protein. J Biol Chem 278:2593–2603
25. Sasaki Y, Sekiguchi K, Nagano Y, Matsuno R (1993) Chloroplast envelope protein encoded by chloroplast genome. FEBS Lett 316:93–98
26. Willey DL, Gray JC (1990) An open reading frame encoding a putative haem-binding polypeptide is cotranscribed with the pea chloroplast gene for apocytochrome f. Plant Mol Biol 15:347–356
27. Rolland N, Dorne AJ, Amoroso G, Sültemeyer DF, Joyard J, Rochaix JD (1997) Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in

the chloroplast of *Chlamydomonas*. EMBO J 16:6713–6726

28. 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 8:174
29. Goremykin VV, Hirsch-Ernst KI, Wölfl 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
30. 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
31. Kahlau S, Aspinall S, Gray JC, Bock R (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of Solanaceous plastid genomes. J Mol Evol 63:194–207
32. 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
33. Cai Z, Guisinger M, Kim H, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK (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
34. Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583–600
35. Weng M-L, Ruhlman TA, Jansen RK (2017) Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol 214:842–851
36. Ruhlman TA, Jansen RK (2018) Aberration or analogy? The atypical plastomes of Geraniaceae. In: Chaw S-M, Jansen RK (eds) Advances in botanical research, plastid genome evolution. Academic, Cambridge, MA, pp 223–262
37. 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
38. Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628
39. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK (2010) Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. J Mol Evol 70:149–166
40. 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
41. Cheon K-S, Kim K-A, Yoo K-O (2017) The complete chloroplast genome sequences of three *Adenophora* species and comparative analysis with Campanuloid species (Campanulaceae). PLoS One 12:e0183652
42. Fajardo D, Senalik D, Ames M, Zhu H, Steffan SA, Harbut R, Polashock J, Vorsa N, Gillespie E, Kron K, Zalapa JE (2013) Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. Tree Genet Gen 9:489–498
43. Martínez-Alberola F, del Campo EM, Lázaro-Gimeno D, Mezquita-Claramonte S, Molins A, Mateu-Andrés I, Pedrola-Monfort-J, Casano LM, Barreno E (2013) Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. PLoS One 8:e79685
44. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol 4:294–306
45. Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Eguiarte LE, Kumar S, Lee HO, Lee J, McMahon M, Steele K, Wing R, Yang TJ, Zwickl D, Wojciechowski MF (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. Am J Bot 102:1115–1127
46. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M (2013) Uncovering the

protein translocon at the chloroplast inner envelope membrane. Science 339:571–574

47. de Vries J, Sousa FL, Bölter B, Soll J, Gould SB (2015) YCF1: a green TIC? Plant Cell 27:1827–1833
48. Iles WJD, Smith SY, Graham SW (2013) A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. In: Early Events in Monocot Evolution, systematics association special volume series. Cambridge University Press, Cambridge, MA
49. Peredo EL, King UM, Les DH (2013) The plastid genome of *Najas flexilis*: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. PLoS One 8:e68591
50. Neyland R, Urbatsch LE (1996) Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndhF* chloroplast gene sequences. Am J Bot 83:1195–1206
51. Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, Cheng C-H, Lin C-Y, Liu S-M, Chang C-C, 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
52. Wu F-H, Chan M-T, Liao D-C, Hsu C-T, Lee Y-W, Daniell H, Duvall MR, Lin C-S (2010) Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. BMC Plant Biol 10:68
53. Ruhlman TA, Chang W-J, Chen JJ, Huang Y-T, Chan M-T, Zhang J, Liao D-C, Blazier JC, Jin X, Shih M-C, Jansen RK, Lin CS (2015) NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. BMC Plant Biol 15:100
54. Blazier JC, Guisinger MM, Jansen RK (2011) Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol Biol 76:263–272
55. Ifuku K, Endo T, Shikanai T, Aro E-M (2011) Structure of the chloroplast NADH dehydrogenase-like complex: nomenclature for nuclear-encoded subunits. Plant Cell Physiol 52:1560–1568
56. 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 U S A 91:9794–9798
57. Braukmann TWA, Kuzmina M, Stefanović S (2009) Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr Genet 55:323–337
58. Lee H, Golicz AA, Bayer PE, Severn-Ellis AA, Chan C-KK, Batley J, Kendrick GA, Edwards D (2018) Genomic comparison of two independent seagrass lineages reveals habitat-driven convergent evolution. J Exp Bot 69:3689–3702
59. Portis AR, Parry MAJ (2007) Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. Photosynth Res 94:121–143
60. Kolodner R, Tewari KK (1979) Inverted repeats in chloroplast DNA from higher plants. Proc Natl Acad Sci U S A 76:41–45
61. Palmer JD, Thompson WF (1981) Rearrangements in the chloroplast genomes of mung bean and pea. Proc Natl Acad Sci U S A 78:5533–5537
62. Palmer JD (1983) Chloroplast DNA exists in two orientations. Nature 301:92–93
63. Stein DB, Palmer JD, Thompson WF (1986) Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. Curr Genet 10:835–841
64. Tassopulu D, Kung SD (1984) *Nicotiana* chloroplast genome. 6. Deletion and hot spot - a proposed origin of the inverted repeats. Theor Appl Genet 67:185–193
65. Goulding SE, Wolfe KH, Olmstead RG, Morden CW (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252:195–206
66. Falcon LI, Magallon S, Castillo A (2010) Dating the cyanobacterial ancestor of the chloroplast. ISME J 4:777–783
67. Criscuolo A, Gribaldo S (2011) Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. Mol Biol Evol 28:3019–3032
68. Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. Am J Bot 91:1846–1862
69. 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

70. Palmer JD, Thompson WF (1982) Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537–550
71. Palmer JD, Osorio B, Aldrich J, Thompson WF (1987) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr Genet 11:275–286
72. Wolfe KH (1988) The site of deletion of the inverted repeat in pea chloroplast DNA contains duplicated gene fragments. Curr Genet 13:97–99
73. 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
74. Downie S, Palmer J (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman & Hall, New York, NY, pp 14–35
75. Blazier JC, Jansen RK, Mower JP, Govindu M, Zhang J, Weng M-L, Ruhlman TA (2016) Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann Bot 117:1209–1220
76. Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK (2017) Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot 104:559–572
77. Zhang J, Ruhlman TA, Sabir JSM, Blazier JC, Weng M-L, Park S, Jansen RK (2016) Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. Genome Biol Evol 8:622–634
78. Guisinger MM Personal communication
79. Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Wm Stevenson D, Zomlefer W, Davis JI (2016) Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytol 209:855–870
80. Wimpee CF, Wrobel RL, Garvin DK (1991) A divergent plastid genome in *Conopholis americana*, an achlorophyllous parasitic plant. Plant Mol Biol 17:161–166
81. Frailey DC, Chaluvadi SR, Vaughn JN, Coatney CG, Bennetzen JL (2018) Gene loss and genome rearrangement in the plastids of five hemiparasites in the family Orobanchaceae. BMC Plant Biol 18:30
82. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM (2013) Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell 25:3711–3725
83. Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci U S A 89:10648–10652
84. Delannoy E, Fujii S, Colas des Francs-Small, C., Brundrett, M., and Small, I. (2011) Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol 28:2077–2086
85. Logacheva MD, Schelkunov MI, Penin AA (2011) Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. Genome Biol Evol 3:1296–1303
86. Wolfe KH, Li W, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A 84:9054–9058
87. Birky CW Jr, Walsh JB (1992) Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics 130:677–683
88. Svab Z, Hajdukiewicz P, Maliga P (1990) Stable transformation of plastids in higher plants. Proc Natl Acad Sci U S A 87:8526–8530
89. Maliga P (1993) Towards plastid transformation in flowering plants. Trends Biotechnol 11:101–107
90. Daniell H, Kumar S, Dufourmantel N (2005) Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol 23:238–245
91. Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J 46:85–94
92. Tungsuchat-Huang T, Sinagawa-García SR, Paredes-López O, Maliga P (2010) Study of plastid genome stability in tobacco reveals that the loss of marker genes is more likely by gene conversion than by recombination between 34-bp loxP repeats. Plant Physiol 153:252–259
93. Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, Bailey CD, Jansen RK (2017) Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome

features across legume subfamilies. J Mol Evol 84:187–203

94. Oldenburg DJ, Bendich AJ (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J Mol Biol 335:953–970
95. Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. New Phytol 186:299–317
96. Martin G, Baurens FC, Cardi C, Aury JM, D'Hont A (2013) The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid monocotyledon evolution. PLoS One 8:e67350
97. Walker JF, Jansen RK, Zanis MJ, Emery NC (2015) Sources of inversion variation in the small single copy (SSC) region of chloroplast genomes. Am J Bot 102:1751–1752
98. Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann Mo Bot Gard 87:482–498
99. Kim K, Lee H (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. Mol Cell 19:104–113
100. Bain JF, Jansen RK (2006) A chloroplast DNA hairpin structure provides useful phylogenetic data within tribe Senecioneae (Asteraceae). Can J Bot 84:862–868
101. Ansell SW, Schneider H, Pedersen N, Grundmann M, Russell SJ, Vogel JC (2007) Recombination diversifies chloroplast *trnF* pseudogenes in *Arabidopsis lyrata*. J Evol Biol 20:2400–2411
102. Morton BR, Clegg MT (1993) A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). Curr Genet 24:357–365
103. Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 106:2–9
104. Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in Apiaceae subfamily Apioideae. Syst Bot 25:648–667
105. 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 8:36–36
106. Davis JI, Soreng RJ (2010) Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae). Am J Bot 97:874–892
107. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus. Silene Genome Biol Evolution 4:294–306
108. Palmer JD, Nugent JM, Herbon LA (1987) Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci U S A 84:769–773
109. Kim KJ, Jansen RK (1994) Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia, Asteraceae*): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Plant Syst Evol 190:157–185
110. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190
111. Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X (2013) The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene 528:120–131
112. 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
113. Mariotti R, Cultrera NGM, Díez CM, Baldoni L, Rubini A (2010) Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. BMC Plant Biol 10:211
114. Oldenburg DJ, Bendich AJ (2015) DNA maintenance in plastids and mitochondria of plants. Front Plant Sci 6:883
115. Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc Natl Acad Sci U S A 84:5818–5822
116. Raubeson LA, Jansen RK (1992) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255:1697–1699
117. 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 U S A 89:1856–1860

118. Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. BMC Evol Biol 4:27
119. Blazier C, Guisinger M, Jansen RK (2011) Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol Biol 76:263–272
120. Weng M-L, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. Mol Biol Evol 31:645–659
121. Park S, Jansen R, Park S (2015) Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. BMC Plant Biol 15:40
122. Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J 10:3073–3078
123. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin 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
124. Ueda M, Fujimoto M, Arimura S, Murata J, Tsutsumi N, Kadowaki K (2007) Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. Gene 402:51–56
125. Jansen RK, Saski C, Lee S, Hansen AK, Daniell H (2011) 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 28:835–847
126. Zhelyazkova P, Sharma CM, Förstner KU, Liere K, Vogel J, Börner T (2012) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. Plant Cell 24:123–136
127. 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 12:279–291
128. 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
129. 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
130. Arthofer W, Schuler S, Steiner FM, Schlick-Steiner BC (2010) Chloroplast DNA-based studies in molecular ecology may be compromised by nuclear-encoded plastid sequence. Mol Ecol 19:3853–3856
131. Lloyd AH, Timmis JN (2011) The origin and characterization of new nuclear genes originating from a cytoplasmic organellar genome. Mol Biol Evol 28:2019–2028
132. Allen JF (2003) The function of genomes in bioenergetic organelles. Philos Trans R Soc B 358:19–38
133. Allen JF, Puthiyaveetil S, Strom J, Allen CA (2005) Energy transduction anchors genes in organelles. BioEssays 27:426–435
134. Daley DO, Whelan J (2005) Why genes persist in organelle genomes. Genome Biol 6:110–110
135. Allen JF (1993) Redox control of gene expression and the function of chloroplast genomes – an hypothesis. Photosynth Res 36:95–102
136. 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. Nucleic Acids Res 31:5324–5331
137. Wright AF, Murphy MP, Turnbull DM (2009) Do organellar genomes function as long-term redox damage sensors? Trends Genet 25:253–261
138. Kanevski I, Maliga P (1994) Relocation of the plastid *rbcL* gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. Proc Natl Acad Sci U S A 91:1969–1973
139. Zerges W (2002) Does complexity constrain organelle evolution? Trends Plant Sci 7:175–182
140. Dekker JP, Boekema EJ (2005) Supramolecular organization of thylakoid membrane proteins in green plants. BBA-Bioenergetics 1706:12–39
141. Akbarova YY, Solovyev VV, Shahmuradov IA (2010) Possible functional and evolutionary role of plastid DNA inserted in rice genome. Appl Comput Math 9:19–33

142. Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, Kadowaki K (2008) Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. Mol Biol Evol 25:1566–1575
143. Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, Nakayama T, Ishida K, Bhattacharya D (2010) Differential gene retention in plastids of common recent origin. Mol Biol Evol 27:1530–1537
144. Rousseau-Gueutin M, Ayliffe MA, Timmis JN (2011) Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol 157:2181–2193
145. Stern DB, Lonsdale DM (1982) Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. Nature 299:698–702
146. 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
147. Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26:99–110
148. Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27:1436–1448
149. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10(1):e1001241
150. Kumar RA, Bendich AJ (2011) Distinguishing authentic mitochondrial and plastid DNAs from similar DNA sequences in the nucleus using the polymerase chain reaction. Curr Genet 57:287–295
151. Maier RM, Schmitz-Linneweber (2004) Plastid genomes. In: Daniell H, Chase C (eds) Molecular biology and biotechnology of plant organelles: chloroplasts and mitochondria. Springer, Dordrecht, pp 115–150
152. Bendich AJ (1987) Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays 6:279–282
153. Rauwolf U, Golczyk H, Greiner S, Herrmann RG (2010) Variable amounts of DNA related to the size of chloroplasts III. Biochemical determinations of DNA amounts per organelle. Mol Genet Genomics 283:35–47
154. Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. Nucleic Acids Res 34:4537–4545
155. Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schottler MA, Bock R (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. Plant Cell 23:3137–3155
156. Maliga P, Carrer H, Kanevski I, Staub J, Svab Z (1993) Plastid engineering in land plants: a conservative genome is open to change. Philos Trans R Soc B 342:203–208
157. Maliga P, Sz-Breznovits A, Marton L, Joo F (1975) Non-Mendelian streptomycin-resistant tobacco mutant with altered chloroplasts and mitochondria. Nature 255:401–402
158. Moller S (2005) Plastid division in higher plants. In: Moller SG (ed) Plastids. Blackwell, Oxford, pp 126–152
159. McCabe MS, Klaas M, Gonzalez-Rabade N, Poage M, Badillo-Corona JA, Zhou F, Karcher D, Bock R, Gray JC, Dix PJ (2008) Plastid transformation of high-biomass tobacco variety Maryland mammoth for production of human immunodeficiency virus type 1 (HIV-1) p24 antigen. Plant Biotechnol J 6:914–929
160. Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AI, Maloney AP, Kavanagh TA, Gray JC, Bock R (2008) High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. Plant Biotechnol J 6:897–913
161. Maenpaa P, Gonzalez EB, Chen L, Khan MS, Gray JC, Aro EM (2000) The *ycf9* (orf62) gene in the plant chloroplast genome encodes a hydrophobic protein of stromal thylakoid membranes. J Exp Bot 51:375–382
162. Baena-Gonzalez E, Gray JC, Tyystjarvi E, Aro EM, Maenpaa P (2001) Abnormal regulation of photosynthetic electron transport in a chloroplast *ycf9* inactivation mutant. J Biol Chem 276:20795–20802
163. Swiatek M, Greiner S, Kemp S, Drescher A, Koop H, Herrmann RG, Maier RM (2003) PCR analysis of pulsed-field gel electrophoresis-purified plastid DNA, a sensitive tool to judge the hetero-/homoplastomic status of plastid transformants. Curr Genet 43:45–53

164. Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genomics 7:222
165. Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P (2012) De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol 12:61
166. Spooner DM, Ruess H, Iorizzo M, Senalik D, Simon P (2017) Entire plastid phylogeny of the carrot genus (*Daucus*, Apiaceae): concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. Am J Bot 104:296–312
167. Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336–351
168. Ma P-F, Zhang Y-X, Guo Z-H, Li D-Z (2015) Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. Sci Rep 5:11608
169. Burke SV, Ungerer MC, Duvall MR (2018) Investigation of mitochondrial-derived plastome sequences in the *Paspalum* lineage (Panicoideae; Poaceae). BMC Plant Biol 18:152
170. Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, Sabir JSM, Jansen RK, Ruhlman TA (2017) Plastome sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. Plant Genome 10:1–14
171. Ku C, Chung W-C, Chen L-L, Kuo C-H (2013) The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. PLoS ONE 8:e68518
172. Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A (2013) Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). Genome Biol Evol 5:1872–1885
173. Klaus SMJ, Huang F, Golds TJ, Koop H (2004) Generation of marker-free plastid transformants using a transiently cointegrated selection gene. Nat Biotechnol 22:225–229
174. Kelchner SA, Wendel JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr Genet 30:259–262
175. Catalano SA, Saidman BO, Vilardi JC (2009) Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. Cladistics 25:93–104
176. Iamtham S, Day A (2000) Removal of antibiotic resistance genes from transgenic tobacco plastids. Nat Biotechnol 18:1172–1176
177. Gray BN, Ahner BA, Hanson MR (2009) Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. Transgenic Res 18:559–572
178. Agrawal AF, Hadany L, Otto SP (2005) The evolution of plastic recombination. Genetics 171:803–812
179. Kavanagh TA, Thanh ND, Lao NT, McGrath N, Peter SO, Horváth EM, Dix PJ, Medgyesy P (1999) Homeologous plastid DNA transformation in tobacco is mediated by multiple recombination events. Genetics 152:1111–1122
180. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016
181. Singh S, Roy S, Choudhury S, Sengupta D (2010) DNA repair and recombination in higher plants: insights from comparative genomics of *Arabidopsis* and rice. BMC Genomics 11:443
182. Cerutti H, Ibrahim HZ, Jagendorf AT (1993) Treatment of pea (*Pisum sativum* L.) protoplasts with DNA-damaging agents induces a 39-kilodalton chloroplast protein immunologically related to *Escherichia coli* RecA. Plant Physiol 102:155–163
183. Cerutti H, Jagendorf AT (1993) DNA strand-transfer activity in pea (*Pisum sativum* L.) chloroplasts. Plant Physiol 102:145–153
184. Cerutti H, Johnson A, Boynton J, Gillham N (1995) Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Escherichia coli* RecA. Mol Cell Biol 15:3003–3011
185. Rowan BA, Oldenburg DJ, Bendich AJ (2010) RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. J Exp Bot 61:2575–2588
186. Kolodner RD, Tewari KK (1975) Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. Nature 256:708–711

187. Krishnan NM, Rao BJ (2009) A comparative approach to elucidate chloroplast genome replication. BMC Genomics 10:237
188. Zampini É, Truche S, Lepage É, Tremblay-Belzile S, Brisson N (2017) Plastid genome stability and repair. In: Somatic genome variation in animals, plants, and microorganisms. John Wiley & Sons, Ltd, Hoboken, NJ, pp 119–163
189. Meeker R, Nielsen B, Tewari KK (1988) Localization of replication origins in pea chloroplast DNA. Mol Cell Biol 8:1216–1223
190. 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
191. Takeda Y, Hirokawa H, Nagata T (1992) The replication origin of proplastid DNA in cultured cells of tobacco. Mol Gen Genet 232:191–198
192. Kunnimalaiyaan M, Shi F, Nielsen BL (1997) Analysis of the tobacco chloroplast DNA replication origin (ori B) downstream of the 23 S rRNA gene. J Mol Biol 268:273–283
193. Kunnimalaiyaan M, Nielsen BL (1997) Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. Nucleic Acids Res 25:3681–3686
194. Mühlbauer SK, Lössl A, Tzekova L, Zou Z, Koop H (2002) Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. Plant J 32:175–184
195. Wang Y, Saitoh Y, Sato T, Hidaka S, Tsutsumi K (2003) Comparison of plastid DNA replication in different cells and tissues of the rice plant. Plant Mol Biol 52:905–913
196. Scharff LB, Koop H (2007) Targeted inactivation of the tobacco plastome origins of replication A and B. Plant J 50:782–794
197. Deng XW, Wing RA, Gruissem W (1989) The chloroplast genome exists in multimeric forms. Proc Natl Acad Sci U S A 86:4156–4160
198. 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
199. Scharff LB, Koop H (2006) Linear molecules of tobacco ptDNA end at known replication origins and additional loci. Plant Mol Biol 62:611–621
200. Shaver JM, Oldenburg DJ, Bendich AJ (2006) Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. Planta 224:72–82
201. Baumgartner BJ, Rapp JC, Mullet JE (1989) Plastid transcription activity and DNA copy number increase early in barley chloroplast development. Plant Physiol 89:1011–1018
202. Rowan B, Oldenburg D, Bendich A (2004) The demise of chloroplast DNA in *Arabidopsis*. Curr Genet 46:176–181
203. Rowan B, Oldenburg D, Bendich A (2009) A multiple-method approach reveals a declining amount of chloroplast DNA during development in *Arabidopsis*. BMC Plant Biol 9:3
204. Zheng Q, Oldenburg DJ, Bendich AJ (2011) Independent effects of leaf growth and light on the development of the plastid and its DNA content in *Zea* species. J Exp Bot 62:2751–2730
205. Li W, Ruf S, Bock R (2006) Constancy of organellar genome copy numbers during leaf development and senescence in higher plants. Mol Gen Genomics 275:185–192
206. Zoschke R, Liere K, Börner T (2007) From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. Plant J 50:710–722
207. Evans IM, Rus AM, Belanger EM, Kimoto M, Brusslan JA (2010) Dismantling of *Arabidopsis thaliana* mesophyll cell chloroplasts during natural leaf senescence. Plant Biol 12:1–12
208. 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 U S A 106:14693–14698
209. Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. Microbiol Mol Biol Rev 71:13–35
210. Cappadocia L, Parent JS, Zampini E, Lepage E, Sygusch J, Brisson N (2012) A conserved lysine residue of plant whirly proteins is necessary for higher order protein assembly and protection against DNA damage. Nucleic Acids Res 40:258–269
211. Parent JS, Lepage E, Brisson N (2011) Divergent roles for the two PolI-like organelle DNA polymerases of *Arabidopsis*. Plant Physiol 156:254–262
212. Xu YZ, Arrieta-Montiel MP, Virdi KS, de Paula WBM, Wildhalm JR, Basset GJ, Davila JI, Elthon TE, Elowsky CG, Sato SJ, Clemente TE, Mackenzie SA (2011) MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell 23:3428–3441
213. Kwon T, Huq E, Herrin DL (2010) Microhomology-mediated and nonhomologous repair of a double-strand break in the

chloroplast genome of *Arabidopsis*. Proc Natl Acad Sci U S A 107:13954–13959

214. McVey M, Lee S (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24:529–538
215. Cappadocia L, Maréchal A, Parent J, Lepage E, Sygusch J, Brisson N (2010) Crystal structures of DNA-whirly complexes and their role in *Arabidopsis* organelle genome repair. Plant Cell 22:1849–1867
216. Hastings PJ, Ira G, Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 5: e1000327
217. Zampini É, Lepage É, Tremblay-Belzile S, Truche S, Brisson N (2015) Organelle DNA rearrangement mapping reveals U-turn-like inversions as a major source of genomic instability in *Arabidopsis* and humans. Genome Res 25:645–654
218. Shuman S, Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol 5:852–861
219. Lieber MR (2010) The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. Annu Rev Biochem 79:181–211
220. Smith DR (2011) Extending the limited transfer window hypothesis to inter-organelle DNA migration. Genome Biol Evol 3:743–748
221. Kohl S, Bock R (2009) Transposition of a bacterial insertion sequence in chloroplasts. Plant J 58:423–436
222. Day A, Madesis P (2007) DNA replication, recombination, and repair in plastids. In: Bock R (ed) Cell and molecular biology of plastids, vol 19. Springer-Verlag, Berlin Heidelberg, pp 65–119
223. Pleines T, Jakob SS, Blattner FR (2008) Application of non-coding DNA regions in intraspecific analyses. Plant Syst Evolution 282:281–294
224. Shen P, Huang HV (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112:441–457
225. Fujitani Y, Yamamoto K, Kobayashi I (1995) Dependence of frequency of homologous recombination on the homology length. Genetics 140:797–809
226. Sidorov VA, Kasten D, Pang SZ, Hajdukiewicz PT, Staub JM, Nehra NS (1999) Technical advance: stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. Plant J 19:209–216
227. Ruf S, Hermann M, Berger IJ, Carrer H, Bock R (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. Nat Biotechnol 19:870–875
228. Zubko MK, Zubko EI, Zuilen KV, Meyer P, Day A (2004) Stable transformation of petunia plastids. Transgenic Res 13:523–530
229. Nguyen TT, Nugent G, Cardi T, Dix PJ (2005) Generation of homoplasmic plastid transformants of a commercial cultivar of potato (*Solanum tuberosum* L.). Plant Sci 168:1495–1500
230. Ruhlman T, Verma D, Samson N, Daniell H (2010) The role of heterologous chloroplast sequence elements in transgene integration and expression. Plant Physiol 152:2088–2104
231. Bock R (2001) Transgenic plastids in basic research and plant biotechnology. J Mol Biol 312:425–438
232. Ruhlman T, Ahangari R, Devine A, Samsam M, Daniell H (2007) Expression of cholera toxin B–proinsulin fusion protein in lettuce and tobacco chloroplasts – oral administration protects against development of insulitis in non-obese diabetic mice. Plant Biotechnol J 5:495–510
233. Verma D, Samson NP, Koya V, Daniell H (2008) A protocol for expression of foreign genes in chloroplasts, *Nat*. PRO 3:739–758
234. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U S A 104:19363–19368
235. Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A 107:4623–4628
236. Gitzendanner MA, Soltis PS, Wong GKS, Ruhfel BR, Soltis DE (2018) Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot 105:291–301
237. Swangpol S, Volkaert H, Sotto RC, Seelanan T (2007) Utility of selected non-coding chloroplast DNA sequences for lineage assessment of *Musa* interspecific hybrids. J Biochem Mol Biol 40:577–587
238. Whitlock BA, Hale AM, Groff PA (2010) Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. PLoS One 5:e11533

239. Waters DLE, Nock CJ, Ishikawa R, Rice N, Henry RJ (2011) Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. Ecol Evol 2:211–217
240. Doorduin L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res 18:93–108
241. Tangphatsornruang S, Birch-Machin I, Newell CA, Gray JC (2011) The effect of different 3′ untranslated regions on the accumulation and stability of transcripts of a *gfp* transgene in chloroplasts of transplastomic tobacco. Plant Mol Biol 76:385–396
242. Shen Y, Danon A, Christopher DA (2001) RNA binding-proteins interact specifically with the *Arabidopsis* chloroplast *psbA* mRNA 5′ untranslated region in a redox-dependent manner. Plant Cell Physiol 42:1071–1078
243. Nickelsen J (2003) Chloroplast RNA-binding proteins. Curr Genet 43:392–399
244. Pfalz J, Bayraktar OA, Prikryl J, Barkan A (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. EMBO J 28:2042–2052
245. Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. Annu Rev Plant Biol 61:125–155
246. Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suarez M, Borner T, Barkan A (2011) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. Nucleic Acids Res 40:3092–3105
247. Stern DB, Gruissem W (1987) Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51:1145–1157
248. Loza-Tavera H, Vargas-Suárez M, Díaz-Mireles E, Torres-Márquez M, González de la Vara L, Moreno-Sánchez R, Gruissem W (2006) Phosphorylation of the spinach chloroplast 24 kDa RNA-binding protein (24RNP) increases its binding to *petD* and *psbA* 3′ untranslated regions. Biochimie 88:1217–1228
249. Borsch T, Quandt D (2009) Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol 282:169–199
250. Matsuoka Y, Yamazaki Y, Ogihara Y, Tsunewaki K (2002) Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. Mol Biol Evol 19:2084–2091
251. Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc Natl Acad Sci U S A 105:18424–18429
252. Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2009) Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. PLoS One 4:e5297
253. Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275–288
254. Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W (2003) Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. J Evol Biol 16:558–576
255. Storchova H, Olson MS (2007) The architecture of the chloroplast *psbA-trnH* non-coding region in angiosperms. Plant Syst Evol 268:235–256
256. 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
257. Harris SA, Ingram R (1991) Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. Taxon 40:393–412
258. Zhang Q, Liu Y, Sodmergen (2003) Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. Plant Cell Physiol 44:941–951
259. Hu Y, Zhang Q, Rao G, Sodmergen (2008) Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. Plant Cell Physiol 49:958–968
260. Medgyesy P, Fejes E, Maliga P (1985) Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc Natl Acad Sci U S A 82:6960–6964
261. Thanh ND, Medgyesy P (1989) Limited chloroplast gene transfer via recombination overcomes plastome-genome incompatibility between *Nicotiana tabacum* and *Solanum tuberosum*. Plant Mol Biol 12:87–93

262. Kung SD, Zhu YS, Chen K, Shen GF, Sisson VA (1981) *Nicotiana* chloroplast genome. Mol Gen Genet 183:20–24
263. Bidani A, Nouri-Ellouz O, Lakhoua L, Sihachakr D, Cheniclet C, Mahjoub A, Drira N, Gargouri-Bouzid R (2007) Interspecific potato somatic hybrids between *Solanum berthaultii* and *Solanum tuberosum* L. showed recombinant plastome and improved tolerance to salinity. Plant Cell Tiss Organ Cult 91:179–189
264. Yadav P, Bhat SR, Prakash S, Mishra LC, Chopra VL (2009) Resynthesized *Brassica juncea* lines with novel organellar genome constitution obtained through protoplast fusion. J Genet 88:109–112
265. Marshall HD, Newton C, Ritland K (2001) Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. Mol Biol Evol 18:2136–2138
266. Vaughn KC (1981) Plastid fusion as an agent to arrest sorting out. Curr Genet 3:243–245
267. Gray JC, Hibberd JM, Linley PJ, Uijtewaal B (1999) GFP movement between chloroplasts. Nat Biotechnol 17:906–909
268. Kwok EY, Hanson MR (2004) GFP-labelled Rubisco and aspartate aminotransferase are present in plastid stromules and traffic between plastids. J Exp Bot 55:595–604
269. Hanson MR, Sattarzadeh A (2008) Dynamic morphology of plastids and stromules in angiosperm plants. Plant Cell Environ 31:646–657
270. Gray JC, Hansen MR, Shaw DJ, Graham K, Dale R, Smallman P, Natesan SKA, Newell CA (2012) Plastid stromules are induced by stress treatments acting through abscisic acid. Plant J 69:387–398
271. Brunkard JO, Runkel AM, Zambryski PC (2015) The cytosol must flow: intercellular transport through plasmodesmata. Curr Opin Cell Biol 35:13–20
272. Pérez-Sancho J, Tilsner J, Samuels AL, Botella MA, Bayer EM, Rosado A (2016) Stitching organelles: organization and function of specialized membrane contact sites in plants. Trends Cell Biol 26:705–717
273. Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc Natl Acad Sci U S A 92:11331–11338
274. Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? J Plant Res 123:201–206
275. Greiner S, Rauwolf U, Meurer J, Herrmann RG (2011) The role of plastids in plant speciation. Mol Ecol 20:671–691
276. Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays 35:354–365
277. Burton RS, Pereira RJ, Barreto FS (2013) Cytonuclear genomic interactions and hybrid breakdown. Annu Rev Ecol Evol Syst 44:281–302
278. Bock DG, Andrew RL, Rieseberg LH (2014) On the adaptive value of cytoplasmic genomes in plants. Mol Ecol 23:4899–4911
279. Jansen RK, Ruhlman TA (2012) Plastid genomes of seed plants. In: Bock R, Knoop V (eds) Advances in photosynthesis and respiration 35, Genomics of chloroplasts and mitochondria. Springer, Dordrecht, pp 103–126
280. Barnard-Kubow KB, Sloan DB, Galloway LF (2014) Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. BMC Evol Biol 14:268
281. Barnard-Kubow KB, McCoy MA, Galloway LF (2017) Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol 213:1466–1476
282. Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu Rev Genet 35:125–148
283. Korpelainen H (2004) The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. Naturwissenschaften 91:505–518
284. Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK (2015) Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. Plant Cell 27:563–573
285. Weng M-L, Ruhlman TA, Jansen RK (2016) Plastid–nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol 8:1824–1838
286. Bogdanova VS, Zaytseva OO, Mglinets AV, Shatskaya NV, Kosterin OE, Vasiliev GV (2015) Nuclear-cytoplasmic conflict in pea (*Pisum sativum* L.) is associated with nuclear and plastidic candidate genes encoding acetyl-coA carboxylase subunits. PLoS ONE 10: e0119835
287. 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. Nucleic Acids Res 36:2366–2378

288. Lysenko EA (2007) Plant sigma factors and their role in plastid transcription. Plant Cell Rep 26:845–859
289. Sloan DB, Triant DA, Wu M, Taylor DR (2014) Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol Biol Evol 31:673–682
290. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious:v5.1. http://www.geneious.com
291. Darling AE, Mau B, Perna NT (2010) Progressive mauve: multiple genome alignment with gene gain, loss, and rearrangement. PLoS One 5:e11147