

# Chapter 5

## Plastid Genomes of Seed Plants

Robert K. Jansen\* and Tracey A. Ruhlman

*Section of Integrative Biology, University of Texas at Austin, 1 University Station  
C0930, Austin, TX 78712, USA*

*and Department of Biological Sciences, Genomics and Biotechnology Section,  
King Abdulaziz University, Jeddah 21589, Saudi Arabia*

Summary .....	103
I. Introduction .....	104
II. Plastid Genome Organization.....	104
A. Overall Organization .....	104
B. Genome Size, Gene/Intron Content.....	106
C. Gene Order.....	109
III. Plastid Inheritance.....	110
IV. Genomic Rearrangements .....	111
A. IR Loss or Expansion/Contraction .....	111
B. Gene and Intron Loss .....	113
C. Gene Order Changes .....	114
V. Patterns and Rates of Nucleotide Substitutions.....	115
A. Sequence Evolution in Coding Regions Versus Intergenic Regions and Introns .....	115
B. Rates of Sequence Evolution in Protein Coding Genes.....	116
C. Correlation Between Rates of Nucleotide Substitutions and Genomic Rearrangements .....	117
VI. Phylogenetic Utility of Plastome Data for Resolving Relationships Among Seed Plants.....	117
VII. Conclusions and Future Directions .....	119
Acknowledgments.....	120
References .....	120

### Summary

The field of comparative plastid genomics has burgeoned during the past decade, largely due to the availability of rapid, less expensive genome sequencing technologies. Currently there are 200 plastid genomes (plastomes) publicly available with 65% of these from seed plants. Comparative analyses have demonstrated that there is an overall stability in plastome architecture, gene and intron content, and gene order across seed plants. However, a number of unrelated lineages of both gymnosperms and angiosperms do not follow this pattern and have experienced considerable genomic upheaval. Within angiosperms these lineages with highly rearranged plastomes exhibit three other phenomena: highly accelerated rates of nucleotide substitutions, an increase in the number of dispersed repeats, many of which are associated with rearrangement endpoints, and biparental plastid inheritance. The correlation

---

\*Author for correspondence, e-mail: jansen@mail.utexas.edu

between genomic upheaval and these other phenomena suggest that aberrant DNA repair mechanisms may be involved in destabilizing these plastid genomes. Experimental studies support this idea because knocking out DNA repair genes destabilizes plastomes. Further studies of nuclear-plastid interactions, especially in seed plant lineages with highly rearranged plastomes, are needed to clarify the causes of the plastome instability. The large number of plastome sequences has also provided valuable data for resolving phylogenetic relationships among seed plants. This is especially true for angiosperms where these data have been instrumental in clarifying relationships among the early diverging clades, an endeavor that had stymied plant biologists for over a decade. The most recent plastome phylogenies clearly identify *Amborella* as the earliest diverging lineage of flowering plants and provide strong support for the position of magnoliids as sister to a large clade that includes eudicots and monocots. This robust phylogenetic estimate provides an evolutionary framework for examining patterns and rates of change in plastid genomes across angiosperms.

## I. Introduction

The field of comparative plastid genomics began in 1986 with the publication of the first two land plant plastid genome sequences (plastomes) for *Nicotiana tabacum* (Shinozaki et al. 1986) and *Marchantia polymorpha* (Ohyama et al. 1986). Following these landmark papers there was a slow and steady increase in the number of completed plastome sequences for about 15 years (Fig. 5.1). The development of less expensive, high-throughput DNA sequencing methods resulted in a rapid rise in the number of publicly available plastome sequences during the past decade. Currently (as of February 16, 2011) there are 205 plastome sequences available on Genbank representing many major lineages of photosynthetic organisms. The vast majority (175) represent green plants (Viridiplantae) with most of these from the two major lineages of seed plants, angiosperms (118) and gymnosperms (16). Within seed plants many of the major lineages of gymnosperms (Wu et al. 2007, 2009; McCoy et al. 2008; Lin et al. 2010; Zhong et al. 2010) and angiosperms (Jansen et al. 2007, 2011; Moore et al. 2007) are now represented, although there is still

limited sampling for some clades, especially among gymnosperms. The increased availability of plastome sequences has provided a wealth of new comparative data for understanding patterns of genome organization, rates of sequence evolution, mechanisms of evolutionary change, and phylogenetic relationships among seed plants.

During the past decade there have been several reviews of plastid genome organization and evolution and the phylogenetic implications of the newly acquired plastome data (Odintsova and Yurina 2003; Raubeson and Jansen 2005; Bock 2007; Ravi et al. 2008; Khan et al. 2010; Gao et al. 2010; Wolf et al. 2011) but many of these were published when there were limited plastome sequences available and none of them focused exclusively on seed plants. In this chapter, we summarize the current knowledge of the organization and evolution of seed plant plastid genomes with a focus on their genome organization, inheritance, rate of nucleotide substitution and genomic rearrangements, and the utility of plastome data for resolving phylogenetic relationships.

## II. Plastid Genome Organization

### A. Overall Organization

In general plastome organization is highly conserved among seed plants with most having a quadripartite structure with two

---

*Abbreviations:* bp – Basepair; indel – Insertion/deletion; IR – Inverted repeat; IRLC – Inverted repeat lacking clade; kb – Kilobase; LSC – Large single copy region; PEP – Plastid encoded polymerase; SSC – Small single copy region

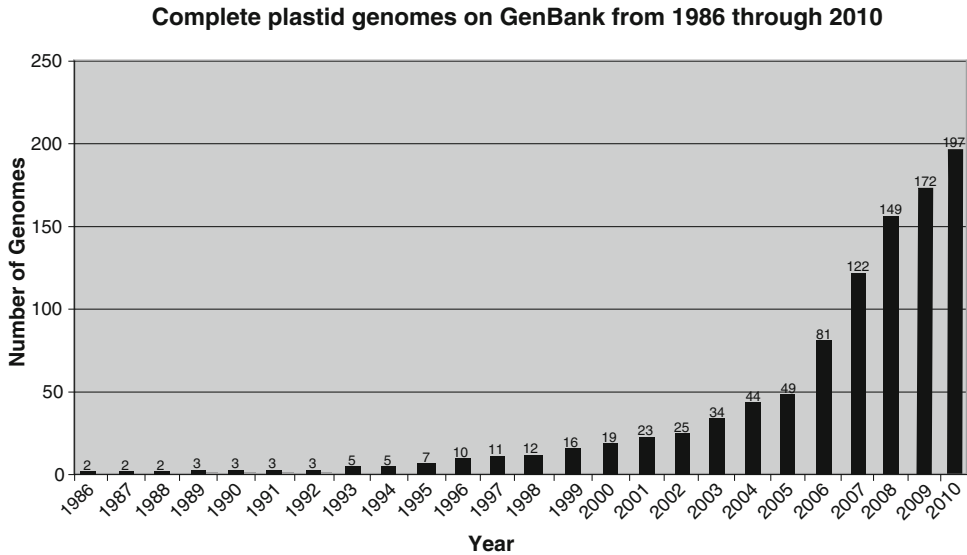


Fig. 5.1. Histogram showing number of plastid genomes available on GenBank from 1986 to December 1, 2010. (<http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid>)

copies of a large inverted repeat (IR) separated by small (SSC) and large (LSC) single copy regions (Fig. 5.2a, b). The two copies of the IR facilitate flip–flop recombination resulting in the presence of isoforms that differ in the orientation of the single copy regions (Palmer 1983). The prevailing view has been that plastid genomes are circular, and this was supported by early electron microscopic comparisons that revealed circular genomes in either monomeric or multimeric form (Kolodner and Tewari 1979). More recently considerable evidence has accumulated that suggests a much more complex structure, with circular, linear, branched, and multimeric configurations that vary during plastid development (Lilly et al. 2001; Bendich 2004; Oldenburg and Bendich 2004; Shaver et al. 2006).

The majority of plastid genes are contained in operons and transcribed as polycistronic units, a feature that reflects the endosymbiotic origin of plastids from a cyanobacterial ancestor (see Fig. 5.2b for operon organization). Among seed plant plastomes there are very few instances of disruption of operons. Exceptions occur in the three angiosperm families Campanulaceae (Cosner et al. 1997;

Haberle 2006; Haberle et al. 2008), Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011), and Fabaceae (Milligan et al. 1989; Cai et al. 2008; Palmer et al. 1988; Perry et al. 2002). In the Geraniaceae, the *rps2-atpA* operon is disrupted in the most recent common ancestor of *Erodium texanum* and *Geranium palmatum* but two other plastomes from this family, *Pelargonium hortorum* and *Monsonia speciosa*, have this operon intact. The highly conserved S10 operon is also disrupted in Geraniaceae. This operon is split into two groups of genes, *rpl23-rps3* and *rpl16-rpoA*, in *E. texanum* and *M. speciosa*, whereas it is split into four pieces (*rpl23*, *rpl2*, *rps19-rpl22*, and *rps3-rpoA*) in *G. palmatum*. *Pelargonium hortorum* has an intact S10 operon except that *rpoA* is so divergent that its functionality is in question. In Campanulaceae, the plastomes have two disrupted operons, *rps2-atpI-atpH-atpF-atpA* and *clpP-5'rps12-rpl20* (Haberle 2006; Haberle et al. 2008). Two operons, *rpoB-rpoC1-rpoC2* and *clpP-5'rps12-rpl20*, are disrupted in *Trifolium* but they are intact in the plastomes of the seven other sequenced legumes, including the closely related genus *Medicago*. The disruption of

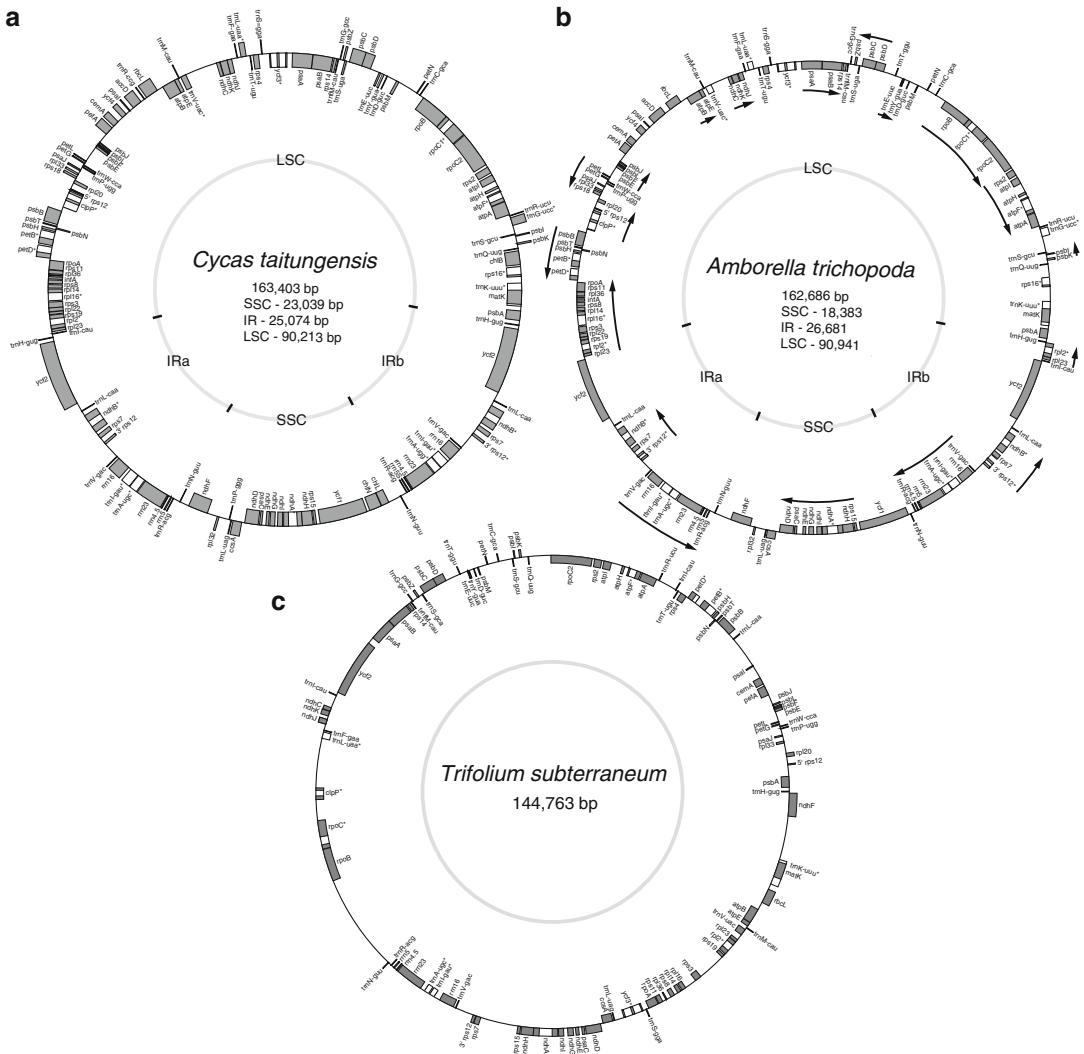


Fig. 5.2. Physical maps of the three seed plant plastid genomes: (a) *Cycas taitungensis* (NC\_009618); (b) *Amborella trichopoda* (NC\_005086); and (c) *Trifolium subterraneum* (NC\_011828). Maps were constructed using GenomeVX (Conant and Wolfe 2008; <http://wolfe.gen.tcd.ie/GenomeVx/>). Genes annotated inside the circle are transcribed clockwise while those outside are transcribed counterclockwise. Arrows in (b) indicate polycistronic transcription units. Introns are annotated as open boxes and genes containing introns are marked with asterisks.

the *rpoB* operon is notable because it includes three of four genes for the plastid-encoded RNA polymerase (PEP), the multi-subunit enzyme that transcribes many plastid genes. In all of these cases, relocated segments of operons must have acquired new promoters to drive gene transcription but experimental studies have not been performed to deter-

mine how these segments are transcribed in their new location.

### B. Genome Size, Gene/Intron Content, and GC Content

Genome size varies considerably among photosynthetic seed plant plastomes (Table 5.1),

Table 5.1. Characteristics of plastid genomes of representative photosynthetic seed plants and the parasitic, non-photosynthetic plant *Epifagus virginica*. Data in this table came from Wu et al. 2009; Lin et al. 2010, and Guisinger et al. 2011 or from analyses using sequences on GenBank (see a–g below)

Genome characteristic	<i>Cycas taitungensis</i> <sup>a</sup>	<i>Ephedra equisetina</i> <sup>b</sup>	<i>Pinus thunbergii</i> <sup>c</sup>	<i>Cryptomeria japonica</i> <sup>d</sup>	<i>Amborella trichopoda</i> <sup>e</sup>	<i>Pelargonium hortorum</i> <sup>f</sup>	<i>Epifagus virginica</i> <sup>g</sup>
Size (bp)	163,403	109,518	119,707	131,810	162,686	217,942	70,028
LSC length (bp)	90,216	59,906	65,696	93,958	90,941	59,710	19,799
SSC length (bp)	23,039	8,104	53,021	37,624	18,383	6,750	4,759
IR length (bp)	25,074	20,754	495	114	26,681	75,741	22,735
Number of different genes/total number of genes	109/124	101/117	114/115	116/118	114/132	109/160	42/53
Number of different protein-coding genes (duplicated in IR)	74 (4)	66 (6)	66 (0)	82 (0)	80 (7)	76 (39)	21 (3)
Number of different tRNA genes (duplicated in IR)	31 (7)	31 (6)	32	32 (1)	30 (7)	29 (8)	17 (4)
Number of different rRNA genes (duplicated in IR)	4 (4)	4 (4)	4 (0)	4 (0)	4 (4)	4 (4)	4 (4)
Number of genes duplicated in IR	15	16	1	2	18	51	11
Number of different genes with introns	18	12	12	17	18	16	4
Percent of genome coding for genes	57.2	72.3	56.5	60.8	56.6	51.5	43
Gene density – total number of genes/genome length including IR (genes/kb)	0.759	1.068	1.01	0.91	0.81	0.734	0.757
GC content (%)	39.4	36.6	38.5	35.4	38.3	39.6	36.0

<sup>a</sup>NC\_009618 (Wu et al. 2007)

<sup>b</sup>NC\_011954 (Wu et al. 2009)

<sup>c</sup>NC\_001631 (Wakasugi et al. 1994)

<sup>d</sup>NC\_010548 (Hirao et al. 2008)

<sup>e</sup>NC\_005086 (Goremykin et al. 2003)

<sup>f</sup>NC\_008454 (Chumley et al. 2006)

<sup>g</sup>NC\_001568 (Wolfe et al. 1992)

ranging from 107,122 bp (*Cathaya argophylla* [NC\_014589]) to 217,942 bp (*Pelargonium hortorum* [NC\_008454]) with an average length of 144,824 bp. The most remarkable example of plastome size variation within a single family occurs in the Geraniaceae with the smallest genome, *Erodium carvifolium*, at 116,935 bp and the largest, *Pelargonium hortorum* at 217,942 bp (Guisinger et al. 2011; Blazier et al. 2011). Plastomes from a few non-photosynthetic, parasitic plants have been sequenced and these genomes are

greatly reduced in size ranging from 59,190 bp in *Rizanthella gardneri* (Delannoy et al. 2011) to 86,744 bp in *Cuscuta gronovii* (Funk et al. 2007). Several factors contribute to this wide variation in genome size. First, expansion/contraction and loss of the IR is one of the most evident causes; it has been recognized for some time that small changes in the extent of the IR are very common due to shifting of the IR/SC boundaries (Goulding et al. 1996). There is considerable variation in IR size across seed plants ranging from absent in the

IR loss clade of legumes (Lavin et al. 1990) and *Erodium texanum* (Guisinger et al. 2011) to 75,741 bp in *Pelargonium hortorum* (Chumley et al. 2006). The second factor contributing to genome size variation is gene loss and additional gene duplications outside of the IR. In gnetophytes (McCoy et al. 2008; Wu et al. 2009) the loss of up to 18 genes has resulted in a more compact genome with gene densities lower than other gymnosperms and most angiosperms (Table 5.1). In Geraniaceae, there has been considerable partial or complete duplication of genes, which could be partly responsible for the larger plastomes of some members of this angiosperm family (Guisinger et al. 2011). The third factor involves downsizing of introns and intergenic spacer regions as was observed for the gnetophyte plastomes (McCoy et al. 2008; Wu et al. 2009).

Seed plant plastomes usually contain 101–118 different genes (Table 5.1) with the majority of these (66–82) coding for proteins involved in photosynthesis and gene expression and several others with miscellaneous functions, 29–32 transfer RNAs, and four ribosomal RNA genes (Fig. 5.2, Table 5.1). The range in the total number of genes is higher (115–160), largely due to duplication of genes in the IR (Table 5.1). The highest number (160 in *Pelargonium hortorum*) is due to the duplication of 39 genes in the 76 kilobase (kb) IR (Chumley et al. 2006). Among photosynthetic seed plants the gnetophytes have the most reduced gene content with up to 18 gene losses, including the absence of all 11 NADH dehydrogenase genes (Wu et al. 2009). The most highly reduced plastomes in terms of gene content are from parasitic plants that have varying degrees of capacity for photosynthesis. The completely non-photosynthetic plants *Epifagus virginica* and *Rhizanthella gardneri* have only 42 and 33 intact genes, respectively (Wolfe et al. 1992; Delannoy et al. 2011), and most of these represent ribosomal genes (tRNAs, rRNAs, and ribosomal proteins). The genus *Cuscuta* has wide variation in adaptations to a parasitic life history. Some species are fully photosynthetic, others are

intermediate with limited photosynthetic capacity and others are completely non-green and parasitic. Recent examinations of plastomes from four *Cuscuta* species with varying photosynthetic capacity have demonstrated a progressive loss of plastid genes with increasing parasitism (Funk et al. 2007; McNeal et al. 2007b).

Intron content of seed plant plastomes is highly conserved; most have 18 genes with introns, six in tRNAs and 12 in protein coding genes (Fig. 5.2a, b). Fifteen of the 18 intron-containing genes have a single intron and three genes, *ycf3*, *clpP*, and *rps12*, have two introns, resulting in a total of 21 introns in most seed plants. Twenty of the introns are group II, whereas *trnL-uua* is the only group I intron in seed plant plastomes. Splicing of exons for 17 of these genes involves cis-splicing. The single exception is *rps12*, which has cis-splicing for exons 2 and 3 and trans-splicing for exon 1 (Hildebrand et al. 1988). Variation in the number of genes with introns occurs among photosynthetic seed plants, ranging from 12 in gnetophytes to 18 in most species (Table 5.1). Similar to the situation in gene content, parasitic plants have a reduced number of introns with only four and six different genes with introns in the holoparasites *Epifagus virginica* and *Rhizanthella gardneri* and 12 in hemiparasitic *Cuscuta* species (Wolfe et al. 1992; Funk et al. 2007; McNeal et al. 2007b; Delannoy et al. 2011).

GC content among seed plant plastomes ranges between 34% and 40% (Goremykin et al. 2003; Kim and Lee 2004; Cai et al. 2006; Raubeson et al. 2007; Guisinger et al. 2011; Table 5.1). There is an uneven distribution of GC content over the plastid genome and this pattern is due primarily to three factors (Cai et al. 2006). First, coding regions have a significantly higher GC content than non-coding regions. Second, the distribution of GC content by regions of the genome varies with the highest in the IR and the lowest in the SSC. The higher GC content in the IR is due primarily to the presence of the four rRNA genes that have highest GC content of any coding regions. The lowest GC content

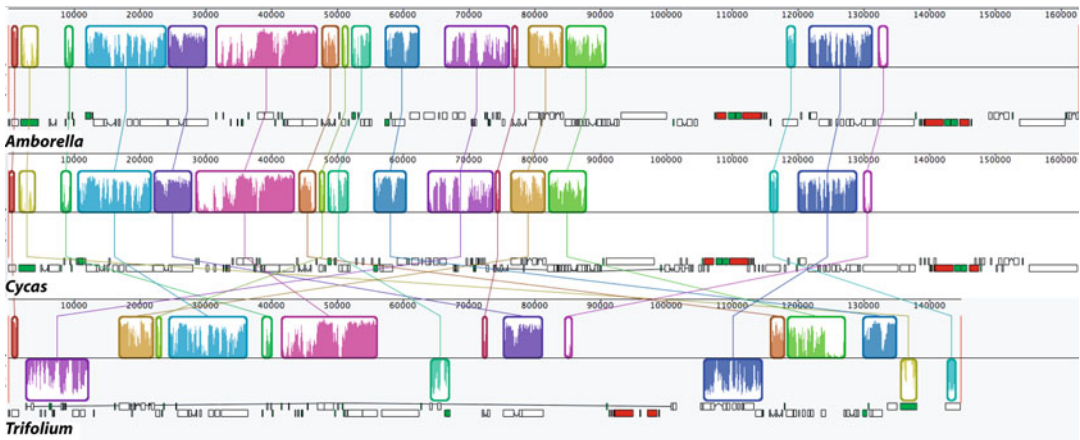


Fig. 5.3. Gene order comparison of two highly conserved and one rearranged plastid genomes. Whole plastid genome sequences were downloaded from Genbank for *Cycas taitungensis* (NC\_009618), *Amborella trichopoda* (NC-005086) and *Trifolium subterraneum* (NC\_011828). Alignments were performed in Geneious Pro (Drummond et al. 2010) with the mauveAligner algorithm (Darling et al. 2010), which aligns synthetic blocks of genes and predicts inversions relative to a reference genome.

in the SSC is caused by the presence of 8 of the 11 NADH dehydrogenase genes, which have the lowest GC content of any functional group. Third, GC content varies by functional groups. Among protein-coding genes, GC content is highest for photosynthetic genes, lowest for NADH genes, with genetic system genes having intermediate values. GC content also varies by codon position in protein-coding genes (Cai et al. 2006; Raubeson et al. 2007; Guisinger et al. 2011). For each of the three classes of genes (photosynthetic, genetic system, and NADH) the third position in the codon has a significant AT bias. This pattern has been attributed to codon usage bias (Shimada and Sugiura 1991; Kim and Lee 2004; Chaw et al. 2004; Liu and Xue 2005). Several studies have examined codon usage of plastid genes to determine if these biases can be attributed to nucleotide compositional bias, selection for translational efficiency, or a balance among mutational biases, natural selection, and genetic drift (Morton 1993, 1994, 1998; Wall and Herbeck 2003). The recent investigation of codon usage and GC content in Geraniaceae and related rosids concluded that codon usage in plastid genes is generally driven by selection and not GC content (Guisinger et al. 2011).

### C. Gene Order

Gene order among seed plant plastomes is generally highly conserved. This is evident by comparison of the gene order of *Cycas* with the basal angiosperm *Amborella* (Fig. 5.3, top two genomes). These two genomes are co-linear, suggesting that the ancestral gene order for seed plants was similar to *Cycas*. Although there is no published plastome sequence for *Ginkgo*, gene mapping studies (Palmer and Stein 1986) indicate that this genome is also co-linear with *Cycas* and basal angiosperms. Despite the high level of conservation in gene order across seed plants, a number of groups, including gnetophytes, conifers, and several lineages of angiosperms have experienced considerable change (see Fig. 5.3 and Sect. IV below).

Three different mechanisms have been suggested to cause gene order changes in seed plant plastomes. First, inversion facilitated by recombination is considered the most common mechanism of plastome rearrangement (Palmer 1991; Raubeson and Jansen 2005). Intramolecular recombination of plastid DNA has been documented in *Oryza*; in this case repeats <15 bp recombine and generate deletions in both coding

and non-coding regions (Kanno et al. 1993; Kawata et al. 1997). Intermolecular recombination between tRNA sequences in *Oryza* was also shown to result in gene order change (Hiratsuka et al. 1989). Recombination between repeats has generated genome rearrangements in transplastomic plants, providing experimental evidence for this mechanism in generating inversions (Rogalski et al. 2006; Gray et al. 2009). Several studies in angiosperms have documented the presence of a large number of repetitive sequences in highly rearranged plastomes, with the highest concentration of repeats occurring at rearrangement endpoints (Haberle et al. 2008; Chumley et al. 2006; Guisinger et al. 2011). The most extensive comparisons are in the Geraniaceae where the size and number of repeats is correlated with the degree of genomic rearrangement (Guisinger et al. 2011). Also, shared families of repeats flank rearrangement endpoints across the four genomes examined in this family. Second, transposition was suggested as a mechanism of plastome rearrangement in *Trachelium* (Cosner et al. 1997) and *Trifolium* (Milligan et al. 1989). However, plastid genome sequences for these species have not confirmed transposition in either of these plastomes (Haberle et al. 2008; Cai et al. 2008). The only case of a plastome transposable element is the degenerate “Wendy” element of the alga *Chlamydomonas* (Fan et al. 1995). Third, expansion and contraction of the IR has been suggested as the cause of gene order changes in the green alga *Chlamydomonas* (Boudreau and Turmel 1995) and the angiosperm families Fabaceae (Perry et al. 2002) and Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011). The Geraniaceae plastomes are the best example of this because of their incredible variation in the size of the IR, ranging from absent to 76 kb.

### III. Plastid Inheritance

Considerable progress has been made in recent years to improve our understanding of modes of plastid inheritance in seed plants

(reviewed in Hagemann 2004; Bock 2007; Hu et al. 2008; Nagata 2010; Kuroiwa 2010). Gymnosperms have not been investigated as extensively as angiosperms and they have been erroneously considered to have almost exclusively paternal inheritance. It turns out that all three modes (biparental, maternal and paternal) have been documented, with cycads, *Ginkgo*, and gnetophytes having maternal inheritance and conifers having substantial variation in mode of inheritance. Most studies have reported paternal plastid inheritance in conifers (Mogensen 1996) but biparental inheritance is known in *Cryptomeria* (Ohba et al. 1971) and progeny from crosses in *Larix* detected a mixture of maternal and paternal plastids (Szmidi et al. 1987). Overall, the prevailing mode of plastid inheritance in conifers is paternal but some species may have maternal or biparental inheritance. Examination of a broader phylogenetic diversity of gymnosperms is needed to fully understand the extent of variation in their mode of plastid inheritance.

Plastids have historically been thought to have largely maternal inheritance among angiosperms (Corriveau and Coleman 1988; Birky 1995; Mogensen 1996; Zhang et al. 2003; Hagemann 2004). The proportion of angiosperms with maternal inheritance is currently estimated at about 80% with the remaining species having biparental inheritance. The only known case of exclusively paternal inheritance is *Actinidia speciosa* (Testolin and Cipriani 1997). The phylogenetic distribution of mode of plastid inheritance in angiosperms indicates that maternal inheritance is ancestral and that there have been repeated conversions to biparental inheritance scattered among more derived lineages (Fig. 5.4; Hu et al. 2008). Furthermore, the phylogenetic distribution also indicates that changes in mode of inheritance are unidirectional because there are no cases of the derivation of maternal inheritance from a biparental ancestor.

Several different mechanisms are known to prevent paternal plastids from being transmitted during fertilization (Hagemann 2004; Bock 2007). Most angiosperms with maternal



inheritance lack plastids in the generative cell, which results in their exclusion in the sperm cells. The presence of plastids in the generative cell does not necessarily result in their transmission to the embryo. Mechanisms to prevent paternal plastid transmission occur in multiple post-fertilization stages from exclusion just prior to fertilization to differential replication of maternal and paternal plastids in the embryo. Several studies have documented a surprising amount of variation in inheritance patterns, including situations where the mode of inheritance varied in progeny from crosses depending on whether the cross is inter- or intraspecific (Cruzan et al. 1993; Soliman et al. 1987; Yang et al. 2000; Lee et al. 1988; Hansen et al. 2007a).

There has been some discussion on why biparental inheritance of plastids has evolved multiple times from maternal inheritance. Zhang and Sodmergen (2010) suggested that biparental inheritance evolved as a mechanism to overcome defective maternal plastids in angiosperms with nuclear plastid incompatibility. This is a tantalizing hypothesis and is consistent with two lines of evidence. First, a number of angiosperms with nuclear plastid incompatibility systems have biparental inheritance, including *Oenothera* (Chiu and Sears 1993), *Passiflora* (Mráček 2005), *Pelargonium* (Metzlaff et al. 1982), *Trifolium* (Pandey et al. 1987), and *Zantedeschia* (Snijder et al. 2007). Second, crossing studies demonstrate that more distant crosses (i.e., interspecific), which are more likely to cause genomic incompatibilities, result in progeny with paternal plastids, whereas progeny from crosses within species have maternal plastids.

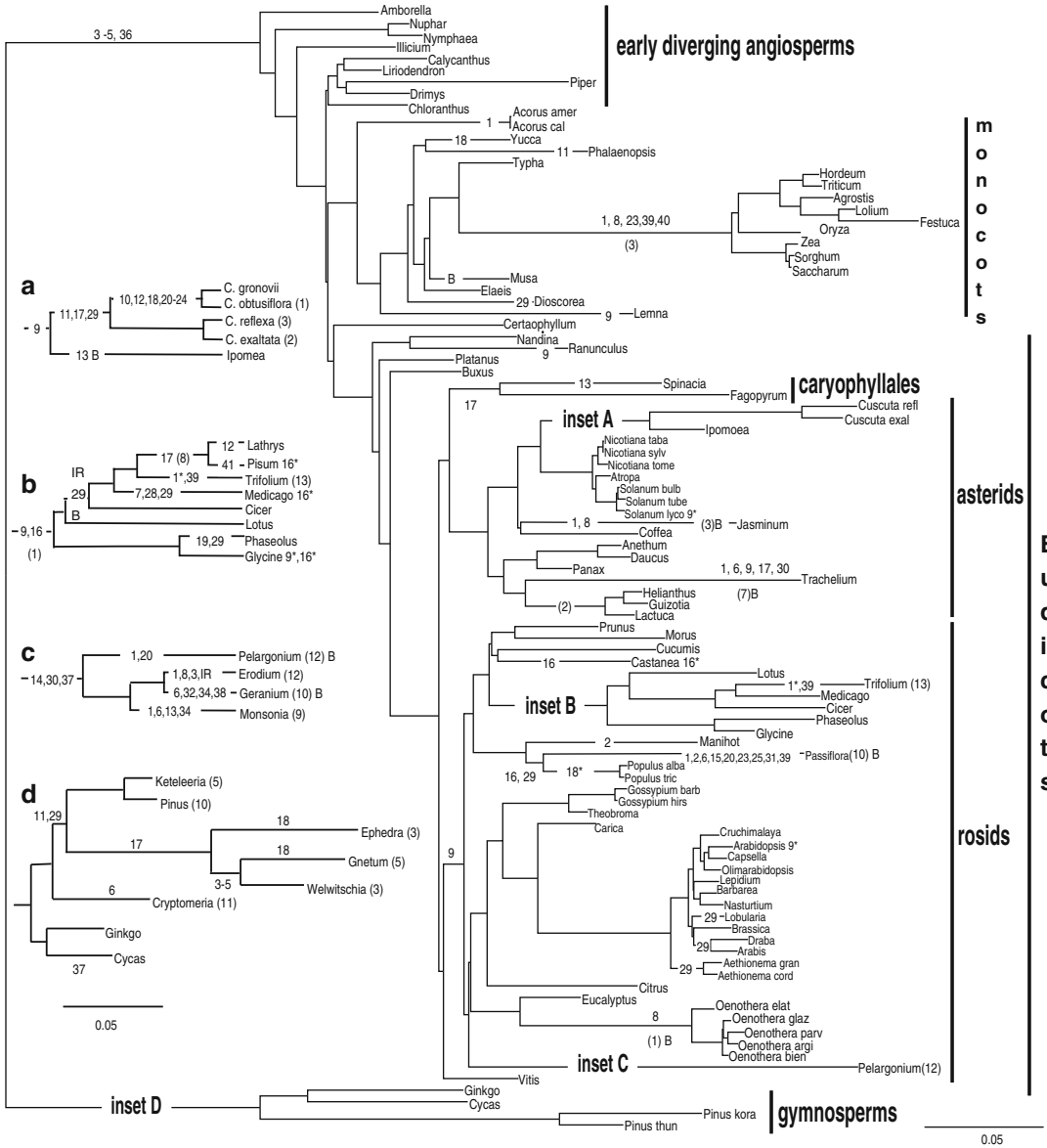
#### IV. Genomic Rearrangements

##### A. IR Loss or Expansion/Contraction

The IR is present in the vast majority of seed plant plastomes, and some have argued that this structure promotes stability for the rest of the genome, largely via intramolecular recombination between the two IR copies,

which limits recombination between the single copy regions (Palmer et al. 1987; Palmer 1991). This idea was supported earlier by the fact that plastomes known to lack the IR experienced more genomic rearrangements, especially some legumes. However, this correlation has not held up as more plastomes have been sequenced. In fact, some of the most highly rearranged seed plant plastomes have retained their IR, including the angiosperm families Campanulaceae (Cosner et al. 2004; Haberle et al. 2008), Lobeliaceae (Knox and Palmer 1999), Oleaceae (Lee et al. 2007), Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011), and the gnephytes among gymnosperms (McCoy et al. 2008; Wu et al. 2009).

The IR has been reported lost at least five times independently in seed plants. Within angiosperms IR loss has occurred at least two times within rosids (Fig. 5.4). The first reported loss was in a large, monophyletic group of papilionoid Fabaceae referred to as the inverted repeat lacking clade (IRLC; Wojciechowski et al. 2004). There have been two independent losses reported in the Geraniaceae in *Erodium texanum* and *Monsonia vanderietieae* (Downie and Palmer 1992) but only one of these has been verified (Guisinger et al. 2011; Blazier et al. 2011). Complete plastome sequences confirmed the IR loss in *Erodium texanum* (Guisinger et al. 2011) and *E. carvifolium* (Blazier et al. 2011), and draft genome sequences of 12 other *Erodium* species indicate that the IR has been lost throughout the genus (C. Blazier and R. Jansen, unpublished). The situation in *Monsonia* is not fully resolved. The complete genome sequence of *Monsonia speciosa* has an IR, although it is greatly reduced to 7 kb (Guisinger et al. 2011). A draft genome sequence of *M. vanderietieae* suggests that there may be a small IR of at least 3 kb, although assembly of this genome is complicated by the large number of rearrangements and repeats (M. Guisinger and R. Jansen, unpublished). Thus, it is likely that there has only been a single IR loss in Geraniaceae. IR losses have been suggested for two genera of Orobanchaceae, *Conophilis* and *Striga*



- |                 |                |                |                |                   |
|-----------------|----------------|----------------|----------------|-------------------|
| 1 accD          | 10 matK        | 18 rpl32       | 26 rps11       | 34 trnG-UCCintron |
| 2 atpFintron    | 11 ndhA-K      | 19 rpl33       | 27 rps12       | 35 trnK-UUU       |
| 3 chlB          | 12 psal        | 20 rpoA        | 28 rps12intron | 36 trnP-GGG       |
| 4 chlL          | 13 rpl2intron  | 21 rpoB        | 29 rps16       | 37 trnT-GGU       |
| 5 chLN          | 14 rpl16intron | 22 rpoC1       | 30 rps16intron | 38 trnV-UAC       |
| 6 clpP          | 15 rpl20       | 23 rpoC1intron | 31 rps18       | 39 ycf1           |
| 7 clp intron1   | 16 rpl22       | 24 rpoC2       | 32 trnA-UGC    | 40 ycf2           |
| 8 clp intron1-2 | 17 rpl23       | 25 rps7        | 33 trnG-GCC    | 41 ycf4           |
| 9 infA          |                |                |                |                   |

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

(Downie and Palmer 1992; Palmer 1991). Draft plastid genome sequences for species from both of these genera confirm the IR loss in *Conopholis* but the situation in *Striga* remains uncertain because of assembly issues (C. dePamphilis, personal communication, 2011). If it turns out that both of these genera lack an IR these would be independent events because phylogenetic analyses of Orobanchaceae indicate that *Striga* and *Conopholis* are not sister genera (Bennett and Mathews 2006). Thus, there would be four independent IR losses in angiosperms, one in Fabaceae, one in Geraniaceae, and two in Orobanchaceae.

The fifth putative IR loss was reported in gymnosperms. Early work on *Pinus thunbergii* suggested that the IR was lost in this genus and that the loss was shared by all conifers (Raubeson and Jansen 1992a). However, it turns out that the IR in *Pinus* has been greatly reduced and consists of a 495 bp repeat that includes *trnI*-*cau* and a portion of *psbA* (Tsudzuki et al. 1992). Plastome sequences of other Pinaceae (Lin et al. 2010) have identified short IRs in three other genera, *Cathaya* – 429 bp, *Cedrus* – 236 bp, and *Keteleeria* – 267 bp. A similar situation occurs in another family of conifers, Cupressaceae (Hirao et al. 2008). The plastome sequence of *Cryptomeria japonica* has a residual IR that is only 114 bp and includes the *trnI*-*cau* gene (Lin et al. 2010).

### B. Gene and Intron Loss

The ancestral genome organization represented by *Cycas* among gymnosperms and

*Amborella* among angiosperms includes the full complement of genes and introns but there have been scattered gene and intron losses across seed plants based on the phylogenetic distribution of these events (Fig. 5.4). Within gymnosperms (Fig. 5.4d) most of these losses occurred in the gnetophytes and Pinaceae, including the loss of all 11 plastid-encoded subunits of NADH dehydrogenase. In the case of gnetophytes these losses are part of an overall downsizing of plastid genomes (McCoy et al. 2008; Wu et al. 2009). In angiosperms there is a high level of conservation of gene and intron content among the basal lineages with repeated bursts of losses in mostly unrelated lineages of monocot and eudicots (Fig. 5.4). In most cases, the causes of these losses or the fate of the genes has not been determined. In the case of intron loss, one mechanism that has been proposed involves reverse transcription of an edited RNA intermediate, followed by homologous recombination between an intron-less cDNA and the original intron-containing copy. This mechanism was suggested for the *atpF* intron loss in the angiosperm order Malphigiales (Daniell et al. 2008).

Although there have been many gene losses documented among seed plants (Fig. 5.4; Raubeson and Jansen 2005; Jansen et al. 2007; Magee et al. 2010) very few of these events have been investigated rigorously. It is widely known that plastid DNA transfer to the nucleus occurs at a high rate (Timmis et al. 2004; Matsuo et al. 2005; Noutsos et al. 2005) but only a few functional gene transfers to the nucleus have been

---

←

*Fig. 5.4.* Angiosperm and gymnosperm phylogenetic trees based on complete plastid genome sequences. The large maximum likelihood phylogram was constructed from 97 taxa based on 81 plastid gene sequences (adapted from Jansen et al. 2011; see this publication for details of the phylogenetic analyses). Scale bars for large tree and inset *d* indicate 0.05 substitutions per site. Inset *a* is a phylogram of gymnosperms adapted from Zhong et al. (2010). Insets *a*, *b*, and *c* are adapted from McNeal et al. (2007b), Magee et al. (2010), and Guisinger et al. (2011), respectively. Gene and intron losses and IR losses plotted on branches are based on Jansen et al. 2007, Magee et al. 2010, and on published papers on each of the sequenced genomes. The number of estimated inversions (in parentheses) is based on Jansen et al. (2007) and on GRIMM (<http://grimm.ucsd.edu/GRIMM/>; Bourque and Pevzner 2002) comparisons with *Cycas* (for gymnosperms) and *Amborella* (for angiosperms) for those taxa not included in Jansen et al. Asterisks indicate reported cases of nuclear transfer. *B* indicates those taxa that have biparental inheritance or the potential for biparental inheritance based on Corriveau and Coleman (1988), Birky (1995), Mogensen (1996), Zhang et al. (2003), Hagemann (2004), and Hu et al. (2008).

characterized in seed plants. The reason for the paucity of documented examples is twofold: (1) once a gene is transferred it must acquire the required sequences to properly regulate nuclear transcription and a transit peptide to target the product back to the plastid; and (2) there have been very few experimental studies to search for nuclear copies. Successful gene transfers to the nucleus in seed plants have been documented for only four genes (Fig. 5.4): *infA* in rosids (Millen et al. 2001), independent transfers of *rpl22* in Fabaceae (Gantt et al. 1991) and Fagaceae (Jansen et al. 2011), *rpl32* in some Salicaceae (Cusack and Wolfe 2007; Ueda et al. 2007), and *accD* in *Trifolium* (Magee et al. 2010). The loss of *rps16* from the plastomes of *Medicago* and *Populus* was determined to be a gene substitution because a nuclear-encoded, mitochondrial-targeted copy is now also targeted to the plastid (Ueda et al. 2008). The acetyl-CoA carboxylase (ACC) subunit D gene (*accD*) has been lost at least seven times among angiosperm plastid genomes (Fig. 5.4) and in one case (*Trifolium*) a copy was found in the nucleus (Magee et al. 2010). The fate of *accD* in grasses is different. In this case, the prokaryotic multisubunit enzyme has been replaced by plastid-targeted eukaryotic ACC (Konishi et al. 1996; Gornicki et al. 1997). A similar situation occurs in *Spinacia oleracea* where the prokaryotic plastid *rpl23* has been replaced by a eukaryotic cytosolic copy of this ribosomal protein (Bubunenko et al. 1994). Therefore, in the few examined cases of gene loss in seed plants, three different pathways have been detected: gene transfer to the nucleus (*infA*, *rpl22*, *rpl32* and *accD*), substitution by a nuclear-encoded, mitochondrial targeted gene product (RPS16), and substitution by a nuclear-encoded protein for a plastid gene product (ACC, RPL23).

### C. Gene Order Changes

As mentioned earlier the majority of seed plant plastomes lack any changes in gene order as is evident by the comparison of the *Cycas* and *Amborella* genomes (Fig. 5.3).

Thus, gene order has been highly conserved over long periods of time during the evolutionary history of seed plants. However, both gymnosperms and angiosperms have experienced multiple bursts of gene order change (Fig. 5.4); in some cases this has resulted from one or few inversions while in others there is evidence for more severe genomic upheaval. This pattern is most evident among angiosperms, partly because the amount of plastome sequence data available is much greater. The phylogenetic distribution of inferred inversions (Fig. 5.4) indicates a long period of genomic stability starting from the early diverging angiosperms, monocots, and eudicots, followed by isolated instances of gene order changes in more derived lineages, especially among eudicots. Three of the most striking examples of extensive gene order changes among photosynthetic angiosperm lineages occur in the Campanulaceae, Fabaceae, and Geraniaceae and are summarized briefly below.

Campanulaceae (sensu APG III 2009, including Lobeliaceae) have experienced a high degree of gene order change. Although only one plastome sequence for *Trachelium caeruleum* has been published (Haberle et al. 2008), draft genomes have been completed for several other genera (Haberle 2006) and restriction site and gene maps have been published for many others (Knox and Palmer 1999; Cosner et al. 2004). The most extensive comparisons included gene maps for 18 genera of Campanulaceae (Cosner 1993; Cosner et al. 2004) and these authors estimated that the gene order changes were due to a minimum of 42 inversions, 18 large insertions (>5 kb) of unknown origin, five IR expansions and contractions, and several putative transpositions. The complete genome sequence for *Trachelium* (Haberle et al. 2008), the least rearranged taxon examined by Cosner et al. (2004), confirmed that at least seven inversions are present in this genome, but it did not identify any evidence for transposition as a mechanism for gene order changes.

Fabaceae are known to exhibit a number of unusual phenomena in their plastomes (Fig. 5.4b), including the loss of the IR in a

large clade of papilionoids (Wojciechowski et al. 2004), transfer of genes to the nucleus (Gantt et al. 1991; Magee et al. 2010), intron losses (Doyle et al. 1995; Jansen et al. 2008), and inversions (Doyle et al. 1996; Bruneau et al. 1990). *Trifolium* has experienced the most extensive genomic reconfigurations within the family, including the loss of the IR, 14–18 inversions, duplication of parts or all of nine genes, and insertions of 20 kb of novel DNA (Milligan et al. 1989; Cai et al. 2008; Figs. 5.3, 5.4).

Geraniaceae have been examined more extensively in terms of complete plastome sequences, which are now available for five species from four of the five genera in the family (Chumley et al. 2006; Guisinger et al. 2011; Blazier et al. 2011). Like Campanulaceae, two mechanisms are responsible for gene order changes in this group, inversions and expansion/contraction of the IR. For *Pelargonium hortorum* Chumley et al. (2006) developed an evolutionary scenario that required a minimum of 12 inversions and eight IR boundary changes. The situation in Geraniaceae is so complex among the five sequenced plastomes that it is as yet not possible to reconstruct an evolutionary model to explain the gene order differences among these plastomes. This will require sequencing many more genomes within each genus so that intermediate stages in gene order can be reconstructed more reliably.

## V. Patterns and Rates of Nucleotide Substitutions

### A. Sequence Evolution in Coding Regions Versus Intergenic Regions and Introns

The relative frequency of base substitutions and insertions/deletions (indels) in plastid genomes has been examined among both closely related and distantly related species. Most studies (e.g., Golenberg et al. 1993; Ingvarsson et al. 2003) compared selected regions of the genome and concluded that indels occur at an equal or slightly higher

rate than nucleotide substitutions, especially in comparisons among closely related species. The availability of complete plastome sequences opened up the opportunity to perform genome-wide comparisons, and several such studies have been completed in different angiosperm families, including Asteraceae (Timme et al. 2007), Poaceae (Masood et al. 2004; Saski et al. 2007; Yamane et al. 2006), Ranunculaceae (Kim et al. 2009), and Solanaceae (Kahlau et al. 2006; Chung et al. 2006; Daniell et al. 2006). All of these comparisons confirmed that the relative frequencies of indels are similar or slightly higher than nucleotide substitutions. In Poaceae, genome-wide comparisons demonstrated that most indels occurred in intergenic spacers (56–64%), with coding regions (10–19%) and introns (25–26%) having less than half the number of indels. In terms of genomic region most indels are concentrated in the LSC (84%), followed by the SSC (12%) and IR (4%). The vast majority of the indels represent single or few bp changes that are likely caused by slip-strand mispairing during DNA replication.

A number of genome-wide comparisons have also examined sequence divergence across plastomes by partitioning the genome into different regions; coding, intron, and intergenic spacer (Timme et al. 2007; Saski et al. 2007; Kim et al. 2009; Daniell et al. 2006). Rates of change varied considerably within these regions in different angiosperm families, however, coding regions were most highly conserved, followed by introns, and intergenic spacers. For example, in the Asteraceae the average p-distance, the proportion of substitution changes between two sequences, was 0.057 for intergenic spacers, 0.030 for introns, 0.022 for protein coding genes, and 0.008 for RNA genes (Timme et al. 2007). Such comparisons have been valuable for deciding which genes or regions to utilize for phylogenetic studies within angiosperms (Shaw et al. 2007; Timme et al. 2007). Comparisons of sequence divergence across intergenic spacers also have important implications for plastid biotechnology; it has been demonstrated that using homologous

flanking and regulatory sequences for plastid transformation significantly increases transgene integration and expression of foreign proteins, respectively (Ruhlman et al. 2010).

### *B. Rates of Sequence Evolution in Protein Coding Genes*

Early evolutionary rate comparisons of plastid coding regions of photosynthetic seed plants were based on a limited number of genes and/or genomes. Several general observations were made from these analyses: synonymous substitution rates are low in plastid DNA relative to nuclear DNA (Wolfe et al. 1987; Gaut 1998); rates can vary among lineages (Wolfe et al. 1987), among codon positions (Gaut et al. 1993), and among genes in different functional groups (Palmer 1991; Gaut et al. 1993); substitution rates in the three regions of the plastome vary, with genes in the IR having a lower rate of synonymous substitutions relative to those in the SSC and LSC regions (Clegg et al. 1984; Wolfe et al. 1987; Gaut 1998; Perry and Wolfe 2002); base composition often plays an important role in plastid DNA sequence evolution (Olmstead et al. 1998; Decker-Walters et al. 2004) resulting in mutations that are spatially biased across the genome. Several earlier comparisons of non-photosynthetic seed plants also focused on a limited number of genes and taxa (dePamphilis et al. 1997; Wolfe et al. 1992; Young and dePamphilis 2000, 2005). Not surprisingly, these studies demonstrated that non-photosynthetic plants have elevated rates of nucleotide substitution, largely due to relaxed selection. For some genes both synonymous ( $dS$ ) and non-synonymous ( $dN$ ) rates increased, suggesting that other forces, including generation time, speciation rate, and population size, may be affecting rates, especially at synonymous sites. More recently there have been a number of genome-wide rate comparisons performed and these have provided a much more comprehensive view of rates of plastome sequence evolution. The most extensive comparisons have been performed in selected gymnosperms (Wu et al. 2007, 2009), the

angiosperm families Poaceae (Chang et al. 2006; Zhong et al. 2009; Guisinger et al. 2010) and Geraniaceae (Guisinger et al. 2008), and the non-photosynthetic genus *Cuscuta* (McNeal et al. 2007a). Some notable observations from two angiosperm families, Poaceae and Geraniaceae, are described briefly below.

Genome-wide comparisons (Chang et al. 2006; Zhong et al. 2009; Guisinger et al. 2010) confirmed that rates of change were accelerated on the branch leading to Poaceae, while internal Poaceae branches have experienced a significant rate deceleration. Furthermore, genes involved in gene expression and photosynthesis metabolism have higher values of  $dN$ , and several genes appear to be under positive selection, i.e. the  $dN/dS$  ratio is greater than one. The precise timing of this rate acceleration is not clear since only two of the 16 families of Poales (Poaceae and Typhaceae) have complete plastome sequences available. Rate heterogeneity in Poaceae could be due to one or more factors, including relaxed or positive selection, mutational bias, altered DNA repair, or differences in levels of gene expression.

The situation in the Geraniaceae is novel for two reasons. First, this is the only seed plant lineage where extreme rate acceleration has been documented in both the mitochondrial (Parkinson et al. 2005; Mower et al. 2007) and plastid genomes (Guisinger et al. 2008). Second, plastid genomes in this family are among the most highly rearranged of any seed plant lineage. Analyses of 72 protein coding genes for nine Geraniaceae and 38 other angiosperms detected both locus and lineage specific rate heterogeneity (Guisinger et al. 2008). Values of  $dN$  were highly accelerated in the branch leading to the Geraniaceae as well as within several lineages within the family for ribosomal protein and RNA polymerase genes. In addition,  $dN/dS$  ratios were significantly higher for these two functional classes of genes and for ATPase genes. It was hypothesized that these unusual phenomena were caused by a combination of aberrant DNA repair and altered levels of gene expression.

### *C. Correlation Between Rates of Nucleotide Substitutions and Genomic Rearrangements*

A significant positive correlation between rates of nucleotide substitutions and genomic rearrangements (indels, gene/intron losses, and inversions) was previously identified across angiosperms (Jansen et al. 2007). This pattern is evident in Fig. 5.4, which plots the distribution of genomic changes on a phylogram that was constructed using sequences of 81 genes from 97 seed plant plastomes. It is evident that early diverging lineages of angiosperms, eudicots, and monocots had very stable plastomes even though there was rapid diversification in morphology, anatomy, and reproductive biology among these lineages. This plastomic stasis was followed by repeated bursts of change in both rates of nucleotide substitution, gene order and gene content in disparate and more derived eudicot and monocot lineages. More extensive studies of two unrelated angiosperm families, Geraniaceae and Poaceae (Guisinger et al. 2008, 2010, 2011), have identified similar positive correlations between rates of nucleotide substitutions and genomic rearrangements.

A correlation between rates of nucleotide substitution and genomic rearrangements has been previously identified in bacterial (Belda et al. 2005) and animal mitochondrial genomes (Shao et al. 2003; Xu et al. 2006). The mitochondrial studies suggested several possible mechanisms to explain this correlation, but argued that accuracy of DNA replication is the most likely cause. In the case of plastid genomes, it was suggested that accelerated rates of genome rearrangements and nucleotide substitutions were possibly caused by aberrant DNA repair mechanisms (Jansen et al. 2007; Guisinger et al. 2008, 2010, 2011). Four classes of nuclear-encoded genes have been implicated in DNA repair in plastids of angiosperms: chloroplast mutator (CHM/MSH1), RecA-like homologs, OSBs (organellar single-stranded DNA-binding proteins), and the Whirlies (reviewed in Maréchal and Brisson 2010). These genes produce proteins that

suppress recombination between repeated DNA sequences, and thus provide stability to the genome by preventing illegitimate recombination. Mutations in genes encoding either Whirly (Maréchal et al. 2009) or RecA (Rowan et al. 2010) have been shown to generate plastome rearrangements. We suggest that plastomes with accelerated rates of nucleotide substitutions and genomic rearrangements may result from mutations in nuclear-encoded DNA repair and/or replication genes. The prevalence of large numbers of dispersed repeats in highly rearranged plastomes (Chumley et al. 2006; Haberle et al. 2008; Cai et al. 2008; Guisinger et al. 2011) is consistent with this idea.

There is also a correlation between lineages with accelerated rates of change and biparental inheritance. Although this correlation has not been tested rigorously, it is evident from the distribution of biparental inheritance on the plastome phylogram for angiosperms (Fig. 5.4). The evolutionary significance of this correlation is not obvious. One possibility is that biparental inheritance provides a mechanism for bringing together multiple plastid types, and these could undergo intermolecular recombination to produce plastomes with novel organization. There is limited evidence available demonstrating plastid recombination (e.g., Medgyesy et al. 1985). Clearly more detailed investigations are needed to confirm this correlation and to examine its possible role in enhancing plastome diversity.

## **VI. Phylogenetic Utility of Plastome Data for Resolving Relationships Among Seed Plants**

Most molecular phylogenetic investigations of seed plant relationships have relied on features of the plastid genome (reviewed in Raubeson and Jansen 2005). Early studies from 1985 to 1995 used restriction site and gene mapping comparisons to examine phylogenetic relationships at a wide range of taxonomic levels. Restriction site/fragment analyses were mostly utilized at the generic

level or below, whereas gene map comparisons were valuable for defining major clades. Several early examinations of gene order identified one or a few inversions that were extremely valuable in defining major lineages of seed plants. This included a 22 kb inversion that identified the subfamily Barnadesioideae as the earliest diverging lineage in the largest angiosperm family Asteraceae (Jansen and Palmer 1987), a 30 kb inversion that placed lycophytes as the earliest clade of land plants (Raubeson and Jansen 1992b), and 50 kb inversion that supports the monophyly of a major clade of papilionoid legumes (Doyle et al. 1996). In addition to inversions, several other structural changes of plastomes were utilized to define major groups of seed plants. The loss of the IR was used to support the monophyly of conifers (Raubeson and Jansen 1992a) and the legume IRLC (Lavin et al. 1990; Wojciechowski et al. 2004). A number of gene and intron losses were also identified and in some cases these were powerful phylogenetic markers (e.g., *rpl22* gene loss in all legumes; Doyle et al. 1995), whereas in other instances such changes were shown to occur multiple times (*rpoC1* intron loss multiple times across angiosperms; Downie et al. 1996), limiting their utility as phylogenetic characters. In a few groups extensive structural changes in their plastomes have been utilized for phylogenetic analyses. The best example is the angiosperm family Campanulaceae. Gene maps for 18 genera identified a total of 84 structural changes, including inversions, putative transpositions, insertions, and gene and intron losses. Despite the extreme genomic complexity phylogenetic trees generated from these data exhibited very little homoplasy and were congruent with trees generated from DNA sequence data for the same taxa (Cosner et al. 2004).

During the past 10 years the field of plastid molecular phylogenetics has changed dramatically due the availability of rapid, less expensive methods for amassing large quantities of DNA sequence data. Thus, rather than relying on generating sequences

for only a handful of markers or using the limited data from restriction site and gene mapping, it is now possible to produce large amounts of genomic data for phylogenetic studies. This has resulted in the production of very large data sets both in terms of number of genes and taxa for examining phylogenetic relationships (Jansen et al. 2007, 2011; Moore et al. 2007, 2010; Lin et al. 2010; Zhong et al. 2010). These studies focused entirely on plastome sequencing by utilizing isolated plastid DNA and either standard Sanger sequencing or 454 pyrosequencing. More recently, plastid genome sequencing for phylogeny reconstruction has shifted to sequencing platforms, such as Illumina, that utilizes shorter reads of up to 75–100 bp (Cronn et al. 2008; Parks et al. 2009). This approach, combined with multiplexing samples, has greatly reduced the cost for generating draft plastid genome sequences. Another recent development in plastome sequencing is the use of total genomic DNA as template for next generation sequencing (Nock et al. 2010; Atherton et al. 2010). The outcome of these new developments is that we have made huge improvements in our understanding of phylogenetic relationship among seed plants and some of these are described below.

One of the most controversial, remaining issues in seed plant phylogeny concerns the position of the three morphologically unique genera of gnetophytes, *Ephedra*, *Gnetum*, and *Welwitschia* (reviewed in Burleigh and Mathews 2004; Mathews 2009). Morphological studies suggested that gnetophytes were sister to angiosperms (Anthophyte hypothesis; Doyle and Donoghue 1986) but most molecular phylogenetic studies do not support this relationship. The situation became more contentious because molecular phylogenetic studies supported three different hypotheses of relationships depending on which genes and taxa were included in the analyses: gnetophytes sister to conifers (Gnetifer hypothesis); gnetophytes sister to Pinaceae (Gnepine hypothesis); gnetophytes sister to Cupressaceae (Gnecup hypothesis). Several recent papers have utilized plastome sequences to try to



resolve this issue but in all cases the number of genomes available was limited and the issue remains unresolved (Wu et al. 2007; McCoy et al. 2008). The major problem with such limited sampling is that it can cause artifacts in phylogenetic tree construction, often arising from long branch attraction. Previous phylogenetic studies using complete plastome sequences in angiosperms has shown that increased taxon sampling can alleviate issues associated with long branch attraction (Leebens-Mack et al. 2005).

The most comprehensive phylogenetic analysis of complete plastome sequences to resolve the position of gnetophytes included only eight gymnosperms and five outgroups (see inset in Fig. 5.4; Zhong et al. 2010). This analysis involved 56 protein coding genes shared among these 13 taxa. Initial results from analyses of these data supported the Gnecup hypothesis but the relationship was the result of long branch attraction between the single Cupressaceae genome and gnetophytes. Removal of the fastest evolving proteins from the dataset, many of which had parallel amino acid substitutions between gnetophytes and Cupressaceae, resulted in trees that supported the Gnepine hypothesis. This hypothesis is also supported by the fact that both gnetophytes and Pinaceae have lost all 11 NADH dehydrogenase genes (Braukmann et al. 2009) and *rps16* (Wu et al. 2007, 2009). Clearly additional gymnosperm plastome sequences are needed, especially from conifers, before the position of gnetophytes can be resolved.

Plastome sequences have been utilized extensively to examine phylogenetic relationships among the major clades of angiosperms. The earliest studies utilized 61 protein coding genes from a limited number of plastome sequences and suffered from some of the same issues mentioned above, especially long branch attraction (Goremykin et al. 2003, 2004, 2005). As more plastome sequences were completed it became evident that these data provide a valuable resource for resolving phylogenetic relationships among angiosperms (Leebens-Mack et al. 2005; Cai et al. 2006; Ruhlman et al. 2006; Hansen et al. 2007b). The most comprehen-

sive studies examined up to 97 plastid genomes using 81–83 genes. These studies have provided strong support for resolving relationships among all major clades of angiosperms (Fig. 5.4; Jansen et al. 2007, 2011; Moore et al. 2007, 2010), including the placement of *Amborella* as the earliest diverging lineage, the position of magnoliids as sister to Chloranthaceae and this group sister to a large clade that includes both eudicots and monocots, placement of *Ceratophyllum* sister to eudicots, sister relationship between monocots and eudicots, and resolution of relationships among many of the orders within both monocots and eudicots. A number of ongoing plastome sequencing projects in these major clades will provide much new data for resolving the angiosperm tree of life.

## VII. Conclusions and Future Directions

Rapid improvement in DNA sequencing technology at a much lower cost has generated a glut of plastome data for plant biologists. For plant evolutionary biologists, plastome sequences have provided reams of data for resolving phylogenetic relationships among the major clades of seed plants and for examining rates and patterns of sequence evolution. These two endeavors are closely intertwined since an understanding of how sequences evolve is essential for using them correctly for making phylogenetic inferences. Some lineages, especially among gymnosperms, are still underrepresented but projects are underway that will fill these gaps and ultimately generate a tree of life for seed plants. The knowledge we have gained about the diversity of plastome organization among seed plants is providing the framework for examining the mechanisms of change in these genomes. We have confirmed that across most seed plant lineages there is an incredible stability of genome organization in terms of overall architecture, size, gene/intron content, and gene order. However, several unrelated groups have experienced

genomic upheaval and these taxa are positioned to illuminate the mechanisms of change in plastomes. Future investigations should focus on these extraordinary lineages by sequencing more representative in these groups, and by examining their nuclear-plastid interactions. Such studies will reveal critical insights into how these genomes have co-evolved to control the many biochemical processes that are coordinated between nuclear and plastid genomes in seed plants. Comparisons of these natural mutant lineages, combined with experimental studies of plastid-nuclear interactions using plastid genetic engineering, will lead to new insights into compartmental crosstalk, which is critical for plant cells to function properly.

## Acknowledgments

Support was provided by the National Science Foundation (DEB-0717372 to RKJ and IOS-1027259 to RKJ and TAR) and the Fred G. Gloeckner Foundation (to TAR and RKJ). The authors thank J. Chris Blazier and Mao-Lun Weng for comments on an earlier version of the manuscript.

## References

- Angiosperm Phylogeny Group [APG] (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161:105–121
- Atherton RA, McComish BJ, Shepherd LD, Berry LA, Albert NW, Lockhart PJ (2010) Whole genome sequencing of enriched chloroplast DNA using the Illumina GAI platform. *BMC Plant Methods* 6:22
- Belda E, Moya A, Silva FJ (2005) Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria. *Mol Biol Evol* 22:1456–1467
- Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16:1661–1666
- Bennett JR, Mathews S (2006) Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *Am J Bot* 93:1039–1051
- Birky CWJR (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc Natl Acad Sci USA* 92:11331–11338
- Blazier JC, Guisinger MM, Jansen RK (2011) Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Mol Biol* 76:263–272
- Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) *Cell and molecular biology of plastids*. Springer, Berlin/Heidelberg, pp 1610–2096
- Boudreau E, Turmel M (1995) Gene rearrangements in *Chlamydomonas* chloroplast DNAs are accounted for by inversions and by expansion/contraction of the inverted repeat. *Plant Mol Biol* 27:351–364
- Bourque G, Pevzner P (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res* 12:26–36
- Braukmann TW, Kuzmina M, Stefanovic 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
- Bruneau A, Doyle JJ, Palmer JD (1990) A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae). *Syst Bot* 15:378–386
- Bubunenko MG, Schmidt J, Subramanian AR (1994) Protein substitution in chloroplast ribosome evolution. A eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. *J Mol Biol* 240:28–41
- Burleigh JG, Mathews S (2004) Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *Am J Bot* 91:1599–1613
- Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson J, dePamphilis CW, Jansen RK (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogeny of magnoliids. *BMC Evol Biol* 6:77
- Cai Z, Guisinger M, Kim HG, 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
- 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 S-M (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
- Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol* 58:424–441
- Chiu WL, Sears BB (1993) Plastome-genome interactions affect plastid transmission in *Oenothera*. *Genetics* 133:989–997

- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol* 23:2175–2190
- Chung H-J, Jung JD, Park H-W, Kim J-H, Cha HW, Min SR, Jeong W-J, Liu JR (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with *Solanaceae* species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Rep* 25:1369–1379
- Clegg MT, Brown ADH, Whitfield PR (1984) Chloroplast DNA diversity in wild and cultivated barley: implications for genetic conservation. *Genet Res* 43:339–343
- Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24:861–862
- Corriveau JS, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *Am J Bot* 75:1443–1458
- Cosner ME (1993) Phylogenetic and molecular evolutionary studies of chloroplast DNA variation in the Campanulaceae. PhD thesis, The Ohio State University, Columbus
- Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet* 31:419–429
- Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evol Biol* 4:27
- Cronn R, Liston A, Parks M, Gernandt D, Shen R, Mockler T (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res* 36:e122
- Cruzan MB, Arnold ML, Carney SE, Wollenberg KR (1993) CpDNA inheritance in interspecific crosses and evolutionary inference in Louisiana irises. *Am J Bot* 80:344–350
- Cusack BP, Wolfe KH (2007) When gene marriages don't work: divorce by subfunctionalization. *Trends Genet* 23:270–272
- Daniell H, Lee S-B, Grevich J, Saski C, Guda C, Tompkins J, Jansen RK (2006) Complete chloroplast genome sequences of *Solanum tuberosum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor Appl Genet* 112:1503–1518
- Daniell H, Wurdack KJ, Kanagaraj A, Lee S-B, Saski C, Jansen RK (2008) The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and multiple losses of the *atpF* intron in Malpighiales. *Theor Appl Genet* 116:723–737
- Darling AE, Mau B, Perna NT (2010) Progressive mauve: multiple genome alignment with gene gain, loss, and rearrangement. *PLoS One* 5:e11147
- Decker-Walters DS, Chung SM, Staub JE (2004) Plastid sequence evolution: a new pattern of nucleotide substitutions in the Cucurbitaceae. *J Mol Evol* 58:606–614
- Delannoy E, Fujii S, des Frances CC, Brundrett M, Small I (2011) Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28:2077–2086
- dePamphilis CW, Young ND, Wolfe AD (1997) Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. *Proc Natl Acad Sci USA* 94:7367–7372
- Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ (eds) *Molecular systematics of plants*. Chapman and Hall, New York, pp 14–35
- Downie SR, Llanas E, Katz-Downie DS (1996) Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNAs. *Syst Bot* 21:135–151
- Doyle JA, Donoghue MJ (1986) Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Bot Rev* 52:321–431
- Doyle JJ, Doyle JL, Palmer JD (1995) Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Syst Bot* 20:272–294
- Doyle JJ, Doyle JL, Palmer JD (1996) The distribution and phylogenetic significance of a 50-kb chloroplast DNA inversion in the flowering plant family Leguminosae. *Mol Phylogenet Evol* 5:429–438
- 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>
- Fan WH, Woelfle MA, Mosig G (1995) Two copies of a DNA element, 'Wendy', in the chloroplast chromosome of *Chlamydomonas reinhardtii* between rearranged gene clusters. *Plant Mol Biol* 29:63–80
- Funk HT, Berg S, Krupinska K, Maier UG, Krause K (2007) Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol* 7:45

- 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
- Gao L, Su Y-J, Wang T (2010) Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *J Syst Evol* 48:77–93
- Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK (ed) *Evolutionary biology*. Plenum Press, New York, pp 93–120
- Gaut BS, Muse SV, Clegg MT (1993) Relative rates of nucleotide substitution in the chloroplast genome. *Mol Phylogenet Evol* 2:89–96
- Golenberg EM, Clegg MT, Durbin ML, Doebley J, Ma DP (1993) Evolution of a noncoding region of the chloroplast genome. *Mol Phylogenet Evol* 2:52–64
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 20:1499–1505
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH (2004) The chloroplast genome of *Nymphaea alba*: whole-genome analyses and the problem of identifying the most basal angiosperm. *Mol Biol Evol* 21:1445–1454
- Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Mol Biol Evol* 22:1813–1822
- Gornicki P, Faris J, King I, Podkowinski J, Gill B, Haselkorn R (1997) Plastid localized acetyl-Co-A carboxylase of bread wheat is encoded by a single gene on each of the three ancestral chromosome sets. *Proc Natl Acad Sci USA* 94:14179–14184
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet* 252:195–206
- 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
- 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 USA* 105:18424–18429
- 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
- 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
- Haberle RC (2006) *Phylogeny and comparative chloroplast genomics of the Campanulaceae*. PhD thesis, University of Texas, Austin
- Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol* 66:350–361
- Hagemann R (2004) The sexual inheritance of plant organelles. In: Daniell H, Chase C (eds) *Molecular biology and biotechnology of plant organelles*. Springer, Heidelberg, pp 93–113
- Hansen AK, Escobar LK, Gilbert LE, Jansen RK (2007a) Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): implications for phylogenetic studies. *Am J Bot* 94:42–46
- Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl V, Boore JL, Jansen RK (2007b) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol Phylogenet Evol* 45:547–563
- Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988) Trans-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*. *Proc Natl Acad Sci USA* 85:372–376
- Hirao T, Watanabe A, Kurita M, Kondo T, Takata K (2008) Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genome structure of coniferous species. *BMC Plant Biol* 8:70
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice *Oryza sativa* chloroplast genome – intermolecular recombination between distinct transfer-RNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol Gen Genet* 217:185–194
- Hu YC, Zhang Q, Rao GY, 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
- Ingvarsson PK, Ribstein S, Taylor DR (2003) Molecular evolution of insertions and deletions in the chloroplast genome of *silene*. *Mol Biol Evol* 20:1737–1740

- Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc Natl Acad Sci USA* 84:5818–5822
- Jansen RK, Cai Z, Raubeson LA, Daniell H, de Pamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104:19369–19374
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, Daniell H (2008) Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Fabaceae). *Mol Phylogenet Evol* 48:1204–1207
- Jansen RK, Saski C, Lee S-B, 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
- 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
- Kanno A, Watanabe K, Nakamura I, Hirai A (1993) Variations in chloroplast DNA from rice (*Oryza sativa*): differences between deletions mediated by short direct-repeat sequences within a single species. *Theor Appl Genet* 86:579–584
- Kawata M, Harada T, Shimamoto Y, Oono K, Takaiwa F (1997) Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). *Curr Genet* 31:179–184
- Khan A, Khan IA, Asif H, Azim MK (2010) Current trends in chloroplast genome research. *Afr J Biotechnol* 9:3494–3500
- Kim K-J, Lee H-L (2004) Complete chloroplast genome sequence from Korean Ginseng (*Panax schiseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res* 11:247–261
- Kim Y-K, Park C-W, Kim K-J (2009) Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthus saniculifolia*, and its evolutionary implications. *Mol Cells* 27:365–381
- Knox EB, Palmer JD (1999) The chloroplast genome arrangement of *Lobelia thuliana* (Lobeliaceae): expansion of the inverted repeat in an ancestor of the Campanulales. *Plant Syst Evol* 214:49–64
- Kolodner R, Tewari KK (1979) Inverted repeats in chloroplast DNA of higher plants. *Proc Natl Acad Sci USA* 76:41–45
- Konishi T, Shinohara K, Yamada K, Sasaki Y (1996) Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both prokaryotic and eukaryotic forms of this enzyme. *Plant Cell Physiol* 37:117–122
- Kuroiwa T (2010) Review of cytological studies on cellular and molecular mechanisms of uniparental (maternal or paternal) inheritance of plastid and mitochondrial genomes induced by active digestion of organelle nuclei (nucleoids). *J Plant Res* 123:207–230
- Lavin M, Doyle JJ, Palmer JD (1990) Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* 44:390–402
- Lee DJ, Blake TK, Smith SE (1988) Biparental inheritance of chloroplast DNA and the existence of heteroplasmic cells in alfalfa. *Theor Appl Genet* 76:545–549
- Lee H-L, Jansen RK, Chumley TW, Kim K-J (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol Biol Evol* 24:1161–1180
- Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, dePamphilis CW (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Biol Evol* 22:1948–1963
- 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
- Lin C-P, Huang J-P, Wu C-S, Hsu C-Y, Chaw S-M (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol Evol* 2:504–517
- Liu Q, Xue Q (2005) Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J Genet* 84:55–62
- 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
- Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytol* 186:299–317
- Maréchal A, Parent J, Véronneau-Lafortune F, Joyeux A, Lang BF, Brisson N (2009) Whirly proteins

- maintain plastid genome stability in *Arabidopsis*. Proc Natl Acad Sci USA 106:14693–14698
- Masood MS, Nishikawa T, Fukuoka S, Njenga PK, Tsudzuki T, Kadowaki K (2004) The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene 340:133–139
- Mathews S (2009) Phylogenetic relationships among seed plants: persistent questions and the limits of molecular data. Am J Bot 96:228–236
- 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
- McCoy SR, Kuehl JV, Boore JL, Raubeson LA (2008) The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evol Biol 8:130
- McNeal JR, Arumugunathan K, Kuehl JV, Boore JL, de Pamphilis CW (2007a) Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). BMC Biol 5:55
- McNeal JR, Kuehl JV, Boore JL, dePamphilis CW (2007b) Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant Biol 7:57
- Medgyesy P, Fejes E, Maliga P (1985) Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc Natl Acad Sci USA 82:6960–6964
- Metzlaff M, Pohlheim F, Börner T, Hagemann R (1982) Hybrid variegation in the genus *Pelargonium*. Curr Genet 5:245–249
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermlin LS, Wolfe KH (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13:645–658
- Milligan BG, Hampton JN, Palmer JD (1989) Dispersed repeats and structural reorganization in subclonal chloroplast DNA. Mol Biol Evol 6:355–368
- Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. Am J Bot 83:383–404
- 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 USA 104:19363–19368
- 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 USA 107:4623–4628
- Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. J Mol Evol 37:273–280
- Morton BR (1994) Codon use and the rate of divergence of land plant chloroplast genes. Mol Biol Evol 11:231–238
- Morton BR (1998) Selection on the codon bias of chloroplast and cyanobacterial genes in different plant and algal lineages. J Mol Evol 46:449–459
- Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 7:135
- Mráček J (2005) Investigation of interspecific genome–plastome incompatibility in *Oenothera* and *Passiflora*. PhD dissertation, University of Munich, Germany
- Nagata N (2010) Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. J Plant Res 123:193–199
- Nock C, Waters D, Edwards M, Bowen W, Rice N, Cordeiro G, Henry R (2010) Chloroplast genome sequences from total DNA for exploring plant relationships. Plant Biotechnol J 9:328–333
- Noutsos C, Richly E, Leister D (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res 15:616–628
- Odintsova MS, Yurina NP (2003) Plastid genomes of higher plants and algae: structure and functions. Mol Biol 37:649–662
- Ohba K, Iwakawa M, Ohada Y, Murai M (1971) Paternal transmission of a plastid anomaly in some reciprocal crosses of Suzi, *Cryptomeria japonica* D. Don. Silvae Genet 210:101–107
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umeshima K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of Liverwort *Marchantia-Polymorpha* chloroplast DNA. Nature 322:572–574
- Oldenburg DJ, Bendich AJ (2004) Changes in the structure of DNA molecules and the amount of DNA per plastid during chloroplast development in maize. J Mol Biol 344:1311–1330
- Olmstead RG, Reeves PA, Yen AC (1998) Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants II: DNA sequencing. Kluwer Academic, Norwell, pp 164–187
- Palmer JD (1983) Chloroplast DNA exists in two orientations. Nature 301:92–93

- Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Hermann RG (ed) *The molecular biology of plastids. Cell culture and somatic cell genetics of plants*. Springer, Vienna, pp 5–53
- Palmer JD, Stein DB (1986) Conservation of chloroplast genome structure among vascular plants. *Curr Genet* 10:823–833
- Palmer JD, 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
- Palmer JD, Osorio B, Thompson WF (1988) Evolutionary significance of inversions in legume chloroplast DNAs. *Curr Genet* 14:65–74
- Pandey K, Grant J, Williams E (1987) Interspecific hybridisation between *Trifolium repens* and *T. uniflorum*. *Aust J Bot* 35:171–182
- Parkinson CL, Mower JP, Qiu Y-Q, Shirk AJ, Song K, Young ND, de Pamphilis CW, Palmer JD (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evol Biol* 5:73
- Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7:84
- Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J Mol Evol* 55:501–508
- Perry AS, Brennan S, Murphy DJ, Kavanagh TA, Wolfe KH (2002) Evolutionary re-organization of a large operon in Adzuki Bean chloroplast DNA caused by inverted repeat movement. *DNA Res* 9:157–162
- Raubeson LA, Jansen RK (1992a) A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochem Syst Ecol* 20:17–24
- Raubeson LA, Jansen RK (1992b) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* 255:1697–1699
- Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ (ed) *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. CABI, Cambridge, MA, pp 45–68
- 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
- Ravi V, Khurana JP, Tyagi AK, Khurana P (2008) An update on chloroplast genomes. *Plant Syst Evol* 271:101–122
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* 34:4537–4545
- Rowan BA, Oldenburg DJ, Bendich AJ (2010) RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *J Exp Bot* 61:2575–2588
- Ruhlman T, Lee S-B, Jansen RK, Hostetler JB, Tallon LJ, Town CD, Daniell H (2006) Complete chloroplast genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics* 7:222
- 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
- Saski C, Lee S-B, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera* and comparative analyses with other grass genomes. *Theor Appl Genet* 115:571–590
- Shao R, Dowton M, Murrell A, Barker SC (2003) Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Mol Biol Evol* 20:1612–1619
- Shaver JM, Oldenburg DJ, Bendich AJ (2006) The structure of chloroplast DNA molecules and the effects of light on the amount of chloroplast DNA during development in *Medicago truncatula*. *Plant Physiol* 146:1064–1074
- 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
- Shimada H, Sugiura M (1991) Fine-structural features of the chloroplast genome – comparison of the sequenced chloroplast genomes. *Nucleic Acids Res* 19:983–995
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchishinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide-sequence of the tobacco chloroplast genome – its gene organization and expression. *EMBO J* 5:2043–2049
- Snijder RC, Brown FS, van Tuyk JM (2007) The role of plastome-genome incompatibility and biparental plastid inheritance in interspecific hybrids in the genus *Zantedeschia* (Araceae). *Floricult Ornament Biotechnol* 1:150–157
- Soliman K, Fedak G, Allard RW (1987) Inheritance of organelle DNA in barley and *Hordeum X Secale* intergeneric hybrids. *Genome* 29:867–872

- Szmidt AE, Alden T, Hallgren JE (1987) Paternal inheritance of chloroplast DNA in *Larix*. *Plant Mol Biol* 9:59–64
- Testolin R, Cipriani G (1997) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. *Theor Appl Genet* 94:897–903
- Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparison of the first two sequenced chloroplast genomes in Asteraceae: Lettuce and Sunflower. *Am J Bot* 94:302–312
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5:123–135
- Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M, Wakasugi T, Sugiura M (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol Gen Genet* 232:206–214
- Ueda M, Fujimoto M, Arimura S-I, Murata J, Tsutsumi N, Kadowaki K-I (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
- Ueda M, Fujimoto M, Takahashi H, Arimura S-I, Tsutsumi N, Kadowaki K-I (2008) Substitution of the gene for chloroplast *rps16* was assisted by generation of dual targeting signal. *Mol Biol Evol* 25:1566–1575
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA* 91:9794–9798
- Wall DP, Herbeck JT (2003) Evolutionary patterns of codon usage in the chloroplast gene *rbcL*. *J Mol Evol* 56:673–688
- 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
- Wolf PG, Der JP, Duffy AM, Davidson JB, Grusz AL, Pryer KM (2011) The evolution of chloroplast genes and genomes in ferns. *Plant Mol Biol* 76:251–261
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058
- Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci USA* 89:10648–10652
- Wu CS, Wang YN, Liu SM, Chaw SM (2007) Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. *Mol Biol Evol* 24:1366–1379
- Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol* 52:115–124
- Xu W, Jameson D, Tang B, Higgs PG (2006) The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *J Mol Evol* 63:375–392
- Yamane K, Yano K, Kawahara T (2006) Pattern and rate of indel evolution from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Res* 13:197–204
- Yang TW, Yang YA, Xiong Z (2000) Paternal inheritance of chloroplast DNA in interspecific hybrids in the genus *Larrea* (Zygophyllaceae). *Am J Bot* 87:1452–1458
- Young ND, dePamphilis CW (2000) Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol Biol Evol* 17:1933–1941
- Young ND, dePamphilis CW (2005) Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evol Biol* 5:16
- Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? *J Plant Res* 123:201–206
- 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
- Zhong BJ, Yonezawa T, Zhong Y, Hasegawa M (2009) Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. *PLoS One* 4:e5297
- Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2010) The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. *Mol Biol Evol* 27:2855–2863