Chapter 4

Plastomes of Bryophytes, Lycophytes and Ferns

Paul G. Wolf*
Department of Biology, Utah State University, Logan, UT 84322, USA

Kenneth G. Karol
The Lewis B. and Dorothy Cullman Program for Molecular Systematics Studies, The New York Botanical Garden, Bronx, NY 10458, USA

Summary

We review current progress in our understanding of chloroplast genomes (plastomes) of liverworts, mosses, hornworts, lycophytes and monilophytes. We briefly cover some of the methods used to obtain complete nucleotide sequences of plastomes and we summarize the published sequences from the plant groups above. We explore some of the evolutionary changes that have occurred in terms of gene content, introns and position of the inverted repeat boundaries. We also discuss RNA editing, which is especially high in plastome genes of some non-seed land plants. We finish with a phylogenetic analysis of available plastome genes and we suggest some possible directions for future research.

I. Introduction

Land plants have a chloroplast (plastid) genome (plastome) with a basic canonical organization that is similar to that of their algal ancestors (see Chap. 3). This represents one of the most evolutionary conserved genomic structures in nature. However, from this basic organization, several structural changes have occurred on

*Author for correspondence, e-mail: wolf@biology.usu.edu

R. Bock and V. Knoop (eds.), Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration 35, pp. 89–102,
DOI 10.1007/978-94-007-2920-9_4, © Springer Science+Business Media B.V. 2012

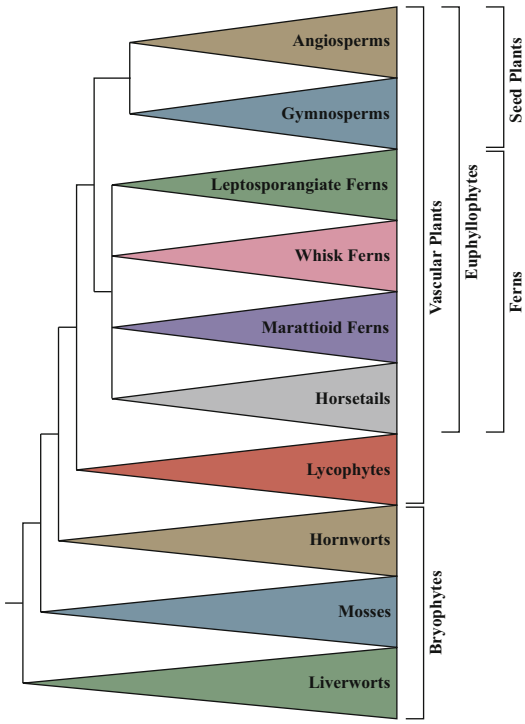

Fig. 4.1. Our current understanding of relationships among major land plant lineages. The extant bryophytes represent a grade of three lineages with liverworts shown sister to all other extant land plants and hornworts shown sister to extant vascular plants. Vascular plants include the lycophytes, monilophytes and seed plants. Four major monilophyte lineages are shown as an unresolved polytomy sister to seed plants.

various evolutionary branches. Here we review aspects of plastomes of extant land plants, except for seed plants (see next chapter). The main lineages include the nonvascular bryophyte lineages (hornworts, liverworts and mosses), the lycophytes and monilophytes. The latter, which include leptosporangiate ferns and horsetails, are also referred to elsewhere as 'ferns' (e.g., Pryer et al. 2004; Schneider et al. 2009). Seed plants appear to be the sister to monilophytes (Pryer et al. 2001). Our current understanding of relationships among these lineages is depicted in Fig. 4.1. We begin with an overview of the taxa and structural aspects of plastomes. We then summarize the major events of gene and intron loss in plastomes of non-seed land plants. Next we discuss the phenomenon of RNA editing, a process that occurs at much higher rates in non-seed land plants than in seed plants.

Abbreviations: IR – Inverted repeat; kb – Kilobases; LSC – Large single copy; mya – Million years ago; PCR – Polymerase chain reaction; PPR – Pentatricopeptide repeat; SSC – Small single copy

II. Techniques and Overall Plastome Organization

Until about the mid-1990s, restriction site mapping was the main approach to inferring plastome organization. The technique involves digesting DNA with restriction endonucleases, separating the DNA fragments on an agarose gel and then transferring them to a membrane. The fragments on this membrane are then probed with labeled plastid DNA from a well-characterized species, or fragments of the same species cut with a different restriction enzyme. After careful analysis, a coarse-scale map of the plastome can be constructed. The first such physical map was that of the *Zea mays* plastome (Bedbrook and Kolodner 1979). Mapping studies also indicated that, within plant cells, the plastome exists in two orientations (Palmer 1983), a pattern that is maintained by a form of homologous recombination (so-called flip-flop recombination; Stein et al. 1986). Subsequently, plastomes of many species were mapped (reviewed by Palmer 1985), verifying that in most (but not all) lineages, plastomes map to a circle with a large single copy region (LSC) and a small single copy region (SSC) separated by two copies of an inverted repeat (IR), which include the ribosomal RNA genes (Palmer 1985). Fine-scale mapping requires nucleotide sequencing, which is easier and cheaper with today's techniques. The first two plastomes to be completely sequenced were those of the flowering plant tobacco (*Nicotiana tabacum*; Shinozaki et al. 1986), and the liverwort *Marchantia polymorpha* (Ohyama et al. 1986). These data confirmed the earlier inferences on overall

plastome organization that had been deduced from mapping studies.

Most green plant plastomes map to a circle of about 150 kb. However, the largest reported plastome, that of the green alga *Floydiella terrestris*, is more than 500 kb (Brouard et al. 2010). Most plant cells contain many copies of the plastome; even plants with a single plastid (e.g., the unicellular green alga *Chlamydomonas reinhardtii*) can contain many copies of the plastome. At the other extreme, wheat cells have more than 50 plastids per cell and more than 300 plastome copies per plastid (Boffey and Leech 1982). Thus, although the plastome is a small genome compared to its nuclear counterpart, plastid DNA makes up a significant proportion of total cellular DNA, as much as 20% in some species (Boffey and Leech 1982).

Plastid DNA is not assembled into chromosomes and it does not reside in the plastid as a population of free circular molecules. Rather, several plastomes are organized, with proteins and RNA, into structures known as nucleoids (Sato et al. 2003). Most nucleoids are attached to the envelope membrane, but mature chloroplasts can also have nucleoids associated with the thylakoid membrane (Sato et al. 2003). It is likely that nucleoid structure plays an important role in plastome replication, transcription and post-transcriptional modification. However, the general relationships between plastome packaging and these processes remain poorly understood (Bock 2007).

Although plastomes are typically depicted as circles, most plastid DNA is not in this form in a living plant cell (Bendich 2004; Bock 2007). Researchers have found linear plastomes, concatenated pieces representing multiple plastomes (sometimes circular) (Bendich 2004), and even branched forms (Oldenburg and Bendich 2004a). This variety of possible conformations is likely a function of both phylogenetic divergence and stage of plastome replication. The plastome replication process itself is also poorly understood (Bock 2007), and several mechanisms have been proposed. Early models involved bidirectional replication similar to that in bacteria, resulting in displacement (D) loops (Kolodner and Tewari 1975b). Rolling circle amplification (RCA) could also be used to achieve additional replication (Kolodner and Tewari 1975a). A double D-loop mechanism has also been proposed (Kunnimalaiyaan and Nielsen 1997). However, these models have been challenged, based on the degree of linear DNA observed (Bendich 2004), and a recombination-dependent mechanism was instead proposed (Oldenburg and Bendich 2004b). The challenge of studying replication is making observations during the actual process. Alternatively, researchers can examine the signature of replication, which can be deduced from variation in base composition. Studies of mitochondrial genomes found that regions accumulate adenine-to-guanine transitions due to deamination during the single-stranded phase of replication. This is because A → G transitions accumulate evenly over time whereas the accumulation of C → T substitutions is complex and asymptotic (Krishnan et al. 2004). Thus, gradients in A/G composition, especially for non-coding DNA, is a function of total amount of time spent in the single-stranded phase, and therefore can reveal origins and directions of replication. This approach was used recently to show that A/G composition gradients are most consistent with the earlier models (bidirectional and RCA) across a wide range of published green plant plastomes (Krishnan and Rao 2009). Direct testing of these models is now needed. Meanwhile, evidence continues to accumulate for a role of recombination-dependent replication in *Arabidopsis*, especially as a repair process for maintaining plastome integrity (Rowan et al. 2010). Clearly, the evidence suggests that more than one replication process appears to be operating, and the result is a complex population of molecules representing the plastome. Regardless, most land plant plastomes map to a circle and have a fairly conserved set of protein and RNA encoding genes. The map of the plastome of the whisk fern *Psilotum nudum* is depicted in linear fashion in Fig. 4.2 as a guide to this overall structure.

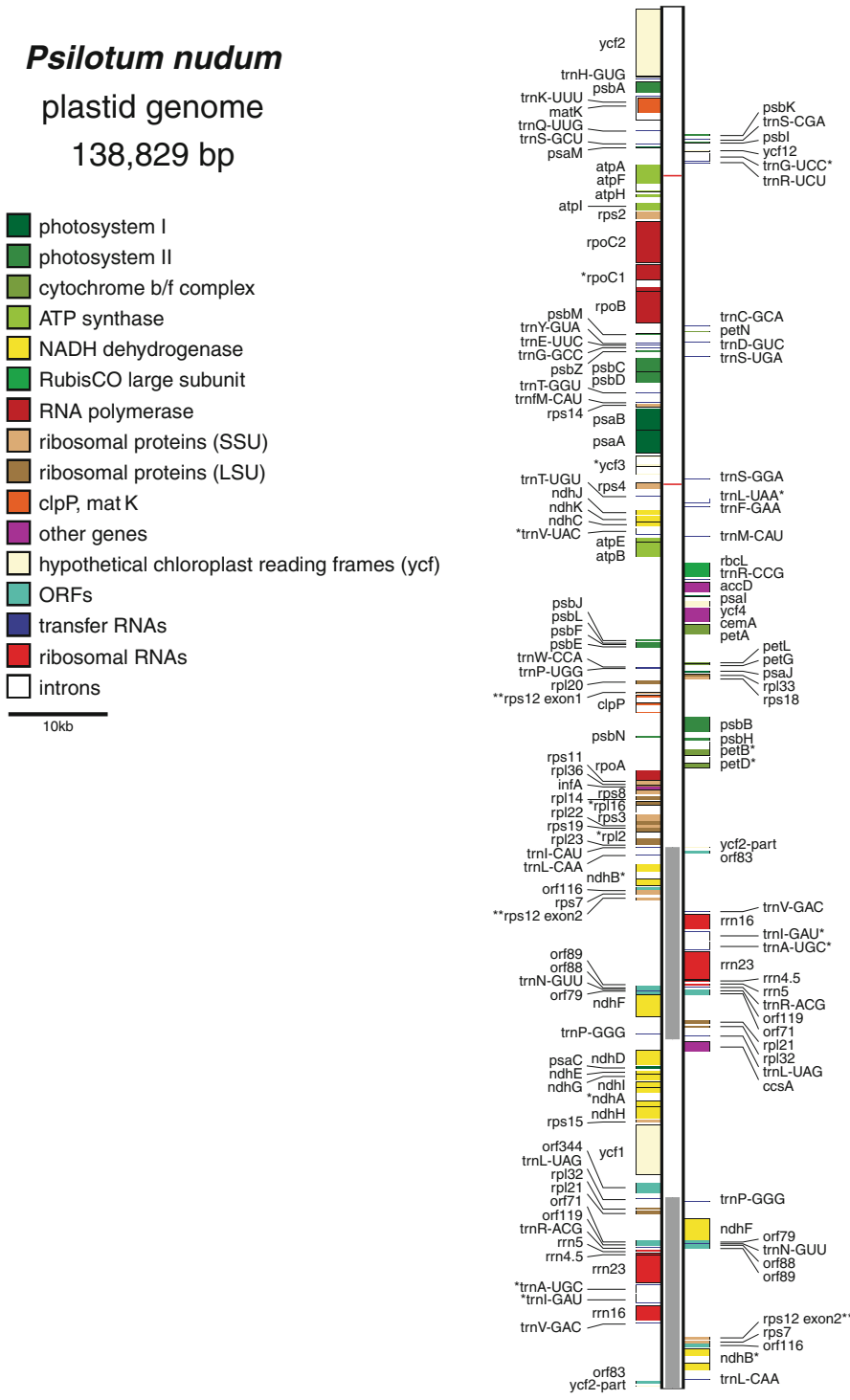

Fig. 4.2. Psilotum nudum plastid genome structure. Genes (*colored boxes*) on the *right side* of the map are transcribed in the *top down* direction, whereas those on the *left side* are transcribed *bottom up*. The tRNA genes are indicated by the three-letter amino acid code followed by the anticodon. Intron-containing genes are show with an *asterisk* (*); the trans-spliced gene *rps12* is shown with *two asterisks* (**). The two horizontal *red lines* along the genome indicate the insertion/deletion events unique to all monilophytes and the two *grey* boxes along the genome indicate the inverted repeats. Note a fragment of *ycf2* is found in the inverted repeat.

A. Bryophytes

The bryophytes represent a grade of three extant lineages (Fig. 4.1, Mishler and Churchill 1984; Nickrent et al. 2000; Renzaglia et al. 2007; Shaw and Renzaglia 2004). Several phylogenetic analyses lead to the hypothesis that liverworts are sister to all other extant land plants (Qiu et al. 1998) and hornworts are sister to extant vascular plants (e.g., Groth-Malonek et al. 2005; Qiu et al. 2006, 2007; Qiu 2008). Recent findings of cryptospores from the early Middle Ordovician (c. 473–471 mya Rubinstein et al. 2010) may represent liverworts or at least their ancestors. A broad-scale phylogenetic analysis of liverworts reveals several key lineages. The earliest branching lineage, Haplomitriopsida, is the sister to all remaining extant liverworts. There is then a major split between the complex thalloid liverworts (Marchantiopsida) and a heterogenous clade (Jungermanniopsida) which includes two clades (Metzgeriidae and Pellidae) of simple thalloid taxa (which is therefore a paraphyletic group) and a monophyletic "leafy" clade (Jungermanniidae) which excludes a few taxa previously considered as leafy (Forrest et al. 2006). The complex-thalloid liverwort *Marchantia polymorpha* was the first plant for which the chloroplast genome was sequenced (Ohyama et al. 1986). Later, the complete mitochondrial genome of *M. polymorpha* was also sequenced (Oda et al. 1992), providing yet another important genomic resource for nonvascular plants. A second liverwort plastome was recently sequenced (Wickett et al. 2008b), that of the only known parasitic bryophyte, the simple-thalloid liverwort *Aneura mirabilis*. Nonphotosynthetic plants often lose plastid genes that are associated with photosynthetic functions (Wickett et al. 2008b; Wolfe et al. 1992). Indeed, *A. mirabilis* has lost some of the same genes as has the parasitic angiosperm *Epifagus virginiana* (Wolfe et al. 1992). However, the loss of only a subset of these genes in *A. mirabilis* suggests that this liverwort is in an earlier stage of acquiring a parasitic life history stage (Wickett et al. 2008a, b).

Mosses are a diverse clade of more than 12,000 species, representing about eight main extant lineages (Cox et al. 2004; Goffinet and Buck 2004; Newton et al. 2000; Wahrmund et al. 2009, 2010). Two complete moss plastomes have been sequenced: that of the model species for molecular genetic studies, *Physcomitrella patens* (Sugiura et al. 2003) and the desiccation-tolerant species *Syntrichia (= Tortula) ruralis* (Oliver et al. 2010). These plastomes differed by a large (71 kb) inversion in the large single copy (LSC) region, with *S. ruralis* possessing the apparently ancestral organization. Further analysis revealed that the inversion is unique to the Funariidae (Goffinet et al. 2007). This inversion is the largest plastome reorganization reported to date for land plant plastomes, and appears to represent a single evolutionary event (Goffinet et al. 2007).

Hornworts represent the third main clade of nonvascular land plants, with about 400 extant species (Bateman et al. 1998). Hornworts are probably sister group to the vascular plants (Groth-Malonek et al. 2005; Qiu et al. 2007). Phylogeny within the hornworts has been examined by Duff and coworkers (Duff et al. 2004; 2007). Currently, there is only a single published complete plastome sequence of a hornwort, *Anthoceros formosae* (Kugita et al. 2003b). This plastome has a very high level of RNA editing (Kugita et al. 2003a), as do several mitochondrial and plastid genes in most hornworts studied (Duff and Moore 2005; Duff 2006). More details on RNA editing are provided later in this chapter.

B. Lycophytes

The lycophytes include a large assemblage of both extant and extinct lineages. Extant groups include the heterosporous Isoetopsida with about 150 species of *Isoetes* (quillworts) and about 700 species of *Selaginella* (spikemosses). The remaining extant lineage is the homosporous Lycopodiopsida (clubmosses) of which about 300 species are known, including

Lycopodium, *Huperzia* and related genera. Extinct lineages include many fossil species, especially from the late Silurian (about 420 mya) through the Carboniferous (about 300 mya, Kenrick and Crane 1997). Ancient representative of this group of plants formed many of the fossil coal beds. Photosynthesis in these plants harnessed the sun's energy, which is now used as one major source of fossil fuels. These extinct lycophytes were large plants; some reached 30 m, whereas today's species are less than 1 m. As a group the lycophytes appear to be a sister group to Euphyllophytes (monilophytes plus seed plants, see below). This early split is supported both by analysis of morphology in fossil taxa (Kenrick and Crane 1997) and extant taxa (Kranz and Huss 1996). However, an additional convincing piece of evidence comes from analysis of plastome organization. Monilophytes and seed plants possess a 30 kb inversion in the LSC relative to lycophytes and bryophytes (Raubeson and Jansen 1992). Further details of the organization of lycophyte plastomes came from restriction site mapping of an *Isoetes* plastome (Duff and Schilling 2000), which confirmed the overall similarity of the lycophyte and bryophyte plastomes. The first complete plastome sequence of a lycophyte was that of *Huperzia lucidula* (Wolf et al. 2005). Since then, additional plastomes have been sequenced from the heterosporous genera, *Selaginella moellendorffii, S. uncinata* and *Isoetes flaccida* (Karol et al. 2010; Tsuji et al. 2007). Although lycophytes share structural similarities with bryophytes, the former do have some unique features. For example, *ycf2* normally resides in the LSC in most plastomes, but has been translocated to the SSC in *I. flaccida*, with the 5′ end now incorporated into the IR. In addition, the *chlL/chlN* gene cluster has been inverted in *I. flaccida* so that it is now adjacent to *ycf2* rather than *ycf1* as in *H. lucidula*. The *ycf2* translocation and the *chlL/chlN* inversion occur in neither of the *Selaginella* plastomes. Both *Selaginella* plastomes differ considerably in gene order from other plastomes (Karol et al. 2010). An approximately 14-kb region has been translocated from the LSC to the IR/SSC in both *Selaginella* plastomes. The genes included in this translocation differ slightly between *S. uncinata* and *S. moellendorffii*. In addition, *rps4* is in the IR in *Selaginella* and marks one endpoint of the translocated segment. The other endpoint resides in the SSC and is marked by *psbD* in *S. moellendorffii*. In *S. uncinata*, the same endpoint includes three additional genes *(trnE-UUC, trnY-GUA and trnD-GUC*), which remain in the LSC adjacent to *ycf2* in *S. moellendorffii. Selaginella uncinata* also has a ~20-kb LSC inversion (*psbI to rpoB-trnC-GAC*), a duplication of the *psbK/trnQ-UUG* region, and translocation of *petN* from the LSC to the SSC. These features appear to be unique to *S. uncinata* (Karol et al. 2010). Because complete plastome sequences are available from only four species of lycophytes, it is not yet possible to infer the phylogenetic extent of all plastome changes. Additional taxon sampling will be needed to understand more fully how recent and extensive these changes are.

C. Monilophytes (Ferns)

Monilophytes represent another group of vascular plants with an extensive fossil history. Here we consider four main extant lineages: (1) leptosporangiate ferns (about 11,000 species), (2) a clade that includes whisk ferns (*Psilotum* and *Tmesipteris*) and the Ophioglossales, (3) Marattioid ferns and (4) Horsetails (*Equisetum*). Data from plastid and nuclear gene sequences (Pryer et al. 2001) and morphology (Kenrick and Crane 1997; Schneider et al. 2009) find support for monophyly of a clade that includes these four lineages. Together the clade is called moniliformopses (Kenrick and Crane 1997), monilophytes (Pryer et al. 2004), or ferns *sensu lato* (Schneider et al. 2009). Further resolution of relationships among these four groups has not yet been achieved. Although monophyly of monilophytes has support from analyses of extant taxa, analyses that include fossil taxa has questioned this idea (Rothwell and Nixon 2006).

The first monilophyte to have a plastome sequenced was *Psilotum nudum* (GenBank accession #AP004638 from 2002, see Fig. 4.2). Several phylogenetic studies support inclusion of the ophioglossoid ferns with the whisk ferns (Pryer et al. 2001, 2004; Qiu et al. 2007), but so far no complete plastome from the ophioglossoid ferns has been published. Complete plastome sequences are available from one horsetail (Karol et al. 2010), one marattioid fern (Roper et al. 2007) and four leptosporangiate ferns (Der 2010; Gao et al. 2009; Wolf et al. 2003, 2011).

An inversion in the LSC involving *trnG-GCC* to *trnT-GGU* is found in all fern plastomes and no other land plant plastomes (Karol et al. 2010), thus providing further evidence for monilophyte monophyly. Within the leptosporangiate ferns, a series of additional inversions has occurred, two of which (18 kb and 21 kb respectively) result in a reverse gene order within the IR (Wolf et al. 2010). An additional pair of inversions occurred more recently in the LSC of a large clade of ferns (the "polypods", Wolf et al. 2010).

III. The Inverted Repeat Boundaries

Plastome IRs from most plants typically house a similar gene content, which includes primarily rRNA and tRNA genes (Jansen et al. 2007; Palmer and Stein 1986; Turmel et al. 2007). This is seen also in some leptosporangiate ferns where, except for a few early-diverging clades, the IR itself is inverted (Wolf et al. 2003). Most of the variation in IR gene content occurs at the ends of the IR. This "ebb and flow" of the IR boundaries into and out of the LSC and SSC regions has been attributed to effects of recombination and gene conversion (Goulding et al. 1996). Effects of these positional changes have been seen in related species at the nucleotide level in several species of *Nicotiana* (Goulding et al. 1996). Furthermore, when comparing distantly related lineages of land plants, several plastomes exhibit unique IR boundaries that differ from the basic theme (Karol et al. 2010). But this is not always the case: other distantly related taxa have very similar IR boundaries. For example, *Marchantia polymorpha*, two mosses and *Equisetum arvense* were identical in gene content at both ends of the IR. This suggests that whereas the ends of the IR clearly ebb and flow in some lineages, in other lineages they appear to be rather stable, at least at the scale of gene order (Karol et al. 2010).

IV. Changes in Gene and Intron Content

Most plastomes sequenced to date contain a very similar repertoire of genes. The most significant exceptions are plastomes from parasitic plants in which many photosynthetic genes are lost or pseudogenized (Wickett et al. 2008b; Wolfe et al. 1992). Overlaid on the basic pattern are found a few genes that are absent in some sequenced plastomes. Some of these genes seem to have been lost multiple times based on their phylogenetic distribution (See Fig. 4.3). These include *infA* and *ycf1*. Other genes appear to be distinctly present or absent in particular clades. Here we briefly list these latter patterns based on what we know is a very limited sample of plastomes (especially for non-seed land plant clades). We ignore many that are specific to only one plastome, except where that plastome is the sole representative (such as the single published hornwort plastome).

The genes *ccsA* and *rpoA* are absent from the plastomes of two mosses (*Syntrichia ruralis* and *Physcomitrella patens*, Oliver et al. 2010; Sugiura et al. 2003), *petN* is lacking in *S. ruralis*, and *cemA* is absent from both *Selaginella* plastomes. Mosses and liverworts lack *rps16*, but the gene is present in hornworts and some vascular plants. The genes *matK* and *rps15* are pseudogenes in the hornwort (Kugita et al. 2003b). The gene cluster *chlB, chlL*, and *chlN* is absent from *Psilotum nudum* and angiosperms. The gene *psaM* is lacking from the three polypod ferns (*Adiantum capillus-veneris, Cheilanthes lindheimeri* and *Pteridium aquilinum*), as

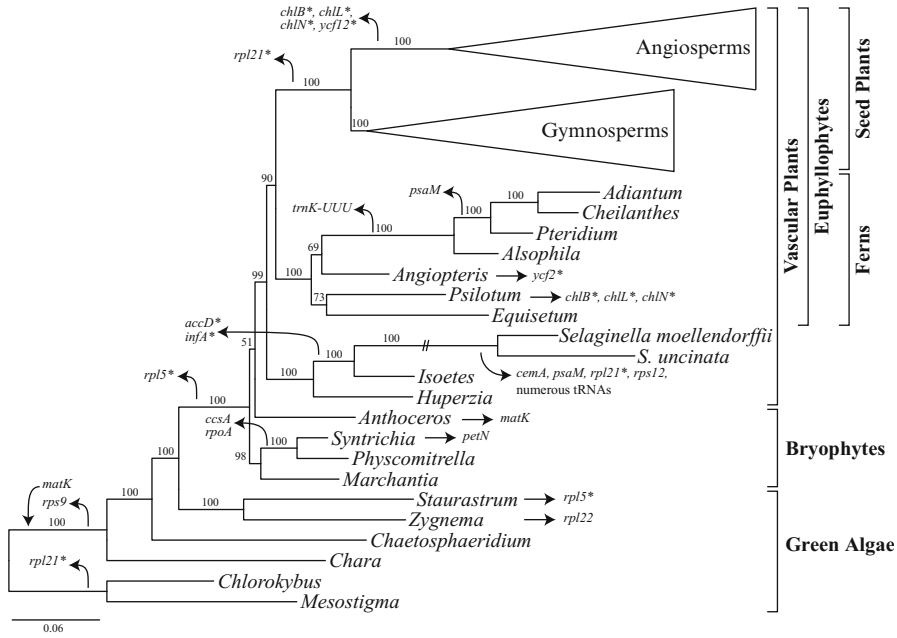

Fig. 4.3. Phylogenetic results using nucleotide data. Phylogenetic analyses were performed using 49 plastome gene sequences from 45 completely sequenced plastomes, including 39 land plants and six charophycean algae. The nucleotide alignment from Karol et al. (2010) was used as a starting point (49 genes from 43 taxa). To this we incorporated into the alignment sequence data from two new leptosporangiate fern plastomes: *Cheilanthes lindheimeri* (Wolf et al. 2011) and *Pteridium aquilinum* (Der 2010). Maximum likelihood analyses were performed on the Cyberinfrastructure for Phylogenetic Research (CIPRES) Portal (v.3.1, Miller et al. 2009) using RAxML-HPC (v. 7.2.7, Stamatakis 2006, 2008) with 200 bootstrap replicates. Third codon positions were excluded to avoid problems associated with relatively rapidly evolving sites. The best tree (−ln = 195205.737395) is shown with bootstrap proportions drawn above branches. The relationships among major fern lineages are weakly supported, though monophyly of the ferns is strongly supported. The branch leading to *Selaginella* was drawn to one-half scale to accommodate this figure. Note that the sister relationship of the liverwort and mosses is strongly supported and is in contrast to the cladogram shown in Fig. 4.1. This relationship was also recovered by Karol et al. (2010) when divergent taxa (*Selaginella* spp.) were excluded from phylogenetic analyses. Furthermore, Renzaglia and Garbary (2001) concluded that characters related to sperm cell development were compelling evidence for the monophyly of liverworts plus mosses, a clade they referred to as Setaphytes. Names of lost plastid genes are shown with *arrows*. An *asterisk* (*) following a gene name indicates that this gene has been lost in at least two lineages independently. The rare gain of a plastid gene (*matK*) is also indicated in the *green algae*.

well as from the two *Selaginella* plastomes and the majority of seed plant plastomes. Seed plant plastomes lack *rpl21,* as do the two *Selaginella* plastomes. The parasitic liverwort *A. mirabilis* has lost several genes (including several *ndh* genes) and many others exist as possibly recent pseudogenes (Wickett et al. 2008a).

A group II intron, along with its encoded maturase gene (*matK*) invaded the *trnK-UUU* gene in charophycean algae after the divergence of chlorophytes and charophytes. All chloropycean algae and some early diverging charophycean algae (Mesostigmatophyceae and Chlorokybophyceae) do not contain this intron. More derived charophycean algae (Charophyceae, Coleocheatophyceae and Zygnematophyceae) have the intron. There is one lineage (Klebsormidiophyceae, which is sister to Charophyceae, Coleocheatophyceae, Zygnematophyceae and land plants) where we do not yet fully know the condition of *trnK*. A large clade of leptosporangiate ferns has subsequently lost *trnK-UUU* and its

intron (Wolf et al. 2010, 2011), yet *matK* remains. The introns of *clpP* are variable across land plants, with some plastomes having two, and others having one intron in this gene, but there appears to be no distinct phylogenetic pattern (Karol et al. 2010).

Thus, although plastome gene content tends to be well-conserved among land plant lineages, several clade-specific gene losses are apparent.

V. RNA Editing

The central dogma of molecular genetics requires conservation of information from genomic DNA through messenger RNA to the final amino acid sequence of a protein. However, detailed studies of the various products of transcription and translation have found exceptions to this conservation. Considerable post-translational modification occurs to proteins. In addition to the various aspects of RNA processing that occur, an independent post-transcriptional stage is RNA editing. This process alters the nucleotides in the primary transcript so that the messenger RNA differs from the genomic encoding sequence (See Chap. 13). RNA editing is found throughout eukaryotes, and is especially common in organellar genomes (reviewed by Tillich et al. 2006). In plastome genes from seed plants, the process occurs at fewer than 40 sites and about ten times that number have been reported in ferns and hornworts. In most cases, cytosines are edited to uracils, but in hornworts and ferns, additional uracil-to-cytosine edits have been reported (Kugita et al. 2003a; Wolf et al. 2004).

RNA editing requires both cis- and trans-acting factors. Cis-acting factors include the actual site to be edited. Other cis-acting factors include upstream and downstream recognition sequences (Kobayashi et al. 2008). However, the latter appear to have no obvious pattern across sites. This might be because the trans-acting factors (nuclear-encoded proteins) are likely to be of several types (Hammani et al. 2009). To date, over 20 different nuclear factors have been associated with RNA editing in *Arabidopsis* (see Stern et al. 2010), most of which are pentatricopeptide repeat (PPR) proteins (Kotera et al. 2005; Okuda and Shikanai 2008). These proteins are characterized by tandem repeats of a degenerate 35 amino acid motif, and several PPR gene subfamilies are found across eukaryotic lineages.

The functions of RNA editing are not obvious. Several authors have argued that RNA editing repairs errors in genomic sequences (Jobson and Qiu 2008; Stern et al. 2010). However, this seems far less efficient than a simple nucleotide substitution at the DNA level of the genome, which would require no further action. An additional role has been implicated in gene regulation, whereby RNA editing varies with developmental stage and could be used to restore correct translation when the gene product is needed (Hirose et al. 1999). This has been observed in a few cases in animals, but seems to play a minor role in plants (Stern et al. 2010). It seems more likely that the enzymes that edit RNA have evolved for other cellular functions and their editing ability then releases selective constraints for the edited sites in genes. In fact, some of these other functions of editing enzymes are known. In primates, the APOBEC family of RNA editing enzymes includes cytosine deaminases that act to restrict infection from retroviruses (Bransteitter et al. 2009). Further research is needed on the RNA editing factors of *Arabidopsis* and other plants if we are to understand further the function and cellular significance of RNA editing.

RNA editing can cause problems for comparative analyses of nucleotide sequences. Most phylogenetic analyses are based on alignment of orthologous genomic sequences. However, if RNA editing occurs, these DNA sequences represent the unedited versions. Should one use the genomic sequences or the edited versions? The latter can only be inferred accurately by using mature RNA transcripts to generate cDNA. Until this is done, one does not know which sites have been edited. For analyses of seed plants, this

dilemma is trivial because RNA editing rates are so low. But in ferns, lycophytes and some bryophytes, the effect on the outcomes of analyses can be significant. In hornworts, RNA editing rates are so high that the same site can be C to U edited in some taxa and U to C edited in other taxa (Duff and Moore 2005). When phylogenetic analyses of hornworts use cDNA sequences, the results are different from those from genomic sequences (Duff and Moore 2005; Duff 2006). Removal of edited sites does not help, because that reduces the amount of potentially useful phylogenetic signal. The solution can only be attained once we know the evolutionary stability of RNA editing itself. If relatively stable, then the fact that a site is edited provides an evolutionary marker. If sites come and go rapidly, then RNA editing sites are homoplastic and the results of phylogenetic analysis of cDNA sequences will be misleading. The answer will depend on the relative levels of homoplasy in genomic sequences versus RNA editing sites, and this is likely to vary across clades of land plants.

VI. Phylogenetic Analyses

Over the last few decades single gene phylogenetic analyses have served as powerful tools for reconstructing the evolutionary history of every major lineage of life on Earth (Donoghue and Cracraft 2004). Reduced costs and improvements in sequencing technologies have allowed several genes to be sequenced across a broad range of taxa for phylogenetic reconstruction (Holton and Pisani 2010; Nickrent et al. 2000; Qiu et al. 2007; Shalchian-Tabrizi et al. 2008). Indeed, with new second-generation sequencing technologies, complete plastome sequences are now being generated at an ever increasing rate (Cronn et al. 2008; Wolf et al. 2011). We reanalyzed the plastome alignment of Karol et al. (2010) and included two new leptosporangiate fern taxa (*Cheilanthes lindheimeri* and *Pteridium aquilinum*). This analysis included 49 plastome genes from 45 green plant taxa and the results are shown in Fig. 4.3. The overall topology is consistent with results presented in Karol et al. (2010), with the two new fern taxa found in a monophyletic leptosporangiate clade. Relationships among the major monilophyte lineages remained weakly supported. Most of the currently available land plant plastome sequences are from seed plants, with very few available from the presumed sister clade, monilophytes. With additional data from other fern representatives, including ophioglossoid ferns, it will become possible to gain further insight into early land plant evolution as well as the patterns and processes that shape the evolution of plastomes.

VII. Future Directions

Currently, the distribution of complete plastome sequence data is biased toward angiosperms. In general, clades more distantly related to angiosperms are less well sampled. There are especially critical clades in the algae for which no representative plastome sequence is available (e.g., Klebsormidiophyceae, *Coleochaete*). Although obtaining the actual DNA sequence is relatively easy, limiting steps in plastome sequencing mostly involve isolating plastome DNA. Although this can be done through various centrifugation and other procedures (Jansen et al. 2005), there are some alternative approaches. If the plastome component of total DNA is high then a total genomic shotgun sequence can provide sufficient data from which the plastome sequence can be assembled (Wolf et al. 2011). A more cost-effective approach involves multiplex sequencing-by-synthesis on the Illumina platform (Cronn et al. 2008). In this protocol, more than a hundred plastomes can be sequenced simultaneously. However, custom probes or PCR-primers will be needed for each major clade, the range of these depending on sequence divergence levels. One problem with the shotgun genome approach is that it may not be possible to distinguish genuine reads of plastome DNA from those that are plastid DNA that has been transferred to

the nucleus (Bock and Timmis 2008). To some extent, this is a problem for all approaches to plastome studies, but the problem is exacerbated by short reads and the use of total genomic DNA extractions. Regardless, the prospects seem good for filling many of the critical clade gaps in the next few years. This should ease the trend away from recent exemplar studies (with a few, though critical taxa) toward more taxon-dense studies with broad phylogenetic breadth. Although such a trend may not always uncover much new in terms of phylogenetic hypotheses, it is sure to show us more details of the evolution of plastomes themselves.

Acknowledgements

This work was supported by NSF DEB1020660 and DEB1036466 to KGK and NSF DEB-0228432 to PGW.

References

Bateman RM, Crane PR, DiMichele WA, Kenrick PR, Rowe NP, Speck T, Stein WE (1998) Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu Rev Ecol Syst 29:263–292

Bedbrook JR, Kolodner R (1979) Structure of chloroplast DNA. Annu Rev Plant Physiol Plant Mol Biol 30:593–620

Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. Plant Cell 16:1661–1666

Bock R (2007) Structure, function, and inheritance of plastid genomes. Cell Mol Biol Plastids 19:29–63

Bock R, Timmis JN (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. Bioessays 30:556–566

Boffey SA, Leech RM (1982) Chloroplast DNA levels and the control of chloroplast division in light-grown wheat leaves. Plant Physiol 69:1387–1391

Bransteitter R, Prochnow C, Chen XJS (2009) The current structural and functional understanding of APOBEC deaminases. Cell Mol Life Sci 66:3137–3147

Brouard JS, Otis C, Lemieux C, Turmel M (2010) The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. Genome Biol Evol 2:240–256

Cox CJ, Goffinet B, Shaw AJ, Boles SB (2004) Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. Syst Bot 29:234–250

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res 36:e122

Der JP (2010) Genomic perspectives on evolution in bracken fern. Utah State University, Logan

Donoghue MJ, Cracraft J (2004) Charting the tree of life. In: Donoghue MJ, Cracraft J (eds) Assembling the tree of life. Oxford University Press, New York, pp 1–4

Duff RJ (2006) Divergent RNA editing frequencies in hornwort mitochondrial *nad5* sequences. Gene 366:285–291

Duff RJ, Moore FBG (2005) Pervasive RNA editing among hornwort *rbcL* transcripts except *Leiosporceros*. J Mol Evol 61:571–578

Duff RJ, Schilling EE (2000) The chloroplast genome structure of the vascular plant *Isoetes* is similar to that of the liverwort *Marchantia*. Am Fern J 90:51–59

Duff RJ, Cargill DC, Villarreal JC, Renzaglia KS (2004) Phylogenetic relationships of the hornworts based on *rbcL* sequence data: novel relationships and new insights. Monogr Syst Bot Ann Missouri Bot Gard 98:41–58

Duff RJ, Villarreal JC, Caagill DC, Renzaglia KS (2007) Progress and challenges toward developing a phylogeny and classification of the hornworts. Bryologist 110:214–243

Forrest LL, Davis EC, Long DG, Crandall-Stotler BJ, Clark A, Hollingsworth ML (2006) Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. Bryologist 109:303–334

Gao L, Yi X, Yang YX, Su YJ, Wang T (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. BMC Evol Biol 9:130

Goffinet B, Buck WR (2004) Systematics of the Bryophyta (Mosses): from molecules to a revised classification. Monog Syst Bot Missouri Bot Gard 98:203–223

Goffinet B, Wickett NJ, Werner O, Ros RM, Shaw AJ, Cox CJ (2007) Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). Ann Bot 99:747–753

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet 252:195–206

Groth-Malonek M, Pruchner D, Grewe F, Knoop V (2005) Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol 22:117–125

Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, Shikanai T, Small I (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. Plant Cell 21:3686–3699

Hirose A, Kusumegi T, Tsudzuki T, Sugiura M (1999) RNA editing sites in tobacco chloroplast transcripts: editing as a possible regulator of chloroplast RNA polymerase activity. Mol Gen Genet 262:462–467

Holton TA, Pisani D (2010) Deep genomic-scale analyses of the Metazoa reject Coelomata: evidence from single- and multigene families analyzed under a supertree and supermatrix paradigm. Genome Biol Evol 2:310–324

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol 395:348–384

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee S-B, 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

Jobson RW, Qiu YL (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? Biol Direct 3:43

Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KDE, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, Olmstead RG, Renzaglia KS, Wolf PG (2010) Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol Biol 10:321

Kenrick P, Crane PR (1997) The origin and early diversification of land plants: a cladistic study. Smithsonian Press, Washington, DC

Kobayashi Y, Matsuo M, Sakamoto K, Wakasugi T, Yamada K, Obokata J (2008) Two RNA editing sites with cis-acting elements of moderate sequence identity are recognized by an identical site-recognition protein in tobacco chloroplasts. Nucleic Acids Res 36:311–318

Kolodner R, Tewari KK (1975a) Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. Nature 256:708–711

Kolodner R, Tewari KK (1975b) Presence of displacement loops in covalently closed circular chloroplast deoxyribonucleic-acid from higher-plants. J Biol Chem 250:8840–8847

Kotera E, Tasaka M, Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433:326–330

Kranz HD, Huss VAR (1996) Molecular evolution of pteridophytes and their relationship to seed plants: evidence from complete 18S rRNA gene sequences. Plant Syst Evol 202:1–11

Krishnan NM, Rao BJ (2009) A comparative approach to elucidate chloroplast genome replication. BMC Genomics 10:237

Krishnan NM, Seligmann H, Raina SZ, Pollock DD (2004) Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. DNA Cell Biol 23:707–714

Kugita M, Yamamoto Y, Fujikawa T, Matsumoto T, Yoshinaga K (2003a) RNA editing in hornwort chloroplasts makes more than half the genes functional. Nucleic Acids Res 31:2417–2423

Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K (2003b) The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res 31:716–721

Kunnimalaiyaan M, Nielsen BL (1997) Chloroplast DNA replication: mechanism, enzymes and replication origins. J Plant Biochem Biotech 6:1–7

Miller MA, Holder MT, Vos R, Midford PE, Liebowitz T, Chan L, Hoover P, Warnow T (2009) The CIPRES Portals. http://www.phylo.org/sub_sections/portal

Mishler BD, Churchill SP (1984) A cladistic approach to the phylogeny of the bryophytes. Brittonia 36:406–424

Newton AE, Cox CJ, Duckett JG, Wheeler JA, Goffinet B, Hedderson TAJ, Mishler BD (2000) Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. Bryologist 103:187–211

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol Biol Evol 17:1885–1895

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Okuda K, Shikanai T (2008) PPR proteins function as a trans-factor in chloroplast RNA editing. In: Allen JF, Gantt E, Golbeck J (eds) Photosynthesis energy from the sun. Springer, New York, pp 1211–1214

Oldenburg DJ, Bendich AJ (2004a) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J Mol Biol 335:953–970

Oldenburg DJ, Bendich AJ (2004b) 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

Oliver MJ, Murdock AG, Mishler BD, Kuehl JV, Boore JL, Mandoli DF, Everett KDE, Wolf PG, Duffy AM, Karol KG (2010) Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. BMC Genomics 11:143

Palmer JD (1983) Chloroplast DNA exists in 2 orientations. Nature 301:92–93

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD, Stein DB (1986) Conservation of chloroplast genome structure among vascular plants. Curr Genet 10:823–833

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409:618–622

Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, Cranfill R (2004) Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Amer J Bot 91:1582–1598

Qiu YL (2008) Phylogeny and evolution of charophytic algae and land plants. J Syst Evol 46:287–306

Qiu YL, Cho YR, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674

Qiu YL, Li LB, Wang B, Chen ZD, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA 103:15511–15516

Qiu YL, Li LB, Wang B, Chen ZD, Dombrovska O, Lee J, Kent L, Li RQ, Jobson RW, Hendry TA, Taylor DW, Testa CM, Ambros M (2007) A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. Int J Plant Sci 168:691–708

Raubeson LA, Jansen RK (1992) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255:1697–1699

Renzaglia KS, Garbary DJ (2001) Motile gametes of land plants: diversity, development, and evolution. Crit Rev Plant Sci 20:107–213

Renzaglia KS, Schuette S, Duff RJ, Ligrone R, Shaw AJ, Mishler BD, Duckett JG (2007) Bryophyte phylogeny: advancing the molecular and morphological frontiers. Bryologist 110:179–213

Roper JM, Hansen SK, Wolf PG, Karol KG, Mandoli DF, Everett KDE, Kuehl J, Boore JL (2007) The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). Am Fern J 97:95–106

Rothwell GW, Nixon KC (2006) How does the inclusion of fossil data change our conclusions about the phylogenetic history of euphyllophytes? Int J Plant Sci 167:737–749

Rowan BA, Oldenburg DJ, Bendich AJ (2010) *RecA* maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. J Exp Bot 61:2575–2588

Rubinstein CV, Gerrienne P, de la Puente GS, Astini RA, Steemans P (2010) Early middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol 188:365–369

Sato N, Terasawa K, Miyajima K, Kabeya Y (2003) Organization, developmental dynamics, and evolution of plastid nucleoids. Int Rev Cytol Surv Cell Biol 232:217–262

Schneider H, Smith AR, Pryer KM (2009) Is morphology really at odds with molecules in estimating fern phylogeny? Syst Bot 34:455–475

Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, Jakobsen KS, Cavalier-Smith T (2008) Multigene phylogeny of Choanozoa and the origin of animals. PLoS One 3:e2098

Shaw J, Renzaglia K (2004) Phylogeny and diversification of bryophytes. Amer J Bot 91:1557–1581

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web-servers. Syst Biol 57:758–771

Stein DB, Palmer JD, Thompson WF (1986) Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. Curr Genet 10:835–841

Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA metabolism. Annu Rev Plant Biol 61:125–155

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from chloroplast to the nucleus. Nucleic Acids Res 31:5324–5331

Tillich M, Lehwark P, Morton BR, Maier UG (2006) The evolution of chloroplast RNA editing. Mol Biol Evol 23:1912–1921

Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K (2007) The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. J Plant Res 120:281–290

Turmel M, Pombert J-F, Charlebois P, Otis C, Lemieux C (2007) The green algal ancestry of land plants as revealed by the chloroplast genome. Int J Plt Sci 168:679–689

Wahrmund U, Rein T, Müller KF, Groth-Malonek M, Knoop V (2009) Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. Plant Syst Evol 282:241–255

Wahrmund U, Quandt D, Knoop V (2010) The phylogeny of mosses – addressing open issues with a new mitochondrial locus: group I intron cobi420. Mol Phylogen Evol 54:417–426

Wickett NJ, Fan Y, Lewis PO, Goffinet B (2008a) Distribution and evolution of pseudogenes, gene losses, and a gene rearrangement in the plastid genome of the nonphotosynthetic liverwort, *Aneura mirabilis* (metzgeriales, jungermanniopsida). J Mol Evol 67:111–122

Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, dePamphilis CW, Boore JL, Goffinet B (2008b) Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol Biol Evol 25:393–401

Wolf PG, Rowe CA, Sinclair RB, Hasebe M (2003) Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res 10:59–65

Wolf PG, Rowe CA, Hasebe M (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. Gene 339:89–97

Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL (2005) The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350:117–128

Wolf PG, Roper JM, Duffy AM (2010) The evolution of chloroplast genome structure in ferns. Genome 53:731–738

Wolf PG, Der JP, Duffy AM, Jacobson JB, Grusz AL, Pryer KM (2011) The evolution of chloroplast genes and genomes in ferns. Plant Mol Biol 76: 251–261

Wolfe KH, Morden CW, Palmer JD (1992) Functions and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA 89:10648–10652