Functional Gene Losses Occur with Minimal Size Reduction in the Plastid Genome of the Parasitic Liverwort *Aneura mirabilis*

Norman J. Wickett, Yan Zhang,† S. Kellon Hansen,‡ Jessie M. Roper,‡ Jennifer V. Kuehl,§ Sheila A. Plock,† Paul G. Wolf,‖ Claude W. dePamphilis,† Jeffrey L. Boore,§ and Bernard Goffinet‡*

*Department of Ecology and Evolutionary Biology, University of Connecticut; †Department of Biology, Penn State University; ‡Genome Project Solutions, Hercules, California; §Department of Energy Joint Genome Institute and University of California Lawrence Berkeley National Laboratory, Walnut Creek, California; and ‖Department of Biology, Utah State University

Aneura mirabilis is a parasitic liverwort that exploits an existing mycorrhizal association between a basidiomycete and a host tree. This unusual liverwort is the only known parasitic seedless land plant with a completely nonphotosynthetic life history. The complete plastid genome of *A. mirabilis* was sequenced to examine the effect of its nonphotosynthetic life history on plastid genome content. Using a partial genomic fosmid library approach, the genome was sequenced and shown to be 108,007 bp with a structure typical of green plant plastids. Comparisons were made with the plastid genome of *Marchantia polymorpha*, the only other liverwort plastid sequence available. All *ndh* genes are either absent or pseudogenes. Five of 15 *psb* genes are pseudogenes, as are 2 of 6 *psa* genes and 2 of 6 *pet* genes. Pseudogenes of *cys*A, *cys*T, *ccs*A, and *ycf*3 were also detected. The remaining complement of genes present in *M. polymorpha* is present in the plastid of *A. mirabilis* with intact open reading frames. All pseudogenes and gene losses co-occur with losses detected in the plastid of the parasitic angiosperm *Epifagus virginiana*, though the latter has functional gene losses not found in *A. mirabilis*. The plastid genome sequence of *A. mirabilis* represents only the second liverwort, and first mycoheterotroph, to have its plastid genome sequenced. We observed a pattern of genome evolution congruent with functional gene losses in parasitic angiosperms but suggest that its plastid genome represents a genome in the early stages of decay following the relaxation of selection pressures.

Introduction

Plastids are organelles that originated as a photosynthetic, endosymbiotic cyanobacterium of early eukaryotic cells (Margulis 1970; Schwartz and Dayhoff 1978; Reith and Munholland 1993; Douglas 1999). As the once free-living, autotrophic cyanobacterium was integrated into its host, genes essential to its autonomy, indeed, the majority of its genes, were transferred to the host's nucleus (Gray 1983; Palmer 1985; Reith and Munholland 1993; Martin and Hermann 1998; Martin et al. 2002). However, plastids retain a remnant genome containing, primarily, genes integral to arguably the most important metabolic process on earth: photosynthesis. As green algae and their descendents diversified, the structure and content of the plastid genome have remained remarkably conserved, particularly so in embryophytes (Raubeson and Jansen 2005). Embryophyte plastid genome maps are generally assembled as 2 inverted repeat (IR) regions that separate a large single-copy (LSC) region from a small single-copy (SSC) region. The genome usually contains 110–130 unique genes, the majority of which code for proteins involved in photosynthesis or gene expression, with the remaining genes coding for transfer and ribosomal RNA (Raubeson and Jansen 2005).

Genes have been lost independently several times throughout the evolution of plastid-containing lineages (Baldauf et al. 1990; Palmer 1991; Martin et al. 1998; Goffinet et al. 2005), but evidence suggests that purifying selection acts on a core complement of genes to be retained in the plastid genome (Palmer 1985; Race et al. 1999; Bungard 2004). If, however, the selective constraints on the plastid genome are relaxed, we would expect to see an accumulation of pseudogenes and gene losses. Such is the case with the plastid genome of *Epifagus virginiana*, an obligate root parasite of beech trees (dePamphilis and Palmer 1990; Wolfe et al. 1992). The decay of the plastid genome in *E. virginiana* likely occurred relatively quickly with rapid accumulation of short insertions and deletions and an increase in mutation rate (dePamphilis and Palmer 1990; Wolfe et al. 1992). We would expect to observe a similar pattern of genome reduction in other plants that no longer directly rely on photosynthesis to acquire carbon. One such plant is the only known heterotrophic lineage of bryophytes.

Aneura mirabilis (Malmb., formerly known as *Cryptothallus mirabilis*; Wickett and Goffinet 2007) is a simple thalloid (Metzgeriales) parasitic liverwort. Unlike haustorial parasites that directly penetrate host tissue, *A. mirabilis* "cheats" a mycorrhizal association between a basidiomycete, *Tulasnella* sp., and a host tree such as pine and birch, a life history commonly known as mycoheterotrophy or epiparasitism (Bidartondo et al. 2003; Bidartondo 2005; Leake 2005). The absence of roots in bryophytes precludes the ability to form a haustorium, and Ligrone et al. (1993) outline evidence to suggest that *A. mirabilis* is, in fact, parasitic; however, the use of "parasite" in this case is not well established. In *A. mirabilis*, fixed carbon is obtained indirectly from the host via the fungus that is simultaneously ectomycorrhizal on the host's roots and endophytic in the liverwort (Bidartondo et al. 2003). This exploitation of an existing symbiosis allows *A. mirabilis* to maintain its subterranean existence. Living buried in damp humus found up to 30 cm under damp mats of *Sphagnum* peat or other bryophytes, the enigmatic liverwort is clearly able to exist in the absence of sunlight (Williams 1950; Schuster 1992). With an alternative source of energy, the plastid genome of *A. mirabilis* may be released from selective constraints, and we would expect to observe a trend of genome reduction as seen in *E. virginiana*.

Key words: *Aneura mirabilis*, *Cryptothallus*, parasitic plants, liverworts, bryophytes, chloroplast genome.

E-mail: norman.wickett@uconn.edu.

Mol. Biol. Evol. 25(2):393–401. 2008
doi:10.1093/molbev/msm267
Advance Access publication December 1, 2007

In the holoparasite *E. virginiana*, the plastid genome is 70,028 bp (Wolfe et al. 1992) compared with 155,939 bp in tobacco (Wakasugi et al. 1998). Four ribosomal RNA genes, 17 transfer RNA genes, and only 21 protein-coding genes remain, of which only 4 are not directly involved in plastid gene expression. Evidence that these genes are evolving under purifying selection, as well as evidence that nuclear-encoded proteins are still targeted to the plastid, suggests that there is a plastid function beyond the bioenergetic processes of photosynthesis and chlororespiration, perhaps, in fatty acid biosynthesis (dePamphilis and Palmer 1990; Schnurr et al. 2002). Similar patterns of gene loss and retention have been observed in the hemi- and holoparasitic plant genus *Cuscuta* (Haberhausen et al. 1992; Krause et al. 2003); however, little is known about nonphotosynthetic plastid genome evolution outside the angiosperms or in mycoheterotrophs (Bidartondo 2005). *Aneura mirabilis* presents a unique case for broadening the phylogenetic breadth of our understanding of plastid genome structure and content in nonphotosynthetic embryophytes.

Liverworts are sister to all other land plants (Qiu et al. 2006), and one of the earliest plastid genome sequences reconstructed was that of *Marchantia polymorpha*, in 1986 (Ohyama et al. 1986). Currently, this remains the only liverwort plastid genome sequence available, whereas over 70 have been published for the sister group (GenBank). In this study, we present the second plastid sequence of a liverwort, *A. mirabilis*, allowing us to address plastid evolution in an independent lineage of nonphotosynthetic embryophytes.

Materials and Methods

Plastid DNA was prepared for sequencing using a partial genomic fosmid library. This method, including DNA extraction, fosmid library construction, and clone selection and preparation, is described in detail in McNeal et al. (2006). Six grams of fresh *A. mirabilis* was used, and a voucher specimen is deposited at the University of Connecticut herbarium (CONN). Plastid-specific probes were produced by labeling, with [α–^{32}P]dATP, a pool of polymerase chain reaction (PCR) products generated from the plastid genes *rps*4, *rps*14, *trn*L, *psa*C, *psb*A, the *atp*B-*rbc*L spacer, and the *ndh*B region between *rps*7 and *trn*L-CAA (in *M. polymorpha*). Positive hybridizations indicated clones containing plastid inserts. Twenty-six positive hybridizations were observed and then selected from the original 384-well plate. Each fosmid insert was end sequenced using the T7 forward primer and the pEpiFos reverse primer (see manufacturer's manual for primer sequences) using an ABI 3100 genetic analyzer. Cycle sequencing reactions were performed with 1 µl DNA template, 1 µl each of 10 mM primer, 2 µl of ABI BigDye terminator, with water added to a total volume of 10 µl. Each sequence was submitted to a BlastN search to determine the location and direction of each insert relative to the *M. polymorpha* genome. PCR tests confirmed the identity of each insert, and a minimally overlapping set of 4 clones covering the complete plastid genome was selected for shotgun sequencing.

Prepared fosmid clones were mechanically sheared into random fragments of approximately 3 kb using a Hydroshear device (GeneMachines, San Carlos, CA). These fragments were enzymatically end repaired, and 3-kb fragments were purified by gel electrophoresis. Fragments were ligated into dephosphorylated pUC18 vector, transformed into *Escherichia coli* using standard techniques (Sambrook et al. 1989), and arrayed into 384-well plates, one for each prepared fosmid clone. The clones were robotically processed through rolling circle amplification of plasmids and end sequenced. These sequences were processed, trimmed, screened for vector sequence, and assembled using Phred and Phrap (Ewing and Green 1998; Ewing et al. 1998). Quality scores and assemblies were viewed and verified using Consed 12 (Gordon et al. 1998). A short gap in coverage, approximately 600 bp, was closed by PCR amplification using newly designed primers (psbMF: GGAAGTTAATGTCTCAGCATTTG and trnCR: GGTGACATGGCCAAGTGGGAAGGC) and sequenced using an ABI 3100 genetic analyzer. Automated assembly methods cannot distinguish between the 2 IRs; therefore, manual input was required to reconstruct part of one IR. The final assembly has an average depth of coverage of 8× and an average quality score per base of 85.7. We assembled the sequence as a circular map with 2 copies of the IR, an organization verified by the occurrence of individual IR boundaries in 3 of the 4 fosmid clones selected for shotgun sequencing. We annotated the genome using Dual Organellar GenoMe Annotator (DOGMA) available on the web at http://dogma.ccbb.utexas.edu/ (Wyman et al. 2004). Genes were located by using a database of previously published chloroplast genomes, from which Blast searches (Altschul et al. 1997) are used to find approximate gene positions. From this initial annotation, we located hypothetical starts, stops, and intron positions based on comparisons to homologous genes in other chloroplast genomes.

Gene content conservation between 3 other bryophyte plastid genomes and 6 tracheophyte plastid genomes was visualized using MultiPipMaker (Schwartz et al. 2000) available on the web at http://pipmaker.bx.psu.edu/pipmaker/. MultiPipMaker (Schwartz et al. 2000) is a powerful tool for visualizing genome-level conservation across multiple genomes, irrespective of gene order (see Maul et al. 2002 for a detailed description of its utility). We chose *Physcomitrella patens* (Sugiura et al. 2003; GenBank accession number AP005672) as the reference genome because its annotation is more reflective of current gene nomenclature than that of the early genome of *M. polymorpha*. Other plastid genomes included in the analyses were *M. polymorpha* (Ohyama et al. 1986; NC_001319), the hornwort *Anthoceros formosae* (Kugita et al. 2003; NC_004543), the lycophyte *Huperzia lucidula* (Wolf et al. 2005; AY660566), the pines *Pinus thunbergii* (Wakasugi et al. 1994; NC_001631) and *Pinus koraiensis* (Noh EW, Lee JS, Choi YI, Han MS, Yi YS, Han SU, unpublished data), *Arabidopsis thaliana* (Sato et al. 1999; NC_000932), *Nicotiana tabacum* (Shinozaki et al. 1986; NC_001879), and *E. virginiana* (Wolfe et al. 1992; NC_001568). We used the program Mulan (Ovcharenko et al. 2005), available on the web at http://mulan.dcode.org/, to visualize gene order

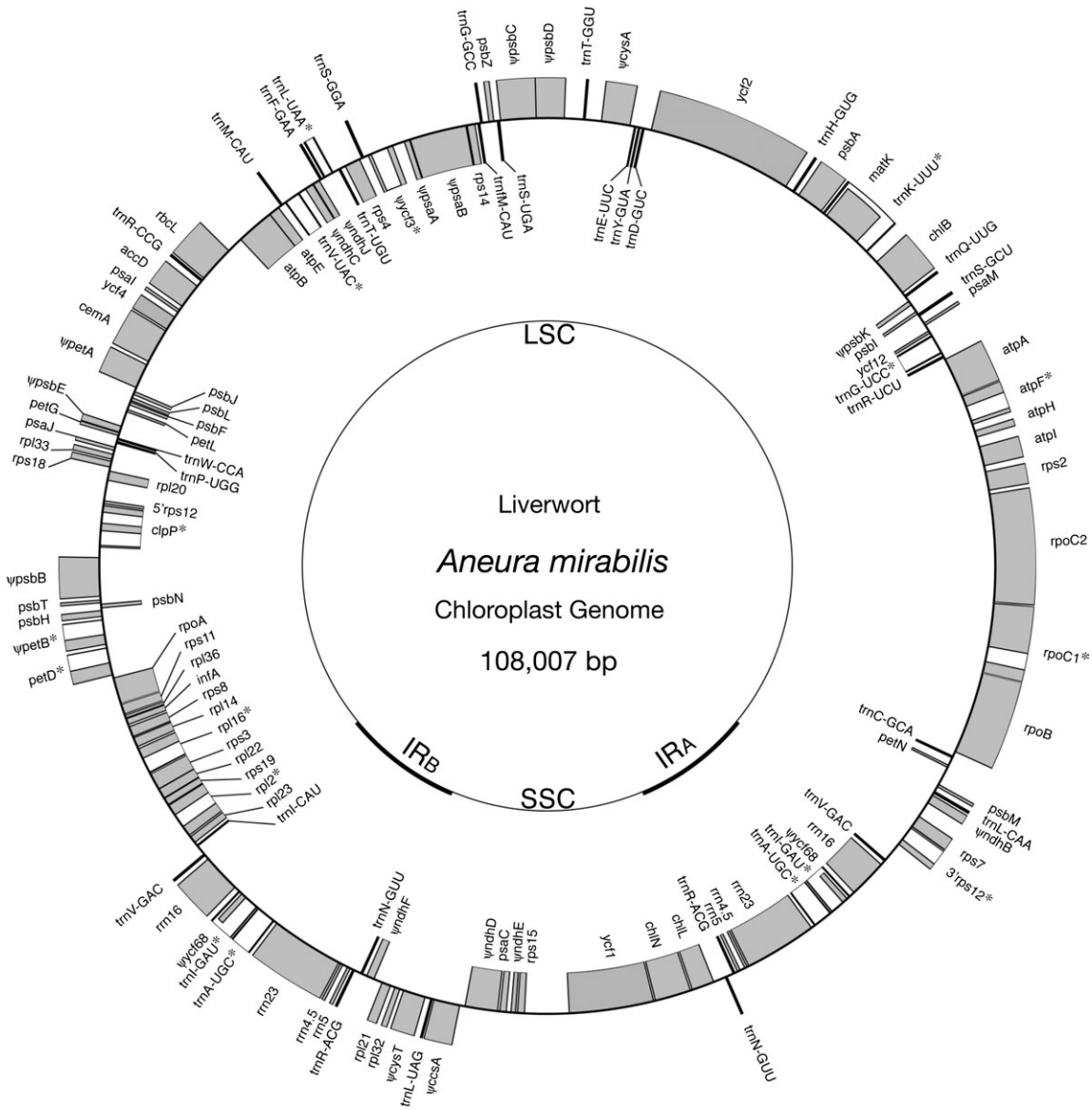

FIG. 1.—Plastid genome map of *Aneura mirabilis*. Genes on the outside of the circle are transcribed in the counterclockwise direction, and genes on the inside of the circle are translated in the clockwise direction. Structural components are labeled on the inner circle as LSC and SSC regions and IR. Asterisks denote split genes. Pseudogenes are notated with a ψ.

conservation between *M. polymorpha* and *A. mirabilis*, with the latter used as the reference genome.

Results

The plastid genome of *A. mirabilis* is 108,007 bp with an LSC region of 77,553 bp separated from a 13,974-bp SSC region by 2 IRs, each of 8,240 bp. The complete nucleotide sequence and annotation are deposited in GenBank (EU043314). The genome was drawn as a circular map with orientations of the LSC and SSC regions consistent with *M. polymorpha* (fig. 1). All genes in both *M. polymorpha* and *A. mirabilis* are in the same order and transcribed in the same direction, with the exception of an inversion of *psb*E and *pet*L in the LSC region of *A. mirabilis* (fig. 2). The gene content and gene order of the IR is identical to that of *M. polymorpha*. A truncated pseudogene of the unknown protein *ycf*68 in *A. mirabilis* was detected between the 2 exons of *trn*I-GAU; however, Raubeson et al. (2007) concluded that levels of sequence divergence for this gene are consistent with noncoding regions in the IR and this is likely not a functional gene. As with *M. polymorpha*, the IR boundaries are between *trn*V-GAC and *trn*I-CAU (IR_B-LSC), between *trn*V-GAC and 3′*rps*12 (IR_A-LSC), between *trn*N-GUU and *ndh*F (IR_B-SSC), and between *trn*N-GUU and *chl*L (IR_A-SSC).

Of the 116 unique genes (i.e., including only one copy of the IR) present in the plastid genome of *A. mirabilis*, 31 are transfer RNAs, 4 are ribosomal RNAs, 4 are proteins of unknown function (*ycf* genes), and the remaining 77 genes code for proteins of known function. Nineteen of the protein-coding genes are pseudogenes. Internal stop codons

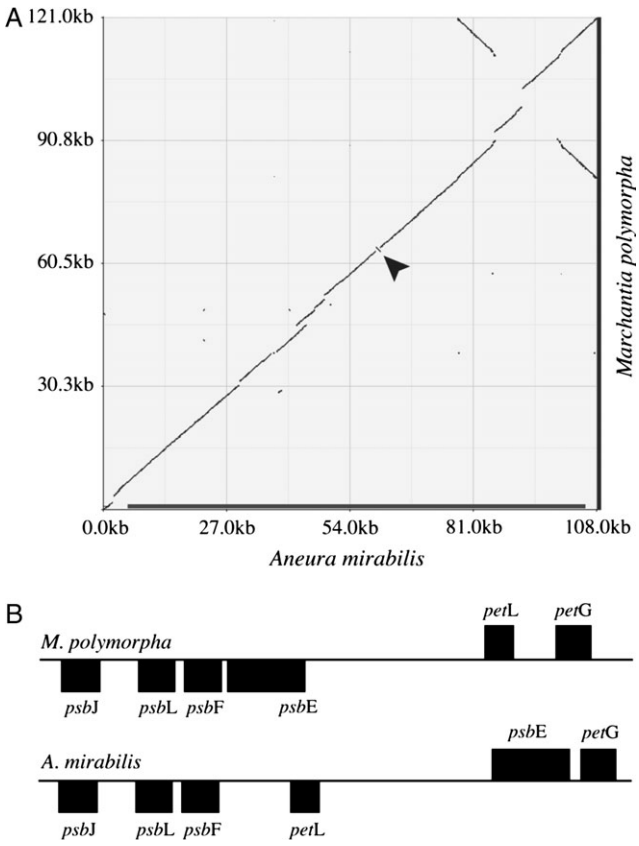

FIG. 2.—The 2 complete chloroplast genomes of liverworts *Aneura mirabilis* and *Marchantia polymorpha* are conserved with respect to gene order with the exception of a small inversion. (*A*) A PipMaker output from a Mulan analysis. Points along the positive slope are in the same orientation in both genomes, whereas points along the negative slope indicate sequences that can be aligned with each other but are oriented in opposite directions. The 2 groups of points that fall along the negative slope in the upper right corner represent the IRs. The short group of points along the negative slope indicated by the arrow represents an inversion of *psb*E and *pet*L shown in (*B*).

were detected in several genes due to short insertions or deletions that disrupt the reading frame. Insertions were verified by inspecting the original sequence reads spanning the insertion sites. In all cases, the read depth was greater than 8× and the insertion sites were unambiguous. Both photosystems I and II possess pseudogenes in *A. mirabilis*: *psa*A, *psa*B, *psb*B, *psb*C, *psb*D, *psb*E, and *psb*K. Two of 6 genes that code for cytochrome *b6/f* (*pet*), *pet*A and *pet*B are pseudogenes, as is *ccs*A, which is involved with heme attachment of cytochrome *c*. Of the 7 chlororespiratory (*ndh*) genes present in the SSC region of *M. polymorpha*, 4 are deleted in *A. mirabilis*, and the remaining 3, *ndh*D, *ndh*E and *ndh*F, are truncated pseudogenes. Four *ndh* genes are located in the LSC region of *M. polymorpha*; in *A. mirabilis*, *ndh*K is absent, and *ndh*B, *ndh*C, and *ndh*J are pseudogenes. The remaining pseudogenes include one protein of unknown function, *ycf*3, and 2 probable transport proteins, *cys*A and *cys*T. One protein of unknown function absent from *A. mirabilis*, *ycf*66, is not annotated in the plastid genome of *M. polymorpha*; however, this gene is recovered when the plastid genome sequence of *M. polymorpha* is annotated with DOGMA. Fifteen genes are split genes in *A. mirabilis*, 6 of which are transfer RNA genes. Split genes are consistent with those in *M. polymorpha* with the exception of *ndh*A and *ndh*B due to the partial or complete loss of these genes.

Figure 3 highlights the regions of the genome that are either missing or conserved between *A. mirabilis* and other embryophytes through a comparison with the genome of the moss *P. patens*. Box D, for example, shows the loss of *ndh* genes in the SSC of *A. mirabilis*, both *Pinus* genomes, and *E. virginiana*. The genomic region containing *psa*A, *psa*B, and *ycf*3 is highly conserved across all embryophytes, except in *A. mirabilis* and *E. virginiana* (box B). Box A includes *pet*A, which is a pseudogene in *A. mirabilis*. This region is conserved at the 75–100% level in the 3 bryophytes (excluding *A. mirabilis*) and *H. lucidula*, suggesting that selective constraints have been relaxed on this portion of the genome in *A. mirabilis*.

Discussion

The plastid genome of *A. mirabilis* is characterized by genome reduction and the complete or partial loss of 25 genes. Although gene loss from the chloroplast genome of land plants is not uncommon (see e.g., from bryophytes, Goffinet et al. 2005 and Sugiura et al. 2003), it is generally associated with single events and transfer to the nucleus,

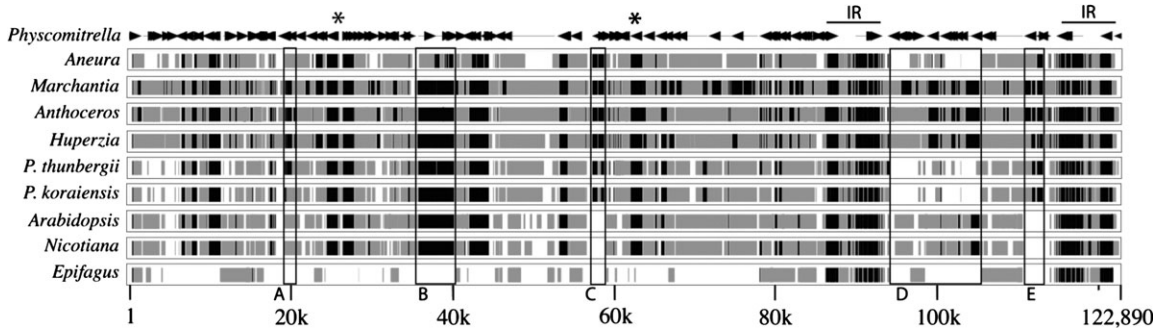

FIG. 3.—MultiPipMaker analysis of 10 sequenced land plant chloroplast genomes using *Physcomitrella patens* as the reference genome. The reference genome is represented as the uppermost line with gene orientations shown as arrows, and the IRs are shown as horizontal lines. Numbers across the bottom edge of the figure indicate positions on the linearized reference genome. Regions that align with the *P. patens* genome, irrespective of gene order, are represented as vertical black (75–100% identity) and gray (50–75% identity) bars. Boxed areas and asterisks are referred to in the text.

Table 1
Genes Present, Genes Absent, and Pseudogenes in *Aneura mirabilis* and *Epifagus virginiana* Compared with *Nicotiana tabacum* and *Marchantia polymorpha*, Respectively

	E. virginiana Compared with *N. tabacum*		*A. mirabilis* Compared with *M. polymorpha*	
	Gene Present	Gene Absent or Pseudogene (ψ)	Gene Absent or Pseudogene (ψ)	Gene Present
Photosynthesis				
Photosystem I		*psa*A, B, C, I, J ψ*psb*A, ψB, C, D, E, F, H, I, J, K, L, M	ψ*psa*A, ψB	*psa*C, I, J, M *psb*A, F, H, I, J, L, M, N, T, Z
Photosystem II			ψ*psb*B, ψC, ψD, ψE, ψK	
Cytochrome *b6f*		*pet*A, B, D, G	ψ*pet*A, ψB	*pet*N, L, G, D
ATP synthase		ψ*atp*A, ψB, E, F, H, I		*atp*I, H, F, A, B, E
Rubisco		ψrbcL		*rbc*L
Chlororespiration		*ndh*A, ψB, C, D, E, F, G, H, I, J, K	*ndh*A, ψB, ψC, ψD, ψE, ψF, G, H, I ψJ, K	
Gene expression				
rRNA	16S, 23S, 4.5S, 5S			16S, 23S, 4.5S, 5S
Ribosomal protein	*rps*2, 3, 4, 7, 8, 11, 14, 18, 19; *rpl*2, 16, 20, 33, 36	*rps*15, 16, ψ14; *rpl*22, 23, 32		*rps*12, 7, 2, 14, 4, 18, 11, 8, 3, 19, 15; *rpl*33, 20, 36, 14, 16, 22, 2, 23, 21, 32
Transfer RNA	D_{GUC}, E_{UUC}, F_{GAA}, H_{GUG}, I_{CAU}, L_{CAA}, L_{UAG}, M_{CAU}, N_{GUU}, P_{UGG}, Q_{UUG}, R_{ACG}, S_{GCU}, S_{UGA}, W_{CCA}, Y_{GUA}, fM_{CAU}	ψA_{UGC}, ψC_{GCA}, G_{GCC}, G_{UCC}, ψI_{GAU}, K_{UUU}, L_{UAA}, ψR_{UCU}, ψS_{GGA}, T_{GGU}, T_{UGU}, V_{GAC}, V_{UAC}		A_{UGC}, C_{GCA}, D_{GUC}, E_{UUC}, F_{GAA}, G_{GCC}, G_{UCC}, H_{GUG}, I_{CAU}, I_{GAU}, K_{UUU}, L_{CAA}, L_{UAA}, L_{UAG}, M_{CAU}, N_{GUU}, P_{UGG}, Q_{UUG}, R_{UCU}, R_{ACG}, R_{CCG}, S_{GCU}, S_{GGA}, S_{UGA}, T_{UGU}, T_{GGU}, V_{UAC}, V_{GAC}, W_{CCA}, Y_{GUA}, fM_{CAU}
RNA polymerase		ψ*rpo*A, B, C1, C2		*rpo*A, B, C1, C2
Maturase	*mat*K			*mat*K
Initiation factor	*inf*A			*inf*A
Other protein-coding genes	*clp*P, *acc*D, *ycf*2, *ycf*1	ORF29, 31, 34, 62, 168, 184, 229, 313	ψycf3, ψ*ccs*A, ψ*cys*A, ψ*cys*T	*ycf*12, *chl*B, *ycf*2, *acc*D, *ycf*4, *cem*A, *clp*P, *ycf*1, *chl*N, *chl*L

rather than wholesale genome reduction (Raubeson and Jansen 2005). Overall reduction in genome content and function has been shown in angiosperms that no longer rely on the major bioenergetic process of the plastid: photosynthesis (e.g., Wolfe et al. 1992). *Aneura mirabilis* is the only bryophyte known to obtain carbon from a source other than photosynthesis (Bidartondo et al. 2003). Our findings from the plastid genome sequence of *A. mirabilis* indicate that the pattern of genome reduction seen in nonphotosynthetic angiosperms is not limited to flowering plants. The plastid genome of *E. virginiana* is substantially more reduced than that of *A. mirabilis* (fig. 3), but there are no genes lost in the liverwort that are retained in beechdrops, indicating a common directionality of these losses (table 1). We also see that the losses in *A. mirabilis* are concentrated in genes and functional groups that are completely lost from *E. virginiana* (table 1). It is likely that the less extensive genome reduction of *A. mirabilis* compared with *E. virginiana* reflects a recent loss of photosynthesis, which is consistent with amounts of DNA sequence divergence from its photosynthetic sister taxon (Wickett and Goffinet 2007). In the recently sequenced plastid genome of *Cuscuta gronovii*, a converse pattern of evolution was observed. Unlike *A. mirabilis*, *C. gronovii* exhibits a reduction in the transcriptional apparatus of the plastid, rather than in plastid-encoded genes of the photosynthetic apparatus (Funk et al. 2007). Though both plants are obligately heterotrophic and exhibit an intermediate level of genome decay between a fully photosynthetic plant and *E. virginiana*, the question of whether the relative importance of host-independent life-history stages accounts for this discrepancy must be raised.

The pattern of gene loss, or loss of function, is not surprising given the role of these genes. Twenty-one of the 25 gene losses are in functional groups associated with the bioenergetic processes of photosynthesis. Both photosystems I and II are made up of subunits encoded by the plastid genes *psa* and *psb*, respectively (Blankenship 2002), and both of these subunit groups possess pseudogenes in *A. mirabilis*. The 2 genes encoding the core reaction center of photosystem I appear to no longer be functional; *psa*B contains internal stop codons and only a minute fraction of *psa*A remains at all. Two of the 5 genes that encode subunits of the core reaction center of photosystem II, *psb*D and *psb*E, are pseudogenes, as are both core antenna protein-coding genes *psb*B and *psb*C. The other pseudogenes of photosystem II subunits are *psb*K and *psb*M that are involved in the oxidation of water and stability of the photosystem (Blankenship 2002). The cytochrome *b6f* complex is intimately involved with noncyclic electron flow

(Blankenship 2002), a process integral to oxygenic photosynthesis. This complex is composed of subunits encoded by the *pet* genes. Both of the major subunits of this complex, *pet*A and *pet*B, coding for cytochrome *f* and *b6*, respectively, are pseudogenes in *A. mirabilis*. The other major subunit, the Rieske protein, is encoded by the nuclear *pet*C gene; detection and sequencing of this gene from the nucleus is required to determine whether this gene has also lost its functionality in *A. mirabilis*, as is the case with other nuclear-encoded genes whose products are targeted to the plastid. Furthermore, the gene controlling heme attachment to cytochrome *c*, *ccs*A, which is also involved in noncyclic electron transport (Stoebe et al. 1998), is a pseudogene. With the core reaction centers of both photosystems disrupted, in addition to the more extensive disruption of photosystem II and electron flow between the photosystems, it appears that the selective constraints on the genetic photosynthetic apparatus have been relaxed.

Chlororespiratory genes are the most affected of any functional category in *A. mirabilis*, with the partial or complete loss of all genes. These genes were initially named as chlororespiratory genes based on their sequence similarity with mitochondrial respiration genes (Wakasugi et al. 1994). The 11 *ndh* genes of the chloroplast encode subunits of the reduced form of nicotinamide adenine dinucleotide–plastoquinone oxidoreductase complex, thought to reduce photooxidative stress, which occurs when high light intensities cause the capacity of the photosynthetic apparatus to be exceeded, resulting in the production of damaging reactive oxygen species (Quiles 2006). Bungard (2004) reviews the evidence that suggests that these genes are lost early and easily in the evolution of a nonphotosynthetic life history. *Cuscuta reflexa*, a holoparasitic plant in the Convovulaceae, appears to have an intact chloroplast genome with the exception of the functional loss of all *ndh* genes (Haberhausen and Zetsche 1994; Funk et al. 2007). The loss of these genes, despite variation in the amount of overall plastid genome reduction, occurs in other holoparisitic plants surveyed thus far (dePamphilis and Palmer 1990; Delavault et al. 1996; Funk et al. 2007). Interestingly, the chlororespiratory genes have been lost in a photosynthetic plant: *P. thunbergii* (Wakasugi et al. 1994), suggesting that another mechanism, perhaps anatomical or physiological, is in place to reduce photooxidative stress or that these genes have been transferred to the nucleus. Unlike *P. thunbergii*, there is no evidence that *A. mirabilis* is tolerant of high light intensities, suggesting that the loss of *ndh* genes in the latter is likely due to the lack of photooxidative stress in the absence of photosynthesis. The chlororespiratory gene *ndh*F, a truncated pseudogene in *A. mirabilis*, is a pseudogene in several species of a subfamily of orchids, the Epidendroideae (Neyland and Urbatsch 1996). Although these orchids are free living, they do rely extensively on fungal interaction for at least part of their life history (Bungard 2004). Furthermore, Chang et al. (2006) reported the loss of 3 *ndh* genes (*ndh*A, *ndh*F, and *ndh*H) from the orchid *Phalaenopsis aphrodite*. The remaining 8 *ndh* genes were reported to be pseudogenes. If chlororespiratory genes are lost in the early stages of a life-history shift toward heterotrophy, before the complete loss of photosynthesis, then a plant that has recently become fully mycoheterotrophic might display a pattern of gene loss similar to *A. mirabilis*: extensive loss of *ndh* genes with only minimal loss of other bioenergetic genes. However, the relative ease with which these genes are discarded from the plastid genome in plants representing the entire spectrum of photosynthetic capacity, including a nonparasitic plant, raises the compelling question of why these genes are retained in the majority of plant lineages.

Of the remaining 4 pseudogenes in the plastid genome of *A. mirabilis*, *cys*A and *cys*T have a function predicted to be transport proteins and are found in *M. polymorpha* and *A. formosae* (only *cys*A) (Kugita et al. 2003) but not in vascular plants or in the moss *P. patens* (Sugiura et al. 2003). Their absence from most sequenced land plant plastid genomes indicates that these genes may not be part of the minimal complement of genes upon which functional constraints act, making it unsurprising that they are pseudogenes in *A. mirabilis*. The loss of *ycf*3 in *A. mirabilis*, but not in photosynthetic plants, suggests that this unknown protein is linked to photosynthesis.

Despite the deletion, or loss of function, of 25 genes, the plastid genome of *A. mirabilis* is remarkably colinear with its distant relative, *M. polymorpha* (fig. 2). Further studies are required to elucidate the phylogenetic significance of the *psb*E–*pet*L inversion. However, this is not an indication that all liverworts share a conserved gene order. Inversions may characterize higher level lineages of plants, such as legumes (Saski et al. 2005) and mosses (Goffinet et al. 2007). There may be unique gene orders in groups of liverworts whose plastid genome sequences have not yet been reconstructed.

When compared with tobacco, the plastid genome of *E. virginiana* reveals an elevated number of insertions and deletions (dePamphilis and Palmer 1990; Wolfe et al. 1992). An accumulation of short deletions may account for the dramatic reduction in genome size; however, the absence of the majority of bioenergetic genes makes it difficult to observe this pattern in either the LSC or the SSC regions. An unprecedented amount of the plastid genome sequence of *A. mirabilis* is occupied by pseudogenes, which may allow for a closer observation of the pattern of genome decay in bioenergetic genes. Of the 71,536 bp of coding sequence in the genome, pseudogenes account for 20.36% of the sequence "space" and are annotated as such due in large part to short insertions and deletions that disrupt the reading frame. Disregarding the *ndh* genes, which are generally either lost entirely or severely truncated, 11 short (<35 bp) deletions and 7 short insertions (when compared with *M. polymorpha* using *P. patens* as the outgroup) account for the observed functional loss in all but 2 (*psb*K and *cys*T) of the observed pseudogenes. The disruption of the reading frame in these genes does not seem to co-occur with an increase in sequence divergence, as evidenced by high sequence identity between *A. mirabilis* and *M. polymorpha*. For example, *pet*A, *psa*B, *psb*B, and *psb*C have sequence identities of 79%, 84%, 82% and 83%, respectively, when the alignment is adjusted to take indels into account. These sequence identities are comparable with those for genes that are conserved both in the nonphotosynthetic liverwort and in *E. virginiana* (*acc*D, 84%; *rps*4, 74%). Accumulation of reading frame–disrupting indels,

perhaps, due to errors in genome replication, may account for the rapid decay of the plastid genome in parasitic plants.

Aneura mirabilis also displays an unexpectedly high level of plastid sequence conservation when compared with *M. polymorpha*, or even more distantly related land plants, as shown in figure 3. Despite the functional losses of several genes in *A. mirabilis*, some regions of the genome are remarkably conserved given their bioenergetic functions—for example, *atp*B, *rbc*L, and *atp*A, which are indicated by asterisks in figure 3. Some of this conservation may be due to evolutionary constraints on portions of the genome required for functions apart from photosynthesis. A transcriptionally active plastid is maintained in most parasitic angiosperms, and in the case of *E. virginiana*, nuclear gene products are targeted to the plastid (Bungard 2004). Some biosynthetic functions, such as the production of fatty acids (dePamphilis and Palmer 1990; Schnurr et al. 2002), may be carried out in the plastid despite not having a role in photosynthesis. However, the intact portions of the genome may simply be due to a recent shift to heterotrophism in *A. mirabilis*. Given the loss of most bioenergetic genes, including *atp*B, *atp*A, and *rbc*L, from *E. virginiana*, the latter explanation may be preferred.

Two genomic regions present a striking pattern in figure 3. Box C corresponds to *chl*B and box E corresponds to *chl*N and *chl*L. The former appears to be highly conserved across all nonangiosperm embryophytes, and the latter appears to be highly conserved, in part, across the same groups of plants. These genes encode 3 subunits of light-independent protochlorophyllide oxidoreductase (Armstrong 1998) that allows photosynthetic organisms to synthesize chlorophyll in the dark. Angiosperms are unable to produce chlorophyll if grown in the dark, which is consistent with the absence of these genes in flowering plants (Kusumi et al. 2006). Light-dependent chlorophyll synthesis (encoded by the nuclear *por* gene) may be more important than light-independent chlorophyll synthesis, as suggested by the absence of these genes in *Psilotum nudum*, *Welwitschia mirabilis*, some algae, and the functional loss in 4 species of *Thuja* (Kusumi et al. 2006). Though *A. mirabilis* is an albino liverwort, the shoot calyptra that protects the developing sporophyte and the spores themselves take on a greenish tint. However, *A. mirabilis* develops in the absence of light and must therefore rely on light-independent chlorophyll synthesis, and subsequently the *chl* genes, to produce the observed pigment. Sequence identity of *chl* sequences between *A. mirabilis* and the distantly related *M. polymorpha* (35) are 82%, 81%, and 78% for *chl*B, *chl*L, and *chl*N, respectively, suggesting that selection may act to retain these genes in the nonphotosynthetic liverwort. The role of chlorophyll in the underground liverwort is unclear, though sporophytes produced in the laboratory appear to grow toward the surface of the overlying vegetation, suggesting that there may be a period in its life history where it is exposed to light.

Bidartondo (2005) describes the study of mycorrhizal cheaters like *A. mirabilis* as a field in its "infancy," and much of our understanding of plastid genome evolution in nonphotosynthetic plants is limited to the angiosperm families Orobanchaceae and Convovulaceae, which possess strictly parasitic plants rather than mycoheterotrophs. This study presents the first plastid genome sequence of a mycoheterotroph, the first of a nonangiosperm nonphotosynthetic plant, and only the second plastid sequence of a liverwort. Functional gene losses from the plastid are consistent with predictions based on plastid evolution in holoparasitic angiosperms, suggesting that genome evolution under the relaxation of photosynthesis-mediated constraints follows broad universal patterns, rather than being lineage specific.

Supplementary Material

The GenBank accession number for this study is EU043314.

Acknowledgments

This research was funded by grants from the National Science Foundation (EF-0531557 to B.G., DEB0120709 to C.W.D., DEB0228432 to P.G.W., and a doctoral dissertation improvement grant DEB0408043); Sigma Xi; and the University of Connecticut (R. A. Bamford endowment to Ecology and Evolutionary Biology and a Doctoral Dissertation Fellowship). Cecilia Sérgio provided the plant material used for DNA extraction. The authors thank Louise Lewis and Kent Holsinger for comments on an early version of the manuscript and Joel McNeal for assistance in the construction of fosmid libraries.

Literature Cited

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Armstrong GA. 1998. Greening in the dark: light-independent chlorophyll biosynthesis from anoxygenic photosynthetic bacteria to gymnosperms. J Photochem Photobiol B Biol. 43:87–100.

Baldauf SL, Manhart JR, Palmer JD. 1990. Different fates of the chloroplast *tuf*A gene following its transfer to the nucleus in green algae. Proc Natl Acad Sci USA. 87:5317–5321.

Bidartondo MI. 2005. The evolutionary ecology of mycoheterotrophy. New Phytologist. 167(2):335–352.

Bidartondo MI, Bruns TD, Weiss M, Sérgio C, Read DJ. 2003. Specialized cheating of the ectomycorrhizal symbiosis by an epiparasitic liverwort. Proc R Soc Lond B Biol Sci. 270(1517):835–842.

Blankenship RE. 2002. Molecular mechanisms of photosynthesis. Oxford: Blackwell.

Bungard RA. 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. BioEssays. 26: 235–247.

Chang CC, Lin HC, Lin IP, et al. (11 coauthors). 2006. The chloroplast genome of *Phalanopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol. 23(2):279–291.

Delavault PM, Russo NM, Lusson NA, Thalouarn PA. 1996. Organization of the reduced plastid genome of *Lathraea clandestina*, and achlorophyllous parasitic plant. Physiol Plant. 96:674–682.

dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature. 348:337–339.

Douglas SE. 1999. Evolutionary history of plastids. Biol Bull. 196:397–399.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred II. Error probabilities. Genome Res. 8:186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred I. Accuracy assessment. Genome Res. 8:175–185.

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 Biology. 7:45.

Goffinet B, Wickett NJ, Shaw AJ, Cox CJ. 2005. Phylogenetic significance of the *rpo*A loss in the chloroplast genome of mosses. Taxon. 54:353–360.

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.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.

Gray MW. 1983. The bacterial ancestry of plastids and mitochondria. Bioscience. 33:693–699.

Haberhausen G, Valentin K, Zetsche K. 1992. Organisation and sequence of photosynthetic genes from the plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. Mol Gen Genet. 232:154–161.

Haberhausen G, Zetsche K. 1994. Functional loss of all *ndh* genes in an otherwise relatively unaltered plastid genome of the holoparasitic flowering plant *Cuscuta reflexa*. Plant Mol Biol. 24:217–222.

Krause K, Berg S, Krupinska K. 2003. Plastid transcription in the holoparasitic plant genus *Cuscuta*: parallel loss of the rrn16 PEP-promoter and of the *rpo*A and *rpo*B genes coding for the plastid-encoded RNA polymerase. Planta. 216:815–823.

Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res. 31(2):716–721.

Kusumi J, Sato A, Tachida H. 2006. Relaxation of functional constraint on light-independent protochlorophyllide oxidoreductase in *Thuja*. Mol Biol Evol. 23(5):941–948.

Leake JR. 2005. Plants parasitic on fungi: unearthing the fungi in myco-heterotrophs and debunking the "saprophytic" plant myth. Mycologist. 19:113–122.

Ligrone R, Duckett JG, Renzaglia KS. 1993. The gametophyte-sporophyte junction in land plants. Adv Bot Res. 19:231–317.

Margulis L. 1970. Origin of eukaryotic cells. New Haven (CT): Yale University Press.

Martin WF, Hermann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. 118:9–17.

Martin WF, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. 2002. Evolutionary analysis of *Arabidopsis*, cynobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci USA. 99(19):12246–12251.

Martin WF, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature. 393:162–165.

Maul JE, Lilly JW, Cui L, DePamphilis CW, Miller W, Harris EH, Stern DB. 2002. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. The Plant Cell. 14:1–21.

McNeal JR, Leebens-Mack JH, Arumuganathan K, Kuehl JV, Boore JL, DePamphilis CW. 2006. Using partial genomic fosmid libraries for sequencing complete organellar genomes. Biotechniques. 41(1):69–73.

Neyland R, Urbatsch LE. 1996. Phylogeny of subfamily Epidendroideae (Orchidaceae) inferred from *ndh*F chloroplast gene sequences. Am J Bot. 83(9):1195–1206.

Ohyama K, Fukuzawa H, Kohchi T, et al. (13 coauthors). 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature. 322:572–574.

Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W. 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. Genome Res. 15:184–194.

Palmer JD. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae. In: MacIntyre RJ, editor. Monographs in evolutionary biology; molecular evolutionary genetics. New York: Plenum. p. 131–240.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. In: Bogorad LJ, Vasil IK, editors. Cell culture and somatic cell genetics of plants, volume 7A: the molecular biology of plastids. San Diego (CA): Academic Press, Inc. p. 5–53.

Qiu Y-L, Lia L, Wang B, et al. (21 coauthors). 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc Nat Acad Sci USA. 103:15511–15516.

Quiles MJ. 2006. Stimulation of chlororespiration by heat and high light intensity in oat plants. Plant Cell Environ. 29:1463–1470.

Race HL, Hermann RG, Martin WF. 1999. Why have organelles retained genomes? Trends Genet. 15(9):364–370.

Raubeson LA, Jansen RK. 2005. Chloroplast genomes of plants. In: Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. Wallingford (UK): CAB International. p. 45–68.

Raubeson LA, Peery R, Chumley T, 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.

Reith M, Munholland J. 1993. A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. The Plant Cell. 5:465–475.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Springs Harbor (NY): CSH Laboratory Press.

Saski C, Lee SB, Daniell H, Wood TC, Tomkins J, Kim HG, Jansen RK. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Mol Biol. 59:309–322.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 6(5):283–290.

Schnurr JA, Schockey JM, Deboer GJ, Browse JA. 2002. Fatty acid export from the chloroplast. Molecular characterization of a major plastidial acyl-coenzyme A synthetase from *Arabidopsis*. Plant Physiol. 129(4):1700–1709.

Schuster RM. 1992. The Hepaticae and Anthocerotae of North America. Chicago: Field Museum of Natural History.Vol. V.

Schwartz RM, Dayhoff MO. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science. 199:395–403.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller R. 2000. PipMaker—a

webserver for aligning two genomic DNA sequences. Genome Res. 10(4):577–586.
Shinozaki K, Ohme M, Tanaka M, et al. (24 coauthors). 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO J. 5:2043–2049.
Stoebe B, Martin W, Kowallik KV. 1998. Distribution and nomenclature of protein-coding genes in 12 sequenced chloroplast genomes. Plant Mol Bio Rep. 16:243–255.
Sugiura C, Kobayashi Y, Aoki S, Sugita S, Sugita M. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpo*A from the chloroplast to the nucleus. Nucleic Acids Res. 31:5324–5331.
Wakasugi T, Sugita M, Tsudzuki T, Sugiura M. 1998. Updated gene map of tobacco chloroplast DNA. Plant Mol Biol Rep. 16:231–241.
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(21):9794–9798.
Wickett NJ, Goffinet B. Forthcoming 2007. Origin and relationships of the myco-heterotrophic liverwort *Cryptothallus mirabilis* Malmb. (Metzgeriales, Marchantiophyta). Bot J Linn Soc. 154.
Williams S. 1950. The occurrence of *Cryptothallus mirabilis* v. Malmb. in Scotland. Trans Br Bryol Soc. 1:357–366.
Wolf PG, Karol KG, Mandoli DF, Kuehl JV, 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(2):117–128.
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.
Wyman SK, Boore JL, Jansen RK. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252–3255.

Charles Delwiche, Associate Editor

Accepted November 20, 2007