Plant Mol Biol (2011) 76:263–272
DOI 10.1007/s11103-011-9753-5

Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae)

J. Chris Blazier · Mary M. Guisinger · Robert K. Jansen

Received: 21 July 2010 / Accepted: 31 January 2011 / Published online: 16 February 2011

Abstract Plastid genomes in the flowering plant family Geraniaceae are known to be highly rearranged based on complete sequences representing the four major genera *Erodium*, *Geranium*, *Monsonia*, and *Pelargonium*. In this paper we report on the genome sequence of a second species of *Erodium*, *E. carvifolium*, representing the second major clade (clade II) in the phylogeny of this genus. Comparison of this genome sequence to the previously published sequence of *E. texanum* from clade I demonstrates that the plastid genomes of these two species encode the same number of proteins but differ greatly in their relative degree of rearrangement; 14 kb of additional sequence in *E. texanum* contains complex repeats associated with rearrangement endpoints, whereas the plastid genome of *E. carvifolium* is streamlined at 116 kb and displays no unique alterations in gene order. Furthermore, these species from both major clades of *Erodium* contain intact NADH dehydrogenase (*ndh*) genes, but the 11 *ndh* genes are represented as pseudogenes in a small clade of 13 species. It is unclear whether plastid-encoded *ndh* genes have been lost entirely or functionally transferred to the nucleus. This is the third report of the absence of functional *ndh* genes, and the current study describes the most recent loss of these genes among photosynthetic seed plants and the second such loss among angiosperms. The other *ndh* losses from Pinaceae/Gnetales and Orchidaceae are much more ancient. Comparative biochemistry between *Erodium* species with and without plastid-encoded *ndh* genes may elucidate changes in photosynthetic function and the role of the Ndh complex.

Keywords Plastid genomics · *ndh* genes · Molecular evolution · Geraniaceae · *Erodium*

Guest Editor: Jacqueline Nugent.

J. Chris Blazier (✉) · M. M. Guisinger
Section of Integrative Biology, University of Texas, Austin, TX 78712, USA
e-mail: blaziermail@mail.utexas.edu

M. M. Guisinger
Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

R. K. Jansen
Section of Integrative Biology and Institute of Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA

Introduction

With very few exceptions the plastid genomes (plastomes) of angiosperms typically contain 79 protein-coding genes, 30 tRNA genes and four rRNA genes (reviewed in Raubeson and Jansen 2005; Bock 2007). Broadly, genes are involved in photosynthesis or housekeeping. Plastome organization is identical in the majority of angiosperms: two single copy regions, the large (~85 kb; kilobase) and small (~15 kb) single copy regions (LSC, SSC), are separated by two large (~25 kb) inverted repeats (IR) designated IRa and IRb. The ribosomal operon lies within the IR. Plastid DNA appears to be present as isomers, differing in the orientation of the single copy regions (Palmer 1983). A few seed plant lineages have lost one copy of the IR, but the vast majority of angiosperms retain this quadripartite structure. Gene order is generally the same across angiosperms, most exceptions being inversions in the LSC as in the grasses, some legumes and sunflower. Very seldom do inversions interrupt conserved transcriptional units (Raubeson and Jansen 2005; Bock 2007).

Relatively few plastid genes have been functionally transferred to the nucleus subsequent to the diversification of angiosperms (Bock and Timmis 2008). Recent transfers or functional substitutions have been documented for only five genes thus far, but for a few of these, such as *infA*, there have been repeated, independent transfers (Millen et al. 2001). Certain ribosomal protein genes appear to have been lost and presumably functionally transferred independently across angiosperms (Jansen et al. 2007), but most other classes of plastid genes are seldom or never lost from photosynthetic seed plants. For example, with few exceptions, genes involved in photosynthesis, electron transport, and ATP synthesis have not been found to be missing from the plastome of a photosynthetic seed plant (Bock 2007; Magee et al. 2010). The 11 plastid-encoded NADH dehydrogenase (*ndh*) genes are unique in being lost or retained only as a full set.

The 11 plastid-encoded *ndh* genes are only commonly lost in nonphotosynthetic plants (Martin and Sabater 2010). In fact, only two photosynthetic seed plant lineages have been documented to lack these genes, Pinaceae/Gnetales (Wakasugi et al. 1994; McCoy et al. 2008; Wu et al. 2009) and a large clade within the Orchidaceae (Neyland and Urbatsch 1996; Chang et al. 2006; Wu et al. 2010). In these two lineages, as in parasitic plants, the entire suite of 11 plastid-encoded *ndh* genes is lost together. Both of these losses of plastid-encoded *ndh* genes are relatively ancient—the divergence of Pinaceae has been estimated at approximately 140 million years ago (MYA) (Wang et al. 2000), and loss of *ndh* genes necessarily predates this divergence, especially if it represents a synapomorphy for Pinaceae and Gnetales (Braukmann et al. 2009). The sister relationship between Pinaceae and Gnetales—the so-called "gnepine" hypothesis—is still controversial (Zhong et al. 2010). However, whether *ndh* gene loss in these lineages represents a single or two separate events, the loss is relatively ancient. Loss of *ndh* genes in orchids is likely also ancient—fossil-calibrated molecular data suggests a crown radiation of Orchidaceae 76-84 MYA (Ramirez et al. 2007). The taxonomic group confirmed to have lost *ndh* genes encompasses at least 70 orchid genera with over 1,000 species (Wu et al. 2010), suggesting that the loss of *ndh* genes is not recent.

It is not known whether Ndh function has been lost or functionally replaced in these two lineages or the genes have been functionally transferred to the nucleus; however, experimental evidence suggests that the thylakoid-bound Ndh complex is a vital intermediary of linear/cyclic electron flow and is apparently indispensable to photosynthesis under a wide range of stress conditions (Endo et al. 1999; Martin et al. 2004; Rumeau et al. 2005; Casano et al. 2001). Unfortunately, neither the composition nor the functions of the Ndh complex have been fully characterized. At least three additional subunits are located in the nuclear genome of angiosperms (Rumeau et al. 2005).

A few angiosperm lineages have been identified with highly rearranged plastomes lacking a variety of genes and introns. One of these lineages, the geranium family (Geraniaceae), contains four major genera, each with distinctly rearranged plastomes (Chumley et al. 2006; Guisinger et al. 2011). All members of the family share the loss of *trnT*-ggu and of two introns (*rps16* and *rpl16*). At 217 kb, the garden geranium (*Pelargonium x hortorum*) has the largest and most rearranged plastid genome among angiosperms. Most of the expansion in the *P.* × *hortorum* plastid genome results from massive expansion of the IR into the single copy regions and from proliferation of complex repeats. While *P.* × *hortorum* has an IR triple the normal size for an angiosperm at 76 kb, the IR has been lost completely in *Erodium texanum* and reduced in *Geranium palmatum* (11 kb) and *Monsonia speciosa* (7 kb). *Monsonia speciosa* is the only land plant plastome in which the IR does not include the entire rRNA operon. *Geranium*, *Monsonia*, and *Erodium* all share the loss of the two large ycfs, *ycf1* and *ycf2*, present in *Pelargonium* and in most angiosperm plastid genomes. Families of large dispersed repeats associated with rearrangement endpoints are found in the plastomes of all four Geraniaceae genera (Guisinger et al. 2011).

We chose to survey the genus *Erodium* (with 74 species) to improve our understanding of the extent and diversity of plastome abnormalities in a single lineage within Geraniaceae. The focus of this paper is to compare plastid genome organization of species representing the two major clades and to report the loss of plastid-encoded *ndh* genes in a single divergent clade of 13 species. This clade represents the most recent loss of *ndh* genes yet identified among photosynthetic seed plants. Because the subgenera and sections in *Erodium* have been found to be polyphyletic and have not yet been revised, this clade has no official taxonomic status (Fiz et al. 2006). Here we refer to the *ndh*-lacking clade as the long-branch clade (LBC), as it is separated from the rest of the genus by a long branch in phylogenetic reconstructions (Guisinger et al. 2008).

Materials and methods

Taxon sampling and sample preparation

Based on a molecular phylogeny of the genus (Fiz et al. 2006), *E. carvifolium* was chosen to represent Clade II—the published sequence of *E. texanum* represents Clade I (Guisinger et al. 2011)—and three additional taxa,

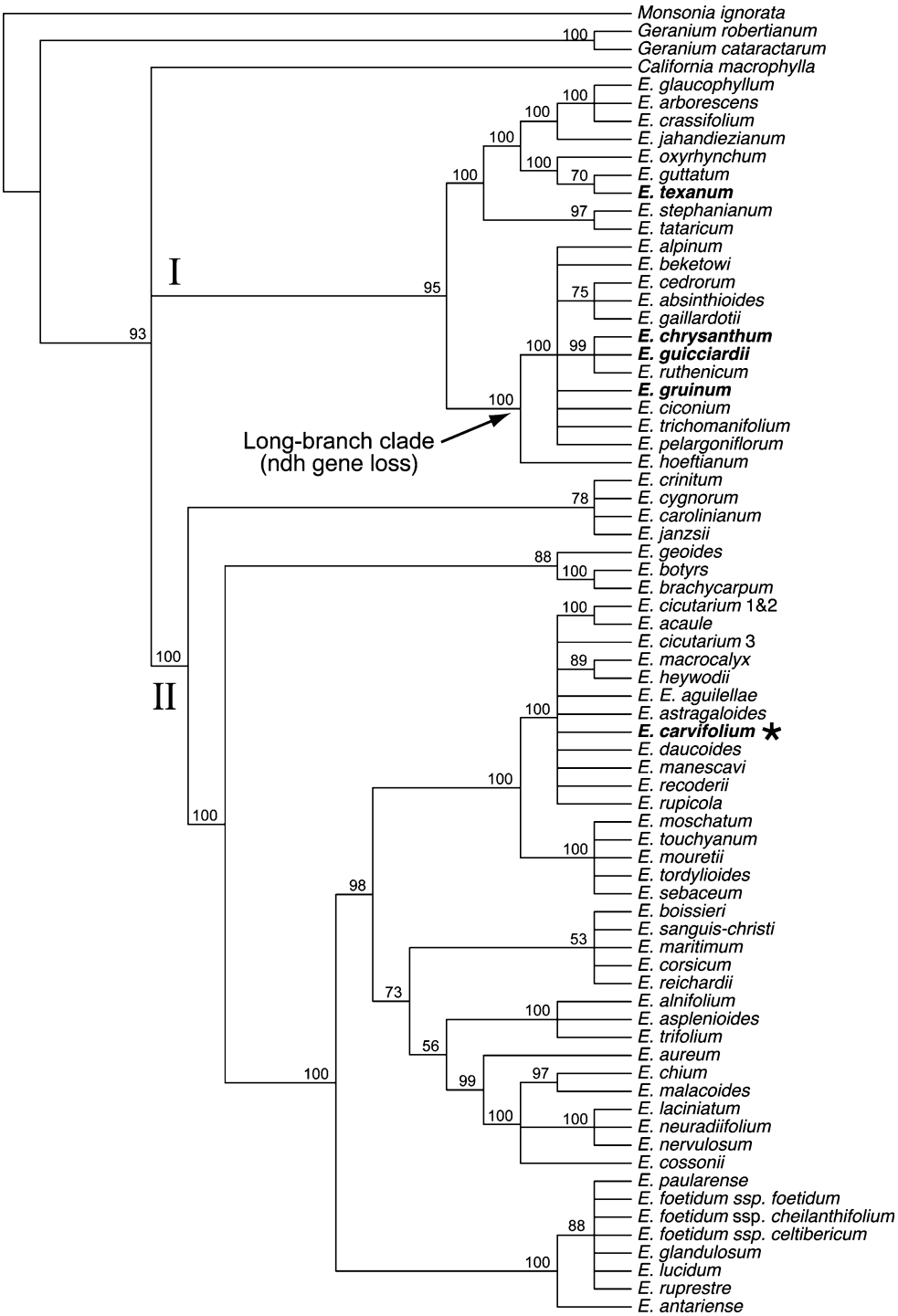

Fig. 1 Phylogeny of *Erodium* adapted from Fiz et al. (2006). Clades I, II, and the long-branch clade (LBC) are indicated. *Numbers* at nodes are Bayesian posterior probabilities as percentages. Taxa in *bold* represent those with complete (*E. texanum*, *E. carvifolium*) or draft plastid genome sequences. The *asterisk* indicates the new genome sequence of *E. carvifolium* reported in this paper

E. chrysanthum, *E. guicciardii* and *E. gruinum*, were chosen from a group of species designated as the long-branch clade (LBC; Fig. 1). Protocols for plastid isolation have been previously described (Jansen et al. 2005). Five to 10 g of fresh young leaf tissue from a single plant was used in each plastid isolation; plants were obtained from commercial sources (Geraniaceae.com; B&T World Seeds, Paguignan, France) and cultivated in the greenhouses at the Brackenridge Field Laboratory at University of Texas at Austin. Vouchers are deposited at TEX.

Genome amplification, sequencing, and finishing

Plastid DNA was amplified using rolling circle amplification (RCA; Qiagen GmbH, Hilden, Germany) utilizing bacteriophage Phi29 polymerase and random hexamer

primers (Dean et al. 2001). An EcoRI restriction digest was performed and visualized with ethidium bromide on 1% agarose gel to verify the purity and quantity of plastid DNA. Amplified DNA was either sent to the DOE Joint Genome Institute (Walnut Creek, CA) for Sanger sequencing or the W. M. Keck Center for Comparative and Functional Genomics at University of Illinois for 454 pyrosequencing.

Genomic data was assembled into contigs de novo in the native 454 assembler (Newbler) under default settings, as well as with open-source assembler MIRA (Chevreux 2009) using the "accurate" setting. Putative gene identifications in each contig were performed in DOGMA (Wyman et al. 2004), and contigs were assembled in Geneious (Drummond et al. 2009). Final annotations were made using DOGMA. A circular map was created for *E. carvifolium* using GenomeVx (Conant and Wolfe 2007).

Extraction and analysis of *ndh* pseudogenes

Due to long complex repeats including the apparent duplication of the *atpB/E* transcriptional unit (data not shown), the plastomes of *E. chrysanthum*, *E. guicciardii*, and *E. gruinum* have not yet been assembled into single contigs. To investigate the presence/absence of *ndh* genes, contigs for each species were converted to a custom BLAST database and queried with intact *ndh* genes from *E. texanum* and *E. carvifolium*.

Protein-coding genes were previously extracted from the *E. chrysanthum* genome assembly (Guisinger et al. 2008), suggesting that the assembly represents the complete plastid genome. The assembly is comprised of 3034 Sanger reads in three large contigs (37, 37, and 32 kb) and three small contigs (6, 3, and 2 kb) for a concatenated length of 118,538 bp. The *E. guiccardii* and *E. gruinum* datasets comprise of 27,620 and 43,067 454 Titanium pyrosequencing reads of 370 and 379 bp average length, respectively; the largest contigs obtained were 101 and 98 kb respectively. Both assemblies contain the same set of protein-coding genes previously found in *E. chrysanthum* (data not shown). Besides missing *ndh* genes, all three assemblies show the loss of the *rpoC1* intron, which is present in both *E. texanum* and *E. carvifolium*.

Putative *ndh* pseudogenes were aligned against intact genes from *E. texanum* and *E. carvifolium* in Geneious under default settings; indels relative to the intact genes were characterized. Only indels >3 basepairs (bp) were tabulated in order to avoid counting as indels any sequencing errors due to homopolymer runs. For degraded, unalignable pseudogenes the sizes of BLAST hits from *E. texanum ndh* genes were tabulated.

Primers were designed to amplify an approximately 600 bp region of *ψndhD* containing four deletions shared among the three LBC genome taxa (*ψndhD*-Forward TCC GCAGGTTCCTTCATTTGT; *ψndhD*-Reverse TCTCCGC GAGTGTCTGGTAAC). This *ψndhD* region was amplified for all 13 LBC taxa to confirm the presence of pseudogenes with shared indels throughout the clade. LBC DNAs were provided by J. J. Aldasoro. Sanger sequencing of PCR products was performed on an ABI 3730 platform at The University of Texas at Austin.

Two non-*ndh* genes formerly transcribed with *ndh* genes, *rps15* and *psaC*, were also extracted from LBC genome assemblies using BLAST in the manner described above for *ndh* pseudogenes. The *rps15* and *psaC* genes were aligned with those from *E carvifolium* and *E. texanum* using MUSCLE (Edgar 2004) under default settings as implemented in Geneious.

Results

Genome organization in *Erodium*

The plastome of *E. carvifolium* is 116,934 bp and contains 75 protein-coding genes, 28 tRNA genes and the standard four rRNA genes (Fig. 2; accession number NC_015083). Gene content of *E. carvifolium* is nearly identical to that of *E. texanum* (Table 1), the only difference being the putative loss of *trnK*-uuu in the latter (Guisinger et al. 2011). Both exons of *trnK*-uuu appear intact in *E. carvifolium*. Relative to *Arabidopsis* the first and second exons of *trnK*-uuu in *E. carvifolium* both have 94% identity, compared to 83.8% and 63.9% for the *E. texanum* exons, respectively. The losses of the *trnT*-ggu gene and the introns in *rpl16* and *rps16* were previously described for the family Geraniaceae, and we confirm these losses in *E. carvifolium*. The genes *ycf1*, *ycf2*, and *trnG*-gcc were shown to be lost in three of the four major Geraniaceae genera (*Erodium*, *Geranium*, and *Monsonia*) and these genes also appear to be lost in *E. carvifolium*. The gene *accD* is absent from *E. texanum*, and we did not detect the gene in *E. carvifolium.*

Given the high degree of genomic rearrangement found in representatives of all major Geraniaceae genera (Chumley et al. 2006; Guisinger et al. 2011), the plastome of *E. carvifolium* is remarkably unrearranged, displaying no unique inversions. Just two inversions separate *E. carvifolium* from the ancestral angiosperm gene order typified by tobacco (Raubeson and Jansen 2005), and both inversions are shared among all Geraniaceae plastomes. Among photosynthetic angiosperms *E. carvifolium* has the smallest reported plastid genome at 116,934 bp. Extensive rearrangement in *E. texanum* is associated with the proliferation of long, complex repeats in intergenic regions,

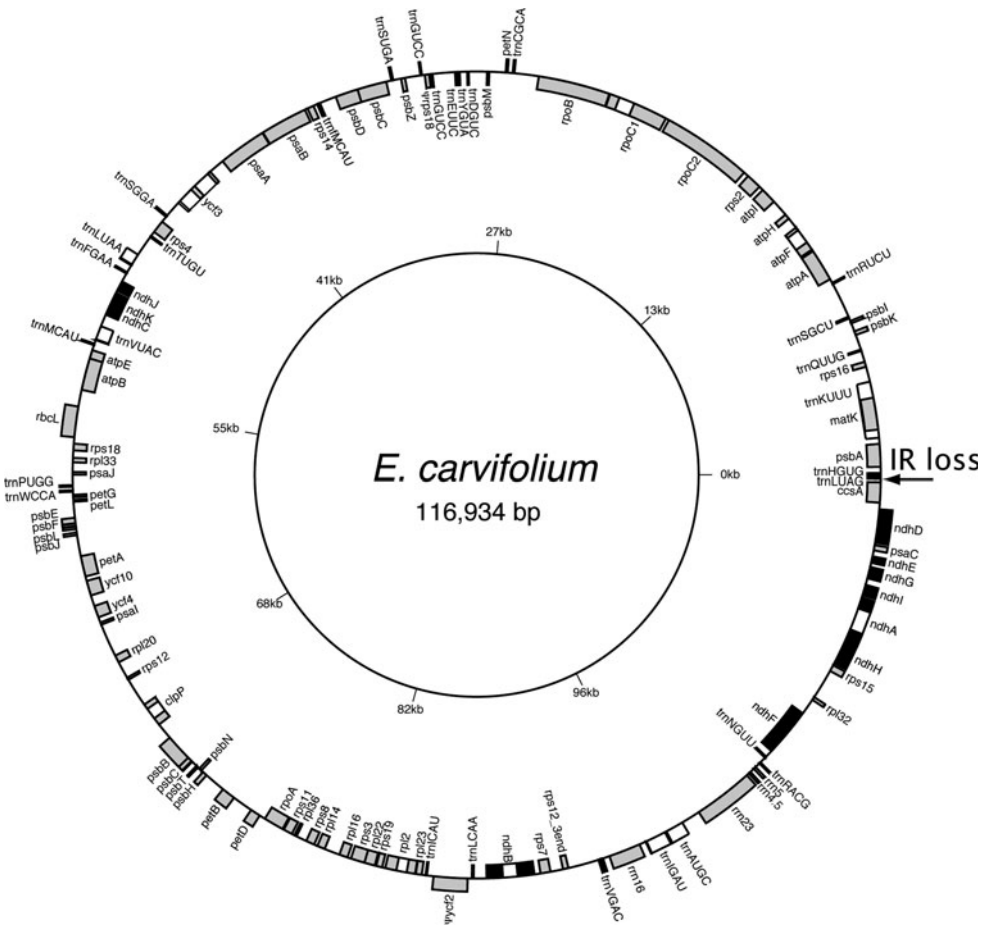

Fig. 2 Circularized gene map of the *E. carvifolium* plastid genome. Genes on the outside of map are transcribed in the counterclockwise direction and genes on the inside of the map are transcribed in the clockwise direction. *ndh* genes are in *black*; all others in *gray*. The *arrow* indicates IR deletion location

Table 1 Comparison of *E. texanum* (clade I) and *E. carvifolium* (clade II) plastid genomes

Genome characteristic	*E. texanum*	*E. carvifolium*
Size (bp)	130,812	116,934
Number of different protein-coding genes	75	76
Number of different tRNA genes	27	28
Number of different rRNA genes	4	4
Number of genes with introns	14	14
GC content	39.50%	39%

contributing to the 14 kb difference in genome size despite nearly identical gene content.

The IR adjacent to the *trnH*-gug—*psbA* LSC junction has been cleanly deleted in *E. carvifolium*, leaving only 233 bp between *trnH*-gug and *trnI*-cau, formerly LSC and SSC junctions, respectively. Comparison of *E. texanum* and *E. carvifolium* reveals that movement of the IR boundary prior to its loss is not the cause of rearrangement in *Erodium*. The two inversions present in *E. carvifolium* are present in genomes of all four major genera of Geraniaceae, and the endpoints are associated with the loss of *trnT*-ggu and *trnG*-gcc (Guisinger et al. 2011). The tandem duplication of *trnfM*-cau between the rearranged *psaA/B psbC/D* units is also found in *Geranium*. The three additional duplications of *trnfM*-cau found in *E. texanum* are absent from *E. carvifolium*.

The only unique area of rearrangement in *E. carvifolium* is located between the rearranged *psaC-D* genes and a conserved cluster of three tRNAs (*trnE*-ucc/*trnY*-gua/*trnD*-guc). This intergenic region contains a pseudogene of *rps18* flanked by two intact copies of *trnG-ucc*. The *rps18* pseudogene shows 76% pairwise identity with the intact gene and contains eight premature stop codons. The two copies of *trnG*-ucc are located on opposite strands. Duplication of *rps18* and *trnG*-ucc are not associated with further genomic rearrangement.

Loss of plastid-encoded *ndh* genes in *Erodium*

The complete plastomes of *E. texanum* and *E. carvifolium*, representing the two major clades of *Erodium*, both contain intact copies of all 11 plastid-encoded *ndh* genes (Fig. 1). The draft genome of *E. chrysanthum*, however, was suggested to lack intact copies of these genes (Guisinger et al. 2008). To verify this unusual loss and determine its

phylogenetic distribution, two additional species were chosen for plastome sequencing from the LBC (Fig. 1): *E. guicciardii* and *E. gruinum*.

No intact open reading frames (ORFs) were found for *ndh* genes in *E. chrysanthum*, *E. guicciardii*, or *E. gruinum*. For larger *ndh* subunits (e.g., *ndhF*, *ndhD*, *ndhJ*, *ndhA*, and *ndhB*) pseudogenes were found by BLAST searches (accession numbers HQ730936–HQ730953), using intact *ndh* genes from *E. texanum*. These larger subunits could be aligned with confidence against intact genes from *E. texanum* and E. *carvifolium*, revealing that *ndh* pseudogenes retain high sequence identity to intact genes (~90% pairwise identity) despite disruption by copious indels inducing frameshifts (Table 2). For other, primarily shorter *ndh* genes, pseudogenes could not be aligned with confidence with intact *ndh* genes. Only short BLAST hits (<50 bp) were obtained when queried with intact genes from *E. texanum* (Table 3). With few exceptions, the same genes are unalignable or completely degraded in the three draft genome assemblies of *E. chrysanthum*, *E. guicciardii* and *E. gruinum*. *Erodium gruinum* contains longer pseudogenes for a few genes degraded in *E. chrysanthum* and *E. guicciardii* (Table 3); this is consistent with the sister relationship between *E. chrysanthum* and *E. guicciardii* (Fig. 1).

Analysis of indel events in *ndh* pseudogenes reveals a remarkably consistent pattern (Table 2): the great majority of indel events are deletions (86%), and a large proportion of these events (74%) is shared among all three taxa. To determine the extent of *ndh* gene loss in *Erodium*, we sequenced a disrupted, deletion-rich region of *ψndhD* for all 13 species in the LBC (Fig. 1). The sequenced region contained the same four deletions, three of which cause frameshifts, in all 13 taxa (Fig. 3a, b), indicating that all 13 members of this clade share the dispersed deletions associated with *ndh* gene loss (accession numbers HQ730926–HQ730935).

Two genes formerly co-transcribed with *ndh* genes in an operon, *rps15* and *psaC*, are intact in *Erodium* genomes lacking *ndh* genes. *psaC* encodes a subunit of photosystem I and *rps15* encodes a ribosomal protein. The 5′ end of *rps15* is variable in *Erodium*, and there are two indels unique to those taxa also lacking *ndh* genes (accession numbers HQ730922-HQ730923). However, *psaC* is highly conserved and shows no indels or nonsynonymous substitutions in the *Erodium* taxa examined (accession numbers HQ730924–HQ730925).

Discussion

Plastome organization in *Erodium*

Despite encoding almost the same number of genes, genome organization is strikingly different in *E. texanum* (clade I) and *E. carvifolium* (clade II). The four published Geraniaceae plastomes representing each major genus, *P.* × *hortorum*, *M. speciosa*, G. *palmatum*, and *E. texanum* are all highly rearranged and show extreme variation in IR size, from 76 kb in *Pelargonium* to IR loss in *Erodium*. In addition to inversions, expansion and contraction of the IR has been proposed as a mechanism of genomic rearrangement (Chumley et al. 2006). However, comparison of *E. carvifolium* and the highly rearranged *E. texanum*, both lacking the IR, suggests that the deletion of the IR occurred before the divergence of the two major clades, and that no rearrangements occurred concurrent with the deletion event. In *E. carvifolium*, the IR deletion site lies between two tRNAs, leaving less than 250 bp between former single copy region junctions. The *E. carvifolium* genome, with

Table 2 Number of indels present in *ndh* pseudogenes that can be aligned with confidence with intact genes from *E. texanum* and *E. carvifolium*

	E. gruinum	*E. chrysanthum*	*E. guicciardii*	Shared indels
ndhA exon 1				
indels	2	2	2	2 of 2
ndhB exon 1				
indels	4	5	5	4 of 5
ndhB exon 2				
indels	3	4	3	3 of 4
ndhD				
indels	7	7	7	6 of 8
ndhF				
indels	17	13	14	13 of 17
ndhJ				
indels	1	3	1	1 of 3
				29/39 = 74%

Full-length pseudogenes were found for these coding sequences. Of the 39 indel events, 29 (74%) are shared among the three LBC taxa. The great majority (84%) are deletions

Table 3 Size of BLAST hits (in base pairs) returned for *ndh* pseudogenes that could not be aligned with confidence

Gene	Gene length (bp)	*E. gruinum* (bp)	*E. chrysanthum* (bp)	*E. guicciardii* (bp)
ndhC	363	351	26	30
ndhE	306	81	19	14
ndhG	540	23	22	22
ndhH	1,182	29	29	29
ndhI	516	47	21	55
ndhK	684	668	20	20

The gene length is from functional *E. texanum* sequences. The majority of BLAST hits are <50 bp, indicating substantial degradation. *Erodium gruinum* contains long, identifiable pseudogenes for two *ndh* genes degraded in *E. chrysanthum* and *E. guicciardii*. Short BLAST hits (20–30 bp long) are frequently spurious and are not necessarily pseudogenes

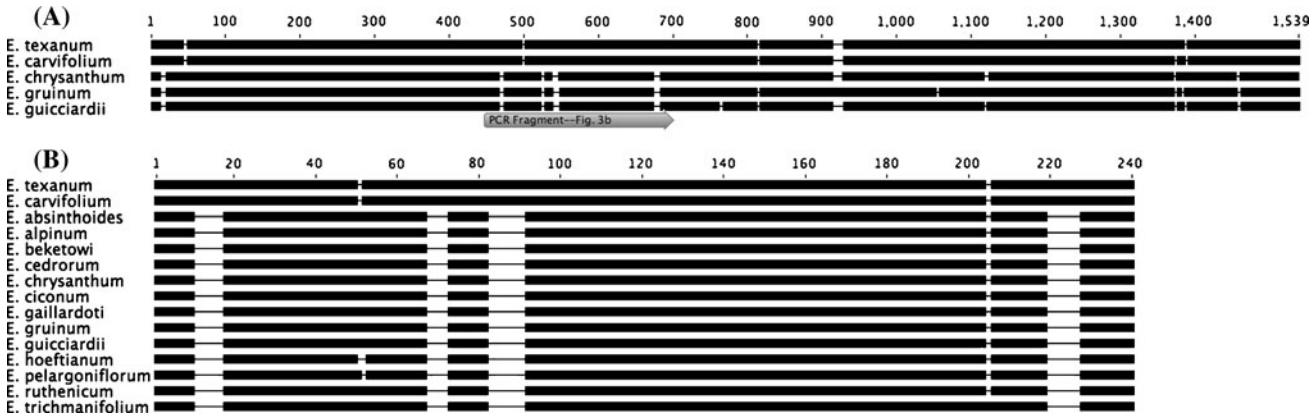

Fig. 3 **a** Alignment of *ndhD* regions from *C. macrophylla*, *E. carvifolium*, and *E. texanum* with *ψndhD* from 3 LBC taxa. The *ψndhD* region amplified for all 13 LBC taxa (Fig. 3b) is marked with an *arrow*. **b** Alignment of *ψndhD* fragments from all 13 LBC taxa against intact *ndhD* genes from *E. texanum* and *E. carvifolium*. The lengths of the shared deletions, in bp, from *left* to *right* are 7, 5, 9, and 8

a 'clean' deletion of the IR but no other associated rearrangements (despite multiple rearrangements in related species) is reminiscent of the situation in legumes, in *Medicago* relative to *Pisum* (Palmer et al. 1988).

The paucity of rearrangements in *E. carvifolium* is striking given the extensive rearrangement evident in other published Geraniaceae plastid genomes (Chumley et al. 2006; Guisinger et al. 2011). The two inversions that distinguish *E. carvifolium* from the ancestral angiosperm gene order are common to all four major genera and thus a synapomorphy of Geraniaceae. Associated with these inversions are the loss of *trnT*-ggu and the duplication of *trnfM*-cau, raising the possibility of illegitimate recombination between tRNAs, a mechanism that has been suggested in at least two other angiosperm groups (Hiratsuka et al. 1989; Haberle et al. 2008). The loss of an IR is a derived character in *Erodium*, and the lack of other unique rearrangements allows us to hypothesize an ancestral gene order for Geraniaceae that resembles *E. carvifolium* prior to the deletion of the IR. Having a model for the organization of the simplest possible Geraniaceae plastome should greatly assist in reconstructing the rearrangement history of each genus. This model suggests that the rearrangement in each genus has occurred independently, with the exception of two inversions shared by all genera. It also suggests that expansion and contraction of the IR was not a force in genomic rearrangement in all genera, because loss of the IR preceded any unique genomic rearrangements in *Erodium*. Finally, the conservation of the S10 operon in *E. carvifolium* suggests that its fragmentation in *E. texanum* and *G. palmatum* represents two independent events (Guisinger et al. 2011).

Phylogenetic distribution and timing of *ndh* gene loss

The loss of *ndh* genes in the long-branch clade of *Erodium* (Fig. 1) is the most recent and phylogenetically restricted among photosynthetic seed plants. The presence of intact *ndh* genes in both major clades of *Erodium* has been established with the publication of *E. texanum* (Guisinger et al. 2011) and in *E. carvifolium* in this paper; the absence of intact *ndh* genes has been confirmed in the long-branch

clade for all 13 species by plastid genome sequencing (3 taxa) or targeted PCR (10 taxa) (Fig. 3a, b). Except for a difference in chromosome number—$n = 8$ or 9 in the LBC compared with $n = 10$ in the rest of the genus—no morphology, life history or other trait appears to distinguish this clade from other members of the genus (Fiz et al. 2006). Indeed, homoplasy in leaf morphology caused the LBC to be grouped into the polyphyletic section *Absinthoidea* (Guittonneau 1990; Fiz et al. 2006). Alignment of pseudogenes from three LBC taxa with intact *ndh* genes from *E. texanum* and *E. carvifolium* reveals that the great majority of indels present in pseudogenes are shared by all three LBC taxa (74%). This consistency is a strong indication that loss of plastid-encoded *ndh* genes occurred prior to the diversification of the clade.

Parkinson et al. (2005) used the *cox1* mitochondrial marker to date the divergence of the major clades in Geraniaceae. The inferred divergence times for *Erodium* (from *Geranium*) and the LBC (from the rest of *Erodium* clade I) are at approximately 22 MYA and 7 MYA, respectively. Another estimate using plastid markers indicated slightly earlier divergence times for *Erodium* and the LBC (from the rest of clade I) at approximately 15.45 MYA and 5 MYA respectively (Fiz et al. 2008). These divergence time estimates are extremely recent compared with the age of the other two known losses of *ndh* genes from photosynthetic seed plants. The loss of *ndh* genes in Pinaceae/Gnetales, if indeed this represents a single loss and a synapomorphy for the clade proposed under the "gnepine" hypothesis (Bowe et al. 2000; Chaw et al. 2000), is ancient. Within gymnosperms, divergence of Pinaceae has been estimated at 140 MYA (Wang et al. 2000), and loss of *ndh* genes necessarily predates this divergence estimate if it represents a synapomorphy for Pinaceae and Gnetales (Braukmann et al. 2009).

Loss of *ndh* genes in orchids is likely also ancient compared to that in *Erodium*. An estimate based on fossil-calibrated molecular data suggests a crown radiation of Orchidaceae 76-84 MYA (Ramirez et al. 2007). The taxonomic group recently confirmed to have lost *ndh* genes encompasses 70 orchid genera and over 1,000 species, suggesting that the loss of *ndh* genes is not recent (Wu et al. 2010). Orchidaceae is among the largest angiosperm families with ~30,000 species (Atwood 1986), and the full extent of *ndh* gene loss in Orchidaceae has not been determined. For example, many members of the subfamily Epidendroideae (*Brassavola*, *Meiracyllium*, *Cattlea*, *Epidendrum*, *Encyclia*, *Stanhopea*, *Phalaenopsis*, and *Oncidium*), the largest subfamily with more than 10,000 species, have deletions in *ndhF* inducing frameshifts. These data suggest that the loss of plastid-encoded *ndh* genes may be a synapomorphy for much of Orchidaceae (Neyland and Urbatsch 1996; Chang et al. 2006; Wu et al. 2010).

Fate of *ndh* genes: loss, replacement, or transfer?

It is unknown whether Ndh function has been lost entirely in these three seed plant lineages—in gymnosperms, orchids, and Geraniaceae—or whether the genes have been functionally transferred to the nucleus or otherwise replaced. No common features such as parasitism, mycotrophism, or epiphytic lifestyle are associated with *ndh* gene loss in *Erodium*. This discovery of a recent loss of *ndh* genes in *Erodium* presents an opportunity to investigate changes in photosynthetic function through comparative biochemistry between *Erodium* species with and without plastid-encoded *ndh* genes.

The chlorophyll fluorescence assay used to detect *ndh* mutants in tobacco (Horváth et al. 2000) can be used as a first step to assess the presence or absence of Ndh function in LBC species. Detection of Ndh function in LBC taxa could indicate an unprecedented recent burst of gene transfer from the plastid genome to the nuclear genome. Functional replacement of the Ndh complex by a nuclear gene or genes is another possibility; for example, in grasses, the multisubunit acetyl-Co carboxylase complex encoded by the plastid gene *accD* and nuclear-encoded subunits was replaced by a single-subunit complex of bacterial origin (Konishi et al. 1996). A failure to detect Ndh function in LBC taxa will suggest that regulation of electron flow between the two photosystems occurs through an alternative mechanism. Experimental data indicates that the Ndh complex is indispensable to the photosynthetic stress response (Endo et al. 1999; Martin et al. 2004; Rumeau et al. 2007; Casano et al. 2001). However, the consistent pattern of deletions found among LBC *ndh* pseudogenes suggests that loss of gene function occurred prior to the diversification of the clade >5MYA, and the subsequent diversification and persistence of this clade suggest that photosynthetic stress response has not been greatly compromised. Because the electron donor and functions of the Ndh complex are still unresolved it is difficult to evaluate the relative likelihood of gene loss, functional replacement, and functional gene transfer of *ndh* genes in *Erodium*.

Conclusion

The causal relationships among the evolutionary rate acceleration, indels, and loss of *ndh* genes in the LBC remain unclear. It is possible that the loss or functional transfer of *ndh* genes was promoted by the acceleration in the rate of evolution, if the rate in the plastome surpassed that in the nucleus, as has been recently suggested by Magee et al. (2010) for a hypervariable region in legume plastomes. A full evolutionary rate analysis of the genes

from each functional class across eight additional *Erodium* species and monotypic sister species *California macrophylla* is underway to disentangle the relative roles of evolutionary rate accelerations, gene losses, indel frequency, and genomic rearrangements in *Erodium* plastome evolution. Finally, the recentness and restricted distribution of *ndh* gene loss in *Erodium* raises the possibility of finding additional recent losses among angiosperms, perhaps within lineages known to display plastome rearrangements.

Acknowledgments Support for this work was provided by grants from NSF DEB-0717372 to RKJ and a NSF GRF predoctoral fellowship to JCB. We thank Juan José Aldasoro for providing DNAs of selected *Erodium* species and Tracey Ruhlman, Katie Hansen, and Mao-Lun Weng for critical comments on an earlier version of this manuscript.

References

Atwood JT (1986) The size of the Orchidaceae and the systematic distribution of epiphytic orchids. Selbyana 9:171–186

Bock R (2007) Structure, function, and inheritance of plastid genomes. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Berlin, pp 29–63

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

Bowe LM, Coat G, dePamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc Natl Acad Sci USA 97:4092–4097

Braukmann TWA, Kuzmina M, Stefanovic S (2009) Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Curr Genet 55:323–337

Casano LM, Martin M, Sabater B (2001) Hydrogen peroxide mediates the induction of chloroplast Ndh complex under photooxidative stress in barley. Plant Physiol 125:1450–1458

Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SH, Chang CC, Chaw SM (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. Mol Biol Evol 23:279–291

Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD (2000) Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc Natl Acad Sci USA 97:4086–4091

Chevreux B (2009) MIRA: an automated genome and EST assembler. Available at http://chevreux.org/thesis/index.html. Accessed 22 Aug 2009

Chumley TM, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

Conant GC, Wolfe KH (2007) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24:861–862

Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. Gen Res 11:1095–1099

Drummond AJ, Ashton B, Buxton S, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A (2009) Geneious v4.8. Available from http://www.geneious.com

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res 32:1792–1797

Endo T, Shikanai T, Takabayashi A, Asada K, Sato F (1999) The role of chloroplastid NAD(P)H dehydrogenase in photoprotection. FEBS Lett 457:5–8

Fiz O, Vargas P, Alarcón ML, Aldasoro JJ (2006) Phylogenetic relationships and evolution in *Erodium* (Geraniaceae) based on *trnL-trnF* sequences. Syst Bot 31:739–763

Fiz O, Vargas P, Alarcón ML, Aedo C, Garcia JL, Aldasoro JJ (2008) Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. Syst Bot 33:326–342

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc Natl Acad Sci USA 105:18424–18429

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28(1):583–600

Guittonneau GG (1990) Taxonomy, ecology, and phylogeny of genus *Erodium* l'Her. in the Mediterranean region. In: Vorster P (ed) Proceedings of international Geraniaceae symposium, University of Stellenbosch, Sept., pp 69–91

Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol 66:350–361

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C, Meng B, Li Y, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Horváth EM, Peter SO, Joët T, Rumeau D, Cournac L, Horváth GV, Kavanaugh TA, Schäfer C, Peltier G, Medgyesy P (2000) Targeted inactivation of the plastid *ndhB* gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. Plant Physiol 123:1337–1350

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Meth 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, 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

Konishi T, Shinohara K, Yamada K, Sasaki Y (1996) Acetyl-CoA carboxylase in higher plants: most plants other than Gramineae have both the prokaryotic and eukaryotic forms of this enzyme. Plant Cell Physiol 37:117–122

Magee AM, Aspinall S, Rice DW, Cusak BP, Sémon M, Perry AS, Stefanovic S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res 20:1700–1710

Martin M, Sabater B (2010) Plastid *ndh* genes in plant evolution. Plant Physiol and Biochem 48:636–645

Martin M, Casano LM, Zapata JM, Guéra A, Del Campo EM, Schmitz-Linneweber C, Maier RM, Sabater B (2004) Role of thylakoid Ndh complex and peroxidase in the protection against photo-oxidative stress: fluorescence and enzyme activities in wild-type and *ndhF*-deficient tobacco. Physiol Plant 122:442–452

McCoy SR, Kuehl JV, Boore JL, Raubeson LA (2008) The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evol Biol 8:130

Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanaugh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe KH (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 13:645–658

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

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

Palmer JD, Osorio B, Thomson WF (1988) Evolutionary significance of inversions in legume chloroplast DNA. Curr Genet 14:65–74

Parkinson CL, Mower JP, Qiu Y, Shirk AJ, Song K, Young ND, dePamphilis CW, Palmer JD (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol 5:73

Ramirez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE (2007) Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. Nature 448:1042–1045

Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry R (ed) Diversity and evolution of plants-genotypic and phenotypic variation in higher plants. CABI Publishing, Wallingford, pp 45–68

Rumeau D, Bécuwe-Linka N, Beyly A, Louwagie M, Garin J, Peltier G (2005) New subunits NDH-M, -N, and -O, encoded by nuclear genes, are essential for plastid Ndh complex functioning in higher plants. Plant Cell 17:219–232

Rumeau D, Peltier G, Cornac L (2007) Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant Cell Environ 30:1041–1051

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci USA 91:9794–9798

Wang X, Tank DC, Sang T (2000) Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol 17:773–781

Wu F, Chan M, Liao D, Hsu C, Lee Y, Daniell H, Duvall MR, Lin C (2010) Complete chloroplast genome of *Oncidium* Gower Ramsey and evaluation of molecular markers for identification and breeding of Oncidiinae. BMC Plant Bio 10:68

Wu SW, Lai YT, Lin CP, Wang YN, Chaw SM (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. Mol Phylo Evol 52:115–124

Wyman SK, Boore JL, Jansen RK (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Zhong B, Yonezawa T, Zhong Y, Hasegawa M (2010) The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. Mol Biol Evol. doi:10.1093/molbev/msq170