Mobile Elements Shape Plastome Evolution in Ferns

Tanner A. Robison[1,*], Amanda L. Grusz[2,3], Paul G. Wolf[1], Jeffrey P. Mower[4], Blake D. Fauskee[2], Karla Sosa[5], and Eric Schuettpelz[3]

[1]Department of Biology, Utah State University

[2]Department of Biology, University of Minnesota Duluth

[3]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, District of Colombia

[4]Department of Agronomy, Center for Plant Science Innovation, University of Nebraska

[5]Department of Biology, Duke University

*Corresponding author: E-mail: robison.tanner@gmail.com.

Accepted: August 23, 2018

Data deposition: This project has been deposited at NCBI GenBank under the accessions MG953517 and MH173066—MH173091.

Abstract

Plastid genomes display remarkable organizational stability over evolutionary time. From green algae to angiosperms, most plastid genomes are largely collinear, with only a few cases of inversion, gene loss, or, in extremely rare cases, gene addition. These plastome insertions are mostly clade-specific and are typically of nuclear or mitochondrial origin. Here, we expand on these findings and present the first family-level survey of plastome evolution in ferns, revealing a novel suite of dynamic mobile elements. Comparative plastome analyses of the Pteridaceae expose several mobile open reading frames that vary in sequence length, insertion site, and configuration among sampled taxa. Even between close relatives, the presence and location of these elements is widely variable when viewed in a phylogenetic context. We characterize these elements and refer to them collectively as Mobile Open Reading Frames in Fern Organelles (MORFFO). We further note that the presence of MORFFO is not restricted to Pteridaceae, but is found across ferns and other plant clades. MORFFO elements are regularly associated with inversions, intergenic expansions, and changes to the inverted repeats. They likewise appear to be present in mitochondrial and nuclear genomes of ferns, indicating that they can move between genomic compartments with relative ease. The origins and functions of these mobile elements are unknown, but MORFFO appears to be a major driver of structural genome evolution in the plastomes of ferns, and possibly other groups of plants.

Key words: fern, horizontal gene transfer, inversion, organelle, Pteridaceae.

Introduction

Plastid genomes (plastomes) are a rich source of molecular sequence data and have proved to be especially useful in explorations of plant evolutionary history. From single-gene analyses to full plastome phylogenomics, important evolutionary insights can be gleaned from these relatively small, highly conserved, and minimally repetitive chromosomes (Taberlet et al. 1991; Chaw et al. 2004; Pryer et al. 2004; Yoon et al. 2004; Shaw et al. 2005; Givnish et al. 2010; Moore et al. 2010; Leliaert et al. 2012; Ruhfel et al. 2014; Gitzendanner et al. 2018). Plastomes contain high proportions of protein coding genes compared with plant nuclear genomes, with many of these genes being essential to photosynthesis (Wicke et al. 2011). Consequently, plastomes experience relatively low nucleotide substitution rates, especially in the inverted repeats (IRs), making them extremely stable over evolutionary time (Wolfe et al. 1987; Li et al. 2016; Zhu et al. 2016).

The plastomes of land plants seem to be especially resistant to changes in gene content, which, along with gene order, generally varies little between distantly related lineages—even after hundreds of millions of years (Palmer 1985). Relatively few plastome genes have been lost, except in heterotrophic lineages in which photosynthetic genes are typically not required for survival (Bungard 2004). Even rarer is the acquisition of new genes (Timmis et al. 2004). Since the gain of *ycf1* and *ycf2* in the algal ancestors, very few new genes have been incorporated into land plant plastomes (Timmis et al. 2004; de Vries et al. 2015). However, some groups do show exceptional variability in plastome structure, even among closely

related taxa (Chumley et al. 2006; Cai et al. 2008; Haberle et al. 2008; Hirao et al. 2008; Guisinger et al. 2011). Notable among these exceptional lineages is Campanulaceae, within which a prolific group of inserted open reading frames (ORFs) appear to have driven over 125 large inversions across the family (Knox 2014).

Overall stability in plastome structure across land plants contrasts, strikingly, with punctuated and/or persistent genomic rearrangements that are apparent in certain lineages (Chumley et al. 2006; Cai et al. 2008; Haberle et al. 2008; Hirao et al. 2008; Guisinger et al. 2011; Mower and Vickrey 2018). Ferns are among these, showing evidence of multiple genomic inversions since their initial diversification (Stein et al. 1992; Wolf et al. 2015; Zhu et al. 2016; Labiak and Karol 2017). An increasing number of fern plastome sequences are beginning to reveal a dynamic organellar genome, shaped in large part by structural inversions and/or shifts in gene content from single copy regions into the IR (Wolf et al. 2010). Genomic inversions like these have been associated with shifts in molecular evolutionary rate (Blazier et al. 2016; Li et al. 2016; Zhu et al. 2016) and may be moderated by selective constraints related to gene synteny and gene expression (Cui et al. 2006; Wicke et al. 2011).

Despite the great strides that have been made in our understanding of plastome evolution in ferns over the last decade (Mower and Vickrey 2018), dense taxonomic sampling is almost always lacking. Instead, studies of plastome evolution in ferns have focused on disparate groups of deeply divergent taxa (Gao et al. 2011; Wei et al. 2017), or on comparisons of only a few closely related species (Labiak and Karol 2017). Here, we aim to bridge this gap with the first family-scale comparative analysis of plastome structure and content in ferns.

Our focus is on Pteridaceae, an early-diverging family of polypod ferns that comprises roughly 1,200 species and accounts for well over 10% of extant fern diversity (PPG I 2016). The family is cosmopolitan in distribution and occupies a wide array of niches, from shaded forests to xeric and even aquatic habitats (Tryon 1990). Members exhibit a range of reproductive modes and some groups are noteworthy for undergoing frequent whole genome duplication. Among the most striking evolutionary patterns in the family is a dramatic shift in molecular evolutionary rate that has been documented across plastid and nuclear genomes of the so-called "vittarioid" ferns (Rothfels and Schuettpelz 2014; Grusz et al. 2016). In this study, we leverage genome skimming data to assemble and analyze 27 new plastomes from across the Pteridaceae in an effort to: 1) examine plastome variation in Pteridaceae; 2) gain insight into genomic shifts within members of the Vittarioideae; and 3) reevaluate the phylogenetic relationships among major lineages comprising the family. Our data expose a massive plastome inversion and a group of mobile elements—newly characterized here—that appear to be a particularly dynamic component of fern plastomes, as evidenced from within Pteridaceae and beyond.

Materials and Methods

Taxon Sampling

Taxonomic sampling included 29 ingroup species, representing all major clades within the Pteridaceae (Schuettpelz et al. 2007), as well as three outgroup taxa (Table 1). Increased sampling from within subfamily Vittarioideae was undertaken in an effort to better understand the molecular evolutionary rate heterogeneity between the two main subclades therein, *Adiantum* and the vittarioid ferns.

DNA Extraction, Library Prep, and Sequencing

Whole genomic DNA for all newly sampled ingroup taxa (27 total) was extracted from silica-dried leaf tissue using the Qiagen DNeasy Plant Mini Kit (Germantown, Maryland) following the manufacturer's protocol. Whole genomic DNA for 26 samples (all except *Vittaria appalachiana*) was sent to the Duke University Center for Genomic and Computational Biology for in-house library preparation and sequencing. There, individual genomic libraries (~300 bp) were prepared using the Kappa Hyper Prep Kit (Wilmington, Massachusetts). In total, 32 samples (26 included in this study) were multiplexed and pooled over one complete flowcell (8 lanes) on the Illumina HiSeq 2000/2500 platform for 125 bp paired-end sequencing. The *V. appalachiana* DNA was sent to BGI (Shenzhen, China) and sequenced on the Illumina HiSeq 2000 platform, generating 5 Gb of 100 bp paired-end reads from an ~800 bp library.

Genome Assembly and Annotation

Raw sequence reads from *V. appalachiana* were assembled using Velvet version 1.2.03 (Zerbino and Birney 2008) according to previously described procedures (Guo et al. 2014; Sigmon et al. 2017). Genome assembly for all other ingroup taxa was performed using NOVOPlasty (Dierckxsens et al. 2017). NOVOPlasty implements a seed-based, de novo genome assembly, which can lessen structural assembly biases that may otherwise mask inferred rearrangements. NOVOPlasty employs the seed to retrieve a given sequence from the target genome, which is then extended and circularized (Dierckxsens et al. 2017).

In most cases, the *rbcL* gene from *Adiantum capillus-veneris* (NC_004766; Table 1) was used as a seed sequence, but in select cases, if *rbcL* had inadequate coverage, the entire plastome of *A. capillus-veneris* was used instead. Raw, unfiltered Illumina reads were subsampled to ≤30 million reads to reduce memory requirements. The default k-mer of 39 was used unless there was low organellar genome coverage (<1%), in which case the k-mer was reduced to 23–30. In

Table 1
Taxonomic Sampling and Voucher Information for Samples Used in This Study

Taxon	Voucher or Citation	Genbank Accession
Adiantum aleuticum (Rupr.) C. A. Paris	Rothfels 4097 (DUKE)	MH173079
Adiantum capillus-veneris L.	Wolf et al. (2004)	NC004766
Adiantum tricholepis Fée	Rothfels 08-094 (DUKE)	MH173071
Antrophyum semicostatum Blume	Schuettpelz 1561 (BO)	MH173087
Bommeria hispida (Mett. ex Kuhn) Underw.	Beck 1130 (DUKE)	MH173074
Calciphilopteris ludens (Wall. ex Hook.) Yesilyurt & H. Schneid.	Huiet s.n. (DUKE)	MH173084
Ceratopteris cornuta (P. Beauv.) Lepr.	Rothfels 4298 (DUKE)	MH173082
Ceratopteris richardii Brongn.	Marchant et al. unpub.	KM052729
Cheilanthes bolborrhiza Mickel & Beitel	Rothfels 3294 (DUKE)	MH173073
Cheilanthes micropteris Sw.	Prado 2132 (DUKE)	MH173078
Cryptogramma acrostichoides R. Br.	Rothfels 4195 (DUKE)	MH173081
Cystopteris chinensis (Ching) X. C. Zhang & R. Wei	Wei et al. (2017)	KY427337
Dryopteris decipiens (Hook.) Kunze	Wei et al. (2017)	KY427348
Gastoniella chaerophylla (Desv.) Li Bing Zhang & Liang Zhang	Prado 2178 (SP)	MH173080
Haplopteris elongata (Sw.) E. H. Crane	Schuettpelz 1559 (BO)	MH173086
Hemionitis subcordata (D. C. Eaton ex Davenp.) Mickel	Rothfels 3163 (DUKE)	MH173072
Jamesonia brasiliensis Christ	Schuettpelz 1444 (SP)	MH173077
Llavea cordifolia Lag.	Schuettpelz 1744 (US)	MH173088
Myriopteris covillei (Maxon) Á. Löve & D. Löve	Schuettpelz 443 (DUKE)	MG953517
Myriopteris lindheimeri (Hook.) J. Sm.	Schuettpelz 450 (DUKE)	NC014592
Myriopteris scabra (C. Chr.) Grusz & Windham	Windham 3495 (DUKE)	MH173083
Notholaena standleyi Maxon	Schuettpelz 435 (DUKE)	MH173067
Onychium japonicum (Thunb.) Kunze	Schuettpelz 1057 (DUKE)	MH173069
Pellaea truncata Goodd.	Schuettpelz 430 (DUKE)	MH173066
Pentagramma triangularis (Kaulf.) Yatsk., Windham & E. Wollenw.	Schuettpelz 1332 (DUKE)	MH173070
Pityrogramma trifoliata (L.) R. M. Tryon	Rothfels 3658 (DUKE)	MH173075
Pteridium aquilinum (L.) Koon.	Der et al. unpub.	NC014348
Pteris vittata L.	Schuettpelz 893 (DUKE)	MH173068
Scoliosorus ensiformis (Hook.) T. Moore	Schuettpelz 1782 (US)	MH173090
Tryonia myriophylla (Sw.) Schuettp., J. Prado & A. T. Cochran	Schuettpelz 1434 (SP)	MH173076
Vaginularia trichoidea Fée	Schuettpelz 1553 (BO)	MH173085
Vittaria appalachiana Farrar & Mickel	Stevens OH-p1-s11 (PUR)	MH173091

cases involving long repetitive regions, higher k-mers of 45–55 were used. For *Jamesonia brasiliensis* and *Cheilanthes bolborrhiza*, complete assembly of plastomes was not possible, but we were able to get contigs of considerable size, which have been included in this study.

Following assembly, genomes were annotated in Geneious 11.1 (Kearse et al. 2012), using the gene sequences of *A. capillus-veneris* as a reference. Putative RNA editing sites were annotated to retain conserved ORFs (Wolf et al. 2003, 2004). Intergenic sequences that differed dramatically from *A. capillus-veneris* were queried against the NCBI Nucleotide database using BLAST (Altschul et al. 1990) to ensure that they did not result from the false assembly of mitochondrial or nuclear sequences. Assembly errors were further assessed by mapping raw reads to the newly assembled genome using Bowtie2 (Langmead and Salzberg 2012), looking for dips in read depth. Additionally, overall plastome assembly quality was assessed for each sample using Pilon (Walker et al. 2014). In all cases, changes proposed by Pilon were relatively minor (<10 nucleotides) suggesting that the quality and accuracy from NOVOPlasty assembly was high.

Plastome Phylogenomic Analyses

In total, 32 plastomes were included in our phylogenomic analyses; 26 were new to this study and the remaining 6 were obtained from Genbank (Table 1). Each annotated plastome was opened in Geneious and all CDS/gene regions were extracted in FASTA format. Using these taxon-specific FASTA files, containing all CDS/gene regions, we then compiled a FASTA file for each locus. Sequences for each region were aligned using MAFFT 7.394 (Kuraku et al. 2013; Katoh et al. 2017) and alignments of all loci concatenated using Sequence Matrix (Vaidya et al. 2011). The resulting concatenated matrix, comprising 76 loci and 68,047 nucleotide sites, partitioned by gene, was processed through PartitionFinder2 (Lanfear et al. 2012, 2017) on the CIPRES Science Gateway version 3.3 (Miller et al. 2010) using the following settings:

Branchlengths = linked, models = GTR, GTR + G, GTR + I + G, and model_selection = AICc. PartitionFinder2 identified 36 unique model partitions spanning the concatenated matrix.

Phylogenomic analyses of the concatenated, partitioned data set were implemented using maximum likelihood (ML) and Bayesian optimality criteria on the CIPRES Science Gateway version 3.3 (Miller et al. 2010). ML searches were conducted in RAxML V.8 using multiparametric bootstrapping (−b; 1,000 replicates) and our previously described partitioned model (−q). Bayesian inference was performed using MrBayes 3.2.6 and comprised four independent runs, each with four chains (one cold, three heated) and otherwise default (i.e., flat) priors, with the exception that rates of evolution were allowed to vary among loci (ratepr = variable). Chains were run for 10 million generations and trees were sampled from the cold chain every 1,000 generations. To determine at which point the analysis had reached stationarity, the standard deviation of split frequencies among the independent runs (as calculated by MrBayes) was examined and the output parameter estimates were plotted using Tracer v1.6 (Rambaut et al. 2014). Based on convergence diagnostics, the first 2.5 million generations were excluded before obtaining a consensus phylogeny and clade posterior probabilities with the "sumt" command (contype = allcompat).

Characterizing MORFFO Elements

To search for MORFFO-like sequences in GenBank, all MORFFO insertions found in Pteridaceae were aligned using Geneious, and a consensus sequence was generated. All BLAST (Altschul et al. 1990; Madden 2013) queries for MORFFO were performed using this consensus sequence. This consensus sequence was also used when querying both VecScreen (NCBI Resource Coordinators 2017) and RepeatMasker (Smit et al. 2013). Three main search strategies were employed when using NCBI Nucleotide BLAST. Initial queries for MORFFO-like sequences were performed with BLASTN using the default parameters. Then, BLASTX and TBLASTN searches were performed using the default parameters and a word size of 3. Additional searches using specific MORFFO sequences rather than a consensus returned equivalent results.

To evaluate the level of selective constraint on the three MORFFO genes, estimates of d_N/d_S were calculated for five species that contained all three MORFFO genes (*Bommeria hispida*, *Hemionitis subcordata*, *Notholaena standleyi*, *Tryonia myriophylla*, and *Vaginularia trichoidea*). First, codon-based alignments were generated using the ClustalW-Codons option in MEGA version 7.0.18 (Kumar et al. 2016). Alignments were trimmed using Gblocks version 0.91b (Castresana 2000) in codon mode with a relaxed set of parameters ($t = c$, $b2 = 3$, $b5 =$ half). For each trimmed gene alignment, branch wise estimates of d_N/d_S were calculated for each species using the GA-branch model (Kosakovsky Pond and Frost 2005), which uses a genetic algorithm to optimize the number of d_N/d_S rate classes across the tree and ML to optimize branch lengths and substitution rates, as implemented on the Datamonkey web server (Delport et al. 2010). For the analysis, the HKY rate matrix was chosen as the substitution rate model based on the Datamonkey model selection tool. To evaluate the influence of the tree topology on d_N/d_S estimates, the analysis was run using either a NJ tree or a user-defined tree that matched organismal relationships shown in figure 1.

Phylogenetic relationships among MORFFO sequences were estimated using ML best tree and bootstrap searches, implemented in RAxML V.8 (Stamatakis 2014) with multiparametric bootstrapping (−b; 1,000 replicates).

Results

Genome Assembly and Annotation

We assembled and annotated 25 complete plastomes from previously unsampled species, representing all major clades within Pteridaceae (Schuettpelz et al. 2007), plus 2 partial plastome sequences for *J. brasiliensis* and *C. bolborrhiza* (139,531 and 39,380 bp, respectively). The average length of complete plastid genome sequences was 153,153 bp (range 145,327–165,631 bp) with an average GC content of 41.46% (range 36.7%–45.3%; Table 2). Gene content remained largely stable across samples, with no losses of protein coding genes relative to *A. capillus-veneris*. We did, however, detect a loss of *trnT* in all vittarioid ferns sampled (Fig. 1), as well as a loss of *trnV* in *Onychium japonicum, Ceratopteris cornuta*, plus all vittarioids with the exception of *Haplopteris elongata*. Across all samples, there were 82 protein coding genes, 33–35 tRNA genes and 4 rRNA genes. Gene order was unchanged across the family—with the exception of a 7,000 bp genomic inversion within the IRs of all vittarioid species except *V. trichoidea*.

Several plastid DNA insertions were recovered from multiple clades within Pteridaceae. The most prominent of these comprised a suite of genomic insertions, here referred to as Mobile Open Reading Frames in Fern Organelles (MORFFO), that were detected in most of the plastomes sampled. These MORFFO clusters are characterized by three large and distinct ORFs that are variably absent, or present in a number of different arrangements (fig. 1). One ~1,300 bp ORF (*morffo1*) is flanked by inverted repeat (IRs) of ~40 bp that are often in the motif TGT CGA TAG, repeated 3–5 times. The amino acid sequences of *morffo1* do not bear similarity to any characterized proteins in GenBank, but do bear similarity to a hypothetical protein found in the fern *Mankuya chejuensis* and the green alga *Roya anglica* (table 3). A larger ORF (*morffo2)* of ~1,700 bp has domains similar to primases associated with mobile

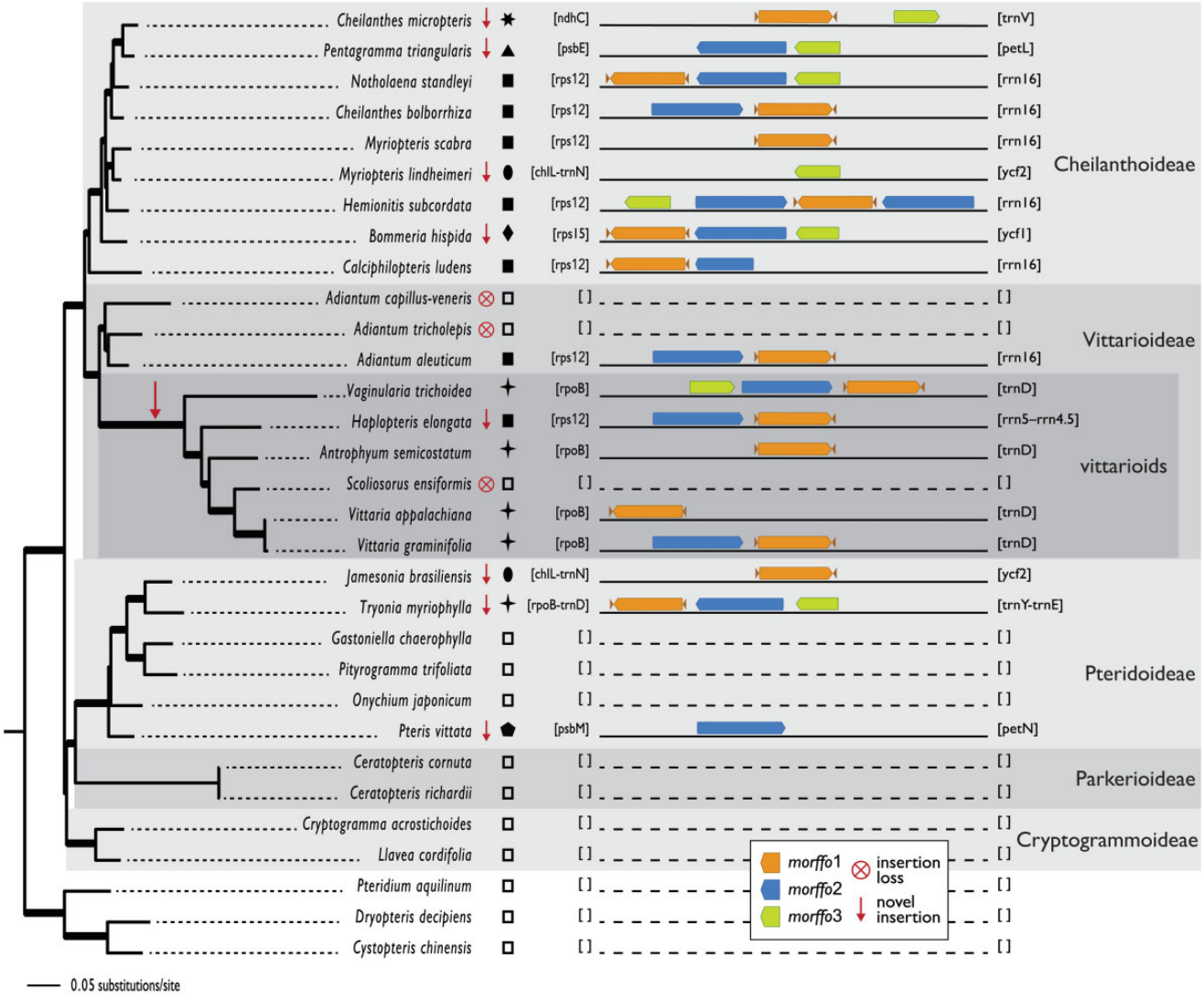

Fig. 1.—Distribution of MORFFO elements across the Pteridaceae phylogeny. Topology results from ML analysis of plastome data (−ln *L*=609991.403586); thickened branches indicate bootstrap/posterior probability support=100/1.0. Symbols highlight shared insertion sites, with empty squares signifying evident lack of a MORFFO insertion. Short arrows flanking *morffo1* indicate short inverted repeats. Novel insertions and losses, as inferred by maximum parsimony, are depicted as arrows or crossed-out circles, respectively.

elements in cyanobacteria and archaea (DN_5 superfamily) when queried using BLASTX (Altschul et al. 1990; table 3). A smaller ORF (*morffo3*) of ~630 bp has no significant similarity to any known genes or proteins; it is found less frequently than the 2 larger ORFs, but is still prevalent. Often, but not always, *morffo1* is found inserted in frame with *morffo2* to form a larger ORF of ~3,500 bp. Importantly, *morffo1, morffo2,* and *morffo3* can be found in a variety of different arrangements, but when present they are always found immediately adjacent to one another (fig. 1).

The location of MORFFO elements varied across the genomes sampled, being included in the large single copy (LSC), the IR, or the small single copy (SSC) regions (fig. 2). As a whole, MORFFO sequences (*morffo1*, *morffo2*, and *morffo3*) were similar among species, ranging from 92% to 45% sequence identity (table 4). For those species with a full set of three MORFFO sequences, the genes appear to encode functional proteins. The coding sequences are intact (no internal stop codons or frameshifting indels) and exhibit nonsynonymous (d_N) to synonymous (d_S) substitution rate ratios that are consistent with selective constraint ($d_N/d_S \ll 1$, ranging from 0.17 to 0.43), with the exception of *morffo1* from *T. myriophylla,* which has at least two frameshifting indels and a d_N/d_S approaching 1 (supplementary table S1, Supplementary Material online). A weak association was observed between the genomic location of

Table 2
Summary of Basic Genomic Features of Plastomes Used in This Study

Species	IR Size (bp)	LSC Size (bp)	Genome Size (bp)	SSC Size (bp)	%GC	*MORFFO*
Adiantum aleuticum	26,289	83,345	157,519	21,596	45	1, 2
Adiantum capillus-veneris	23,448	82,282	150,568	21,390	41	Absent
Adiantum tricholepis	23,233	82,740	150,667	21,461	42	Absent
Antrophyum semicostatum	20,977	87,492	150,274	20,828	40	1, 3
Bommeria hispida	23,142	82,491	156,749	27,974	43	1, 2, 3
Calciphilopteris ludens	26,585	82,423	157,068	21,475	43	1, 2
Ceratopteris cornuta	22,287	83,623	149,424	21,227	37	Absent
Ceratopteris richardii	22,020	83,178	148,444	21,226	35	Absent
Cheilanthes bolborrhiza	~25,000	na	39,380*	Na	44	1, 2
Cheilanthes micropteris	23,306	88,393	157,567	22,562	41	1, 3
Cryptogramma acrostichoides	22,652	83,690	150,162	21,168	42	Absent
Cystopteris chinensis	26,671	83,429*	131,808*	21,708*	40	Absent
Dryopteris decipiens	23,456	82,462	150,978	21,604	42	Absent
Gastoniella chaerophylla	22,657	81,918	148,099	20,867	40	2
Haplopteris elongata	27,188	80,810	156,002	20,816	41	1, 2
Hemionitis subcordata	30,921	82,607	165,631	21,182	43	1, 2, 3
Jamesonia brasiliensis	27,704	na	139,531*	20,941	41	1
Llavea cordifolia	23,208	81,944	149,387	21,027	42	Absent
Myriopteris covillei	25,567	83,093	155,548	21,321	42	2
Myriopteris lindheimeri	25,694	83,059	155,770	21,323	42	2
Myriopteris scabra	27,115	82,874	162,051	24,947	42	1
Notholaena standleyi	27,261	83,769	159,556	21,265	42	1, 2, 3
Onychium japonicum	23,419	82,289	150,156	21,029	41	Absent
Pellaea truncata	23,240	82,865	150,713	21,368	42	Absent
Pentagramma triangularis	23,378	85,675	153,445	21,014	42	1, 3
Pityrogramma trifoliata	22,465	82,321	148,156	20,905	40	Absent
Pteridium aquilinum	23,384	84,335	152,362	21,259	41	2
Pteris vittata	25,275	82,604	154,108	20,954	42	2
Scoliosorus ensiformis	21,078	82,358	145,327	20,813	40	Absent
Tryonia myriophylla	24,141	87,238	156,327	20,807	40	1, 2, 3
Vaginularia trichoidea	21,618	84,026	147,192	19,930	39	1, 2, 3
Vittaria appalachiana	22,185	84,330	149,531	20,831	40	1
Vittaria graminifolia	22,066	86,058	151,035	20,845	40	1, 2

* signifies that the genome was only a partial assembly and therefore the genome size is prone to error.
IR=inverted repeat; LSC=large single copy; SSC=small single copy.

MORFFO elements and phylogenetic position among species sampled (table 4).

Chromosome-wide read depth analyses revealed no shifts in coverage spanning MORFFO insertions or insertion boundaries, indicating that these inserts are not an artifact of genomic library preparation or genome misassembly. Furthermore, MORFFO insertions were detected in *V. appalachiana,* which was sequenced and assembled in a separate lab, using an alternative assembly protocol. We also examined each member of the MORFFO cluster using VecScreen (NCBI Resource Coordinators 2017) and RepeatMasker (Smit et al. 2013), neither of which yielded matches to any known vectors nor transposable elements. We searched for MORFFO sequences against current draft assemblies of nuclear genomes of the ferns *Azolla* and *Salvinia* on FernBase (Li et al. 2018) as well as scaffolds for the draft genomes of *Ceratopteris*, and whereas they were not detected in *Azolla* or *Salvinia*, we did observe the presence of *morffo1* in *Ceratopteris* scaffolds. In addition, we searched for the presence of MORFFO in available transcriptomes from members of Pteridaceae in the 1 kp project (Johnson et al. 2012; Matasci et al. 2014; Wickett et al. 2014; Xie et al. 2014), and found no evidence of transcription of MORFFO.

To test whether MORFFO elements could be of mitochondrial origin, we filtered plastid reads using the mitochondrial option in NOVOPlasty (Dierckxsens et al. 2017), and then assembled the remaining reads using *morffo1* as a seed. This did not generate an assembly of any known mitochondrial sequence. Instead, a seemingly circular 2,139 bp contig was inferred in *Adiantum tricholepis*, containing *morffo1* and *morffo2*, but no known mitochondrial sequences. Furthermore, this contig lacked the flanking inverted repeat

Table 3
Summary of Hits to MORFFO Sequences in NCBI Blast, Using Either BLASTX or TBLASTN

Species	Organism Group and Genome Compartment	Search Strategy	MORFFO Match	Length of Match (aa)	% Identity	Accession Number
Actinostachys pennula	Fern plastome	TBLASTN	*morffo2*	231	34	KU764518.1
Alsophila spinulosa	Fern plastome	TBLASTN	*morffo2*	78	44	FJ556581.1
Angiopteris angustifolia	Fern plastome	TBLASTN	*morffo2*	332	49	KP099647
Angiopteris evecta	Fern plastome	TBLASTN	*morffo2*	331	47	DQ821119.1
Asplenium nidus	Fern mitochondrial genome	TBLASTN	*morffo1*	260	52	AM600641.1
Asplenium prolongatum	Fern plastome	TBLASTN	*morffo2*	110/59	77/64	KY427332.1
Chondrocystis[a]	Cyanobacterium plasmid	TBLASTN	*morffo2*	233	27	AP018284.1
Crocosphaera watsonii[a]	Cyanobacterium	BLASTX	*morffo2*	365	43	WP_007310072.1
Dryopteris fragrans	Fern plastome	TBLASTN	*morffo1/morffo2*	323/355	46/49	KX418656.2
Helminthostachys zeylanica	Fern plastome	TBLASTN	*morffo2*	226	70	KM817788.2
Huperzia lucidula	Lycopod plastome	TBLASTN	*morffo2*	114	35	AY660566.1
Huperzia serrata	Lycopod plastome	TBLASTN	*morffo2*	114	35	KX426071.1
Lepisorus clathratus	Fern plastome	TBLASTN	*morffo1/morffo2*	183/208	50/56	KY419704.1
Lygodium japonicum	Fern plastome	TBLASTN	*morffo2*	152	36	HM021803.1
Myxosarcina sp.[a]	Green alga plastome	BLASTX	*morffo2*	497	27	WP_052055951.1
Nostoc punctiforme[a]	Cyanobacterium	TBLASTN	*morffo2*	245	28	CP001037.1
Ophioglossum californicum	Fern plastome	TBLASTN	*morffo1/morffo2/morffo3*	120/169/65	56/41/42	KC117178.1
Polypodium glycyrrhiza	Fern plastome	TBLASTN	*morffo3*	162	53	KP136832
Prasiola crispa	Green alga plastome	TBLASTN	*morffo2*	416	25	KR017750.1
Roya anglica	Green alga plastome	TBLASTN	*morffo1*	202	30	NC_024168
Roya obtusa	Green alga plastome	TBLASTN	*morffo1*	202	30	KU646496.1
Volvox carteri	Green alga plastome	TBLASTN	*morffo1/morffo2*	200/303	28/25	EU755299.1
Woodwardia unigemmata	Fern plastome	TBLASTN	*morffo1/morffo2*	247/105	51/77	KT599101.1

aPutative DN_5 superfamily conserved domain.

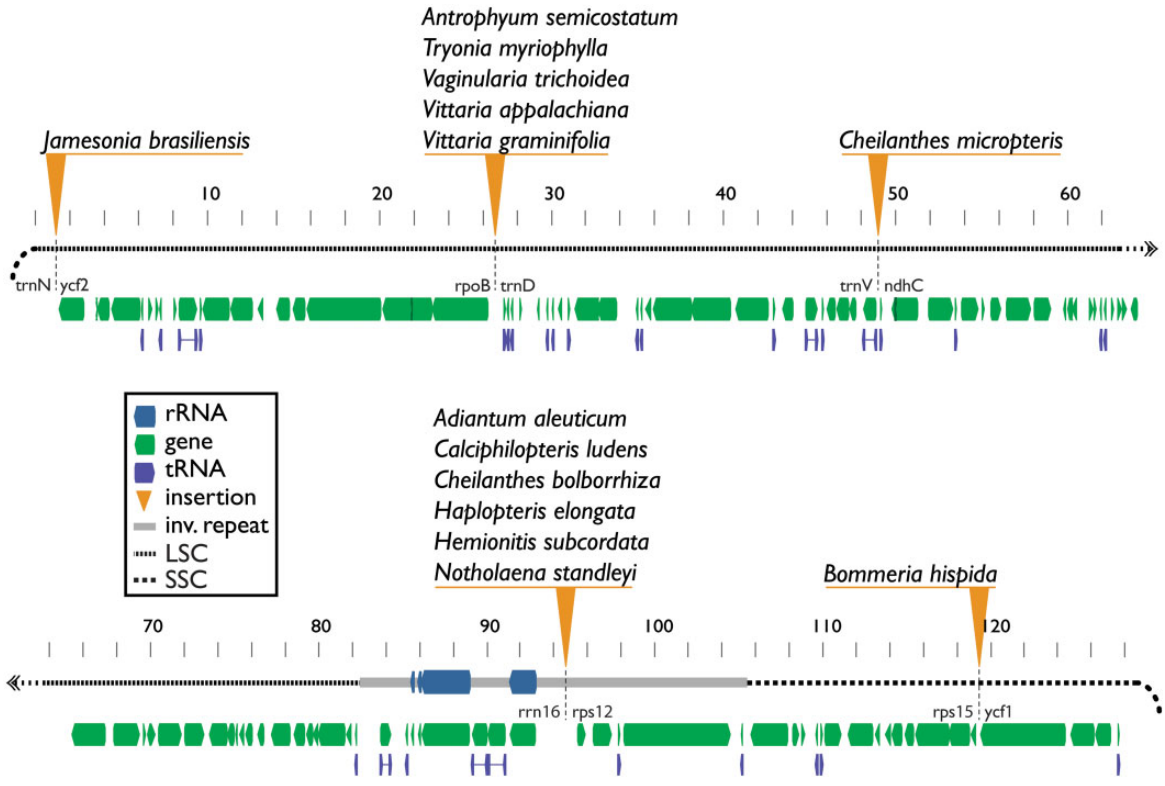

Fig. 2.—Detected insertion sites in plastomes of Pteridaceae, relative to *Adiantum capillus-veneris*. Light gray bar denotes inverted repeat region.

Table 4
Summary of Matches for MORFFO Sequences Within Ferns, Using BLASTN

Species	MORFFO Present	Length of Match (bp)	% Identity	Region	Accession Number
Adiantum aleuticum	*morffo1/morffo2*	1316/ 1851	88/67	IR	MH173079
Alsophila podophylla	*morffo1/morffo2/morffo3*	1250/1778/631	78/76/81	IR	MG262389
Antrophyum semicostatum	*morffo1/ morffo3*	1323/ 629	53/ 57	LSC	MH173087
Asplenium prolongatum	*morffo2*	265	81	IR	KY427332
Bommeria hispida	*morffo1/ morffo2/ morffo3*	1296/ 1708/ 1708	89/ 87/ 84	SSC	MH173074
Calciphilopteris ludens	*morffo1/ morffo2 (truncated)*	1324/ 928	78/ 51	IR	MH173084
Cheilanthes bolborrhiza	*morffo1/ morffo2*	1304/ 1859	92/ 90	IR	MH173073
Cheilanthes micropteris	*morffo1/ morffo3*	1292/ 630	50/ 50	LSC	MH173078
Cibotium barometz	*morffo1/ morffo2*	1296/1850	77/50	IR	NC_037893
Dicksonia squarrosa	*morffo1*	1321	50	IR	KJ569698
Diplopterygium glaucum	*morffo3*	639	55	LSC	KF225594
Drynaria roosii	*morffo1*	1314	51	LSC	KY075853
Haplopteris elongata	*morffo1/ morffo2*	1336/ 1859	56/ 55	IR	MH173086
Hemionitis subcordata	*morffo1/ morffo2/ morffo3*	1296/ 1863/ 665	91/ 75/ 92	IR	MH173072
Hymenasplenium unilaterale	*morffo1/morffo3*	252/ 643	65/61	IR/LSC	KY427350
Jamesonia brasiliensis	*morffo1*	1310	51	IR	MH173077
Mankyua chejuensis	*morffo1/ morffo2*	193/186	68/67	IR	KP205433
Myriopteris covillei	*morffo2*	1886	49	IR	MG953517
Myriopteris lindheimeri	*morffo2*	852	67	IR	HM778032
Myriopteris scabra	*morffo1*	1310	98	IR	MH173083
Notholaena standleyi	*morffo1/ morffo2/ morffo3*	1312/ 1866/ 640	89/ 86/ 91	IR	MH173067
Pentagramma triangularis	*morffo2/ morffo3*	1860/ 630	83/ 84	LSC	MH173070
Plagiogyria glauca	*morffo1*	1305	49	LSC	KP136831
Plagiogyria glauca	*morffo2*	1856	46	Mito	Wolf et al. (2015)
Plagiogyria japonica	*morffo2*	1295	50	LSC	HQ658099
Pteridium aquilinum	*morffo2*	280	65	LSC	HM535629.1
Pteris vittata	*morffo2*	1172	80	LSC	MH173068
Rhachidosorus consimilis	*morffo2*	430	70	IR	KY427356
Tryonia myriophylla	*morffo1/ morffo2/ morffo3*	1326/ 1854/ 641	89/ 84/ 86	LSC	MH173076
Vaginularia trichoidea	*morffo1/ morffo2/ morffo3*	1315/ 1836/ 630	53/ 52/ 62	LSC	MH173085
Vittaria appalachiana	*morffo1*	1316	68	LSC	MH173091

IR=inverted repeat; LSC=large single copy; SSC=small single copy.

normally associated with *morffo1*. A control assembly using the mitochondrial genes *atp1* and *cox1* as seed sequences was also generated using nonplastid raw reads from this species. The assemblies based on mitochondrial genes had markedly lower average coverage depth (86.5) than that of the *morffo1* based assembly (556) and the plastome (359), suggesting that the MORFFO cluster in *A. tricholepis* exists as an independent mobile element that is not an integrated component of the mitochondrial genome. Where this element resides within the cell is unclear.

Relationships among MORFFO sequences were estimated using an ML optimality criterion and 1,000 bootstrap replicates. Each MORFFO element—*morffo1, morffo2, morffo3*—comprises a monophyletic clade, within which some sequences are united with moderate to high bootstrap support (supplementary fig. S1, Supplementary Material online). However, relationships among sequences within each MORFFO clade were not congruent with the accepted species tree.

Plastome Phylogenomic Analyses

Our final, concatenated plastome alignment included 68,047 sites spanning 76 plastid loci for 31 taxa, including three outgroups. PartitionFinder2 returned a most favorable partition model with 36 subsets (AICc: 1211937.74576), from which partition blocks were assigned in RAxML (Stamatakis 2014) and MrBayes (Ronquist et al. 2012). Trees inferred using ML and Bayesian optimality criteria were in full topological agreement with maximum support on all branches, with the exception of the branch subtending *O. japonicum*, *T. myriophylla*, *J. brasiliensis*, *Gastionella chaerophylla*, and *Pityrogramma trifoliata*, which was supported by an ML bootstrap of 95 and a posterior probability of 1.0 (fig. 1).

Discussion

Comparative analyses of plastomes over the past two decades have dramatically improved our understanding of their evolution across land plants. Early data painted a picture of

structural and organizational stability among deeply divergent embryophyte plastomes, punctuated by relatively few large-scale inversions (Ogihara et al. 1988; Hoot and Palmer 1994; Wolf et al. 2010). Recent evidence, however, has begun to expose the plastome as a dynamic molecule that in some lineages undergoes frequent changes in DNA content and structure (Guisinger et al. 2011; Lin et al. 2012; Knox 2014; Cremen et al. 2018). As more information has come to light, many highly rearranged plastomes have also been found to host sizeable insertions, occasionally including ORFs of unknown homology (Knox 2014; Cremen et al. 2018). In some cases, these inserted ORFs appear to encode functional proteins, whereas in others they resemble conserved domains that have undergone extensive rearrangements and/or pseudogenization, comparable to what has been observed within some plastid genes (e.g., *ndhK*, *clpP*, and *ycf2*; Haberle et al. 2008; Lin et al. 2012; Smith 2014; Sun et al. 2016). Several studies have determined that similar, undescribed plastid ORFs are the result of horizontal transfer from mitochondria to plastids (Goremykin et al. 2009; Iorizzo et al. 2012; Ma et al. 2015; Burke et al. 2016; Rabah et al. 2017).

Plastome ORF insertions like these—with no known sequence homology—have not been characterized in ferns, although previous authors have reported large intergenic expansions and insertions in some taxa (Gao et al. 2011; Logacheva et al. 2017). It was speculated that some of these intergenic expansions originated via intracellular transfer from the mitochondrion (Logacheva et al. 2017), but until now, limited sampling in previous studies has obscured the highly mobile nature of these peculiar sequences. Here, we take a focused phylogenetic approach, targeting the fern family Pteridaceae, to reveal a suite of highly mobile ORFs (MORFFO) within a broad sampling of plastomes from across the family. Preliminary analyses indicate that MORFFO elements, which are frequently associated with extensive genomic rearrangements, may be present in lineages well-removed from ferns.

Characterization of MORFFO Elements

Logacheva et al. (2017) established that "hypervariable" sequences of significant length are found in the IR of *Woodwardia unigemmata* as well as the LSC of *Plagiogyria*. Our results are consistent with their findings, and further expose the dynamic nature of these sequences (MORFFO) among a collection of closely related fern plastomes. Searches for MORFFO-like sequences outside of Pteridaceae returned similar, putatively homologous regions in many ferns (tables 3 and 4), but not in seed plants. Significantly, an 8 kb region in the plastome of the fern *M. chejuensis* (Ophioglossaceae) contains an expanded complement of the MORFFO cluster. Additional searches for MORFFO-like sequences outside of vascular plants revealed similar conserved domains in several cyanobacteria plastomes (table 3), as well as domains in the plastomes of the green algae *Prasiola crispa*, *Roya obtusa*, and *R. anglica* (table 3).

Within ferns, we note that MORFFO elements are frequently found adjacent to inferred sites of genomic inversion. For example: 1) *morffo1* is found adjacent to the border of one of two hypothesized inversions in the region spanning *rpoB-psbZ* which occurred in a common ancestor of the core leptosporangiates (fig. 3; PPG I 2016); 2) *morffo3* is found within a 9.7 kb inversion that characterizes leptosporangiate ferns (Kim et al. 2014); 3) *morffo1* and *morffo2* are found inserted adjacent to the 7 kb inversion seen in the plastomes of vittarioid ferns; 4) and *morffo1* and *morffo2* also appear adjacent to an inversion described in filmy ferns (fig. 3; Gao et al. 2011; Wolf et al. 2011; Kim et al. 2014; Kuo et al. 2018). Although MORFFO insertions are frequently associated with inversions, we are unsure why. One possibility is that MORFFO may target nucleotide sites that are prone to inversion. Conversely, the insertion of MORFFO could be directly influencing inversion events. In several taxa, we observe a proliferation of the inverted repeats flanking *morffo1*, possibly caused by replication slippage, or possibly by the repeated insertion and excision of *morffo1*. In other groups, plastome reorganization has been similarly associated with the presence of small dispersed repeats like these (Wicke et al. 2011). Likewise, the relationship between MORFFO insertion sites and inversions is not unlike the insertions seen in other dynamic embryophyte plastomes (Knox 2014).

The variable presence, location, and configuration of MORFFO observed in a phylogenetic context suggests that these ORFs are mobile elements. With a few notable exceptions, plastid genes are not frequently gained or lost, yet our results indicate that MORFFO moves into, out of, and across the plastome in relatively short evolutionary timescales. Although MORFFO sequences have been observed in mitochondrial contigs (Logacheva et al. 2017), it is important to note that they are not found in either of the currently available complete mitochondrial genomes of ferns (Guo et al. 2017). Additionally, the location of these insertions in mitochondrial genomes seems to vary as much as in those of plastids, making it difficult to pinpoint a potential mitochondrial origin. Furthermore, we have also noted the presence of MORFFO in the nuclear genome of *Ceratopteris*. Thus, MORFFO appears to be moving across genomes as readily as within them. However, this does not explain the origin of the elements or the mechanisms of their movement.

Plasmid-like sequences have been observed in the plastomes of diatoms, green algae, dinoflagellates, and red algae (Zhang et al. 1999; La Claire and Wang 2000; Ruck et al. 2014; Lee et al. 2016; Cremen et al. 2018). Although we are not aware of previous work describing chloroplast plasmids in land plants, this is a plausible mechanism to explain the variable presence and location of MORFFO elements. It would also explain the variability in order and direction of MORFFO insertions (fig. 1). As noted above (see Results

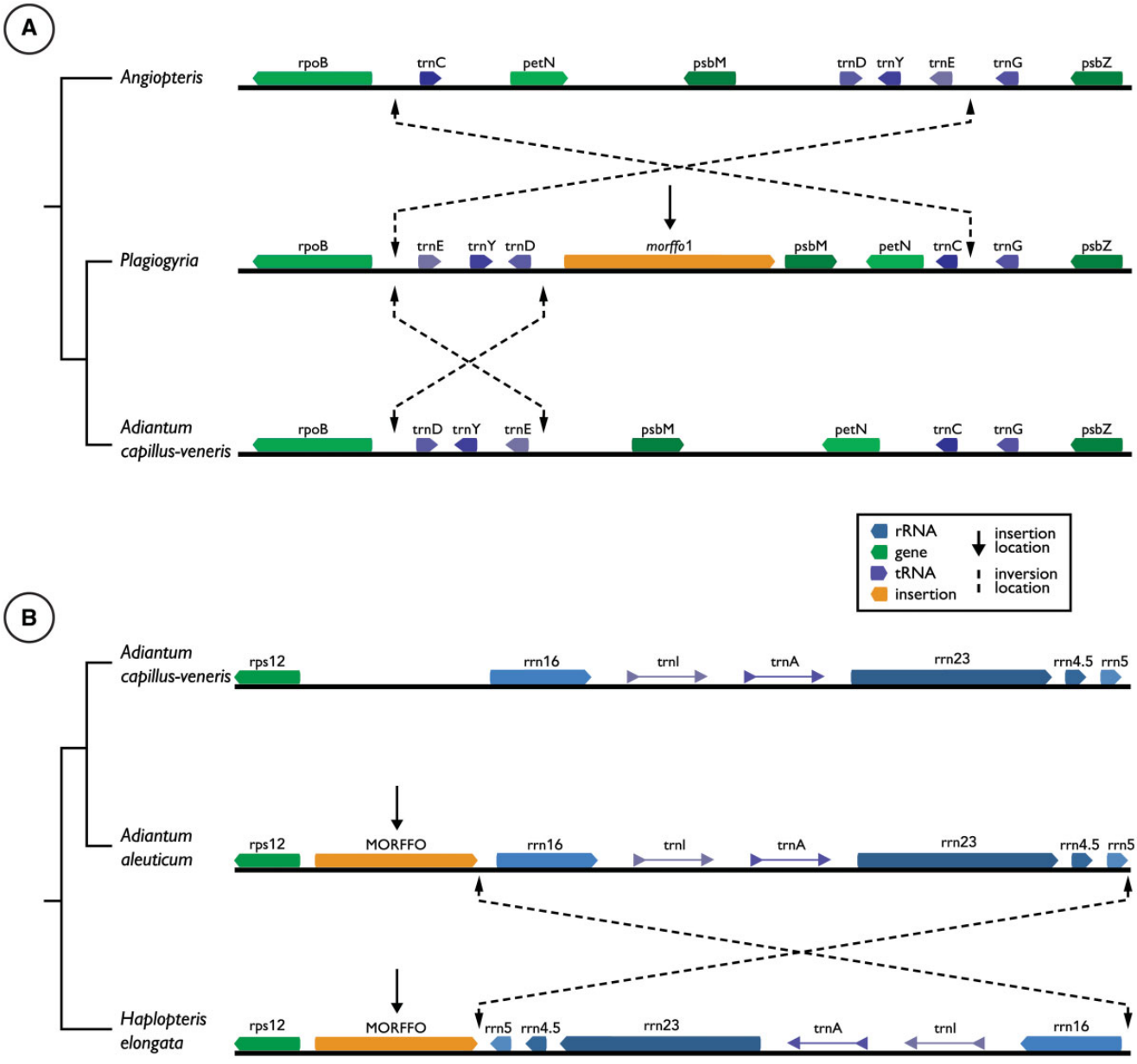

Fig. 3.—Major inversion events uncovered in fern plastomes. (*A*) Depiction of the two inversion events necessary to explain gene order differences between *Angiopteris* and *Adiantum* and their relationship to morffo1. (*B*) Depiction of the inversion events seen in early leptosporangiate and vittarioid ferns, highlighting relationship of *morffo1* to the event.

section), we were able to assemble a circular sequence containing *morffo1* and *morffo2* from *A. tricholepis*, which did not have the MORFFO insert in its plastome. The coverage analysis for this sequence indicates that MORFFO is likely a high copy-number sequence that is independent of both the plastome and the mitochondrial genome. This, combined with the fact that *morffo2* has regions that share similarity to conserved domains (DN_5 superfamily) associated with primase genes found in mobile elements of cyanobacteria, strongly suggests that these sequences could be of plasmid origin, possibly from a plastid plasmid.

Alternatively, MORFFO elements could be of viral origin. In addition to being similar to plasmid primases, the conserved domain found in *morffo2* also resembles primase genes found in phages. Viral origins could explain why MORFFO is found frequently but irregularly in fern plastomes. Likewise, many of the above arguments in favor of a plasmid or plasmid-like origin for MORFFO sequences can also be attributed to viral origin.

The structural similarities that *morffo1* shares with bacterial insertion sequences is noteworthy, especially because insertion sequences are known to cause inversions (Darmon and

Leach 2014). This, along with the clear mobility of these sequences strongly suggests that *morffo1* could be a previously undescribed insertion sequence. The case for *morffo1* being an insertion sequence is made stronger by the fact that of all the MORFFO sequences, it appears to display the most independence. In nonplastid DNA, it is almost always found without the other MORFFO sequences. It is also more frequently seen independent of additional MORFFO sequences in the plastomes of ferns outside of Pteridaceae. Furthermore, copies of *morffo1* were detected in the nuclear genome of *Ceratopteris* and the mitochondrial genome of *Asplenium nidus*, suggesting that it may be a particularly promiscuous mobile element. The relationship of *morffo2* and *morffo3* to *morffo1*, however, remains unclear. If *morffo1* is an independent insertion sequence, then how are *morffo2* and *morffo3* inserted?

Phylogenetic analysis of MORFFO elements reveals three strongly supported, monophyletic groups made up of *morffo1*, *morffo2,* and *morffo3* elements. Relationships within each MORFFO clade do not reflect the accepted species phylogeny, but phylogenetic similarity across clades may reflect shared histories of degradation among MORFFO elements (supplementary fig. S1, Supplementary Material online).

We also note that whereas MORFFO elements are pervasive in Pteridaceae, they appear to be less common in most other groups of ferns. In part, this may be an artifact of the historical reliance on reference-based assemblies, which can be biased toward assembling genomes that appear more similar to their reference, thus reducing the likelihood of detecting significant rearrangements. Based on the few sequences available in GenBank, MORFFO elements may be prevalent in *Plagiogyria* and Ophioglossaceae; however, it remains to be determined how widespread this cluster of ORFs is among other lineages of ferns. More studies at the family level are needed to understand the extent to which MORFFO sequences are moving throughout fern genome space. The current paucity of fern nuclear and mitochondrial genomes makes it difficult to determine the source of these inserts. At the time of writing there were only two fern mitochondrial sequences (Guo et al. 2017) available in GenBank, and no nuclear genomes, although several were in preparation. As more genomes are published in the coming years, the reservoir from which MORFFO clusters are migrating should become clear.

Plastome Variation Across Pteridaceae

Pteridaceae is an ecologically and morphologically diverse family comprising more than 10% of extant fern species (Schuettpelz et al. 2007). Within this group, subfamily Vittarioideae, comprising the genus *Adiantum* and the so-called vittarioid ferns, is especially noteworthy. High levels of molecular substitution rate heterogeneity have been detected between members of the genus *Adiantum* and the vittarioid ferns, in both plastid and nuclear DNA sequences (Rothfels and Schuettpelz 2014; Grusz et al. 2016). As noted above, we also find variation in plastome structure across the family, with MORFFO elements (*morffo1*, *morffo2*, and *morffo3*) being repeatedly gained, lost, and/or rearranged, even among closely related taxa (fig. 1).

The physical position of MORFFO cluster insertions is relatively conserved within the five major clades comprising Pteridaceae (fig. 1), but in some cases the location and composition of these clusters varies widely, even between congeneric relatives (e.g., *Myriopteris lindheimeri* vs. *M. scabra* and *V. appalachiana* vs. *V. graminifolia*; fig. 1). Based on our sampling, we find no evidence of MORFFO elements within Cryptogrammoideae or Parkerioideae. However, unique insertions of MORFFO sequences have taken place in some members of the Pteridoideae, including *J. brasiliensis* (between *trnN* and *ycf2*), *T. myriophylla* (between *trnD* and *trnY)*, and *Pteris vittata* (between *psbM* and *petN*). MORFFO elements were notably absent in the species of *Gastoniella*, *Pityrogramma*, and *Onychium* sampled (fig. 1). Compared with the Pteridoideae, subfamilies Vittarioideae and Cheilanthoideae exhibit relative stability in their MORFFO insertion sites (fig. 1). Altogether, we find at least nine unique MORFFO cluster insertions across the Pteridaceae (there are almost certainly more), not including a multitude of species-specific rearrangements, gains, and losses of individual MORFFO elements (*morffo1*, *morffo2*, and *morffo3*) following cluster insertions.

Within each independent MORFFO cluster insertion, the presence and position of *morffo1*, *morffo2*, and *morffo3* are highly variable among species sampled. For example, a variety of insertions, rearrangements, and losses of all three elements can be found in the MORFFO cluster found between *rpoB* and *trnD* within subfamily Vittarioideae (fig. 1). Likewise, the MORFFO cluster between *rps12* and *rrn16* in Cheilanthoideae shows insertions and losses of all three MORFFO elements, including a duplication of *morffo2* in *H. subcordata* (fig. 1).

Interestingly, vittarioid ferns do not appear to have experienced expansion or contraction of the IR, which have been associated with extensive genomic rearrangements, gene loss, and the proliferation of repetitive regions in other groups (Zhu et al. 2016). In addition to the variable presence of MORFFO elements, we find that their insertion into the ancestral vittarioid IR (fig. 1) may have also coincided with a loss of *trnT*. Most vittarioid ferns share an additional gene loss (*trnV*)—with the exception of *Haplopteris elongata*, in which *trnV* is found intact. Given that *V. trichoidea* and *H. elongata* are inferred to be successively sister to the remaining vittarioid ferns sampled (fig. 1; Schuettpelz et al. 2016), this topology implies the gain of *trnV* into the plastome of *H. elongata*. Nevertheless, a shared ~7 kb inversion between *rrn16* and *rrn5* in all vittarioid ferns except *V. trichoidea* further

supports our phylogenetic inferences based on DNA sequence data (figs. 1 and 3).

The vittarioid ferns are characterized by high levels of plastome rearrangement, elevated molecular substitution rates, a shift to epiphytism, morphological reduction, and shared ancestral whole genome duplication (Pryer et al. 2016). This array of shared traits leads one to ask which (if any) might have driven these changes in vittariod plastome structure and expression. Similarly, frequent rearrangements, insertions, and losses of the MORFFO suite within the plastomes of subfamily Cheilanthoideae coincide with adaptations to extreme xeric environments, extensive whole genome duplications, hybridization, and apomixis—any of which may relate to the changes we detect in plastome structure across this subfamily.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Thanks to Jim Van Etten, Dan Sloan, Aaron Duffy, and Michael McKain for various fruitful discussions, Nico Devos (Duke University Center for Genomic and Computational Biology) for technical assistance, and Sally Chambers for providing the *V. appalachiana* sample. Thanks to Blaine Marchant for generously allowing us to use the *Ceratopteris* nuclear genome in our search for MORFFO. Thanks also to the University of Utah Center for High-Performance Computing, particularly Anita Orendt, for providing computational resources for data analyses, and to the Minnesota Supercomputing Institute. We further acknowledge the Smithsonian Laboratories of Analytical Biology. Funding for this research was provided in part by the United States National Science Foundation (award DEB-1405181 to E.S.) and the Smithsonian Institution.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Blazier JC, et al. 2016. Variable presence of the inverted repeat and plastome stability in Erodium. Ann Bot. 117(7):1209–1220.

Bungard RA. 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. Bioessays 26(3):235–247.

Burke SV, et al. 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). BMC Plant Biol. 16(1):140.

Cai Z, et al. 2008. Extensive reorganization of the plastid genome of Trifolium subterraneum (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol. 67(6):696–704.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol. 58(4):424–441.

Chumley TW, et al. 2006. The complete chloroplast genome sequence of Pelargonium x hortorum: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 23(11):2175–2190.

Cremen MCM, Leliaert F, Marcelino VR, Verbruggen H. 2018. Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in Siphonous Green Algae (Bryopsidales, Chlorophyta). Genome Biol Evol. 10(4):1048–1061.

Cui L, et al. 2006. Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. BMC Evol Biol. 6:13.

Darmon E, Leach DRF. 2014. Bacterial genome instability. Microbiol Mol Biol Rev. 78(1):1–39.

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26(19):2455–2457.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.

Gao L, Zhou Y, Wang Z-W, Su Y-J, Wang T. 2011. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. BMC Plant Biol. 11:64.

Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am J Bot., doi: 10.1002/ajb2.1048

Givnish TJ, et al. 2010. Assembling the Tree of the Monocotyledons: plastome Sequence Phylogeny and Evolution of Poales. Ann Mo Bot Gard. 97(4):584–616.

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. Mol Biol Evol. 26(1):99–110.

Grusz AL, Rothfels CJ, Schuettpelz E. 2016. Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. BMC Genomics 17:692.

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.

Guo W, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within Juniperus plants and have shifted multiple times during cupressophyte evolution. Genome Biol Evol. 6(3):580–590.

Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns Ophioglossum californicum and Psilotum nudum are highly repetitive with the largest organellar introns. New Phytol. 213(1):391–403.

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(4):350–361.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the Cryptomeria japonica D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.

Hoot SB, Palmer JD. 1994. Structural rearrangements, including parallel inversions, within the chloroplast genome of Anemone and related genera. J Mol Evol. 38(3):274–281.

Iorizzo M, et al. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 12:61.

Johnson MTJ, et al. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 7(11):e50226.
Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform., doi:10.1093/bib/bbx108
Kearse M, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma Oxf Engl. 28(12):1647–1649.
Kim HT, Chung MG, Kim K-J. 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. Mol Cells 37(5):372–382.
Knox EB. 2014. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. Proc Natl Acad Sci U S A. 111(30):11097–11102.
Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22(5):1208–1222.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 33(7):1870–1874.
Kuo L-Y, Qi X, Ma H, Li F-W. 2018. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. Am J Bot. doi:10.1002/ajb2.1152.
Kuraku S, Zmasek CM, Nishimura O, Katoh K. 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41:W22–W28.
Labiak PH, Karol KG. 2017. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. Am J Bot., doi: 10.3732/ajb.1700135
La Claire JW, Wang J. 2000. Localization of plasmidlike DNA in giant-celled marine green algae. Protoplasma 213(3–4):157–164.
Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29(6):1695–1701.
Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol. 34(3):772–773.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359.
Lee J, et al. 2016. Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. Sci Rep. 6:23744.
Leliaert F, et al. 2012. Phylogeny and molecular evolution of the green algae. CRC Crit Rev Plant Sci. 31(1):1–46.
Li F-W, et al. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat Plants 4(7):460–472.
Li F-W, Kuo L-Y, Pryer KM, Rothfels CJ. 2016. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. Genome Biol Evol. 8(8):2452–2458.
Lin C-P, Wu C-S, Huang Y-Y, Chaw S-M. 2012. The complete chloroplast genome of Ginkgo biloba reveals the mechanism of inverted repeat contraction. Genome Biol Evol. 4(3):374–381.
Logacheva MD, et al. 2017. Comparative analysis of inverted repeats of polypod fern (Polypodiales) plastomes reveals two hypervariable regions. BMC Plant Biol. 17(Suppl 2):255.
Madden T. 2013. The BLAST Sequence Analysis Tool. National Center for Biotechnology Information (US).
Ma P-F, Zhang Y-X, Guo Z-H, Li D-Z. 2015. Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. Sci Rep. 5:11608.
Matasci N, et al. 2014. Data access for the 1, 000 Plants (1KP) project. Gigascience 3:17.
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). p. 1–8.
Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A. 107(10):4623–4628.
Mower JP, Vickrey TL. 2018. Chapter nine - structural diversity among plastid genomes of land plants. In: Chaw S-M, Jansen RK, editors. Advances in botanical research. Vol. 85. Cambridge (MA): Academic Press. p. 263–292.
NCBI Resource Coordinators 2017. Database resources of the national center for biotechnology information. Nucleic Acids Res. 45(D1):D12–D17.
Ogihara Y, Terachi T, Sasakuma T. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc Natl Acad Sci U S A. 85(22):8573–8577.
Palmer JD. 1985. Comparative organization of chloroplast genomes. Annu Rev Genet. 19:325–354.
PPGI 2016. A community-derived classification for extant lycophytes and ferns. J Syst Evol. 54(6):563–603. PPG I.
Pryer KM, et al. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am J Bot. 91(10):1582–1598.
Pryer KM, Huiet L, Li F-W, Rothfels CJ, Schuettpelz E. 2016. Maidenhair Ferns, Adiantum, are Indeed Monophyletic and Sister to Shoestring Ferns, Vittarioids (Pteridaceae). Syst Bot. 41(1):17–23.
Rabah SO, et al. 2017. Plastome sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. Plant Genome 10(3), doi:10.3835/plantgenome2017.03.0020.
Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6 (visited on 2017-09-01), 2014.
Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542.
Rothfels CJ, Schuettpelz E. 2014. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. Syst Biol. 63(1):31–54.
Ruck EC, Nakov T, Jansen RK, Theriot EC, Alverson AJ. 2014. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. Genome Biol Evol. 6(3):644–654.
Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol. 14:23.
Schuettpelz E, et al. 2016. A revised generic classification of vittarioid ferns (Pteridaceae) based on molecular, micromorphological, and geographic data. Taxon 65(4):708–722.
Schuettpelz E, Schneider H, Huiet L, Windham MD, Pryer KM. 2007. A molecular phylogeny of the fern family Pteridaceae: assessing overall relationships and the affinities of previously unsampled genera. Mol Phylogenet Evol. 44(3):1172–1185.
Shaw J, et al. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot. 92(1):142–166.
Sigmon BA, Adams RP, Mower JP. 2017. Complete chloroplast genome sequencing of vetiver grass (Chrysopogon zizanioides) identifies markers that distinguish the non-fertile 'Sunshine' cultivar from other. Ind Crops Prod., https://www.sciencedirect.com/science/article/pii/S0926669017304922
Smit A, Hubley R, Green P. 2013. 2013–2015. RepeatMasker Open-4.0. https://www.repeatmasker.org
Smith DR. 2014. Mitochondrion-to-plastid DNA transfer: it happens. New Phytol. 202(3):736–738.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9): 1312–1313.
Stein DB, et al. 1992. Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proc Natl Acad Sci U S A. 89(5):1856–1860.
Sun Y, et al. 2016. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. Mol Phylogenet Evol. 96:93–101.
Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol. 17:1105–1109.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5(2):123–135.
Tryon RM. 1990. Pteridaceae. In: Kramer KU, Green PS, editors. Pteridophytes and gymnosperms. Berlin, Heidelberg (Germany): Springer. p. 230–256.
Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27(2):171–180.
de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. 2015. YCF1: a Green TIC? Plant Cell 27(7):1827–1833.
Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.
Wei R, et al. 2017. Plastid phylogenomics resolve deep relationships among Eupolypod II Ferns with rapid radiation and rate heterogeneity. Genome Biol Evol. 9(6):1646–1657.
Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76(3–5):273–297.
Wickett NJ, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A. 111(45): E4859–E4868.
Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 84(24):9054–9058.
Wolf PG, et al. 2015. An exploration into fern genome space. Genome Biol Evol. 7(9):2533–2544.
Wolf PG, et al. 2011. The evolution of chloroplast genes and genomes in ferns. Plant Mol Biol. 76(3–5):251–261.
Wolf PG, Roper JM, Duffy AM. 2010. The evolution of chloroplast genome structure in ferns. Genome 53(9):731–738.
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, 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(2):59–65.
Xie Y, et al. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30(12):1660–1666.
Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol. 21(5):809–818.
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18(5):821–829.
Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. Nature 400(6740):155–159.
Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209(4):1747–1756.

Associate editor: Daniel Sloan