

Received Date:

Revised Date:

Accepted Date:

Article Type: MS - Regular Manuscript

Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat

Jeffrey P. Mower^{1,2*}, Peng-Fei Ma^{1,3}, Felix Grewe⁴, Alex Taylor⁵, Todd P. Michael⁶, Robert VanBuren⁷, Yin-Long Qiu⁵

¹ Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA

² Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

³ Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

⁴ Integrative Research Center, Science and Education, Field Museum of Natural History, Chicago, IL 60605, USA

⁵ Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

⁶ J. Craig Venter Institute, La Jolla, CA 92037, USA

⁷ Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA

* **Corresponding Author:** Jeffrey P. Mower, jpmower@unl.edu, +1 402-472-2130

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/nph.15650](https://doi.org/10.1111/nph.15650)

This article is protected by copyright. All rights reserved

Received: 27 August 2018

Accepted: 10 December 2018

ORCIDs: JPM (0000-0002-1501-5809), TPM (0000-0001-6272-2875), RVB (0000-0003-2133-2760)

Summary

- Lycophytes are a key group for understanding vascular plant evolution. Lycophyte plastomes are highly distinct, indicating a dynamic evolutionary history, but detailed evaluation is hindered by the limited availability of sequences.
- Eight diverse plastomes were sequenced to assess variation in structure and functional content across lycophytes.
- Lycopodiaceae plastomes have remained largely unchanged compared with the common ancestor of land plants, whereas plastome evolution in *Isoetes* and especially *Selaginella* is highly dynamic. *Selaginella* plastomes have the highest GC content and fewest genes and introns of any photosynthetic land plant. Uniquely, the canonical inverted repeat was converted into a direct repeat (DR) via large-scale inversion in some *Selaginella* species. Ancestral reconstruction identified additional putative transitions between an inverted and DR orientation in *Selaginella* and *Isoetes* plastomes. A DR orientation does not disrupt the activity of copy-dependent repair to suppress substitution rates within repeats.
- Lycophyte plastomes include the most archaic among vascular plants and the most reconfigured among land plants. These evolutionary trends correlate with the mitochondrial genome, suggesting shared underlying mechanisms. Copy-dependent repair for DR-localized genes indicates that recombination and gene conversion are not inhibited by the DR orientation. Gene relocation in lycophyte plastomes occurs via overlapping inversions rather than transposase/recombinase-mediated processes.

Key Words: evolutionary stasis, gene loss, inversion, *Isoetes* (quillworts), Lycopodiophyta (lycophytes), Lycopodiaceae (clubmosses), plastid genome (plastome), *Selaginella* (spikemosses)

Author Manuscript

Introduction

Across the diversity of land plants, the plastid genome (plastome) is distinguished by its overall conservation in size, structure, and content. Plastomes of photoautotrophic land plants typically range from 120 to 160 kb in length, with ~80 protein-coding genes, 4 rRNAs, and ~30 tRNAs arranged into a structure that usually includes large and small single-copy (SC) regions (termed LSC and SSC) separated by a large inverted repeat (IR) in two copies (Wicke *et al.*, 2011; Jansen & Ruhlman, 2012; Mower & Vickrey, 2018). One of the most notable effects of this genomic structure is that substitution rates are several times lower in the IR relative to SC regions (Wolfe *et al.*, 1987; Perry & Wolfe, 2002; Li *et al.*, 2016; Zhu *et al.*, 2016). This reduction in IR substitution rates has been attributed to a copy-dependent repair mechanism such as biased gene conversion (Birky & Walsh, 1992), which may be facilitated by frequent intramolecular recombination between IR copies that produces two isomeric forms of the plastome (Bohnert & Löffelhardt, 1982; Palmer, 1983).

Comparative analysis of plastome structures from diverse land plant representatives demonstrated that inversions and IR boundary shifts (via expansion or contraction) are the primary processes of structural rearrangements, with varying frequency among lineages (reviewed in Mower & Vickrey, 2018). More rarely, a third type of rearrangement (variously termed transposition or translocation) has been invoked to explain the intramolecular relocation of one or more genes within the plastome (e.g., Milligan *et al.*, 1989; Cosner *et al.*, 1997; Cosner *et al.*, 2004; Chumley *et al.*, 2006; Tsuji *et al.*, 2007; Karol *et al.*, 2010; Knox, 2014). Nearly all such examples come from more rearranged plastomes, making it difficult to reconstruct whether the relocation was direct (implying the involvement of a translocase or site-specific recombinase) or the indirect result of some combination of overlapping inversions and/or differential IR expansions/contractions.

Among nonvascular land plants, plastomic rearrangements are rare, with just a few examples of inversions and IR boundary shifts and no putative cases of transposition/translocation. Indeed, plastomes from most species are fully collinear, including duplications of four ribosomal RNAs (rRNAs) and five transfer RNAs (tRNAs) within the IR (Grosche *et al.*, 2012; Bell *et al.*, 2014; Villarreal *et al.*, 2018), indicating that this structural arrangement was stably inherited from the common ancestor of all land plants (Mower & Vickrey, 2018). The rare exceptions of structural change involve a ~70 kb inversion that characterizes mosses within the Funariidae (Sugiura *et al.*, 2003; Goffinet *et al.*, 2007), a 1 kb inversion unique to the parasitic liverwort *Aneura mirabilis*

(Wickett *et al.*, 2008), and a ~7 kb IR expansion that is specific to hornworts in Anthocerotaceae (Villarreal *et al.*, 2013).

By contrast, inversions and IR boundary shifts are recurrent themes in the evolution of euphyllophyte (i.e., angiosperm, gymnosperm, and fern) plastomes (Zhu *et al.*, 2016; Mower & Vickrey, 2018). For example, a ~35 kb inversion is shared by all species, suggesting that it occurred in the euphyllophyte ancestor (Raubeson & Jansen, 1992). Major expansions of the IR distinguish angiosperm and gymnosperm plastomes from all other land plant lineages, which were followed by further episodes of inversions and IR boundary shifts affecting one to many genes in particular descendant lineages (Wicke *et al.*, 2011; Jansen & Ruhlman, 2012; Mower & Vickrey, 2018). Gene transitions/translocations have been invoked for some highly rearranged plastomes (Milligan *et al.*, 1989; Cosner *et al.*, 1997; Cosner *et al.*, 2004; Chumley *et al.*, 2006; Knox, 2014), but dense sampling of plastomes suggested that overlapping inversions provide an equally good or better explanation for most, but not necessarily all, examples (Lee *et al.*, 2007; Cai *et al.*, 2008; Haberle *et al.*, 2008; Knox, 2014; Weng *et al.*, 2014). Among ferns, a small number of plastomic inversions are shared by all living species (Gao *et al.*, 2011; Grewe *et al.*, 2013; Mower & Vickrey, 2018), while descendant lineages have experienced either multiple IR expansions from a plesiomorphic arrangement or multiple IR contractions following an early expansion (Grewe *et al.*, 2013; Zhu *et al.*, 2016; Mower & Vickrey, 2018). The relocation of several genes from the LSC to the IR of most leptosporangiate ferns was inferred to result from two overlapping inversions (Hasebe & Iwatsuki, 1992; Stein *et al.*, 1992; Raubeson & Stein, 1995). Other structural changes among fern plastomes may have been facilitated by the presence of mobile elements, termed MORFFO, within the plastome (Robison *et al.*, 2018).

Given that lycophytes are the sister lineage to all other extant vascular plants and comprise three distinctive lineages (clubmosses, spikemosses, and quillworts) that have each existed for >300 million years (Larsén & Rydin, 2016; Testo *et al.*, 2018), it is perhaps unsurprising that their plastomes exhibit diverse characteristics. Plastomes from *Huperzia lucidula* and *Huperzia serrata* are nearly identical to one another (Z. Y. Guo *et al.*, 2016), and, except for a small expansion of the IR, they are collinear with nonvascular plant plastomes, implying a plesiomorphic retention of gene order from the land plant ancestor (Wolf *et al.*, 2005; Mower & Vickrey, 2018). By contrast, *Isoetes* and *Selaginella* harbor plastomes that are more rearranged (Tsuji *et al.*, 2007; Smith, 2009; Karol *et al.*, 2010; Xu *et al.*, 2018). The plastome from *Isoetes flaccida* exhibits two diagnostic changes, including an inversion affecting *chlL* and *chlN* and a

putative translocation of *ycf2* from the LSC to the SSC (Karol *et al.*, 2010). Plastomes from three *Selaginella* species (*S. moellendorffii*, *S. tamariscina*, and *S. uncinata*) are more diverse due to multiple inversions, putative translocations, and gene losses (Tsuji *et al.*, 2007; Smith, 2009; Xu *et al.*, 2018). The IR has evolved differently in all three genera, exemplified by minor expansions in *Huperzia* that captured portions of *ndhF* and exon 2 of *rps12* (Wolf *et al.*, 2005; Karol *et al.*, 2010), by an independent and slightly larger IR expansion in *Isoetes* that assimilated *rps7* and exons 2 and 3 of *rps12* (Karol *et al.*, 2010), and by distinct IR expansions in different *Selaginella* species accompanied by multiple losses of tRNA genes (Tsuji *et al.*, 2007; Smith, 2009; Xu *et al.*, 2018).

The variable nature of plastome sequences among lycophytes indicates that multiple evolutionary events must have occurred, but the limited availability of complete plastomes has hampered attempts to reliably reconstruct the series of events that produced these changes. To date, only six species from three genera have been sequenced from lycophytes, a group that comprises ~1300 species and 18 genera (Schuettpehl *et al.*, 2016). To improve plastome sampling among lycophytes, we generated complete plastome sequences from seven species (*Dendrolycopodium obscurum*, *Diphasiastrum digitatum*, *Isoetes malinverniana*, *Isoetes piedmontana*, *Lycopodium clavatum*, *Selaginella kraussiana*, and *Selaginella lepidophylla*) and resequenced the plastome of an eighth species (*H. lucidula*). These new sequences were used to assess the range of plastomic diversity among species, and from this information, the evolutionary history of plastome change in lycophytes was reconstructed. These analyses uncovered a pattern of evolutionary stasis among Lycopodiaceae genomes contrasted with major genomic upheaval in *Selaginella*, including a novel genomic arrangement in which the large rRNA-containing repeat copies shifted from an inverted to a direct orientation. The origins of this novel arrangement and its potential effects on the rate of nucleotide substitution were evaluated.

Materials and Methods

Sample collection and DNA sequencing

Tissue was collected from six species of lycophytes: *De. obscurum*, *Di. digitatum*, *H. lucidula*, *I. malinverniana*, *I. piedmontana* and *L. clavatum*. Total genomic DNA was extracted using a simplified CTAB procedure (Doyle & Doyle, 1987) or the Qiagen DNeasy Plant mini kit and then sequenced on the Illumina HiSeq or NextSeq platforms. DNA for *S. lepidophylla* was extracted and then sequenced using the PacBio platform, as described previously (VanBuren *et al.*,

2018). For *S. kraussiana*, Illumina sequencing data were obtained from the NCBI Sequence Read Archive (accession number SRR2037123). Details of sample collection and genomic sequencing are listed in Supporting Information Table S1.

Plastome assembly

The *S. lepidophylla* plastome was identified from contigs in the PacBio assembly produced by Canu v1.7 from the published nuclear genome project (VanBuren *et al.*, 2018). To assemble plastomes for the remaining species, a full data set of all raw Illumina sequence reads and a reduced data set including a subset of 20 million raw reads (which often contains sufficient plastomic coverage for high-quality assembly) were generated for each species. The full and reduced data sets were each independently assembled using Velvet v1.2.10 (Zerbino & Birney, 2008) and SPAdes v3.11.1 (Bankevich *et al.*, 2012). For both assemblers, kmers were set dependent upon the sequenced read length: kmers 79, 95, 111 and 127 were used with 151 bp reads; kmers 69, 83, 97 and 111 were used with 125 bp reads; and kmers 61, 71, 81 and 91 were used with 101 bp reads. For SPAdes assemblies, the careful flag was activated and coverage cutoff was set either to 10 (for the full-read data set) or 5 (for the 20-million-read data set). For Velvet assemblies, independent runs were executed without scaffolding using pairwise combinations of kmers and expected coverage values (set to 50, 100, 200, 500, and 1000), as previously described (Guo *et al.*, 2014; Sigmon *et al.*, 2017).

For each of the draft assemblies produced from the two data sets and two assemblers, plastid contigs were identified by using blastn searches with *H. lucidula* (GenBank accession number AY660566), *I. flaccida* (GU191333) and *S. moellendorffii* (HM173080) plastome sequences as queries. For each species, a minimum of three of the best draft assemblies (regardless of assembler or data set used) were identified based on maximal average length of identified plastid contigs. These best assemblies were manually aligned, from which a final consensus sequence was generated.

PCR and Sanger sequencing were used to verify the sequence and structure of the plastome at several notable positions. The direct orientation of the large repeat in *S. kraussiana* was verified by sequencing PCR amplicons that spanned each boundary of the SC and repeat regions (Supporting Information Fig. S1). In addition, the plastomes of *De. obscurum*, *Di. digitatum*, *H. lucidula* and *L. clavatum* contain several repetitive regions composed of tandem repeats, with unit sizes ranging from one to several hundred nucleotides. The size and complexity of these repetitive regions were verified using primers anchored in non-repetitive flanking regions. PCR reactions included an initial denaturation step for 3 min at 94 °C, followed by 34 cycles of

denaturation (94 °C for 30 s), annealing (50–54 °C for 30 s), and elongation (72 °C for 90–150 s), and concluding with a final elongation step for 10 min at 72 °C. PCR products were sequenced on both strands by GenScript (Piscataway, NJ).

Plastome annotation

The *I. piedmontana* plastome was initially annotated using the Live Annotation feature in Geneious v10.2.2 (Kearse *et al.*, 2012), with the annotation of *I. flaccida* (GU191333) as a template. All other plastomes were initially annotated using the online implementation of GeSeq (Tillich *et al.*, 2017), with annotated plastomes from *H. lucidula* (AY660566), *H. serrata* (KX426071), *I. flaccida* (GU191333), *S. moellendorffii* (HM173080) and *S. uncinata* (AB197035) selected as references. Annotation of tRNAs was performed with ARAGORN (using the plant chloroplast genetic code) and tRNAscan-SE (set for organellar tRNAs) as implemented by GeSeq.

To ensure that all protein-coding genes, tRNAs, rRNAs, introns and splicing junctions were detected and properly annotated in all lycophyte plastomes evaluated in this study (including all eight newly generated plastomes and the six plastomes obtained from GenBank), a blastn search was conducted against each plastome using a query set of all plastid genes (including intron sequences when present) collected from a diversity of land plant plastomes. In addition, for protein-coding genes, tblastn searches were conducted for increased sensitivity. These blast validations identified several exons and introns that were often missed by GeSeq, involving the short first exons in *psbB*, *psbD*, and *rpl16* and the *trans*-spliced *rps12* gene, as well as several genes and introns that were missing or incorrectly annotated in some plastomes in GenBank. These annotation errors were corrected prior to comparative analyses. Finally, to assess whether MORFFO-like mobile elements were present in lycophyte plastomes, blastn and tblastn searches were also conducted using intact *morffo* gene sequences as queries. All newly sequenced plastomes were deposited in GenBank under accession numbers MH549637–MH549643 and MK089531.

Comparative genomic analyses

All comparative analyses were performed using the corrected annotations for all 14 examined lycophyte plastomes. Thus, any variation among genes and introns is likely to reflect true variation in genomic content rather than annotation errors. The presence of genes and introns was taken from the corrected annotations, all of which were verified by the blastn and tblastn searches performed during intron annotation. The absence of genes was confirmed by an absence of any hits in the blastn and tblastn searches. Pseudogenes were scored when reading

frames had frameshifts and/or substantial truncations. Additionally, because no U-to-C editing has been detected in *Selaginella* (Hecht *et al.*, 2011; Oldenkott *et al.*, 2014), the presence of premature stop codons was used as another indicator of pseudogenization in *Selaginella* plastomes. Collinearity among plastomes was assessed by comparison of linear genome maps generated by OgDraw (Lohse *et al.*, 2007). Ancestral reconstruction of inversion events and gene and intron losses was performed manually using the maximum parsimony criterion, which minimizes the number of inferred changes in a phylogenetic context.

Substitution rate calculations

Protein-coding and rRNA genes were collected from the plastomes of 12 different species of lycophytes and a diverse sampling of seven other land plants: the angiosperm *Magnolia tripetala* (KJ408574), the gymnosperm *Ginkgo biloba* (KP099648), the ferns *Angiopteris angustifolia* (KP099647) and *Ophioglossum californicum* (KC117178), the hornwort *Anthoceros formosae* (AB086179), the moss *Physcomitrella patens* (KY126308), and the liverwort *Marchantia polymorpha* (LC192146). Individual protein genes were aligned by codon with TranslatorX v1.1 (Abascal *et al.*, 2010) using default settings, while individual rRNA genes were aligned with Muscle v3.8.31 (Edgar, 2004) using default settings. Three separate protein-gene data sets, containing either the 51 protein genes that are single copy in all lycophytes, the three protein genes (*ndhB*, *psbM*, *rps7*) that are duplicated in *S. kraussiana*, or the *rps4* gene that is duplicated in *S. moellendorffii* and *S. tamariscina*, were created with Gblocks v0.91b (Castresana, 2000) in codon mode with gaps allowed in up to half of sequences (-t=c -b5=h -a=y). An rRNA data set containing all four rRNA genes was generated with Gblocks in DNA mode with gaps allowed in up to half of sequences (-t=d -b=h -a=y). All four trimmed data sets are provided in Supporting Information Dataset S1.

A maximum likelihood tree was generated from the 51-gene data set with PhyML v3.0 (Guindon *et al.*, 2010) using a GTR+G+I nucleotide model, which was determined as the best-fitting model by the ModelTest (Posada & Crandall, 1998) implementation in HyPhy v2.2.4 (Pond *et al.*, 2005), and all model parameters were estimated during the PhyML run. The recovered topology fully agreed with recent phylogenetic analyses of lycophytes (Field *et al.*, 2016; Larsén & Rydin, 2016; Zhou *et al.*, 2016). Sequence divergence was calculated with HyPhy for each branch in a tree that was constrained according to the maximum likelihood topology from the phylogenetic analysis of 51 SC genes. Branch lengths representing synonymous sequence divergence (d_s) were estimated from each of the protein-gene data sets using a local MG94CustomF3x4 codon model. The best-fitting codon model was determined by the

CodonModelCompare analysis in HyPhy to be 012232 for the 51-gene data set, 012312 for the *ndhB+psbM+rps7* data set, and 012012 for the *rps4* data set. Branch lengths representing total rRNA sequence divergence (d) were estimated from the concatenated rRNA data set using a GTR+G+I nucleotide model, which was determined to be the best fit by the HyPhy implementation of ModelTest.

Results

Lineage-specific variation in plastome size and GC content among lycophytes

Plastome size and GC content among the three major lineages of lycophytes vary in a lineage-specific manner (Table 1). Plastomes from Lycopodiaceae are largest (152–161 kb) with the lowest GC content (35–36%), *Isoetes* plastomes are intermediate in size (145–146 kb) and GC content (38%), and *Selaginella* plastomes are smallest (115–144 kb) and the most GC rich (51–55%). The size of the LSC correlates directly with overall genome size, being largest in Lycopodiaceae and smallest in *Selaginella*, whereas SSC size exhibits an inverse trend. The two-fold variation in size of the large plastid repeat, at 7.3–18 kb, is driven primarily by expansions, contractions, gene losses, and inversions (described in detail in the remainder of Results). GC content within the IR is consistently higher than the SC regions in all lycophytes, due at least in part to the elevated GC content of rRNA genes.

Extensive variation in gene and intron content among lycophyte plastomes

The repertoire of genes and introns varies widely among lycophyte plastomes, which again follows a lineage-specific trend (Table 1). For protein-coding genes (Fig. 1a), all six sequenced Lycopodiaceae plastomes contain the same set of 87 intact (and presumably functional) protein-coding genes, a count that includes the recently described ORF named *ycf94* (Song *et al.*, 2018). Additionally, a frameshift within the *ycf2* gene was characteristic of Lycopodiaceae (*De. obscurum*, *Di. digitatum*, and *L. clavatum*), suggesting that this gene may now operate as two separate reading frames (hereafter referred to as *ycf2n* and *ycf2c*). The three *Isoetes* species have a slightly reduced repertoire of 82 intact protein genes in their plastomes. Of the five lost genes, nonfunctional remnants remain for *accD*, *infA*, *rps2* and *rps16* but no fragment of *ycf94* was detected in any *Isoetes* plastome.

The five species of *Selaginella* contain many fewer intact protein-coding genes, indicating substantial gene loss (Fig. 1a). Gene loss is most pronounced in the recently published plastome from *S. tamariscina* and the newly elucidated *S. lepidophylla* plastome, with only 58 and 64 intact protein genes remaining, respectively. Notably, both of these plastomes lack

functional copies of the *ndh* complex (Figure 1a; Xu *et al.*, 2018), similar to observations in a variety of other land plants (Wakasugi *et al.*, 1994; Wickett *et al.*, 2008; Blazier *et al.*, 2011), while the *S. kraussiana* and *S. lepidophylla* plastomes lack the maturase gene *matK*, which encodes an intron splicing factor (Vogel *et al.*, 1997; Zoschke *et al.*, 2010). All *Selaginella* plastomes have also lost numerous ribosomal protein genes, as well as a variety of other genes, that are universally present in all other lycophytes. The status of *infA* is uncertain in some *Selaginella* species; an intact but highly divergent copy exists in *S. uncinata*, which was tentatively scored as functional in this study, whereas it has probably become a pseudogene in *S. moellendorffii* (due to the presence of two internal stop codons which cannot be eliminated because U-to-C RNA editing is absent from *S. moellendorffii*). In ferns, particular MORFFO mobile elements have led to structural rearrangements (Robison *et al.*, 2018); however, no homology to any *morffo*-like genes were detected in any of the 14 lycophyte plastomes.

Among RNA genes, all lycophytes have retained the standard set of four ribosomal RNA (rRNA) genes (Fig. 1b), whereas transfer RNA (tRNA) content is more variable among lineages (Fig. 1c). All three *Isoetes* plastomes have retained 32 tRNA genes, including a *trnT-GGU* gene that is common throughout land plants (Grosche *et al.*, 2012; Grewe *et al.*, 2013; Bell *et al.*, 2014; Guo *et al.*, 2014; Villarreal *et al.*, 2018). Plastomes from Lycopodiaceae have retained nearly the same set of tRNAs, except for the loss of *trnT-GGU*. By contrast, all five species of *Selaginella* have lost the majority of their plastid tRNAs. This loss is most pronounced in *S. tamariscina* and *S. kraussiana*, which harbor only 6 and 10 detectable tRNAs in their plastomes, respectively.

Intron content also varies widely among lycophytes (Fig. 1d). All six Lycopodiaceae species contain a complete set of 22 plastid introns that are widely present across land plants. *Isoetes* plastomes lack just one of these introns, and its absence can be attributed to the loss of function and subsequent degradation of the host *rps16* gene. By contrast, 16 different introns have been variously lost from one or more *Selaginella* species, most extensively from *S. tamariscina* and *S. lepidophylla* plastomes. The absence of 13 introns (from *ndhA*, *ndhB*, *rpl16*, *rps12*, *rps16*, *ycf66*, and six tRNAs) can be attributed to loss of their host genes. For the remaining three intron losses (from *clpP* and *ycf3*), the host plastid genes are still present and presumably functional.

Plesiomorphic retention of plastome structure in Lycopodiaceae

Plastome structure and gene order are highly conserved among the Lycopodiaceae species examined in this study (Fig. 2). All six plastomes share the loss of *trnT-GGU* and all but *L.*

clavatum have an expanded IR that includes nearly all of the *ndhF* gene. Based on current sampling, these two evolutionary events can be most parsimoniously inferred to have occurred in the common ancestor of Lycopodiaceae. In *L. clavatum*, however, only about 200 bp of the *ndhF* gene are located within the IR, suggesting that this genome experienced a more recent IR contraction. Both *Huperzia* species share another minor IR expansion that moved exon 2 and part of intron 2 of *rps12* into the IR, whereas the split of *ycf2* into *ycf2n* and *ycf2c* is specific to the Lycopodiaceae (*De. obscurum*, *Di. digitatum*, and *L. clavatum*). Other than these minor changes, plastome gene order in Lycopodiaceae is collinear with plastomes from nonvascular land plants, and this conserved synteny enabled a robust reconstruction of the ancestral lycophyte plastome (Fig. 2, top).

Multiple shifts between inverted and direct orientations of the large plastid repeat in *Selaginella*

In contrast to the evolutionary stasis of Lycopodiaceae plastomes, major structural changes have reorganized *Selaginella* plastomes (Fig. 3). In the *S. kraussiana* and *S. tamariscina* plastomes, the most conspicuous change is the relative orientation of the large rRNA-containing repeat copies, which are arranged as a pair of direct repeats (DR) instead of the canonical IR orientation observed in other land plant plastomes. Comparison to the ancestral lycophyte plastome demonstrates that this unique IR-to-DR transition occurred by a single inversion event that fully spanned the IR (Inversion S1 in Fig. 3), with one endpoint in the ancestral LSC (between *trnF-GAA* and *rps4*) and the other in the ancestral SSC (just beyond the IR/SSC boundary). The direct orientation of the *S. kraussiana* repeat was independently confirmed by PCR and Sanger sequencing across all four DR/SC junctions (Supporting Information Fig. S1), while the DR structure for *S. tamariscina* was confirmed in another study (Xu *et al.*, 2018).

Evolutionary reconstruction of the plastomes of *S. moellendorffii*, *S. lepidophylla* and *S. uncinata* revealed that this same Inversion S1 also affected these plastomes, demonstrating that this event occurred in the common ancestor of all five *Selaginella* species (Fig. 3; Supporting Information Fig. S2). In the *S. lepidophylla* and *S. moellendorffii* plastomes, Inversion S1 was followed by a second inversion (Inversion S2 for *S. lepidophylla* and Inversion S3 for *S. moellendorffii*) that also fully spanned the rRNA-containing repeat, which independently reverted the DR back to a canonical IR orientation in the two species (Fig. 3). In the *S. uncinata* plastome, Inversion S1 was followed by at least four additional inversions and a tandem duplication of the *trnQ* and *psbK* genes, which collectively account for all of the various relocations and rearrangements in this plastome relative to the lycophyte ancestor (Supporting

Information Fig. S2). Three of these *S. uncinata* inversions also spanned the rRNA repeat (Inversions S5, S7 and S8 in Supporting Information Fig. S2), which sequentially caused the orientation to revert from a DR to an IR, then invert back to a DR, and then revert once again to an IR. Additional plastomic inversions in *S. tamariscina* (Inversion S4 in Fig. 3) and *S. uncinata* (Inversion S6 in Supporting Information Fig. S2) did not affect the orientation of the repeat but instead inverted gene order in other regions of the plastome.

In addition to inversion events, these ancestral plastome reconstructions allowed us to infer the relative timing of IR expansions as well as gene and intron losses among *Selaginella* species (Fig. 3; Supporting Information Fig. S2). Many genes and introns are missing from all sampled *Selaginella* species, so their loss can be most parsimoniously interpreted as having occurred in the common ancestor of the five species, while other gene and intron losses occurred in particular descendant lineages. Most IR expansions were inferred to have occurred independently in the five species, except that the IR expansion that captured *rps4* probably occurred in the ancestor of *S. moellendorffii*, *S. tamariscina* and *S. uncinata*.

Additional shifts in repeat orientation in *Isoetes* plastomes

Compared with the ancestral lycophyte plastome, the main distinguishing features of the three sequenced *Isoetes* plastomes are the relocation of *ycf2* from the LSC into the SSC, an inversion within the SSC, and a minor IR expansion (Fig. 4). The relocation of *ycf2* can be attributed to two overlapping inversions (Inversion I1 and I2 in Fig. 4). Inversion I1 extended from *ycf2* to just beyond the IR/SSC boundary, while Inversion I2 endpoints involved the same region except it did not include *ycf2*, which effectively stranded *ycf2* in the SSC. Importantly, both of these inversion events in the *Isoetes* plastome fully spanned the IR region. Thus, as also detected for *Selaginella*, the *Isoetes* lineage is inferred to have experienced a period in which the IR existed in a DR orientation, with a subsequent reversion back to an IR orientation. A third inversion (Inversion I3 in Fig. 4) was localized within the SSC that inverted most genes (from *ycf1* to *ndhF*) relative to the *chlN–chlL–ycf2* gene cluster. A small IR expansion resulted in the capture of a small part of *ycf2* from the SSC and exons 2 and 3 of *rps12* and *rps7* from the LSC. Finally, *trnD-GUC* was tandemly duplicated specifically in *I. malinverniana* (Fig. 4).

Substitution rates for genes in single-copy and repeat regions in *Selaginella* plastomes

The IR exhibits reduced substitution rates in plant plastomes due to a copy-dependent repair mechanism driven by homologous recombination, but it is unknown whether repeats in a DR orientation can promote homologous recombination or copy-dependent repair. To explore this

issue, we first established the relative substitution rate for SC genes by calculating synonymous sequence divergence (d_s) of 51 genes that are single-copy in all lycophyte plastomes (Fig. 5a). Overall, SC d_s is substantially higher for *Selaginella* compared with other lycophytes. Since the split between *Isoetes* and *Selaginella*, SC d_s is 5–7 times higher in *Selaginella* relative to *Isoetes*, with the highest SC d_s in *S. kraussiana*, which is consistent with previous studies reporting elevated sequence divergence in *Selaginella* (Korall & Kenrick, 2004; Zhou *et al.*, 2016).

To test whether genes within the DR have reduced substitution rates relative to SC genes, SC d_s branch lengths were compared with corresponding d_s branches for the concatenated *ndhB+psbM+rps7* genes (which are duplicated in the DR of *S. kraussiana*) and for *rps4* (which is duplicated in the DR of *S. tamariscina*). For the concatenated *ndhB+psbM+rps7* genes (Fig. 5b), the terminal *S. kraussiana* branch (representing DR d_s) was one-third the length of the corresponding SC d_s branch for *S. kraussiana* from the 51-SC-gene tree (0.184 versus 0.544 substitutions/site). Similarly, in the *rps4* tree (Fig. 5c), the terminal *S. tamariscina* branch (representing DR d_s) was less than one-half the length of the corresponding SC d_s branch from the 51-SC-gene tree (0.141 versus 0.285). These results are consistent with a reduced substitution rate for DR genes in *S. kraussiana* and *S. tamariscina*. As a contrast, d_s branch lengths were roughly equivalent between the *ndhB-psbM-rps7* tree and the 51-SC-gene tree for *S. moellendorffii* (0.161 versus 0.128), *S. tamariscina* (0.217 versus 0.285), and *S. lepidophylla* (0.258 versus 0.228), consistent with the SC status of *ndhB-psbM-rps7* in these three species. Likewise, d_s branch lengths were roughly equivalent in the trees from *rps4* (which is SC in *S. kraussiana* and *S. lepidophylla*) and the 51 SC genes for *S. kraussiana* (0.523 versus 0.544) and *S. lepidophylla* (0.170 versus 0.228).

As an additional test, SC d_s from the 51-gene tree was compared with overall sequence divergence (d) from a tree constructed from all four rRNA genes (Fig. 5d). Given that rRNA genes are duplicated in all *Selaginella* plastomes, and under the assumption that effects of selection on rRNA genes are roughly equivalent for each species, we would expect rRNA d to correlate with SC d_s in an equivalent manner for all species if the DR experiences copy-correction activity. Indeed, the rRNA d was 8–12 times lower than SC d_s for all *Selaginella* species, indicating that they have experienced similar reductions in rRNA substitution rate relative to SC rate despite differences in repeat orientation among species.

Discussion

Evolutionary stasis in Lycopodiaceae plastomes contrasts with exceptional evolutionary eccentricities in *Selaginella* plastomes

Comparative analysis of 14 plastomes (eight newly generated for this study) from species spanning the breadth of extant lycophytes has revealed substantial variation in size, content, structure and substitution rate, driven primarily by evolutionary eccentricities in *Selaginella* plastomes. The 115–161 kb range of lycophyte plastome sizes, due mostly to variation among *Selaginella* plastomes (Table 1), closely aligns with the typical 120–160 kb range observed for most other photoautotrophic vascular plants (Wicke *et al.*, 2011; Jansen & Ruhlman, 2012; Mower & Vickrey, 2018). The 51–55% GC content among the five *Selaginella* plastomes (Table 1) is substantially higher than for all other land plant plastomes sequenced to date, as previously noted for *S. moellendorffii* and *S. uncinata* (Tsuji *et al.*, 2007; Smith, 2009), while the plastomic GC content of 35–38% for Lycopodiaceae and *Isoetes* (Table 1) is unremarkable compared with other vascular plants (Jansen & Ruhlman, 2012; Grewe *et al.*, 2013; Chaw *et al.*, 2018). Similarly, substitution rates are substantially higher in *Selaginella* plastid genes compared with other lycophytes and a representative sampling of other land plants (Fig. 5). Another unique characteristic of the *Selaginella* plastome is an extreme level of cytidine-to-uracil (C-to-U) RNA editing, with over 3400 sites for *S. uncinata*, the highest of any land plant, whereas RNA editing in other lycophytes was predicted to be much lower, at 340 sites in *H. lucidula* and 460 sites in *I. flaccida* (Oldenkott *et al.*, 2014).

With respect to plastid gene and intron content, *Selaginella* is again an extreme outlier due to unprecedented gene and intron losses, which contrasts sharply with the minimal losses from Lycopodiaceae and *Isoetes* (Table 1; Fig. 1). The total unique repertoire (i.e., not counting duplicates in the IR) of 68–93 genes and 7–11 introns in *Selaginella* plastomes is the lowest among all photoautotrophic land plants, which typically harbor 110–120 genes and 18–22 introns in their plastomes (e.g., Grosche *et al.*, 2012; Jansen & Ruhlman, 2012; Grewe *et al.*, 2013; Bell *et al.*, 2014; Villarreal *et al.*, 2018). This major reduction in plastid gene content was driven primarily by the loss of 60–80% of the 32 tRNAs as well as many ribosomal protein genes, particularly from *S. tamariscina* and *S. kraussiana* plastomes, and all *ndh* genes from the *S. tamariscina* and *S. lepidophylla* plastomes. Homologs to some, but not all, of the lost plastid genes were identified in the nuclear genomes of *S. tamariscina* and *S. moellendorffii*, indicating functional intracellular transfer of these genes to the nuclear genome (Smith, 2009; Xu *et al.*, 2018). The lack of detection of nuclear homologs for the other lost plastid genes may indicate a

complete loss of function from the organism, as observed previously for *ndh* genes in many seed plants (Ruhlman *et al.*, 2015), or an inability to detect nuclear homologs for generally short and fast-evolving genes (Smith, 2009). Gene losses from *Selaginella* and *Isoetes* plastomes were the primary cause for most of the intron losses, whereas intron losses from functional *ycf3* and *clpP* genes suggest a role for retroprocessing, which is a process involving the reverse transcription and genomic reintegration of a spliced transcript. Retroprocessing has also been postulated as the primary mechanism of intron loss from plant mitochondrial genomes (Cuenca *et al.*, 2016; Grewe *et al.*, 2016; W. Guo *et al.*, 2016).

Plastome structure also varies substantially among lycophytes. The *S. kraussiana* and *S. tamariscina* plastomes stand alone among land plants due to their unique DR orientation (Fig. 3). The only comparable plastomes with DR orientations are found among various red algae in the Bangiophyceae and Florideophyceae (Lee *et al.*, 2016). The *S. uncinata* plastome is the most rearranged among lycophytes, with at least five inversions and a tandem duplication required to reconcile its plastome structure (Supporting Information Fig. S2). In contrast, Lycopodiaceae plastomes are unique among vascular plants for their similarity to nonvascular plant plastomes; indeed, the retention of most plastid genes and introns and conserved synteny compared with nonvascular land plants indicate that plastome structure has remained largely unchanged in Lycopodiaceae and most nonvascular land plants since the last common ancestor of all land plants ~500 million years ago (Magallon *et al.*, 2013; Morris *et al.*, 2018).

Collectively, these results show that *Selaginella* plastomes exhibit a wide variety of atypical features compared with all other land plants, whereas Lycopodiaceae plastomes are remarkable for their lack of change. The confinement of these extreme evolutionary eccentricities to *Selaginella* plastomes indicates that they probably arose within the *Selaginella* lineage. The extent of these genomic changes varies among the five sampled *Selaginella* species. The DR plastomes of *S. kraussiana* and *S. tamariscina* tend to have smaller sizes (Table 1), fewer genes and introns (Table 1; Fig. 1), and the fastest substitution rates (Fig. 5), although the *S. lepidophylla* plastome with an IR structure is even smaller and has lost many genes as well. Denser sampling of *Selaginella* plastomes is needed to more fully assess the evolutionary dynamics of these genomic features within the genus, and to evaluate whether a DR structure promotes other types of structural change.

Broadly correlated evolutionary dynamics of lycophyte plastomes and mitochondrial genomes

Intriguingly, the *Selaginella* mitochondrial genome exhibits many of the same extreme evolutionary trends as the plastome. This includes a small genome with a high GC content, in fact the highest among eukaryotes (Hecht *et al.*, 2011). Mitochondrial gene content is also substantially lower for *Selaginella* relative to *Isoetes* and *Phlegmariurus* (Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012). Mitochondrial gene loss was primarily due to loss of tRNAs and ribosomal protein genes (Hecht *et al.*, 2011), paralleling the predominant pattern of gene loss from the plastome. Furthermore, analyses of a few mitochondrial genes and introns indicate that the mitochondrial substitution rate is substantially higher for *Selaginella* relative to other lycophytes (Wikstrom & Pryer, 2005; Guo & Mower, 2013), while the *Selaginella* mitochondrial transcriptome has the highest amount of C-to-U RNA editing relative to all other land plants (Hecht *et al.*, 2011). The *Selaginella* mitochondrial genome is also highly recombinogenic (Hecht *et al.*, 2011), while the *Phlegmariurus* genome is the least rearranged among vascular plants (Liu *et al.*, 2012). However, one major inconsistency between the two organellar genomes relates to introns: Mitochondrial intron content for all three sequenced lycophyte genomes is very high compared with other land plants (Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012), with more mitochondrial introns in *Selaginella* than in any other land plant, which is opposite to the pattern of massive intron loss from *Selaginella* plastomes.

A correlation between GC content and RNA editing was previously noted (Malek *et al.*, 1996; Tsuji *et al.*, 2007; Smith, 2009; Hecht *et al.*, 2011), although it is unclear which is the causal factor or if both phenomena are the effect of some other driving force. For example, a GC bias for the increased rate of nucleotide substitution in *Selaginella* could potentially drive up both GC content and the number of edited sites in *Selaginella*. A higher mutation rate environment in the organellar genomes may have also facilitated the transfer of genes to the nuclear genome, although the extreme level of editing would likely require that these genes be transferred as RNA or as a DNA transfer of a retroprocessed gene. Correlated evolution of mitochondrial and plastid features could also be coordinated by the effect of one or more of the many nuclear factors involved in genomic maintenance and transcriptional processing that are shared between the two organelles (Delannoy *et al.*, 2007; Takenaka *et al.*, 2012; Gualberto & Newton, 2017). Ultimately, many more lycophyte organellar genomes are needed to assess the true degree of correlation of these features between the mitochondrial and plastid genomes, and the putative underlying causes of these correlations.

DR orientation and implications for recombination and repair

The direct orientation of the rRNA-containing repeat in some *Selaginella* plastomes has intriguing implications with respect to the activities of recombination and repair. In plastomes with an IR, it is now well established that the repeat copies undergo frequent homologous recombination, resulting in isomeric forms of the plastome that differ in the relative orientation of the LSC and SSC (Bohnert & Löffelhardt, 1982; Palmer, 1983). This recombinational activity is also known to promote copy-dependent repair via gene conversion, reducing the substitution rate of genes that reside in the IR relative to those in the SC regions (Wolfe *et al.*, 1987; Perry & Wolfe, 2002; Zhu *et al.*, 2016). However, due to the novelty of plastomes with a DR arrangement, no studies have yet assessed whether homologous recombination occurs or promotes copy-dependent repair in DR plastomes.

Our study demonstrates that genes within the DR of *S. kraussiana* and *S. tamariscina* have reduced substitution rates relative to SC genes (Fig. 5), consistent with the continued activity of copy-dependent repair despite a DR arrangement. This finding implies that recombination is indeed occurring between the two repeat copies in these DR plastomes. Assuming a circular chromosome, recombination between the DR copies would result in a pair of subgenomic circles, each containing either the LSC or SSC and one copy of the plastomic repeat. However, many studies have raised questions about the circular chromosome model for plastomes, suggesting instead that the major form of the plastome is linear (e.g., Bendich, 2004; Oldenburg & Bendich, 2016). If the *Selaginella* plastome exists predominantly as linear molecules, then frequent recombination between repeats would generate a diverse population of linear molecules in which the repeat might separate the LSC and SSC, or two copies of the LSC, or two copies of the SSC. Studies using electron microscopy and electrophoretic separation are needed to assess the *in vivo* structure of the plastome in *Selaginella*.

Overlapping inversions as the mechanism for gene relocation in lycophytes

Finally, our results demonstrate that there is no need to invoke direct transposition or intragenomic translocation, as mediated by a translocase or recombinase, for the relocation of plastid genes from the LSC to the SSC in *S. uncinata* and *S. moellendorffii*. Instead, overlapping inversions provide a more likely explanation, particularly because the intermediate step was captured in the plastomes from *S. kraussiana* and *S. tamariscina*, which produced the DR orientation in these two genomes (Fig. 3). The retention of this intermediate inversion step in the *S. kraussiana* and *S. tamariscina* plastomes provides strong evidence that inversion S1 occurred in the common ancestor of all five sampled *Selaginella* species, prior to any

subsequent overlapping inversions. Subsequent inversion events, which broadly overlapped inversion S1, must then be inferred for *S. lepidophylla*, *S. moellendorffii* and *S. uncinata* to explain the reversion of the IR orientation. Importantly, the endpoints of the subsequent inversions were different in each of these plastomes, providing a clear explanation for the slightly different suite of genes that were ultimately relocated between the LSC and SSC of these species. Denser sampling of the many species in the clades represented by *S. lepidophylla*, *S. moellendorffii* and *S. uncinata* should further pinpoint the relative timing of these subsequent inversions, and may reveal species with plastomes that are even more rearranged. Overall, among land plants it seems that there are no strong candidates for gene relocation via direct transposition or translocation. As with *Selaginella*, gene relocations affecting core leptosporangiate ferns were better explained by overlapping inversions (Hasebe & Iwatsuki, 1992; Stein *et al.*, 1992; Raubeson & Stein, 1995). In that case, evidence for the first inversion comes from the fact that the orientation of the rRNA operon has flipped and is now transcribed towards the LSC in core leptosporangiate ferns, instead of towards the SSC as seen in nearly all other land plant plastomes. Likewise, overlapping inversions with distinct evolutionary intermediates were shown to be better explanations than transposition/translocation for the suite of genomic rearrangements observed among species in the Oleaceae (Lee *et al.*, 2007). It thus seems prudent to assume that the LSC-to-SSC relocation of *ycf2* in all sampled *Isoetes* genomes (Fig. 4) resulted from the pair of overlapping inversions rather than direct transposition/translocation. Currently, *Isoetes* sampling is restricted to species from clades C (*I. malinverniana*) and E (*I. flaccida* and *I. piedmontana*), following the classification of Larsén and Rydin (2016). Sampling of *Isoetes* species from clades A and B, which fall outside of the C+E clade, may identify species having only one of the two inferred inversions, which would provide strong support for the overlapping inversion hypothesis in *Isoetes*.

Acknowledgements

The authors gratefully acknowledge Jay Bolin, Lytton Musselman, Peter Schafran, Weston Testo, and Trisha Vickrey for assistance with plant collection, Isabel Distefano and Weishu Fan for assistance with DNA extraction, Linda Bonen for assistance with intron classification, and Alan Christensen for discussions on translocation. This research was supported in part by funding from the University of Nebraska-Lincoln (faculty seed grant IC-2699 to JPM), the China Scholarship Council (a visiting scholar award to P-FM), the Field Museum of Chicago (a Tryon Grant to FG), and the University of Michigan (a Margaret and Herman Sokol Faculty Award to Y-LQ).

Author Contributions

JPM planned and designed research. JPM, FG, AT, and Y-LQ collected plant specimens and prepared samples for sequencing. JPM, P-FM, FG, TPM and RV performed experimental and/or computational analyses. All authors interpreted results. JPM wrote the manuscript, with input from all other authors.

References

- Abascal F, Zardoya R, Telford MJ. 2010.** TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**: W7-13.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD et al. 2012.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455-477.
- Bell NE, Boore JL, Mishler BD, Hyvönen J. 2014.** Organellar genomes of the four-toothed moss, *Tetraphis pellucida*. *BMC Genomics* **15**: 383.
- Bendich AJ. 2004.** Circular chloroplast chromosomes: the grand illusion. *Plant Cell* **16**: 1661-1666.
- Birky CW, Walsh JB. 1992.** Biased gene conversion, copy number, and apparent mutation-rate differences within chloroplast and bacterial genomes. *Genetics* **130**: 677-683.
- Blazier CJ, Guisinger MM, Jansen RK. 2011.** Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). *Plant Molecular Biology* **76**: 263-272.
- Bohnert HJ, Löffelhardt W. 1982.** Cyanelle DNA from *Cyanophora paradoxa* exists in 2 forms due to intramolecular recombination. *FEBS Letters* **150**: 403-406.
- Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. 2008.** Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *Journal of Molecular Evolution* **67**: 696-704.
- Castresana J. 2000.** Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540-552.
- Chaw S-M, Wu C-S, Sudianto E 2018.** Evolution of gymnosperm plastid genomes. In: Chaw S-M, Jansen RK eds. *Advances in Botanical Research*: Academic Press, 195-222.
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. 2006.** The complete chloroplast genome sequence of *Pelargonium x hortorum*: Organization

- and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution* **23**: 2175-2190.
- Cosner ME, Jansen RK, Palmer JD, Downie SR. 1997.** The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Current Genetics* **31**: 419-429.
- Cosner ME, Raubeson LA, Jansen RK. 2004.** Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology* **4**: 27.
- Cuenca A, Ross TG, Graham SW, Barrett CF, Davis JI, Seberg O, Petersen G. 2016.** Localized retroprocessing as a model of intron loss in the plant mitochondrial genome. *Genome Biology and Evolution* **8**: 2176-2189.
- Delannoy E, Stanley WA, Bond CS, Small ID. 2007.** Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochemical Society Transactions* **35**: 1643-1647.
- Doyle JJ, Doyle JL. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.
- Field AR, Testo W, Bostock PD, Holtum JAM, Waycott M. 2016.** Molecular phylogenetics and the morphology of the Lycopodiaceae subfamily Huperzioidae supports three genera: *Huperzia*, *Phlegmariurus* and *Phylloglossum*. *Molecular Phylogenetics and Evolution* **94**: 635-657.
- Gao L, Zhou Y, Wang ZW, Su YJ, Wang T. 2011.** Evolution of the *rpoB-psbZ* region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. *BMC Plant Biology* **11**: 64.
- Goffinet B, Wickett NJ, Werner O, Ros RM, Shaw AJ, Cox CJ. 2007.** Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). *Annals of Botany* **99**: 747-753.
- Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. 2013.** Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evolutionary Biology* **13**: 8.

- Grewe F, Viehoveer P, Weisshaar B, Knoop V. 2009.** A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Research* **37**: 5093-5104.
- Grewe F, Zhu A, Mower JP. 2016.** Loss of a *trans*-splicing *nad1* intron from Geraniaceae and transfer of the maturase gene *matR* to the nucleus in *Pelargonium*. *Genome Biology and Evolution* **8**: 3193-3201.
- Grosche C, Funk HT, Maier UG, Zauner S. 2012.** The chloroplast genome of *Pellia endiviifolia*: gene content, RNA-editing pattern, and the origin of chloroplast editing. *Genome Biology and Evolution* **4**: 1349-1357.
- Gualberto JM, Newton KJ. 2017.** Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annual Review of Plant Biology* **68**: 225-252.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307-321.
- Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP. 2014.** Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biology and Evolution* **6**: 580-590.
- Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, Mower JP. 2016.** *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Molecular Biology and Evolution* **33**: 1448-1460.
- Guo W, Mower JP. 2013.** Evolution of plant mitochondrial intron-encoded maturases: frequent lineage-specific loss and recurrent intracellular transfer to the nucleus. *Journal of Molecular Evolution* **77**: 43-54.
- Guo ZY, Zhang HR, Shrestha N, Zhang XC. 2016.** Complete chloroplast genome of a valuable medicinal plant, *Huperzia serrata* (Lycopodiaceae), and comparison with its congener. *Applications in Plant Sciences* **4**: 1600071.
- 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. *Journal of Molecular Evolution* **66**: 350-361.
- Hasebe M, Iwatsuki K. 1992.** Gene localization on the chloroplast DNA of the maiden hair fern; *Adiantum capillus-veneris*. *Botanical Magazine Tokyo* **105**: 413-419.
- Hecht J, Grewe F, Knoop V. 2011.** Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of

- frequent plant mtDNA recombination in early tracheophytes. *Genome Biology and Evolution* **3**: 344-358.
- Jansen RK, Ruhlman TA 2012.** Plastid genomes of seed plants. In: Bock R, Knoop V eds. *Genomics of Chloroplasts and Mitochondria*: Springer Netherlands, 103-126.
- Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD et al. 2010.** Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evolutionary Biology* **10**: 321.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. 2012.** Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Knox EB. 2014.** The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proceedings of the National Academy of Sciences USA* **111**: 11097-11102.
- Korall P, Kenrick P. 2004.** The phylogenetic history of Selaginellaceae based on DNA sequences from the plastid and nucleus: extreme substitution rates and rate heterogeneity. *Molecular Phylogenetics and Evolution* **31**: 852-864.
- Larsén E, Rydin C. 2016.** Disentangling the phylogeny of *Isoetes* (Isoetales), using nuclear and plastid data. *International Journal of Plant Sciences* **177**: 157-174.
- Lee HL, Jansen RK, Chumley TW, Kim KJ. 2007.** Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution* **24**: 1161-1180.
- Lee J, Cho CH, Park SI, Choi JW, Song HS, West JA, Bhattacharya D, Yoon HS. 2016.** Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology* **14**: 75.
- Li FW, Kuo LY, Pryer KM, Rothfels CJ. 2016.** Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biology and Evolution* **8**: 2452-2458.
- Liu Y, Wang B, Cui P, Li L, Xue JY, Yu J, Qiu YL. 2012.** The mitochondrial genome of the lycophyte *Huperzia squarrosa*: the most archaic form in vascular plants. *PLoS One* **7**: e35168.

- Lohse M, Drechsel O, Bock R. 2007.** OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* **52**: 267-274.
- Magallon S, Hilu KW, Quandt D. 2013.** Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100**: 556-573.
- Malek O, Lüttig K, Hiesel R, Brennicke A, Knoop V. 1996.** RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO Journal* **15**: 1403-1411.
- Milligan BG, Hampton JN, Palmer JD. 1989.** Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Molecular Biology and Evolution* **6**: 355-368.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018.** The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences USA* **115**: E2274-E2283.
- Mower JP, Vickrey TL 2018.** Structural diversity among plastid genomes of land plants. In: Chaw S-M, Jansen RK eds. *Advances in Botanical Research, Volume 85*. Cambridge, MA, USA: Academic Press, 263-292.
- Oldenburg DJ, Bendich AJ. 2016.** The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Current Genetics* **62**: 431-442.
- Oldenkott B, Yamaguchi K, Tsuji-Tsukinoki S, Knie N, Knoop V. 2014.** Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. *RNA* **20**: 1499-1506.
- Palmer JD. 1983.** Chloroplast DNA exists in two orientations. *Nature* **301**: 92-93.
- Perry AS, Wolfe KH. 2002.** Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution* **55**: 501-508.
- Pond SLK, Frost SD, Muse SV. 2005.** HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676-679.
- Posada D, Crandall KA. 1998.** MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.
- Raubeson LA, Jansen RK. 1992.** Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255**: 1697-1699.
- Raubeson LA, Stein DB. 1995.** Insights into fern evolution from mapping chloroplast genomes. *American Fern Journal* **85**: 193-204.

- Robison TA, Grusz AL, Wolf PG, Mower JP, Fauskee BD, Sosa K, Schuettpelz E. 2018.** Mobile elements shape plastome evolution in ferns. *Genome Biology and Evolution* **10**: 2558-2571.
- Ruhlman TA, Chang WJ, Chen JJ, Huang YT, Chan MT, Zhang J, Liao DC, Blazier JC, Jin X, Shih MC et al. 2015.** NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. *BMC Plant Biology* **15**: 100.
- Schuettpelz E, Schneider H, Smith AR, Hovenkamp P, Prado J, Rouhan G, Salino A, Sundue M, Almeida TE, Parris B et al. 2016.** A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution* **54**: 563-603.
- 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 accessions. *Industrial Crops and Products* **108**: 629-635.
- Smith DR. 2009.** Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Molecular Biology* **71**: 627-639.
- Song M, Kuo LY, Huiet L, Pryer KM, Rothfels CJ, Li FW. 2018.** A novel chloroplast gene reported for flagellate plants. *American Journal of Botany* **105**: 117-121.
- Stein DB, Conant DS, Ahearn ME, Jordan ET, Kirch SA, Hasebe M, Iwatsuki K, Tan MK, Thomson JA. 1992.** Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. *Proceedings of the National Academy of Sciences USA* **89**: 1856-1860.
- Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M. 2003.** Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic Acids Research* **31**: 5324-5331.
- Takenaka M, Zehrmann A, Verbitskiy D, Kugelman M, Hartel B, Brennicke A. 2012.** Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proceedings of the National Academy of Sciences USA* **109**: 5104-5109.
- Testo W, Field A, Barrington D. 2018.** Overcoming among-lineage rate heterogeneity to infer the divergence times and biogeography of the clubmoss family Lycopodiaceae. *Journal of Biogeography* **45**: 1929-1941.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017.** GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6-W11.

- Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K. 2007.** The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. *Journal of Plant Research* **120**: 281-290.
- VanBuren R, Wai CM, Ou S, Pardo J, Bryant D, Jiang N, Mockler TC, Edger P, Michael TP. 2018.** Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. *Nature Communications* **9**: 13.
- Villarreal JC, Forrest LL, Wickett N, Goffinet B. 2013.** The plastid genome of the hornwort *Nothoceros aenigmaticus* (Dendrocerotaceae): phylogenetic signal in inverted repeat expansion, pseudogenization, and intron gain. *American Journal of Botany* **100**: 467-477.
- Villarreal JC, Turmel M, Bourgouin-Couture M, Laroche J, Salazar Allen N, Li FW, Cheng S, Renzaglia K, Lemieux C. 2018.** Genome-wide organellar analyses from the hornwort *Leiosporoceros dussii* show low frequency of RNA editing. *PLoS One* **13**: e0200491.
- Vogel J, Hubschmann T, Borner T, Hess WR. 1997.** Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for MatK as an essential splice factor. *Journal of Molecular Biology* **270**: 179-187.
- 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*. *Proceedings of the National Academy of Sciences USA* **91**: 9794-9798.
- Weng ML, Blazier JC, Govindu M, Jansen RK. 2014.** Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution* **31**: 645-659.
- Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011.** The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273-297.
- Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, DePamphilis CW, Boore JL, Goffinet B. 2008.** Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* **25**: 393-401.
- Wikstrom N, Pryer KM. 2005.** Incongruence between primary sequence data and the distribution of a mitochondrial *atp1* group II intron among ferns and horsetails. *Molecular Phylogenetics and Evolution* **36**: 484-493.

- Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL. 2005.** The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene* **350**: 117-128.
- Wolfe KH, Li WH, Sharp PM. 1987.** Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* **84**: 9054-9058.
- Xu Z, Xin T, Bartels D, Li Y, Gu W, Yao H, Liu S, Yu H, Pu X, Zhou J et al. 2018.** Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Molecular Plant* **11**: 983-994.
- Zerbino DR, Birney E. 2008.** Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821-829.
- Zhou XM, Rothfels CJ, Zhang L, He ZR, Le Pechon T, He H, Lu NT, Knapp R, Lorence D, He XJ et al. 2016.** A large-scale phylogeny of the lycophyte genus *Selaginella* (Selaginellaceae: Lycopodiopsida) based on plastid and nuclear loci. *Cladistics* **32**: 360-389.
- 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 Phytologist* **209**: 1747-1756.
- Zoschke R, Nakamura M, Liere K, Sugiura M, Borner T, Schmitz-Linneweber C. 2010.** An organellar maturase associates with multiple group II introns. *Proceedings of the National Academy of Sciences USA* **107**: 3245-3250.

Tables

Table 1. General characteristics of lycophyte plastomes

	Lycopodiaceae						Isoetaceae			Selaginellaceae				
	Dobs	Ddig	Lcla	Hluc1	Hluc2	Hser	lfla	lmal	lpie	Skra	Slep	Smoe	Stam	Sunc
Length (bp)	160877	159614	151819	154373	154368	154176	145303	145535	145030	129971	114693	143775	126399	144170
Repeat (bp)	17742	16885	12417	15314	15314	15313	13118	13217	13042	14597	7308	12114	12831	12789
LSC (bp)	105928	106400	105643	104088	104083	103892	91862	91715	91748	54728	80625	83665	53170	77706
SSC (bp)	19465	19444	21342	19657	19657	19658	27205	27386	27198	46049	19452	35882	47567	40886
GC content (%)	35.0	35.7	34.5	36.2	36.3	36.3	37.9	38.0	38.0	52.3	51.9	51.0	54.0	54.8
Repeat GC (%)	42.3	43.4	46.6	44.9	44.9	44.9	48.0	47.9	48.0	56.5	57.1	55.7	55.2	57.5
LSC GC (%)	33.1	33.8	32.3	34.4	34.4	34.4	36.4	36.5	36.5	50.9	51.2	49.9	53.4	54.3
SSC GC (%)	31.9	32.9	31.3	32.8	32.8	32.8	33.4	33.3	32.7	51.5	51.3	50.5	54.0	54.3
Genes	122	122	122	122	122	122	118	118	118	85	80	93	68	93
Protein	87	87	87	87	87	87	82	82	82	71	64	76	58	77
rRNA	4	4	4	4	4	4	4	4	4	4	4	4	4	4
tRNA	31	31	31	31	31	31	32	32	32	10	12	13	6	12
Introns	22	22	22	22	22	22	21	21	21	9	7	11	7	11

LSC, large single copy. SSC, small single copy. Dobs, *Dendrolycopodium obscurum*. Ddig, *Diphasiastrum digitatum*. Lcla, *Lycopodium clavatum*. Hluc1 and Hluc2, *Huperzia lucidula*. Hser, *Huperzia serrata*. lfla, *Isoetes flaccida*. lmal, *Isoetes malinverniana*. lpie, *Isoetes piedmontana*. Skra, *Selaginella kraussiana*. Slep, *Selaginella lepidophylla*. Smoe, *Selaginella moellendorffii*. Stam, *Selaginella tamariscina*. Sunc, *Selaginella uncinata*.

Figure Legends

Fig. 1. Functional content of lycophyte plastomes. **(a)** Protein-coding genes. **(b)** rRNA genes. **(c)** tRNA genes. **(d)** Introns. Total counts are listed at the bottom of each functional category. Annotations for specific genes and introns that were corrected in this study are marked with a dark gray circle.

Fig. 2. Structural evolution of Lycopodiaceae plastomes. The loss of *trnT-GGU* is marked with a ψ . Inverted repeat (IR) expansions are denoted with red arrows, and genes affected by the IR expansion are listed in red text. The two genes arising from the split of *ycf2* are labeled.

Fig. 3. Structural evolution of *Selaginella* plastomes. Pseudogenes and lost genes are marked with a ψ and listed in black text. Inversion endpoints are marked by brown dotted lines, and the genes closest to each endpoint are listed in brown text. Inverted repeat (IR) expansions are denoted with red arrows, and genes affected by the IR expansion are listed in red text.

Fig. 4. Structural evolution of *Isoetes* plastomes. Pseudogenes and lost genes are marked with a ψ and listed in black text. Inversion endpoints are marked by brown dotted lines, and the genes closest to each endpoint are listed in brown text. Inverted repeat (IR) expansions are denoted with red arrows, and genes affected by the IR expansion are listed in red text.

Fig. 5. Substitution rate variation among lycophyte species and genes. **(a)** Synonymous divergence (d_s) tree estimated from 51 concatenated genes that are single copy (SC) in all lycophytes. **(b)** d_s tree estimated from concatenated *ndhB+psbM+rps7* genes that are duplicated in *Selaginella kraussiana*. **(c)** d_s tree estimated from *rps4*, which is duplicated in *S. moellendorffii* and *S. tamariscina*. **(d)** Overall divergence (d) tree estimated from 4 rRNA genes that are duplicated in all *Selaginella* species. Species with duplicated genes are shown in red. Terminal branch lengths for all *Selaginella* species are listed, and a ratio of branch lengths is given for each species in **(b–d)** relative to branch lengths in **(a)**. All trees are drawn to the scale shown at bottom. IR, inverted repeat; DR, direct repeat.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Verification of the direct repeat (DR) orientation in the *Selaginella kraussiana* plastome.

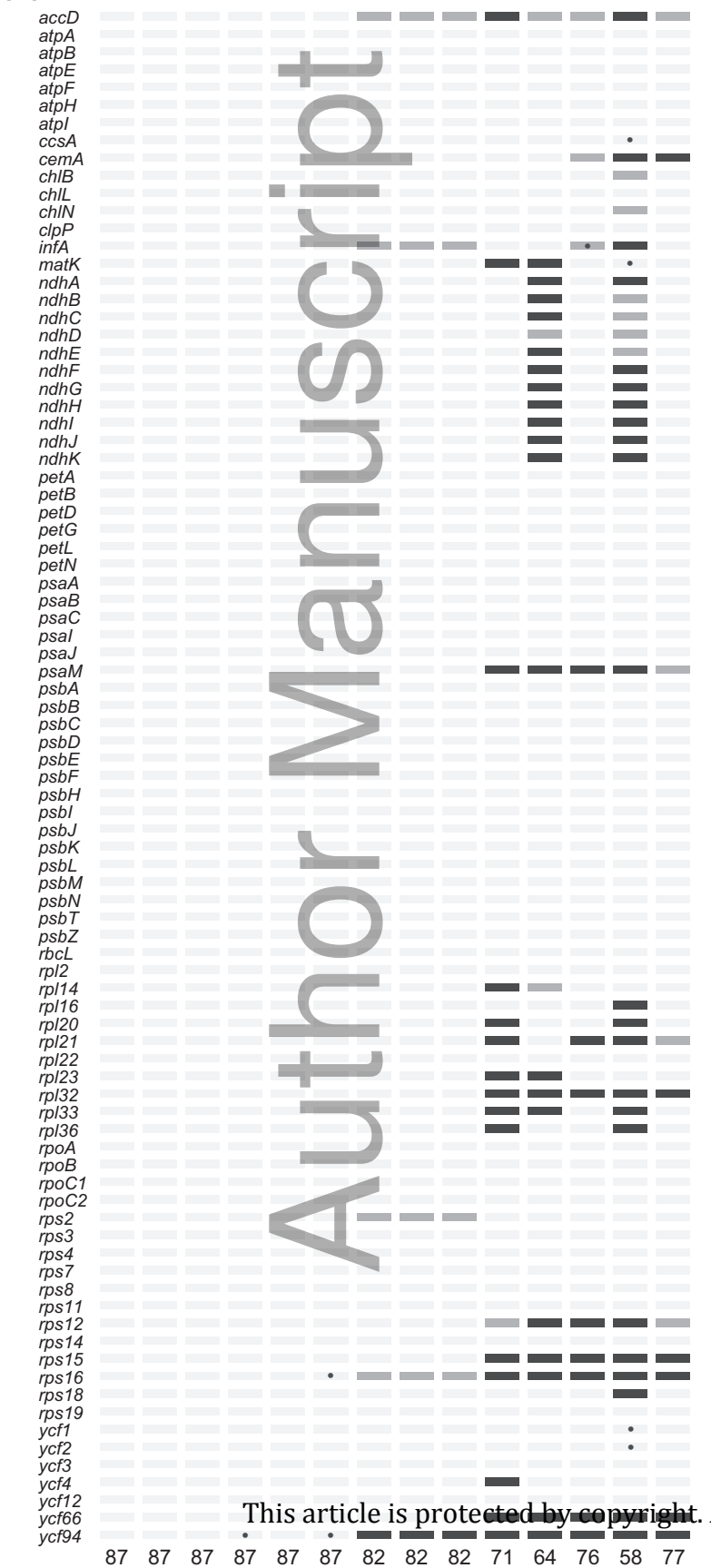
Fig. S2 Structural evolution of the *Selaginella uncinata* plastome.

Table S1 Sample and sequencing information for lycophyte plastomes.

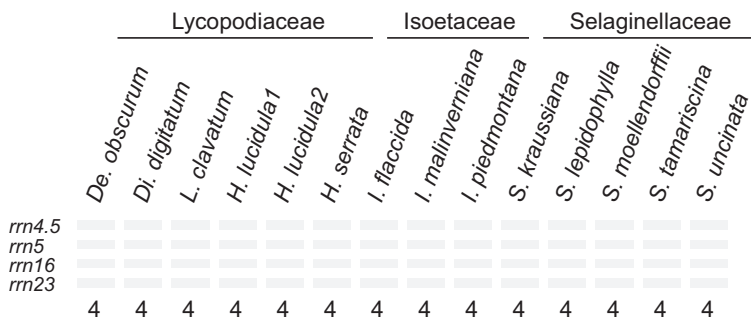
Dataset S1 DNA sequence alignments for 51 single-copy genes, 4 rRNA genes, *ndhB+psbM+rps7*, and *rps4*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

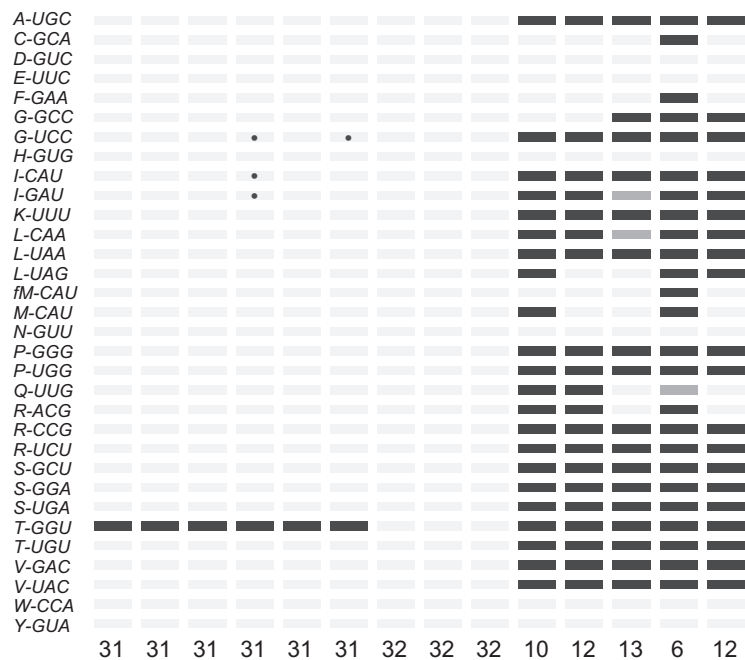
(a)



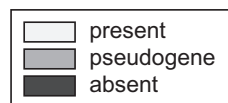
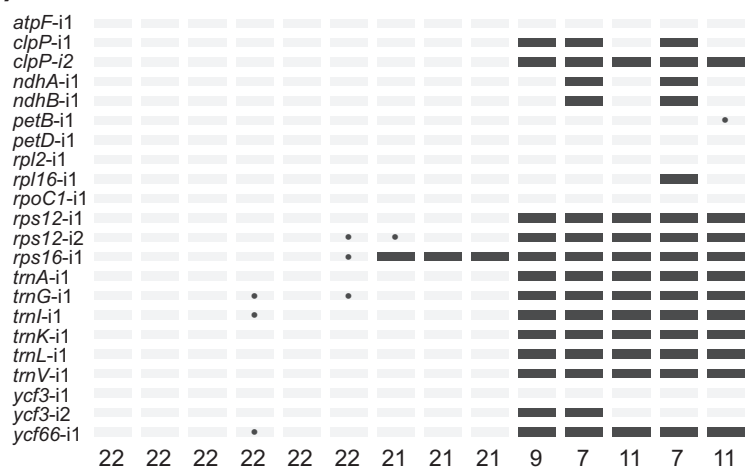
(b)



(c)



(d)



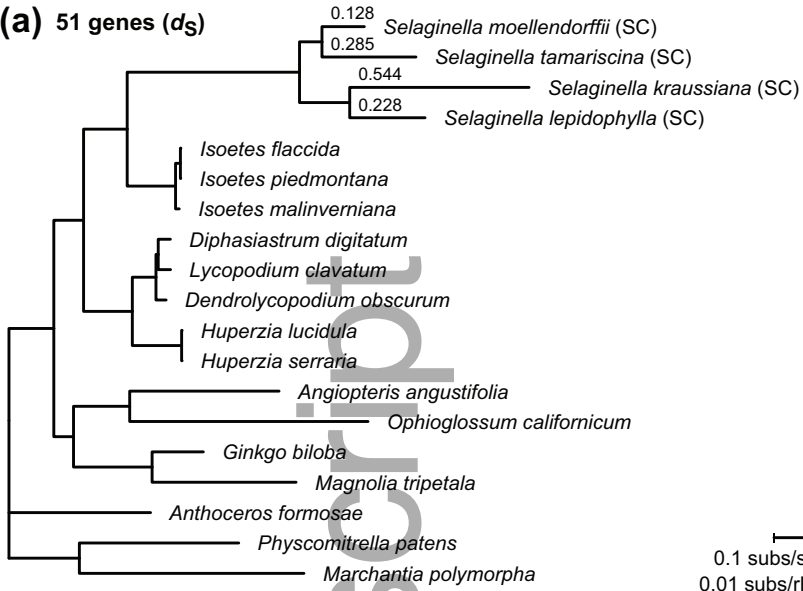
This article is protected by copyright. All rights reserved

Author Manuscript

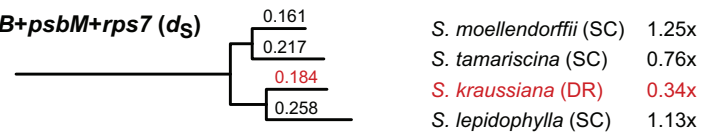
Author Manuscript

Author Manuscript

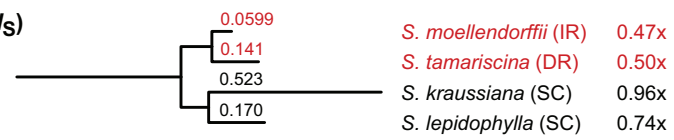
(a) 51 genes (d_S)



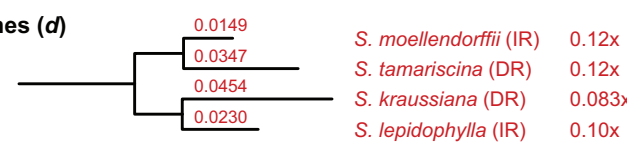
(b) $ndhB+psbM+rps7$ (d_S)



(c) $rps4$ (d_S)



(d) 4 rRNA genes (d)



0.1 subs/syn site
0.01 subs/rRNA site

nph_15650_f5.eps