University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

2017

Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns

Wenhu Guo

Andan Zhu

Weishu Fan

Jeffrey P. Mower

Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns

Wenhu Guo[1,2,3], Andan Zhu[1,4], Weishu Fan[1,4] and Jeffrey P. Mower[1,4]

[1]Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA; [2]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA; [3]ACGT Inc., Wheeling, IL 60090, USA; [4]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

Authors for correspondence:
Wenhu Guo
Tel: +1 847 520 9162 ext. 104
Email: wenhuguo@gmail.com

Jeffrey P. Mower
Tel: +1 402 472 2130
Email: jpmower@unl.edu

Summary

• Currently, complete mitochondrial genomes (mitogenomes) are available from all major land plant lineages except ferns. Sequencing of fern mitogenomes could shed light on the major evolutionary transitions that established mitogenomic diversity among extant lineages.

• In this study, we generated complete mitogenomes from the adder's tongue fern (*Ophioglossum californicum*) and the whisk fern (*Psilotum nudum*).

• The *Psilotum* mitogenome (628 kb) contains a rich complement of genes and introns, some of which are the largest of any green plant organellar genome. In the *Ophioglossum* mitogenome (372 kb), gene and intron content is slightly reduced, including the loss of all four mitochondrial *ccm* genes. Transcripts of nuclear *Ccm* genes also were not detected, suggesting loss of the entire mitochondrial cytochrome *c* maturation pathway from *Ophioglossum*. Both fern mitogenomes are highly repetitive, yet they show extremely low levels of active recombination. Transcriptomic sequencing uncovered ~1000 sites of C-to-U RNA editing in both species, plus a small number (< 60) of U-to-C edit sites.

• Overall, the first mitochondrial genomes of ferns show a mix of features shared with lycophytes and/or seed plants and several novel genomic features, enabling a robust reconstruction of the mitogenome in the common ancestor of vascular plants.

Introduction

Mitochondrial genomic diversity among flowering plants is well characterized owing to > 70 complete mitochondrial genomes (mitogenomes) that have been sequenced (as of December 2015). They can vary greatly in size, gene and intron content, RNA editing abundance, nucleotide substitution rates, DNA turnover rates (i.e. the loss of shared, homologous DNA over time), and the frequency of recombination across repeats (Sloan *et al.*, 2012; Rice *et al.*, 2013; Richardson *et al.*, 2013; Skippington *et al.*, 2015; Guo *et al.*, 2016). This rich collection of genomic data provides a basis with which to infer ancestral angiosperm features and subsequent evolutionary trends that have established the present-day diversity in mitogenome structure and content (Mower *et al.*, 2012b; Richardson *et al.*, 2013; Skippington *et al.*, 2015). By contrast, the number of sequenced mitogenomes from other vascular plant lineages is limited. Currently, only three gymnosperms (*Cycas taitungensis*, *Ginkgo biloba* and *Welwitschia mirabilis*) and three lycophytes (*Isoetes engelmannii*, *Selaginella moellendorffii* and *Huperzia squarrosa*) have been fully sequenced (Chaw *et al.*, 2008; Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012; Guo *et al.*, 2016).

Despite the limited number of mitogenomes from nonflowering vascular plants, the available genomes have provided valuable insights into the extent of mitogenomic diversity among these lineages and the evolutionary trends that have generated this diversity. For example, extensive variation in gene and RNA editing content were shown in both gymnosperms and lycophytes (Chaw *et al.*, 2008; Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012; Guo *et al.*, 2016). Intron content also is highly variable in these two groups, including the presence of numerous previously unknown group I and group II introns in lycophytes, several of which are found in novel *trans*-spliced arrangements (Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012). Regarding structural evolution, extensive rearrangements were inferred for *Isoetes*, *Selaginella* and *Welwitschia* (Grewe *et al.*, 2009; Hecht *et al.*, 2011; Guo *et al.*, 2016), whereas mitochondrial DNA turnover rates were found to be extremely slow in *Cycas*, *Ginkgo* and *Huperzia* (Liu *et al.*, 2012; Guo *et al.*, 2016).

Ferns (Monilophyta) include over 10 000 species classified into five major groups: Equisetales, Psilotales, Ophioglossales, Maratiales and Polypodidae (Christenhusz & Chase, 2014). As ferns are the sister lineage to seed plants, knowledge of fern mitogenomic features is essential for a comprehensive understanding of

evolution and diversification of vascular plant mitogenomes through comparative analyses, yet they are the only one of the seven major clades of land plants (i.e. angiosperms, gymnosperms, ferns, lycophytes, hornworts, mosses and liverworts) that lack a completely sequenced mitogenome. In a previous whole genome shotgun sequencing project, mitochondrial contigs were reported for six ferns (Wolf *et al.*, 2015). However, these incompletely sequenced and assembled mitochondrial DNAs (mtDNAs) are insufficient to address questions regarding the structure, content and evolution of plant mitogenomes in ferns.

To fill in this remaining phylogenetic gap in land plants, we sequenced and completed the first fern mtDNAs, from the adder's tongue fern (*Ophioglossum californicum*) and the whisk fern (*Psilotum nudum*). To evaluate their features in a broadscale phylogenetic context, we compared the two mitogenomes with those of lycophytes, gymnosperms and several gene-rich, representative angiosperms. These comparative genomic analyses demonstrated that the two fern mitogenomes possess both lycophyte-like and seed plant-like features. They also show several novel genomic features, including a very high proportion of repetitive sequences yet extremely low levels of active recombination, and the largest organellar genes and introns of any green plant.

Materials and Methods

Genome sequencing, assembly and annotation

Organelle-enriched DNAs of *Ophioglossum californicum* Prantl and *Psilotum nudum* (L.) P. Beauv. were isolated and extracted as described in Grewe *et al.* (2013). In addition to the sequencing data obtained in Grewe *et al.* (2013), the organelle-enriched DNAs of each species also were sequenced at BGI (Shenzhen, China) on an Illumina HiSeq 2000 platform from a single 5-kb mate-pair library, which generated 40.4 and 41.8 million 2×100 bp reads for *Ophioglossum* and *Psilotum*, respectively. The *Psilotum* organelle-enriched DNA was also 454 sequenced at the University of Nebraska Core for Applied Genomics and Ecology, producing 149 474 single-end reads with an average length of 473 bp.

The paired-end and mate-pair Illumina sequencing reads were assembled separately with VELVET v.1.210 (Zerbino & Birney, 2008). Multiple VELVET assemblies were constructed for each dataset using different pairwise combinations of Kmer values and expected coverage values, as described previously (Grewe *et al.*, 2014; Zhu *et al.*, 2014). For all runs, the scaffolding option was turned off. The 454 sequencing data were assembled by using Roche's GS *de novo* ASSEMBLER v.2.3 ('Newbler') with default parameters. SSPACE v.3.0 (Boetzer *et al.*, 2011) was used for iteratively scaffolding all assemblies with the read-pair information from the Illumina 800-bp paired-end and 5-kb mate-pair libraries. GAPFILLER v.1.10 (Nadalin *et al.*, 2012) was used *in silico* to fill most of the gaps. Remaining gaps in the assemblies were due to long mononucleotide repeats (10–20 nt in length) and were finished by Sanger sequencing. Completed

mitogenomes were deposited in GenBank under accession numbers KX171637 (*Ophioglossum*), KX171638 (*Psilotum* chromosome 1) and KX171639 (*Psilotum* chromosome 2).

Depth of sequencing coverage for the completed genomes (Supporting Information Fig. S1) was evaluated by mapping Illumina paired-end reads onto the genome sequences using BOWTIE v.2 as described previously (Guo *et al.*, 2014; Zhu *et al.*, 2014). Genes and introns were annotated using established procedures (Grewe *et al.*, 2014; Zhu *et al.*, 2014). To detect mitochondrial *ccm* and nuclear *Ccm* genes in other eusporangiate ferns, *Psilotum* mitochondrial *ccm* and *Arabidopsis* nuclear *Ccm* amino acid sequences were queried against the 1-kp database (https://www.bioinfodata.org/Blast4OneKP/) with TBLASTN searches and an e-value cut-off of 1×10^{-10} (Table S1).

Transcriptome sequencing and analysis

Organelle-enriched RNAs of both *Ophioglossum* and *Psilotum* were isolated, prepared and sequenced as described in Guo *et al.* (2015). Adapter and low-quality bases of the RNA sequencing data were trimmed as described in Guo *et al.* (2015). To eliminate the effect of plastid-derived sequences (MIPTs), the trimmed reads were mapped simultaneously to both mitochondrial and plastid genomes of *Ophioglossum* and *Psilotum*, respectively, using TOPHAT v.2.0.13 (Kim *et al.*, 2013) with relaxed parameters (-N 4 -read-gap-length 0 -read-edit-dist 4 -max-insertion-length 0 -max-deletion-length 0 -I 15000 -coverage-search). RNA edit sites then were identified using the established bioinformatics pipeline described in Guo *et al.* (2015) except the minimum read depth was increased to 10×.

Substitution rate analysis

Substitution rates were calculated using 41 mitochondrial protein-coding genes (Fig. 1a) from *Ophioglossum*, *Psilotum*, three gymnosperms (*Cycas taitungensis*, *Ginkgo biloba* and *Welwitschia mirabilis*), four slowly evolving and gene-rich angiosperms (*Amborella trichopoda*, *Liriodendron tulipifera*, *Phoenix dactylifera* and *Vitis vinifera*) and two lycophyte outgroups (*Huperzia squarrosa* and *Isoetes engelmannii*). Genes were individually aligned with MUSCLE v.3.8.31 (Edgar, 2004) using default parameters. To mitigate the confounding effects of C-to-U and U-to-C RNA editing on substitution rate calculations, empirical editing data was used for *Ophioglossum*, *Psilotum* and *Isoetes*. For seed plants and *Huperzia*, empirical editing data are not available for all genes from all species. Therefore, predicted RNA sequences from Guo *et al.* (2016) or generated by the PREP-Mt webserver (Mower, 2009) were used in the alignments. Because these species are used as outgroups relative to ferns, the use of predicted data should have no effect on rates estimated for ferns.

Poorly aligned regions were eliminated from each alignment using GBLOCKS v.0.91b (Castresana, 2000) with relaxed settings ($-t = c$, $-b1 = h$, $-b2 = h$, $-b4 = 5$, $-b5 = h$). Filtered alignments were concatenated with FASCONCAT v.1.0 (Kück & Meusemann, 2010), generating a final alignment of 33 639 bp.

(a)

Gene	Lycophytes — Huperzia	Isoetes	Selaginella	Ferns — Ophioglossum	Psilotum	Gymnosperms — Cycas	Ginkgo	Welwitschia	Angiosperms — Amborella	Liriodendron	Phoenix	Vitis
atp1	●	●	●	●	●	●	●	●	●	●	●	●
atp4	●	●	–	●	●	●	●	●	●	●	●	●
atp6	●	●	●	●	●	●	●	●	●	●	●	●
atp8	●	●	●	●	●	●	●	●	●	●	●	●
atp9	●	●	●	●	●	●	●	●	●	●	●	●
ccmB	–	–	–	–	●	●	●	●	●	●	●	●
ccmC	–	–	–	–	●	●	●	●	●	●	●	●
ccmF$_C$	ψ	–	–	–	●	●	●	●	●	●	●	●
ccmF$_N$	–	–	–	–	●	●	●	●	●	●	●	●
cob	●	●	●	●	●	●	●	●	●	●	●	●
cox1	●	●	●	●	●	●	●	●	●	●	●	●
cox2	●	●	●	●	●	●	●	●	●	●	●	●
cox3	●	●	●	●	●	●	●	●	●	●	●	●
matR	●	–	–	●	●	●	●	●	●	●	●	●
mat-atp9i87	–	●	●	●	●	–	–	–	–	–	–	–
mat-atp9i95	●	–	–	●	●	–	–	–	–	–	–	–
mat-cobi787	●	–	–	●	●	–	–	–	–	–	–	–
mttB	●	●	●	●	●	●	●	●	●	●	●	●
nad1	●	●	●	●	●	●	●	●	●	●	●	●
nad2	●	●	●	●	●	●	●	●	●	●	●	●
nad3	●	●	●	●	●	●	●	●	●	●	●	●
nad4	●	●	●	●	●	●	●	●	●	●	●	●
nad4L	●	●	●	●	●	●	●	●	●	●	●	●
nad5	●	●	●	●	●	●	●	●	●	●	●	●
nad6	●	●	●	●	●	●	●	●	●	●	●	●
nad7	–	●	●	●	●	●	●	●	●	●	●	●
nad9	●	●	●	●	●	●	●	●	●	●	●	●
rpl2	●	–	–	●	●	●	●	–	●	●	●	●
rpl5	●	●	–	●	●	●	●	–	●	●	●	●
rpl6	●	–	–	●	●	–	–	–	–	–	–	–
rpl10	●	–	–	–	–	●	●	●	–	●	●	●
rpl16	●	–	–	●	●	●	●	ψ	●	●	●	●
rps1	ψ	ψ	–	●	●	●	●	–	●	●	●	●
rps2	●	●	–	●	●	●	●	–	●	●	●	–
rps3	●	●	–	●	●	●	●	●	●	●	●	●
rps4	●	●	–	●	●	●	●	●	●	●	●	●
rps7	–	–	–	●	●	●	●	–	●	●	●	●
rps8	●	–	–	–	–	–	–	–	–	–	–	–
rps10	●	–	–	–	–	●	●	–	●	●	●	●
rps11	●	–	–	–	–	●	●	–	●	●	●	–
rps12	●	–	–	●	●	●	●	●	●	●	●	●
rps13	●	–	–	●	●	●	●	–	●	●	●	●
rps14	●	–	–	●	●	●	●	–	●	●	●	●
rps19	●	–	–	–	–	●	●	–	●	●	●	●
sdh3	●	●	–	●	●	●	●	–	●	●	–	●
sdh4	●	–	–	●	●	●	●	●	●	●	ψ	●
total (●)	**38**	**25**	**19**	**37**	**41**	**41**	**41**	**29**	**40**	**41**	**39**	**39**

(b)

	Nuclear — CcmA	CcmE	CcmH	Mitochondrial — ccmB	ccmC	ccmF$_C$	ccmF$_N$
Equisetales — *Equisetum*	gray	gray	gray	gray	gray	gray	gray
Marattiales — *Angiopteris*	white	gray	gray	gray	gray	gray	gray
Marattia	gray	gray	gray	gray	gray	gray	gray
Ophioglossales — *O. californicum*	white	gray	gray	gray	gray	gray	gray
O. vulgatum	white	gray	gray	gray	gray	gray	gray
Botrypus	gray	gray	gray	gray	gray	white	gray
Sceptridium	gray	gray	gray	gray	gray	white	gray
Psilotales — *Psilotum*	gray	gray	gray	gray	gray	gray	gray
Tmesipteris	gray	gray	gray	gray	gray	gray	gray

(c)

Evolutionary timing of mitochondrial gene losses (chronogram, 500–0 Ma):

- **46 Genes** (ancestral node)
- ccmB, ccmC, ccmF$_C$, ccmF$_N$, rps1, rps7 → (lycophyte lineage)
- mat-atp9i87, nad7 → *Huperzia*
- atp4, rpl5, rps2, rps3, rps4, sdh3 → *Selaginella*
- matR, mat-atp9i95, mat-cobi787, rpl2, rpl6, rpl10, rpl16, rps8, rps10, rps11, rps12, rps13, rps14, rps19, sdh4 → *Isoetes*
- ccmB, ccmC, ccmF$_C$, ccmF$_N$ → *Ophioglossum*
- rpl10 → *Psilotum*
- mat-atp9i87, mat-atp9i95, mat-cob787, rps8 (seed plant lineage)
- rpl6 (internal)
- rpl2, rpl5, rpl16, rps1, rps2, rps7, rps10, rps11, rps13, rps14, rps19, sdh3 → *Welwitschia*
- rpl10 → *Amborella*
- rps2, rps11 → *Vitis*
- sdh3, sdh4 → *Phoenix*

Taxa: *Phoenix*, *Vitis*, *Liriodendron*, *Amborella*, *Cycas*, *Ginkgo*, *Welwitschia*, *Ophioglossum*, *Psilotum*, *Isoetes*, *Selaginella*, *Huperzia*, Nonvascular Plants.

Time axis: 500 400 300 200 100 0 Ma

Fig. 1 (a) Mitochondrial protein-coding gene content among representative vascular plants. ●, indicates presence of an intact gene; ψ, a pseudogene; –, gene loss. (b) Nuclear *Ccm* and mitochondrial *ccm* gene content in eusporangiate ferns. Gray boxes, presence of a gene; white boxes, absence of a gene. (c) Evolutionary timing of mitochondrial gene losses in vascular plants. Gene losses were mapped onto a chronogram using maximum parsimony. The 46 ancestral genes are listed in (a). Ma, million years.

Trees representing synonymous (d_S) and nonsynonymous (d_N) branch lengths were calculated in HYPHY v.2.2.1 (Pond *et al.*, 2005) with a local MG94xHKY85 codon model. Topologies were constrained according to version 13 of the Angiosperm Phylogeny Website (http://www.mobot.org/mobot/research/apweb/). Absolute rates of synonymous (R_S) and nonsynonymous (R_N)

substitution were calculated by dividing d_S or d_N branches by their estimated divergence times reported by the TimeTree web service (http://www. timetree.org/).

Repeats and repeat-mediated recombination

Repeats in both mitogenomes were identified with BLAST searches using different sets of parameters. For comparison of repeat content across vascular plants, a word size of 7 and an e-value cut-off of 1×10^{-6} was used, which is very sensitive and can detect repeats as short as 30 bp (perfect match) or 33 bp (one mismatch). For additional comparisons with several highly repetitive mitogenomes, a word size of 7 and an e-value cut-off of 1 or a minimum raw score of 19 was used to detect repeats down to 19 bp. In addition to the BLAST analyses, short dispersed repeats up to several hundred bp in length were classified into repeat families using REPEATSCOUT v.1.0.5 (Price *et al.*, 2005) and REPEATMASKER v.4.0.5 (http://www.repeatmasker.org). Short tandem repeats up to 30 bp were identified using MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html), with the following search parameters: minimum number of repeats for mononucleotide and dinucleotide was set to 5, for trinucleotide to octal-nucleotide was set to 3 and for all others was set to 2.

Repeat-mediated recombination was evaluated for all nonoverlapping repeats between 100 bp and 5 kb in both mitogenomes by using established procedures (Guo *et al.*, 2016). As an independent measure of overall recombination, we counted the frequency of all improperly mapped read pairs, which did not map to the genome in the expected tail-to-tail orientation at a distance of 5 kb ($\pm 100\%$). Excluded from this count were all read pairs that mapped in a head-to-head orientation at a distance of < 1 kb, which are likely derived from unbiotinylated fragments of contiguous DNA that were not washed away completely during the library preparation procedure (http://support.illumina.com/downloads/mate_pair_v2_sample_prep_guide_for_2-5_kb_libraries_15008135.html).

Gene order and DNA turnover analyses

Gene orders for *Physcomitrella*, *Megaceros*, *Huperzia*, *Ophioglossum* and *Psilotum* were taken from their annotated GenBank files. Genes were color-coded according to the ancestral land plant gene orders determined by Liu *et al.* (2012).

Rates of DNA turnover were estimated by plotting the amount of shared DNA between two mitogenomes as a function of divergence time, as described previously (Guo *et al.*, 2016). Briefly, to obtain the shared amount of mtDNA between *Ophioglossum* and *Psilotum*, the two mitogenomes were aligned using BLASTN with a word size of 7 and an e-value cut-off of 1×10^{-6}. The divergence time between *Ophioglossum* and *Psilotum* was estimated to 305 Myr ago (Ma) using the TimeTree website (http://timetree.org/). These shared DNA and divergence time values were plotted with data obtained previously (Guo *et al.*, 2016).

Results

Genome size

The *Ophioglossum* mitogenome was assembled into a single circular molecule, with a genome size of 372 kb, whereas the *Psilotum* mitogenome was assembled into two circular molecules with sizes of 364 and 264 kb (Table 1; Fig. S1). The two chromosomes in *Psilotum* mitochondria share two large repeats (2.7 and 3.0 kb), and have essentially the same sequencing depth (Fig. S1B). Despite the shared presence of two large repeats, read-pair mapping shows that the two chromosomes rarely recombine with each other (see the 'Genome structural dynamics' section below), providing strong support for the two chromosome conformation rather than the typical single chromosome assembled in most plant mitogenomic studies. The *Ophioglossum* mitogenome is similar in size to that of the lycophyte *Huperzia* (414 kb), the gymnosperms *Cycas* (415 kb) and *Ginkgo* (347 kb), and many angiosperms, whereas the mitogenome size of *Psilotum* is larger than > 75% of sequenced vascular plants (as of December 2015). In comparison with *Ophioglossum*, the larger mitogenome in *Psilotum* is due primarily to a major increase of repetitive sequences (138 kb vs 331 kb), and, to a lesser extent, to increased gene space (34 kb vs 51 kb) and intronic regions (58 kb vs 102 kb). MIPTs (mitochondrial DNA of plastid origin) are low in both species (1.8 kb vs 0.1 kb), contributing very little to overall size (Table 1).

Gene content

The *Ophioglossum* and *Psilotum* mitogenomes have similar coding content to other vascular plants, with a few notable distinctions (Fig. 1a). Compared with the 41 protein-coding genes inferred to be present in the ancestral mitogenome of seed plants, *Ophioglossum* and *Psilotum* have an additional *rpl6* gene that is shared with *Huperzia*, but both species lack the *rpl10* gene that is present in most vascular plants. In addition, *Ophioglossum* lacks all four *ccm* genes (*ccmB*, *ccmC*, *ccmF_C* and *ccmF_N*). Transcripts also were not detected for any of the nucleus-encoded members of this pathway (i.e. the nuclear *Ccm* genes) in *Ophioglossum*, whereas most or all of these genes were detected in genomic and transcriptomic data from other members of Ophioglossales and more distantly related ferns (Fig. 1b; Table S1). Thus, the entire pathway may have been lost specifically from *Ophioglossum*.

Lycophytes have a few additional mitochondrial genes not found in other vascular plants (Fig. 1a), including several novel maturase genes (Guo & Mower, 2013) as well as a full-length *rps8* gene that was reported previously as a pseudogene in *Huperzia* (Liu *et al.*, 2012). However, the *Huperzia rps8* gene is found within a highly conserved cluster of ribosomal protein genes that is syntenic with liverworts and mosses. In addition, by assuming C-to-U and U-to-C RNA editing, processes which are active in the *Huperzia* mitogenome (Liu *et al.*, 2012), both a start codon can be created and a premature stop codon can be removed (Fig. S2). The gene also is transcribed, based on the recovery of an expressed sequence tag from a *Huperzia* cDNA

Table 1 General mitogenomic features of representative vascular plants

	Lycophytes			Ferns		Gymnosperms			Angiosperms	
	Huperzia	*Isoetes*	*Selagin.*	*Ophiogl.*	*Psilotum*	*Cycas*	*Ginkgo*	*Welwits.*	*Lirioden.*	*Phoenix*
Size (bp)	413 530	> 57 571	> 183 000	372 339	628 553	414 903	346 544	978 846	553 721	715 001
GC%	44.2	48.7	68.1	52.2	51.2	46.9	50.4	53.0	47.7	45.1
Genes[a]	67	41	21	63	68	67	64	40	65	64
tRNA	27	13	0	23	24	23	20	8	21	22
rRNA	3	3	2	3	3	3	3	3	3	3
Protein coding	37	25	19	37	41	41	41	29	41	39
Introns	32	30	37	20	26	26	25	10	25	25
No. of large repeats (> 1 kb)	6	NA	NA	4	5	5	3	0	5	1
Repeats (kb)	114.1	NA	NA	138.1	331.4	113.4	39.1	72.5	86.0	28.4
Tandem repeats (kb)	6.3	NA	NA	40.1	24.7	22.4	3.6	24.4	2.0	1.8
Plastid-derived DNA (kb)	0.0	12.0	0.0	1.8	0.1	18.0	0.5	8.5	28.0	74.0
No. of C-to-U edit sites	~200	1560	2152	1014	965	1214	1306	226	>781	592
No. of U-to-C edit sites	~200	222	0	58	19	0	0	0	0	0

[a]Duplicate gene copies were not counted.
Selagin., *Selaginella*; *Ophiogl.*, *Ophioglossum*; *Welwits.*, *Welwitschia*; *Lirioden.*, *Liriodendron*; NA, not available.

library (GenBank accession number GO912990). These lines of evidence indicate that *rps8* probably is functional in *Huperzia*.

Altogether, there are 46 protein-coding genes present in at least one lineage of vascular plants (Fig. 1a). Because orthologs for all 46 genes also are detected in at least some nonvascular plants, this implies that all of these genes were present in the common ancestor of vascular plants, followed by lineage-specific loss of some genes from particular vascular plant lineages (Fig. 1c). Many of the gene losses affect ribosomal protein genes and *sdh* genes, consistent with previous observations for angiosperms (Adams *et al.*, 2002).

With regard to RNA gene content (Table S2), *Ophioglossum* and *Psilotum* have the typical set of three rRNA genes that also are found in nearly all land plants, with two copies of *rrn18* and *rrn26* present in *Psilotum* due to a 25-kb repeat. The tRNA content includes 24 genes in *Ophioglossum* and 27 genes in *Psilotum*, most of which also are present in other vascular plants, except for a *trnR*-UCG gene that is found only in the two ferns. The closest homolog to this *trnR*-UCG gene was detected in several bacteria from Chlamydiales, suggesting that this fern tRNA was acquired via horizontal transfer, as postulated also for several other mitochondrial tRNAs in some lycophytes and ferns (Knie *et al.*, 2015).
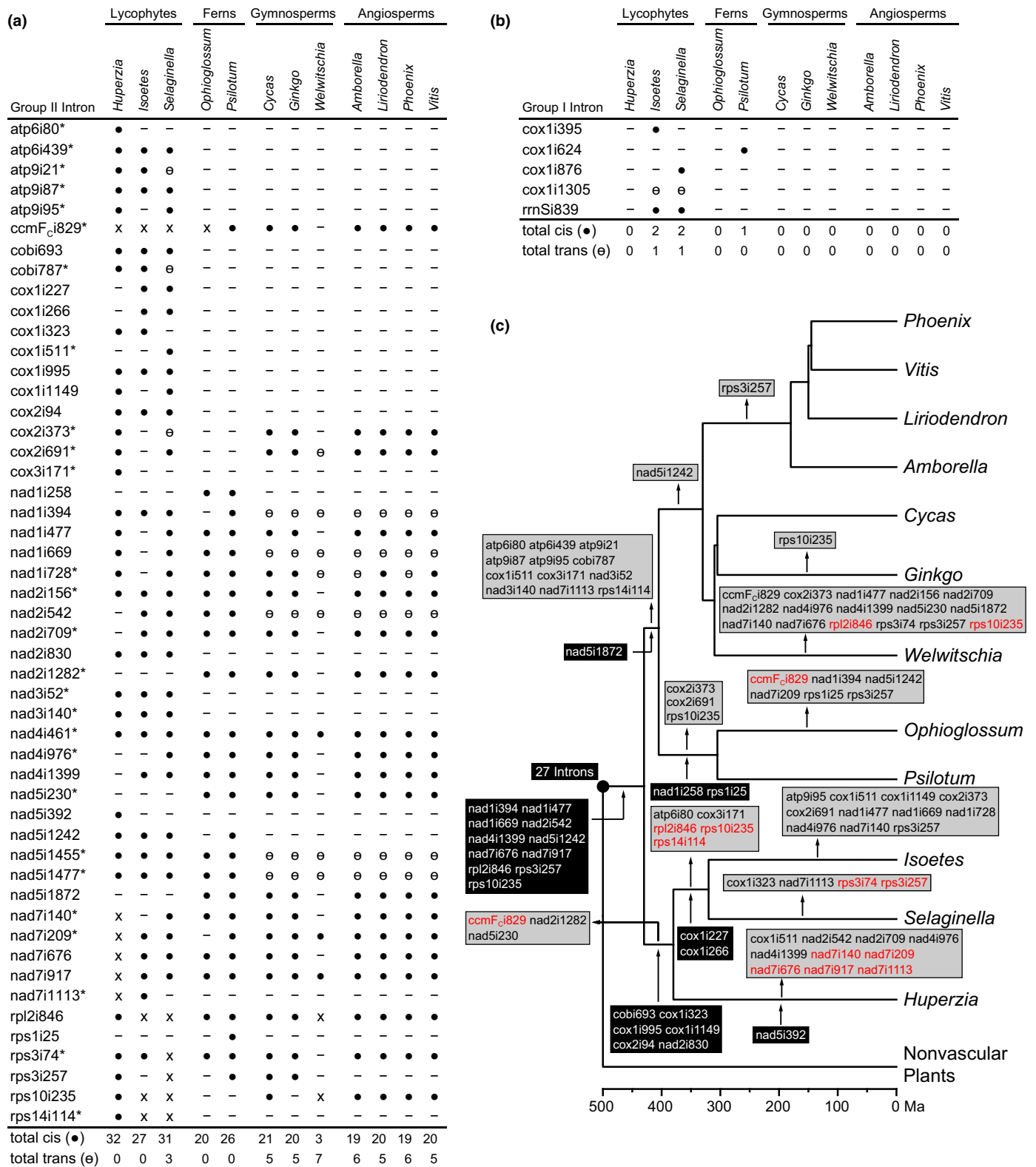
Intron content

Mitochondrial intron content shows a more variable pattern among vascular plants, due mainly to many novel introns in lycophytes (Fig. 2). The overall intron distribution pattern in ferns is more similar to seed plants than to lycophytes. Among group II introns (Fig. 2a), the predominant type of intron in land plants, there is only a single intron from ferns (*Psilotum* nad5i1242) that is shared with lycophytes but not seed plants. The two ferns contain a novel group II intron (nad1i258), and *Psilotum* contains another novel intron (rps1i25), both of which were shown

previously to be unique to ferns (Dombrovska & Qiu, 2004; Knie *et al.*, 2016). Group I introns are much more restricted in vascular plants (Fig. 2b). *Psilotum* contains a group I intron (cox1i624) that is not shared with any vascular plants, although an ortholog is present in mosses and liverworts (Liu *et al.*, 2011, 2014).

Mapping group II intron gains and losses in a phylogenetic context shows that intron loss predominated during vascular plant evolution (Fig. 2c). Several introns, including nad1i394, nad5i1242, nad7i209 and rps3i257, are absent from *Ophioglossum* but present in *Psilotum* and many other vascular plants, indicating a loss from the *Ophioglossum* lineage. The ccmF$_C$i829 intron was also lost from *Ophioglossum* due to loss of its host gene. Gain and loss patterns are more difficult to infer for group I introns. The much more restricted distribution of group I introns, coupled with the sporadic presence of homologs in nonvascular plants or green algae, makes it uncertain whether their patchy distribution is due to inheritance from a common ancestor coupled with extensive loss or to horizontal intron transfer, as proposed for the acquisition of *cox1* group I intron in angiosperms from fungi (Vaughn *et al.*, 1995; Adams *et al.*, 1998; Sanchez-Puerta *et al.*, 2008).

RNA editing abundance and efficiency

Using a previously established bioinformatics pipeline (Guo *et al.*, 2015) with slightly modified parameters, a total of 1014 C-to-U and 58 U-to-C edit sites were detected in the *Ophioglossum* mitochondrial transcriptome, whereas 965 C-to-U and 19 U-to-C edit sites were identified in the *Psilotum* mitochondrial transcriptome (Tables 2, S3). In both species, the majority of edit sites are found in coding regions, and about half of those are located at second codon positions. For multiple genes, C-to-U RNA editing is required to create start codons and/or stop codons, and numerous internal stop codons were removed by

(a)

Group II Intron	Lycophytes			Ferns		Gymnosperms			Angiosperms			
	Huperzia	Isoetes	Selaginella	Ophioglossum	Psilotum	Cycas	Ginkgo	Welwitschia	Amborella	Liriodendron	Phoenix	Vitis
atp6i80*	●	–	–	–	–	–	–	–	–	–	–	–
atp6i439*	●	●	●	–	–	–	–	–	–	–	–	–
atp9i21*	●	●	ө	–	–	–	–	–	–	–	–	–
atp9i87*	●	●	●	–	–	–	–	–	–	–	–	–
atp9i95*	●	–	●	–	–	–	–	–	–	–	–	–
ccmF$_C$i829*	x	x	x	x	●	●	●	–	●	●	●	●
cobi693	●	●	●	–	–	–	–	–	–	–	–	–
cobi787*	●	●	ө	–	–	–	–	–	–	–	–	–
cox1i227	–	●	●	–	–	–	–	–	–	–	–	–
cox1i266	–	●	●	–	–	–	–	–	–	–	–	–
cox1i323	●	●	–	–	–	–	–	–	–	–	–	–
cox1i511*	–	●	●	–	–	–	–	–	–	–	–	–
cox1i995	●	●	●	–	–	–	–	–	–	–	–	–
cox1i1149	●	–	●	–	–	–	–	–	–	–	–	–
cox2i94	●	●	●	–	–	–	–	–	–	–	–	–
cox2i373*	●	–	ө	–	–	●	●	–	●	●	●	●
cox2i691*	●	–	●	–	–	●	●	ө	●	●	●	●
cox3i171*	●	–	–	–	–	–	–	–	–	–	–	–
nad1i258	–	–	–	●	●	–	–	–	–	–	–	–
nad1i394	●	●	●	–	●	ө	ө	ө	ө	ө	ө	ө
nad1i477	●	●	●	●	●	●	●	–	●	●	●	●
nad1i669	●	–	●	●	●	ө	ө	ө	ө	ө	ө	ө
nad1i728*	●	●	●	–	●	–	●	ө	●	●	●	●
nad2i156*	–	●	●	●	●	●	●	–	●	●	●	●
nad2i542	–	●	●	●	●	ө	ө	ө	ө	●	ө	ө
nad2i709*	–	●	●	●	●	●	●	●	●	●	●	●
nad2i830	●	●	●	–	–	–	–	–	–	–	–	–
nad2i1282*	–	–	–	●	●	●	●	●	●	●	●	●
nad3i52*	●	●	●	–	–	–	–	–	–	–	–	–
nad3i140*	●	●	●	●	●	–	–	–	–	–	–	–
nad4i461*	●	●	●	●	●	●	●	–	●	●	●	●
nad4i976*	●	●	●	●	●	ө	ө	–	ө	ө	ө	ө
nad4i1399	–	●	●	●	●	●	●	–	●	●	●	●
nad5i230*	–	–	●	●	●	●	●	●	●	●	●	●
nad5i392	●	–	●	●	●	–	–	–	–	–	–	–
nad5i1242	●	●	●	●	●	–	–	–	–	–	–	–
nad5i1455*	●	●	●	●	●	ө	ө	ө	ө	ө	ө	ө
nad5i1477*	●	●	●	●	●	ө	ө	ө	ө	ө	ө	ө
nad5i1872	–	–	–	●	●	●	●	●	●	●	●	●
nad7i140*	x	–	●	●	●	●	●	–	●	●	●	●
nad7i209*	x	●	●	–	●	●	●	●	●	●	●	●
nad7i676	x	●	●	●	●	●	●	–	●	●	●	●
nad7i917	x	●	●	●	●	●	●	–	●	●	●	●
nad7i1113*	x	●	–	●	●	●	●	–	●	●	●	●
rpl2i846	●	x	x	●	●	●	●	x	●	●	●	●
rps1i25	–	–	–	–	●	–	–	–	–	–	–	–
rps3i74*	●	●	x	●	●	●	●	–	●	●	●	●
rps3i257	●	–	x	●	●	●	●	–	●	●	●	●
rps10i235	●	x	x	–	–	●	–	x	●	●	●	●
rps14i114*	●	x	x	–	–	–	–	–	–	–	–	–
total cis (●)	32	27	31	20	26	21	20	3	19	20	19	20
total trans (ө)	0	0	3	0	0	5	5	7	6	5	6	5

(b)

Group I Intron	Lycophytes			Ferns		Gymnosperms			Angiosperms			
	Huperzia	Isoetes	Selaginella	Ophioglossum	Psilotum	Cycas	Ginkgo	Welwitschia	Amborella	Liriodendron	Phoenix	Vitis
cox1i395	–	●	–	–	–	–	–	–	–	–	–	–
cox1i624	–	–	–	–	●	–	–	–	–	–	–	–
cox1i876	–	–	●	–	–	–	–	–	–	–	–	–
cox1i1305	–	ө	ө	–	–	–	–	–	–	–	–	–
rrnSi839	–	●	●	–	–	–	–	–	–	–	–	–
total cis (●)	0	2	2	0	1	0	0	0	0	0	0	0
total trans (ө)	0	1	1	0	0	0	0	0	0	0	0	0

(c) *Evolutionary timing of intron gains and losses (chronogram).*

Tips (top to bottom): Phoenix, Vitis, Liriodendron, Amborella, Cycas, Ginkgo, Welwitschia, Ophioglossum, Psilotum, Isoetes, Selaginella, Huperzia, Nonvascular Plants.

Time axis: 500 – 400 – 300 – 200 – 100 – 0 Ma

Boxes along the tree:

- rps3i257
- nad5i1242
- (gray) atp6i80 atp6i439 atp9i21 atp9i87 atp9i95 cobi787 cox1i511 cox3i171 nad3i52 nad3i140 nad7i1113 rps14i114
- (gray) rps10i235
- (gray) ccmF$_C$i829 cox2i373 nad1i477 nad2i156 nad2i709 nad2i1282 nad4i976 nad4i1399 nad5i230 nad5i1872 nad7i140 nad7i676 *rpl2i846* rps3i74 rps3i257 *rps10i235*
- (black) nad5i1872
- (gray) cox2i373 cox2i691 rps10i235
- (gray) *ccmF$_C$i829* nad1i394 nad5i1242 nad7i209 rps1i25 rps3i257
- ● 27 Introns
- (black) nad1i258 rps1i25
- (gray) atp6i80 cox3i171 *rpl2i846 rps10i235 rps14i114*
- (gray) atp9i95 cox1i511 cox1i1149 cox2i373 cox2i691 nad1i477 nad1i669 nad1i728 nad4i976 nad7i140 rps3i257
- (black) nad1i394 nad1i477 nad1i669 nad2i542 nad4i1399 nad5i1242 nad7i676 nad7i917 rpl2i846 rps3i257 rps10i235
- (gray) *ccmF$_C$i829* nad2i1282 nad5i230
- (gray) cox1i323 nad7i1113 *rps3i74 rps3i257*
- (black) cox1i227 cox1i266
- (gray) cox1i511 nad2i542 nad2i709 nad4i976 nad4i1399 *nad7i140 nad7i209 nad7i676 nad7i917 nad7i1113*
- (black) cobi693 cox1i323 cox1i995 cox1i1149 cox2i94 nad2i830
- (black) nad5i392

Fig. 2 Mitochondrial group II intron (a) and group I intron (b) content among representative vascular plants. ●, indicates presence of a *cis*-splicing intron; ө, a *trans*-splicing intron; x, indicates intron loss due to gene loss. *, Intron has homologs in nonvascular plants and/or streptophytic green algae, indicating that the intron was acquired before the divergence of nonvascular plants and vascular plants. (c) Evolutionary timing of mitochondrial intron gains (shaded in black background) and losses (gray background) in vascular plants. Red text, intron losses were due to loss of the host gene. Intron gains and losses were mapped onto a chronogram using maximum parsimony. The 27 ancestral introns are marked with * in (a). Ma, million years.

U-to-C RNA editing (Table S3). The majority of coding edit sites are nonsilent and they are more efficiently edited on average than silent edit sites, consistent with a greater level of selective constraint to maintain nonsilent editing (Fig. 3a).

RNA editing was less common in noncoding regions (Table 2). Most of the noncoding edit sites affect intergenic regions, presumably within UTRs or intergenic spacers, whereas

a small fraction is located within introns. In *Psilotum*, all three ribosomal RNA genes were altered, each by a single C-to-U editing event, whereas no editing events were detected in any *Ophioglossum* ribosomal RNAs. No RNA editing was detected in any transfer RNAs of either species. In general, most of these noncoding edit sites are edited very inefficiently, as also reported previously for the angiosperm *Silene noctiflora* (Wu *et al.*, 2015), suggesting a lower level of selective constraint to maintain them (Fig. 3a).

Of the 862 and 731 edit sites in coding regions of *Ophioglossum* and *Psilotum*, respectively, only 186 sites are shared at homologous positions (Table S4). This highly lineage-specific distribution of edit sites, which was also observed in the plastid transcriptomes of both species (Guo *et al.*, 2015), affects all of the mitochondrial genes to a large degree, indicating a rapid turnover rate due to frequent gain and loss of RNA edit sites during fern diversification. The enrichment of shared edit sites in the high editing efficiency category (75–100%) suggests a higher level of functional importance for the shared sites (Fig. 3b, $P < 1 \times 10^{-9}$, χ^2 test).

Nucleotide composition and DNA substitution rates

The *Ophioglossum* and *Psilotum* mitogenomes are GC-rich, with GC percentages (GC%) of 52.2% and 51.2%, respectively (Table 1). These mitochondrial GC% values are higher than all other land plants except *Welwitschia* (53.0%) and *Selaginella*

Table 2 Summary of mitochondrial RNA editing events in *Ophioglossum* and *Psilotum*

	Ophioglossum		*Psilotum*	
Total	1072		984	
C-to-U	1014	94.6%	965	98.1%
U-to-C	58	5.4%	19	1.9%
Coding	862	80.4%	731	74.3%
1st	225	21.0%	213	21.6%
2nd	479	44.7%	371	37.7%
3rd	158	14.7%	147	14.9%
Start created	10	0.9%	6	0.6%
Conventional stop restored	7	0.7%	6	0.6%
Premature stop removed	39	3.6%	12	1.2%
Premature stop introduced	1	0.1%	4	0.4%
Noncoding	210	19.6%	253	25.7%
Intron	23	2.1%	44	4.5%
rRNA	0	0.0%	3	0.3%
tRNA	0	0.0%	0	0.0%
UTR/intergenic	187	17.4%	206	20.9%

Fig. 3 (a) Frequency of RNA editing in *Ophioglossum* and *Psilotum* at nonsilent, silent, and noncoding sites. (b) Shared and lineage-specific RNA edit sites as a function of editing efficiency. (c) Correlation between editing abundance and genome GC%, gene GC%, and second codon position GC%. Linear regression analyses either included (orange line, upper equation) or excluded (blue line, lower equation) the *Selaginella* data point, which is a possible outlier.

(68.1%) (Hecht *et al.*, 2011; Guo *et al.*, 2016). The similar GC % between the two ferns indicates that their common ancestor also probably possessed a GC-rich mitogenome. Because previous studies have shown that editing abundance positively correlates with either genomewide GC% (Malek *et al.*, 1996; Smith, 2009; Hecht *et al.*, 2011) or GC% at second codon positions (Guo *et al.*, 2016), we correlated editing abundance in selected ferns, lycophytes and seed plants with GC% in the whole genome, in gene sequences and at second codon positions (Fig. 3c). These results demonstrate that editing abundance correlates more closely with second position GC% ($R^2 = 0.78$) and gene GC% ($R^2 = 0.63$) than with genomewide GC% ($R^2 = 0.48$). Excluding *Selaginella*, which may be an outlier in the data, weakens the correlation of editing abundance with genomewide GC% ($R^2 = 0.07$) and gene GC% ($R^2 = 0.41$), but the correlation with second position GC% remains strong ($R^2 = 0.79$).

Using the 41 protein-coding genes present in the majority of vascular plants examined in this study (Fig. 1a), rates of synonymous and nonsynonymous substitution were estimated for *Ophioglossum* and *Psilotum*. The *Ophioglossum* mitogenome has experienced slightly higher levels of both synonymous (d_S) and nonsynonymous (d_N) sequence divergence than *Psilotum* (Fig. 4). With an estimated divergence time of 305 Ma, absolute synonymous (R_S) and nonsynonymous (R_N) substitution rates are estimated to be 0.72 and 0.17 substitutions per site per billion years for *Ophioglossum*, and 0.58 and 0.15 substitutions per site per billion years for *Psilotum*. These R_S and R_N values in ferns are slightly lower than rates in *Welwitschia* but 2–3 times higher than rates in *Cycas* and *Ginkgo* (Guo *et al.*, 2016), and similar to rates observed in many angiosperm lineages (Mower *et al.*, 2007; Richardson *et al.*, 2013).

Repeat content and expanded mitochondrial introns

The *Ophioglossum* and *Psilotum* mitogenomes are highly repetitive (Tables 1, S5). Regardless of the specific criteria used for repeat identification, it is clear that the *Psilotum* mitogenome has one of the highest proportion of repetitive sequence (52.7–63.3%) of any land plant, whereas *Ophioglossum's* mitochondrial repeat content (37.1–44.3%) is similar to some of the most repeat-rich mitogenomes of land plants, such as *Silene conica* (40.8%, Sloan *et al.*, 2012), *Viscum scurruloideum* (39%, Skippington *et al.*, 2015), *Cucurbita pepo* (38%, Alverson *et al.*, 2010) and *Cucumis sativus* (36%, Alverson *et al.*, 2011).

The *Ophioglossum* mitogenome has four large (> 1 kb) repeats with sizes of 1.0, 1.3, 4.0 and 11.3 kb, whereas the *Psilotum* mitogenome has five large repeats of 1.1, 1.5, 2.7, 3.0 and 25.4 kb in size. All of these large repeats are present in two copies, and all are identical or nearly so (99.5–100% sequence identity) except for the 1.0-kb and 1.3-kb repeats in *Ophioglossum*, which have 97.0% and 78.8% sequence identities. Also, an indel of ~100 bp is present in each of the 4.0-kb and 11.3-kb repeats in *Ophioglossum*. Smaller (< 1 kb) repeats are composed of a complex array of tandem and dispersed subrepeat families that are present in many copies throughout each genome. Many of these repeats partially overlap, making it difficult to quantify precisely the number of repeats in the two genomes. Using REPEATSCOUT and REPEATMASKER, we identified 27 families in *Ophioglossum* and 41 families in *Psilotum* (Fig. S3). None of these repeat families are shared between the two species, and they do not share any obvious similarity to the Bpu elements previously reported for *Cycas* and *Ginkgo* (Chaw *et al.*, 2008; Guo *et al.*, 2016).

Tandem repeats make up a substantial proportion of repeat content, accounting for *c.* 11% of the *Ophioglossum* mitogenome and 4% of the *Psilotum* mitogenome (Table 1). Numerous short tandem repeats (STRs) were also detected in both mitogenomes (Table S6). Some notable high-frequency STRs include ACC/ GGT (28 copies), CCA/TGG (40), TTAAA/TTTAA (38), AAC TCCTCC/GGAGGAGTT (58), ACTCCTCCC/GGGAGG AGT (45) and AGGAGTTGG/CCAACTCCT (34) in the *Ophioglossum* mitogenome, and AAC/GTT (58), AACC/GGTT (45), AAGC/GCTT (110), GCAA/TTGC (140), CTCCC/ GGGAG (30), CTTCC/GGAAG (26) and GGGGA/TCCCC (84) in the *Psilotum* mitogenome. Nearly all of these STRs are located in noncoding regions. The few exonic STRs are in all cases trinucleotide repeats present in few tandem copies (3–5), and are therefore not expected to affect functionality of the translated product.

The abundance of repeats in the fern mitogenomes has resulted in a substantial increase in the size of many introns (Fig. 5 and Table S7). Of the 262 *cis*-spliced introns examined from the 12 sampled mitogenomes, the six largest introns are found in the ferns, including five *Psilotum* introns and one *Ophioglossum* intron that are all > 6 kb in length. Overall, the average mitochondrial intron size in *Psilotum* is substantially larger than in all other vascular plants, whereas *Ophioglossum* ranks second, slightly ahead of the lycophyte *Huperzia*. The increased repeat content in both fern mitogenomes accounts for a large portion of the increased intron sizes (Table S8); between 46% and 80% of these six largest introns are composed of repetitive elements. These large introns in the *Psilotum* mitogenome have also resulted in extremely long total lengths for several intron-containing genes. The longest gene, *nad1,* is > 21 kb from start to stop codon, and the genes *nad2, nad4, nad5* and *nad7* are 15–20 kb in length.
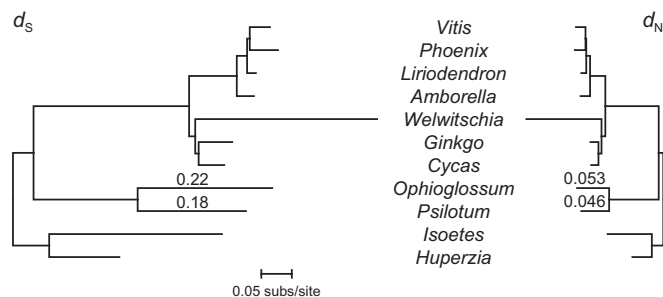
Fig. 4 Phylograms representing synonymous (d_S) and nonsynonymous (d_N) sequence divergence based on the 41 mitochondrial genes present in the majority of sampled vascular plants. The d_S and d_N branch lengths are shown for those branches that were used for the substitution rate calculations for *Ophioglossum* and *Psilotum*.
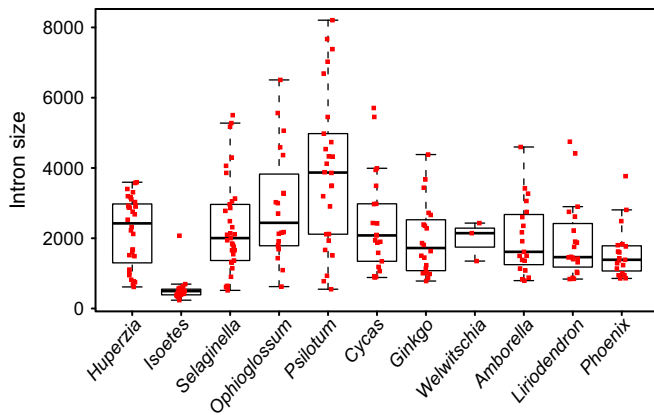
Fig. 5 Distribution of intron sizes in vascular plant mitogenomes. Variation is represented using box and whisker plots, generated in R v3.2.1. Each data point corresponds to a single intron. The boxes show the 25% and 75% quartiles of the data points, and the horizontal lines within the boxes indicate the means. The dotted whisker lines encompass the range of all non outlier data points.

Genome structural dynamics

We examined the recombinational activity of all nonoverlapping repeats between 100 bp and 5 kb in the *Ophioglossum* and the *Psilotum* mitogenomes by mapping the 5-kb mate-pair reads to the genome sequences. For the 4.0-kb (nearly identical except for a 130 bp indel) and 1.0-kb (97.0% sequence identity) inverted repeats in *Ophioglossum*, only 24.5% and 0.1% read pairs that span each repeat support the recombinant repeat forms that would generate an alternative conformation (AC) of the genome. In *Psilotum*, because one copy of the 3.0-kb (99.5% sequence identity) and 2.7-kb (100% sequence identity) repeats are located on each chromosome, recombination at these repeats would generate a single chromosome. However, only 0.3% of the read pairs spanning the 3.0-kb repeat and none of the read pairs spanning the 2.7-kb repeat support a recombination event, suggesting that the *Psilotum* mitogenome is present predominantly as two autonomous circular chromosomes and may only be substoichiometrically present as a single chromosomal arrangement. For the 1.5-kb repeat in *Psilotum* (100% sequence identity), both copies are located on chromosome 2, and only 0.7% of the spanning read pairs support the AC. The 1.3-kb direct repeats in *Ophioglossum* and the 1.1-kb direct repeats in *Psilotum* were located <5 kb from each other and thus were not suitable for analyzing recombinational activity. Most of the medium repeats (100 and 1000 bp) in both *Ophioglossum* and *Psilotum* exhibited very little (< 1%) to no recombination, with just a few repeats having up to 2.3% recombination frequency. The limited amount of repeat-mediated rearrangement in the two fern mitogenomes is supported further by the fact that only 6.6% (3195 of 48 474) of *Ophioglossum* mapped read pairs and 2.2% (8942 of 405 171) of *Psilotum* mapped read pairs were consistent with any kind of AC, despite the presence of so many repeats.

Although recombination at large mitochondrial repeats appears to be low in both *Ophioglossum* and *Psilotum*, gene order in their mitogenomes is not well conserved, in contrast to the highly syntenic mitogenomes of bryophytes (Fig. S4). The *Megaceros* and *Physcomitrella* mitogenomes have retained four major clusters of genes, with just a single translocation of the *Megaceros nad1* gene out of its ancestral position. As previously demonstrated (Liu *et al.*, 2012), *Huperzia* has retained quite a few gene adjacencies from these ancestral gene clusters, including many of the genes in the ribosomal protein cluster, the ribosomal RNA cluster, the *sdh3-sdh4-nad4L* cluster, and the *rpl10-nad5* and *atp4-cox1* clusters. The two ferns have retained many fewer gene adjacencies, including just a few fragmented subsets of the ancestral ribosomal protein cluster and the *shd4-mttB* pairing. Thus, even though repeat-mediated recombination is low in ferns relative to seed plants, the recombination rate in *Huperzia* and bryophytes may be even lower to account for the higher degree of synteny between their mitogenomes.

We also estimated the rates of DNA turnover in *Ophioglossum* and *Psilotum* mitogenomes using methods established previously (Guo *et al.*, 2016). The *Ophioglossum* and *Psilotum* share ~55 kb of mtDNA. Assuming again a divergence time of 305 Ma, these values fit very well to the regression line previously estimated from 14 pairs of phylogenetically independent seed plants (Fig. 3b of Guo *et al.*, 2016), indicating that DNA turnover rates in these two ferns are similar to most seed plants.

Discussion

The lack of a complete mitochondrial genome from ferns, a phylogenetically pivotal group as the sister lineage of seed plants, has impeded a comprehensive understanding of the origin and evolution of mitogenomic diversity in vascular plants. In the present study, we sequenced and analyzed the first complete mitogenomes from two ferns, *Ophioglossum californicum* and *Psilotum nudum*, and then compared them with other seed plant and lycophyte genomes. These comparative analyses have provided substantial insight on the mitogenomic features of the common ancestor of vascular plants, contributed to a broader and more comprehensive understanding of all aspects of plant mitogenome evolution, and identified several remarkable features of these fern mitogenomes that expand the range of known mitogenomic diversity in plants.

The mitogenome of the ancestral vascular plant was moderately sized but rich in genes, introns and edit sites

Although mitogenomic sizes are highly variable among seed plants (e.g., Sloan *et al.*, 2012; Skippington *et al.*, 2015; Guo *et al.*, 2016), more than half of the sequenced angiosperm genomes are 300–600 kb in size (as of December 2015), and two of the three available gymnosperm mitogenomes are ~400 kb in size, suggesting that the size of the ancestral seed plant mitogenome was also ~400 kb (Guo *et al.*, 2016). The two newly sequenced fern mitogenomes are similar in size to the majority of seed plants, as is the only fully assembled lycophyte mitogenome from *Huperzia squarrosa*, at 414 kb (Liu *et al.*, 2012). Taken together, the abundance of ~400-kb mitogenomes in vascular plant lineages suggests that their common ancestor may also be

around this size, followed by independent genome expansions and contractions in specific descendant lineages. However, sequencing of additional gymnosperm, fern and lycophyte mitogenomes is necessary to ensure that this inference is not influenced by ascertainment bias in these lineages of smaller genomes that are easier to be fully assembled.

Comparative analyses have indicated that 41 protein-coding genes were present in the mitogenome of the common ancestor of seed plants (Guo *et al.*, 2016), whereas in the vascular plant ancestor, the mitogenome appears to be more gene-rich (Fig. 1c). In particular, the presence of a mitochondrial *rpl6* gene in *Huperzia* and both newly sequenced ferns indicates that this gene was also likely to be present in the vascular plant ancestor, followed by loss in other lycophytes and seed plants. Mitochondrial maturases, RNA splicing factors that assist in the removal of introns during transcriptional maturation, probably also were more abundant in the common ancestor of vascular plants. In addition to *matR*, which is a mitochondrial gene in the majority of vascular plants, lycophyte mitogenomes have several other maturase genes, including *mat-atp9i87* in *Isoetes* and *Selaginella*, and *mat-atp9i95* and *mat-cobi787* in *Huperzia* (Guo & Mower, 2013). Notably, intact or pseudogenized orthologs of these lycophyte maturases also are found in one or more nonvascular land plants, strongly suggesting that these maturases were present in the vascular plant ancestor, maintained in some lycophytes, and functionally lost from most other vascular plants, including the ferns sequenced in this study. *Huperzia* also contains pseudogenes for several additional maturases, but the lack of functional orthologs in any vascular plants suggests that these genes were not functional in their ancestor. Altogether, comparative analysis of gene content suggests that the common ancestor of vascular plants had at least 46 protein-coding genes in its mitogenome, including the 41 genes in seed plants plus *rpl6*, *rps8*, *mat-atp9i87*, *mat-atp9i95* and *mat-cobi787*. The absence of *rps8* and these additional maturases and their host introns from all seed plants and the two newly sequenced ferns suggests that they were lost from the common ancestor of euphyllophytes (i.e. ferns plus seed plants), although additional fern mitogenomes are needed to better assess this inference.

Ancestral reconstruction of intron content in vascular plants is more complicated due to the large number of distinct introns in lycophytes relative to ferns and seed plants (Fig. 2c). In total, 27 of the group II introns in vascular plants have homologs in nonvascular plants and/or streptophytic green algae (Hecht *et al.*, 2011; Liu *et al.*, 2012; Turmel *et al.*, 2013), which signifies that probably they were acquired early in plant evolution and retained in the ancestral vascular plant. Another 11 group II introns are present in both lycophytes and euphyllophytes but not any nonvascular plants or green algae, indicating that they were acquired in the common ancestor of vascular plants and inherited by descendant lineages. The remaining 11 group II introns are more restricted in their distribution, suggesting relatively recent gains in specific lineages, although for all of these introns, it is only slightly less parsimonious to infer that they were actually acquired by the vascular plant ancestor and then lost once or a few times in descendant lineages. The evolutionary origin of the five group

I introns in lycophytes or ferns (cox1i395, cox1i624, cox1i876, cox1i1305, rrnSi839) is less clear due to their highly lineage-specific distribution in vascular plants, coupled with the sporadic distribution of homologs in nonvascular plants and/or green algae. This is akin to the situation for the cox1i729 intron, which is present in many angiosperms and liverworts and was shown to be horizontally acquired from fungi (Cho *et al.*, 1998; Sanchez-Puerta *et al.*, 2008). Deeper taxon sampling is needed to determine whether these fern/lycophyte group I introns were obtained through horizontal gene transfer or vertical transfer coupled with many losses in the majority of other land plants. Overall, it is apparent that the ancestral mitogenome of vascular plants was intron-rich, containing at least 38 group II introns and up to five group I introns. This high ancestral count implies that intron losses have been more frequent than intron gains in vascular plants, particularly in euphyllophytes which have substantially fewer mitochondrial introns presently.

Reconstruction of RNA editing content is even more difficult due to the low number of conserved sites among lineages. In the ferns *Psilotum* and *Ophioglossum*, which diverged *c.* 305 Myr ago (Ma), both species have about ~1000 C-to-U edit sites and a small number of U-to-C sites (Table 2). However, only ~25% of the edit sites in coding regions are conserved between them (Table S4), suggesting a high rate of gain and/or loss in these two fern lineages. Between the gymnosperms *Ginkgo* and *Cycas*, which also diverged *c.* 300 Ma, *c.* 70% of the ~1300 edit sites (all C-to-U) are shared, suggesting a somewhat lower rate of gain and/or loss compared with ferns (Guo *et al.*, 2016). In lycophytes, RNA editing content and type ranges widely, from a few hundred C-to-U and U-to-C sites in *Huperzia* (Liu *et al.*, 2012) to > 1500 C-to-U sites and > 200 U-to-C sites in *Isoetes* (Grewe *et al.*, 2011) and > 2100 sites (all C-to-U) in *Selaginella* (Hecht *et al.*, 2011). Editing content in hornworts, the presumed sister lineage to vascular plants, was also inferred to be high and to include both C-to-U and U-to-C sites (Xue *et al.*, 2010), although precise editing counts are not known because complete transcriptomes were not examined. Taken together, it seems likely that the vascular plant ancestor had a large number of C-to-U edit sites given that > 1000 edit sites are found in most ferns, gymnosperms and lycophytes. Because U-to-C editing has been detected in ferns, lycophytes and hornworts, this process was almost certainly active in the vascular plant ancestor as well. Additional transcriptomes are needed to more accurately estimate the ancestral editing content of vascular plants.

Recurrent loss of mitochondrial *ccm* genes from land plants suggests functional co-option by alternative cytochrome c maturation pathways

In plants, the mitochondrial *ccm* genes encode products involved in the attachment of a heme group to the cytochrome c protein (Giege *et al.*, 2008). The absence of any functional *ccm* genes in *Ophioglossum* but their presence in *Psilotum* and many other eusporangiate ferns signifies that these genes were lost from the *Ophioglossum* mitogenome. Across land plants, the

mitochondrial *ccm* genes have been completely lost or pseudogenized at least four times independently, from the liverwort *Treubia lacunosa* (Liu *et al.*, 2011), the hornworts *Megaceros aenigmaticus* and *Phaeoceros laevis* (Li *et al.*, 2009; Xue *et al.*, 2010), the lycophytes *Isoetes*, *Selaginella* and *Huperzia* (Grewe *et al.*, 2009; Hecht *et al.*, 2011; Liu *et al.*, 2012), and the fern *Ophioglossum* (this study). From a broader perspective, this cytochrome *c* maturation pathway appears to have been acquired from the ancestral proto-mitochondrion and lost repeatedly in many eukaryotic lineages (Allen *et al.*, 2008; Babbitt *et al.*, 2015).

The correlated loss of all four mitochondrial genes in multiple land plant lineages argues against a model whereby all four genes were transferred more or less simultaneously from the mitochondrion to the nucleus. Instead, it is possible that the function of the *ccm* gene products in these lineages has been co-opted by one of the alternative cytochrome *c* maturation pathways, which also was postulated to have occurred in other eukaryotic lineages (Allen *et al.*, 2008; Babbitt *et al.*, 2015). Consistent with the functional co-option hypothesis, none of the nuclear-encoded or mitochondrial-encoded members of this mitochondrial cytochrome *c* maturation pathway were detected in the *Ophioglossum* transcriptome (Fig. 1b; Table S1). Likewise, homology searches of the *Selaginella* nuclear genome (Banks *et al.*, 2011) did not identify any homologs of the missing mitochondrial *ccm* genes or nuclear *Ccm* genes, although this negative result should be interpreted cautiously given the generally higher rate of sequence evolution in plant nuclear genomes compared with mitogenomes (Wolfe *et al.*, 1987; Drouin *et al.*, 2008; Zhu *et al.*, 2014) and the long divergence time (> 400 Myr) between *Selaginella* and euphyllophytes, whose genes were used for homology assessment. Searches of nuclear genomes from additional species lacking these genes are needed to test more robustly for the possibility of functional co-option.

Fern mitogenomes have the largest organellar introns and intron-encoded genes among green plants

One notable feature of the fern mitogenomes is the abundance of very large introns, several of which are > 6 kb in size. The largest *Psilotum* intron (nad7i917) is 8206 bp in length, which, to our knowledge, is the largest intron ever reported from the mitochondrial or plastid genome of any eukaryote. The increased repeat content appears to be the main driver of intron expansion in the fern mitogenomes (Table S8). Outside of ferns, the two largest reported introns in land plant mitogenomes are likely to be erroneous. A 6.9-kb intron was reported in the *Selaginella* mitogenome study (Hecht *et al.*, 2011), but the identity of this intron is unknown as the largest intron in the genome annotation is 5.5 kb (nad1i1477). In the annotated *Huperzia* genome, the atp9i87 intron was reported to be 5.8 kb (Liu *et al.*, 2012). However, the very short downstream exon was later inferred to be annotated incorrectly (Guo & Mower, 2013); after correction of this exon position, the size of atp9i87 is 3.6 kb. Thus, compared with ferns, the largest mitochondrial introns in other land plants

are the *Cycas* nad4i976 intron at 5.7 kb followed by the *Selaginella* nad1i1477 intron at 5.5 kb. There are no confirmed introns > 6 kb in any sequenced land plant mitogenomes other than ferns.

Many of the largest introns in the *Psilotum* mitogenome are found in the intron-rich *nad* genes, which makes several of these genes very long. In fact, the *Psilotum* nad1 gene, at 21 474 bp, is the longest organellar gene of any green plant. In total, over 95% of this gene is intronic. Several other *nad* genes (*nad2*, *nad4*, *nad5* and *nad7*) in *Psilotum* are also long (15.7–19.3 kb) and mostly intronic (89–94%), raising intriguing questions about the potential regulatory roles or possible detrimental effects of these *nad* introns during transcription. Longer organellar genes are known only from fungal mitogenomes, such as the extremely intron-rich *cox1* gene, which is more than 29 kb in *Agaricus bisporis* (Ferandon *et al.*, 2010).

Many repeats but extremely low levels of repeat-mediated recombination

Repeats, both large (> 1 kb) and small (< 1 kb), are nearly ubiquitously present in the mitogenomes of vascular plants. In this study, we determined that the mitogenomes of the ferns *Ophioglossum* and *Psilotum* are highly repetitive (Tables 1, S5). Despite this abundance of repeats, however, recombinational activity is unusually quiescent in these fern genomes. This result contrasts sharply with the situation in most angiosperms, in which the larger repeats undergo frequent recombination resulting in nearly equimolar ratios of various genomic forms created by the repeat-mediated recombination events (Maréchal & Brisson, 2010; Mower *et al.*, 2012a). In gymnosperms, recombinational activity in *Ginkgo* and *Welwitschia* was found to be somewhat lower at large and small repeats compared with most angiosperms (Guo *et al.*, 2016). In lycophytes, the evidence for repeat-mediated recombination is more indirect. The many repeats and genomic arrangements in *Isoetes* and *Selaginella* mitogenomes suggest a highly active recombination system (Grewe *et al.*, 2009; Hecht *et al.*, 2011), whereas the mitogenomes of *Huperzia* and nonvascular plants exhibit a high degree of synteny (Liu *et al.*, 2012), suggesting much lower recombinational activity in *Huperzia* and most nonvascular plants. Thus, as with many features of plant mitochondrial genomes, there appears to be a wide range of variation among species with respect to the frequency of repeat-mediated recombination.

Author contributions

J.P.M. designed the research. W.G. and W.F. performed experiments. W.G., A.Z. and J.P.M. analyzed the data. W.G. and J.P.M. interpreted the results and wrote the paper. All authors read and approved the final manuscript.

References

Adams KL, Clements MJ, Vaughn JC. 1998. The *Peperomia* mitochondrial *cox1* group I intron: timing of horizontal transfer and subsequent evolution of the intron. *Journal of Molecular Evolution* **46**: 689–696.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences, USA* **99**: 9905–9912.

Allen JW, Ferguson SJ, Ginger ML. 2008. Distinctive biochemistry in the trypanosome mitochondrial intermembrane space suggests a model for stepwise evolution of the MIA pathway for import of cysteine-rich proteins. *FEBS Letters* **582**: 2817–2825.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* **23**: 2499–2513.

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* **27**: 1436–1448.

Babbitt SE, Sutherland MC, Francisco BS, Mendez DL, Kranz RG. 2015. Mitochondrial cytochrome c biogenesis: no longer an enigma. *Trends in Biochemical Sciences* **40**: 446–455.

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA *et al.* 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* **332**: 960–963.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

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, Chun-Chieh Shih A, Wang D, Wu Y-W, Liu S-M, Chou T-Y. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short Interspersed elements, Bpu Sequences, and abundant RNA editing sites. *Molecular Biology and Evolution* **25**: 603–615.

Cho Y, Qiu Y-L, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences, USA* **95**: 14244–14249.

Christenhusz MJ, Chase MW. 2014. Trends and concepts in fern classification. *Annals of Botany* **113**: 571–594.

Dombrovska O, Qiu YL. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Molecular Phylogenetics and Evolution* **32**: 246–263.

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* **49**: 827–831.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Ferandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M, Barroso G. 2010. The *Agaricus bisporus cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS ONE* **5**: e14048.

Giege P, Grienenberger JM, Bonnard G. 2008. Cytochrome c biogenesis in mitochondria. *Mitochondrion* **8**: 61–73.

Grewe F, Edger PP, Keren I, Sultan L, Pires JC, Ostersetzer-Biran O, Mower JP. 2014. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion* **19**(Part B): 135–143.

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 Evolution Biology* **13**: 8.

Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B, Knoop V. 2011. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Research* **39**: 2890–2902.

Grewe F, Viehoever 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.

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, Grewe F, Mower JP. 2015. Variable frequency of plastid RNA editing among ferns and repeated loss of Uridine-to-Cytidine editing from vascular plants. *PLoS ONE* **10**: e0117075.

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.

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.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36.

Knie N, Grewe F, Knoop V. 2016. Monilophyte mitochondrial *rps1* genes carry a unique group II intron that likely originated from an ancient paralog in *rpl2*. *RNA*. doi: 10.1261/rna.056572.116.

Knie N, Polsakiewicz M, Knoop V. 2015. Horizontal gene transfer of chlamydial-like tRNA genes into early vascular plant mitochondria. *Molecular Biology and Evolution* **32**: 629–634.

Kück P, Meusemann K. 2010. FASconCAT: convenient handling of data matrices. *Molecular Phylogenetics and Evolution* **56**: 1115–1118.

Li L, Wang B, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *Journal of Molecular Evolution* **68**: 665–678.

Liu Y, Medina R, Goffinet B. 2014. 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. *Molecular Biology and Evolution* **31**: 2586–2591.

Liu Y, Wang B, Cui P, Li L, Xue J-Y, Yu J, Qiu Y-L. 2012. The mitochondrial genome of the lycophyte *Huperzia squarrosa*: the most archaic form in vascular plants. *PLoS ONE* **7**: e35168.

Liu Y, Xue JY, Wang B, Li L, Qiu YL. 2011. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: dynamic and conservative evolution. *PLoS ONE* **6**: e25836.

Malek O, Lattig K, Hiesel R, Brennicke A, Knoop V. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO Journal* **15**: 1403–1411.

Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **186**: 299–317.

Mower JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Research* **37**: W253–W259.

Mower JP, Case AL, Floro ER, Willis JH. 2012a. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biology and Evolution* **4**: 670–686.

Mower JP, Sloan DB, Alverson AJ. 2012b. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, ed. *Plant genome diversity, vol 1*. Vienna, Austria: Springer, 123–144.

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology* **7**: 135.

Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* **13**(Suppl. 14): S8.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.

Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**: i351–i358.

Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW *et al.* 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**: 1468–1473.

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* **11**: 29.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Molecular Biology and Evolution* **25**: 1762–1777.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proceedings of the National Academy of Sciences, USA* **112**: E3515–E3524.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology* **10**: e1001241.

Smith DR. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Molecular Biology* **71**: 627–639.

Turmel M, Otis C, Lemieux C. 2013. Tracing the evolution of streptophyte algae and their mitochondrial genome. *Genome Biology and Evolution* **5**: 1817–1835.

Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric *cox1* gene of *Peperomia*. *Journal of Molecular Evolution* **41**: 563–572.

Wolf PG, Sessa EB, Marchant DB, Li F-W, Rothfels CJ, Sigel EM, Gitzendanner MA, Visger CJ, Banks JA, Soltis DE *et al.* 2015. An exploration into fern genome space. *Genome Biology and Evolution* **7**: 2533–2544.

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.

Wu Z, Stone JD, Štorchová H, Sloan DB. 2015. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC Genomics* **16**: 938.

Xue JY, Liu Y, Li L, Wang B, Qiu YL. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics* **56**: 53–61.

Zerbino DR, Birney E. 2008. Velvet: Algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Research* **18**: 821–829.

Zhu A, Guo W, Jain K, Mower JP. 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Molecular Biology and Evolution* **31**: 1228–1236.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Mitogenome maps and depth of coverage analysis of *Ophioglossum californicum* and *Psilotum nudum*.

Fig. S2 Alignment of *rps8* amino acid sequences.

Fig. S3 Highly repetitive families of dispersed repeats identified in the *Ophioglossum* and *Psilotum* mitogenomes.

Fig. S4 Comparison of gene order in mitogenomes of ferns and other land plants.

Table S1 TBLASTN search results for transcribed *ccm* genes in fern transcriptomes

Table S2 Number of RNA genes in representative mitogenomes of vascular plants

Table S3 Description of RNA edit sites in the *Ophioglossum* and *Psilotum* mitogenomes

Table S4 Number of mitochondrial RNA edit sites shared between *Ophioglossum* and *Psilotum*

Table S5 Amount of repetitive DNA in highly repetitive mitogenomes of vascular plants

Table S6 Number of short tandem repeats in fern mitogenomes

Table S7 Intron sizes in vascular plant mitogenomes

Table S8 Composition of the largest fern introns

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.