

Extensive shifts from *cis* to *trans* splicing of gymnosperm mitochondrial introns

Wenhu Guo^{†,1,2}, Andan Zhu^{†,1,3,4}, Weishu Fan^{1,3,4}, Robert P. Adams⁵, Jeffrey P. Mower^{*,1,3}

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

² School of Biological University of Sciences, Nebraska, Lincoln, NE 68588, USA

³ Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

⁴ Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, Yunnan, China

⁵ Biology Department, Baylor University, Waco, TX 76798, USA

[†] These authors contributed equally to this work

*** Corresponding author:** Jeffrey P. Mower. Address: 1901 Vine St, Beadle Center E128, University of Nebraska, Lincoln, NE 68588. Phone: +1 402-472-2130. Email: jpmower@unl.edu

Keywords: *cis* splicing, gene loss, gymnosperms, intron evolution, mitochondrial genomes, *trans* splicing

Running Head: Extensive *trans* splicing in gymnosperm mitogenomes

ABSTRACT

Hundreds of plant mitogenomes have been sequenced from angiosperms, but relatively few mitogenomes are available from its sister lineage, gymnosperms. To examine mitogenomic diversity among extant gymnosperms, we generated draft mitogenomes from 11 diverse species and compared them with four previously published mitogenomes. Examined mitogenomes from Pinaceae and cycads retained all 41 protein genes and 26 introns present in the common ancestor of seed plants, whereas gnetophyte and cupressophyte mitogenomes experienced extensive gene and intron loss. In Pinaceae and cupressophyte mitogenomes, an unprecedented number of exons are distantly dispersed, requiring *trans* splicing of 50–70% of mitochondrial introns to generate mature transcripts. RNAseq data confirms *trans* splicing of these dispersed exons in *Pinus*. The prevalence of *trans* splicing in vascular plant lineages with recombinogenic mitogenomes suggests that genomic rearrangement is the primary cause of shifts from *cis* to *trans* splicing in plant mitochondria.

INTRODUCTION

Across eukaryotes, the mitochondrial genome (mitogenome) exhibits an amazing diversity of genomic architectures, with large variations in size, structure, gene and intron content, and nucleotide composition (Smith and Keeling 2015). The mitogenomes of land plants encapsulate much of this architectural extremism. Land plants tend to have 30 to 40 mitochondrial protein genes, although coding content can range from fewer than 20 protein genes in mistletoe (Skippington et al. 2015) to more than 50 protein genes (including intron-encoded endonucleases and maturases) in some liverworts (Mower et al. 2012). Additionally, plant mitochondrial transcripts require extensive post-transcriptional modifications, including intron splicing and RNA editing, to manage the variable numbers of introns and RNA edit sites distributed in lineage-specific patterns across land plants (Bonen 2012; Mower et al. 2012; Ichinose and Sugita 2017). Most plant mitochondrial introns are classified as group II introns, while a minority are classified as group I, based on differences in structure and splicing mechanism (Bonen 2012).

Plant mitochondrial introns are typically removed by *cis* splicing, whereby the exons to be joined are within the same primary transcript, separated by the intron to be removed. *Trans* splicing, which describes the joining of exons from distinct transcripts, is much less common, although examples exist for group I and II introns from eukaryotic organelles and prokaryotes (e.g., Malek and Knoop 1998; Burger et al. 2009; Glanz and Kück 2009; Grewe et al. 2009; Pombert et al. 2013; LaRoche-Johnston et al. 2018). Among land plants, mitochondrial *trans* splicing is absent from nonvascular plants but has evolved several times in particular lineages of vascular plants, most abundantly in angiosperms. *Trans* splicing is typically required for five or six (ca. 25%) angiosperm mitochondrial introns, of which five (nad1i394, nad1i669, nad2i542, nad5i1455, nad5i1477) likely became *trans*-spliced in the seed plant common ancestor by fragmentation of a *cis*-spliced arrangement (Malek and Knoop 1998; Groth-Malonek et al. 2005; Guo et al. 2016). Additional shifts from *cis* to *trans* splicing are more restricted in scope among vascular plants, such as atp9i21 and cob1787 in the lycophyte *Selaginella moellendorffii* (Hecht et al. 2011) and cox2i691 in the gymnosperm *Welwitschia mirabilis* (Guo et al. 2016). Some introns have experienced repeated, independent transitions from *cis* to *trans* splicing, including nad1i728 in diverse seed plants (Qiu and Palmer 2004; Guo et al. 2016) and cox2i373 in some *Allium* species (Kim and Yoon 2010) and *S. moellendorffii* (Hecht et al. 2011).

Hundreds of mitogenomes have been sequenced from angiosperms, but many fewer are available from gymnosperms, the sister lineage to angiosperms. Gymnosperms comprise ~1000

species divided into Pinaceae (conifers I), cupressophytes (conifers II), gnetophytes, cycads, and ginkgo, and relationships among these groups continue to be confounded by the uncertain placement of gnetophytes, which varies depending on data set and methodology used (Ran et al. 2018; One Thousand Plant Transcriptomes 2019). Complete mitogenomes have been sequenced for a cycad (*Cycas taitungensis*, Chaw et al. 2008), a gnetophyte (*W. mirabilis*, Guo et al. 2016) and the ginkgo tree (*Ginkgo biloba*, Guo et al. 2016), and a draft mitogenome sequence is available from *Picea glauca* in Pinaceae (Jackman et al. 2016). Notably, there is an increased requirement for *trans* splicing of seven introns in the *W. mirabilis* mitogenome (Guo et al. 2016), and our preliminary examination of exon distribution in the *P. glauca* mitochondrial assembly suggested an even greater requirement for *trans* splicing. This raised the strong possibility that *trans* splicing is more prevalent in particular gymnosperms in comparison to all other land plants.

RESULTS

To assess the diversity of gene and intron content and the relative frequency of *cis* versus *trans* splicing among gymnosperm mitochondria, draft mitogenomes (Table S1) were assembled from five cupressophytes (*Araucaria heterophylla*, *Cupressus sempervirens*, *Hesperocyparis glabra*, *Podocarpus macrophyllus*, *Taxus baccata*), three Pinaceae species (*Abies sibirica*, *Picea abies*, *Pinus strobus*), two cycads (*Ceratozamia hildae*, *Zamia integrifolia*), and one gnetophyte (*Gnetum gnemon*). Estimated mitogenomic sizes range from < 500 kb in *T. baccata* and cycads to >6 Mb in *P. strobus* and *P. abies*, although the true mitogenomic size is likely to be higher due to unresolved repeats in these draft assemblies. GC% ranges from <45% in *P. strobus* and *P. abies* to >50% in Cupressales, which includes *C. sempervirens*, *H. glabra*, *T. baccata*. Comparison of mitochondrial gene and intron content among these 11 draft mitogenomes and four previously published mitogenomes (from *C. taitungensis*, *G. biloba*, *P. glauca*, and *W. mirabilis*) revealed widespread retention in some lineages and extensive loss in others (Fig. 1). Mitogenomes from three cycads and four Pinaceae species contain the 41 protein-coding genes and 26 introns inferred to be present in the common ancestor of seed plants (Guo et al. 2016), and *G. biloba* mitogenome content is similar except for loss of the *rps10* intron. By contrast, cupressophytes have lost 8–9 genes and 11–12 introns, while gnetophytes have lost 12 genes and either 4 (*G. gnemon*) or 16 (*W. mirabilis*) introns.

Moreover, there is extensive variation in the *cis* versus *trans* splicing arrangements of many mitochondrial introns (Fig. 1B). In ginkgo and cycads, *trans* splicing is required for just the five aforementioned introns (nad1i394, nad1i669, nad2i542, nad5i1455, nad5i1477) that shifted to a

trans configuration in the seed plant ancestor. In other gymnosperm lineages, however, we discovered a broader requirement for mitochondrial *trans* splicing, affecting in total 7 gnetophyte introns, 10–11 cupressophyte introns, and 13 Pinaceae introns. Overall, mitochondrial *trans* splicing is required for 50% of Pinaceae introns and 70% of introns from cupressophytes and *W. mirabilis*, but only for 20% of introns in ginkgo and cycads. Importantly, no repeat-mediated recombinant products were detected close to the intron breakpoints of any *trans*-spliced intron in any assembly, indicating the absence of alternative mitogenomic arrangements containing a *cis*-spliced version of these introns. Furthermore, *de novo* assembly and read mapping of RNAseq data demonstrated that *Pinus taeda* mitochondrial exons are correctly spliced for all 10 intron-containing genes, verifying a *trans*-splicing mechanism to remove 13 of the 26 introns (Fig. 2). Correct splicing was also demonstrated for all *cis*- and *trans*-spliced introns from *G. biloba* and *W. mirabilis* (Guo et al. 2016; Fan et al. 2019).

Of the 15 mitochondrial introns that are *trans*-spliced in at least one gymnosperm, eight have only been identified from gymnosperms, including *cox2i691* in Pinaceae and gnetophytes, *nad2i1282* and *nad4i461* in Pinaceae and cupressophytes, *nad4i976* and *nad4i1399* in cupressophytes, and *nad7i209*, *nad7i917* and *rpl2i846* in Pinaceae. For *cox2i373*, however, the shift to *trans* splicing in cupressophytes and Pinaceae has also occurred independently in *Allium* (Kim and Yoon 2010) and *S. moellendorffii* (Hecht et al. 2011). Likewise, the *nad1i728* intron, which became *trans* spliced numerous times in angiosperms (Qiu and Palmer 2004), has shifted to *trans* splicing in cupressophytes, gnetophytes and Pinaceae.

Trans splicing of *nad1i728* is particularly interesting because this intron contains the *matR* gene, encoding a maturase that assists in intron splicing (Sultan et al. 2016). In angiosperms, when the intron became *trans*-spliced, *matR* remained associated with either the upstream or downstream exon, depending on the position of the intron break (Qiu and Palmer 2004). In Pinaceae and *P. macrophyllus*, the intron break occurred upstream of *matR*, which remained adjacent to the downstream exon, as exemplified by *P. abies* (Fig. 3). In gnetophytes and the other four cupressophytes, however, intron breaks occurred both upstream and downstream of *matR*, such that *matR* is no longer near either exon, although *matR* is still flanked on both sides by fragments of the *trans*-spliced intron, as previously described for *W. mirabilis* (Guo et al. 2016) and shown here for *A. heterophylla* (Fig. 3). Thus, *nad1i728* may function as a tripartite *trans*-spliced intron in these gymnosperms, which was previously observed for mitochondrial intron *nad5i1477* in *Oenothera berteriana* (Knoop et al. 1997) and a plastid intron in *Chlamydomonas reinhardtii* (Goldschmidt-Clermont et al. 1991).

DISCUSSION

Here, we detailed extensive variation in mitochondrial gene and intron content among gymnosperm lineages, and uncovered the highest requirement for *trans* splicing among all plant and green algal organellar genomes. Indeed, such high levels of *trans* splicing (using distinct mechanisms) are otherwise known from diplomid mitogenomes (Marande and Burger 2007) and nuclear genomes of some dinoflagellates, euglenozoans, and nematodes (Lei et al. 2016). Unambiguously reconstructing the precise evolutionary history of gymnosperm gene losses, intron losses, and *cis*-to-*trans* splicing shifts is difficult due to the uncertain phylogenetic position of gnetophytes; nevertheless, parsimony-based reconstructions indicated many independent or convergent events of gene and intron loss and *cis*-to-*trans*-splicing shifting irrespective of phylogeny (Fig 4). Of particular note is the inference of loss of one or more *trans*-spliced introns, including nad5i1477 from *Podocarpus* and (depending on reconstruction) cox2i373 from gnetophytes or cox2i691 from cupressophytes. Across eukaryotes, loss of a *trans*-spliced organellar intron is unusual, with only two other examples known (Grewe et al. 2016; Fučíková et al. 2019).

Regardless of phylogeny, convergent evolution must be inferred for the shared losses of *sdh4* and six introns among some (but not all) cupressophytes and gnetophytes. Furthermore, depending on phylogenetic relationships, convergent evolution must also be invoked to explain the gene and intron losses shared among all cupressophytes and gnetophytes and/or the *cis*-to-*trans* splicing changes shared among all cupressophytes and Pinaceae. More generally, convergent evolution is rampant in other vascular plant mitogenomes for various gene and intron losses and *cis*-to-*trans* splicing shifts (Joly et al. 2001; Adams et al. 2002; Qiu and Palmer 2004; Guo et al. 2017). Collectively, these findings suggest that a simple parsimony-based inference scheme is likely to underestimate the true number of evolutionary changes in vascular plant mitogenomes.

The evolution of *trans*-spliced introns from *cis*-spliced ancestors presumably results from DNA rearrangement, most likely driven by mutagenic repair of double-strand breaks or recombination at small repeats (Qiu and Palmer 2004; Bonen 2012). At a broad level, shifts to *trans* splicing correlate well with relative rearrangement rates. Angiosperm mitogenomes are notorious for their frequent rearrangement (Palmer and Herbon 1988; Woloszynska 2010), and until this study, angiosperms were notable among land plants for having the highest number of *trans*-spliced introns. In stark contrast, no *trans* splicing has yet been detected in any nonvascular

plant mitogenomes, which tend to be much less rearranged (Wang et al. 2009; Xue et al. 2010; Liu et al. 2014; Dong et al. 2018; Myszczyński et al. 2018).

The sparser data from vascular plants other than angiosperms also supports a correlation between *trans* splicing and rearrangement. No mitochondrial *trans*-spliced introns were identified in two ferns or in the lycophyte *Phlegmariurus squarrosus*, which have limited rearrangement (Liu et al. 2012; Guo et al. 2017), whereas *S. moellendorffii* has four *trans*-spliced introns and an extraordinarily recombinogenic mitogenome (Hecht et al. 2011). Since the seed plant ancestor, ginkgo and cycads have not acquired any new *trans*-spliced introns, and their mitogenomes are much less rearranged than other seed plants, whereas *W. mirabilis* appears more recombinogenic and has acquired two additional *trans*-spliced introns (Guo et al. 2016). Collectively, these results strongly support the role of DNA rearrangement in causing *cis*-to-*trans* splicing shifts during plant evolution. The increased mitochondrial *trans* splicing in cupressophytes and Pinaceae predicts that these mitogenomes will be more extensively rearranged relative to other gymnosperms. Indeed, several as-yet unpublished Pinaceae mitogenomes suggest extraordinary amounts of rearrangement (Jackman et al. 2019; Sullivan et al. 2019), correlating with the many *trans*-spliced introns arising in this family.

MATERIALS AND METHODS

Collection and sequencing details for all newly evaluated plants are provided in Table S2. Briefly, mitochondria-enriched DNA was isolated from *A. heterophylla* and *P. strobus* as described previously (Mower et al. 2010). Total cellular DNA was isolated from *C. hildae*, *C. sempervirens*, *G. gnemon*, *H. glabra*, *P. macrophyllus* and *Z. integrifolia* using a simplified CTAB procedure (Doyle and Doyle 1987). For each species, leaves/needles were sampled from a single individual. *A. heterophylla*, *P. strobus* and *P. macrophyllus* DNAs were sequenced at BGI (Shenzhen, China) on an Illumina HiSeq 2500 machine. *C. sempervirens*, *G. gnemon* and *H. glabra* DNAs were sequenced at the Center for Genomics and Bioinformatics (Indiana University, Bloomington, IN) on an Illumina MiSeq machine. *C. hildae* and *Z. integrifolia* DNAs were sequenced at ACGT, INC. (Wheeling, IL) on an Illumina NextSeq 500 machine. Illumina sequencing data were obtained from the NCBI SRA for *A. sibirica*, *P. abies* and *T. baccata*.

Illumina reads were assembled with Velvet version 1.203 (Zerbino and Birney 2008) using a range of kmer values and expected coverage values, as described previously (Guo et al. 2016). Reads were also assembled with SPAdes 3.13 (Bankevich et al. 2012) using a range of kmer values shown in Table S1, as described previously (Mower et al. 2019). Mitochondrial exons

and introns were identified by default blastn homology searches using *C. taitungensis* and *G. biloba* sequences as queries. Introns were scored as *trans*-spliced when the intron was fragmented in the SPAdes and Velvet assemblies, the flanking exons were not within 5 kb of a contig end in at least one assembly, and there were no repeats within 5 kb of the broken intron ends. To identify additional mitochondrial scaffolds that lack mitochondrial genes, we first calculated average mitochondrial GC% and kmer coverage (weighted by contig length) from all annotated mitochondrial scaffolds for each species. Scaffolds >1 kb in length were scored as mitochondrial if GC% was within 3% and kmer coverage was between one-half and thrice the annotated mitochondrial average. Annotated mitochondrial scaffolds were deposited in GenBank under accessions MN965050–MN965384.

Illumina RNAseq reads for *P. taeda* were obtained from the NCBI SRA (SRR1200343, SRR1200412, SRR1200414, SRR1200421, SRR1200424, SRR1200432). RNAseq reads were assembled using Trinity v2.8.4 (Grabherr et al. 2011) with default parameters. Reads were mapped onto the assembled transcripts using Hisat2 aligner version 2.1.0 (Kim et al. 2019).

ACKNOWLEDGMENTS

We thank Yizhong Zhang for assistance with mitochondria-enriched DNA extractions, Volker Knoop for providing tissue for *Gnetum gnemon*, Jeff Palmer for comments on an early version of the manuscript, and members of the Mower lab for helpful discussions. This work was supported by the University of Nebraska (School of Biological Sciences Research Award to W.G.), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (startup funding 2017-LSF-GBOWS-02 to A.Z.), Baylor University (award BU 0324512 to R.P.A), and the National Science Foundation (awards IOS 1027529 and MCB 1125386 to J.P.M.).

REFERENCES

- 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. *Proc Natl Acad Sci U S A* 99: 9905-9912.
- 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. *J Comput Biol* 19: 455-477.
- Bonen L. 2012. Evolution of mitochondrial introns in plants and photosynthetic microbes. *Adv Bot Res* 63: 155-186.
- Burger G, Yan Y, Javadi P, Lang BF. 2009. Group I-intron *trans*-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet* 25: 381-386.
- Chaw SM, Shih AC, Wang D, Wu YW, Liu SM, Chou TY. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol Biol Evol* 25: 603-615.

- Dong SS, Xue JY, Zhang SZ, Zhang L, Wu H, Chen ZD, Goffinet B, Liu Y. 2018. Complete mitochondrial genome sequence of *Anthoceros angustus*: conservative evolution of the mitogenomes in hornworts. *Bryologist* 121: 14-22.
- Fan W, Guo W, Funk L, Mower JP, Zhu A. 2019. Complete loss of RNA editing from the plastid genome and most highly expressed mitochondrial genes of *Welwitschia mirabilis*. *Sci China Life Sci* 62: 498-506.
- Fučíková K, Lewis PO, Neupane S, Karol KG, Lewis LA. 2019. Order, please! Uncertainty in the ordinal-level classification of Chlorophyceae. *PeerJ* 7: e6899.
- Glanz S, Kück U. 2009. *Trans*-splicing of organelle introns—a detour to continuous RNAs. *Bioessays* 31: 921-934.
- Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. 1991. A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. *Cell* 65: 135-143.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644-652.
- 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 Res* 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 Biol Evol* 8: 3193-3201.
- Groth-Malonek M, Pruchner D, Grewe F, Knoop V. 2005. Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol Biol Evol* 22: 117-125.
- 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. *Mol Biol Evol* 33: 1448-1460.
- Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytol* 213: 391-403.
- 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 Biol Evol* 3: 344-358.
- Ichinose M, Sugita M. 2017. RNA editing and its molecular mechanism in plant organelles. *Genes-Basel* 8: 5.
- Jackman SD, Coombe L, Warren RL, Kirk H, Trinh E, McLeod T, Pleasance S, Pandoh P, Zhao Y, Coope RJ, et al. 2019. Largest complete mitochondrial genome of a gymnosperm, sitka spruce (*Picea sitchensis*), indicates complex physical structure. *bioRxiv*: 601104.
- Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, Raymond A, Pleasance S, Coope R, Wildung MR, et al. 2016. Organellar genomes of white spruce (*Picea glauca*): assembly and annotation. *Genome Biol Evol* 8: 29-41.
- Joly S, Brouillet L, Bruneau A. 2001. Phylogenetic implications of the multiple losses of the mitochondrial *coxII.i3* intron in the angiosperms. *Int J Plant Sci* 162: 359-373.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37: 907-915.
- Kim S, Yoon MK. 2010. Comparison of mitochondrial and chloroplast genome segments from three onion (*Allium cepa* L.) cytoplasm types and identification of a *trans*-splicing intron of *cox2*. *Curr Genet* 56: 177-188.
- Knoop V, Altwasser M, Brennicke A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet* 255: 269-276.
- LaRoche-Johnston F, Monat C, Coulombe S, Cousineau B. 2018. Bacterial group II introns generate genetic diversity by circularization and *trans*-splicing from a population of intron-invaded mRNAs. *PLoS Genet* 14: e1007792.
- Lei Q, Li C, Zuo Z, Huang C, Cheng H, Zhou R. 2016. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol Evol* 8: 562-577.

- Liu Y, Medina R, Goffinet B. 2014. 350 My of mitochondrial genome stasis in mosses, an early land plant lineage. *Mol Biol Evol* 31: 2586-2591.
- 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.
- Malek O, Knoop V. 1998. *Trans*-splicing group II introns in plant mitochondria: The complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. *RNA* 4: 1599-1609.
- Marande W, Burger G. 2007. Mitochondrial DNA as a genomic jigsaw puzzle. *Science* 318: 415.
- Mower JP, Ma PF, Grewe F, Taylor A, Michael TP, VanBuren R, Qiu YL. 2019. Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. *New Phytol* 222: 1061-1075.
- Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics*. Vienna: Springer. p. 123-144.
- Mower JP, Stefanović S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol* 8: 150.
- Myszczyński K, Górski P, Ślipiko M, Sawicki J. 2018. Sequencing of organellar genomes of *Gymnomitrium concinnatum* (Jungermanniales) revealed the first exception in the structure and gene order of evolutionary stable liverworts mitogenomes. *BMC Plant Biol* 18: 321.
- One Thousand Plant Transcriptomes I. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679-685.
- Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol* 28: 87-97.
- Pombert JF, Otis C, Turmel M, Lemieux C. 2013. The mitochondrial genome of the prasinophyte *Prasinoderma coloniale* reveals two *trans*-spliced group I introns in the large subunit rRNA gene. *PLoS One* 8: e84325.
- Qiu Y-L, Palmer JD. 2004. Many independent origins of *trans* splicing of a plant mitochondrial group II intron. *J Mol Evol* 59: 80-89.
- Ran JH, Shen TT, Wang MM, Wang XQ. 2018. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc R Soc B* 285: 20181012.
- 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. *Proc Natl Acad Sci U S A* 112: E3515-3524.
- Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A* 112: 10177-10184.
- Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, Keech O, Öberg L, Møller IM, Arvestad L, et al. 2019. The mitogenome of Norway spruce and a reappraisal of mitochondrial recombination in plants. *bioRxiv*: 682732.
- Sultan LD, Milesina D, Grewe F, Rolle K, Abudraham S, Głodowicz P, Niazi AK, Keren I, Shevtsov S, Klipcan L, et al. 2016. The reverse transcriptase/RNA maturase protein MatR Is required for the splicing of various group II introns in Brassicaceae mitochondria. *Plant Cell* 28: 2805-2829.
- Wang B, Xue J, Li L, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Curr Genet* 55: 601-609.
- Woloszynska M. 2010. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes-though this be madness, yet there's method in't. *J Exp Bot* 61: 657-671.
- 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. *Curr Genet* 56: 53-61.
- Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821-829.

Fig. 1. Mitochondrial gene content and intron content among gymnosperms and selected angiosperms. Species relationships were taken from the One Thousand Transcriptomes Project Initiative (2019). Gnet. = gnetophytes; Angio. = angiosperms. A) Protein gene content. “●” indicates presence of an intact gene, “Ψ” indicates a pseudogene, and “–” indicates gene loss. The 27 genes present in all sampled species include *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rps3*, *rps4*, and *rps12*. B) Intron content. “●” and “θ” indicate presence of *cis*- and *trans*-spliced introns, respectively. Intron losses are shaded, “–” indicates intron loss from an intact gene, and “x” indicates intron loss due to gene loss.

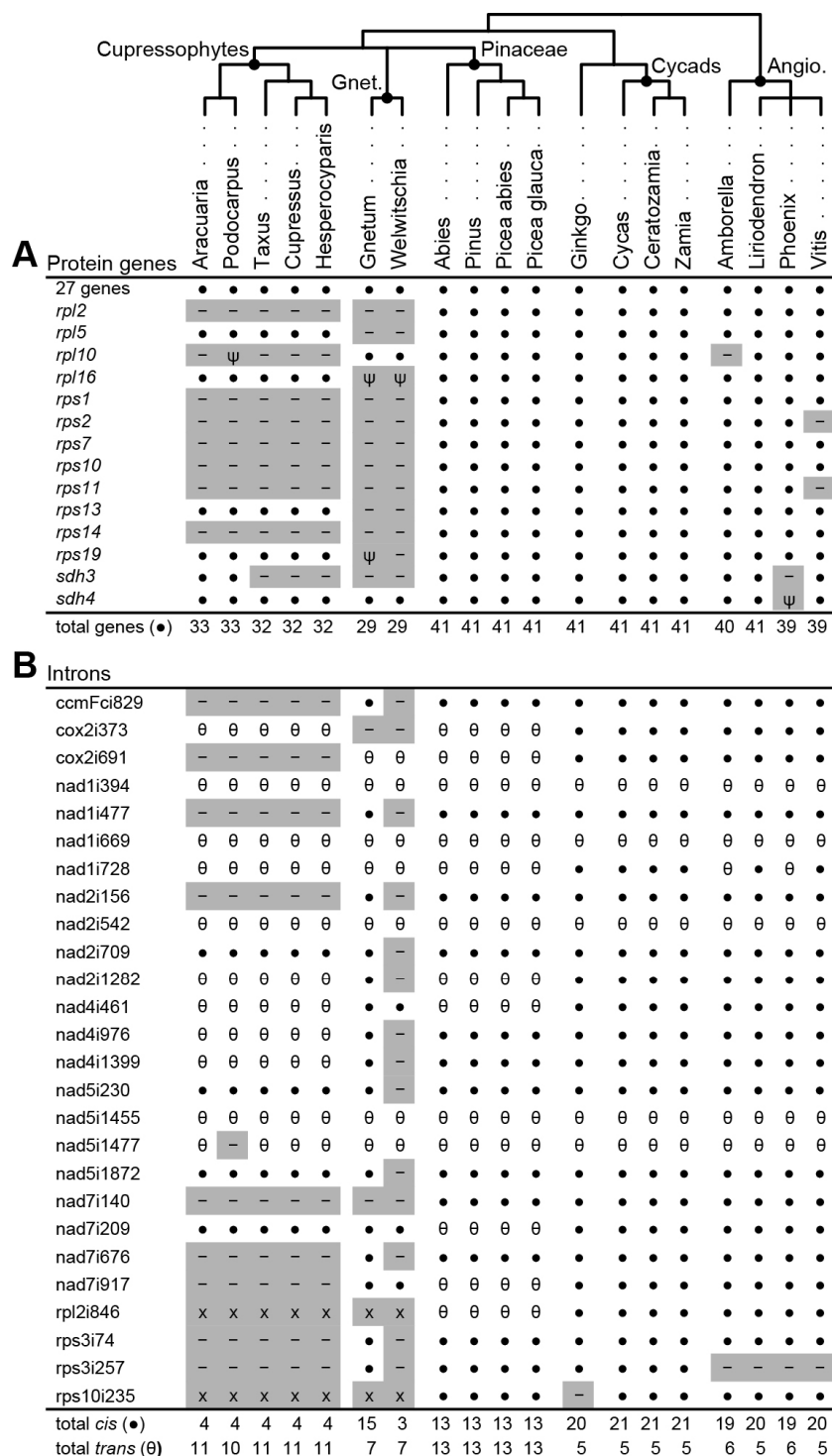


Fig. 2. Mature mitochondrial transcripts of ten intron-containing genes from *Pinus taeda*. Locations of all 26 spliced introns are marked with a vertical line, and the number of read pairs that span the splice junction are shown above each intron. Depth of coverage of each transcript is plotted and shaded in gray.

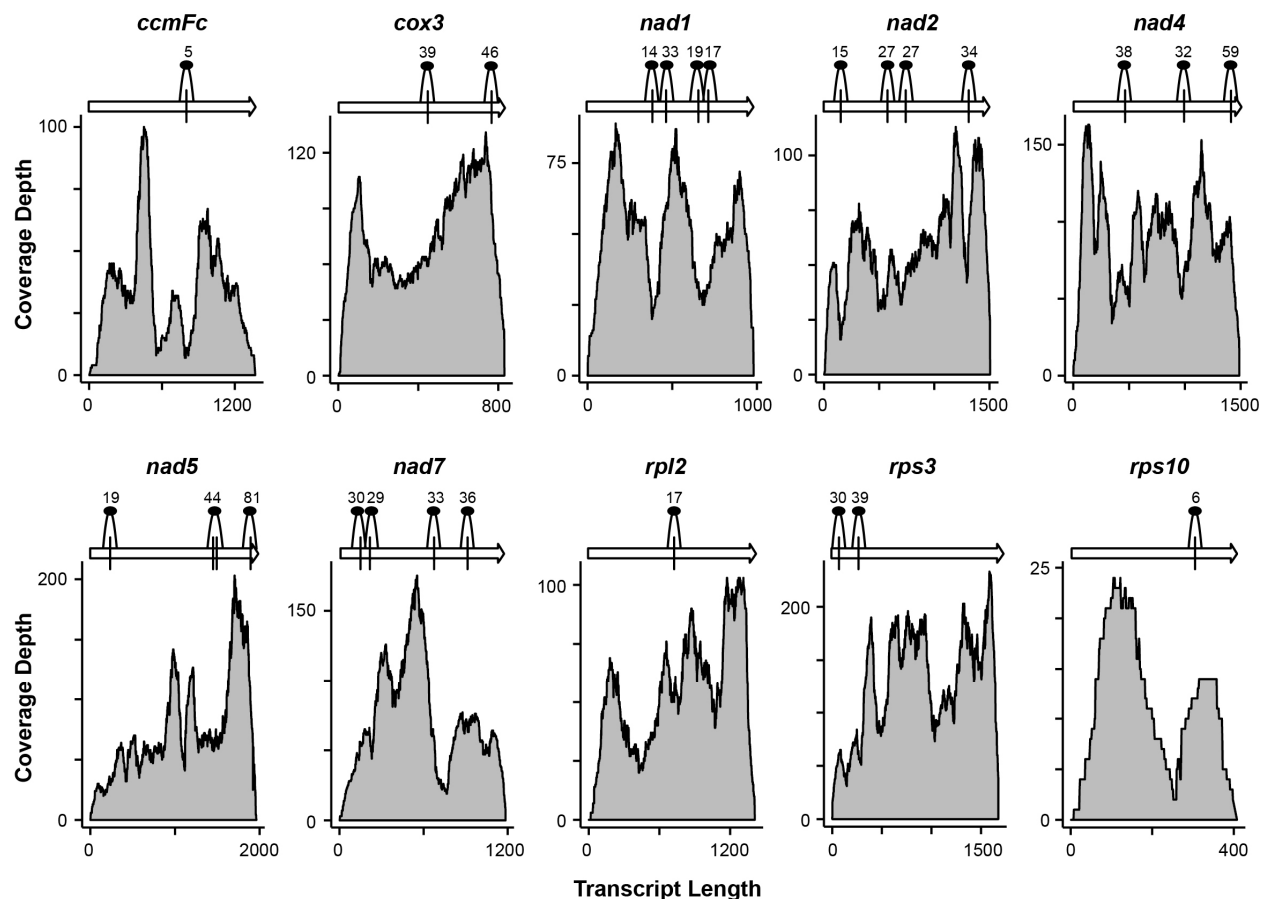


Fig. 3. Structure of intron nad1i728 and its association with maturase *matR* in gymnosperms representing the different *cis*- and *trans*-spliced intron arrangements. Genome coordinates are given for each genome fragment. Flanking exons are labeled “e4” and “e5”. Gray shading indicates segments in *P. glauca* or *A. heterophylla* that are homologous to *G. biloba*, based on blastn comparisons.

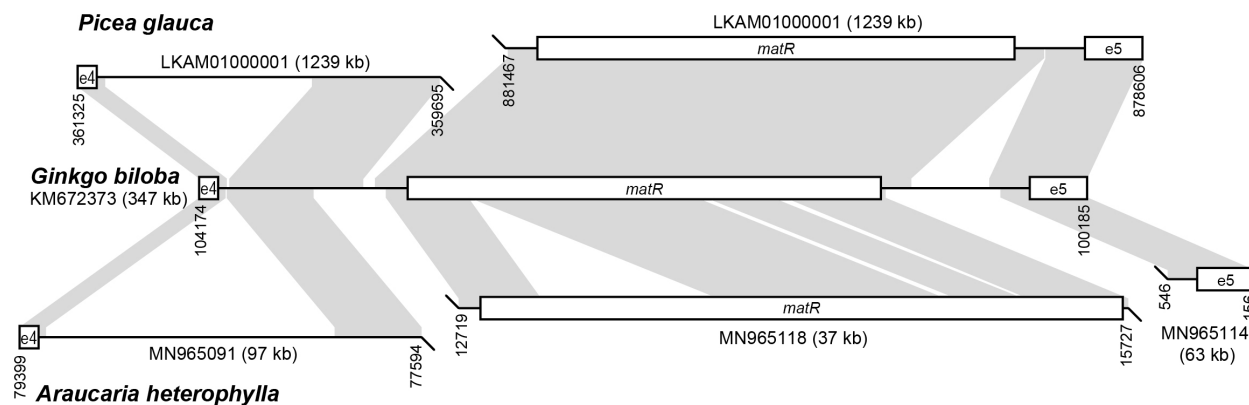
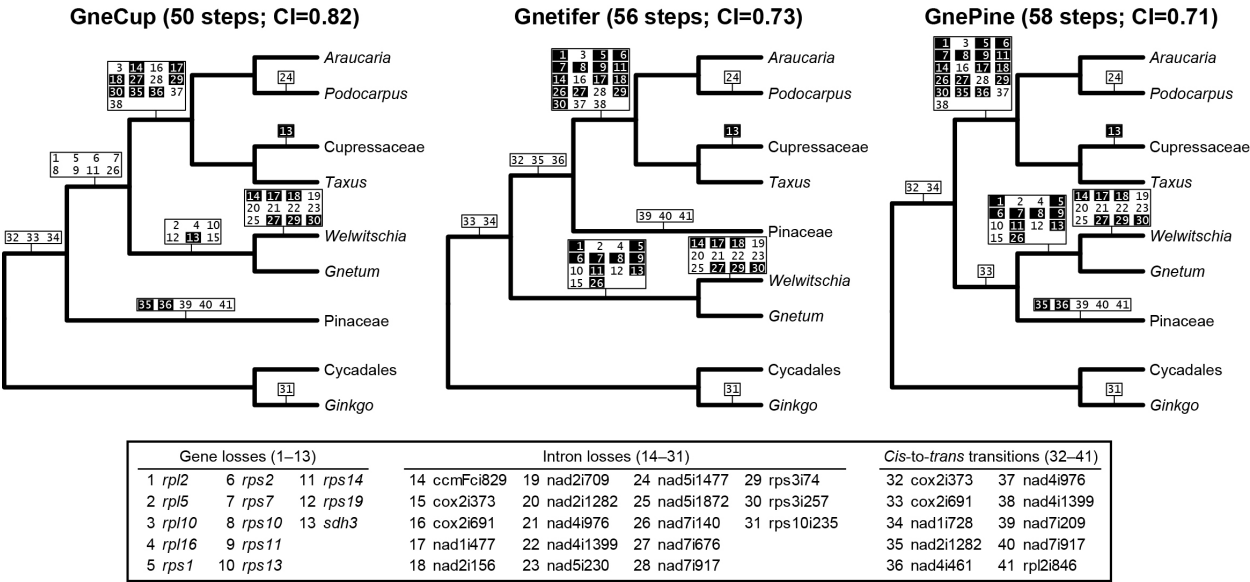


Fig. 4. Phylogenetic mapping of gene losses, introns losses, and *cis*-to-*trans*-splicing shifts. Genomic changes are mapped onto three alternative topologies (GneCup, Gnetifer, GnePine) representing the three possible positions of gnetophytes among gymnosperms. Particular evolutionary events are labeled by numbers shown in key at bottom. Convergent events are mapped using white numbering with black background.



SUPPLEMENTARY DATA

Table S1. SPAdes mitochondrial assembly statistics.

Table S2. DNA sequencing data for gymnosperms examined in this study.