Dynamics of group II introns in plant mitogenomes

and rickettsial DNA invasions in the mitogenome of *Haplopteris ensiformis*

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

Vorgelegt von

Simon Maria Zumkeller

aus

Köln, Deutschland

angefertigt am

Institut für Zelluläre und Molekulare Botanik (IZMB)

Bonn, Oktober 2022

°

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Volker Knoop {Institut für Zelluläre und Molekulare Botanik)

2. Gutachter: Prof. Dr. Walter Witke {Institut für Genetik)

Tag der Promotion:

27.02.2023

Erscheinungsjahr:

2023


UNI
BONN

Table of contents

Dynamics of group II introns in plant mitogenomes and rickettsial DNA invasions in the mitogenome of *Haplopteris ensiformis*

Table of figures

Table of abbreviations

Abbreviation	*Translation*
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
Bya	Billion years ago
cDNA	Complementary DNA
cpDNA	Chloroplast DNA
D/En	DNA binding/ Endonuclease Domain
DNA	Deoxyribonucleic acid
EGT	Endosymbiontic gene transfer
e.g.	Exempli gratia
HGT	Horizontal gene transfer
IEP	Intron encoded protein
mtDNA	Mitochondrial DNA
Mat	Maturase
mB	Monophyletic Bryophytes
Mya	Million years ago
NCBI	National Center for Biotechnology Information
ncDNA	Nuclear DNA
NLE	Non-liverwort Embryophytes
OXPHOS	Oxidative phosphorylation
PCR	Polymerase Chain Reaction
RDR	Repeat directed recombination
RT	Reverse Transcriptase
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
tRNA	Transfer RNA
tss	Template switch sensitive
xDNA	Xenologous DNA

Summary

After land plant split into bryophytes and vascular plants, the latter evolved mitochondrial genomes (mtDNA) that can be highly complex (Knoop, 2012, 2013). The outstanding complexity of vascular plant mitogenomes attributes to a combination of (1) great variability in size due to the accumulation of repetitive sequence of non-coding DNA in intergenic regions (Schuster and Brennicke, 1994; Christensen, 2018), including lateral DNA transfer of promiscuous DNA, e.g. of chloroplast ancestry (Smith, 2011; Wang, Rousseau-Gueutin and Timmis, 2012; Knoop, 2013). (2) The mtDNA structural organization or mtDNA topology is not uniform as it can appear linear or circular, as one plasmid or multi-chromosomal (Knoop, 2012; Sloan *et al.*, 2012; Sloan, 2013). (3) Gene synteny declines dramatically in vascular plants due to frequent events of genomic rearrangements over repeats (Mower, Sloan and Alverson, 2012). (4) In combination with high mitogenomic intron content (Mower, 2020), genomic rearrangements can disrupt genes that further need separate transcription of introns and exons and spliced in *trans* to join RNAs for translation (Bonen, 2008). (5) Finally, coding sequences might appear mutated on DNA level, as land plants are able to recover evolutionary conserved sites via RNA editing, the post-transcriptional exchange of pyrimidine bases (Knoop, 2011).

Ferns are one of the taxonomically largest groups of vascular plants behind flowering plants but only two mtDNAs are completely sequenced overall (Guo *et al.*, 2017). Especially the ecologically and economically relevant fern lineage of Polypodiales does not have a reference mtDNA available to date. For the family of the shoestring ferns, the Vittariaceae, previous studies have indicated outstandingly high rates of mutations (Grusz, Rothfels and Schuettpelz, 2016) and both types of RNA editing, C to U and U to C (Knie *et al.*, 2016), accompanied by rather recent dynamics in mitochondrial group II

intron content (Knie, Grewe and Knoop, 2016; Zumkeller, Knoop and Knie, 2016). Consequently, we chose the vittariacean fern *Haplopteris ensiformis* to be deep sequenced using Illumina technology for DNA and RNA.

The mitogenome of *Haplopteris ensiformis* meets all attributes of a complex vascular plant mtDNA described above (Zumkeller, Polsakiewicz and Knoop, 2022). *H. ensiformis* mtDNA is distributed onto at least nine circular chromosomes that recombine over repeats. Surprisingly, the large ribosomal RNA gene *rrnL* requires *trans*-splicing of a novel group I intron. Over a thousand events of RNA editing are required to re-establish, amongst many others, start-codons (C to U) or remove pre-emptive stops (U to C) of the generally well conserved gene complement. Most intriguingly, *H. ensiformis* contains a large amount of DNA that is not of direct mitochondrial ancestry but originates from chloroplasts and rickettsial bacteria likely via horizontal gene transfer (HGT). The latter findings are highly interesting regarding previously observed events of HGT among fern nuclear genomes and their mitochondria (Knie, Polsakiewicz and Knoop, 2015; Wickell and Li, 2020).

Assembly of the mtDNA *Haplopteris ensiformis* enabled the capture of a more complete set of mitochondrial group II introns in ferns and track intron distribution amongst all land plant lineages and related streptophytes algae, in parallel. The origin and the biological role of mitochondrial group II introns in land plants are only poorly understood. Mitochondrial group II introns, for example, generally lack proteins that are known to be intron encoded and responsible for splicing and mobility among prokaryotes (Barkan, 2007; Lambowitz and Belfort, 2015). On the other hand, among flowering plant model organisms a growing set of diverse proteins of different evolutionary origins that are nuclear-encoded, affect intron splicing in the mitochondria and chloroplast (Eckardt, 2007; Brown, des Francs-Small and Ostersetzer-Biran, 2014). It is unknown, however, equally to reason why there are introns among organellar genomes in the first place, how the nuclear-encoded sets of the protein responsible for organellar splicing coevolved.

Group II introns sequences of the major streptophyte algal and land plant lineages were systematically collected and sequence similarities were investigated. In this way, 29 close evolutionary relationships for 104 of the 161 mitochondrial group II introns, so called group II intron families, were discovered (Zumkeller and Knoop, 2022). The results

display a framework for the evolutionary origin of cytonuclear interdependency behind mitochondrial intron splicing and mobility, as paralogues of the same family might be associated with identical nuclear encoded splicing factors. Alongside, rare intron configurations like introns-within-introns (twintrons) and yet undescribed zombie-twintrons were identified in the vascular plant family of Lycopodiaceae and the bryophyte *Anthoceros agrestis* indicating evolutionary recent group II intron dynamics (Zumkeller, Gerke and Knoop, 2020).

1.0 Introduction

1.1 Chloroplast and mitochondria are organelles of green plants {Chloroplastida)

Mitochondria and chloroplast are eukaryotic compartments that evolved after two independent evolutionary events from autonomous eubacterial endosymbionts to semi-autonomous organelles with reduced genomes. Both organelles are intricately linked regarding their physiological role and evolution.

The chloroplast is the site of eukaryotic photosynthesis. The sunlight-driven conversion of water and reduction of carbon dioxide into carbohydrates generates a source of energy the production of oxygen (Finkeldey and Gailing, 2013). Mitochondria are responsible for oxidative phosphorylation (OXPHOS). Here the carbohydrate pyruvate is oxidized to carbon dioxide under the consumption of oxygen to generate a chemiosmotic gradient via an electron transport that fuels ATP-synthesis (Mitchell and Moyle, 1967).

Mitochondria are argued to be the determining factor of eukaryotic life. They have been described as “powerhouses of the cell”: They are the source of biological energy that just enabled the evolution complex eukaryotic nuclear genomes (rev. in Lane and Martin, 2010). Mitochondria have also been described as “overlords the cell” due to their controlling role for the cell cycle, deciding about a cells life, death and sex (rev. in Garg and Martin, 2016). The existence of sexes, or mating types, is often considered to be an evolutionary necessity to achieve uniparental organelle inheritance that prevents any accumulation of selfish cytoplasmic traits that might drive genetic variances and mutations among a population of chloroplast or mitochondrial genomes (Greiner, Sobanski and Bock, 2015). This is likely because genetic variances among the cytoplasmic organellar genomes can lead to lethal incompatibilities with the nuclear genome and exert species barriers (Greiner and Bock, 2013; Zupok *et al.*, 2021).

Keeping organellar genomes genetically identical (homoplasmic) is crucial to maintain functionality of chloroplast and mitochondria. Maintenance and Inheritance of both organelles is, however, mainly controlled by proteins encoded in the nuclear genome of eukaryotes.

1.1.1 Plant chloroplast and mitochondria rely on cyto-nuclear interdependency

In contrast to much larger and more complex nuclear genomes, mitochondrial (mtDNA) and plastid genomes (cpDNA) are generally displayed as a single, circular

chromosome. They are much smaller in size and contain only a small portion of genes compared to their free-living bacterial relatives. Genes that appear missing from chloroplast or mitochondrial genomes, might be identified among the nuclear genome of all eukaryotes (Maier *et al.*, 2013).

The closest free-living relatives of mitochondria are a-proteobacteria that encode around 3 to 5 thousand protein-coding genes (Boussau *et al.*, 2004). Ancestral cyanobacteria, closest free-living relatives to chloroplasts, likely encoded around 1.6 to 3.3 thousand protein-coding genes (Larsson, Nylander and Bergman, 2011). In contrast, chloroplasts only contain 120 protein-coding genes on average (Wicke *et al.*, 2011; Civan *et al.*, 2014). Likewise, the "most-bacterial-like" mitochondrial genome of the protist *Andalucia godoyi* encodes only around 100 genes (Gray *et al.*, 2020; Valach *et al.*, 2021). The massive discrepancy in gene content between organellar genomes in green plants and their free-living relatives is explained by the loss of an immense number of mitochondrial and chloroplast genes, caused by endosymbiotic gene transfer (EGT) to the nuclear genome after endosymbiosis (Gray, 1992). Hence, both organelles rely functionally on a nuclear-encoded gene complement, which must be imported to the mitochondria (Whelan 2014) (Figure 1A).

In total, the mitochondrial and chloroplast proteome of flowering plant consist of over 3000 proteins respectively (Millar *et al.*, 2005; Zoschke and Bock, 2018; Fuchs *et al.*, 2020). This makes the multienzyme complexes of chloroplasts and mitochondria, like the photosynthetic apparatus or the OXPHOS complexes a mixture of protein subunits respectively expressed within the organelle and the nucleus (Bock and Timmis, 2008; Knoop, 2012; Timmis and Wang, 2013) (Figure 1B). The chloroplast genome of autotrophic land plants encodes ribosomal RNAs (4), tRNAs (30) and approximately 80 protein-coding genes (Wicke *et al.*, 2011; Jansen and Ruhlman, 2012). Mitochondrial gene content varies greatly among eukaryotic lineages but always resembles a sub-set of genes found in the early branching eukaryotic protist *Andalucia godoyi* (Gray, 2012; Gray *et al.*, 2020). In land plants this sub-set includes around 40 subunits for the respiratory chain complexes, amongst others (Mower, 2020).

Orchestration of both photosynthesis and OXPHOS is interdependent on multiple metabolic and regulatory levels that involve the nuclear and the two organellar genomes

(Braun, 2020). Signalling exerted by the redox state of organelles and cytosol, for example, is translated into constant anterograde (nucleus-to-organelle) and retrograde (organelle-to-nucleus) import and export of proteins or metabolites (Pfannschmidt, 2010; Greiner and Bock, 2013). Accordingly, the metabolic activity and gene expression in mitochondria and chloroplast must be tightly coordinated during day-and-night shifts (Shameer, Ratcliffe and Sweetlove, 2019) or e.g. the germination of seeds, when the plant metabolism shift from breakdown of storage carbo-hydrates to autotrophic energy from photosynthesis (Best, Mizrahi and Ostersetzer-Biran, 2020).

1.1.2 Complex mitogenomes require a versatile set of proteins

The genomes of chloroplast (cpDNA) and mitochondria (mtDNA) have a lot in common but can display extreme differences as well. The highly conserved cpDNA, for example, provide excellent molecular characteristics to study evolutionary ancient phylogenetic relationships between land plants and green algae (Wicke *et al.*, 2011). Among non-vascular plants like bryophytes and streptophyte algae, mtDNAs are less conserved, but mitogenomes of vascular are enormously divergent and complex (Knoop, 2012).

Plant mitochondria require a large and likely highly specified set of nuclear encoded proteins for e. g. DNA replication and recombination for DNA repair and signalling after DNA damage (Boesch *et al.*, 2011) and RNA processing like intron splicing and RNA editing (Bonen and Vogel, 2001; Small *et al.*, 2020) (Figure 1C). The gene family that is responsible for group II intron splicing and mobility, the maturase, can be encoded within the organellar- or nuclear DNA (Guo and Mower, 2013; Brown, des Francs-Small and Ostersetzer-Biran, 2014). Due to its enigmatic gains and losses, maturase evolution has been described as the most dynamic history of all organellar genes of land plants (Guo and Mower, 2013). In addition, organellar intron splicing is contributed by a variety of nuclear encoded proteins of independent evolutionary ancestry (Brown, des Francs-Small and Ostersetzer-Biran, 2014; Schmitz-Linneweber *et al.*, 2015). It can be expected that proteins responsible for organellar intron splicing are highly diverse as intron content of mtDNA of land plant lineages varies greatly. Land plant mitochondria encode an enigmatically distributed set of 13 group I introns and over 100 different group II introns (Mower, 2020).

Gene content of land plant mitochondria is relatively high compared to metazoan mitochondrial genomes which universally only encode 13 protein coding genes (Boore, 1999). Most land plant mitochondria retained a rich gene complement of 30 to 42 protein coding genes (Mower, 2020). Apart from maturases and intron content, and in contrast to chloroplast genomes, several rather recent events of EGT from the mtDNA to the nucleus have been documented along land plant evolution. To name a few examples of variability in gene content, losses of OXPHOS genes include *atp4* and all mitochondrial tRNA genes from the early branching vascular plant *Selaginella* (Hecht, Grewe and Knoop, 2011), *atp8* in the non-vascular plants of hornworts (Li *et al.*, 2009; Xue *et al.*, 2010), *cox2* from some flowering plant legumes (Nugent and Palmer, 1991; Alverson, Zhuo, *et al.*, 2011), the entire suite of complex I genes in the hemiparasitic *Viscum* species (Petersen *et al.*, 2015; Skippington *et al.*, 2015) and the evolutionary independently re-occurring case of *nad7* across all land plant lineages (Groth-Malonek *et al.*, 2007; Xue *et al.*, 2010; Liu *et al.*, 2012; Bell *et al.*, 2014; Goryunov *et al.*, 2018),.

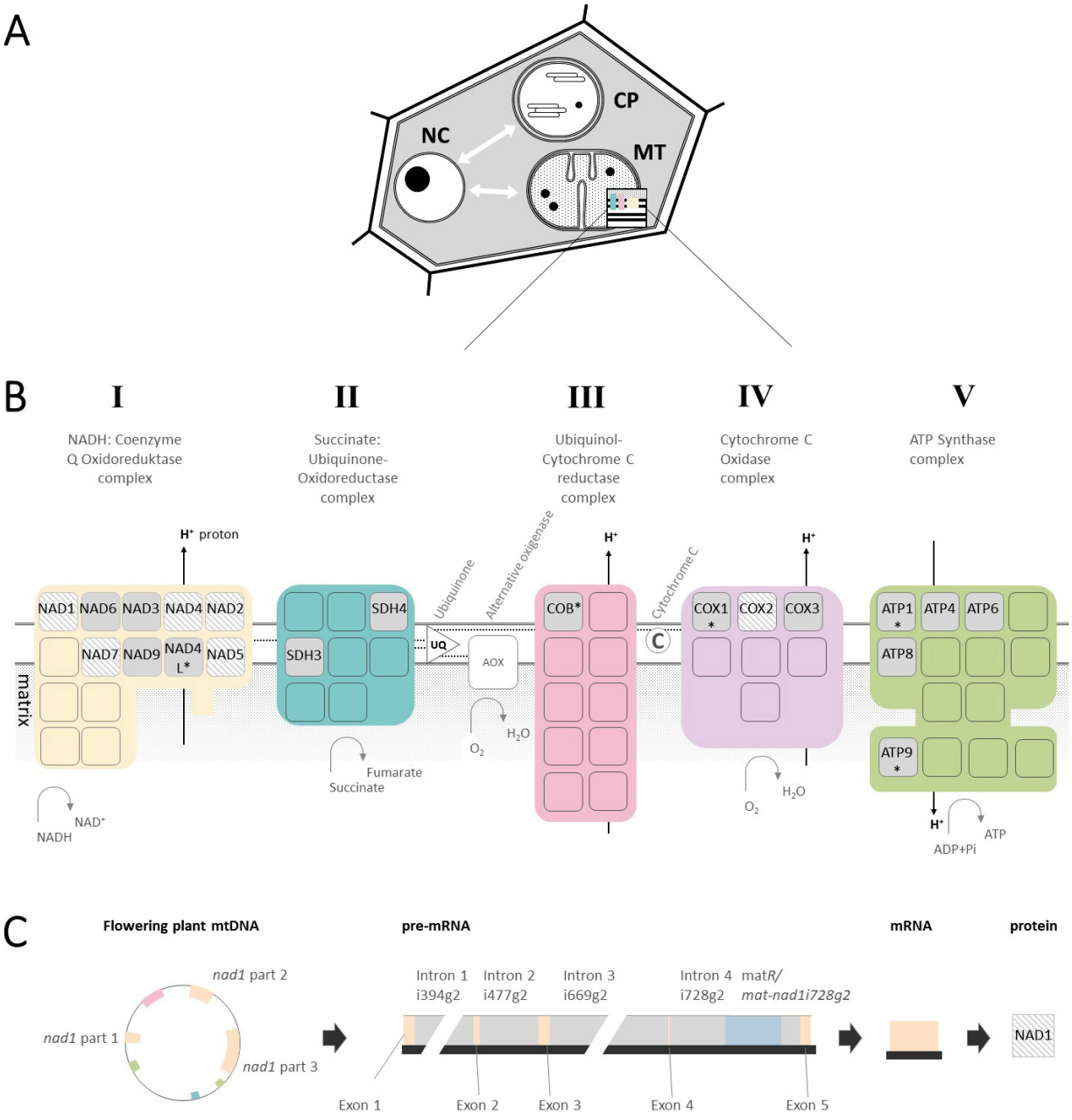


Figure 1 The three genetic compartments of a flowering plant cell and the mitochondrial genes encoding subunits of the respiratory chain complexes. A. *Nucleus (NC), chloroplast (CP) and mitochondria (MT) possess nucleoids (black dots) where DNA is organized. White arrows show anterograde (nucleus-to-organelle) and retrograde (organelle-to-nucleus) exchange of proteins.* B. *The mitochondrial OXPHOS complexes I to V (roman numerals) consist of nuclear encoded proteins (filled, coloured boxes) and mitochondrially encoded proteins (grey boxes)* (Braun *et al.*, 2014; Huang *et al.*, 2019; Braun, 2020; Zancani *et al.*, 2020)*. The respiratory chain complexes generate a proton gradient through the trans-membrane transport of electrons (blue dots) to produce ATP from ADP and phosphate ions (Pi). Mitochondrially encoded respiratory chain genes occupied by introns in flowering plants like* Arabidopsis thaliana *or* Zea mays *require splicing (grey, white striped). Non-vascular plants like the liverwort* Marchantia polymorpha *share the same set of mitochondrially encoded genes displayed here but are occupied by a different set of introns (asterisks).* C. *The gene* nad1 *(yellow) of flowering plants is separated into three*

parts across the mtDNA and occupied by two cis *arranged (grey boxes) and two* trans-*arranged (grey box interrupted) group II introns, respectively. Maturation of* nad1 *pre-mRNA requires intron splicing. Mitochondrial group II intron splicing is assisted by intron encoded* matR/ mat-nad1i728g2 *(light blue box) and different nuclear encoded proteins in* Arabidopsis thaliana (Brown, des Francs-Small and Ostersetzer-Biran, 2014; Sultan *et al.*, 2016). *

1.1.3 Fusion and fission are a necessity for mtDNA maintenance

Within the cell, the number of plastid or mitochondria, respectively, are dynamic. Thus, plastids fission, but mitochondria also fuse. lt is hypothesized that this main differences of the cellular control over the number of both organelles are responsible for the observed high rates of recombination and consequently the complexity and multivarious appearance of the vascular plant mtDNA.

The number of chloroplasts per cell is closely tied with cell division, differentiation of the respective tissue and environmental influences (Possingham and Saurer, 1969; Nagy-Déri *et al.*, 2011). Similarly, the number of mitochondria per plant cell varies and is related to the metabolic activity of a certain tissue. Plant mitochondria can rapidly change their organization. They may form scattered spheres or encage the nucleus during cell division in meristematic cells (Seguí-Simarro, Coronado and Staehelin, 2008), closely colocalize with chloroplasts in pavements cells (Barton *et al.*, 2018) and can even move from cell to cell (Gurdon *et al.*, 2016). In contrast, chloroplasts may change their shape, create stromules (Hanson and Conklin, 2020) or change cellular localization and fission, but rather not undergo fusion (Kowald and Kirkwood, 2011).

Mitochondrial dynamics are rooted in constant fusion and fission, which has been observed in the cells of early branching liverworts *Marchantia polymorpha* and likewise, the flowering plant *Arabidopsis thaliana* (Arimura et al. 2004; Sheahan, McCurdy, and Rose 2005; Sheahan, Rose, and McCurdy 2004; Paszkiewicz et al. 2017). Both processes are strictly regulated by nuclear encoded proteins, conserved in all major eukaryotic lineages (Arimura, 2018).

One reason for mitochondrial fusion is when and where DNA replication is active (Arimura, 2018). Notably, this involves DNA replication and recombination for DNA repair after DNA damage (rev. in Boesch *et al.*, 2011). DNA maintenance, involving replication and recombination is a constant necessity in metabolic active tissues. Proteins involved in organellar DNA maintenance are nuclear encoded proteins, e.g. DNA polymerases Pol1A

*: lt is an alternative oxidase (AOX) in figure 1 not an oxygenase (Schonbaum et al., 1971).

and Pol1B or recombinase protein RecA2 (Morley, Ahmad and Nielsen, 2019). These proteins factors are dually targeted to chloroplast and mitochondria, where they form so-called nucleoids in association with the organellar DNA (Sato *et al.*, 2003; Morley, Ahmad and Nielsen, 2019). Depending on physiological and developmental demands the number of chloroplast or mitochondrial nucleoids, as well as genome copies per organelle, can vary drastically (Kucej *et al.*, 2008; Arimura, 2018; Greiner *et al.*, 2020). In plant mitochondria, for example, mitochondrial fission can occur quicker than DNA replication in somatic cells, which results in mitochondria completely devoid of the mtDNA or left with only a portion of it (Preuten *et al.*, 2010; Gao *et al.*, 2018; Shen *et al.*, 2019).

Pulse field gel electrophoresis and electron-microscopy of organellar DNA in flowering plants demonstrated genome-sized circular and linear molecules, as well as sub-genome-sized circular, linear, branched and rosette-like structures (Bendich, 1993; Backert, Lurz and Börner, 1996; Oldenburg and Bendich, 2004, 2016; Cheng *et al.*, 2017). The overabundance of sub-genome-sized linear and branched molecules that have been physically observed, are interpreted as a result of high activity of DNA recombination (Gualberto *et al.*, 2014; Oldenburg and Bendich, 2015). DNA recombination relies on distinct feature of the organellar genomes, like repetitive sequences for repeat directed recombination (RDR).

1.2 Streptophytes have different mitogenome sizes, but variance in structural organizations and DNA transfer is restricted to vascular plants

Large differences in size of the mtDNA are attributed to high content of repeats among land plant lineages (Wynn and Christensen, 2019). The ancestors of land plants, the streptophyte algal mtDNA displays some variability in size too, but just like bryophytes, have generally low content of intergenic repeats. Land plants are monophyletic and include the major lineages of bryophytes*:* liverworts, mosses, hornworts, and vascular plants: lycophytes, ferns, gymnosperms and flowering plants (Figure 2A).

Zygnematophyceaen algae belong to the streptophyte algae and are likely the closest living relatives of land plants (Civan *et al.*, 2014; Wickett *et al.*, 2014; Su *et al.*, 2021). These feature mtDNAs strongly varying sizes of 215 Kbp and 69 Kbp in *Zygnema*

circumcarinatum and *Roya obtusa,* respectively (Turmel, Otis and Lemieux, 2013; Orton *et al.*, 2020). Other streptophyte algal lineages like representants of the Coleochaetophyceaen lineage, *Coleochaete scutata* and *Chaetosphaeridium globosum*, display mtDNA sizes from 400 kbp to around 55 kbp, respectively as well (Turmel, Otis and Lemieux, 2002a, 2019). Mitochondrial gene synteny, however, is well conserved on unipartite circular genomes of lineages of Charophyceaen, Coleochaetophyceaen and Zygnematophyceaen algae (Turmel, Otis and Lemieux, 2013). These three lineages belong to the monophyletic lineage of Phragmoplastophyta together with land plants (Wickett *et al.*, 2014; Su *et al.*, 2021).

Among the earliest land plant lineage, the liverwort *Marchantia polymorpha* display high similarities concerning monopartite circular structure, conservation of gene synteny, polycistronic regions and gene content to Charophyceaen-algae (Oda *et al.*, 1992; Lemieux, Otis and Turmel, 2000, 2016; Turmel, Otis and Lemieux, 2002b, 2003; Terasawa *et al.*, 2007). According to the similarity of the mtDNA, the placement of *Marchantia* as one of the earliest land plants is highly suggestive. In general, bryophyte mtDNAs are conserved and unipartite. Here, genome sizes range between 105 kbp in the model moss *Physcomitrium patens* to 150 Kbp in the liverwort *Treubia lacunosa*, but also reach over 228 Kbp in the case of hornwort *Anthoceros agrestis* (Oda *et al.*, 1992; Terasawa *et al.*, 2007; Y. Liu *et al.*, 2011; Gerke *et al.*, 2020). Further, bryophyte mitochondrial genomes generally feature low numbers of repeats (Wynn and Christensen, 2019) and conserve gene synteny (Knoop, Qiu and Yoshinaga, 2004; Y. Liu *et al.*, 2011; Liu, Medina and Goffinet, 2014).

Vascular plants do not necessarily have highly complex mitochondrial genome. Flowering plants model organism like *Arabidopsis thaliana* or *Zea mays* possess “more regular“ unipartite circular mtDNAs with sizes of 367 Kbp and 570 Kbp, respectively (Unseld *et al.*, 1997; Clifton *et al.*, 2004)*. On the other end of flowering plant mtDNA sizes, there is *Viscum scurruloideum* which has only a miniaturized mitogenome size of 66 Kbp (Petersen *et al.*, 2015; Skippington *et al.*, 2015). Within the genus *Viscum* the reduction of mtDNA size came together with complete loss of complex I mitochondrial respiratory genes and acceleration of mutational rates, only after their specialization to a parasitic

*: The mtDNAs of *Arabidopsis thaliana* can be displayed as a unipartite circular genome, but two large repeats are active in recombination and result in two subgenomic circular molecules (Klein et al., 1997). Regarding the makeup of the mtDNA of *Zea mays* in comparison to closely related *Oryza sativa,* Clifton and colleagues mention occurences for rather evolutionary recent recombination events involving native and promiscouos DNA from the chloroplast.

lifestyle (Petersen *et al.*, 2015). This strongly emphasizes, how evolutionary adaptation influences mitochondrial genomes.

On the other end, however, large and multichromosomal mitogenomes for flowering plants have been documented, too. There is for example the early branching flowering plant *Amborella trichopoda* with five mtDNA chromosomes with a net size of more than 4 Mbp (Bergthorsson *et al.*, 2004). For comparison, this makes the *Amborella* mtDNA around twice as large as the complete genome of a free-living bacteria, encoding only less than 60, in contrast to thousands of genes, respectively (Giovannoni *et al.*, 2005). This is odd, but similar sizes are around for the mitochondrial genome of *Silene conica* which, on top, consists of varying numbers of 59 to 63 different circular sub-genomic chromosomes (Wu *et al.*, 2015).

1.2.2 There is an ongoing transfer of DNA between organelles and the nucleus

The massive multipartite mtDNA size of *A. trichopoda* does not stem from repetitive sequences alone. DNA sequences can originate from horizontal gene transfer (HGT) that terms the process of exchange of genetic material across species boundaries.

Around two thirds of all mitochondrial protein coding genes in *A. trichopoda* have one or more, predominantly pseudogenized, xenologous copies that contribute to the complete mitochondrial genome (Bergthorsson *et al.*, 2004). The xenologous genes could be traced back to other flowering plant donor species, three different moss donors (Bergthorsson *et al.*, 2004) and three algal donors (Rice *et al.*, 2013). This includes a nearly full-length mitochondrial DNA of a moss species most similar to the genus *Anomodon* (Rice *et al.*, 2013).

The transfer and incorporation of foreign DNA into the nucleus from organelles or *vice versa* has been documented, as well. Most cases documenting promiscuous DNA of chloroplast ancestry within the mtDNA of vascular plants (Timmis *et al.*, 2004; Bock, 2010; Smith, 2011; Wang, Rousseau-Gueutin and Timmis, 2012). In the flowering plant, *Nicotiana tabacum* chloroplast-to-nulceus DNA transfer was determined to occur at an astonishingly high frequency of around 2×10^{-5} per cell per generation (Stegemann *et al.*, 2003). Transfer of genetic material from the nucleus to the organelles has only been found to occur in mitochondria (Knoop *et al.*, 1996; Notsu *et al.*, 2002; Alverson, Rice, *et al.*, 2011; Liu *et al.*, 2013; Chen *et al.*, 2017). Findings of promiscuous DNA with

mitochondrial ancestry in the cpDNA are few, but notable (Iorizzo, Grzebelus, *et al.*, 2012; Iorizzo, Senalik, *et al.*, 2012; Straub *et al.*, 2013; Ma *et al.*, 2015; Rabah *et al.*, 2017; Raman *et al.*, 2019). The high frequency of chloroplast derived promiscuous DNA within the mitogenome likely results from the high recombinational activity of the mitochondrial genome and DNA leakage from organelles after environmental stress, as foreign DNA sequences are integrated by microhomology-break-induced replication or non-homologous end joining (Hastings *et al.*, 2009; P. Liu *et al.*, 2011; Wang, Lloyd and Timmis, 2012)*.

1.2.3 Implications of land plant phylogeny on the evolution of the mitochondrial genome

The occurrence of large and complex mitochondrial genomes is restricted to vascular plants. This makes the evolutionary split between bryophytes and vascular plants highly interesting regarding changes to habitus and life cycle.

What it takes to maintain a vascular-plant-like complex mtDNA, in contrast to a bryophyte-like conserved mtDNA, might be best explored among early branching vascular plant linages and their closest bryophyte relatives (Knoop, 2013). Lycophytes and ferns, formerly known as paraphyletic Pteridophytes are early branching vascular plants (Figure 2A). Among them, only five complete mtDNAs are publicly available in the year 2022. However, regarding mtDNA size, the lycophytes *Isoetes, Selaginella* and *Phlegmariurus,* display intriguing divergence already by having sizes of 57 Kbp, 183 Kbp and 413 Kbp, respectively (Grewe *et al.*, 2009; Hecht, Grewe and Knoop, 2011; Liu *et al.*, 2012). While the smaller ones are likely highly recombinative active and multipartite, devoid of any gene synteny, resemble *Phlegmariurus* "the most archaic form" of a vascular plant mtDNA in the matter of topology and synteny like bryophytes. The two eusporangiate fern mtDNA feature outstandingly high repeat content on 628 Kbp bipartite mtDNA of *Psilotum nudum* and the 372 Kbp unipartite mtDNA of *Ophioglossum californicum* (Guo *et al.*, 2017).

The phylogeny of bryophyte and their placement in relation to vascular plants is currently conflicted and debated. There are two major theories (Figure 2B): A monophyletic bryophyte (mB) clade being the sister lineage to vascular plants with hornworts and setaphytes, a joint clade of liverworts and mosses, sharing a common monophyletic ancestor (Nishiyama *et al.*, 2004; Goremykin and Hellwig, 2005; Karol *et al.*,

*: NHEJ is a DNA repair mechanism independent of homologous recombination (Moore et al., 1996). Accordingly, the integration of promiscuous DNA within the mitogenome is likely a result from DNA repair after stress, involving NHEJ and homologous recombination.

2010; Renzaglia, Villareal Aguilar and Garbarry, 2018; Sousa *et al.*, 2019, 2020; Harris *et al.*, 2020; Su *et al.*, 2021). Notably, recent studies, supporting the mB-topology, moreover, included poly-A enriched transcriptome data or chloroplast coding sequence.

The other hypothesis, contrary to the mB-topology, describes bryophytes as paraphyletic, with liverworts as sister to all extant land plants and a joint non-liverwort clade (NLE) with hornworts as the sister lineage to tracheophytes (Qiu *et al.*, 2006; Qiu, 2008; Ruhfel *et al.*, 2014). This topology has rather been supported by phylogenetic data involving balanced and rich taxon sampling of coding and non-coding DNA, and specifically the use of selected mitochondrial genes under Qiu and colleagues.

The serious difference between the mB and NLE hypotheses is that whatever determines he biology of the mtDNA of land plants, it must have evolved whether juxtaposed or rather in a transition over around 500 million years (Morris *et al.*, 2018; Su *et al.*, 2021) of bryophyte-like conserved and convergent, and vascular plant-like complex and divergent mitogenomes*. The mB-topology supports that the common ancestor of land plants was likely already rather complex, more similar to Lycophytes than mosses or liverworts (Harris *et al.*, 2022).

1.2.4 Life cycle changed along streptophyte evolution multiple times

Streptophytes inherit their organellar genomes sexually following oogamy, where an externally flagellated male gamete fertilizes an immobile female gamete. Thus, streptophytes have two altering phases in their life cycle, one that is the sexually active haploid gametophyte and another that is the meiotic active diploid sporophyte (Figure 1C). Among land plants the zygote is dependent on a multicellular female gametophyte and develops towards a multicellular embryo, which is dormant until a sporophyte is produced (Niklas and Kutschera, 2010). This makes the embryo a synapomorphy of all land plants. Accordingly land plants are diplobiontic, in contrast, to streptophyte algae which are haplontic organisms, where a diploid stage only exists very briefly (Niklas and Kutschera, 2010). Bryophytes have a haplo-dominant lifecycle where the embryo is matrotrophic, while seed plants or spermatophytes have a diplo-dominant lifecycle where the gametes are matrotrophic (Qiu, Taylor and McManus, 2012). Ferns, however, are exceptional in this case, as they can generate sporophyte morphologies directly from their gametophytic cells (apogamy) and gametophyte morphologies from their

sporophyte cells (apospory) (Niklas and Kutschera, 2010). This allows both habitus to be free living (Qiu, Taylor and McManus, 2012). There is even one fern known, *Vittaria appalachiana* which was often mistaken for a liverworts, as it has lost its sporophyte life stage secondarily completely (Pinson and Schuettpelz, 2016).

*: The mB and NLE phylogenies are two different ways of depicting the evolutionary relationships of land plants, with most notable conflict on the evolution of the bryophyte linage. This can lead to conflicting conclusions about the characteristics of the last common ancestor of land plants. One specific example of this conflict is with respect to the evolution of the mtDNA. The NLE topology suggests that most likely the ancestor of land plants had a mtDNA that was similar to that of streptophyte algae and bryophytes. This cannot necessarily be concluded from the mB topology.

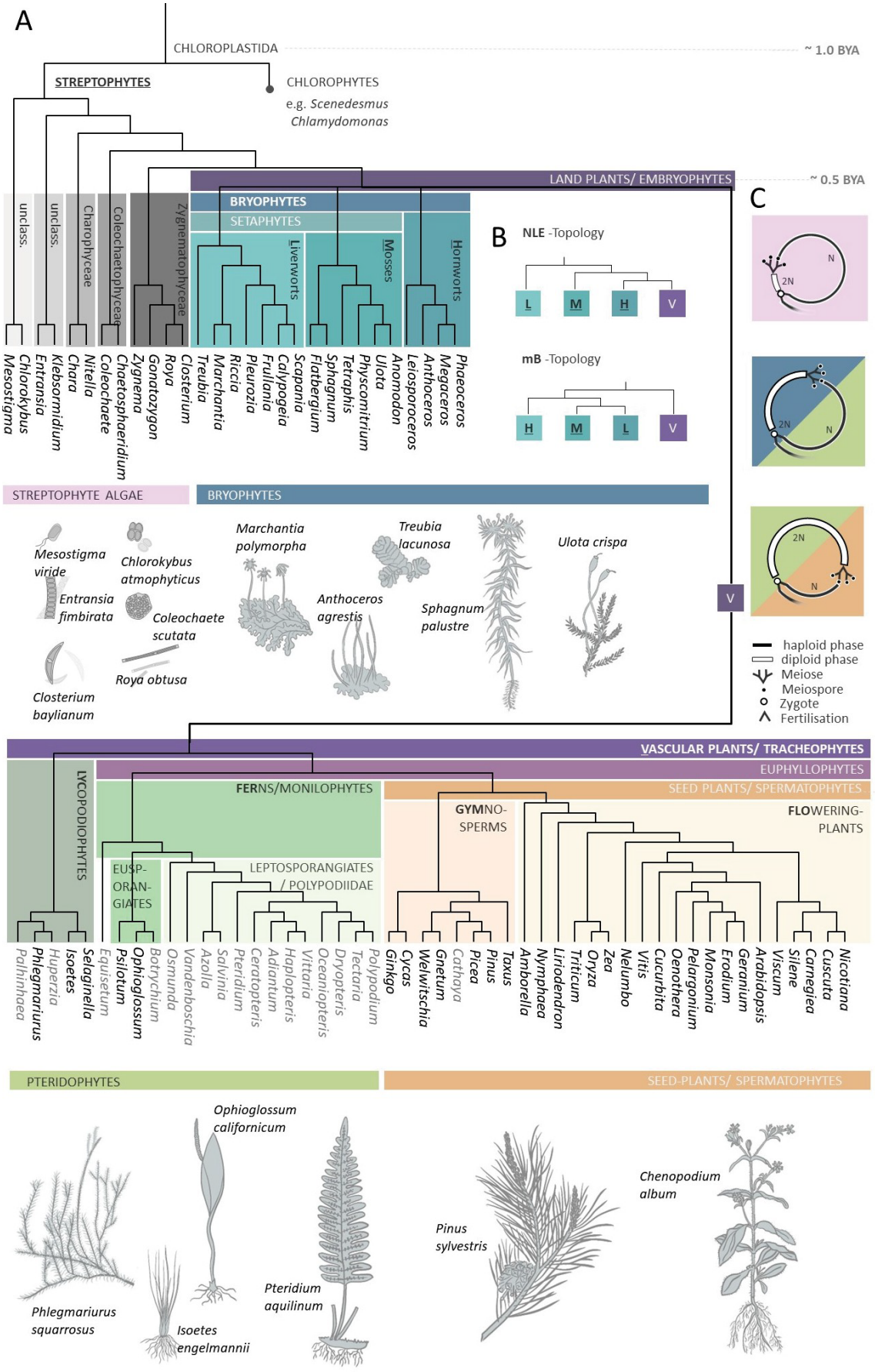

A
CHLOROPLASTIDA
~ 1.0 BYA
STREPTOPHYTES
CHLOROPHYTES
e.g. Scenedesmus
Chlamydomonas
LAND PLANTS/ EMBRYOPHYTES
~ 0.5 BYA
BRYOPHYTES
SETAPHYTES
unclass.
unclass.
Charophyceae
Coleochaetophyceae
Zygnematophyceae
Liverworts
Mosses
Hornworts
Mesostigma
Chlorokybus
Entransia
Klebsormidium
Chara
Nitella
Coleochaete
Chaetosphaeridium
Zygnema
Gonatozygon
Roya
Closterium
Treubia
Marchantia
Riccia
Pleurozia
Frullania
Calypogeia
Scapania
Flatbergium
Sphagnum
Tetraphis
Physcomitrium
Ulota
Anomodon
Leiosporoceros
Anthoceros
Megaceros
Phaeoceros
B
NLE -Topology
L
M
H
V
mB -Topology
H
M
L
V
C
N
2N
2N
N
2N
N
V
haploid phase
diploid phase
Meiose
Meiospore
Zygote
Fertilisation
STREPTOPHYTE ALGAE
BRYOPHYTES
Mesostigma viride
Chlorokybus atmophyticus
Entransia fimbirata
Coleochaete scutata
Roya obtusa
Closterium baylianum
Marchantia polymorpha
Anthoceros agrestis
Treubia lacunosa
Sphagnum palustre
Ulota crispa
VASCULAR PLANTS/ TRACHEOPHYTES
EUPHYLLOPHYTES
SEED PLANTS/ SPERMATOPHYTES
LYCOPODIOPHYTES
FERNS/MONILOPHYTES
EUSP-ORAN-GIATES
LEPTOSPORANGIATES / POLYPODIIDAE
GYMNO-SPERMS
FLOWERING-PLANTS
Palhinhaea
Phlegmariurus
Huperzia
Isoetes
Selaginella
Equisetum
Psilotum
Ophioglossum
Botrychium
Osmunda
Vandenboschia
Azolla
Salvinia
Pteridium
Ceratopteris
Adiantum
Haplopteris
Vittaria
Oceaniopteris
Dryopteris
Tectaria
Polypodium
Ginkgo
Cycas
Welwitschia
Gnetum
Cathaya
Picea
Pinus
Taxus
Amborella
Nymphaea
Liriodendron
Triticum
Oryza
Zea
Nelumbo
Vitis
Cucurbita
Oenothera
Pelargonium
Monsonia
Erodium
Geranium
Arabidopsis
Viscum
Silene
Carnegiea
Cuscuta
Nicotiana
PTERIDOPHYTES
SEED-PLANTS/ SPERMATOPHYTES
Ophioglossum californicum
Phlegmariurus squarrosus
Isoetes engelmannii
Pteridium aquilinum
Pinus sylvestris
Chenopodium album

Figure 2 Cladogram of selected streptophyte genera relevant for this thesis. The cladogram topology follows the consensus of molecular Streptophyte phylogenies (Qiu *et al.*, 2006; Leebens-Mack *et al.*, 2019; Sousa *et al.*, 2020). A. *Streptophytes are one of the two eukaryotic lineages that contain chloroplasts, Chloroplastida, besides Chlorophytes, originating more than 1 BYA* (Su *et al.*, 2021). *The cladogram of all major streptophyte lineages includes the most relevant genera that are addressed within this thesis or have been the substance of analyses amongst publications. Black lettered genera have a complete mtDNA and cpDNA available.* B. *Conflicting topologies are emphasized, due to results supporting the non-liverwort embryophyte lineage (NLE)* (Qiu *et al.*, 2006) *and a monophyletic bryophyte tree (mB)*(Leebens-Mack *et al.*, 2019; Sousa *et al.*, 2020). *The Streptophyte lineage includes early branching algae that have not been taxonomically classified, given to only recent phylogenetic positioning. The Zygnematophyceae are most likely the sister lineage of land plants* (Cheng *et al.*, 2019). *Land plants colonized land around 515 to 473 MYA* (Morris *et al.*, 2018), *after the development of desiccation tolerant zygote, the embryo* (Niklas and Kutschera, 2010). C. *All land plants are diplobiontic and have alternating haplo- and diplo- biontic stages along their life cycle. Streptophyte algae have a haplontic lifecycle, where diploidy exists only briefly as a unicellular zygote (rose)* (Wang *et al.*, 2021). *Bryophytes have a haplo-dominant lifecycle (blue) and seed- plants have a diplo-dominant lifecycle (yellow). Among pteridophytes haplo- and diplo dominant life stages do exist in parallel (green).*

1.3 The conserved streptophyte chloroplast genome

In sheer contrast to the variability of the land mitogenome can the chloroplast genome sequence easily be displayed as a unipartite circle. The cpDNA is distinctively organized and consists of two repeats in inverted orientation (IRa and IRb) that separate a large and small single-copy region (LSC and SSC), respectively, to very few exceptions (Mower and Vickrey, 2018). These IRs allow the cpDNA to appear in two isomeric, sub-genomic circular forms via homologous recombination (Kolodner and Tewari, 1979; Palmer, 1983). Streptophyte cpDNAs have a conserved size of 120-160 Kbp, with early-branching liverworts at the lower-, and ferns, gymnosperms and flowering plants at the higher end (Lemieux, Otis and Turmel, 2016; Mower and Vickrey, 2018).

There are, however, outliers within seed plants, *Cathaya argyrophylla* and *Pelargonium hortorum* which display sizes at the minimum and maximum end with 107kbp and 218kbp, respectively (Chumley *et al.*, 2006; Lin *et al.*, 2010). Also streptophyte algae cpDNA sizes range from around 107 Kbp in *Coleochaete scutata* to over 201 Kbp in *Closterium baillyanum* (Lemieux, Otis and Turmel, 2016). Although the size and gene content of the IR remained relatively stable during land-plant evolution (Zhu et al., 2016), expansions and contractions in intergenic regions mainly contributed to observed variances. Smaller, but in summation notable variances in genome size result from rare

events of gene and/or intron loss or the occurrence of small repetitive sequences in intergenic regions (rev. in Mower and Vickrey 2018). For example, a group II intron encoding the chloroplast gene *matK,* located at the tRNA gene *trnK-UUU* is an apomorphy of the joint clade of Klebsormidiophyceae and all other extant streptophytes (Lemieux, Otis and Turmel, 2016). This intron with its intron encoded gene is absent from Mesostigmatophyceae and Chlorokybophyceae.

Throughout land-plant evolution only a few major genome rearrangements of the mainly conserved gene synteny and polycistronic regions of the plasmid genome did occur (rev in. Mower and Vickrey, 2018). Most notably, there are inversion events in the ancestor of vascular plants and ferns that have changed gene synteny but not the quadripartite organization into IRs and SCs (Wolf, Roper and Duffy, 2010). For example a 30 Kbp inversion within the LSC distinguishes all euphyllophyte compared to lycophyte and bryophyte cpDNAs (Raubeson and Jansen, 1992).

Sequencing of the early branching drought tolerant vascular plant *Selaginella tamariscina* ,however, identified a chloroplast genome with two direct repeats, accompanied by other genomic rearrangements (Xu *et al.*, 2018; J. Park *et al.*, 2020). Additionally, there are three exceptional cases known among flowering plants, lacking one of the two IRs: the genus *Erodium* (Geraniaceae), the legume family Fabaceae and the saguaro cactus (*Carnegiea gigantea*) (Palmer and Thompson, 1982; Palmer *et al.*, 1987; Sanderson *et al.*, 2015; Blazier *et al.*, 2016). Nevertheless, rare exceptions can be found among streptophyte algae, as well. For example, one copy of the IRs is independently lost in *Coleochaete* and several species of Zygnematophyceae, too {Civan *et al.*, 2014; Lemieux, Otis and Turmel, 2016). Loss or inversion of one of the IRs in flowering plants, *Selaginella* or streptophyte algae*,* respectively, coincide with additional genome rearrangements and effects of genome stability, e.g. loss of small ribosomal protein (*rps*) genes 4,7,12 and 14 in *Coleochaete* (Lemieux, Otis and Turmel, 2016)*,* the complete NADH-dehydrogenase-ferredoxin-oxygenase (*ndh)* suite in *Carnegiea gigantea* (Sanderson *et al.*, 2015) or all tRNAs genes in *Selaginella* (Xu *et al.*, 2018; J. Park *et al.*, 2020).

1.4 The availability organellar DNAs is taxonomically imbalanced and led to the barcoding paradox

Currently, searching the National Center for Biotechnology Information (NCBI) nucleotide database for complete circular chloroplast genomes reveal around 24,000 entries, but only 597 for complete circular mtDNAs, of which an overwhelmingly majority belongs to flowering plants (National Center for Biotechnology Information, NCBI; GenBank database, date 21/05/2022). Intriguingly, for the vascular plant lineage of ferns there are 337 complete circular cpDNAs, but only two mtDNAs. For completion, around 45,000 entries for complete circular metazoan mtDNAs are currently available.

Having complete nucleotide sequence data for organellar genomes of land plants can be highly important for scientific research e.g. outlined in the chapters above, but for practical reasons too. Substitution of herbal food or medicine with lower quality cultivars or uncertain species is dangerous. Food fraud is a common problem and costs the worldwide industry approximately 30-40 billion USD per year (Steinberg and Engert, 2019). To guarantee food traceability and quality, DNA barcoding of plant organellar genes is more and more frequently used for automatized determination of taxonomy (Galimberti *et al.*, 2013; H. S. Park *et al.*, 2020; Thongkhao *et al.*, 2020).

Notably, complete cpDNA of flowering plants are well covered taxonomically and overall conserved DNA is homoplasmic, where the population of all plastid genomes of one organism is identical (Finkeldey and Gailing, 2013). With taxonomically highly distinctively intergenic sequence, this makes chloroplast genomes highly appealing for DNA barcoding (Li *et al.*, 2015). However, only recently has DNA barcoding led to a rather problematic detection of food fraud around the “best-selling functional food” *Cynanchum* in South Korea (Park *et al.*, 2020). It was claimed that *Cynanchum wilfordii* has been substituted with an adulterant, based on the detection of variations of chloroplast genes, followed by a lawsuit at the Korean Supreme Court and a massive plummet in sales. Eventually, it turned out to be a false claim, because the tested chloroplast *matK* gene had an ancestral non-functional copy within the mitogenome, too, misleading the consumer advocates (Park *et al.*, 2020). An event of EGT of nearly 35% of the chloroplast to the mitochondrial genome among the genus had deceived the South Korean public and was termed the bar-coding paradox.

1.4.1 Ferns are relevant ecologically and economically, but underrepresented

The above described "*Cynanchum*"-case underlines the importance of having complete nucleotide data (cpDNA, mtDNA and ncDNA) available for species, which are commercially or ecologically relevant or those, which are yet to be labelled as such. The fern lineage is largely underrepresented among the vascular plants, regarding publicly available sequence data.

Ferns are a species-rich lineage of ancient vascular plants with true leaves and no seeds or flowers. Ferns hold the second most species rich lineage of land plants, consisting of evolutionary old horse-tails (Equisetales), and the younger eusporangiate- and leptosporangiate ferns (Polypodiidae or also "real ferns") (Christenhusz and Chase, 2014). Here, polypodiid ferns hold the taxonomic majority of the around 11,000 species (Christenhusz and Chase, 2014). Their highly dissected shiny fronds, for some following fractal symmetry, make the crown group of true ferns, the polypodiales so visually appealing that they are commonly sold as ornamentals (Sharpe, Mehltreter and Walker, 2010). Besides this, ferns deserve appreciation for their diverse ecological and economical importance, regarding land-use and agriculture. For example, aquatic *Salvinia* or epipteric *Pteris* are promising candidate species to remediate heavily polluted water or soils (Yan *et al.*, 2019; Delgado-González *et al.*, 2021). In addition, the aquatic nitrogen-fixing genus *Azolla* is used as green manure for rice fields and is subsequently fed to life-stock in south-east Asia and India.

Ferns are edible for humans too. Consumption is documented by the conserved gastrointestinal tract of the glacier mummy of the neanderthal Ötzi dating back to the copper age, already (Maixner *et al.*, 2018). But also today, in the Americas and Asia the characteristic fiddleheads of the polypodiid bracken or ostrich ferns (*Pteridium aquilium* and *Matteuccia struthiopteris*) are widely used in popular dishes like Japanese warabimochi or Korean bibimbap (Kim, 2007). Here, especially the young croziers of ferns display a promising natural resource of an effective variety of antioxidants with high content of different omega 3- fatty acids (Nekrasov and Svetashev, 2021), outcompeting e.g. synthetic Vitamin E (Trolox) derivates regarding their fixation of oxygen radicals (Langhansova *et al.*, 2021).

There are, on the other hand, severe concerns regarding human consumption, as bracken ferns and other relatives can contain carcinogens like ptaquiloside (Rasmussen *et al.*, 2003). This further includes the unwelcomed consumption of fern toxins, due to their accumulation in milk or meat of life-stock (Virgilio *et al.*, 2015). Here, fern-specific compounds already induce heavy economic damage in rural areas causing polioencephalomalacia in cattle or sheep, equally to Beriberi or Wernicke-encephalopathy in humans (Chick *et al.*, 1981). This makes the polypodiid bracken fern one of the top concerning poisonous plants worldwide (Pakeman *et al.*, 1996). Intriguingly, there are, cultivars or developmental stages of bracken ferns that contain less to no ptaquiloside (Rasmussen, 2021). Consequently, a distinction between plant species or cultivars and growth stages might enable wider consumption of ferns in the future. This distinction can easily be achieved by DNA sequencing technologies. For example, DNA sequence analyses recently demonstrated that Chamomile could effectively be distinguished from poisonous morphologically highly similar adulterants and eventually, prevent the accumulation of toxic alkaloids in chamomile tea (Mahgoub *et al.*, 2022).

1.4.2 Ferns hold an evolutionary and genetically rich history

Around 64 MYA, ferns, like dinosaurs, went nearly extinct after the global mass-extinction set caused by the famous chicxulub impact (Vajda, Raine and Hollis, 2001; Schulte *et al.*, 2010). The lineage of ferns survived, thanks to mainly polypodiid ferns rapidly re-conquering earth hereafter, in parallel to the emergence of seed and flowering plants (Schneider *et al.*, 2004; Schuettpelz and Pryer, 2009).

A key factor to this interesting turn in history of ferns is that polypodiid ferns feature an extra photoreceptor, called “neochrome”, that optimizes plant-growth in shady habitats, which was likely acquired via HGT from hornworts, roughly coinciding with the main radiation-events of today’s ferns (Kawai *et al.*, 2003; Li *et al.*, 2014). Another remarkable example, involving HGT to polypodiid ferns is linked to increased resistance to insect herbivory. Recently, insecticidal protein Tma12 was isolated from polypodiid fern *Tectaria macrodonta* (Shukla *et al.*, 2016). Tma12 belongs to AA10 protein family of lytic polysaccharide monooxygenases (Yadav *et al.*, 2019), effectively degrading chitin (Shukla *et al.*, 2016) and has previously only been known to occur in fungi or bacteria (Busk and

Lange, 2015). Notably, Tma12 homologs were identified among transcriptomes of other polypodiid ferns, including one of the two nearly completely assembled fern genomes of *Salvinia* and *Azolla,* where it could be phylogenetically linked to orthologs in actinobacteria (Li *et al.*, 2018).

The challenges revolving around molecular genetics in ferns are the extreme sizes of the nuclear genomes (Clark *et al.*, 2016). In addition, ferns often tend to have polyploid sets of chromosomes, e. g. the record-holder *Ophioglossum reticulatum* contains 720 pairs of chromosomes (Khandelwal, 1990) or the royal fern genus *Osmunda* can have more than two or four sets of chromosomes (Schneider *et al.*, 2015). Consequently, most modern NGS sequencing methods struggle with overcoming the sheer redundancy of sequence information, increasing the error rate and the chance of overlooked sequences. Recent advances, however, assembled the rather small for a fern nuclear genomes of the polypodiid ferns *Ceratopteris richardii, Adiantum capillus-veneris, Alsophila spinulosa* and *Azolla filiculoides* (F.-W. Li *et al.*, 2018; Fang *et al.*, 2022; Huang *et al.*, 2022; Marchant *et al.*, 2022). All the yet assembled fern genomes, however, lack a complete mtDNA, implying an overall unknown complexity here that obstructs standardized organellar DNA assembly methods.

1.5 Organellar group II introns

Splicing of organellar introns is fundamental for functional chloroplasts and mitochondria. Group II introns are considered to be older than eukaryotes, likely first evolving in bacteria with other ribozymes in the primordial "RNA world" (Doolittle, 2013)*. Subsequently, group II introns within organelles are hypothesized to be a characteristic prokaryotic feature that has been inherited after endosymbiosis (Cavalier-Smith, 1991; Palmer and Logsdon, 1991). Group II introns and spliceosomal introns likely share a common ancestry, as well, suggesting that nuclear spliceosomal introns are descendants of group II introns (Lambowitz and Belfort, 2015).

Despite their common terminology, organellar group I and group II introns are not related and can be distinguished by their characteristic RNA secondary structures (Haugen, Simon and Bhattacharya, 2005; Zimmerly and Semper, 2015). Group II introns are commonly characterized by a conserved RNA-secondary structure of six helical domains (Michel, Kazuhiko and Haruo, 1989; Blocker *et al.*, 2005) (Figure 3A). In addition, conserved 5'- and

*: The hypothetical primordial "RNA world" predates the evolution of prokaryotes and eukaryotes. It depicts the idea that before the origin of cellular lifeforms, biological process relied on the catalytic activity of ribonucleoproteins and associated ribozymes, e.g. the ribosome. After the finding of autocatalytic splicing (forward and reverse) group I and II introns related to the eukaryotic spliceosome, their role as potential contributors the "RNA world" became highly debated (Doolittle, 2013).

3'terminal sequences and/or RNA secondary structure domain V sequence can be used for identification (Knoop et al., 1994).

1.5.1 Prokaryotic group II introns are mobile genetic ribozymes

Group II introns are an ancient class of ribozymes that in combination with an intron encoded reverse transcriptase (RT) like protein, the maturase, forms a highly invasive retroelement (Lambowitz and Zimmerly, 2011). In 2012 group II introns could be counted in around ¼ of completely sequenced bacterial genomes (Candales *et al.*, 2012). Typically low in number, but largely located on transmittable DNAs like plasmids, IS elements or pathogenicity islands, prone for bacterial HGT, the widespread occurrence of group II introns in bacteria does not appear too surprising (Waldern *et al.*, 2020). Apart from this, the distribution of group II introns in bacteria generally appears to be restricted to regions flanking genes or after *rho* elements, where effects on the host fitness are expectedly low (Dai and Zimmerly, 2002; Robart, Seo and Zimmerly, 2007; Simon *et al.*, 2008; Waldern *et al.*, 2020). Phylogenetic analyses of intron encoded maturases indicate a common origin of bacterial, chloroplast and mitochondrial group II introns (Toor, Hausner and Zimmerly, 2001; Toro and Nisa-Martínez, 2014). However, other than the "core-set" of organellar protein-coding genes, described in the previous chapter (chapter 1.1.1) there is no distinct set of group II introns that is inherited from early endosymbionts to their descendants.

1.5.2 Homologies in structure and reaction mechanism point to a common evolutionary origin of group II intron and the nuclear spliceosome

The terminal regions of group II introns resemble those of nuclear spliceosomal introns and especially domain V, as the catalytic centre, displays sequential and functional similarity to the spliceosomal snRNA U6 (Lambowitz & Belfort, 2015; Smathers & Robart, 2019). Like spliceosomal introns, group II introns splice in two transesterification reactions that usually remove the intron as a lariat and result in ligated exons (Figure 3B). The splicing reaction in both cases is highly homologous.

Group II intron splicing and nuclear spliceosomal splicing are catalysed by group II intron domain V and spliceosomal snRNA U6, respectively, each binding two divalent metal cations, initiating a nucleophilic transesterification of the intron's first nucleotide (Peebles

et al., 1986; Schmelzer and Schweyen, 1986; van der Veen *et al.*, 1986; Lambowitz and Zimmerly, 2004). Domain IV does not directly participate in the splicing reaction but can harbour an open reading frame (ORF) that encodes an intron associated protein (IEP), the maturase (Figure 3C) (Michel and Ferat, 1995; Mohr and Lambowitz, 2003). In yeasts phylogenetic similarities found to the ubiquitously conserved spliceosomal factor PRP8 and maturase further support common evolutionary origin {Dlakic and Mushegian, 2011} (Figure 3C). PRP8 is essential for chaperoning the ribozymic core of the spliceosome (Liu, Query and Konarska, 2007), similar to group II intron associated maturase (Zhao and Pyle, 2017). The numerous structural and functional similarities between group II introns and the spliceosomal PRP8 (Smathers and Robart, 2019) indicate that a progenitor maturase was likely to have expanded its intron target range and evolved into one of the largest molecular complexes with more than 150 proteins, defining the nuclear eukaryote genome (Wahl, Will and Lührmann, 2009).

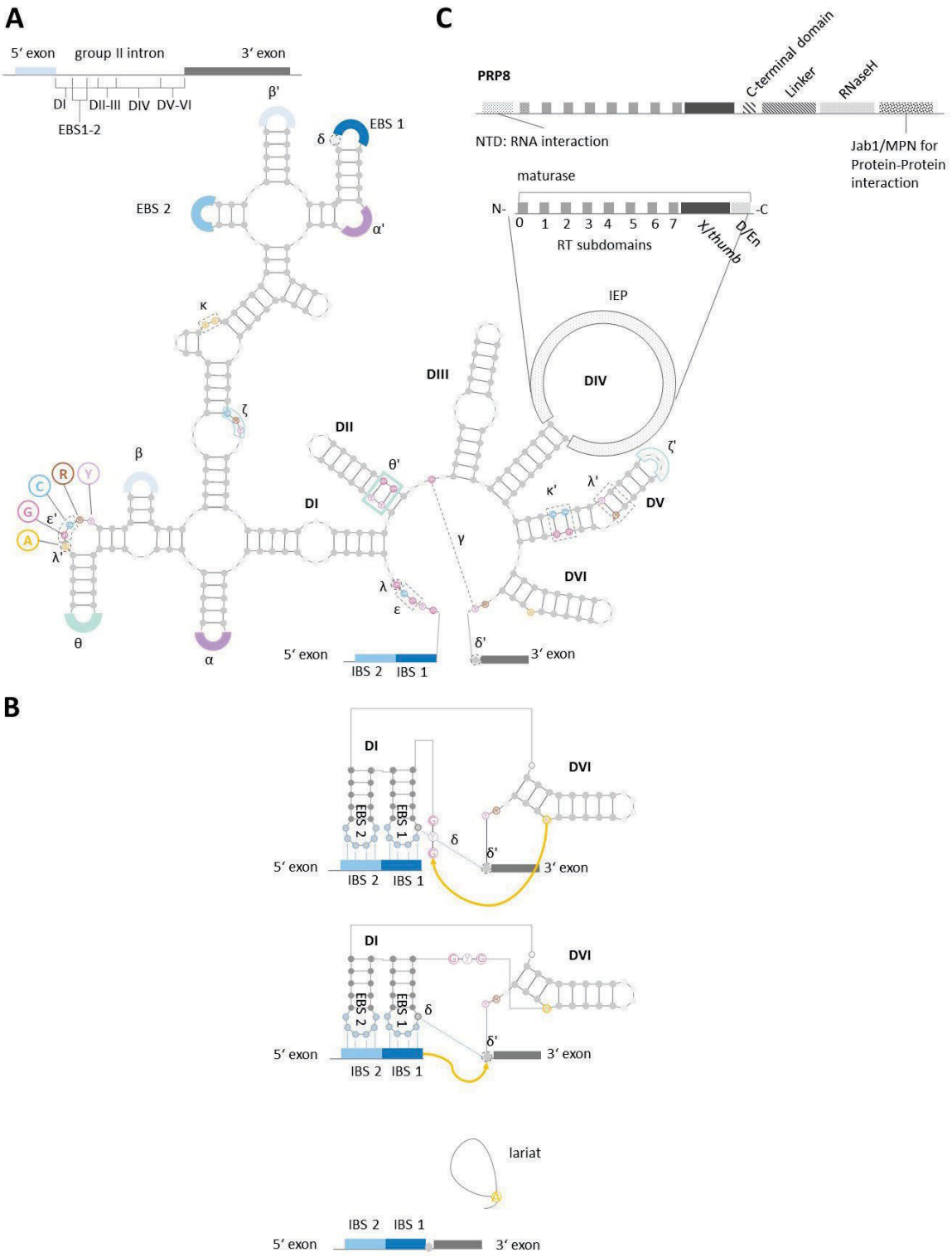


Figure 3 Group II intron structure, splicing and associated proteins. Characteristic RNA secondary structure of six domains (roman numbering) of mitochondrial group IIA intron (A) *and corresponding two-step transesterification with lariat formation (B).* A. *Coloured boxes indicate tertiary interaction sites (greek letters), exon binding sites (EBS, light and dark blue boxes) and intron binding sites (IBS, light and dark blue boxes). Highlighted*

nucleotides are generally conserved features of group II intron mitochondrial class A1. The loop of group II introns domain VI of land plants rarely encodes an intron encoded protein (IEP), the maturase (mat) which consist of a N-terminal reverse transcriptase (RT, RT-subdomains RT0-7) and C-terminal X-domain/thumb (X-domain/thumb-domain), which assists in splicing in vivo. In addition, group II intron maturase can also feature a C-terminal DNA-binding /endonuclease (D/En) domain. B. *Group II intron splicing usually follows a two-step transesterification reaction (yellow arrows). The first step of splicing involves a bulged adenosine (yellow ovals) located in domain VI, whereas its 2'-hydroxy group attacks the 5' splicing site phosphodiester in a nucleophilic attack. Hereby, the group II intron's first nucleotide is covalently linked to the bulged DVI-adenosine and the 5'-exon is freed from the transcript. In the second step of splicing EBS/IBS and delta/delta' complementary bindings (dotted blue lines) guide the 3'-end of the freed 5'-exon into a transesterification with the 5'-end of the 3'-exon. Consequently, both exons are ligated by the intron RNA which is released in is characteristically lariat formation, eventually.* C. *Group II intron encoded proteins (IEPs) are related to Yeast spliceosomal protein PRP8. Background information about the group II intron splicing reaction, conserved nucleotide positions and associated proteins were taken and edited from review articles* (Lambowitz and Zimmerly, 2011; Lambowitz and Belfort, 2015; Schmitz-Linneweber *et al.*, 2015).

1.5.3 Group I and group II introns are hypothesized as selfish genetic elements which secondarily acquired regulatory functions

Organellar group I and group II introns are hypothesized to be descendants of prokaryotic homologs that can splice and invade intron-free alleles and ectopic sites (Hausner, Hafez and Edgell, 2014; Zimmerly and Semper, 2015). Accordingly, group I and II introns fall into the category of selfish genetic elements as they enhance their transmission at the expense of other genes in the genome, even if this has no or a negative effect on organismal fitness (Doolittle and Sapienza, 1980; Orgel and Crick, 1980).

In the long run, selfish elements might undergo subsequent adaption as regulatory elements after autonomous spread throughout genomes (Britten and Davidson, 1969; Orgel and Crick, 1980). Compared to intron-free alleles, introns naturally impact

transcription passively as RNA-polymerization is a cost- and time intensive process and introns generate a lag in the feedback of gene expression that results in stronger up-regulation (Chorev and Carmel, 2012). Current data on plant mitochondrial transcription suggests a continuously, but relaxed and loosely controlled process, (rev. in Hammani and Giegé, 2014), implicating that those post-transcriptional mechanisms like intron splicing are significant for the number of mature RNAs, (rev. in Hammani and Giegé 2014).

There is a group I and group II introns in T4 phage and prokaryotes, which have been described to self-splice independent of proteins and under distinct environmental conditions like oxidizing agents, reactive oxygen, temperature, or DNA damage (rev. in Belfort, 2017). For example, the chlorophyte *Chlamydomonas reinhardtii* encodes a group I intron within the cpDNA gene of photosystem II protein D1 (*psbA*) that self-splices depending on light availability and/or the redox state of the chloroplast (Deshpande, Bao and Herrin, 1997). In the dark, the splicing efficiency of this group I intron is reduced, accumulating unspliced precursors of the *psbA* transcript, whereas exposure to light increases splicing efficiency, making mature *psbA* transcripts available for correct translation to replace photo-damaged proteins (Deshpande, Bao and Herrin, 1997). Another example of an ancestrally mobile group I intron being domesticated as regulatory element is found among intron paralogues within *cox1* and *nad5* genes of metazoan Hexacorallia (Johansen *et al.*, 2020). The *nad5* transcript of hexacoral can be spliced in *cis, trans* or circularly.

Similar observations have been made for mitochondrial group II introns of wheat (*Triticum aestivum)*. Here, the ratio of un-spliced to spliced mitochondrial transcripts varies under different developmental stages, like seed development (Li-Pook-Than, Carrillo and Bonen, 2004) or environmental stimuli, like temperature (Dalby and Bonen, 2013). Interestingly, different temperatures did also affect the group II intron splicing mechanistically, producing unconventional linear or circular intron RNAs, or conventional lariat intron RNAs. The authors Dalby and Bonen hypothesize that mitochondrial group II intron splicing might play a regulatory function during germination, when seeds feed energy produced by mitochondria heterotrophically.

1.5.4 Land plant mitochondrial introns are highly diverse but appear lineage specific

Understanding why likely ancestral mobile organellar introns have survived during the evolution of streptophyte organelles must be approached from different angles. This work will focus on the largest and most diverse group of organellar introns: group II introns. Firstly, group II introns have a long evolutionary history and for streptophyte mitochondrial genomes, the observed intron diversity and distribution remain to be elucidated (Turmel, Otis and Lemieux, 2013; Mower, 2020). Secondly, group II intron splicing and mobility is a process that relies on the catalytic capabilities of the RNA itself, and additionally on protein factors, which are the maturase or nuclear encoded splicing factors (Brown, des Francs-Small and Ostersetzer-Biran, 2014). The complex interaction of nuclear factor and mitochondrial group II intron splicing might reveal important regulatory potential (Dalby and Bonen, 2013) and important parallels to the evolution of the spliceosome (Schmitz-Linneweber *et al.*, 2015).

The distribution of group II introns in streptophyte chloroplast and mitochondrial genomes do indicate two different fates in evolution. In land-plant chloroplasts, a set of 20 group II intron is conserved and can be efficiently used for phylogenetic analyses (Kelchner, 2002). The distinct set of chloroplast group II introns emerged along early streptophyte evolution, with only a few extraordinary gain/loss dynamics in Zygnematophyceae (Lemieux et al., 2016). In contrast, mitochondrial genomes of streptophytes display the largest number and diversity of group II introns besides bacteria (Zimmerly and Semper, 2015). Especially land-plant mitochondria contain a phylogenetically rich and diverse set of group II introns ranging from 20 to 40 introns for each of the major land-plant lineages (Knoop, 2004).

Among more the 100 different mitochondrial group II introns of land plants, however, only one intron within the *atp9* gene is shared among the three bryophyte and the tracheophyte lineages (Mower, 2020). The enigmatic distribution of mitochondrial group II intron in land-plant has previously been explained by numerous gain events at the earliest stages of the land-plant lineages, followed by a phase of stagnation and losses (Qiu *et al.*, 1998, 2006; Pruchner *et al.*, 2002; Dombrovska and Qiu, 2004; Qiu and Palmer, 2004; Groth-Malonek *et al.*, 2005; Volkmar and Knoop, 2010; Volkmar *et al.*, 2012). Accordingly, the ancestral mitochondrial genome of land plants must have had

large gene sizes, resulting from the hypothetical superset of group II introns. However, the mitochondrial genomes of streptophyte algae are generally smaller than 100 Kbp and harbour low numbers of group II introns, e.g. 3 and 14 in the earliest branching streptophyte algae *Mesostigma viride* and *Chlorokybus atmophyticus,* respectively (Turmel, Otis and Lemieux, 2002b, 2007). Consequently, the hypothetical large intron-rich land-plant mitochondrial genome must have emerged after the split from streptophyte algae, but none of the major land plant lineages supports such a phase throughout evolution.

Sequence similarities beyond generally conserved regions like domain V have been recognized before for mitochondrial group II introns among the liverwort lineage and flowering plant lineages and have not been investigated any further (Ohyama *et al.*, 1993; Lippok, Brennicke and Wissinger, 1994). The comparative analyses of flowering plant group II introns and bacterial counterparts indicated their incapability of autocatalytic splicing or transfer-mobility due to a degeneration of intron structure and maturases (Barkan, 2007; Zimmerly and Semper, 2015).

The analyses of mitochondrial genes in ferns indicated that group II introns within the *rps1* and *atp1* gene might originate from copy events of intron within the *rpl2* and *rps3* gene, respectively, however (Knie, Grewe and Knoop, 2016; Zumkeller, Knoop and Knie, 2016). One reason was that the two latter group II introns are present to also other vascular plant lineages than ferns, while the other, became clear with a nearly completely resolved molecular phylogeny for the fern lineage, placing horsetails as a sister group to all other extant ferns and Marattiales as sister group to all true ferns (Knie *et al.*, 2015). Accordingly, the distribution of group II introns within the *rps1* and *atp1* gene displayed a molecular synapomorphy of all extant ferns, but horsetails and the lineage of Marattiales and true ferns, respectively (Wikström and Pryer, 2005; Knie, Grewe and Knoop, 2016; Zumkeller, Knoop and Knie, 2016). Both introns gained late within the evolution of ferns and significant sequence similarity, in addition, supported the hypothesis that the *atp1* and *rps1* intron originated from lateral intron transfer, despite an associated maturase that could promote mobility or splicing was not known.

1.6 Mitochondrial group II introns require multiple proteins for splicing

In prokaryotes, the group II intron splicing reaction is assisted by an intron-encoded maturase, preferentially acting in cis (Zimmerly and Semper, 2015). Land plant chloroplast genomes, however, encode and express a single well-studied maturase that acts in *trans* on a broader intron target range, the *matK* (Hausner *et al.*, 2006; Barthet and Hilu, 2007). Like *matK,* there is one maturase protein frequently encoded in euphyllophyte mitochondria, the *mat-nad1i728g2*, also known as *matR*. The *matR* from *A. thaliana* and close relatives was shown to be involved in the splicing of numerous mitochondrial group II introns, rather than acting only on its well-conserved host intron nad1i728g2 (Sultan *et al.*, 2016).

In *Arabidopsis*, four nuclear genes encoding for maturases (nMAT1-4) have been identified so far (Mohr and Lambowitz, 2003). GFP localization studies, followed by reverse genetical experiments demonstrated that the nuclear encoded maturases facilitate the splicing of different subsets of mitochondrial group II introns (Keren *et al.*, 2009, 2012; Cohen *et al.*, 2014; Shevtsov-Tal *et al.*, 2020). The list of nuclear-derived proteins that are involved in organellar group II intron splicing is diverse and growing. There are for example pentatricopeptide (PPR) proteins, that are known as key factors for organellar RNA maturation, with over 450 known members in flowering plants (Lurin *et al.*, 2004; O'Toole *et al.*, 2008). Other protein families that have been identified as splicing factors are the chloroplast RNA splicing and ribosome maturation (CRM) proteins (Barkan *et al.*, 2007), plant organellar RNA recognition proteins (PORR) (Kroeger *et al.*, 2009) or DEAD-box helicases (Putnam and Jankowsky, 2013). Factors like mitochondrially localized DEAD-box helicase pmh2 in *A. thaliana* are involved in the splicing of multiple group II introns (Köhler, Schmidt-Gattung and Binder, 2010). Interestingly, *pmh2* affects the same sub-set of introns as *nmat2* does and the double *pmh2*/*nmat2* mutant shows increased splicing deficiency and a more severe phenotype than knock-out mutants of the respective maturase alone (Zmudjak *et al.*, 2017). Similarly, RNAi experiments on CRM-type protein *mcsf1* also resulted in impaired splicing along with the incorrect assembly of respiratory complex I and IV (Zmudjak *et al.*, 2013). Additional experiments in *Z. mays* could show that the mcsf1 ortholog recruits PPR protein SMR1 for correct splicing of mitochondrial introns (Chen *et al.*, 2019).

1.7 Land plant organellar RNA editing

The term “RNA editing” combines all processes that occur *in vivo* to change any RNA transcript by insertions, deletions or substitutions (Knoop, 2011). Processes that extend RNA characteristics over the four standard nucleotides and do not bind complementarily to one of those are excluded from that definition. P-type PPR proteins consist of PPR motifs of 35 amino acid length and are mostly involved in RNA transcript splicing, cleavage or stabilization, whereas an sub-variant of PPRs, the PLS-type (L for long motifs, S for short motifs) is likely mainly responsible for plant organellar RNA editing (rev. in Small *et al.*, 2020).

The most abundant type of RNA editing in land-plant organelles is the deamination of cytosines to uridines and *vice versa* (rev. in Schmitz-Linneweber and Small, 2008; Schallenberg-Rüdinger and Knoop, 2016; Small *et al.*, 2020). RNA editing occurs predominantly in mRNAs, but is observed in untranslated regions (UTRs) (Hiesel *et al.*, 1989), tRNAs (Marchfelder, Brennicke and Binder, 1996; Grewe *et al.*, 2009), rRNAs (Hecht, Grewe and Knoop, 2011) and organellar introns (Carrillo, Chapdelaine and Bonen, 2001; Castandet *et al.*, 2010; Farré *et al.*, 2012; Oldenkott *et al.*, 2014; Knie, Grewe and Knoop, 2016; Zumkeller, Knoop and Knie, 2016). Most abundantly occurring in codon sequences, C-to-U and U-to-C RNA editing act as a correction mechanism to restore evolutionarily conserved codon identities, mutated on DNA level, resulting e.g. in the removal of pre-emptive stop-codons (U to C; UAA, UAG and UGA to CAA, CAG and CGA, respectively) or introduction of start codons (C to U; ACG to AUG) (Maier *et al.*, 2008).

Historically initially recognized C-to-U (forward) RNA editing likely originated in the common ancestor of land-plants, as it can be found in all land plant organelles, except the marchantiid liverworts which likely lost it secondarily (Schallenberg-Rüdinger and Knoop, 2016). The mechanistic link between C to U RNA editing and PLS-PPR proteins with a C-terminally encoded DYW motif, highly resembling a cytidine deaminase has been most extensively studied in the model moss *Physcomitrium patens* formerly known as *Physcomitrella.* In the moss *Physcomitrium patens* ten nuclear encoded DYW-PLS-PPR proteins are known to be responsible for 2 and 11 RNA editing sites in the chloroplast and mitochondria, respectively (Rüdinger *et al.*, 2009). Whereas one, PPR43, is a splicing factor, nine were characterized as RNA editing factors and could be specifically assigned

to the 13 RNA editing sites in *Physcomitrium* organelles (Rüdinger *et al.*, 2011; Ichinose *et al.*, 2013; Ichinose, Uchida and Sugita, 2014). No other components were so far identified to be necessary for RNA editing in the model moss suggesting that it represents the ancestral variant of the C-to-U editing machinery. Recently, this has been underlined by functional expression of *P. patens* PLS-DYW- type PPR56 and PPR65 for C to U RNA editing expressed in bacteria *E. coli* (Oldenkott *et al.*, 2019) and human cells (Lesch *et al.*, 2022).

An explanation for the evolution U-to-C (reverse) RNA editing is more challenging, as it is only present to hornworts, lycophytes, and ferns. According to the NLE-topology reverse RNA editing has been likely gained once in the common ancestor of hornworts and vascular plants and was independently lost among seed-plants. According to the mB-topology reverse RNA editing has likely been gained amongst the common ancestor of land plants and was independently lost at least twice in seed-plants and setaphytes. In addition, a cognate factor, like the ancestral PLS-DYW subtype for forward RNA Editing, has not been identified for reverse RNA editing. Among the hornworts *Anthoceros agrestis* nuclear and organellar genome analyses has identified putative candidate PPR proteins of the KPAxA subtype based on target analyses (Gerke *et al.*, 2020). Nonetheless, more reference species are required for comparative genomics and target analyses. Ferns might offer an opportunity to further study nuclear encoded PPRs candidate proteins and their targets, as ferns feature much higher taxonomic richness than hornworts and feature very low to highest rates of reverse RNA editing exceeding forward RNA editing (Knie *et al.*, 2016).

2.0 Results

2.1 Summary: A functional twintron, "zombie" twintrons and a hypermobile group II intron invading itself in plant mitochondria

The manuscript *„A functional twintron, "zombie" twintrons and a hypermobile group II intron invading itself in plant mitochondria* "(Zumkeller, Gerke and Knoop, 2020) have been submitted successfully for publication in the journal of *Nucleic Acids Research*. It describes the discovery of group I and group II intron configurations like twintrons, introns-within-introns of mitochondrial genomes in the plant lineage.

Identification of twintrons among land plant mitogenomes required careful sequence inspection, taxon sampling, experimental enrichment of pre-mature organellar

RNA and dedicated PCR amplification strategies on the cDNA level. Hereby, it could be confirmed that land plant mitochondrial twintrons consist of an external and internal intron that are being spliced whether serially or spliced in one, due to a degenerated internal intron. Where internal intron splicing is not necessary for splicing of the complete twintron anymore, the term “zombie” twintrons (half-dead, half-alive) has been proposed.

Serially spliced twintrons and zombie-twintrons have been identified in the hornwort lineage and the lycophyte order of Lycopodiaceae. Likely responsible for the formation of on zombie-twintron in Lycopodiaceae mitogenomes, a hypermobile group II intron invading ten different loci including itself has been identified. Similarly, a distinct group of introns gave rise to multiple introns and two zombie twintrons in hornworts. These findings are highly interesting regarding the transition between ancient bryophyte and vascular plant lifestyles and coinciding changes in the appearance of the mitogenomes here. Among the rather conserved mtDNA of hornworts, the here reported dynamics of ancient group II intron mobility appear surprising.

Simon Zumkeller and Phillipp Gerke did analysis of sequence data, Simon Zumkeller did nucleic acid preparations, PCR amplifications and molecular cloning, Volker Knoop designed and supervised the study and wrote the manuscript, Simon Zumkeller and Volker Knoop figures and edited the final manuscript.

2.2 Summary: Rickettsial DNA invasions and a scrambled rRNA cluster with a trans-splicing group I intron: The highly unorthodox mitogenome of the fern Haplopteris ensiformis

The manuscript "*Rickettsial DNA invasions and a scrambled rRNA cluster with a trans-splicing group I intron: The highly unorthodox mitogenome of the fern Haplopteris ensiformis*" has been submitted to the journal *Communications Biology* (Zumkeller, Polsakiewicz and Knoop, 2022). It describes the highly complex mitochondrial genome of *Haplopteris ensiformis* a polypodiid fern belonging to the shoestring fern family, Vittariaceae. The manuscript provided here is currently available as a pre-print.

A combination of deep Illumina sequencing, a metagenome-assembly strategy and specifically developed template-switch-sensitive PCR strategies allowed the complete assembly of *H. ensiformis* organellar genomes. Hereby, several characteristics were determined that explain why the assembly of leptosporangiate fern mtDNA presents as difficult. Both the chloroplast and mitochondrial genome of *H. ensiformis* feature exceptionally high rates of C-to-U and, most notably, U-to-C RNA editing, obscuring the identification of organellar genes on the DNA level alone due to altered and undetectable start and/or stop codons. However, transcriptome analyses revealed that *H. ensiformis* has a well conserved gene and intron complement, both for cpDNA and mtDNA. The mitogenome only lacks the *ccm* gene suite and three group II introns which have been found in eusporangiate ferns before (Guo *et al.*, 2017). In addition, *H. ensiformis* mitochondrial gene large ribosomal RNA *rrnL* is occupied by a novel group I intron that is spliced in *trans.*

The findings presented here are surprising to some extent, as the mtDNA of *H. ensiformis* documents the high activity of recombination across a likely multichromosomal genome, shaped by a massive number of insertions of chloroplast DNA and bacterial DNA, following recent and ancient events of lateral and horizontal gene transfer, respectively. The insertion of bacterial DNA includes previously recognized chlamydial tRNAs (Knie, Polsakiewicz and Knoop, 2015) along 40 insertions of bacterial DNA, partly highly similar to Rickettsial DNA. The insights into the *H. ensiformis* mitogenome are an important addition to the field regarding its taxonomical position and its highly unusual molecular features.

Simon Zumkeller did design experiments, wet lab work and established bioinformatic pipelines. Monika Polsakiewicz helped with nucleic acid preparations and molecular cloning. Simon Zumkeller and Volker Knoop analysed data and prepared figures. Volker Knoop wrote the manuscript and all authors edited and approved the final manuscript version.

2.3 Summary: Categorizing 161 plant mitochondrial group II introns into 29 families of related paralogues finds only limited links between intron mobility and intron-borne maturases

The manuscript *„Categorizing 161 plant mitochondrial group II introns into 29 families of related paralogues finds only limited links between intron mobility and intron-borne maturases*" (Zumkeller and Knoop, 2022) has been submitted for publication in BMC Ecology and Evolution. It describes how most of the diversity of land plant mitochondrial group II introns originates from lateral copying events for more than half a billion years of evolution. Maturases found among land plant mitochondria, however, overly lack functional domains responsible for promoting group II intron mobility. The manuscript provided here is currently available as a pre-print.

This study required taxonomically balanced sampling of high quality complete mitochondrial genomes across the streptophyte lineage, extraction of group II intron sequence content and similarity search using the BLAST tool suite (Altschul *et al.*, 1990). Based on criteria defined within this manuscript and phylogenetic analyses, 104 of 161 mitochondrial streptophyte group II introns were clustered into 29 families that reflect a common evolutionary relationship, respectively. In addition, an extensive search for mitochondrial group II introns led to the detection of yet overlooked functional introns and fossilized introns within intergenic space. While most mitochondrial group II introns lack an intron-borne maturase that could promote cognate splicing or mobility intron, paralogues of one group II intron family, however, can share a set of common nuclear encoded splicing factors of different evolutionary backgrounds that have meantime been identified in model organisms.

Simon Zumkeller collected sequence data and established bioinformatic pipelines. Simon Zumkeller and Volker Knoop analysed data and prepared figures. Volker Knoop wrote the manuscript and all authors edited and approved the final manuscript version.

3.0 Discussion

3.1 From the first mitogenome of a polypodiid fern to the next?

Ferns have a rich potential in molecular biology and their occupation of diverse ecological niches. Most notably, and even with only very few genomic studies at hand, there are already numerous indications for an overly common occurrence of horizontal gene transfers in the fern lineage that could even be relevant for future biotechnological applications (rev. in Wickell and Li, 2020). With the first completely assembled polypod fern mitochondrial genome of *H. ensiformis* this work adds another chapter to the highly interesting evolutionary story of this ecologically crucial but taxonomically understudied lineage. Due to its manifold complexities within the mtDNA, *Haplopteris* may allow quicker assembly of a complete mtDNA for the already four available nuclear genomes (F.-W. Li *et al.*, 2018; Fang *et al.*, 2022; Huang *et al.*, 2022; Marchant *et al.*, 2022).

3.1.1 The technical and biological pitfalls of the assembly of *H. ensiformis* mtDNA

The mitochondrial genome of *Haplopteris ensiformis* can be mapped to at least nine circular sub-genomic chromosomes with a total size of 1.4 Mbp. The chromosomes consist of single-copy- and repetitive sequence and there is clear evidence for recombinogenic activity between the chromosomes. The elaborated configuration of these nine chromosomes (Zumkeller, Polsakiewicz, and Knoop 2022, Figure 4) likely displays only one variant of the mitochondrial genome of the other possible arrangements which likely exist in *H. ensiformis*. Apart from its large size, it is composed of sequences of different genetic origins, which additionally complicated the assembly of the mtDNA of *H. ensiformis*.

The assembly of the mitochondrial genome of *H. ensiformis* had to mainly address problems with respect to discriminating between authentic chloroplast contigs and the numerous promiscuous chloroplast DNA inserts in the mitogenome. The verification of the latter necessitated carefully optimized template-switch sensitive PCR strategies to corroborate the mtDNA assemblies (Zumkeller, Polsakiewicz, and Knoop 2022, Figure 5). Different assembly strategies of the paired-end 250 bp Illumina DNA and RNA/cDNA reads alone, always resulted in several open end contigs and repeat configurations could not be discriminated. Alternatively arranged mtDNA isoforms could be identified via

optimized PCR strategies indeed, but experimentation can be time consuming and include typical flaws of the PCR strategy like template switch over repeat sequences, if not performed with caution. However, full capture of vascular plants complex mtDNA is necessary to avoid missing sequence data. To determine recombinational activity of the enormous multipartite 11 Mbp mtDNA of *Silene noctiflora,* for example*,* paired-end reads from 454 sequencing have been used (Sloan *et al.*, 2012). In general, 454 sequencing produces longer reads (>500 bp) than Illumina HiSeq (~250 bp). In a recent study dedicated to mtDNA recombinational activity of lettuce, PacBio reads were effectively used to capture a complete picture of the mtDNA isoforms produced by recombination (Kozik *et al.*, 2019). It should be noted that upcoming nanopore sequencing technologies, like the Oxford Minion nanopore are capable to determine structural variants, mainly because it allows continuous sequencing of DNA molecules larger than 100 kbp (Xie *et al.*, 2021). PacBio RSII and/or nanopore sequencing technologies, which produce long contiguous readouts around 13- 20 kbp, are recommended for future sequencing of complex mitochondrial genomes like those of *Haplopteris* and other leptosporangiate ferns (Rhoads and Au, 2015; Lu, Giordano and Ning, 2016). Parallel Illumina sequencing will likely still be necessary to guarantee for high qualities of primary sequences by stochastically ruling out sequencing errors. An overall comparison of the tssPCR to bioinformatical mapping experiments across alternatively arranged mtDNA isoforms derived from different strategies sequencing technologies like Oxford Minion nanopore or 454 sequencing should be helpful to approximate each performance.

In accordance with the outcome of the tssPCR that avoids template switches *H. ensiformis* contigs have been concatenated across repeats and open ends. Visualisation of potential connections between contigs can by utilized with bioinformatic tools like the Semi-Automated Graph-Based Assembly Curator (SAGBAC), which was specifically designed for the complex mtDNA of evening primroses after experimentally mitochondrial isolation to reduce cpDNA contamination (Fischer *et al.*, 2022). To a certain extent, the interactive *de novo* assembly visualisation tool BANDAGE is, in terms of the bioinformatic procedure, analogous to this (Wick *et al.*, 2015). By now, there are also DNA assembly programs that have been more and more adapted for complex mtDNAs of land plants. These include strategies to e.g. distinguish between reads of chloroplast,

mitochondrial or nuclear origin artificially, by coverage and subsequent assembly (Dierckxsens, Mardulyn and Smits, 2017).

Above all, is the necessity of profound knowledge about the peculiar novelties of the *Haplopteris ensiformis* mtDNA that eases upcoming mitogenome assemblies. Notable discoveries in the mtDNA of *H. ensiformis,* like first time detection of a trans-spliced *rrnL* gene or prevalence of plenty pre-emptive stop-codons that must be removed via U-to-C RNA editing within small genes like *rps11* are crucial information for future analyses for complete genome projects among polypod ferns. The mtDNA of *H. ensiformis* contains all expected subunits of the respiratory chain complexes I-V, ribosomal proteins and a rich set of tRNAs (Zumkeller, Polsakiewicz, and Knoop 2022, Table 1). Further, prominent C to U and U to C RNA editing reconstitutes functionality of transcribed genes. In total, 1,618 RNA editing events were detected in the mitochondrial transcriptome of *H. ensiformis* (Zumkeller, Polsakiewicz, and Knoop 2022, Suppl. Table 2), of which 233 events account for the removal of pre-emptive stop-codons. In general, high rates of both types of RNA editing can be expected among leptosporangiate ferns (Knie *et al.*, 2016). Consequently, deep sequencing of RNA and dedicated downstream transcriptome analyses as performed in this study are recommended for future studies of the organellar DNA of ferns. To determine RNA editing sites, a DNA/RNA read mapping pipeline for SNP-Calling has been modified using the program JACUSA (Piechotta *et al.*, 2017), which has already been established for the upcoming model organisms in the field of RNA editing like *Anthoceros agrestis* (Gerke *et al.*, 2020). JACUSA is known to identify RNA editing sites with high efficiency but comes with the problem of mis-calling certain sites, as well. Ultimately, this calls for biological and technical replicates of RNAseq data to improve such weaknesses, just as it has been applied in recent studies that introduced RNA editing outside of the plant kingdom (Oldenkott *et al.*, 2020; Lesch *et al.*, 2022).

In combination with JACUSA, Illumina read mapping can successfully determine partial RNA editing, ranging from 1%-100%, in both chloroplast and mitochondrial transcriptomes of *H. ensiformis* (Zumkeller, Polsakiewicz and Knoop, 2022). More than half of all partial RNA editing sites in *H. ensiformis* are silent, where editing does not affect the amino acid codon identity. This does include first position RNA editing of Leucine codons CUN and UUR and codons with pyrimidine bases in the third codon

position. Intriguingly however, RNA editing rates below 50% could be identified for the crucial introduction of start codons, e.g. rps12eU2TM (29%) or rps2eUTM (34%) or removal of pre-emptive stop codons, e.g. atp8eC34*Q (19%) or cox1eC1048*C (24%) (Zumkeller, Polsakiewicz, and Knoop 2022, Suppl. Table 1 and 2). These results are alarming when automatized gene annotation after *de novo* transcriptome assembly is planned for fern organelles, using standardized tools like Velvet, SPAdes or Trinity (Zerbino and Birney, 2008; Grabherr *et al.*, 2011; Bankevich *et al.*, 2012). Partially edited positions might be ruled out for contig assembly as they appear as sequencing errors for the de Bruijn graph (Zerbino and Birney, 2008). Consequently, it is important to extend post-assembly analyses on the differentiation between errors and polymorphisms, generated by partial RNA editing. Else, organellar genes might be overlooked or confused with pseudogenes.

Nevertheless, it is highly important to perform careful comparison with related species if genome data is available. Only distantly related completely sequenced mtDNA counterparts are available, e.g. the eusporangiate ferns *Psilotum nudum* and *Ophioglossum californicum* for a comparative analysis (Guo *et al.*, 2017). Intriguingly, the mitochondrial genome of *H. ensiformis* features a plethora of new exceptional characteristics compared with the early-branching representatives. As it could be expected from previous analyses, the cpDNA of *H. ensiformis* resembles the one of other polypod ferns, more specifically of the Pteridaceae and in particular, other Vittariaceae with high sequence similarities, equal gene complements and similar gene synteny (Wolf, Roper and Duffy, 2010; Robison *et al.*, 2018).

3.1.2 Repeats and recombinational activity - The polypodiid fern *Haplopteris ensiformis* has a highly complex mtDNA like other vascular plants

With the mitochondrial genome of *H. ensiformis,* a large and complex multi-chromosomal genome has now been identified in the fern lineage. Hence, multichromosomal mtDNAs have now been discovered in eusporangiate ferns with the bipartite mtDNA of *Psilotum* (Guo *et al.*, 2017) and leptosporangiate fern. Examples of large multichromosomal, recombinogenic active genomes among spermatophytes can be found in the gymnosperm *Picea sitchensis* (~5.5 Mbp), *P. abies* (~4.9 Mbp) or among flowering plants with *Cucurbita pepo* (~0.9 Mbp), the genus *Silene* (~7 Mbp) or *Amborella*

trichopoda (~3.8 Mbp) (Alverson, Rice, *et al.*, 2011; Sloan *et al.*, 2012; Rice *et al.*, 2013; Sullivan *et al.*, 2019; Jackman *et al.*, 2020). Already within the earliest branching vascular lineage, the lycophytes *Selaginella moellendorffii* and *Isoetes engelmannii* exhibit two highly complex examples, where high content of repeats and, consequentially, recombination produce numerous possible genome arrangements (Grewe *et al.*, 2009; Hecht, Grewe and Knoop, 2011).

Contrary to the diverse complexity into which several tracheophyte mtDNAs have evolved, all mitochondrial genomes of bryophytes (i.e. liverworts, mosses and hornworts) have relatively conserved sizes (~100-200 Kbp), conserved gene syntenies and less complex circular organisation (Knoop, 2012; Mower, Sloan and Alverson, 2012). Bryophyte mitochondrial genomes are almost completely devoid of repeated sequences (Wynn and Christensen, 2019). However, the low activity of recombination detected in liverworts involves repeats of sizes from 50-250 bp were indicating that recombinational activity was already ancestrally present in all mitochondrial genomes of land plants (Dong *et al.*, 2019). Rather exceptional evolutionary old mtDNA rearrangements within the bryophyte lineages could be linked to such activity. In marchantiid liverworts for example, recombination in the mtDNA changed gene synteny and is involved in pseudogenisation of the genes *cob* for cytochrome b and *nad7* NADH dehydrogenase subunit 7 (Groth-Malonek *et al.*, 2007; Wahrmund, Groth-Malonek and Knoop, 2008).

Fern mtDNA, together with other examples from vascular plants now stand in clear contrast to that. The repeat content of the *H. ensiformis* mitochondrial genome lies in between those of *O. californicum* and *P. nudum.* All three completely sequenced fern mitochondrial genomes share neglectable gene syntenies in comparison to bryophytes, the lycophyte *Phlegmariurus squarrosus* or amongst each other, indicating heavy rearrangements of the mtDNA during fern evolution. Notably, more than the half of the mitochondrial genome of the eusporangiate fern *Psilotum nudum* consists of repeats (Guo *et al.*, 2017). However, the mitochondrial genome structure of *P. nudum* is present as two autonomous circular chromosomes and mapping of Illumina reads against putative alternative arrangements indicates nearly to none (>3%) recombinational activity.

The mitochondrial genome of the eusporangiate fern *O. californicum* with a repeat content of ~40% is similar to other repeat-rich flowering plants like *Silene conica* or *Cucurbita pepo* (Alverson *et al.*, 2010; Sloan *et al.*, 2012). *Silene conica* or *Cucurbita pepo* are good examples for recombinative active mitogenomes among flowering plants, but similar to *P. nudum*, the mtDNA of *O. californicum* shows no or very low indication for recombinational activity (Guo *et al.*, 2017).

The flowering plant mitochondrial genomes of the Geraniaceae family show particularly well how repeat content and recombinational activity evolve in reciprocal dependence. In the genus *Monsonia* mitogenomic rearrangement rates were analysed on a relatively low phylogenetical and evolutionary time scale (Cole *et al.*, 2018). Here the "hyperactive" mitochondrial genome of *Monsonia ciliata* showed 10 times higher rearrangement rates than its sister species, despite having the lowest number of repeats among the genus (Cole *et al.*, 2018). Consequently, content of repeat sequence alone cannot be simply correlated with activity of DNA recombination and explain its complexity. Accordingly, the repetitive composition of *H. ensiformis* is a predisposition for recombinative activity and loss of gene syntenies are strong indication for ancient recombinative activity.

3.1.3 *H. ensiformis rrnL* requires trans-splicing

Tracheophyte mitochondrial genomes generally show only low numbers of group I introns (Mower, 2020). The mitochondrial genome of *H. ensiformis* contains four introns of this type: cox1i395g1, rrnLi825g1, rrnLi1897g1 and i1928g1 (Zumkeller, Polsakiewicz and Knoop, 2022). There are only two cases documented of *trans-spliced* group I introns in plant mitochondrial genomes, the here reported *H. ensiformis* rrnli825g1 and cox1i395g1 in *Isoetes engelmannii* (Grewe *et al.*, 2009). Intriguingly, we now identified a *cis-spliced* variant of cox1i395g1 in *H. ensiformis,* indicating a complex evolutionary history with multiple gain/loss events, considering its occurrence in liverworts, the lycophyte *Isoetes*, Equisetales and polypod ferns (Bégu and Araya, 2009; Farré *et al.*, 2012). Notably, a *cis*-arranged orthologue, albeit with very low overall sequence similarity, could also be identified for rrnLi819g1 in the streptophyte alga *Chara vulgaris* (Turmel, Otis and Lemieux, 2003). In the case of the mitochondrial flowering plant intron cox1i726g1, its patchy distribution and sequence similarity with fungal group I intron

paralogues indicate the involvement of HGT (Seif *et al.*, 2005; Sanchez-Puerta *et al.*, 2008, 2011).

Increased recombinogenic activity is hypothesized as reason for the frequent occurrence of *tans-spliced* introns in land plant mitogenome evolution (Grewe *et al.*, 2009; Hecht, Grewe and Knoop, 2011; Guo *et al.*, 2020). With rrnLi819g1 in *H. ensiformis* we discovered a first *trans*-spliced intron in the fern lineage (Zumkeller, Polsakiewicz, and Knoop 2022, Figure 7). Orthologous *cis*-arranged group II intron orthologues had been identified as ancestors of *trans-spliced* introns in flowering plant mitogenomes in ferns and bryophytes (Malek and Knoop, 1998; Volkmar *et al.*, 2012). Further evidence for such *cis-to-trans* shifts in intron splicing have later been documented for gymnosperms (Guo *et al.*, 2020). Likely rrnLi819g1 has been gained in the fern lineage on *cis*-spliced variant and shifted to *trans*-splicing after intra-genic recombination, secondarily.

Just like group II introns, group I introns in mtDNA of land plants appear to follow a complex evolutionary history involving frequent loss-and-gain dynamics, lateral transfers, and variability in post-transcriptional processing regarding *trans*-splicing or RNA editing. The concept of group II intron "families" (Zumkeller and Knoop, 2022) tracking lateral intron transfers as presented here should eventually allow for an extended understanding on the evolution of those retro-mobile elements. A similar approach for mitochondrial group I introns of streptophyte plants may likewise prove valuable despite their overall paucity in vascular plant mitogenomes.

Nonetheless is the discovery of a *trans-spliced* group I intron noteworthy. The autocatalytic splicing capabilities of group I introns are being used in a biotechnological application to produce large circular exon RNAs *in vitro* and *in vivo* (Petkovic and Müller, 2015). Circular RNAs display promising therapeutic agents, mainly due to their higher stability compared to linear isoforms (Holdt, Kohlmaier and Teupser, 2018). Common strategies are permuted introns and exons (PIE strategy) or the so called RNA cyclase ribozyme (Puttaraju and Been, 1992; Ford and Ares, 1994). The RNA cyclase ribozyme is derived from a T4 phage group I intron, while the PIE strategy is derived from an autocatalytic group I intron of a tRNA gene of the protist *Anabaena.* Both strategies rely on complementary binding of terminal intron ends, analogous to spliceosomal back-splicing or group I intron *trans*-splicing (Yu and Kuo, 2019). However, the appliance of the

autocatalytic ribozymes comes with unwanted RNA-hydrolysis in heterologous systems (Li and Breaker, 1999).

The large ribosomal RNA (23S) of plant mitochondria is one of the most prominent RNAs of the cell. Accordingly, *trans*-splicing of rrnLi819g1 must be highly efficient. Further investigation of *Haplopteris* rrnLi819g1s splicing capabilities *in vivo*, regarding assisting protein factors and/or *in vitro* ribozyme activity, can certainly provide complementary information on the optimization of circular RNA production using ribozymes.

3.2 Massive integration of xenologous DNA into the *Haplopteris* mtDNA

There are several interesting cases already discovered for horizontal gene transfer (HGT) into the nuclear genome of ferns, see (chapter 1.5). This is explained, for amongst others, by the fact that ferns have “naked” sporangia and, hence, are more vulnerable for infections that affect the germline (rev. in Wickell and Li 2020). Horizontally acquired DNA must be introduced into the germline, where evolutionary fixation could be exerted by positive selection over generations, for example by a gain of function. Else, DNA insertions into the mtDNA will degrade and eventually be lost. A good starting point in identifying a gain of function is when a “novel” gene is conserved amongst a large group of organisms.

3.2.1 Xenologous tRNA genes in fern mitogenomes

The mitochondrial genome of *H. ensiformis* contains well-conserved tRNAs of chlamydial origin (Zumkeller, Polsakiewicz and Knoop, 2022). Chlamydial-like *trnN-GUU* has also been documented in other tracheophytes: polypodiid ferns and *Angiopteris*, the lycophyte *Phlegmariurus squarrosus* and gymnosperm *Cycas* (Knie, Polsakiewicz and Knoop, 2015). The other chlamydial-like *trnR-UCG*, is also found in eusporangiate ferns *O. californicum* and *P. nudum* (Guo *et al.*, 2017). In both cases, the xenologous *trnN-GUU* and *trnR-UCG* genes were likely gained once, in the ancestor of tracheophytes and the common ancestor of the Ophioglossales-Psilotales clade and leptosporangiate ferns, respectively (Knie, Polsakiewicz and Knoop, 2015; Guo *et al.*, 2017). A conservation of both tRNA genes within the mtDNA over the evolution of ferns strongly supports a functional adaption, which however, could not be shown yet, as transcription of tRNAs could not investigated for *H. ensiformis*, due to Illumina read sizes. Nonetheless, transcription and processing of a likely horizontally acquired *trnC-GCA* into the mtDNA of

flowering plants could be demonstrated in *Beta vulgaris, Spinacea oleracea* and *Cucumis sativus* (Kitazaki *et al.*, 2011).

Another, yet more complex example is the incomplete set of transfer RNAs (tRNAs), insufficient for autonomous translation, encoded by land-plant mitochondrial genomes (rev. in A. Schneider 2011). This is complemented by long-known active import of cytosolic tRNAs into the mitochondria of flowering-plants (Maréchal-Drouard, Weil and Guillemaut, 1988; Small *et al.*, 1992; Kumar *et al.*, 1996; Maréchal-Drouard *et al.*, 1999). However, the mechanisms of tRNA import are variable between species and not completely understood (Salinas-Giegé, Giegé and Giegé, 2015). Interestingly, chloroplast-like tRNAs have been identified in the mitochondrial genome of flowering plants *Arabidopsis thaliana*, *Triticum* (wheat) or *Oryza* (rice) that are transcribed (Joyce and Gray, 1989; Miyata, Nakazono and Hirai, 1998; Duchêne and Maréchal-Drouard, 2001). On top, likely those chloroplast-derived tRNAs even replaced native mitochondrial counterparts along land-plant evolution (Marechal-Drouard *et al.*, 1990; Fey *et al.*, 1997; Miyata, Nakazono and Hirai, 1998; Duchêne and Maréchal-Drouard, 2001). The functional uptake of xenologous tRNAs appears more plausible as these genes are very small, transcription in plant mitochondria is organized polycistronic and that tRNAs processing works rather universal (Warren and Sloan, 2020).

Apart from the tRNA genes described above, larger DNA insertions of xenologous origin could be identified in the *Haplopteris* mitogenome, which are not transcribed and apparently not conserved in the closely related species *Vittaria lineata* (Zumkeller, Polsakiewicz, and Knoop 2022, Figure 9). In addition, no evidence for transcription was found. Hence, these xenologous DNA inserts were likely acquired recently on an evolutionary timescale and have no function.

3.2.2 Rickettsial DNA insertions in *H. ensiformis* mtDNA as a document of symbiosis and/or pathogeny

Xenologous DNA insertions in *H. ensiformis* mtDNA mainly seem to originate from a-proteobacteria closely related to obligate intracellular endosymbiontal parasites like e.g. Cand. *Caedibacter acanthamoeba* (Suzuki *et al.*, 2015). Endosymbionts like *Caedibacter caryophilus, C. biaurelia* or related C. *acanthamoeba* were shown to transmit genetic material to their host's nucleus, which is *Paramecium* or the protist

Acanthomoeba, respectively (Horn *et al.*, 1999; Grosser *et al.*, 2018; Schrallhammer, Castelli and Petroni, 2018). Related obligate intracellular parasites like *Rickettsia massaliae* or *R. bellii* have been demonstrated to actively transfer DNA amongst each other via conjugation, to outcompete intracellular competitors (Blanc *et al.*, 2007). The xenologous bacterial DNA regions identified in the *H. ensiformis* mitogenome include notable similarities to IS481 family transposase and/or virB8 and virB9 homologs (Zumkeller, Polsakiewicz, and Knoop 2022, Figure 9). The latter genes are commonly found among mobile IS elements and are associated with conjugative gene transfer (Gelvin and Habeck, 1990; Siguier, Gourbeyre and Chandler, 2014). Possibly, the mitochondrial genome of *H. ensiformis* was victim to the genetic warfare between intracellular pathogens.

Alternatively, the high recombinational activity of the mitogenome led to the incorporation of xenologous DNA after successful defence of an infection, similar to the acquisition of promiscuous cpDNA after DNA leakage from the chloroplast under stress in *Tobacco* (Hastings *et al.*, 2009; P. Liu *et al.*, 2011; Wang, Lloyd and Timmis, 2012). In general, acquisition of xenologous DNA into the host genome is assumed to depend on NHEJ recombination during DNA repair (Lieber, 2010; Husnik and McCutcheon, 2017). In this regard, our findings of xenologous DNA insertions in the mitochondrial genome of *H. ensiformis* fit with the overall elevated complexity of the mitochondrial genome, due to its recombinative activity.

The identification of a donor for the xenologous DNA inserts among the Rickettsiales closely related to endosymbiotic or endoparasitic organisms makes the findings of DNA insertions in the mtDNA of *H. ensiformis* appear more plausible. Rickettsia-like DNA insertion might indicate a putative close interaction between donor and recipient. Studies of the a-proteobacterial rickettsia *Wolbachia* show that intracellular parasites can effectively survive within taxonomically distantly related hosts. For example, endoparasitic *Wolbachia* are transmitted among angiosperms like *Gossypium* and herbivory arthropods like whiteflies (Bing *et al.*, 2014; Li *et al.*, 2017) or the oak tree *Quercus* and gall wasps, respectively (Schuler *et al.*, 2018). Nonetheless, bacterial-to-mitochondrial gene transfer in plants is an exceptional finding and has only been documented for *trnS-GCU, trnL-GAG, trnS-GGA* and *trnN-GUU* in some vascular

plants (Knie, Polsakiewicz and Knoop, 2015). In contrast, horizontal DNA transfer from land plant organelle-to-mitochondria has been recognized in flowering plants much more often, e.g. from epiphytes to *Amborella* or the plant *parasitic Cuscuta* (Bergthorsson *et al.*, 2004; Richardson and Palmer, 2007; Goremykin *et al.*, 2009; Mower *et al.*, 2010; Skippington *et al.*, 2015).

The observation of HGT form bacteria to fern mitochondria in this quantity is highly surprising and allows new speculations on the evolution of ferns and their mitochondria. Bacterial endosymbionts or endoparasites, like e.g., *Wolbachia* or Rickettsia, have not been described in ferns yet, but findings of the xenologous DNA insertions in the mtDNA of *H. ensiformis* probably indicate such symbiotic or pathogenic relations. Close symbiotic relation between bacteria and ferns or, in general, land plants, are long known and have been investigated on molecular level in case of the water fern *Azolla* or hornwort *Anthoceros* with cyanobacterial genus *Nostoc* (F. W. Li *et al.*, 2018; Li *et al.*, 2020) or the legume *Medicago* and bacterial *Sinorhizobium* (Toro, Martínez-Abarca and Fernández-López, 2016). Next to the observation of bacteria-to-mitochondria HGT, the mtDNA of *H. ensiformis* is likely documenting a bacterial-*Haplopteris* parasitic/ symbiotic relationship. Ultimately, these findings must be investigated further, as *Agrobacterium tumefaciens* gives an outstanding example for the potential application of domesticated prokaryotic plant parasites as mutagenic tools (Sardesai and Subramanyam, 2018).

3.3 Group II introns in plant mitochondria come in families, documenting ancient retro-mobile activity

In comparison to 318 different group II introns detected in prokaryotes, the mitochondrial genomes of streptophytes contain the second largest set of diverse group II introns counting currently 161 (Candales *et al.*, 2012; Zumkeller and Knoop, 2022). With our systematic taxonomical screening covering mitogenomes of all major streptophyte lineages presently available, we provide an overall nearly to complete record to date. Intriguingly, the 161 mitochondrial group II introns stand in sharp contrast to the 34 chloroplast group II introns that are known of along streptophyte evolution (Lemieux, Otis and Turmel, 2016).

3.3.1 Dynamics of group II intron evolution in land plant mitochondria: Gains are and losses are frequent

The major lineages of streptophytes have characteristic, unique sets of mitochondrial group II introns. Intersections between the major clades range from zero to a maximum of eight group II introns shared between hornworts, mosses and vascular plants with 40, 24 and 47 introns, respectively (Zumkeller and Knoop 2022, Figure 1). Only one mitochondrial intron, atp9i87g2, is shared among liverworts, mosses, hornworts and tracheophytes (Knoop, 2010, 2013; Mower, 2020). Changes of group II intron content within the major lineages of land plants only occurs rarely. In particular, there are only five cases documented yet, where phylogenetic distribution clearly indicates late gain of a group II intron within only one major land-plant lineage: (1) nad1i258g2 which is exclusively found in ferns (Dombrovska and Qiu, 2004), (2) atp1i361g2 in Marattiales and polypod ferns (Wikström and Pryer, 2005; Zumkeller, Knoop and Knie, 2016), (3) rps1i25g2 in all extant ferns except basal Equisetales (Knie, Grewe and Knoop, 2016), (4) cox2i98g2 in the hornwort genus *Anthoceros* (Dong *et al.*, 2018; Gerke *et al.*, 2020) and (5) rrnSi1065g2 in marchantiid liverworts and the liverwort *Haplomitrium* (Oda *et al.*, 1992; Dong *et al.*, 2019).

Loss events of mitochondrial group II introns occur much more frequently in comparison to gain events. For example, introns cox2i373g2, cox2i691g2 and nad4i976g2 are independently lost among spermatophytes multiple times, e.g.in *Arabidopsis thaliana, Beta vulgaris*, *Magnolia tripelata,* the genus *Silene,* the monocotyledonous order of Alismatales and Cupressophytes (Unseld *et al.*, 1997; Kubo *et al.*, 2000; Itchoda *et al.*, 2002; Kudla *et al.*, 2002; Sloan *et al.*, 2010; Hepburn, Schmidt and Mower, 2012; Guo *et al.*, 2020). Among ferns alone intron atp1i361g2 is independently lost four times in the family of Pteridaceae (Zumkeller, Knoop and Knie, 2016). The gymnosperm *Welwitschia mirabilis,* the hemi-parasitic flowering plant *Viscum album,* as well as for example *Geranium bryceri* underwent moderate to extreme mitochondrial intron losses, losing about half of their intron content (Parkinson *et al.*, 2005; Park *et al.*, 2015; Guo *et al.*, 2016). Along land plant backbone phylogeny and terminal major land plant lineages 83 to 91 loss events could be determined, depending on the NLE or mB phylogeny, respectively (Zumkeller and Knoop 2022, Figure 10).

3.3.2 The enigmatic distribution of group II introns: Were there ancestrally intron-rich mitochondria or did mobile group II introns break species boundaries via HGT?

The mitochondrial group II intron distribution in land plants may be best explained by an initial phase rich in intron gains at the emergence of the major lineages, followed by a stagnation and clade-specific loss phase (Qiu *et al.*, 1998, 2006; Pruchner *et al.*, 2002; Groth-Malonek *et al.*, 2005; Volkmar and Knoop, 2010; Volkmar *et al.*, 2012). This hypothesis implies that mitochondrial group II intron evolution has been largely inactive for the last 400-500 MYA.

There are only a few isolated examples for mitochondrial group II introns in the early-branching metazoan lineages with placozoans like *Trichoplax adhaerens* with cox1i652g2 and sponges like *Axinella verrucosa* with cox1i972g2 and cox1i1147g2 (referred to as intron 699 and 1141) (Dellaporta *et al.*, 2006; Signorovitch, Buss and Dellaporta, 2007; Huchon *et al.*, 2015). In contrast, similarly to land plant mitochondrial genomes, fungi, diatoms and rhodophytes are known for diverse distribution of group II intron within their organelles, too (Paquin *et al.*, 1997; Ehara, Watanabe and Ohama, 2000; Nielsen *et al.*, 2004; Seif *et al.*, 2005; Brigham *et al.*, 2018; Funk *et al.*, 2018; Preuss *et al.*, 2021). Here, comparative genomic studies indicate multiple gain/loss scenarios for group II introns, without clear indication for active group II intron mobility within the respective lineages (Guillory *et al.*, 2018; Pogoda, Keepers, Nadiadi, *et al.*, 2019; Preuss *et al.*, 2021). Like the assumptions for the mitochondrial group II intron distribution in land plants, it is assumed that the ancestral organellar genomes of fungi, rhodophytes or stramenopiles were intron-rich after extensive group II intron radiation and subsequently lost orthologous introns (Paquin *et al.*, 1997; Ehara, Watanabe and Ohama, 2000; Nielsen *et al.*, 2004; Seif *et al.*, 2005; Brigham *et al.*, 2018; Funk *et al.*, 2018; Preuss *et al.*, 2021).

Among the identified group II introns mentioned above, most group II introns surprisingly occupy the respective *cox1* genes, followed by the *rrnL* genes (Guillory *et al.*, 2018). Interestingly, group II introns here even share the same insertion sites over wide phylogenetic distances, with for example cox1i972g2 in sponges, rhodophytes and stramenopiles (Ehara, Watanabe and Ohama, 2000; Nielsen *et al.*, 2004; Huchon *et al.*, 2015). In addition to that, we identified *Coleochaete scutata* mitochondrial cox1i1147g2 as solitary-type intron without paralogous in streptophyte mitochondria and orf762 as

mat-cox1i1147g2 (Zumkeller and Knoop, 2022, Suppl. Figure 2). We found that homologs based on similarity of the intron encoded maturases with identical insertion sites also in the chlorophyceaen *Microspora stagnorum* (54% Aa identity), the brown algae *Pylaeiella littoralis* (39% Aa identity), the ascomyceaen fungi *Candida sojae* (41 % Aa identity), together with, the sponge *Axinella verrucosa* (39% Aa identity) (Zumkeller and Knoop, 2022, Suppl. Figure 2). Thus, fungi, sponges, placozoans and brown, chlorophyceaen and coleochaetophyceaen algae share a likely orthologous mitochondrial group II intron with cox1i1147g2 (Huchon *et al.*, 2015; Zumkeller and Knoop, 2022). The last common ancestors of those taxa would be the ancestor of all eukaryotes itself, making cox1i1147g2 and *mat-cox1i1147g2* a putative "core-member" of the mtDNA and its evolution.

Group II intron distributions can be explained by independent gain events due to horizontal gene transfer (HGT) across different species, alternatively (Kamikawa *et al.*, 2009; Guillory *et al.*, 2018). Especially among prokaryotes horizontal group II intron transfer is prominent, as retro-mobile group II introns can be actively transmitted from one organism to another by conjugation amongst bacteria at least. Well studied *Lactococcus lactis* LtrA group II intron resides within a relaxase gene *ltrB* on a conjugative plasmid pRS01 (Mills, Mckay and Dunny, 1996). The pRS01 plasmid and namely the relaxes gene usually promote conjugal HGT (Novikova *et al.*, 2014). When expressed, LtrA is effectively spliced and inhibits the relaxase host gene expression by intron RNA – mRNA interaction, actively reducing mRNA levels (Qu *et al.*, 2018). Consequently, accuracy and efficiency of splicing is deciding over conjugation in the host and the recipients genome (Klein *et al.*, 2004; Belhocine *et al.*, 2007). In addition, LtrA group II intron not only can down-regulate its host gene, but also invade itself and ectopic targets via reverse splicing on RNA level, increasing transcript diversity after transcription (LaRoche-Johnston *et al.*, 2018). Strikingly, intergenic *trans*-splicing among LtrA invaded genes can then create functional protein chimeras, *in vivo* (LaRoche-Johnston, Bosan and Cousineau, 2021).

The occurrence of group II introns in Archaea is rare, phylogenetically patchy and there is some evidence for their acquisition by HGT (Rest and Mindell, 2003). Archaean group II introns match the requisitions to be retro-mobile elements: Archaean group II introns encode full conserved maturases with RT-X and D/En domain, Archaea have

twintrons, introns-within-intron that indicate recent intra-genomic mobility, accompanied by significant sequence similarity to eu-bacterial counterparts (Dai and Zimmerly, 2002; Rest and Mindell, 2003; Simon *et al.*, 2008). There are also examples of putative group II intron HGT to organelles, based on sequence similarity of intron-encoded maturases are proposed for cyanobacteria-to-mitochondria in the red alga *Porphyra purpurea* (70% nucleotide identity) (Burger *et al.*, 1999), cyanobacteria-to-chloroplast in *Euglena myxocylindracea* (43% nucleotide identity) (Sheveleva and Hallick, 2004) or from diatoms to the mitochondria of eukaryotic freshwater algae *Chattonella* (93% identity) (Kamikawa *et al.*, 2009). Further, the isolated occurrence of a mitochondrial group II intron cox1i721g2 of the bilaterian marine catworm *Nephtys sp.* is considered to results by HGT from an yet unknown source with around only 25% sequence similarity of its IEP to bacterial *Proteus mirabilis* (Vallès, Halanych and Boore, 2008). However, it is not clear how these group II introns have been transmitted into organelles.

The distribution of cox1i1147g2 could also easily be explained by HGT as well, but this would neglect that no common organismal vector or clear mechanism for transmission into the organelles of coleochaetophyceaen algae, sponges and fungi is known. Our analyses of mitochondrial group II introns led to the discovery of mitochondrial group II intron families, which could explain an enigmatic distribution of group II introns by a combination of intra-genomic lateral intron transfers and frequent losses over 500 Mya of evolution in streptophytes (Zumkeller and Knoop, 2022). This *modus operandi* exemplarily shown here for streptophyte mitochondrial group II intron might be applied to fungi, rhodophytes and diatoms as well.

3.3.3 Solitary type group II introns are evolutionary ancient lone survivors of families or taxonomically isolated

For 57 solitary-type (S-type) streptophyte mitochondrial group II introns no paralogues were identified. Phylogenetic analyses of mitochondrial S-type group II introns maturases further indicate that 14 S-type share common origin and have been captured in one of the 27 Superfamilies (Zumkeller and Knoop, 2022, Figure 5). With respect to their distribution and conservation, a notable difference between S- and F-type group II introns was observed.

To the exception of family F21 group II introns, exclusively present in *Coleochaete scutata,* every other streptophyte mitochondrial group II intron family has paralogues in more than one lineage. Consequently, group II intron families are evolutionary old. In stark contrast, only 11 “truly” S-type group II introns (cox2i104g2, cox2i691g2, nad1i394g2, nad2i542g2, nad2i709g2, nad3i52g2, nad4Li283g2, nad5i1872g2, nad7i140g2, nad7i917g2 and trnN-GUUi38g2) were found in more than one of the major streptophyte lineages of streptophyte algae, liverworts, mosses, hornworts, lycophytes, ferns, gymnosperms and spermatophytes. The remaining 41 S-type group II introns occur rather taxonomically isolated in early branching streptophytes algae (12), Charophyceae (3), Coleochaetophyceae (10), Zygnematophyceae (5), Liverworts (1), Mosses (3), Hornworts (5) and Lycophytes (2). There are few cases that have fossilized paralogous in case of atp1i1050g2, cobi692g2, cox1i732g2, cox1i995g2, nad2i709g2 documenting some degree of intron mobility (Zumkeller and Knoop, 2022, Suppl. Figure 1).

S-type group II intron clade specific “signature” introns are prevalent. There are for example atp1i805g2, atp1i1019g2, cobi838g2, cox1i1298g2 and cox2i281g2 in hornworts only (Zumkeller and Knoop, 2022, Figure 5). This indicate that these introns were gained late from an unknown origin within their lineage. However, putative donors outside the streptophyte lineage could not be identified, to the exception of cox1i1147g2. Also, cox1i1147g2 is the only S-type with a functionally well conserved maturase that is not closely related to any other streptophyte paralogue. Consequently, most S-type group II introns do not show any capabilities for recent mobility involving an associated maturase promoting invasiveness.

Isolated S-type group II introns, like previously overlooked cox3i34g2 and nad2i89g2 are only present in Zygnematophyceaen *Gonatozygon brebissonii* and *Zygnema circumcarinatum* and have just been identified Zygnematophyceaen algae are morphologically, taxonomically and likely also genomically extremely diverse (>4,000 spp.) (Guiry, 2010; Zhou and von Schwartzenberg, 2020). Therefore, it is necessary to further sample for high quality genomes with high attention gene annotation as potential paralogues might yet to be discovered.

Another reason for the occurrence of S-type group II introns likely lies within the classification of group II intron families based on sequence similarity estimation via

BLASTN. Identification of putative paralogues is hampered by sequence divergence over long evolutionary distances or lineage-specific elevated mutational rates. Most notably, mitochondrial group II introns of the lycophytes *I. engelmannii* or *S. moellendorffii* show unusual reduction in sizes of group II introns or highly specified substitutions accompanied by extreme rates of RNA editing, respectively, that significantly reduce sequence similarities among orthologs (Grewe *et al.*, 2009; Hecht, Grewe and Knoop, 2011). For example, the mitochondrial genome of both species encodes S-type introns cox1i227g2 and cox1i266g2 that were likely gained in the common ancestor of *Isoetes* and *Selaginella.* However, both introns do not share significant similarity to their orthologues.

Ultimately, remaining S-type group II introns simply might not have been copied or represent the “last survivors” of ancient group II intron families, where shared sequence similarity to any paralogues have been lost. The identification of S-type group II introns does not contradict the concept that mitochondrial. Group II intron evolution in streptophytes is mainly driven by intra-genomic lateral intron transfer and frequent loss events, consequently.

3.3.4 Land plant mitochondrial group II introns mobility – head-to-head with prokaryotes

For the 161 mitochondrial group II introns of streptophytes no significant sequence similarities were found in prokaryotes, fungi, diatoms or any other archaeplastida on nucleotide level (Zumkeller & Knoop 2022). Alignment and phylogenetic analyses of intron encoded maturases, however, allow to understand group II intron evolution across the different kingdoms of life (Toor, Hausner and Zimmerly, 2001). Examples like *mat-cox1i1147g2* in *Coleochaete scutata* might be a promising addition to address the question around the origin of group II introns in eukaryotes. In contrast, most mitochondrial group II introns of land plants do not contain an intron encoded maturase, nor do their hitherto identified nuclear-encoded splicing factors encode an highly conserved reverse transcriptase and/or DNA binding domain with endonuclease properties (D/En) (Guo and Mower, 2013). So far, evidence for retro-mobile activity of mitochondrial group II introns in land plants is rare and relies on comparative sequence similarity search alone.

Broad-scale analyses on the dynamics of active retro-element group II introns have been conducted in different prokaryotes and indicate rapid gain-loss-dynamics (Lambowitz and Zimmerly, 2011). For example, group II introns like *E.c.*I4 from *E.coli* can be found as almost identical copies (>90% nucleotide identities) with varying copy numbers of 1-15 that are distributed on chromosomes and plasmids among Enterobacteriales (Leclercq and Cordaux, 2012). In addition, group II introns with 93% nucleotide identity to *E.c.I4* have also been found in other bacteria like *Yersinia pestis* or *Shigella berneri* (Dai and Zimmerly, 2002). Mainly these *E.c.*I4 group II introns paralogues occupy genes of IS elements within the respective species, whereby they are able to spread via HGT, at least to closely related bacteria (Dai and Zimmerly, 2002). The retro-mobile group of *E.c.I4* paralogues found in different bacterial species have been described as a group II intron family before (Dai and Zimmerly, 2002). Similarly, the symbiotic Rhizobiales bacteria of legumes show varying copy numbers of group II intron RmInt1, ranging from 1 to around 26 copies in strains of *Sinorhizobium adhaerens* and *Sinorhizobium meliloti* (Fernández-López *et al.*, 2005; Toro, Martínez-Abarca and Fernández-López, 2016). RmInt1 copy numbers also differ dramatically among 20 analysed *S. meliloti* strains and populations (Fernández-López *et al.*, 2005; Toro, Martínez-Abarca and Fernández-López, 2016). The retro-mobility capacities of RmInt1 have been studied in detail in heterologous systems like *E. coli* (Molina-Sánchez, García-Rodríguez and Toro, 2016).

These exemplarily studies on prokaryotic group II intron families demonstrate their potential for rapid expansion. In comparison, similarities found among the land plant group II intron families are lower on average in comparison to above 90% identity of paralogues of *E.c.I4* or *RmInt1.* Lower sequence similarities among mitochondrial land-plant group II intron families, however, are owed to the long evolutionary distances, patchy distribution and likely old lateral intron transfer events as overall sequence similarities between paralogues slowly disappeared among land plant mtDNAs. However, we identified exceptionally cases with F08 rps14i114g2 transfer to atp9i87g2i1114g2 in *Phlegmariurus* (85 % nucleotide identity), F10 cox1i1149g2 transfer into itself cox1i1149g2ii652g2 in *Phlegmariurus squarrosus* (81 % nucleotide identity) or F04 atp1i1050g2ii1536g2's likely transfer to cox2i98g2 in *Anthoceros* (97.5% nucleotide

identity) (S. Zumkeller, Gerke, and Knoop 2020, Figure 2, Figure 6). Although internal atp1i1050g2ii1536g2 was not detected as spliced on cDNA level in the zombie-twintron, the high sequence similarity with cox2i98g2 indicate a lateral intron transfer from atp1i1050g2ii1536g2 into *cox2*. In addition, four additional F04 group II intron paralogues (nad6i444g2, nad9i502g2, nad5i881g2, cox1i1116g1ii209g2) were identified in hornworts.

Intron cox1i1149g2ii635g2 (F10) is of particular interest, as the internal intron likely resulted by invasion of cox1i1149g2 into itself. Further, we identified an additional twintron sdh3i249g2ii42g2 in overlooked 5' terminal region of the *sdh3* gene, together with eight additional insertion sites within the mitochondrial genome of *P. squarrosus* (Zumkeller, Gerke and Knoop, 2020). Consequently, cox1i1149g2 displays a case of a "hypermobile" mitochondrial group II intron. Starting from the detection of cox1i1149g2, cox1i1149g2652g2, sdh3i249g2ii42g2 and further paralogs in related taxa in the family Lycopodiaceae. Also, we identified family 10 nad1i699g2 as putative donor for cox1i1149g2 in Lycopodiaceae (Zumkeller, Gerke and Knoop, 2020). Intron nad1i699g2 is ancestrally present in tracheophyte and the fresh-water algae *Coleochaete scutata,* encoding well a conserved IEP with reverse-transcriptase (RT) and endonuclease domain (D/En) (supplementary material Table 1). Clear remnants of *mat-nad1i699g2* can still be found within the fern *Osmunda regalis* (Genebank: Y17815) (Malek and Knoop, 1998), indicating that *mat-nad1i699g2* has been ancestrally present in the vascular plant lineage. Family 10 group II introns match the criteria for mobility and high "invasiveness", in regard of copy numbers (over 10 copy events in the mtDNA of *P. squarrosus*) or presence of maturases that feature functional RT and D/En domains (paralogous *mat-nad1i699g2* and *mat-cobi824g2*). Also, the mtDNA streptophyte algae of *C. scutata* contains four functional and three fossilized paralogues of family 21, including one well conserved *mat-nad5i1725g2* that lacks a D/En domain (supplementary material Table 1). Similar examples for taxon-specific expansions of group II intron families can be found in Coleochaetophyceae (F21), Zygnematophyceae (F24), Liverworts (F05), mosses (F01) and hornworts (F03 and F04).

Few cases of recent evolutionary mobility of land plant mitochondrial group II introns show similar sequence similarities like similar events in prokaryotes. The occurrence of

twintrons, most notably, in case of group II intron family 10 and the number of created paralogs here match the observations found in bacteria now identified in the mtDNA of *P. squarrosus* (Zumkeller, Gerke and Knoop, 2020). Only in euphyllophyte mitochondria no rapid an expansion of a group II intron family could be found and not even one group II intron gain within spermatophytes. Consequently, the previous assumption that embryophyte mitochondrial group II introns cannot undergo lateral transfer (Barkan, 2007; Lambowitz and Belfort, 2015), has likely been biased by incomplete taxon-sampling, eventually and has to be re-evaluated, due to our discoveries in streptophyte algae, bryophytes and pteridophytes.

Likely expansion of group II intron families might be evolutionary old and can only be detected to the overall much more slower evolving mitogenome compared to prokaryotic genomes*. Surprisingly however, mitochondrial group II introns can invade highly demanding house-keeping genes, like the respiratory chain complexes or ribosomal RNA genes without obvious negative effects on the host fitness**. Among bacteria, group II introns rather invade loci where effects on the host fitness are expectedly low (Dai and Zimmerly, 2002; Robart, Seo and Zimmerly, 2007; Simon *et al.*, 2008; Waldern *et al.*, 2020). This indicates that whether spontaneous mitochondrial group II intron mobility led to negative selection of its host, due to cyto-nuclear incompatibilities and exerting species barriers that makes evolutionary survivors rare. Alternatively, mitochondrial group II intron mobility is evolutionary rare, but can be balanced by a rather forgiving splicing apparatus acting simultaneously on paralogues.

3.3.5 Group II intron families likely favored evolution towards generality of maturases and other splicing factors

All 161 introns, except the defect internal introns of zombie-twintrons, meantime identified in the mitochondrial genomes of land plants must be spliced to obtain functional mitochondria. The origin of any group II intron in the first place in an organism is generally explained by invasion of the new genomic location via retro-transposition, normally based on an intron-encoded maturase mediated reverse transcription mechanism (Zimmerly *et al.*, 1995; Zhong and Lambowitz, 2003). Once a group II intron has occupied a new locus, splicing must be guaranteed. Consequently, the novel group

*,** In comparison of group II intron mobility between bacterial and streptophyte mitochondria two points are especially worth mentioning. First, the capture of group II intron mobility in streptophyte mitogenomes covers a evolutionary much larger timescale than the mobility observed in bacteria, due to higher sequence conservation in the mtDNA. Second, the mobility of group II introns among highly conserved and demanded genes of e.g. the respiratory chain within the mtDNA of streptophyte plants is surprising compared to the mobility among bacteria.

II intron must be able to self-splice or must have an associated maturase that maintains splicing.

The model organism *Arabidopsis thaliana* encodes mitochondrial *mat-nad1i728g2* and four nuclear encoded maturases in contrast to 24 mitochondrial group II introns (Mohr and Lambowitz, 2003; Guo and Mower, 2013). To explain the discrepancy, RT/maturases have likely underwent evolutionary adaptations into generalization and/or co-option, accompanied by other protein factors like RNA-binding PPRs, to address splicing of more than one intron (Barkan, 2007; Guo and Mower, 2013; Schmitz-Linneweber *et al.*, 2015). For example, mitochondrially encoded *mat-nad1i728g2* lacks RT0-1 and the D/En subdomain in spermatophytes and *mat-nad1i728g2* not only associates with it selves intron RNA in RNA-Chip and knockdown experiments, but also rpl2i846g2 (F09), rps3i74g2 (F01), nad1i669g2 (F10), nad5i1477g2 (F3) and nad7i209g2 (F01) (Sultan *et al.*, 2016). In fact, many studies on mitochondrial organellar group II intron splicing in flowering plant model organisms support the hypothesis of generalized splicing machinery (Brown, des Francs-Small and Ostersetzer-Biran, 2014). The discovery of mitochondrial group II intron families is likely crucial to understand the evolution of the mitochondrial splicing machinery, as paralogous introns with high sequence similarity could easily be addressed by a shared maturase, hypothetically. Accordingly, knockdown studies of mat-nad1i728g2 (nad1i728g2 F17) of *A. thaliana* show also partial effects on nad4i461g2 (F17) splicing (Sultan *et al.*, 2016).

Group II introns nad1i728g2, nad4i461g2 and cobi399g2 are now defined to belong to intron family F17. Interestingly, the mtDNA of the streptophyte alga *Coleochaete scutata* contains *mat-nad4i461g2c,* which highly resembles *mat-nad1i728g2* that can be found in vascular plant mitochondria (Zumkeller and Knoop, 2022, Figure 9B). Phylogenetic analyses strongly support that maturases of nad1i728g2 and nad4i461g2 share a common ancestry. In comparison to *A. thaliana mat-nad1i728g2*, *mat-nad4i461g2c* of *C. scutata* features subdomains RT0-7, including significant amino acid-motifs crucial for reverse transcriptase activity and DNA-binding/ endonuclease domain (see supplementary material Table 1). One notable criterion for the identification of active RTs is the conservation of the “YADD” or “YXDD” motif located in subdomain RT5, involved in recruiting dNTPs during RNA to DNA transcription (Wakefield, Jablonski and Morrow,

1992; Boyer *et al.*, 2001; Zhao, Liu and Pyle, 2018). Interestingly, 72 codon identities are changed via RNA editing in the fern *Haplopteris ensiformis* across RT0-7 and X-domain of *mat-nad1i728g2c* - among them are three edits within RT5 subdomain that change "*HADDF*" motif into "RYADDFR" (Zumkeller, Polsakiewicz and Knoop, 2022, Suppl. Table 2). Also *mat-nad1i728g2* in the lycophyte *Phlegmariurus squarrosus* and even pseudogene *mat-nad1i728g2* in the hornwort *Leiosporoceros dussii* indicate conservation of subdomains RT0-7, while within the latter, signatures of an D/En domain can be identified still (see supplementary material Table 1). Thus, the ancestral version of *mat-nad1i728g2* in the common ancestor of hornworts and tracheophytes likely also consisted of well conserved RT0-7 subdomains, including, e.g. the YADD motif and a D/En domain, resembling a retro-mobile type of maturase ancestrally. Subsequently, mat-nad1i728g2 has lost the sub-domain responsible for retro-mobility. Among embryophytes, paralogous F17 nad1i728g2 and nad4i461g2 are both simultaneously present in mosses, hornworts, lycophytes, ferns and spermatophyte mitochondria) (Mower, 2020; Zumkeller and Knoop, 2022). It is possible that nad4i461g2 has been retro-transferred into nad1i728g2 exerted by a fully conserved *mat-nad4i461g2* among the ancestor of mosses, hornworts, and vascular plants (Figure 4 B1). The maturase of nad4i461g2 likely degenerated over time, while *mat-nad1i728g2/ matR* persisted (Figure 4 C3).

In the lycophyte *Isoetes engelmannii* and the flowering plant *Pelargonium spp.* nad1i728g2 together with *mat-nad1i728g2* has been lost from the mitochondria, but retained nad4i461g2 (Grewe *et al.*, 2009; Grewe, Zhu and Mower, 2016). In *Pelargonium* a mat-nad1i728g2 homolog with targeting signals predicted for mitochondrial activity has been identified in its nucleus (Grewe, Zhu and Mower, 2016). Likely, the mitochondrial derived nuclear encoded maturase in *Pelargonium* is still responsible for splicing (Figure 4 B4).

In the completely sequenced early land plant model organism *Physcomitrium patens*, 24 group II introns and two maturase genes (i.e., *mat-cox1i732g*2 S-type and *mat-nad5i1455g2* F25) are found in the mitochondrial genome, whereas 12 genes with high similarity to maturases are identified in the nuclear genome (Terasawa *et al.*, 2007; Rensing *et al.*, 2008; Guo and Mower, 2013). Here, the *P. patens* nMAT4 and nMAT5

display significant sequence similarity to the maturase gene of *mat-atp9i87g2* F07 respectively. Subdomains nMAT4 and 5 resemble RT0- RT7 and the X-subdomain, respectively (Bachelor thesis Wartmann, 2019; Zumkeller and Knoop, 2022, Suppl. Figure 2). The *mat-atp9i87g2* gene has likely been transferred to the nucleus in the ancestors of mosses and likely underwent gene fragmentation, after Sphagnales and extent mosses split, from one to two functional proteins, both independently transcribed and predicted for mitochondrial targeting (Bachelor thesis Wartmann, 2019). F07 paralogous atp9i87g2 and nad2i156g2 exist without an intron-encoded maturase in moss mitochondria.

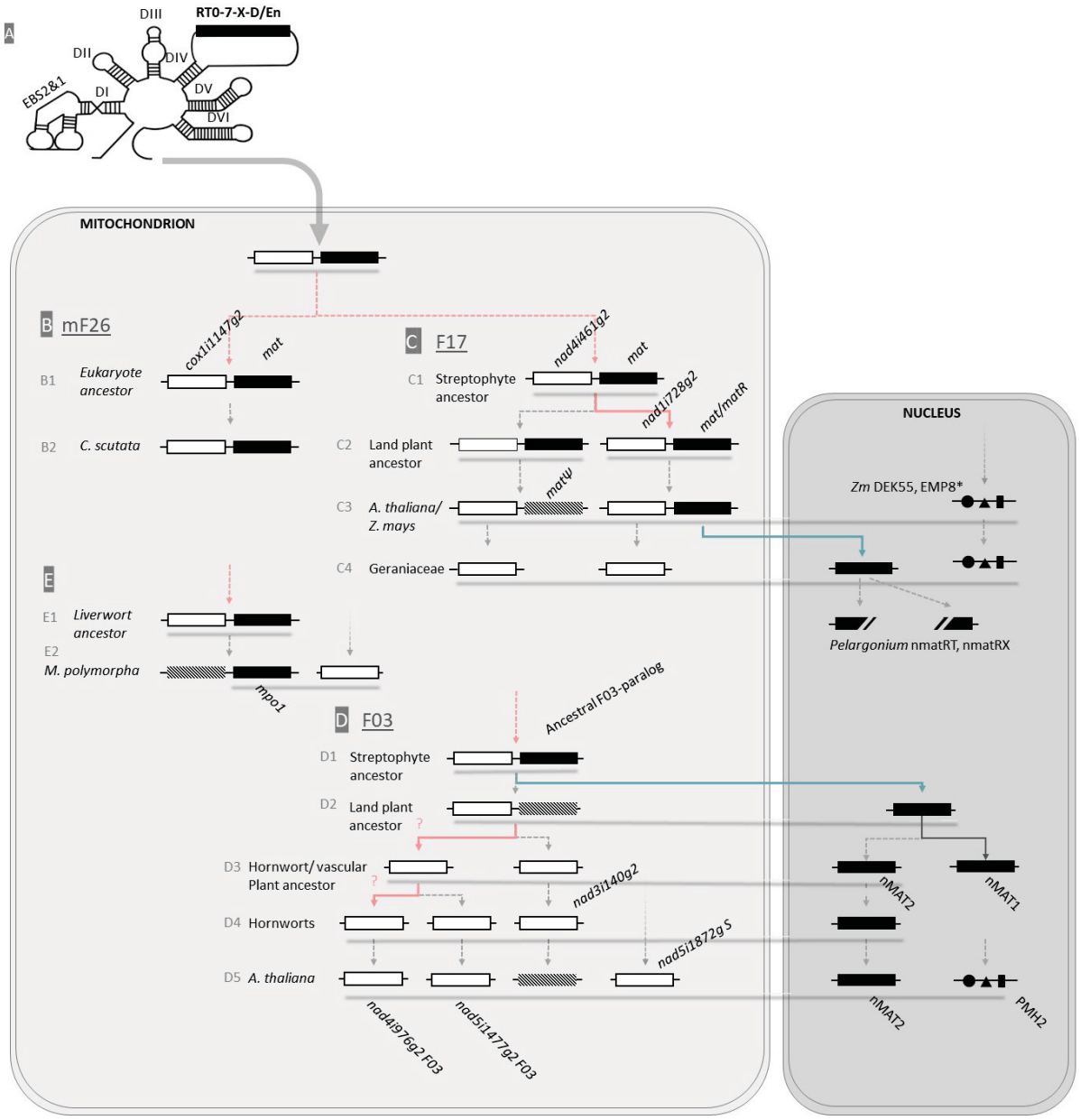


Figure 4 Modes of land plant mitochondrial group II intron splicing complexes evolving. A. *A hypothetical ancient group II intron (white boxes) with a conserved maturase (mat, black boxes) consists of reverse transcriptase domains (RT0-7), splicing core domain (x) and DNA binding domain with an DNA binding/ endonuclease domain (D/En). The ancestral group II intron entered the mitochondrial lineage and created ancient paralogues (dotted red lines). The model for group II intron complexes assumes that ancestral mitochondrial group II intron splicing must have been associated with an intron-encoded maturase protein to guarantee splicing. Correct splicing is a condition for inheritance of group II intron (grey dotted arrows). Association between intron and protein factor are illustrated by grey underlines below boxes. A complete intron-encoded maturase (RT0-7, X, D/En) might also promote intron mobility (red arrows). It is unknown if a degenerated maturase (striped boxes), which acts on multiple targets or has been transferred to the nucleus (blue arrows) can promote lateral intron transfer events as well (question marks).* B. *Solitary type group II introns lost significant sequence similarity to*

other paralogues and/or did simply not created paralogues. Well-conserved maturases of S-type group II introns can still be used for phylogenetic analyses, like for e.g. mF26 cox1i1147g2 and its cognate mat-cox1i1147g2*, which is only present in the genus of* Coleochaete scutata (C. scutata) *among Streptophytes*. C. *Mitochondrial group II intron family F17 was likely gained among the common ancestor of Coleochaetophyceaen algae and land plants with intron nad4i461g2 and its cognate well-conserved maturase* mat-nad4i461g2 *(C1). Among the common ancestor of land plants, nad4i461g2 was laterally transferred to nad1i728g2 together with* mat-nad1i728g2 *(also known as* matR*).* MATR *is the only mitochondrial encoded maturase amongst ferns and seed-plants (C1). Maturase of nad4i461g2 has been degenerated or lost among land plants. In flowering plant model organism* Arabidopsis thaliana (A. thaliana) *it has been demonstrated that mat-nad1i728g2 supports splicing of multiple mitochondrial group II introns, including both F17 paralogues (C2)* (Sultan *et al.*, 2016)*. In addition, nuclear encoded factors like PPRs* DEK55 *and* EMP8 *and* mTERF *protein* SM3 *from* Zea mays (Z. mays) *assist F17 paralogue splicing, specifically* (Sun *et al.*, 2018; Pan *et al.*, 2019; Ren *et al.*, 2020)*. Among flowering plant family* Geraniaceae mat-nad1i728g2 *has been transferred to the nuclear genome (blue arrow) from where it is likely exported back into the mitochondria to work as splicing factor (C3). The nuclear encoded* mat-nad1i728g2 *homolog of* Pelargonium spp. *is disrupted into to two independently expressed proteins called nmatRT and nmatRX, regarding its split between the domains of RT and an X* (Grewe, Zhu and Mower, 2016)*.* D. *Mitochondrial group II intron family F03 was likely gained among the common ancestor of* Klebsormidium *and Phragmoplastophyta, due to the presence of e.g. cox1i862g2 in Klebsormidium, cox2i550g2 in Coleochaete or nad3i140g2 in Chara or cox2i250g2 in* Marchantia *with well-conserved maturases. Phylogenetic analyses of maturases show that F03 maturases are related to nuclear encoded* nmat1 *and* nmat2 *in flowering plants* (Guo and Mower, 2013; Zumkeller and Knoop, 2022)*. Nuclear* nmat1 *and* nmat2 *resulted likely from a duplication (D3). In flowering plants* nMAT2 *assists splicing in the mitochondria of* A. thaliana *including F03 paralogues introns nad4i976g2 and nad5i1477g2, but also e.g. S-type nad5i1872g2* (Keren *et al.*, 2012; Zmudjak *et al.*, 2017) *(D5). It is unclear if a nuclear encoded maturase can promote mitochondrial group II intron copy events (D3-D4).* E. *Mitochondrial group II intron maturases can evolve into free-standing individual genes. There is for example* mpo1 *in* Marchantia *and other liverworts. The ancestral intron carrying* mpo1 *is unknown. It is unknown if maturase* mpo1 *is involved in splicing of different mitochondrial introns, but its conservation among liverworts suggests this. Land plant chloroplast genomes, encode and express a single well-studied maturase that acts in trans on a broader intron target range, the matK* (Hausner *et al.*, 2006; Barthet and Hilu, 2007)*. Among ferns matK has evolved into a free-standing maturase without the flanking intron, like* mpo1*.*

Phylogenetic analyses of four nuclear encoded maturases nmat1-4 from model organism *Arabidopsis thaliana* and *Amborella trichopoda* show that these were ancestrally mitochondrially encoded and are derived from distinct mitochondrial group II introns (Guo and Mower, 2013). The maturase-dataset of 2013 was taxonomically extended with our updated dataset and phylogenetic analyses were conducted, which

were consistent with previous results. Interestingly, flowering plant nmat1 and nmat2 are related to streptophyte algal F03-type maturases, *Klebsormidium flaccidum mat-cox1i862g2, Coleochaete scutata mat-cox2i550g2, Chara vulgaris mat-nad3i140g2* and liverwort *mat-cox2i250g2c* (Figure 4). Also, nuclear encoded maturases nmat3 and nmat4 form a joint clade with F10-type maturases from liverwort *mat-cobi824g2c* and *Coleochaete scutata mat-nad1i669g2* (Zumkeller and Knoop, 2022, Figure 9B). Interestingly, nmat2, together with an DNA-helicase protein pmh2 has been shown to be involved in splicing of F03 paralogues nad4i976g2 and nad5i1455g2 in *Arabidopsis* (Zmudjak *et al.*, 2017). Nuclear encoded nmat1 is associated with mitochondrial splicing of nad4i976g2, too (Keren *et al.*, 2012). Similarly, association with splicing performance could be found in *Arabidopsis* for nmat3 and nmat4 regarding F10 paralogue nad1i669g2, amongst others (Cohen *et al.*, 2014; Shevtsov-Tal *et al.*, 2020).

That mitochondrial group II intron families share a common set of protein factors underlines the evolutionary relevance of mitochondrial group II intron mobility. Whether it be mitochondrially encoded *mat-nad1i728g2* of F10, nuclear encoded nmats1 and 2 associated with F03 or the diverse set of nuclear encoded protein discussed above and in Zumkeller and Knoop 2022 chapter “Nuclear splicing factor acting on paralogues in the same family and an outlook”, the apparatus responsible for splicing had likely been extended from one intron to its paralogous copy. Accordingly, mitochondrial group II intron mobility is a crucial aspect to the evolution from splicing factors acting specifically on one target, to a factor acting generally on many targets. Understanding the evolution of mitochondrial group II intron splicing apparatus might help to understand, how the eukaryotic spliceosome has evolved billions of years ago.

For comparison with flowering plant models, further studies on early branching land plants are necessary. Mosses like *Physcomitrium patens* show a higher abundance of both mitochondrial and nuclear-encoded maturases compared to vascular plants (Guo and Mower, 2013). Besides *P. patens,* also the mitochondria of liverworts contain a high number of intron-encoded maturase genes. There are seven maturase genes in frame with the host gene harbouring well conserved reverse-transcriptase domains (*mat-cox1i44g2c S-type*, *mat-atp9i87g2c F07*, *mat-cox2i250g2c F03*, *mat-cobi824g2c F10*, *mat-atp1i1050g2c S-type*, *mat-cox1i178g2c F13*), one more in the late-gained *mat-*

rrnSi1065g2 S-type and one free standing maturase (*mat-mpo1)* gene in the mitochondrial genome alone (Master thesis Dmitriev, 2021) (Figure 4). Intriguingly, no nuclear-encoded maturases has been identified for liverworts yet (Guo and Mower, 2013). It will be highly interesting to study the capabilities of mitochondrial group II introns and their associated factors regarding splicing and mobility in this bryophyte species (Master thesis Dmitriev, 2021).

3.3.6 The causes of mitochondrial group II intron lateral transfer are unknown

The capture of 25 core families and 29 maturase superfamilies underlines ancient events of group II intron mobility along the mitochondrial evolution of streptophytes. Surprisingly, land plant mitochondrial group II introns lack well conserved intron-borne maturases that feature sub-domains like the D/En responsible for DNA invasion (Lambowitz and Zimmerly, 2011). Just like splicing, mitochondrial group II intron mobility has likely a multifaceted, more complex explanation to offer compared to prokaryotic counterparts*.

With the identification of group II intron fossils, further interesting details about the propagation of group II introns are provided, which extend beyond a sampling bias that focuses on conserved mitochondrial genes alone (Zumkeller and Knoop, 2022). We documented that the hypermobile intron cox1i1149g2 F10 inserted itself (at least eight times) in the orientation of the flanking genes, except for the case of the *cox3-trnF* IGS copy (Zumkeller, Gerke and Knoop, 2020). This could indicate some sort of strand-specificity for mitochondrial group II intron invasions. Group II introns can make use of a transposition pathway independent of the D/En domain present in some maturases, thus invading an intron-free RNA allele and being subsequently recombined in RNA mediated gene conversion mechanism or use the nascent DNA strand during replication for integration (Cousineau *et al.*, 2000; Fernández-López *et al.*, 2005). Recombination-mediated group II intron transfer and invasion of nascent DNA strands comes with co-conversion of the flanking exons and/or strand bias, respectively (Cousineau *et al.*, 2000; Fernández-López *et al.*, 2005). This kind of lateral intron transfer is RNA-mediated and recombination dependent. Group II intron transposition in the mitochondria of land plants may not actively propagate involving a cognate functionally restricted and intron encoded reverse-transcriptase but may rather partially “hijack” other mechanism present

*: Streptophyte mitochondria use a different set of splicing proteins for group ll intron splicing compared to bacteria, where one protein splices one intron. Specific intron-encoded maturases are rare in streptophyte mitochondria, unlike in bacteria where they are common. lt is unclear whether a cognate maturase is necessary for reverse-splicing and reverse-transcription in streptophyte mitochondria or if other factors are involved.

in streptophyte mitochondria involving active reverse transcription. This would imply, however, that there is some positive selection on a to maintain capabilities of reverse transcription.

Like liverwort mitochondria, the mitogenomes of *Chlamydomonas spp.* conserve a likely group II intron derived free-standing reverse transcriptase gene (Boer and Gray 1988). Its role is currently debated, but it has been hypothesized to synthesize primers from RNA for replication of the linear telomeric ends of the mitochondrial genomes of *Chlamydomonas* species (Vahrenholz et al. 1993; Smith and Craig 2021). Replication of land plant mitochondria is only scarcely understood, but rolling-circle replication, independent of primers, appears rather likely (Morley, Ahmad and Nielsen, 2019). Alternatively, and independently of group II intron derived RTs, DNA polymerases have been described to facilitate reverse transcription activity over short stretches of RNA in humans recently (Chandramouly et al. 2021; Y. Su et al. 2019) or in the case of Pol I in *E. coli* (Ricchetti and Buc, 1993). Especially, DNA polymerase 8 is well conserved among eukaryotes and generally involved in end-joining of dsDNA breaks (Wyatt et al. 2016; Hanscom and McVey 2020}. Surprisingly, DNA polymerase 8 promotes RNA-templated DNA synthesis over micro-homologies (Chandramouly et al. 2021). RNA-DNA hybrids are a common phenomenon in dsDNA repair, to e.g. prevent translocations during end joining or guiding the RecA homologue Rad52 recombinase mediated strand exchange in human and yeast (Mazina *et al.*, 2017; Cohen *et al.*, 2018; McDevitt *et al.*, 2018). In general, DNA polymerase y is responsible for mitochondrial DNA replication, but also polymerase 8 has recently been found in human mitochondria, likely being involved in DNA maintenance (Wisnovsky, Jean, and Kelley 2016; Krasich and Copeland 2017).

The exposition of these recently described processes, shall not serve as an alternative explanation for the occurrence of reverse transcription events in land plant mitochondria, based on DNA polymerase. It rather displays an excellent example pointing to a putative crucial involvement of reverse transcription in RNA mediated DNA homeostasis in land plant mitochondria. A putatively similar evolution is documented in the model organism *Saccharomyces cerevisiae,* where homologous recombination is efficient, RNA was shown to mediate recombination indirectly, through a cDNA intermediate generated by a

reverse transcriptase of a secondarily acquired ancestrally invasive Ty LTR-retro-transposon (Derr and Strathern, 1993; Lesage and Todeschini, 2005).

Apart from our observation on intron mobility, reverse transcription likely plays an additional evolutionary crucial role in gene conversion events, involving loss of introns and RNA editing throughout land plant mitochondrial evolution (Cuenca *et al.*, 2016). Within the flowering plant lineage of Alismatales independent group II intron losses occur frequently. Here, loss of introns is accompanied by loss of RNA editing sites in the flanking region what is interpreted as an localized event of retro-processing, involving recombination of a maturated cDNA copy with the DNA counterpart (RNA-mediated gene conversion) (Cuenca *et al.*, 2016). Similar findings involve taxonomically independent loss of RNA editing sites and intron atp1i361g2 in Pteridaceae ferns (Zumkeller, Knoop and Knie, 2016), intron rps3i249g2 in gymnosperms conifers (Ran, Gao and Wang, 2010), intron nad4i976g2 in the lycophyte *Isoetes* (Grewe *et al.*, 2011) or simultaneous loss of atp1i989g2 and atp1i1050g2 in liverworts genus *Calypogeia* {Slipiko *et al.*, 2017). However, loss of intron is an easy-to-notice molecular feature, localized RNA-mediated gene conversion events are found also independently in for example in *Silene* (Sloan *et al.*, 2010) or *Welwitschia,* where it likely made RNA editing within the mitochondrial genome nearly almost completely obsolete (Fan *et al.*, 2018).

Altogether, retro-processing and retro-transposition of mitochondrial group II introns document a growingly large number of evolutionary events involving reverse transcription throughout land plant evolution. The protein and its native, however, is unknown. A maturase with reverse-transcriptase activity is the best candidate so far.

3.3.7 Ambiguous proteins are involved in group II intron splicing, DNA maintenance and regulation of OXPHOS gene expression in flowering plant mitochondria

There is an intriguing overlap between two mandatory processes within the plant mitochondria. There is little evidence that mitochondrial RNA splicing is regulated in correspondence with mtDNA maintenance and other stimuli to produce mitochondrial reactive oxygen species (ROS).

Studies on *RecA* and *RecG*, organellar nucleoid-associated proteins, involved in DNA maintenance and DNA recombination of *Physcomitrium patens* show a link between group II introns and mitochondrial genome evolution (Odahara *et al.*, 2009, 2015). Here,

for example, KO of *RecA1* or *RecG* results in repeat-mediated genome instability in mitochondria, namely deleterious recombination over small repeats (>100 bp) that cause serious defects in plant growth (Odahara *et al.*, 2009, 2015). Interestingly, the involved repeat regions inducing detrimental effects are paralogous group II introns from the same family, respectively: nad4i461g2 and nad1i728g2 of F17, atp9i87g2 and nad2i156g2 of F07 and nad9i283g2 and nad7i209g2 of F01. However, any group II intron-related specificity is highly unlikely, as *RecG* and *RecA* act on small repeats rather "blindly" after dsDNA breaks and are dually targeted to the chloroplast and mitochondrial genome of *P. patens* and *A. thaliana,* respectively (Odahara *et al.*, 2007, 2015; Miller-Messmer *et al.*, 2012; Wallet *et al.*, 2015).

A rather direct association between group II introns and the DNA repair machinery has been documented with RAD52-like protein ODB1 in *A. thaliana* (Gualberto *et al.*, 2015). RAD52 is well conserved among eukaryotes, where it initiates homologous recombination of ssDNA ends for DNA repair (Symington, 2002). Alternative splicing of *A. thaliana* RAD52-homologue ODB1 produces one isoform that is targeted to the chloroplast and the other to the mitochondrion (Samach *et al.*, 2011). In the mitochondrial genome, ODB1 mediates homologous recombination after genotoxic stress and co-precipitates in nucleo-protein complexes with other proteins involved in DNA repair and maintenance, notably *RecA*, *RecG* or *msh1* (Janicka *et al.*, 2012). Intriguingly, T-DNA insertion lines of *odb1* also increase the abundance of un-spliced *nad2* pre-mRNA (Gualberto *et al.*, 2015). Notably, ODB1 is crucial for the removal of *cis*-spliced nad2i156g2 in *A. thaliana* and additionally of nad1i477g2 (Gualberto *et al.*, 2015). Interestingly, nad1i477g2 and nad2i156g2 both belong to group II intron family F07 (Zumkeller and Knoop, 2022, Figure 8).

One highly interesting example that might particularly show another connection between mitochondrial intron splicing and mtDNA maintenance at the regulatory level. Mitochondrial group II intron splicing involves RUG3 of the regulator of chromosome condensation family (RCC1) in *A. thaliana*. RUG3 KO-lines show mis-assembly and accumulation of complex I proteins, likely exerted by splicing deficiency in nad2i548g2 S-type and nad2i709g2 S-type (Kühn *et al.*, 2011). Further, RUG3 interacts with a key sensor of DNA damage, termed Ataxia Telangiectasia Mutated ATM (Maréchal and Zou, 2013;

Su, Yuan, *et al.*, 2017). ATM is activated by DNA damage and can block the cell cycle at phases G1/s to G2/M prior to and posterior to DNA replication (Maréchal and Zou, 2013). Interestingly, RUG3 mutants show impeded meristematic growth, which indicates cell cycle arrest (Su, Zhao, *et al.*, 2017). Further, RUG3 and ATM double knock-out lines show increased splicing defects in comparison to a single KO of RUG3 only on nad2i548g2 in *A. thaliana* (Su, Zhao, *et al.*, 2017). The authors conclude that RUG3 is working synergistically with ATM, acting as upstream regulators of retrograde signalling by modulating ROS production by induced impairment of *nad2* splicing in case of DNA damage (Su, Zhao, *et al.*, 2017).

Another highly interesting example is the P-type PPR proteins Resistance To Phytophtora parasitica 7 (RTP7) and BSO-insensitive roots 6 (BIR6) involved in the splicing of S-type nad7i140g2 (Koprivova *et al.*, 2010; Yang *et al.*, 2022). Both, PTP7 and BIR6 mutants show enhanced resistance to oxidative stresses under salt, osmotic stress and a similar response in pathogenicity assays of e.g. *Phytophora parasitica* in *A. thaliana* (Koprivova *et al.*, 2010; Yang *et al.*, 2022). Most intriguingly, exposing wild-type *A. thaliana* to an *P. parasitica* infection resulted in the accumulation of non-spliced *nad7* transcripts that eventually led to increased ROS levels, likely as an immune response (Yang *et al.*, 2022). These cases exemplify the regulatory involvement of mitochondrial group II intron splicing. Evolution of these regulatory elements will be highly interesting to study regarding cognate mitochondrial group II intron gains and losses.

It must be noted, however, that mitochondrial RNA processing proteins operate in protein complexes as documented in *A. thaliana* (Rugen *et al.*, 2019; Fuchs *et al.*, 2020). PPR-proteins involved in mitochondrial group II intron *cis*- and *trans*-splicing are at the core of large multi-enzyme-complexes together with the ribosome, for example (Rugen *et al.*, 2019). Accordingly, a complex network of protein-protein interactions in RNA procession in the plant mitochondria can be expected. Complexes involved in RNA processing or DNA maintenance might affect the assembly of each other's reciprocally, directly, or indirectly. Splicing defects of mitochondrial group II introns in plant model organisms result in severe phenotypes. Correspondingly, false or insufficient assembly of the respiratory chain complexes may just coincide with increased ROS production and higher sensitivity of the plant hormone abscisic acid ABA arresting meristematic growth, independent of

mitochondrial splicing defects, too (Schwarzländer *et al.*, 2012; Wang *et al.*, 2018). Mitochondrial RNA processing proteins might have multiple roles, as mutated RTP7, for example, also comes with reduced efficiency of one RNA editing site in *nad1,* two in *nad3* and five in *nad6* (Yang *et al.*, 2022). Mitochondrial group II introns of land plants may only display one hub of many, for all kind of protein interactions and diverse phenotypical responses under the control of the nuclear genome.

4.0 Conclusion and Outlook

The complete assembly of the fern *Haplopteris ensiformis* organellar DNA[s], chloroplast (cpDNA) and mitochondria (mtDNA), provides a reference for the taxonomically diverse and ecologically rich group of polypodiid/ true ferns. Most notably, regarding the high rates of plant C to U and U to C RNA editing in both organelles, the presence of astounding *rrnL* trans-splicing and frequent insertions of promiscuous DNA from different genetic origins within the mtDNA complicated identification of coding sequences and its assembly (Zumkeller, Polsakiewicz and Knoop, 2022). These features of *H. ensiformis* mtDNA are highly peculiar and display now identified pitfalls, which might apply to other true ferns as well. Modern long-read sequencing method, deep RNA sequencing and bioinformatical tools outlined in chapter 3.1.1 should be used in combination for effective mtDNA assembly of fern species.

Our findings of pseudogenized genes and other non-functional remnants of rickettsial DNAs following recent horizontal gene transfer (HGT) within *H. ensiformis* mtDNA are outstanding and require further investigation. The xenologous rickettsial DNAs indicate close habitual contact between *H. ensiformis* and a yet undiscovered probably endosymbiotic/ endoparasitic Rickettsia bacteria and the potential of stable integration of large DNA sequences into an mtDNA of the prokaryotic source (chapter 3.2.3).

Investigating mtDNAs of other closely related fern species of *H. ensiformis* is highly promising regarding exploration of nuclear and mitochondrial dependencies outside the lineage of flowering plants. Like *H. ensiformis, Adiantum capillus-veneris* belongs to a the similar sub-family of Adiantoids but has a completely assembled nuclear genome (Fang *et al.*, 2022). Accordingly, assembly of *A. capillus-veneris* mtDNA should be next allowing linkage of e.g. nuclear encoded splicing or PPR factors and mitochondrially encoded introns and RNA-editing sites, respectively. Especially as *A. capillus-veneris* mtDNA

features group II intron paralogs F06 rps3i249g2 and atp1i361g2, with structurally convergent evolving domains due to a likely shared splicing factors (Zumkeller, Knoop and Knie, 2016).

It has been confirmed for flowering plant model organisms like *Arabidopsis thaliana* and *Zea mays* nuclear encoded splicing proteins target mitochondrial group II intron paralogues of the same family (Zumkeller and Knoop, 2022). Amongst those proteins are ambiguous PPR proteins involved in hormone and ROS signalling like RUG3 and RTP7 or proteins like MSH1 involved in mitochondrial DNA maintenance (chapter 3.3.7). However, it is difficult to track the ever-growing number and variety of proteins associated with organellar splicing among flowering plants in annual review articles, as these become incomplete quickly. Consequently, a curated database for plant organellar splicing proteins appears more practicable.

Collection of proteins that are assigned to mitochondrial group II intron splicing and the phylogenetic capture of their orthologs in non-model organisms displays a promising follow up project. Especially, as splicing proteins targeted to mitochondria and their cognate group II intron likely co-evolve, plant lineages where intron have been lost or gained are of high interest (chapter 3.3.1). For example, atp1i361g2 among Pteridaceaen ferns is lost at least four times independently (Zumkeller, Knoop and Knie, 2016). Or, there is the re-occurring loss of the only group II intron occupying complex III subunits among flowering plants, cox2i373g2, which is absent in *Arabidopsis thaliana*, but present in *Oryza sativa* (Mower, 2020; Zumkeller and Knoop, 2022). Most notably, regarding mitochondrial group II introns gains, among mitogenomes of early vascular plant family of Lycopodiaceae and the hornworts, high-mobility group II intron families F10 and F04 have been identified (chapter 3.3.4, Zumkeller, Gerke and Knoop, 2020). The hornwort *Anthoceros agrestis* has a complete nuclear genome available that can be screened for nuclear encoded proteins (Li *et al.*, 2020). The protein machinery that promotes group II intron mobility along land plant mitochondrial evolution is unknown (chapter 3.3.6).

Ultimately, the findings about group II introns families associated with nuclear encoded splicing factors can be extended across to group I introns and other lineages of the tree of life. Fungi, red algae, diatoms and few metazoans feature mitochondrial group

II introns as well (Huchon *et al.*, 2015; Pogoda, Keepers, Hamsher, *et al.*, 2019; Mukhopadhyay and Hausner, 2021). With cox1i1147g2 and its intron encoded maturase a common group II intron present in fungi, sponges, placozoans and brown, chlorophyceaen and coleochaetophyceaen algae, an ancestral lineage to land plants was identified (Zumkeller and Knoop, 2022).

5.0 Acknowledgement

Completion of this thesis has been difficult. In fact, it took some time and gave a noteworthy group of people a chance to accumulate in the meantime that backed this work. This thesis involved mentors, colleagues, family, and friends. I feel deeply grateful for being this privileged having both, the chance to complete a doctoral thesis at the university of Bonn and for being consistently, unconditionally supported, during this demanding time. First, I want to thank every committed person that an initial interest in biology and evolution, could become a renowned career.

I thank Prof. Dr. Volker Knoop. I am highly grateful for his trust in me and the projects I chose to pursue. His well-considered criticism and persistent interest shaped this thesis, my scientific career and most notably, my curiosity. I am proud of this time and the common achievements, which I consider of invaluable worth.

I want to thank Prof. Dr. Walter Witke, as well, for the recommendation, applying at the AG Molecular Evolution and his entertaining lectures that primed me for genetics, in the first place. In addition, I'd like to express my gratitude for him co-reviewing this thesis. In addition, I'd like to thank Prof. Dr. Raoul-Martin Memmesheimer and Prof. Dr. Ute Weisz that kindly agreed to participate on the examination board.

Great appreciation is due to my co-authors for the brilliant collaboration. Monika Polsakiewicz contributed her immense experience and calmness to the puzzling project of *Haplopteris ensiformis*. Further, I thank Phillipp Gerke for the open and so very fruitful exchange while progressing into the depths of bioinformatics and joint work on twintrons in *Anthoceros*.

I feel grateful, as well, for the helpful institutions without this work could not have been performed. The botanical garden Bonn provided me with plant material and expert expertise, embodied by the kind and informative curator Cornelia Löhne and gardener

Bernd Reinken. In addition, I must thank the patient anonymous helpdesk of the Paderborn University for Parallel Computing PC2.

Also thank you, colleagues, of the AG Knoop for reviewing, commenting, and challenging my ideas and my work. Here I want to especially thank Dr. Mareike Schallenberg-Rüdinger for keeping enduring oversight, Dr. Bastian Oldenkott for his clear honesty, Elena Lesch for the inspiring energy and ultimately, Dietrich Dmitriev, who read it all and did not lose interest. Thank you for pushing this thesis to its best.

Last, I thank my very own F01 including my brother Severin, my parents Petra and Michael, and their mothers, Rosi and Mathilde for the unwavering trust and support all along this long road. I also want to thank the more recently gained F02 of Mirka and Leszek for being so very welcoming and ultimately, as now forming a super-family F01-F02 due to our son Janosch, my deepest gratitude belongs to the incredibly patience and goodwill of Victoria.

So long and thanks for all, the *Fich*.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. 1990. “Basic Local Alignment Search Tool.” Journal of Molecular Biology 215 (3): 403–10. https://doi.org/10.1016/S0022-2836{05)80360-2.

Alverson, Andrew J., Danny W. Rice, Stephanie Dickinson, Kerrie Barry, and Jeffrey D. Palmer. 2011. "Origins and Recombination of the Bacterial-Sized Multichromosomal Mitochondrial Genome of Cucumber." Plant Cell 23 (7): 2499–2513. https://doi.org/10.1105/tpc.111.087189.

Alverson, Andrew J., Xiaoxin Wei, Danny W. Rice, David B. Stern, Kerrie Barry, and Jeffrey D. Palmer. 2010. "Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of Citrullus Lanatus and Cucurbita Pepo {Cucurbitaceae)." Molecular Biology and Evolution 27 {6): 1436–48. https://doi.org/10.1093/molbev/msq029.

Alverson, Andrew J., Shi Zhuo, Danny W. Rice, Daniel B. Sloan, and Jeffrey D. Palmer. 2011. "The Mitochondrial Genome of the Legume Vigna Radiata and the Analysis of Recombination across Short Mitochondrial Repeats." PLoS ONE 6 {1). https://doi.org/10.1371/journal.pone.0016404.

Arimura, Shin Ichi. 2018. "Fission and Fusion of Plant Mitochondria, and Genome Maintenance.” Plant Physiology 176 {1): 152–61. https://doi.org/10.1104/PP.17.01025.

Arimura, Shin Ichi, Junko Yamamoto, Gen Paul Aida, Mikio Nakazono, and Nobuhiro Tsutsumi. 2004. "Frequent Fusion and Fission of Plant Mitochondria with Unequal Nucleoid Distribution." Proceedings of the National Academy of Sciences of the United States of America 101 {20): 7805–8. https://doi.org/10.1073/pnas.0401077101.

Backert, Steffen, Rudi Lurz, and Thomas Börner. 1996. "Electron Microscopic Investigation of Mitochondrial DNA from Chenopodium Album {L.)." Current Genetics 29 {5): 427–36. https://doi.org/10.1007/BF02221510.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." Journal of Computational Biology 19 {5): 455. https://doi.org/10.1089/CMB.2012.0021.

Barkan, Alice. 2007. "Intron Splicing in Plant Organelles." In Molecular Biology and Biotechnology of Plant Organelles, 295–322. Springer Netherlands. https://doi.org/10.1007/978-1-4020-3166-3_11.

Barkan, Alice, Larik Klipcan, Oren Ostersetzer, Tetsuya Kawamura, Yukari Asakura, and Kenneth P. Watkins. 2007. "The CRM Domain: An RNA Binding Module Derived from an Ancient Ribosome-Associated Protein." RNA 13 {1): 55–64. https://doi.org/10.1261/rna.139607.

Barthet, Michelle M., and Khidir W. Hilu. 2007. "Expression of MatK: Functional and Evolutionary Implications." American Journal of Botany 94 {8): 1402–12. https://doi.org/10.3732/ajb.94.8.1402.

Barton, Kiah A., Michael R. Wozny, Neeta Mathur, Erica-Ashley Jaipargas, Jaideep Mathur, and Jenny Russinova. 2018. "Chloroplast Behaviour and Interactions with Other Organelles in Arabidopsis Thaliana Pavement Cells." Journal of Cell Science 131 {2). https://doi.org/10.1242/JCS.202275.

Begu, Dominique, and Alejandro Araya. 2009. "The Horsetail Equisetum Arvense Mitochondria Share Two Group I Introns with the Liverwort Marchantia, Acquired a Novel Group II Intron but Lost Intron-Encoded ORFs." Current Genetics 55 {1): 69–79. https://doi.org/10.1007/s00294-008-0225-7.

Belfort, Marlene. 2017. “Mobile Self-Splicing Introns and Inteins as Environmental Sensors.” Current Opinion in Microbiology 38 {August): 51. https://doi.org/10.1016/J.MIB.2017.04.003.

Belhocine, Kamila, Victoria Mandilaras, Bonnie Yeung, and Benoit Cousineau. 2007. "Conjugative Transfer of the Lactococcus Lactis Sex Factor and PRS01 Plasmid to Enterococcus Faecalis." FEMS Microbiology Letters 269 {2): 289–94. https://doi.org/10.1111/j.1574-6968.2007.00641.x.

Bell, Neil E., Jeffrey L. Boore, Brent D. Mishler, and Jaakko Hyvönen. 2014. "Organellar Genomes of the Four-Toothed Moss, Tetraphis Pellucida." BMC Genomics 15 {1): 1–11. https://doi.org/10.1186/1471-2164-15-383.

Bendich, AJ. 1993. “Reaching for the Ring: The Study of Mitochondrial Genome Structure.” Curr Genet 24: 279–90. doi: 10.1007/BF00336777.

Bergthorsson, Ulfar, Aaron O. Richardson, Gregory J. Young, Leslie R. Goertzen, and Jeffrey D. Palmer. 2004. “Massive Horizontal Transfer of Mitochondrial Genes from Diverse Land Plant Donors to the Basal Angiosperm Amborella." Proceedings of the National Academy of Sciences of the United States of America 101 (51): 17747–52. https://doi.org/10.1073/pnas.0408336102.

Best, Corinne, Ron Mizrahi, and Oren Ostersetzer-Biran. 2020. “Why so Complex? The Intricacy of Genome Structure and Gene Expression, Associated with Angiosperm Mitochondria, May Relate to the Regulation of Embryo Quiescence or Dormancy—Intrinsic Blocks to Early Plant Life." Plants 9 (5): 598. https://doi.org/10.3390/plants9050598.

Bing, Xiao-Li, Wen-Qiang Xia, Jia-Dong Gui, Gen-Hong Yan, Xiao-Wei Wang, and Shu-Sheng Liu. 2014. "Diversity and Evolution of the Wolbachia Endosymbionts of Bemisia {Hemiptera: Aleyrodidae) Whiteflies." Ecology and Evolution 4 {13): 2714–37. https://doi.org/10.1002/ECE3.1126.

Blanc, Guillaume, Hiroyuki Ogata, Catherine Robert, Stéphane Audic, Jean Michel Claverie, and Didier Raoult. 2007. “Lateral Gene Transfer between Obligate Intracellular Bacteria: Evidence from the Rickettsia Massiliae Genome." Genome Research 17 {11): 1657–64. https://doi.org/10.1101/gr.6742107.

Blazier, John C., Robert K. Jansen, Jeffrey P. Mower, Madhu Govindu, Jin Zhang, Mao Lun Weng, and Tracey A. Ruhlman. 2016. "Variable Presence of the Inverted Repeat and Plastome Stability in Erodium.” Annals of Botany 117 (7): 1209–20. https://doi.org/10.1093/aob/mcw065.

Blocker, Forrest J.H., Georg Mohr, Lori H. Conlan, Li Qi, Marlene Belfort, and AlanM. Lambowitz. 2005. “Domain Structure and Three-Dimensional Model of a Group II Intron-Encoded Reverse Transcriptase." RNA 11 {1): 14–28. https://doi.org/10.1261/RNA.7181105.

Bock, Ralph. 2010. “The Give-and-Take of DNA: Horizontal Gene Transfer in Plants." Trends in Plant Science. Elsevier. https://doi.org/10.1016/j.tplants.2009.10.001.

Bock, Ralph, and Jeremy N. Timmis. 2008. "Reconstructing Evolution: Gene Transfer from Plastids to the Nucleus." BioEssays. John Wiley and Sons Inc. https://doi.org/10.1002/bies.20761.

Boesch, Pierre, Frederique Weber-Lotfi, Noha Ibrahim, Vladislav Tarasenko, Anne Cosset, Fran&#xfffd;ois Paulus, Robert N. Lightowlers, and Andre Dietrich. 2011. "DNA Repair in Organelles: Pathways, Organization, Regulation, Relevance in Disease and Aging." Biochimica et Biophysica Acta - Molecular Cell Research. Elsevier. https://doi.org/10.1016/j.bbamcr.2010.10.002.

Boer, P. H., and M. W. Gray. 1988. "Genes Encoding a Subunit of Respiratory NADH Dehydrogenase {ND1) and a Reverse Transcriptase-like Protein {RTL) Are Linked to Ribosomal RNA Gene Pieces in Chlamydomonas Reinhardtii Mitochondrial DNA." The EMBO Journal 7 {11): 3501–8. https://doi.org/10.1002/j.1460-2075.1988.tb03226.x.

Bonen, Linda. 2008. “Cis- and Trans-Splicing of Group II Introns in Plant Mitochondria." Mitochondrion 8 {1): 26–34. https://doi.org/10.1016/j.mito.2007.09.005.

Bonen, Linda, and Jörg Vogel. 2001. "The Ins and Outs of Group II Introns." Trends in Genetics. Elsevier Current Trends. https://doi.org/10.1016/S0168-9525(01)02324-1.

Boore, Jeffrey L. 1999. "Animal Mitochondrial Genomes." Nucleic Acids Research 27 {8): 1767–80. https://doi.org/10.1093/nar/27.8.1767.

Boussau, Bastien, E. Olof Karlberg, A. Carolin Frank, Boris Antoine Legault, and Siv G.E. Andersson. 2004. "Computational Inference of Scenarios for a-Proteobacterial Genome Evolution." Proceedings of the National Academy of Sciences of the United States of America 101 {26): 9722–27. https://doi.org/10.1073/pnas.0400975101.

Boyer, Paul L., Hong-Qiang Gao, Patrick K. Clark, Stefan G. Sarafianos, Edward Arnold, and Stephen H. Hughes. 2001. "YADD Mutants of Human Immunodeficiency Virus Type 1 and Moloney Murine Leukemia Virus Reverse Transcriptase Are Resistant to Lamivudine Triphosphate {3TCTP) In Vitro." Journal of Virology 75 {14): 6321–28. https://doi.org/10.1128/jvi.75.14.6321-6328.2001.

Braun, Hans Peter. 2020. "The Oxidative Phosphorylation System of the Mitochondria in Plants." Mitochondrion 53 {July): 66–75. https://doi.org/10.1016/J.MITO.2020.04.007.

Braun, Hans Peter, Stefan Binder, Axel Brennicke, Holger Eubel, Alisdair R. Fernie, Iris Finkemeier, Jennifer Klodmann, et al. 2014. "The Life of Plant Mitochondrial Complex I." Mitochondrion 19 Pt B {PB): 295–313. https://doi.org/10.1016/J.MITO.2014.02.006.

Brigham, Laurel M., Luis M. Allende, Benjamin R. Shipley, Kayla C. Boyd, Tanya J. Higgins, Nicholas Kelly, Carly R. Anderson Stewart, et al. 2018. “Genomic Insights into the Mitochondria of 11 Eastern North American Species of Cladonia." Mitochondrial DNA Part B: Resources 3 {2): 508–12. https://doi.org/10.1080/23802359.2018.1463827.

Britten, Roy J., and Eric H. Davidson. 1969. "Gene Regulation for Higher Cells: A Theory." Science 165 (3891): 349–57. https://doi.org/10.1126/science.165.3891.349.

Brown, Gregory G., Catherine Colas des Francs-Small, and Oren Ostersetzer-Biran. 2014. “Group II Intron Splicing Factors in Plant Mitochondria." Frontiers in Plant Science. Frontiers Research Foundation. https://doi.org/10.3389/fpls.2014.00035.

Burger, Gertraud, Diane Saint-Louis, Michael W. Gray, and B. Franz Lang. 1999. "Complete Sequence of the Mitochondrial DNA of the Red Alga Porphyra Purpurea: Cyanobacterial Introns and Shared Ancestry of Red and Green Algae.” Plant Cell 11 {9): 1675–94. https://doi.org/10.1105/tpc.11.9.1675.

Busk, Peter K., and Lene Lange. 2015. "Classification of Fungal and Bacterial Lytic Polysaccharide Monooxygenases." BMC Genomics 16 {1): 1–13. https://doi.org/10.1186/s12864-015-1601-6.

Candales, Manuel A., Adrian Duong, Keyar S. Hood, Tony Li, Ryan A.E. Neufeld, Runda Sun, Bonnie A. McNeil, Li Wu, Ashley M. Jarding, and Steven Zimmerly. 2012. "Database for Bacterial Group II Introns." Nucleic Acids Research 40 {D1): D187–90. https://doi.org/10.1093/nar/gkr1043.

Carrillo, C., Y. Chapdelaine, and L. Bonen. 2001. "Variation in Sequence and RNA Editing within Core Domains of Mitochondrial Group II Introns among Plants." Molecular and General Genetics 264 {5): 595–603. https://doi.org/10.1007/s004380000345.

Castandet, Beno1t, David Choury, Dominique Begu, Xavier Jordana, and Alejandro Araya. 2010. "Intron RNA Editing Is Essential for Splicing in Plant Mitochondria." Nucleic Acids Research 38 {20): 7112–21. https://doi.org/10.1093/nar/gkq591.

Chandramouly, Gurushankar, Jiemin Zhao, Shane McDevitt, Timur Rusanov, Trung Hoang, Nikita Borisonnik, Taylor Treddinick, et al. 2021. "Pole Reverse Transcribes RNA and Promotes RNA-Templated DNA Repair." Science Advances 7 {24): 1771–82. https://doi.org/10.1126/sciadv.abf1771.

Cavalier-Smith, T. 1991. "Intron Phylogeny: A New Hypothesis." Trends in Genetics 7 {5): 145–48. https://doi.org/10.1016/0168-9525(91)90377-3.

Chen, Z., H. Nie, C. E. Grover, Y. Wang, P. Li, M. Wang, H. Pei, et al. 2017. "Entire Nucleotide Sequences of Gossypium Raimondii and G. Arboreum Mitochondrial Genomes Revealed A-Genome Species as Cytoplasmic Donor of the Allotetraploid Species." Plant Biology 19 {3): 484–93. https://doi.org/10.1111/plb.12536.

Chen, Zongliang, Hong Chun Wang, Jiayu Shen, Feng Sun, Miaodi Wang, Chunhui Xu, and Bao Cai Tan. 2019. "PPR-SMR1 Is Required for the Splicing of Multiple Mitochondrial Introns, Interacts with Zm-MCSF1, and Is Essential for Seed Development in Maize." Journal of Experimental Botany 70 {19): 5245–58. https://doi.org/10.1093/jxb/erz305.

Cheng, Ning, Yih-Shan Lo, Mohammad Israil Ansari, Kuo-Chieh Ho, Shih-Tong Jeng, Na-Sheng Lin, and Hwa Dai. 2017. "Correlation between MtDNA Complexity and MtDNA Replication Mode in Developing Cotyledon Mitochondria during Mung Bean Seed Germination." New Phytologist 213 (2): 751–63. https://doi.org/10.1111/nph.14158.

Cheng, Shifeng, Wenfei Xian, Yuan Fu, Birger Marin, Jean Keller, Tian Wu, Wenjing Sun, et al. 2019. "Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution." Cell 179 {5): 1057-1067.e14. https://doi.org/10.1016/J.CELL.2019.10.019.

Chick, BF, SN Carroll, C Kennedy, and BV McCleary. 1981. "Some Biochemical Features of an Outbreak of Polioencephalomalacia in Sheep." Australian Veterinary Journal 57 {5): 251–52. https://doi.org/10.1111/J.1751-0813.1981.TB02680.X.

Chorev, Michal, and Liran Carmel. 2012. "The Function of Introns." Frontiers in Genetics 3 {APR): 55. https://doi.org/10.3389/fgene.2012.00055.

Christenhusz, Maarten J.M., and Mark W. Chase. 2014. “Trends and Concepts in Fern Classification." Annals of Botany. Oxford University Press. https://doi.org/10.1093/aob/mct299.

Christensen, Alan C. 2018. "Mitochondrial DNA Repair and Genome Evolution." In Annual Plant Reviews Online, 11–32. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119312994.apr0544.

Chumley, Timothy W., Jeffrey D. Palmer, Jeffrey P. Mower, H. Matthew Fourcade, Patrick J. Calie, Jeffrey L. Boore, and Robert K. Jansen. 2006. "The Complete Chloroplast Genome Sequence of Pelargonium x Hortorum: Organization and Evolution of the Largest and Most Highly Rearranged Chloroplast Genome of Land Plants." Molecular Biology and Evolution 23 {11): 2175–90. https://doi.org/10.1093/molbev/msl089.

Civan, Peter, Peter G. Foster, Martin T. Embley, Ana Seneca, and Cymon J. Cox. 2014. “Analyses of Charophyte Chloroplast Genomes Help Characterize Theancestral Chloroplastgenomeof Land Plants." Genome Biology and Evolution 6 {4): 897–911. https://doi.org/10.1093/gbe/evu061.

Clark, James, Oriane Hidalgo, Jaume Pellicer, Hongmei Liu, Jeannine Marquardt, Yannis Robert, Maarten Christenhusz, et al. 2016. "Genome Evolution of Ferns: Evidence for Relative Stasis of Genome Size across the Fern Phylogeny." New Phytologist 210 {3): 1072–82. https://doi.org/10.1111/nph.13833.

Clifton, Sandra W., Patrick Minx, Christiane M.R. Fauron, Michael Gibson, James O. Allen, Hui Sun, Melissa Thompson, et al. 2004. "Sequence and Comparative Analysis of the Maize NB Mitochondrial Genome." Plant Physiology 136 {3): 3486–3503. https://doi.org/10.1104/PP.104.044602.

Cohen, Sarah, Nadine Puget, Yea Lih Lin, Thomas Clouaire, Marion Aguirrebengoa, Vincent Rocher, Philippe Pasero, Yvan Canitrot, and Gaelle Legube. 2018. "Senataxin Resolves RNA:DNA Hybrids Forming at DNA Double-Strand Breaks to Prevent Translocations." Nature Communications 9 {1): 1–14. https://doi.org/10.1038/s41467-018-02894-w.

Cohen, Sigal, Michal Zmudjak, Catherine Colas Des Francs-Small, Sunita Malik, Felix Shaya, Ido Keren, Eduard Belausov, et al. 2014. "NMAT4, A Maturase Factor Required for Nad1 Pre-MRNA Processing

and Maturation, Is Essential for Holocomplex i Biogenesis in Arabidopsis Mitochondria." Plant Journal 78 (2): 253–68. https://doi.org/10.1111/tpj.12466.

Cole, Logan W., Wenhu Guo, Jeffrey P. Mower, and Jeffrey D. Palmer. 2018. “High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants." Molecular Biology and Evolution 35 {11): 2773–85. https://doi.org/10.1093/molbev/msy176.

Cousineau, Benoit, Stacey Lawrence, Dorie Smith, and Mariene Belfort. 2000. "Retrotransposition of a Bacterial Group II Intron." Nature 404 {6781): 1018–21. https://doi.org/10.1038/35010029.

Cuenca, Argelia, T. Gregory Ross, Sean W. Graham, Craig F. Barrett, Jerrold I. Davis, Ole Seberg, and Gitte Petersen. 2016. "Localized Retroprocessing as a Model of Intron Loss in the Plant Mitochondrial Genome." Genome Biology and Evolution 8 {7): 2176–89. https://doi.org/10.1093/gbe/evw148.

Dai, Lixin, and Steven Zimmerly. 2002. "Compilation and Analysis of Group II Intron Insertions in Bacterial Genomes: Evidence for Retroelement Behavior." Nucleic Acids Research. Oxford University Press. https://doi.org/10.1093/nar/30.5.1091.

Dalby, Stephen J., and Linda Bonen. 2013. “Impact of Low Temperature on Splicing of Atypical Group II Introns in Wheat Mitochondria." Mitochondrion 13 {6): 647–55. https://doi.org/10.1016/j.mito.2013.09.001.

Delgado-Gonzalez, Cristian Raziel, Alfredo Madariaga-Navarrete, Jose Miguel Fernandez-Cortés, Margarita Islas-Pelcastre, Goldie Oza, Hafiz M.N. Iqbal, and Ashutosh Sharma. 2021. “Advances and Applications of Water Phytoremediation: A Potential Biotechnological Approach for the Treatment of Heavy Metals from Contaminated Water." International Journal of Environmental Research and Public Health 18 {10). https://doi.org/10.3390/IJERPH18105215.

Dellaporta, Stephen L., Anthony Xu, Sven Sagasser, Wolfgang Jakob, Maria A. Moreno, Leo W. Buss, and Bernd Schierwater. 2006. "Mitochondrial Genome of Trichoplax Adhaerens Supports Placozoa as the Basal Lower Metazoan Phylum." Proceedings of the National Academy of Sciences of the United States of America 103 (23): 8751–56. https://doi.org/10.1073/pnas.0602076103.

Derr, Leslie K., and Jeffrey N. Strathern. 1993. "A Role for Reverse Transcripts in Gene Conversion." Nature 361 {6408): 170–73. https://doi.org/10.1038/361170a0.

Deshpande, N N, Y Bao, and D L Herrin. 1997. "Evidence for Light/Redox-Regulated Splicing of PsbA Pre-RNAs in Chlamydomonas Chloroplasts." RNA 3 {1): 37. /pmc/articles/PMC1369460/?report=abstract.

Dierckxsens, Nicolas, Patrick Mardulyn, and Guillaume Smits. 2017. "NOVOPlasty: De Novo Assembly of Organelle Genomes from Whole Genome Data." Nucleic Acids Research 45 {4): e18. https://doi.org/10.1093/NAR/GKW955.

Dlakic, Mensur, and Arcady Mushegian. 2011. "Prp8, the Pivotal Protein of the Spliceosomal Catalytic Center, Evolved from a Retroelement-Encoded Reverse Transcriptase." RNA 17 {5): 799. https://doi.org/10.1261/RNA.2396011.

Dmitriev, Dietrich. 2021. "Expression of Recombinant, Plant Mitochondrial Group II Introns in Escherichia Coli.” Master Thesis AG Knoop IZMB.

Dombrovska, Olena, and Yin Long Qiu. 2004. "Distribution of Introns in the Mitochondrial Gene Nad1 in Land Plants: Phylogenetic and Molecular Evolutionary Implications." Molecular Phylogenetics and Evolution 32 {1): 246–63. https://doi.org/10.1016/j.ympev.2003.12.013.

Dong, Shanshan, Jia Yu Xue, Shouzhou Zhang, Li Zhang, Hong Wu, Zhiduan Chen, Bernard Goffinet, and Yang Liu. 2018. "Complete Mitochondrial Genome Sequence of Anthoceros Angustus:

Conservative Evolution of the Mitogenomes in Hornworts." Bryologist 121 {1): 14–22. https://doi.org/10.1639/0007-2745-121.1.014.

Dong, Shanshan, Chaoxian Zhao, Shouzhou Zhang, Li Zhang, Hong Wu, Huan Liu, Ruiliang Zhu, Yu Jia, Bernard Goffinet, and Yang Liu. 2019. "Mitochondrial Genomes of the Early Land Plant Lineage Liverworts {Marchantiophyta): Conserved Genome Structure, and Ongoing Low Frequency Recombination." BMC Genomics 20 {1): 1–14. https://doi.org/10.1186/s12864-019-6365-y.

Doolittle, W. F. 2013. "The Spliceosomal Catalytic Core Arose in the RNA World.. or Did It?" Genome Biology 14 (12): 1–3. https://doi.org/10.1186/GB4145/METRICS.

Doolittle, W. Ford, and Carmen Sapienza. 1980. "Selfish Genes, the Phenotype Paradigm and Genome Evolution." Nature 284 {5757): 601–3. https://doi.org/10.1038/284601A0.

Duchêne, Anne Marie, and Laurence Maréchal-Drouard. 2001. “The Chloroplast-Derived TrnW and TrnM-e Genes Are Not Expressed in Arabidopsis Mitochondria." Biochemical and Biophysical Research Communications 285 (5): 1213–16. https://doi.org/10.1006/bbrc.2001.5303.

Eckardt, Nancy A. 2007. "Chloroplast Intron Splicing Mechanisms." The Plant Cell 19 {12): 3838. https://doi.org/10.1105/TPC.108.191210.

Ehara, Megumi, Kazuo I. Watanabe, and Takeshi Ohama. 2000. “Distribution of Cognates of Group II Introns Detected in Mitochondrial Cox1 Genes of a Diatom and a Haptophyte." Gene 256 {1–2): 157–67. https://doi.org/10.1016/S0378-1119(00)00359-0.

Fan, Weishu, Wenhu Guo, Lexis Funk, Jeffrey P. Mower, and Andan Zhu. 2018. "Complete Loss of RNA Editing from the Plastid Genome and Most Highly Expressed Mitochondrial Genes of Welwitschia Mirabilis." Science China Life Sciences 62 {4): 498–506. https://doi.org/10.1007/s11427-018-9450-1.

Fang, Yuhan, Xing Qin, Qinggang Liao, Ran Du, Xizhi Luo, Qian Zhou, Zhen Li, et al. 2022. "The Genome of Homosporous Maidenhair Fern Sheds Light on the Euphyllophyte Evolution and Defences." Nature Plants 2022 8:9 8 {9): 1024–37. https://doi.org/10.1038/s41477-022-01222-x.

Farré, Jean Claude, Cindy Aknin, Alejandro Araya, and Beno1t Castandet. 2012. "RNA Editing in Mitochondrial Trans-Introns Is Required for Splicing." PLoS ONE 7 {12). https://doi.org/10.1371/journal.pone.0052644.

Fernández-López, Manuel, Estefanía Muñoz-Adelantado, Moniek Gillis, Anne Willems, and Nicolas Toro. 2005. "Dispersal and Evolution of the Sinorhizobium Meliloti Group II RmInt1 Intron in Bacteria That Interact with Plants." Molecular Biology and Evolution 22 {6): 1518–28. https://doi.org/10.1093/molbev/msi144.

Fey, Julien, André Dietrich, Anne Cosset, Thierry Desprez, and Laurence Maréchal-Drouard. 1997. "Evolutionary Aspects of 'chloroplast-like' TrnN and TrnH Expression in Higher-Plant Mitochondria." Current Genetics 32 {5): 358–60. https://doi.org/10.1007/s002940050288.

Finkeldey, R., and O. Gailing. 2013. "Chloroplasts." Brenner's Encyclopedia of Genetics: Second Edition, January, 525–27. https://doi.org/10.1016/B978-0-12-374984-0.00231-X.

Fischer, Axel, Jana Dotzek, Dirk Walther, and Stephan Greiner. 2022. “Graph-Based Models of the Oenothera Mitochondrial Genome Capture the Enormous Complexity of Higher Plant Mitochondrial DNA Organization." NAR Genomics and Bioinformatics 4 {2). https://doi.org/10.1093/NARGAB/LQAC027.

Ford, Ethan, and Manuel Jr. Ares. 1994. "Synthesis of Circular RNA in Bacteria and Yeast Using RNA Cyclase Ribozymes Derived from a Group I Intron of Phage T4." Proceedings of the National Academy of Sciences of the United States of America 91 {8): 3117. https://doi.org/10.1073/PNAS.91.8.3117.

Fuchs, Philippe, Nils Rugen, Chris Carrie, Marlene Elsässer, Iris Finkemeier, Jonas Giese, Tatjana M. Hildebrandt, et al. 2020. "Single Organelle Function and Organization as Estimated from Arabidopsis Mitochondrial Proteomics." Plant Journal 101 (2): 420–41. https://doi.org/10.1111/tpj.14534.

Funk, Erik R., Alexander N. Adams, Sarah M. Spotten, Roxanne A. Van Hove, Kristina T. Whittington, Kyle G. Keepers, Cloe S. Pogoda, James C. Lendemer, Erin A. Tripp, and Nolan C. Kane. 2018. "The Complete Mitochondrial Genomes of Five Lichenized Fungi in the Genus Usnea {Ascomycota: Parmeliaceae)." Mitochondrial DNA Part B: Resources 3 {1): 305–8. https://doi.org/10.1080/23802359.2018.1445485.

Galimberti, Andrea, Fabrizio De Mattia, Alessia Losa, Ilaria Bruni, Silvia Federici, Maurizio Casiraghi, Stefano Martellos, and Massimo Labra. 2013. "DNA Barcoding as a New Tool for Food Traceability." Food Research International 50 {1): 55–63. https://doi.org/10.1016/J.FOODRES.2012.09.036.

Gao, Long, Xue Guo, Xue-Qiong Liu, Li Zhang, Jilei Huang, Li Tan, Zhen Lin, Shingo Nagawa, and Dan-Yang Wang. 2018. “Changes in Mitochondrial DNA Levels during Early Embryogenesis in Torenia Fournieri and Arabidopsis Thaliana." The Plant Journal 95 {5): 785–95. https://doi.org/10.1111/TPJ.13987.

Garg, Sriram G., and William F. Martin. 2016. "Mitochondria, the Cell Cycle, and the Origin of Sex via a Syncytial Eukaryote Common Ancestor." Genome Biology and Evolution 8 {6): 1950. https://doi.org/10.1093/GBE/EVW136.

Gelvin, SB, and LL Habeck. 1990. "Vir Genes Influence Conjugal Transfer of the Ti Plasmid of Agrobacterium Tumefaciens.” Journal of Bacteriology 172 {3): 1600–1608. https://doi.org/10.1128/JB.172.3.1600-1608.1990.

Gerke, Philipp, Peter Szövenyi, Anna Neubauer, Henning Lenz, Bernard Gutmann, Rose McDowell, Ian Small, Mareike Schallenberg-Rüdinger, and Volker Knoop. 2020. “Towards a Plant Model for Enigmatic U-to-C RNA Editing: The Organelle Genomes, Transcriptomes, Editomes and Candidate RNA Editing Factors in the Hornwort Anthoceros Agrestis." New Phytologist 225 {5): 1974–92. https://doi.org/10.1111/nph.16297.

Giovannoni, Stephen J., H. James Tripp, Scott Givan, Mircea Podar, Kevin L. Vergin, Damon Baptista, Lisa Bibbs, et al. 2005. “Genome Streamlining in a Cosmopolitan Oceanic Bacterium.” Science 309 (5738): 1242–45. https://doi.org/10.1126/science.1114057.

Goremykin, V. V., and F. H. Hellwig. 2005. "Evidence for the Most Basal Split in Land Plants Dividing Bryophyte and Tracheophyte Lineages." Plant Systematics and Evolution 2005 254:1 254 {1): 93–103. https://doi.org/10.1007/S00606-005-0337-1.

Goremykin, Vadim V., Francesco Salamini, Riccardo Velasco, and Roberto Viola. 2009. “Mitochondrial DNA of Vitis Vinifera and the Issue of Rampant Horizontal Gene Transfer." Molecular Biology and Evolution 26 {1): 99–110. https://doi.org/10.1093/molbev/msn226.

Goryunov, Denis V., Svetlana V. Goryunova, Oxana I. Kuznetsova, Maria D. Logacheva, Irina A. Milyutina, Alina V. Fedorova, Michael S. Ignatov, and Aleksey V. Troitsky. 2018. "Complete Mitochondrial Genome Sequence of the 'Copper Moss' Mielichhoferia Elongata Reveals Independent Nad7 Gene Functionality Loss." PeerJ 2018 {2). https://doi.org/10.7717/peerj.4350.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, et al. 2011. “Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome." Nature Biotechnology 29 {7): 644–52. https://doi.org/10.1038/NBT.1883.

Gray, Michael W. 1992. "The Endosymbiont Hypothesis Revisited." International Review of Cytology 141 (C): 233–357. https://doi.org/10.1016/S0074-7696{08)62068-9.

———. 2012. "Mitochondrial Evolution." Cold Spring Harbor Perspectives in Biology 4 {9). https://doi.org/10.1101/cshperspect.a011403.

Gray, Michael W., Gertraud Burger, Romain Derelle, Vladimfr Klimes, Michelle M. Leger, Matt Sarrasin, Cestmfr Vlcek, Andrew J. Roger, Marek Elias, and B. Franz Lang. 2020. "The Draft Nuclear Genome Sequence and Predicted Mitochondrial Proteome of Andalucia Godoyi, a Protist with the Most Gene-Rich and Bacteria-like Mitochondrial Genome." BMC Biology 18 (1): 1–35. https://doi.org/10.1186/s12915-020-0741-6.

Gregory R. Schonbaum, Walter D. Bonner, Jr., Bayard T. Storey, James T. Bahr, Specific Inhibition of the Cyanide-insensitive Respiratory Pathway in Plant Mitochondria by Hydroxamic Acids, *Plant Physiology*, Volume 47, Issue 1, January 1971, Pages 124–128, https://doi.org/10.1104/pp.47.1.124

Greiner, Stephan, and Ralph Bock. 2013. "Tuning a Ménage à Trois: Co-Evolution and Co-Adaptation of Nuclear and Organellar Genomes in Plants." BioEssays 35 {4): 354–65. https://doi.org/10.1002/bies.201200137.

Greiner, Stephan, Hieronim Golczyk, Irina Malinova, Tommaso Pellizzer, Ralph Bock, Thomas Börner, and Reinhold G. Herrmann. 2020. "Chloroplast Nucleoids Are Highly Dynamic in Ploidy, Number, and Structure during Angiosperm Leaf Development." Plant Journal 102 {4): 730–46. https://doi.org/10.1111/tpj.14658.

Greiner, Stephan, Johanna Sobanski, and Ralph Bock. 2015. "Why Are Most Organelle Genomes Transmitted Maternally?" BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology 37 (1): 80–94. https://doi.org/10.1002/BIES.201400110.

Grewe, Felix, Stefan Herres, Prisca Viehöver, Monika Polsakiewicz, Bernd Weisshaar, and Volker Knoop. 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 {7): 2890–2902. https://doi.org/10.1093/nar/gkq1227.

Grewe, Felix, Prisca Viehoever, Bernd Weisshaar, and Volker Knoop. 2009. "A Trans-Splicing Group I Intron and TRNA-Hyperediting in the Mitochondrial Genome of the Lycophyte Isoetes Engelmannii." Nucleic Acids Research 37 {15): 5093–5104. https://doi.org/10.1093/nar/gkp532.

Grewe, Felix, Andan Zhu, and Jeffrey P. Mower. 2016. "Loss of a Trans-Splicing Nad1 Intron from Geraniaceae and Transfer of the Maturase Gene Matr to the Nucleus in Pelargonium." Genome Biology and Evolution 8 {10): 3193–3201. https://doi.org/10.1093/gbe/evw233.

Grosser, K, P Ramasamy, AD Amirabad, MH Schulz, G Gasparoni, M Simon, and M Schrallhammer. 2018. "More than the 'Killer Trait': Infection with the Bacterial Endosymbiont Caedibacter Taeniospiralis Causes Transcriptomic Modulation in Paramecium Host." Genome Biology and Evolution 10 {2): 646–56. https://doi.org/10.1093/GBE/EVY024.

Groth-Malonek, Milena, Dagmar Pruchner, Felix Grewe, and Volker Knoop. 2005. "Ancestors of Trans-Splicing Mitochondrial Introns Support Serial Sister Group Relationships of Hornworts and Mosses with Vascular Plants." Molecular Biology and Evolution 22 {1): 117–25. https://doi.org/10.1093/molbev/msh259.

Groth-Malonek, Milena, Ute Wahrmund, Monika Polsakiewicz, and Volker Knoop. 2007. "Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial Nad7 Gene Supports Haplomitrium as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae." Molecular Biology and Evolution 24 {4): 1068–74. https://doi.org/10.1093/molbev/msm026.

Grusz, Amanda L., Carl J. Rothfels, and Eric Schuettpelz. 2016. "Transcriptome Sequencing Reveals Genome-Wide Variation in Molecular Evolutionary Rate among Ferns." BMC Genomics 17 {1). https://doi.org/10.1186/S12864-016-3034-2.

Gualberto, Jose M., Daria Mileshina, Clementine Wallet, Adnan Khan Niazi, Frederique Weber-Lotfi, and Andre Dietrich. 2014. "The Plant Mitochondrial Genome: Dynamics and Maintenance." Biochimie. Elsevier. https://doi.org/10.1016/j.biochi.2013.09.016.

Gualberto, Jose M., Monique Le Ret, Barbara Beator, and Kristina Kühn. 2015. "The RAD52-like Protein ODB1 Is Required for the Efficient Excision of Two Mitochondrial Introns Spliced via First-Step Hydrolysis." Nucleic Acids Research 43 {13): 6500–6510. https://doi.org/10.1093/nar/gkv540.

Guillory, Wilson X., Anastasiia Onyshchenko, Elizabeth C. Ruck, Matthew Parks, Teofil Nakov, Norman J. Wickett, and Andrew J. Alverson. 2018. "Recurrent Loss, Horizontal Transfer, and the Obscure Origins of Mitochondrial Introns in Diatoms {Bacillariophyta)." Genome Biology and Evolution 10 {6): 1504–15. https://doi.org/10.1093/gbe/evy103.

Guiry, D., M. 2010. "AlgaeBase. World-Wide Electronic Publication, National University of Ireland, Galway." Http://Www.Algaebase.Org/. https://ci.nii.ac.jp/naid/10028197296.

Guo, Wenhu, Felix Grewe, Weishu Fan, Gregory J. Young, Volker Knoop, Jeffrey D. Palmer, and Jeffrey P. Mower. 2016. "Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution." Molecular Biology and Evolution 33 {6): 1448–60. https://doi.org/10.1093/molbev/msw024.

Guo, Wenhu, and Jeffrey P. Mower. 2013. "Evolution of Plant Mitochondrial Intron-Encoded Maturases: Frequent Lineage-Specific Loss and Recurrent Intracellular Transfer to the Nucleus." Journal of Molecular Evolution 77 {1–2): 43–54. https://doi.org/10.1007/s00239-013-9579-7.

Guo, Wenhu, Andan Zhu, Weishu Fan, Robert P. Adams, and Jeffrey P. Mower. 2020. "Extensive Shifts from Cis-to Trans-Splicing of Gymnosperm Mitochondrial Introns." Molecular Biology and Evolution 37 {6): 1615–20. https://doi.org/10.1093/molbev/msaa029.

Guo, Wenhu, Andan Zhu, Weishu Fan, and Jeffrey P. Mower. 2017. "Complete Mitochondrial Genomes from the Ferns Ophioglossum Californicum and Psilotum Nudum Are Highly Repetitive with the Largest Organellar Introns." New Phytologist 213 {1): 391–403. https://doi.org/10.1111/nph.14135.

Gurdon, Csanad, Zora Svab, Yaping Feng, Dibyendu Kumar, and Pal Maliga. 2016. "Cell-to-Cell Movement of Mitochondria in Plants." Proceedings of the National Academy of Sciences 113 {12): 3395–3400. https://doi.org/10.1073/PNAS.1518644113.

Hammani, Kamel, and Philippe Giege. 2014. "RNA Metabolism in Plant Mitochondria." Trends in Plant Science. Elsevier Ltd. https://doi.org/10.1016/j.tplants.2013.12.008.

Hanscom, Terrence, and Mitch McVey. 2020. "Regulation of Error-Prone DNA Double-Strand Break Repair and Its Impact on Genome Evolution." Cells. NLM {Medline). https://doi.org/10.3390/cells9071657.

Hanson, Maureen R., and Patricia L. Conklin. 2020. "Stromules, Functional Extensions of Plastids within the Plant Cell." Current Opinion in Plant Biology 58 {December): 25–32. https://doi.org/10.1016/J.PBI.2020.10.005.

Harris, Brogan J., James W. Clark, Dominik Schrempf, Gergely J. Szöllosi, Philip C. J. Donoghue, Alistair M. Hetherington, and Tom A. Williams. 2022. "Divergent Evolutionary Trajectories of Bryophytes and Tracheophytes from a Complex Common Ancestor of Land Plants." Nature Ecology & Evolution 2022, September, 1–10. https://doi.org/10.1038/s41559-022-01885-x.

Harris, Brogan J., C. Jill Harrison, Alistair M. Hetherington, and Tom A. Williams. 2020. "Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata." Current Biology 30 (11): 2001-2012.e2. https://doi.org/10.1016/J.CUB.2020.03.048.

Hastings, P. J., James R. Lupski, Susan M. Rosenberg, and Grzegorz Ira. 2009. "Mechanisms of Change in Gene Copy Number." Nature Reviews Genetics. Nat Rev Genet. https://doi.org/10.1038/nrg2593.

Haugen, Peik, Dawn M. Simon, and Debashish Bhattacharya. 2005. "The Natural History of Group I Introns." Trends in Genetics. Elsevier Current Trends. https://doi.org/10.1016/j.tig.2004.12.007.

Hausner, Georg, Mohamed Hafez, and David R. Edgell. 2014. "Bacterial Group i Introns: Mobile RNA Catalysts." Mobile DNA. BioMed Central Ltd. https://doi.org/10.1186/1759-8753-5-8.

Hausner, Georg, Robert Olson, Dawn Simon, Ian Johnson, Erin R. Sanders, Kenneth G. Karol, Richard M. McCourt, and Steven Zimmerly. 2006. "Origin and Evolution of the Chloroplast TrnK {MatK) Intron: A Model for Evolution of Group II Intron RNA Structures." Molecular Biology and Evolution 23 {2): 380–91. https://doi.org/10.1093/molbev/msj047.

Hecht, Julia, Felix Grewe, and Volker Knoop. 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 (1): 344–58. https://doi.org/10.1093/gbe/evr027.

Hepburn, Nancy J., Derek W. Schmidt, and Jeffrey P. Mower. 2012. "Loss of Two Introns from the Magnolia Tripetala Mitochondrial Cox2 Gene Implicates Horizontal Gene Transfer and Gene Conversion as a Novel Mechanism of Intron Loss." Molecular Biology and Evolution 29 {10): 3111–20. https://doi.org/10.1093/molbev/mss130.

Hiesel, Rudolf, Bernd Wissinger, Wolfgang Schuster, and Axel Brennicke. 1989. "RNA Editing in Plant Mitochondria." Science 246 {4937): 1632–34. https://doi.org/10.1126/science.2480644.

Holdt, Lesca M., Alexander Kohlmaier, and Daniel Teupser. 2018. "Circular RNAs as Therapeutic Agents and Targets." Frontiers in Physiology 9 {OCT): 1262. https://doi.org/10.3389/FPHYS.2018.01262/BIBTEX.

Horn, Matthias, Thomas R. Fritsche, Romesh K. Gautom, Karl-Heinz Schleifer, and Michael Wagner. 1999. "Novel Bacterial Endosymbionts of Acanthamoeba Spp. Related to the Paramecium Caudatum Symbiont Caedibacter Caryophilus." Environmental Microbiology 1 (4): 357–67. https://doi.org/10.1046/j.1462-2920.1999.00045.x.

Huang, Shaobai, Hans Peter Braun, Ryan M.R. Gawryluk, and A. Harvey Millar. 2019. "Mitochondrial Complex II of Plants: Subunit Composition, Assembly, and Function in Respiration and Signaling." The Plant Journal 98 {3): 405–17. https://doi.org/10.1111/TPJ.14227.

Huang, Xiong, Wenling Wang, Ting Gong, David Wickell, Li Yaung Kuo, Xingtan Zhang, Jialong Wen, et al. 2022. "The Flying Spider-Monkey Tree Fern Genome Provides Insights into Fern Evolution and Arborescence." Nature Plants 2022 8:5 8 {5): 500–512. https://doi.org/10.1038/s41477-022-01146-6.

Huchon, Dorothée, Amir Szitenberg, Sigal Shefer, Micha Ilan, and Tamar Feldstein. 2015. "Mitochondrial Group i and Group II Introns in the Sponge Orders Agelasida and Axinellida." BMC Evolutionary Biology 15 {1). https://doi.org/10.1186/s12862-015-0556-1.

Husnik, Filip, and John P. McCutcheon. 2017. "Functional Horizontal Gene Transfer from Bacteria to Eukaryotes." Nature Reviews Microbiology 2017 16:2 16 {2): 67–79. https://doi.org/10.1038/nrmicro.2017.137.

Ichinose, Mizuho, Chieko Sugita, Yusuke Yagi, Takahiro Nakamura, and Mamoru Sugita. 2013. "Two DYW Subclass PPR Proteins Are Involved in RNA Editing of CcmFc and Atp9 Transcripts in the Moss Physcomitrella Patens: First Complete Set of PPR Editing Factors in Plant Mitochondria." Plant and Cell Physiology 54 (11): 1907–16. https://doi.org/10.1093/PCP/PCT132.

Ichinose, Mizuho, Masato Uchida, and Mamoru Sugita. 2014. "Identification of a Pentatricopeptide Repeat RNA Editing Factor in Physcomitrella Patens Chloroplasts." FEBS Letters 588 {21): 4060–64. https://doi.org/10.1016/J.FEBSLET.2014.09.031.

Iorizzo, Massimo, Dariusz Grzebelus, Douglas Senalik, Marek Szklarczyk, David Spooner, and Philipp Simon. 2012. "Against the Traffic." Mobile Genetic Elements 2 {6): 261–66. https://doi.org/10.4161/mge.23088.

Iorizzo, Massimo, Douglas Senalik, Marek Szklarczyk, Dariusz Grzebelus, David Spooner, and Philipp Simon. 2012. "De Novo Assembly of the Carrot Mitochondrial Genome Using next Generation Sequencing of Whole Genomic DNA Provides First Evidence of DNA Transfer into an Angiosperm Plastid Genome." BMC Plant Biology 12 {1): 1–17. https://doi.org/10.1186/1471-2229-12-61.

Itchoda, N., S. Nishizawa, H. Nagano, T. Kubo, and T. Mikami. 2002. "The Sugar Beet Mitochondrial Nad4 Gene: An Intron Loss and Its Phylogenetic Implication in the Caryophyllales." Theoretical and Applied Genetics 104 {2–3): 209–13. https://doi.org/10.1007/s001220100744.

Jackman, Shaun D., Lauren Coombe, René L. Warren, Heather Kirk, Eva Trinh, Tina MacLeod, Stephen Pleasance, et al. 2020. "Complete Mitochondrial Genome of a Gymnosperm, Sitka Spruce {Picea Sitchensis), Indicates a Complex Physical Structure." Genome Biology and Evolution 12 {7): 1174–79. https://doi.org/10.1093/GBE/EVAA108.

Janicka, Sabina, Kristina Kühn, Monique Le Ret, Geraldine Bonnard, Patrice Imbault, Halina Augustyniak, and José M. Gualberto. 2012. “A RAD52-like Single-Stranded DNA Binding Protein Affects Mitochondrial DNA Repair by Recombination." Plant Journal 72 {3): 423–35. https://doi.org/10.1111/j.1365-313X.2012.05097.x.

Jansen, Robert K., and Tracey A. Ruhlman. 2012. "Plastid Genomes of Seed Plants." In , 103–26. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2920-9_5.

Johansen, Steinar Daae, Åse Emblem, Steinar Daae Johansen, and Åse Emblem. 2020. “Mitochondrial Group I Introns in Hexacorals Are Regulatory Genetic Elements." Advances in the Studies of the Benthic Zone, March. https://doi.org/10.5772/INTECHOPEN.91465.

Joyce, Paul B.M., and Michael W. Gray. 1989. "Chloroplast-like Transfer RNA Genes Expressed in Wheat Mitochondria." Nucleic Acids Research 17 {14): 5461–76. https://doi.org/10.1093/nar/17.14.5461.

Kamikawa, Ryoma, Isao Masuda, Mikihide Demura, Kenichi Oyama, Sadaaki Yoshimatsu, Masanobu Kawachi, and Yoshihiko Sako. 2009. “Mitochondrial Group II Introns in the Raphidophycean Flagellate Chattonella Spp. Suggest a Diatom-to-Chattonella Lateral Group II Intron Transfer." Protist 160 {3): 364–75. https://doi.org/10.1016/j.protis.2009.02.003.

Karol, Kenneth G, Kathiravetpillai Arumuganathan, Jeffrey L Boore, Aaron M Duffy, Karin DE Everett, John D Hall, S Hansen, et al. 2010. "Complete Plastome Sequences of Equisetum Arvense and Isoetes Flaccida: Implications for Phylogeny and Plastid Genome Evolution of Early Land Plant Lineages." BMC Evolutionary Biology 10 {1): 321. https://doi.org/10.1186/1471-2148-10-321.

Kawai, Hiroko, Takeshi Kanegae, Steen Christensen, Tomohiro Kiyosue, Yoshikatsu Sato, Takato Imaizumi, Akeo Kadota, and Masamitsu Wada. 2003. “Responses of Ferns to Red Light Are Mediated by an Unconventional Photoreceptor." Nature 421 {6920): 287–90. https://doi.org/10.1038/nature01310.

Kelchner, Scot A. 2002. "Group II Introns as Phylogenetic Tools: Structure, Function, and Evolutionary Constraints.” American Journal of Botany. Botanical Society of America Inc. https://doi.org/10.3732/ajb.89.10.1651.

Keren, Ido, Ayenachew Bezawork-Geleta, Max Kolton, Inbar Maayan, Eduard Belausov, Maggie Levy, Anahit Mett, David Gidoni, Felix Shaya, and Oren Ostersetzer-Biran. 2009. "AtnMat2, a Nuclear-Encoded Maturase Required for Splicing of Group-II Introns in Arabidopsis Mitochondria." RNA 15 (12): 2299–2311. https://doi.org/10.1261/rna.1776409.

Keren, Ido, Liat Tal, Catherine C. Des Francs-Small, Wagner L. Arauoj, Sofia Shevtsov, Felix Shaya, Alisdair R. Fernie, Ian Small, and Oren Ostersetzer-Biran. 2012. "NMAT1, a Nuclear-Encoded Maturase Involved in the Trans-Splicing of Nad1 Intron 1, Is Essential for Mitochondrial Complex i Assembly and Function." Plant Journal 71 {3): 413–26. https://doi.org/10.1111/j.1365-313X.2012.04998.x.

Khandelwal, Shadra. 1990. "Chromosome Evolution in the Genus Ophioglossum L." Botanical Journal of the Linnean Society 102 (3): 205–17. https://doi.org/10.1111/J.1095-8339.1990.TB01876.X.

Kim, Sook He. 2007. "Cultural Perspectives and Current Consumption Changes of Cooked Rice in Korean Diet." Nutrition Research and Practice 1 {1): 8. https://doi.org/10.4162/NRP.2007.1.1.8.

Kitazaki, Kazuyoshi, Tomohiko Kubo, Hiroyo Kagami, Takuma Matsumoto, Asami Fujita, Hiroaki Matsuhira, Muneyuki Matsunaga, and Tetsuo Mikami. 2011. "A Horizontally Transferred TRNACys Gene in the Sugar Beet Mitochondrial Genome: Evidence That the Gene Is Present in Diverse Angiosperms and Its Transcript Is Aminoacylated." Plant Journal 68 {2): 262–72. https://doi.org/10.1111/j.1365-313X.2011.04684.x.

Klein, Joanna R., Yuqing Chen, Dawn A. Manias, Jin Zhuo, Liang Zhou, Craig L. Peebles, and Gary M. Dunny. 2004. "A Conjugation-Based System for Genetic Analysis of Group II Intron Splicing in Lactococcus Lactis." Journal of Bacteriology 186 {7): 1991–98. https://doi.org/10.1128/JB.186.7.1991-1998.2004.

Klein, M., Eckert-Ossenkopp, U., Schmiedeberg, I., Brandt, P., Unseld, M., Brennicke, A. and Schuster, W. {1994), Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. The Plant Journal, 6: 447-455. https://doi.org/10.1046/j.1365-313X.1994.06030447.x

Knie, Nils, Simon Fischer, Felix Grewe, Monika Polsakiewicz, and Volker Knoop. 2015. "Horsetails Are the Sister Group to All Other Monilophytes and Marattiales Are Sister to Leptosporangiate Ferns." Molecular Phylogenetics and Evolution 90 {September): 140–49. https://doi.org/10.1016/J.YMPEV.2015.05.008.

Knie, Nils, Felix Grewe, Simon Fischer, and Volker Knoop. 2016. "Reverse U-to-C Editing Exceeds C-to-U RNA Editing in Some Ferns – a Monilophyte-Wide Comparison of Chloroplast and Mitochondrial RNA Editing Suggests Independent Evolution of the Two Processes in Both Organelles." BMC Evolutionary Biology 16 {1): 134. https://doi.org/10.1186/s12862-016-0707-z.

Knie, Nils, Felix Grewe, and Volker Knoop. 2016. "Monilophyte Mitochondrial Rps1 Genes Carry a Unique Group II Intron That Likely Originated from an Ancient Paralog in Rpl2." RNA 22 {9): 1338–48. https://doi.org/10.1261/rna.056572.116.

Knie, Nils, Monika Polsakiewicz, and Volker Knoop. 2015. "Horizontal Gene Transfer of Chlamydial-like TRNA Genes into Early Vascular Plant Mitochondria." Molecular Biology and Evolution 32 {3): 629–34. https://doi.org/10.1093/molbev/msu324.

Knoop, Volker. 2004. "The Mitochondrial DNA of Land Plants: Peculiarities in Phylogenetic Perspective." Current Genetics. Springer. https://doi.org/10.1007/s00294-004-0522-8.

———. 2010. "<p>Looking for Sense in the Nonsense: A Short Review of Non-Coding Organellar DNA Elucidatingn the Phylogeny of Bryophytes</Strong></P>." Bryophyte Diversity and Evolution 31 {1): 51. https://doi.org/10.11646/bde.31.1.10.

———. 2011. "When You Can't Trust the DNA: RNA Editing Changes Transcript Sequences." Cellular and Molecular Life Sciences. Cell Mol Life Sci. https://doi.org/10.1007/s00018-010-0538-9.

———. 2012. "Seed Plant Mitochondrial Genomes: Complexity Evolving." In , 175–200. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-2920-9_8.

———. 2013. "Plant Mitochondrial Genome Peculiarities Evolving in the Earliest Vascular Plant Lineages." Journal of Systematics and Evolution. Wiley-Liss Inc. https://doi.org/10.1111/j.1759-6831.2012.00228.x.

Knoop, Volker, Sebastian Kloska, and Axel Brennicke. 1994. "On the Identification of Group II Introns in Nucleotide Sequence Data." Journal of Molecular Biology 242 {4): 389–96. https://doi.org/10.1006/jmbi.1994.1589.

Knoop, Volker, Yin-Long Qiu, and Koichi Yoshinaga. 2004. "Molecular Phylogeny of Bryophytes and Peculiarities of Their Chloroplast and Mitochondrial DNAs." In New Frontiers in Bryology, 1–16. Springer Netherlands. https://doi.org/10.1007/978-0-306-48568-8_1.

Knoop, Volker, Michael Unseld, Joachim Marienfeld, Petra Brandt, Sabine Snkel, Henriette Ullrich, and Axel Brennicke. 1996. "Copia-, Gypsy- and Line-like Retrotransposon Fragments in the Mitochondrial Genome of Arabidopsis Thaliana." Genetics 142 {2): 579–85. https://doi.org/10.1093/genetics/142.2.579.

Köhler, Daniela, Stephanie Schmidt-Gattung, and Stefan Binder. 2010. "The DEAD-Box Protein PMH2 Is Required for Efficient Group II Intron Splicing in Mitochondria of Arabidopsis Thaliana." Plant Molecular Biology 72 (4): 459–67. https://doi.org/10.1007/s11103-009-9584-9.

Kolodner, R., and K. K. Tewari. 1979. "Inverted Repeats in Chloroplast DNA from Higher Plants." Proceedings of the National Academy of Sciences 76 {1): 41–45. https://doi.org/10.1073/pnas.76.1.41.

Koprivova, Anna, Catherine Colas Des Francs-Small, Grant Calder, Sam T. Mugford, Sandra Tanz, Bok Rye Lee, Bernd Zechmann, Ian Small, and Stanislav Kopriva. 2010. "Identification of a Pentatricopeptide Repeat Protein Implicated in Splicing of Intron 1 of Mitochondrial Nad7 Transcripts." The Journal of Biological Chemistry 285 (42): 32192–99. https://doi.org/10.1074/JBC.M110.147603.

Kowald, Axel, and Tom B. L. Kirkwood. 2011. "Evolution of the Mitochondrial Fusion–Fission Cycle and Its Role in Aging." Proceedings of the National Academy of Sciences 108 {25): 10237–42. https://doi.org/10.1073/PNAS.1101604108.

Kozik, Alexander, Beth A. Rowan, Dean Lavelle, Lidija Berke, M. Eric Schranz, Richard W. Michelmore, and Alan C. Christensen. 2019. "The Alternative Reality of Plant Mitochondrial DNA: One Ring Does Not Rule Them All." PLOS Genetics 15 {8): e1008373. https://doi.org/10.1371/JOURNAL.PGEN.1008373.

Krasich, Rachel, and William C. Copeland. 2017. "DNA Polymerases in the Mitochondria: A Critical Review of the Evidence." Frontiers in Bioscience - Landmark. Frontiers in Bioscience. https://doi.org/10.2741/4510.

Kroeger, Tiffany S., Kenneth P. Watkins, Giulia Friso, Klaas J. Van Wijk, and Alice Barkan. 2009. "A Plant-Specific RNA-Binding Domain Revealed through Analysis of Chloroplast Group II Intron Splicing." Proceedings of the National Academy of Sciences of the United States of America 106 {11): 4537–42. https://doi.org/10.1073/pnas.0812503106.

Kubo, Tomohiko, Satsuki Nishizawa, Akira Sugawara, Noriko Itchoda, Amy Estiati, and Tetsuo Mikami. 2000. "The Complete Nucleotide Sequence of the Mitochondrial Genome of Sugar Beet {Beta Vulgaris L.) Reveals a Novel Gene for TRNA{Cys){GCA)." Nucleic Acids Research 28 {13): 2571–76. https://doi.org/10.1093/nar/28.13.2571.

Kucej, Martin, Blanka Kucejova, Ramiah Subramanian, Xin Jie Chen, and Ronald A. Butow. 2008. "Mitochondrial Nucleoids Undergo Remodeling in Response to Metabolic Cues." Journal of Cell Science 121 {11): 1861–68. https://doi.org/10.1242/jcs.028605.

Kudla, J., F. Albertazzi, D. Blazevic, M. Hermann, and R. Bock. 2002. “Loss of the Mitochondrial Cox2 Intron 1 in a Family of Monocotyledonous Plants and Utilization of Mitochondrial Intron Sequences for the Construction of a Nuclear Intron." Molecular Genetics and Genomics 267 {2): 223–30. https://doi.org/10.1007/s00438-002-0657-6.

Kühn, Kristina, Chris Carrie, Estelle Giraud, Yan Wang, Etienne H. Meyer, Reena Narsai, Catherine Colas Des Francs-Small, Botao Zhang, Monika W. Murcha, and James Whelan. 2011. “The RCC1 Family Protein RUG3 Is Required for Splicing of Nad2 and Complex i Biogenesis in Mitochondria of Arabidopsis Thaliana." Plant Journal 67 {6): 1067–80. https://doi.org/10.1111/j.1365-313X.2011.04658.x.

Kumar, Raman, Laurence Maréchal-Drouard, Kazuhito Akama, and Ian Small. 1996. "Striking Differences in Mitochondrial TRNA Import between Different Plant Species." Molecular and General Genetics 252 {4): 404–11. https://doi.org/10.1007/BF02173005.

Lambowitz, Alan M., and Steven Zimmerly. 2004. “Mobile Group II Introns.” Annual Review of Genetics. Annual Reviews. https://doi.org/10.1146/annurev.genet.38.072902.091600.

———. 2011. "Group II Introns: Mobile Ribozymes That Invade DNA." Cold Spring Harbor Perspectives in Biology 3 (8): 1–19. https://doi.org/10.1101/cshperspect.a003616.

Lambowitz, Alan M, and Marlene Belfort. 2015. “Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution." In Mobile DNA III, 3:1209–36. American Society of Microbiology. https://doi.org/10.1128/microbiolspec.mdna3-0050-2014.

Lane, Nick, and William Martin. 2010. "The Energetics of Genome Complexity." Nature 2010 467:7318 467 {7318): 929–34. https://doi.org/10.1038/nature09486.

Langhansova, Lenka, Karolina Pumprova, Daniel Haisel, Libor Ekrt, Antonio Pavicic, Marketa Zajfckova, Tomas Vanek, and Marcela Dvorakova. 2021. "European Ferns as Rich Sources of Antioxidants in the Human Diet." Food Chemistry 356 {September): 129637. https://doi.org/10.1016/J.FOODCHEM.2021.129637.

LaRoche-Johnston, Felix, Rafia Bosan, and Benoit Cousineau. 2021. "Group II Introns Generate Functional Chimeric Relaxase Enzymes with Modified Specificities through Exon Shuffling at Both the RNA and DNA Level." Molecular Biology and Evolution 38 {3): 1075–89. https://doi.org/10.1093/molbev/msaa275.

LaRoche-Johnston, Félix, Caroline Monat, Samy Coulombe, and Benoit Cousineau. 2018. “Bacterial Group II Introns Generate Genetic Diversity by Circularization and Trans-Splicing from a Population of Intron-Invaded MRNAs." PLoS Genetics 14 {11). https://doi.org/10.1371/journal.pgen.1007792.

Larsson, John, Johan A.A. Nylander, and Birgitta Bergman. 2011. "Genome Fluctuations in Cyanobacteria Reflect Evolutionary, Developmental and Adaptive Traits." BMC Evolutionary Biology 11 (1): 1–21. https://doi.org/10.1186/1471-2148-11-187.

Leclercq, Sebastien, and Richard Cordaux. 2012. "Selection-Driven Extinction Dynamics for Group II Introns in Enterobacteriales." PLoS ONE 7 {12). https://doi.org/10.1371/journal.pone.0052268.

Leebens-Mack, James H., Michael S. Barker, Eric J. Carpenter, Michael K. Deyholos, Matthew A. Gitzendanner, Sean W. Graham, Ivo Grosse, et al. 2019. "One Thousand Plant Transcriptomes and the Phylogenomics of Green Plants." Nature 2019 574:7780 574 {7780): 679–85. https://doi.org/10.1038/s41586-019-1693-2.

Lemieux, Claude, Christian Otis, and Monique Turmel. 2000. "Ancestral Chloroplast Genome in Mesostigma Viride Reveals an Early Branch of Green Plant Evolution." Nature 403 {6770): 649–52. https://doi.org/10.1038/35001059.

———. 2016. "Comparative Chloroplast Genome Analyses of Streptophyte Green Algae Uncover Major Structural Alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae." Frontiers in Plant Science 7 {MAY2016): 697. https://doi.org/10.3389/fpls.2016.00697.

Lesage, P., and A. L. Todeschini. 2005. "Happy Together: The Life and Times of Ty Retrotransposons and Their Hosts." Cytogenetic and Genome Research. Karger Publishers. https://doi.org/10.1159/000084940.

Lesch, Elena, Maximilian T Schilling, Sarah Brenner, Yingying Yang, Oliver J Gruss, Volker Knoop, and Mareike Schallenberg-Rüdinger. 2022. "Plant Mitochondrial RNA Editing Factors Can Perform Targeted C-to-U Editing of Nuclear Transcripts in Human Cells." Nucleic Acids Research 50 {17): 9966–83. https://doi.org/10.1093/nar/gkac752.

Li-Pook-Than, Jennifer, Catherine Carrillo, and Linda Bonen. 2004. "Variation in Mitochondrial Transcript Profiles of Protein-Coding Genes during Early Germination and Seedling Development in Wheat." Current Genetics 46 {6): 374–80. https://doi.org/10.1007/s00294-004-0544-2.

Li, Fay-Wei, Paul Brouwer, Lorenzo Carretero-Paulet, Shifeng Cheng, Jan de Vries, Pierre-Marc Delaux, Ariana Eily, et al. 2018. "Fern Genomes Elucidate Land Plant Evolution and Cyanobacterial Symbioses." Nature Plants 2018 4:7 4 {7): 460–72. https://doi.org/10.1038/s41477-018-0188-8.

Li, Fay Wei, Paul Brouwer, Lorenzo Carretero-Paulet, Shifeng Cheng, Jan De Vries, Pierre Marc Delaux, Ariana Eily, et al. 2018. "Fern Genomes Elucidate Land Plant Evolution and Cyanobacterial Symbioses." Nature Plants 4 {7): 460–72. https://doi.org/10.1038/s41477-018-0188-8.

Li, Fay Wei, Tomoaki Nishiyama, Manuel Waller, Eftychios Frangedakis, Jean Keller, Zheng Li, Noe Fernandez-Pozo, et al. 2020. "Anthoceros Genomes Illuminate the Origin of Land Plants and the Unique Biology of Hornworts." Nature Plants 6 {3): 259–72. https://doi.org/10.1038/s41477-020-0618-2.

Li, Fay Wei, Juan Carlos Villarreal, Steven Kelly, Carl J. Rothfels, Michael Melkonian, Eftychios Frangedakis, Markus Ruhsam, et al. 2014. "Horizontal Transfer of an Adaptive Chimeric Photoreceptor from Bryophytes to Ferns." Proceedings of the National Academy of Sciences of the United States of America 111 {18): 6672–77. https://doi.org/10.1073/pnas.1319929111.

Li, Libo, Bin Wang, Yang Liu, and Yin Long Qiu. 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 {6): 665–78. https://doi.org/10.1007/s00239-009-9240-7.

Li, Shao-Jian, Muhammad Z Ahmed, Ning Lv, Pei-Qiong Shi, Xing-Min Wang, Ji-Lei Huang, and Bao-Li Qiu. 2017. "Plantmediated Horizontal Transmission of Wolbachia between Whiteflies." The ISME Journal 11 (4): 1019–28. https://doi.org/10.1038/ismej.2016.164.

Li, Xiwen, Yang Yang, Robert J. Henry, Maurizio Rossetto, Yitao Wang, and Shilin Chen. 2015. "Plant DNA Barcoding: From Gene to Genome." Biological Reviews of the Cambridge Philosophical Society 90 (1): 157–66. https://doi.org/10.1111/BRV.12104.

Li, Yingfu, and Ronald R. Breaker. 1999. "Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group." Journal of the American Chemical Society 121 {23): 5364–72. https://doi.org/10.1021/JA990592P.

Lieber, Michael R. 2010. “The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway." Annual Review of Biochemistry 79 {July): 181. https://doi.org/10.1146/ANNUREV.BIOCHEM.052308.093131.

Lin, Ching-Ping, Jen-Pan Huang, Chung-Shien Wu, Chih-Yao Hsu, and Shu-Miaw Chaw. 2010. "Comparative Chloroplast Genomics Reveals the Evolution of Pinaceae Genera and Subfamilies." Genome Biology and Evolution 2 {1): 504–17. https://doi.org/10.1093/gbe/evq036.

Lippok, Bernadette, Axel Brennicke, and Bernd Wissinger. 1994. "Differential RNA Editing in Closely Related Introns in Oenothera Mitochondria." Molecular & General Genetics : MGG 243 {1): 39–46. https://doi.org/10.1007/BF00283874.

Liu, Guozheng, Dandan Cao, Shuangshuang Li, Aiguo Su, Jianing Geng, Corrinne E. Grover, Songnian Hu, and Jinping Hua. 2013. “The Complete Mitochondrial Genome of Gossypium Hirsutum and Evolutionary Analysis of Higher Plant Mitochondrial Genomes." PLoS ONE 8 {8). https://doi.org/10.1371/journal.pone.0069476.

Liu, Li, Charles C. Query, and Maria M. Konarska. 2007. "Opposing Classes of Prp8 Alleles Modulate the Transition between the Catalytic Steps of Pre-MRNA Splicing." Nature Structural & Molecular Biology 2007 14:6 14 {6): 519–26. https://doi.org/10.1038/nsmb1240.

Liu, Pengfei, Ayelet Erez, Sandesh C.Sreenath Nagamani, Shweta U. Dhar, Katarzyna E. Kotodziejska, Avinash V. Dharmadhikari, M. Lance Cooper, et al. 2011. “Chromosome Catastrophes Involve Replication Mechanisms Generating Complex Genomic Rearrangements." Cell 146 {6): 889–903. https://doi.org/10.1016/j.cell.2011.07.042.

Liu, Yang, Rafael Medina, and Bernard Goffinet. 2014. "350 My of Mitochondrial Genome Stasis in Mosses, an Early Land Plant Lineage." Molecular Biology and Evolution 31 {10): 2586–91. https://doi.org/10.1093/molbev/msu199.

Liu, Yang, Bin Wang, Peng Cui, Libo Li, Jia Yu Xue, Jun Yu, and Yin Long Qiu. 2012. "The Mitochondrial Genome of the Lycophyte Huperzia Squarrosa: The Most Archaic Form in Vascular Plants." PloS One 7 {4): 35168. https://doi.org/10.1371/journal.pone.0035168.

Liu, Yang, Jia Yu Xue, Bin Wang, Libo Li, and Yin Long Qiu. 2011. "The Mitochondrial Genomes of the Early Land Plants Treubia Lacunosa and Anomodon Rugelii: Dynamic and Conservative Evolution." PLoS ONE 6 {10). https://doi.org/10.1371/journal.pone.0025836.

Lu, Hengyun, Francesca Giordano, and Zemin Ning. 2016. "Oxford Nanopore MinION Sequencing and Genome Assembly." Genomics, Proteomics & Bioinformatics 14 {5): 265–79. https://doi.org/10.1016/J.GPB.2016.05.004.

Lurin, Claire, Charles Andres, Sebastien Aubourg, Mohammed Bellaoui, Frederique Bitton, Clemence Bruyère, Michel Caboche, et al. 2004. “Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis." Plant Cell 16 {8): 2089–2103. https://doi.org/10.1105/tpc.104.022236.

Ma, Peng Fei, Yu Xiao Zhang, Zhen Hua Guo, and De Zhu Li. 2015. “Evidence for Horizontal Transfer of Mitochondrial DNA to the Plastid Genome in a Bamboo Genus." Scientific Reports 5 {1): 1–9. https://doi.org/10.1038/srep11608.

Mahgoub, Yasmin A., Eman Shawky, Moustafa Eldakak, Mohammed Bahey-El-Din, Fikria A. Darwish, Nadia A. El Sebakhy, and Amr El-Hawiet. 2022. "Plant DNA Barcoding and Metabolomics for Comprehensive Discrimination of German Chamomile from Its Poisonous Adulterants for Food Safety." Food Control 136 {June): 108840. https://doi.org/10.1016/J.FOODCONT.2022.108840.

Maier, Uwe G., Andrew Bozarth, Helena T. Funk, Stefan Zauner, Stefan A. Rensing, Christian Schmitz-Linneweber, Thomas Börner, and Michael Tillich. 2008. "Complex Chloroplast RNA Metabolism: Just

Debugging the Genetic Programme?" BMC Biology 6 {1): 1–9. https://doi.org/10.1186/1741-7007-6-36.

Maier, Uwe G., Stefan Zauner, Christian Woehle, Kathrin Bolte, Franziska Hempel, John F. Allen, and William F. Martin. 2013. "Massively Convergent Evolution for Ribosomal Protein Gene Content in Plastid and Mitochondrial Genomes." Genome Biology and Evolution 5 (12): 2318–29. https://doi.org/10.1093/gbe/evt181.

Maixner, Frank, Dmitrij Turaev, Amaury Cazenave-Gassiot, Marek Janko, Ben Krause-Kyora, Michael R. Hoopmann, Ulrike Kusebauch, et al. 2018. "The Iceman's Last Meal Consisted of Fat, Wild Meat, and Cereals." Current Biology : CB 28 {14): 2348-2355.e9. https://doi.org/10.1016/j.cub.2018.05.067.

Malek, Olaf, and Volker Knoop. 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 {12): 1599–1609. https://doi.org/10.1017/S1355838298981262.

Marchant, D. Blaine, Guang Chen, Shengguan Cai, Fei Chen, Peter Schafran, Jerry Jenkins, Shengqiang Shu, et al. 2022. "Dynamic Genome Evolution in a Model Fern." Nature Plants 2022 8:9 8 (9): 1038–51. https://doi.org/10.1038/s41477-022-01226-7.

Marchfelder, Anita, Axel Brennicke, and Stefan Binder. 1996. "RNA Editing Is Required for Efficient Excision of TRNAPhe from Precursors in Plant Mitochondria." Journal of Biological Chemistry 271 (4): 1898–1903. https://doi.org/10.1074/jbc.271.4.1898.

Maréchal-Drouard, L., A. Dietrich, H. Mireau, N. Peeters, and I. Small. 1999. "The Import of Cytosolic TRNA into Plant Mitochondria." In Mitochondrial Diseases, 317–26. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-59884-5_24.

Marechal-Drouard, Laurence, Pierre Guillemaut, Anne Cosset, Michele Arbogast, Frederique Weber, Jacques Henry Weil, and Andre Dietrich. 1990. "Transfer RNAs of Potato {Solanum Tuberosum) Mitochondria Have Different Genetic Origins." Nucleic Acids Research 18 {13): 3689–96. https://doi.org/10.1093/nar/18.13.3689.

Maréchal-Drouard, Laurence, Jacques Henry Weil, and Pierre Guillemaut. 1988. "Import of Several TRNAs from the Cytoplasm into the Mitochrondria in Bean Phaseolus Vulgaris." Nucleic Acids Research 16 {11): 4777–88. https://doi.org/10.1093/nar/16.11.4777.

Marechal, Alexandre, and Lee Zou. 2013. "DNA Damage Sensing by the ATM and ATR Kinases." Cold Spring Harbor Perspectives in Biology 5 {9): a012716. https://doi.org/10.1101/cshperspect.a012716.

Mazina, Olga M., Havva Keskin, Kritika Hanamshet, Francesca Storici, and Alexander V. Mazin. 2017. "Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair.” Molecular Cell 67 {1): 19-29.e3. https://doi.org/10.1016/j.molcel.2017.05.019.

McDevitt, Shane, Timur Rusanov, Tatiana Kent, Gurushankar Chandramouly, and Richard T. Pomerantz. 2018. "How RNA Transcripts Coordinate DNA Recombination and Repair." Nature Communications 9 {1). https://doi.org/10.1038/s41467-018-03483-7.

Michel, Francois, and Jean Luc Ferat. 1995. "Structure and Activities of Group II Introns." Annual Review of Biochemistry. Annual Reviews Inc. https://doi.org/10.1146/annurev.bi.64.070195.002251.

Michel, Francois, Umesono Kazuhiko, and Ozeki Haruo. 1989. "Comparative and Functional Anatomy of Group II Catalytic Introns - a Review.” Gene 82 (1): 5–30. https://doi.org/10.1016/0378-1119{89)90026-7.

Millar, A. Harvey, Joshua L. Heazlewood, Brian K. Kristensen, Hans Peter Braun, and Ian M. Møller. 2005. "The Plant Mitochondrial Proteome." Trends in Plant Science. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2004.12.002.

Miller-Messmer, Marie, Kristina Kuhn, Marc Bichara, Monique le Ret, Patrice Imbault, and Jose M. Gualberto. 2012. “RecA-Dependent DNA Repair Results in Increased Heteroplasmy of the Arabidopsis Mitochondrial Genome." Plant Physiology 159 {1): 211–26. https://doi.org/10.1104/pp.112.194720.

Mills, David A., Larry L. Mckay, and Gary M. Dunny. 1996. "Splicing of a Group II Intron Involved in the Conjugative Transfer of PRS01 in Lactococci." Journal of Bacteriology 178 {12): 3531–38. https://doi.org/10.1128/jb.178.12.3531-3538.1996.

Mitchell, Peter, and Jennifer Moyle. 1967. "Chemiosmotic Hypothesis of Oxidative Phosphorylation." Nature 213 {5072): 137–39. https://doi.org/10.1038/213137a0.

Miyata, Shin Ichi, Mikio Nakazono, and Atsushi Hirai. 1998. "Transcription of Plastid-Derived TRNA Genes in Rice Mitochondria." Current Genetics 34 {3): 216–20. https://doi.org/10.1007/s002940050389.

Mohr, Georg, and Alan M. Lambowitz. 2003. "Putative Proteins Related to Group II Intron Reverse Transcriptase/Maturases Are Encoded by Nuclear Genes in Higher Plants." Nucleic Acids Research. Nucleic Acids Res. https://doi.org/10.1093/nar/gkg153.

Molina-Sánchez, Maria D., Fernando M. García-Rodrfguez, and Nicolas Toro. 2016. "Functionality of In Vitro Reconstituted Group II Intron RmInt1-Derived Ribonucleoprotein Particles." Frontiers in Molecular Biosciences 0 {SEP): 58. https://doi.org/10.3389/FMOLB.2016.00058.

Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol Cell Biol. 1996 May;16{5):2164-73. doi: 10.1128/MCB.16.5.2164.

Morley, Stewart A., Niaz Ahmad, and Brent L. Nielsen. 2019. "Plant Organelle Genome Replication." Plants 8 {10): 358. https://doi.org/10.3390/plants8100358.

Morris, Jennifer L., Mark N. Puttick, James W. Clark, Dianne Edwards, Paul Kenrick, Silvia Pressel, Charles H. Wellman, Ziheng Yang, Harald Schneider, and Philip C.J. Donoghue. 2018. "The Timescale of Early Land Plant Evolution." Proceedings of the National Academy of Sciences of the United States of America 115 (10): E2274–83. https://doi.org/10.1073/pnas.1719588115.

Mower, Jeffrey P. 2020. "Variation in Protein Gene and Intron Content among Land Plant Mitogenomes." Mitochondrion. Elsevier B.V. https://doi.org/10.1016/j.mito.2020.06.002.

Mower, Jeffrey P., Daniel B. Sloan, and Andrew J. Alverson. 2012. "Plant Mitochondrial Genome Diversity: The Genomics Revolution." In Plant Genome Diversity Volume 1: Plant Genomes, Their Residents, and Their Evolutionary Dynamics, 123–44. Springer-Verlag Vienna. https://doi.org/10.1007/978-3-7091-1130-7_9.

Mower, Jeffrey P., Sasa Stefanovic, Weilong Hao, Julie S. Gummow, Kanika Jain, Dana Ahmed, and Jeffrey D. Palmer. 2010. "Horizontal Acquisition of Multiple Mitochondrial Genes from a Parasitic Plant Followed by Gene Conversion with Host Mitochondrial Genes.” BMC Biology 8 (1): 1–16. https://doi.org/10.1186/1741-7007-8-150.

Mower, Jeffrey P., and Trisha L. Vickrey. 2018. "Structural Diversity Among Plastid Genomes of Land Plants." In Advances in Botanical Research, 85:263–92. Academic Press Inc. https://doi.org/10.1016/bs.abr.2017.11.013.

Mukhopadhyay, Jigeesha, and Georg Hausner. 2021. "Organellar Introns in Fungi, Algae, and Plants." Cells 10 {8). https://doi.org/10.3390/CELLS10082001.

Nagy-Deri, H., E. R. Peli, K. Georgieva, and Z. Tuba. 2011. “Changes in Chloroplast Morphology of Different Parenchyma Cells in Leaves of Haberlea Rhodopensis Friv. during Desiccation and Following Rehydration." Photosynthetica 49 {1): 119–26. https://doi.org/10.1007/s11099-011-0020-4.

Nekrasov, Eduard V., and Vasily I. Svetashev. 2021. "Edible Far Eastern Ferns as a Dietary Source of Long-Chain Polyunsaturated Fatty Acids." Foods 10 {6). https://doi.org/10.3390/FOODS10061220/S1.

Nielsen, Cydney B., Brad Friedman, Bruce Birren, Christopher B. Burge, and James E. Galagan. 2004. "Patterns of Intron Gain and Loss in Fungi." PLoS Biology 2 {12). https://doi.org/10.1371/journal.pbio.0020422.

Niklas, Karl J., and Ulrich Kutschera. 2010. "The Evolution of the Land Plant Life Cycle." New Phytologist 185 (1): 27–41. https://doi.org/10.1111/J.1469-8137.2009.03054.X.

Nishiyama, Tomoaki, Paul G. Wolf, Masanori Kugita, Robert B. Sinclair, Mamoru Sugita, Chika Sugiura, Tatsuya Wakasugi, et al. 2004. "Chloroplast Phylogeny Indicates That Bryophytes Are Monophyletic." Molecular Biology and Evolution 21 {10): 1813–19. https://doi.org/10.1093/MOLBEV/MSH203.

Notsu, Y., S. Masood, T. Nishikawa, N. Kubo, G. Akiduki, M. Nakazono, A. Hirai, and K. Kadowaki. 2002. "The Complete Sequence of the Rice {Oryza Sativa L.) Mitochondrial Genome: Frequent DNA Sequence Acquisition and Loss during the Evolution of Flowering Plants." Molecular Genetics and Genomics 268 {4): 434–45. https://doi.org/10.1007/s00438-002-0767-1.

Novikova, Olga, Dorie Smith, Ingrid Hahn, Arthur Beauregard, and Marlene Belfort. 2014. "Interaction between Conjugative and Retrotransposable Elements in Horizontal Gene Transfer." PLoS Genetics 10 {12). https://doi.org/10.1371/journal.pgen.1004853.

Nugent, Jacqueline M., and Jeffrey D. Palmer. 1991. "RNA-Mediated Transfer of the Gene CoxII from the Mitochondrion to the Nucleus during Flowering Plant Evolution." Cell 66 {3): 473–81. https://doi.org/10.1016/0092-8674{81)90011-8.

O'Toole, Nicholas, Mitsuru Hattori, Charles Andres, Kei Iida, Claire Lurin, Christian Schmitz-Linneweber, Mamoru Sugita, and Ian Small. 2008. "On the Expansion of the Pentatricopeptide Repeat Gene Family in Plants." Molecular Biology and Evolution 25 {6): 1120–28. https://doi.org/10.1093/molbev/msn057.

Oda, Kenji, Katsuyuki Yamato, Eiji Ohta, Yasukazu Nakamura, Miho Takemura, Naoko Nozato, Kinya Akashi, et al. 1992. "Gene Organization Deduced from the Complete Sequence of Liverwort Marchantia Polymorpha Mitochondrial DNA. A Primitive Form of Plant Mitochondrial Genome." Journal of Molecular Biology 223 (1): 1–7. https://doi.org/10.1016/0022-2836{92)90708-R.

Odahara, Masaki, Takayuki Inouye, Tomomichi Fujita, Mitsuyasu Hasebe, and Yasuhiko Sekine. 2007. “Involvement of Mitochondrial-Targeted RecA in the Repair of Mitochondrial DNA in the Moss, Physcomitrella Patens." Genes and Genetic Systems 82 {1): 43–51. https://doi.org/10.1266/ggs.82.43.

Odahara, Masaki, Haruko Kuroiwa, Tsuneyoshi Kuroiwa, and Yasuhiko Sekine. 2009. “Suppression of Repeat-Mediated Gross Mitochondrial Genome Rearrangements by RecA in the Moss Physcomitrella Patens." Plant Cell 21 {4): 1182–94. https://doi.org/10.1105/tpc.108.064709.

Odahara, Masaki, Yuichi Masuda, Mayuko Sato, Mayumi Wakazaki, Chizuru Harada, Kiminori Toyooka, and Yasuhiko Sekine. 2015. "RECG Maintains Plastid and Mitochondrial Genome Stability by Suppressing Extensive Recombination between Short Dispersed Repeats." PLoS Genetics 11 {3): 1005080. https://doi.org/10.1371/journal.pgen.1005080.

Ohyama, K, K Oda, E Ohta, and M Takemura. 1993. "Gene Organization and Evolution of Introns of a Liverwort, Marchantia Polymorpha, Mitochondrial Genome." Plant Mitochondria: With Emphasis on RNA Editing and Cytoplasmic Male Sterility Weinheim; https://agris.fao.org/agris-search/search.do?recordID=US201301782811.

Oldenburg, Delene J., and Arnold J. Bendich. 2004. "Most Chloroplast DNA of Maize Seedlings in Linear Molecules with Defined Ends and Branched Forms." Journal of Molecular Biology 335 {4): 953–70. https://doi.org/10.1016/j.jmb.2003.11.020.

———. 2015. "DNA Maintenance in Plastids and Mitochondria of Plants." Frontiers in Plant Science. Frontiers Research Foundation. https://doi.org/10.3389/fpls.2015.00883.

———. 2016. "The Linear Plastid Chromosomes of Maize: Terminal Sequences, Structures, and Implications for DNA Replication." Current Genetics 62 {2): 431–42. https://doi.org/10.1007/s00294-015-0548-0.

Oldenkott, Bastian, Matthias Burger, Anke Christiane Hein, Anja Jörg, Jennifer Senkler, Hans Peter Braun, Volker Knoop, Mizuki Takenaka, and Mareike Schallenberg-Rüdinger. 2020. “One C-to-U RNA Editing Site and Two Independently Evolved Editing Factors: Testing Reciprocal Complementation with DYW-Type PPR Proteins from the Moss Physcomitrium {Physcomitrella) Patens and the Flowering Plants Macadamia Integrifolia and Arabidopsis." Plant Cell 32 {9): 2997–3018. https://doi.org/10.1105/TPC.20.00311.

Oldenkott, Bastian, Kazuo Yamaguchi, Sumika Tsuji-Tsukinoki, Nils Knie, and Volker Knoop. 2014. "Chloroplast RNA Editing Going Extreme: More than 3400 Events of C-to-U Editing in the Chloroplast Transcriptome of the Lycophyte Selaginella Uncinata." RNA 20 {10): 1499–1506. https://doi.org/10.1261/rna.045575.114.

Oldenkott, Bastian, Yingying Yang, Elena Lesch, Volker Knoop, and Mareike Schallenberg-Rüdinger. 2019. "Plant-Type Pentatricopeptide Repeat Proteins with a DYW Domain Drive C-to-U RNA Editing in Escherichia Coli." Communications Biology 2 {1). https://doi.org/10.1038/s42003-019-0328-3.

Orgel, L. E., and F. H.C. Crick. 1980. "Selfish DNA: The Ultimate Parasite." Nature 284 {5757): 604–7. https://doi.org/10.1038/284604a0.

Orton, Lauren M., Elisabeth Fitzek, Elisabeth Fitzek, Xuehuan Feng, W. Scott Grayburn, Jeffrey P. Mower, Jeffrey P. Mower, et al. 2020. "Zygnema Circumcarinatum UTEX 1559 Chloroplast and Mitochondrial Genomes Provide Insight into Land Plant Evolution." Journal of Experimental Botany 71 {11): 3361–73. https://doi.org/10.1093/JXB/ERAA149.

Pakeman, R. J., R. H. Marrs, D. C. Howard, C. J. Barr, and R. M. Fuller. 1996. "The Bracken Problem in Great Britain: Its Present Extent and Future Changes." Applied Geography 16 {1): 65–86. https://doi.org/10.1016/0143-6228{95)00026-7.

Palmer, Jeffrey D. 1983. "Chloroplast DNA Exists in Two Orientations." Nature 301 {5895): 92–93. https://doi.org/10.1038/301092a0.

Palmer, Jeffrey D., and John M. Logsdon. 1991. “The Recent Origins of Introns.” Current Opinion in Genetics and Development 1 {4): 470–77. https://doi.org/10.1016/S0959-437X(05)80194-7.

Palmer, Jeffrey D., Bernardita Osorio, Jane Aldrich, and William F. Thompson. 1987. "Chloroplast DNA Evolution among Legumes: Loss of a Large Inverted Repeat Occurred Prior to Other Sequence Rearrangements." Current Genetics 11 {4): 275–86. https://doi.org/10.1007/BF00355401.

Palmer, Jeffrey D., and William F. Thompson. 1982. "Chloroplast DNA Rearrangements Are More Frequent When a Large Inverted Repeat Sequence Is Lost." Cell 29 {2): 537–50. https://doi.org/10.1016/0092-8674{82)90170-2.

Pan, Zhenyuan, Xuemei Ren, Hailiang Zhao, Lei Liu, Zengdong Tan, and Fazhan Qiu. 2019. "A Mitochondrial Transcription Termination Factor, ZmSmk3, Is Required for Nad1 Intron4 and Nad4 Intron1 Splicing and Kernel Development in Maize." G3 {Bethesda, Md.) 9 {8): 2677–86. https://doi.org/10.1534/G3.119.400265.

Paquin, Bruno, Marie Josee Laforest, Lise Forget, Ingeborg Roewer, Zhang Wang, Joyce Longcore, and B. Franz Lang. 1997. "The Fungal Mitochondrial Genome Project: Evolution of Fungal Mitochondrial Genomes and Their Gene Expression." Current Genetics 31 {5): 380–95. https://doi.org/10.1007/s002940050220.

Park, Hyun Seung, Murukarthick Jayakodi, Sae Hyun Lee, Jae Hyeon Jeon, Hyun Oh Lee, Jee Young Park, Byeong Cheol Moon, et al. 2020. "Mitochondrial Plastid DNA Can Cause DNA Barcoding Paradox in Plants." Scientific Reports 10 {1): 1–12. https://doi.org/10.1038/s41598-020-63233-y.

Park, Jongsun, Yongsung Kim, Gi-Hun Lee, and Chan-Ho Park. 2020. "The Complete Chloroplast Genome of Selaginella Tamariscina (Beauv.) Spring (Selaginellaceae) Isolated in Korea." Mitochondrial DNA Part B 5 {2): 1654–56. https://doi.org/10.1080/23802359.2020.1715885.

Park, Seongjun, Felix Grewe, Andan Zhu, Tracey A. Ruhlman, Jamal Sabir, Jeffrey P. Mower, and Robert K. Jansen. 2015. "Dynamic Evolution of Geranium Mitochondrial Genomes through Multiple Horizontal and Intracellular Gene Transfers." New Phytologist 208 {2): 570–83. https://doi.org/10.1111/nph.13467.

Parkinson, Christopher L., Jeffrey P. Mower, Yin Long Qiu, Andrew J. Shirk, Keming Song, Nelson D. Young, Claude W. DePamphilis, and Jeffrey D. Palmer. 2005. "Multiple Major Increases and Decreases in Mitochondrial Substitution Rates in the Plant Family Geraniaceae." BMC Evolutionary Biology 5 (December): 1–12. https://doi.org/10.1186/1471-2148-5-73.

Paszkiewicz, Gael, Jose M. Gualberto, Abdelilah Benamar, David Macherel, and David C. Logan. 2017. "Arabidopsis Seed Mitochondria Are Bioenergetically Active Immediately upon Imbibition and Specialize via Biogenesis in Preparation for Autotrophic Growth." The Plant Cell 29 {1): 109–28. https://doi.org/10.1105/TPC.16.00700.

Peebles, C. L., P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell, and H. L. Cheng. 1986. "A Self-Splicing RNA Excises an Intron Lariat." Cell 44 {2): 213–23. https://doi.org/10.1016/0092-8674{86)90755-5.

Petersen, G., A. Cuenca, I. M. M!llller, and O. Seberg. 2015. "Massive Gene Loss in Mistletoe (Viscum, Viscaceae) Mitochondria." Scientific Reports 5 {December). https://doi.org/10.1038/srep17588.

Petkovic, Sonja, and Sabine Müller. 2015. "RNA Circularization Strategies in Vivo and in Vitro." Nucleic Acids Research 43 {4): 2454. https://doi.org/10.1093/NAR/GKV045.

Pfannschmidt, Thomas. 2010. "Plastidial Retrograde Signalling - a True 'Plastid Factor' or Just Metabolite Signatures?" Trends in Plant Science 15 {8): 427–35. https://doi.org/10.1016/j.tplants.2010.05.009.

Piechotta, Michael, Emanuel Wyler, Uwe Ohler, Markus Landthaler, and Christoph Dieterich. 2017. "JACUSA: Site-Specific Identification of RNA Editing Events from Replicate Sequencing Data." BMC Bioinformatics 18 {1): 1–15. https://doi.org/10.1186/S12859-016-1432-8/FIGURES/10.

Pinson, Jerald B., and Eric Schuettpelz. 2016. "Unraveling the Origin of the Appalachian Gametophyte, Vittaria Appalachiana." American Journal of Botany 103 {4): 668–76. https://doi.org/10.3732/AJB.1500522.

Pogoda, Cloe S., Kyle G. Keepers, Sarah E. Hamsher, Joshua G. Stepanek, Nolan C. Kane, and J. Patrick Kociolek. 2019. "Comparative Analysis of the Mitochondrial Genomes of Six Newly Sequenced Diatoms Reveals Group II Introns in the Barcoding Region of Cox1." Mitochondrial DNA Part A 30 {1): 43–51. https://doi.org/10.1080/24701394.2018.1450397.

Pogoda, Cloe S., Kyle G. Keepers, Arif Y. Nadiadi, Dustin W. Bailey, James C. Lendemer, Erin A. Tripp, and Nolan C. Kane. 2019. "Genome Streamlining via Complete Loss of Introns Has Occurred Multiple Times in Lichenized Fungal Mitochondria." Ecology and Evolution 9 {7): 4245–63. https://doi.org/10.1002/ece3.5056.

Possingham, J. V., and W. Saurer. 1969. "Changes in Chloroplast Number per Cell during Leaf Development in Spinach." Planta 86 {2): 186–94. https://doi.org/10.1007/BF00379826.

Preuss, Maren, Heroen Verbruggen, John A. West, and Giuseppe C. Zuccarello. 2021. "Divergence Times and Plastid Phylogenomics within the Intron-rich Order Erythropeltales (Compsopogonophyceae, Rhodophyta)." Journal of Phycology, April. https://doi.org/10.1111/jpy.13159.

Preuten, Tobias, Emilia Cincu, Jörg Fuchs, Reimo Zoschke, Karsten Liere, and Thomas Börner. 2010. "Fewer Genes than Organelles: Extremely Low and Variable Gene Copy Numbers in Mitochondria of Somatic Plant Cells." Plant Journal 64 {6): 948–59. https://doi.org/10.1111/j.1365-313X.2010.04389.x.

Pruchner, Dagmar, Susanne Beckert, Hermann Muhle, and Volker Knoop. 2002. "Divergent Intron Conservation in the Mitochondrial Nad2 Gene: Signatures for the Three Bryophyte Classes (Mosses, Liverworts, and Hornworts) and the Lycophytes." Journal of Molecular Evolution 55 {3): 265–71. https://doi.org/10.1007/s00239-002-2324-2.

Putnam, Andrea A., and Eckhard Jankowsky. 2013. "DEAD-Box Helicases as Integrators of RNA, Nucleotide and Protein Binding." Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. Biochim Biophys Acta. https://doi.org/10.1016/j.bbagrm.2013.02.002.

Puttaraju, M., and Michael Been. 1992. "Group I Permuted Intron-Exon {PIE) Sequences Self-Splice to Produce Circular Exons." Nucleic Acids Research 20 {20): 5357. https://doi.org/10.1093/NAR/20.20.5357.

Qiu, Yin Long. 2008. "Phytogeny and Evolution of Charophytic Algae and Land Plants." Journal of Systematics and Evolution 46 {3): 287–306. https://doi.org/10.3724/SP.J.1002.2008.08035.

Qiu, Yin Long, Yangrae Cho, J. Colin Cox, and Jeffrey D. Palmer. 1998. "The Gain of Three Mitochondrial Introns Identifies Liverworts as the Earliest Land Plants." Nature 394 {6694): 671–74. https://doi.org/10.1038/29286.

Qiu, Yin Long, Libo Li, Bin Wang, Zhiduan Chen, Volker Knoop, Milena Groth-Malonek, Olena Dombrovska, et al. 2006. "The Deepest Divergences in Land Plants Inferred from Phylogenomic Evidence." Proceedings of the National Academy of Sciences of the United States of America 103 {42): 15511–16. https://doi.org/10.1073/pnas.0603335103.

Qiu, Yin Long, and Jeffrey D. Palmer. 2004. "Many Independent Origins of Trans Splicing of a Plant Mitochondrial Group II Intron." Journal of Molecular Evolution 59 {1): 80–89. https://doi.org/10.1007/S00239-004-2606-Y/FIGURES/4.

Qiu, Yin Long, Alexander B. Taylor, and Hilary A. McManus. 2012. "Evolution of the Life Cycle in Land Plants." Journal of Systematics and Evolution 50 {3): 171–94. https://doi.org/10.1111/j.1759-6831.2012.00188.x.

Qu, Guosheng, Carol Lyn Piazza, Dorie Smith, and Marlene Belfort. 2018. "Group II Intron Inhibits Conjugative Relaxase Expression in Bacteria by MRNA Targeting." ELife 7 {June). https://doi.org/10.7554/eLife.34268.

Rabah, Samar O., Chaehee Lee, Nahid H. Hajrah, Rania M. Makki, Hesham F. Alharby, Alawiah M. Alhebshi, Jamal S.M. Sabir, Robert K. Jansen, and Tracey A. Ruhlman. 2017. "Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew." The Plant Genome 10 {3): plantgenome2017.03.0020. https://doi.org/10.3835/plantgenome2017.03.0020.

Raman, Gurusamy, Seongjun Park, Eun Mi Lee, and Seon Joo Park. 2019. "Evidence of Mitochondrial DNA in the Chloroplast Genome of Convallaria Keiskei and Its Subsequent Evolution in the Asparagales." Scientific Reports 9 {1): 1–11. https://doi.org/10.1038/s41598-019-41377-w.

Ran, Jin Hua, Hui Gao, and Xiao Quan Wang. 2010. "Fast Evolution of the Retroprocessed Mitochondrial Rps3 Gene in Conifer II and Further Evidence for the Phylogeny of Gymnosperms." Molecular Phylogenetics and Evolution 54 {1): 136–49. https://doi.org/10.1016/j.ympev.2009.09.011.

Rasmussen, Lars H., Stine Kroghsbo, Jens C. Frisvad, and Hans Christian B. Hansen. 2003. "Occurrence of the Carcinogenic Bracken Constituent Ptaquiloside in Fronds, Topsoils and Organic Soil Layers in Denmark." Chemosphere 51 (2): 117–27. https://doi.org/10.1016/S0045-6535{02)00694-X.

Rasmussen, Lars Holm. 2021. "Presence of the Carcinogen Ptaquiloside in Fern-Based Food Products and Traditional Medicine: Four Cases of Human Exposure." Current Research in Food Science 4 (January): 557–64. https://doi.org/10.1016/J.CRFS.2021.08.004.

Raubeson, Linda A., and Robert K. Jansen. 1992. "Chloroplast DNA Evidence on the Ancient Evolutionary Split in Vascular Land Plants." Science {New York, N.Y.) 255 {5052): 1697–99. https://doi.org/10.1126/SCIENCE.255.5052.1697.

Ren, Ru Chang, Xu Wei Yan, Ya Jie Zhao, Yi Ming Wei, Xiaoduo Lu, Jie Zang, Jia Wen Wu, et al. 2020. "The Novel E-Subgroup Pentatricopeptide Repeat Protein DEK55 Is Responsible for RNA Editing at Multiple Sites and for the Splicing of Nad1 and Nad4 in Maize." BMC Plant Biology 20 {1). https://doi.org/10.1186/S12870-020-02765-X.

Rensing, Stefan A., Daniel Lang, Andreas D. Zimmer, Astrid Terry, Asaf Salamov, Harris Shapiro, Tomoaki Nishiyama, et al. 2008. "The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants." Science 319 {5859): 64–69. https://doi.org/10.1126/science.1150646.

Renzaglia, Karen S., Juan Carlos Villareal Aguilar, and David J. Garbarry. 2018. "Morphology Supports -the Setaphyte Hypothesis: Mosses plus Liverworts Form a Natural Group." Bryophyte Diversity and Evolution 40 {2): 11. https://doi.org/10.11646/bde.40.2.1.

Rest, Joshua S., and David P. Mindell. 2003. "Retroids in Archaea: Phylogeny and Lateral Origins." Molecular Biology and Evolution 20 (7): 1134–42. https://doi.org/10.1093/molbev/msg135.

Rhoads, Anthony, and Kin Fai Au. 2015. "PacBio Sequencing and Its Applications." Genomics, Proteomics & Bioinformatics 13 {5): 278–89. https://doi.org/10.1016/J.GPB.2015.08.002.

Ricchetti, M., and H. Buc. 1993. "E. Coli DNA Polymerase I as a Reverse Transcriptase." EMBO Journal 12 (2): 387–96. https://doi.org/10.1002/j.1460-2075.1993.tb05670.x.

Rice, Danny W., Andrew J. Alverson, Aaron O. Richardson, Gregory J. Young, M. Virginia Sanchez-Puerta, Jeröme Munzinger, Kerrie Barry, et al. 2013. "Horizontal Transfer of Entire Genomes via Mitochondrial Fusion in the Angiosperm Amborella." Science 342 {6165): 1468–73. https://doi.org/10.1126/science.1246275.

Richardson, Aaron O., and Jeffrey D. Palmer. 2007. "Horizontal Gene Transfer in Plants." In Journal of Experimental Botany, 58:1–9. Oxford Academic. https://doi.org/10.1093/jxb/erl148.

Robart, Aaron R., Wooseok Seo, and Steven Zimmerly. 2007. "Insertion of Group II Intron Retroelements after Intrinsic Transcriptional Terminators." Proceedings of the National Academy of Sciences of the United States of America 104 {16): 6620–25. https://doi.org/10.1073/pnas.0700561104.

Robison, Tanner A., Amanda L. Grusz, Paul G. Wolf, Jeffrey P. Mower, Blake D. Fauskee, Karla Sosa, and Eric Schuettpelz. 2018. "Mobile Elements Shape Plastome Evolution in Ferns." Genome Biology and Evolution 10 {10): 2558–71. https://doi.org/10.1093/gbe/evy189.

Rüdinger, Mareike, Helena T. Funk, Stefan A. Rensing, Uwe G. Maier, and Volker Knoop. 2009. "RNA Editing: Only Eleven Sites Are Present in the Physcomitrella Patens Mitochondrial Transcriptome and

a Universal Nomenclature Proposal." Molecular Genetics and Genomics 2009 281:5 281 {5): 473–81. https://doi.org/10.1007/S00438-009-0424-Z.

Rüdinger, Mareike, Peter Szövenyi, Stefan A. Rensing, and Volker Knoop. 2011. "Assigning DYW-Type PPR Proteins to RNA Editing Sites in the Funariid Mosses Physcomitrella Patens and Funaria Hygrometrica." The Plant Journal 67 {2): 370–80. https://doi.org/10.1111/J.1365-313X.2011.04600.X.

Rugen, Nils, Henryk Straube, Linda E. Franken, Hans Peter Braun, and Holger Eubel. 2019. "Complexome Profiling Reveals Association of PPR Proteins with Ribosomes in the Mitochondria of Plants." Molecular and Cellular Proteomics 18 (7): 1345–62. https://doi.org/10.1074/mcp.RA119.001396.

Ruhfel, Brad R, Matthew A Gitzendanner, Pamela S Soltis, Douglas E Soltis, and J Gordon Burleigh. 2014. “From Algae to Angiosperms–Inferring the Phylogeny of Green Plants { Viridiplantae) from 360 Plastid Genomes." BMC Evolutionary Biology 2014 14:1 14 {1): 1–27. https://doi.org/10.1186/1471-2148-14-23.

Salinas-Giege, Thalia, Richard Giege, and Philippe Giege. 2015. "TRNA Biology in Mitochondria." International Journal of Molecular Sciences. MDPI AG. https://doi.org/10.3390/ijms16034518.

Samach, Aviva, Cathy Melamed-Bessudo, Naomi Avivi-Ragolski, Shmuel Pietrokovski, and Avraham A. Levy. 2011. "Identification of Plant RAD52 Homologs and Characterization of the Arabidopsis Thaliana RAD52-like Genes." Plant Cell 23 {12): 4266–79. https://doi.org/10.1105/tpc.111.091744.

Sanchez-Puerta, M. Virginia, Yangrae Cho, Jeffrey P. Mower, Andrew J. Alverson, and Jeffrey D. Palmer. 2008. "Frequent, Phylogenetically Local Horizontal Transfer of the Cox1 Group I Intron in Flowering Plant Mitochondria." Molecular Biology and Evolution 25 {8): 1762–77. https://doi.org/10.1093/molbev/msn129.

Sanchez-Puerta, Maria V., Cinthia C. Abbona, Shi Zhuo, Eric J. Tepe, Lynn Bohs, Richard G. Olmstead, and Jeffrey D. Palmer. 2011. "Multiple Recent Horizontal Transfers of the Cox1 Intron in Solanaceae and Extended Co-Conversion of Flanking Exons." BMC Evolutionary Biology 11 {1): 1–15. https://doi.org/10.1186/1471-2148-11-277.

Sanderson, Michael J., Dario Copetti, Alberto Burquez, Enriquena Bustamante, Joseph L.M. Charboneau, Luis E. Eguiarte, Sudhir Kumar, et al. 2015. "Exceptional Reduction of the Plastid Genome of Saguaro Cactus {Carnegiea Gigantea): Loss of the Ndh Gene Suite and Inverted Repeat 1." American Journal of Botany 102 (7): 1115–27. https://doi.org/10.3732/ajb.1500184.

Sardesai, Nagesh, and Subhashree Subramanyam. 2018. "Agrobacterium: A Genome-Editing Tool-Delivery System." Current Topics in Microbiology and Immunology 418: 463–88. https://doi.org/10.1007/82_2018_101.

Sato, Naoki, Kimihiro Terasawa, Kazunori Miyajima, and Yukihiro Kabeya. 2003. "Organization, Developmental Dynamics, and Evolution of Plastid Nucleoids." International Review of Cytology 232: 217–62. https://doi.org/10.1016/S0074-7696{03)32006-6.

Schallenberg-Rüdinger, M., and V. Knoop. 2016. "Coevolution of Organelle RNA Editing and Nuclear Specificity Factors in Early Land Plants." Advances in Botanical Research 78 {January): 37–93. https://doi.org/10.1016/bs.abr.2016.01.002.

Schmelzer, Carlo, and Rudolf J. Schweyen. 1986. "Self-Splicing of Group II Introns in Vitro: Mapping of the Branch Point and Mutational Inhibition of Lariat Formation." Cell 46 {4): 557–65. https://doi.org/10.1016/0092-8674{86)90881-0.

Schmitz-Linneweber, Christian, Marie Kristin Lampe, Laure D. Sultan, and Oren Ostersetzer-Biran. 2015. "Organellar Maturases: A Window into the Evolution of the Spliceosome." Biochimica et Biophysica Acta - Bioenergetics. Elsevier B.V. https://doi.org/10.1016/j.bbabio.2015.01.009.

Schmitz-Linneweber, Christian, and Ian Small. 2008. "Pentatricopeptide Repeat Proteins: A Socket Set for Organelle Gene Expression." Trends in Plant Science. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2008.10.001.

Schneider, André. 2011. “Mitochondrial TRNA Import and Its Consequences for Mitochondrial Translation." Annual Review of Biochemistry 80 {July): 1033–53. https://doi.org/10.1146/annurev-biochem-060109-092838.

Schneider, Haerald, Hongmei Liu, James Clark, Oriane Hidalgo, Jaume Pellicer, Shouzhou Zhang, Laura J. Kelly, Michael F. Fay, and Ilia J. Leitch. 2015. “Are the Genomes of Royal Ferns Really Frozen in Time? Evidence for Coinciding Genome Stability and Limited Evolvability in the Royal Ferns." New Phytologist. Blackwell Publishing Ltd. https://doi.org/10.1111/nph.13330.

Schneider, Harald, Eric Schuetipelz, Kathleen M. Pryer, Raymond Cranfill, Susana Magall6n, and Richard Lupia. 2004. "Ferns Diversified in the Shadow of Angiosperms." Nature 428 {6982): 553–57. https://doi.org/10.1038/nature02361.

Schrallhammer, M, M Castelli, and G Petroni. 2018. "Phylogenetic Relationships among Endosymbiotic R-Body Producer: Bacteria Providing Their Host the Killer Trait." Systematic and Applied Microbiology 41 (3): 213–20. https://doi.org/10.1016/J.SYAPM.2018.01.005.

Schuettpelz, Eric, and Kathleen M. Pryer. 2009. "Evidence for a Cenozoic Radiation of Ferns in an Angiosperm-Dominated Canopy." Proceedings of the National Academy of Sciences of the United States of America 106 {27): 11200–205. https://doi.org/10.1073/pnas.0811136106.

Schuler, Hannes, Scott P. Egan, Glen R. Hood, Robert W. Busbee, Amanda L. Driscoe, and James R. Ott. 2018. "Diversity and Distribution of Wolbachia in Relation to Geography, Host Plant Affiliation and Life Cycle of a Heterogonic Gall Wasp." BMC Evolutionary Biology 2018 18:1 18 {1): 1–15. https://doi.org/10.1186/S12862-018-1151-Z.

Schulte, Peter, Laia Alegret, Ignacio Arenillas, Jose A. Arz, Penny J. Barton, Paul R. Bown, Timothy J. Bralower, et al. 2010. “The Chicxulub Asteroid Impact and Mass Extinction at the Cretaceous-Paleogene Boundary." Science. American Association for the Advancement of Science. https://doi.org/10.1126/science.1177265.

Schuster, W., and A. Brennicke. 1994. "The Plant Mitochondrial Genome: Physical Structure, Information Content, RNA Editing, and Gene Migration to the Nucleus." ANNU.REV.PLANT PHYSIOL.PLANT MOL.BIOL. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA . https://doi.org/10.1146/annurev.pp.45.060194.000425.

Schwarzländer, Markus, Ann Christine König, Lee J. Sweetlove, and Iris Finkemeier. 2012. “The Impact of Impaired Mitochondrial Function on Retrograde Signalling: A Meta-Analysis of Transcriptomic Responses." Journal of Experimental Botany 63 {4): 1735–50. https://doi.org/10.1093/jxb/err374.

Seguí-Simarro, José M., María José Coronado, and L. Andrew Staehelin. 2008. “The Mitochondrial Cycle of Arabidopsis Shoot Apical Meristem and Leaf Primordium Meristematic Cells Is Defined by a Perinuclear Tentaculate/Cage-Like Mitochondrion." Plant Physiology 148 {3): 1380. https://doi.org/10.1104/PP.108.126953.

Seif, Elias, Jessica Leigh, Yu Liu, Ingeborg Roewer, Lise Forget, and B. Franz Lang. 2005. "Comparative Mitochondrial Genomics in Zygomycetes: Bacteria-like RNase P RNAs, Mobile Elements and a Close Source of the Group I Intron Invasion in Angiosperms." Nucleic Acids Research 33 {2): 734–44. https://doi.org/10.1093/nar/gki199.

Shameer, Sanu, R. George Ratcliffe, and Lee J. Sweetlove. 2019. "Leaf Energy Balance Requires Mitochondrial Respiration and Export of Chloroplast NADPH in the Light." Plant Physiology 180 {4): 1947. https://doi.org/10.1104/PP.19.00624.

Sharpe, Joanne M., Klaus Mehltreter, and Lawrence R. Walker. 2010. “Ecological Importance of Ferns.” Fern Ecology, January, 1–21. https://doi.org/10.1017/CBO9780511844898.002.

Shen, Jia, Yuejian Zhang, Michael J. Havey, and Weisong Shou. 2019. "Copy Numbers of Mitochondrial Genes Change during Melon Leaf Development and Are Lower than the Numbers of Mitochondria." Horticulture Research 6 {1): 1–9. https://doi.org/10.1038/s41438-019-0177-8.

Sheahan, Michael B., David W. McCurdy, and Ray J. Rose. 2005. “Mitochondria as a Connected Population: Ensuring Continuity of the Mitochondrial Genome during Plant Cell Dedifferentiation through Massive Mitochondrial Fusion." The Plant Journal : For Cell and Molecular Biology 44 {5): 744–55. https://doi.org/10.1111/J.1365-313X.2005.02561.X.

Sheahan, Michael B., Ray J. Rose, and David W. McCurdy. 2004. "Organelle Inheritance in Plant Cell Division: The Actin Cytoskeleton Is Required for Unbiased Inheritance of Chloroplasts, Mitochondria and Endoplasmic Reticulum in Dividing Protoplasts." The Plant Journal : For Cell and Molecular Biology 37 (3): 379–90. https://doi.org/10.1046/J.1365-313X.2003.01967.X.

Sheveleva, Elena V., and Richard B. Hallick. 2004. “Recent Horizontal Intron Transfer to a Chloroplast Genome." Nucleic Acids Research 32 {2): 803–10. https://doi.org/10.1093/nar/gkh225.

Shevtsov-Tal, Sofia, Corinne Best, Roei Matan, Sam Aldrin Chandran, Gregory G. Brown, and Oren Ostersetzer-Biran. 2020. "NMAT3 Is an Essential Maturase Splicing Factor Required for Holo-Complex I Biogenesis and Embryo-Development in Arabidopsis Thaliana Plants." BioRxiv. bioRxiv. https://doi.org/10.1101/2020.10.20.346734.

Shukla, Anoop Kumar, Santosh Kumar Upadhyay, Manisha Mishra, Sharad Saurabh, Rahul Singh, Harpal Singh, Nidhi Thakur, et al. 2016. "Expression of an Insecticidal Fern Protein in Cotton Protects against Whitefly." Nature Biotechnology 34 {10): 1046–51. https://doi.org/10.1038/nbt.3665.

Signorovitch, Ana Y., Leo W. Buss, and Stephen L. Dellaporta. 2007. "Comparative Genomics of Large Mitochondria in Placozoans." PLoS Genetics 3 {1): 0044–0050. https://doi.org/10.1371/journal.pgen.0030013.

Siguier, Patricia, Edith Gourbeyre, and Mick Chandler. 2014. "Bacterial Insertion Sequences: Their Genomic Impact and Diversity." FEMS Microbiology Reviews 38 {5): 865–91. https://doi.org/10.1111/1574-6976.12067.

Simon, Dawn M., Nicholas A.C. Clarke, Bonnie A. McNeil, Ian Johnson, Davin Pantuso, Lixin Dai, Dinggeng Chai, and Steven Zimmerly. 2008. “Group II Introns in Eubacteria and Archaea: ORF-Less Introns and New Varieties." RNA 14 {9): 1704–13. https://doi.org/10.1261/rna.1056108.

Skippington, Elizabeth, Todd J. Barkmanb, Danny W. Ricea, and Jeffrey D. Palmera. 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 of the United States of America 112 (27): E3515–24. https://doi.org/10.1073/pnas.1504491112.

Slipiko, Monika, Kamil Myszczynski, Katarzyna Buczkowska-Chmielewska, Alina Bc1czkiewicz, Monika Szczecinska, and Jakub Sawicki. 2017. "Comparative Analysis of Four Calypogeia Species Revealed Unexpected Change in Evolutionarily-Stable Liverwort Mitogenomes.” Genes 8 (12). https://doi.org/10.3390/genes8120395.

Sloan, Daniel B. 2013. "One Ring to Rule Them All? Genome Sequencing Provides New Insights into the 'Master Circle' Model of Plant Mitochondrial DNA Structure." New Phytologist 200 {4): 978–85. https://doi.org/10.1111/NPH.12395.

Sloan, Daniel B., Andrew J. Alverson, John P. Chuckalovcak, Martin Wu, David E. McCauley, Jeffrey D. Palmer, and Douglas R. Taylor. 2012. "Rapid Evolution of Enormous, Multichromosomal Genomes in

Flowering Plant Mitochondria with Exceptionally High Mutation Rates." PLoS Biology 10 {1): 1001241. https://doi.org/10.1371/journal.pbio.1001241.

Sloan, Daniel B., Alice H. MacQueen, Andrew J. Alverson, Jeffrey D. Palmer, and Douglas R. Taylor. 2010. “Extensive Loss of RNA Editing Sites in Rapidly Evolving Silene Mitochondrial Genomes: Selection vs. Retroprocessing as the Driving Force." Genetics 185 {4): 1369–80. https://doi.org/10.1534/genetics.110.118000.

Small, I., L. Marechal-Drouard, J. Masson, G. Pelletier, A. Cosset, J. H. Weil, and A. Dietrich. 1992. "In Vivo Import of a Normal or Mutagenized Heterologous Transfer RNA into the Mitochondria of Transgenic Plants: Towards Novel Ways of Influencing Mitochondrial Gene Expression?" EMBO Journal 11 (4): 1291–96. https://doi.org/10.1002/j.1460-2075.1992.tb05172.x.

Small, Ian D., Mareike Schallenberg-Rüdinger, Mizuki Takenaka, Hakim Mireau, and Oren Ostersetzer-Biran. 2020. "Plant Organellar RNA Editing: What 30 Years of Research Has Revealed." Plant Journal. Blackwell Publishing Ltd. https://doi.org/10.1111/tpj.14578.

Smathers, Claire M., and Aaron R. Robart. 2019. “The Mechanism of Splicing as Told by Group II Introns: Ancestors of the Spliceosome.” Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1862 {11–12): 194390. https://doi.org/10.1016/J.BBAGRM.2019.06.001.

Smith, David Roy. 2011. “Extending the Limited Transfer Window Hypothesis to Inter-Organelle DNA Migration." Genome Biology and Evolution 3 {1): 743–48. https://doi.org/10.1093/gbe/evr068.

Smith, David Roy, and Rory J. Craig. 2021. "Does Mitochondrial DNA Replication in Chlamydomonas Require a Reverse Transcriptase?" New Phytologist. Blackwell Publishing Ltd. https://doi.org/10.1111/nph.16930.

Sousa, Filipe, Peter Civa1, Peter G. Foster, and Cymon J. Cox. 2020. "The Chloroplast Land Plant Phylogeny: Analyses Employing Better-Fitting Tree- and Site-Heterogeneous Composition Models." Frontiers in Plant Science 0 {July): 1062. https://doi.org/10.3389/FPLS.2020.01062.

Sousa, Filipe de, Peter G. Foster, Philip C. J. Donoghue, Harald Schneider, and Cymon J. Cox. 2019. "Nuclear Protein Phylogenies Support the Monophyly of the Three Bryophyte Groups {Bryophyta Schimp.)." New Phytologist 222 {1): 565–75. https://doi.org/10.1111/NPH.15587.

Stegemann, Sandra, Stefanie Hartmann, Stephanie Ruf, and Ralph Bock. 2003. “High-Frequency Gene Transfer from the Chloroplast Genome to the Nucleus." Proceedings of the National Academy of Sciences of the United States of America 100 {15): 8828–33. https://doi.org/10.1073/pnas.1430924100.

Steinberg, Pablo, and Stefan Engert. 2019. "A Daring Task: The Battle against Food Crime." Journal of Consumer Protection and Food Safety 2019 14:4 14 (4): 317–18. https://doi.org/10.1007/S00003-019-01259-9.

Straub, Shannon C.K., Richard C. Cronn, Christopher Edwards, Mark Fishbein, and Aaron Liston. 2013. "Horizontal Transfer of DNA from the Mitochondrial to the Plastid Genome and Its Subsequent Evolution in Milkweeds {Apocynaceae)." Genome Biology and Evolution 5 {10): 1872–85. https://doi.org/10.1093/gbe/evt140.

Su, Chao, Jinhong Yuan, Hongtao Zhao, Yankun Zhao, Hongtao Ji, Youning Wang, and Xia Li. 2017. "RUG3 Is a Negative Regulator of Plant Responses to ABA in Arabidopsis Thaliana." Plant Signaling and Behavior, June. https://doi.org/10.1080/15592324.2017.1333217.

Su, Chao, Hongtao Zhao, Yankun Zhao, Hongtao Ji, Youning Wang, Liya Zhi, and Xia Li. 2017. "RUG3 and ATM Synergistically Regulate the Alternative Splicing of Mitochondrial Nad2 and the DNA Damage Response in Arabidopsis Thaliana." Scientific Reports 7 {1): 1–14. https://doi.org/10.1038/srep43897.

Su, Danyan, Lingxiao Yang, Xuan Shi, Xiaoya Ma, Xiaofan Zhou, S Blair Hedges, and Bojian Zhong. 2021. “Large-Scale Phylogenomic Analyses Reveal the Monophyly of Bryophytes and Neoproterozoic Origin of Land Plants." Molecular Biology and Evolution 38 {8): 3332–44. https://doi.org/10.1093/MOLBEV/MSAB106.

Su, Yan, Pratibha P. Ghodke, Martin Egli, Lin Li, Yinsheng Wang, and F. Peter Guengerich. 2019. "Human DNA Polymerase Has Reverse Transcriptase Activity in Cellular Environments." Journal of Biological Chemistry 294 {15): 6073–81. https://doi.org/10.1074/jbc.RA119.007925.

Sullivan, Alexis R., Yrin Eldfjell, Bastian Schiffthaler, Nicolas Delhomme, Torben Asp, Kim H. Hebelstrup, Olivier Keech, et al. 2019. "The Mitogenome of Norway Spruce and a Reappraisal of Mitochondrial Recombination in Plants." Genome Biology and Evolution 12 {1): 3586–98. https://doi.org/10.1093/gbe/evz263.

Sultan, Laure D., Daria Mileshina, Felix Grewe, Katarzyna Rolle, Sivan Abudraham, Pawet Gtodowicz, Adnan Khan Niazi, 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 {11): 2805–29. https://doi.org/10.1105/tpc.16.00398.

Sun, Feng, Xiaoyan Zhang, Yun Shen, Hongchun Wang, Rui Liu, Xiaomin Wang, Dahai Gao, Yan Zhuo Yang, Yiwei Liu, and Bao Cai Tan. 2018. "The Pentatricopeptide Repeat Protein EMPTY PERICARP8 Is Required for the Splicing of Three Mitochondrial Introns and Seed Development in Maize." The Plant Journal : For Cell and Molecular Biology 95 {5): 919–32. https://doi.org/10.1111/TPJ.14030.

Suzuki, Haruo, Amy L. Dapper, Craig E. Jackson, Heewook Lee, Vikas Pejaver, Thomas G. Doak, Michael Lynch, and John R. Preer. 2015. "Draft Genome Sequence of Caedibacter Varicaedens, a Kappa Killer Endosymbiont Bacterium of the Ciliate Paramecium Biaurelia." Genome Announcements 3 {6). https://doi.org/10.1128/genomeA.01310-15.

Symington, Lorraine S. 2002. " Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair ." Microbiology and Molecular Biology Reviews 66 {4): 630–70. https://doi.org/10.1128/mmbr.66.4.630-670.2002.

Terasawa, Kimihiro, Masaki Odahara, Yukihiro Kabeya, Tatsuhiko Kikugawa, Yasuhiko Sekine, Makoto Fujiwara, and Naoki Sato. 2007. "The Mitochondrial Genome of the Moss Physcomitrella Patens Sheds New Light on Mitochondrial Evolution in Land Plants." Molecular Biology and Evolution 24 {3): 699–709. https://doi.org/10.1093/molbev/msl198.

Thongkhao, Kannika, Chayapol Tungphatthong, Thatree Phadungcharoen, and Suchada Sukrong. 2020. "The Use of Plant DNA Barcoding Coupled with HRM Analysis to Differentiate Edible Vegetables from Poisonous Plants for Food Safety." Food Control 109 {March): 106896. https://doi.org/10.1016/J.FOODCONT.2019.106896.

Timmis, Jeremy N., Michael A. Ayliff, Chun Y. Huang, and William Martin. 2004. "Endosymbiotic Gene Transfer: Organelle Genomes Forge Eukaryotic Chromosomes." Nature Reviews Genetics. Nature Publishing Group. https://doi.org/10.1038/nrg1271.

Timmis, Jeremy N., and Dong Wang. 2013. "Endosymbiotic Evolution: The Totalitarian Nucleus Is Foiled Again." Current Biology. Cell Press. https://doi.org/10.1016/j.cub.2012.11.038.

Toor, Navtej, Georg Hausner, and Steven Zimmerly. 2001. "Coevolution of Group II Intron RNA Structures with Their Intron-Encoded Reverse Transcriptases." RNA 7 {8): 1142–52. https://doi.org/10.1017/S1355838201010251.

Toro, Nicolas, Francisco Martfnez-Abarca, and Manuel Fernández-Lpez. 2016. "The Early Events Underlying Genome Evolution in a Localized Sinorhizobium Meliloti Population." BMC Genomics 17 (1): 1–14. https://doi.org/10.1186/s12864-016-2878-9.

Toro, Nicolas, and Rafael Nisa-Martfnez. 2014. "Comprehensive Phylogenetic Analysis of Bacterial Reverse Transcriptases." PLoS ONE 9 {11). https://doi.org/10.1371/journal.pone.0114083.

Turmel, Monique, Christian Otis, and Claude Lemieux. 2002a. "The Complete Mitochondrial DNA Sequence of Mesostigma Viride Identifies This Green Alga as the Earliest Green Plant Divergence and Predicts a Highly Compact Mitochondrial Genome in the Ancestor of All Green Plants." Molecular Biology and Evolution 19 {1): 24–38. https://doi.org/10.1093/oxfordjournals.molbev.a003979.

———. 2002b. "The Chloroplast and Mitochondrial Genome Sequences of the Charophyte Chaetosphaeridium Globosum: Insights into the Timing of the Events That Restructured Organelle DNAs within the Green Algal Lineage That Led to Land Plants." Proceedings of the National Academy of Sciences of the United States of America 99 {17): 11275–80. https://doi.org/10.1073/pnas.162203299.

———. 2003. "The Mitochondrial Genome of Chara Vulgaris: Insights into the Mitochondrial DNA Architecture of the Last Common Ancestor of Green Algae and Land Plants." Plant Cell 15 {8): 1888–1903. https://doi.org/10.1105/tpc.013169.

———. 2007. "An Unexpectedly Large and Loosely Packed Mitochondrial Genome in the Charophycean Green Alga Chlorokybus Atmophyticus." BMC Genomics 8 (May): 137. https://doi.org/10.1186/1471-2164-8-137.

———. 2013. "Tracing the Evolution of Streptophyte Algae and Their Mitochondrial Genome." Genome Biology and Evolution 5 {10): 1817–35. https://doi.org/10.1093/GBE/EVT135.

———. 2019. “Complete Mitogenome of the Streptophyte Green Alga Coleochaete Scutata {Coleochaetophyceae)." Mitochondrial DNA Part B 4 {2): 4209–10. https://doi.org/10.1080/23802359.2019.1693300.

Unseld, Michael, Joachim R. Marienfeld, Petra Brandt, and Axel Brennicke. 1997. "The Mitochondrial Genome of Arabidopsis Thaliana Contains 57 Genes in 366,924 Nucleotides." Nature Genetics. Nat Genet. https://doi.org/10.1038/ng0197-57.

Vahrenholz, Carola, Gudula Riemen, Elke Pratje, Bernard Dujon, and Georg Michaelis. 1993. "Mitochondrial DNA of Chlamydomonas Reinhardtii: The Structure of the Ends of the Linear 15.8-Kb Genome Suggests Mechanisms for DNA Replication." Current Genetics 24 {3): 241–47. https://doi.org/10.1007/BF00351798.

Vajda, V., J. I. Raine, and C. J. Hollis. 2001. "Indication of Global Deforestation at the Cretaceous-Tertiary Boundary by New Zealand Fern Spike." Science 294 {5547): 1700–1702. https://doi.org/10.1126/science.1064706.

Valach, Matus, Jose Angel Gonzalez Alcazar, Matt Sarrasin, B. Franz Lang, Michael W. Gray, and Gertraud Burger. 2021. "An Unexpectedly Complex Mitoribosome in Andalucia Godoyi, a Protist with the Most Bacteria-like Mitochondrial Genome." Molecular Biology and Evolution 38 (3): 788–804. https://doi.org/10.1093/molbev/msaa223.

Valles, Yvonne, Kenneth M. Halanych, and Jeffrey L. Boore. 2008. "Group II Introns Break New Boundaries: Presence in a Bilaterian's Genome." Edited by Jason Stajich. PloS One 3 {1): e1488. https://doi.org/10.1371/journal.pone.0001488.

Veen, R. van der, A. C. Arnberg, G. van der Horst, L. Bonen, H. F. Tabak, and L. A. Grivell. 1986. “Excised Group II Introns in Yeast Mitochondria Are Lariats and Can Be Formed by Self-Splicing in Vitro.” Cell 44 (2): 225–34. https://doi.org/10.1016/0092-8674{86)90756-7.

Virgilio, A, A Sinisi, V Russo, S Gerardo, A Santoro, A Galeone, O Taglialatela-Scafati, and F Roperto. 2015. "Ptaquiloside, the Major Carcinogen of Bracken Fern, in the Pooled Raw Milk of Healthy Sheep

and Goats: An Underestimated, Global Concern of Food Safety." Journal of Agricultural and Food Chemistry 63 {19): 4886–92. https://doi.org/10.1021/ACS.JAFC.5B01937.

Volkmar, U., M. Groth-Malonek, J. Heinrichs, H. Muhle, M. Polsakiewicz, and V. Knoop. 2012. "Exclusive Conservation of Mitochondrial Group II Intron Nad4i548 among Liverworts and Its Use for Phylogenetic Studies in This Ancient Plant Clade." Plant Biology 14 {2): 382–91. https://doi.org/10.1111/j.1438-8677.2011.00499.x.

Volkmar, Ute, and Volker Knoop. 2010. "Introducing Intron Locus Cox1i624 for Phylogenetic Analyses in Bryophytes: On the Issue of Takakia as Sister Genus to All Other Extant Mosses." Journal of Molecular Evolution 70 {5): 506–18. https://doi.org/10.1007/s00239-010-9348-9.

Wahl, Markus C., Cindy L. Will, and Reinhard Lührmann. 2009. "The Spliceosome: Design Principles of a Dynamic RNP Machine." Cell 136 {4): 701–18. https://doi.org/10.1016/J.CELL.2009.02.009.

Wahrmund, Ute, Milena Groth-Malonek, and Volker Knoop. 2008. "Tracing Plant Mitochondrial DNA Evolution: Rearrangements of the Ancient Mitochondrial Gene Cluster TrnA-TrnT-Nad7 in Liverwort Phylogeny." Journal of Molecular Evolution 2008 66:6 66 {6): 621–29. https://doi.org/10.1007/S00239-008-9114-4.

Wakefield, J K, S A Jablonski, and C D Morrow. 1992. "In Vitro Enzymatic Activity of Human Immunodeficiency Virus Type 1 Reverse Transcriptase Mutants in the Highly Conserved YMDD Amino Acid Motif Correlates with the Infectious Potential of the Proviral Genome." Journal of Virology 66 {11): 6806–12. https://doi.org/10.1128/jvi.66.11.6806-6812.1992.

Waldern, Justin, Nicholas J. Schiraldi, Marlene Belfort, Olga Novikova, and Irina Arkhipova. 2020. "Bacterial Group II Intron Genomic Neighborhoods Reflect Survival Strategies: Hiding and Hijacking." Molecular Biology and Evolution 37 {7): 1942–48. https://doi.org/10.1093/molbev/msaa055.

Wallet, Clementine, Monique Le Ret, Marc Bergdoll, Marc Bichara, André Dietrich, and José M. Gualberto. 2015. "The RECG1 DNA Translocase Is a Key Factor in Recombination Surveillance, Repair, and Segregation of the Mitochondrial DNA in Arabidopsisopen." Plant Cell 27 {10): 2907–25. https://doi.org/10.1105/tpc.15.00680.

Wang, Dong, Andrew H. Lloyd, and Jeremy N. Timmis. 2012. "Environmental Stress Increases the Entry of Cytoplasmic Organellar DNA into the Nucleus in Plants." Proceedings of the National Academy of Sciences of the United States of America 109 {7): 2444–48. https://doi.org/10.1073/pnas.1117890109.

Wang, Dong, Mathieu Rousseau-Gueutin, and Jeremy N. Timmis. 2012. "Plastid Sequences Contribute to Some Plant Mitochondrial Genes." Molecular Biology and Evolution 29 {7): 1707–11. https://doi.org/10.1093/molbev/mss016.

Wang, Liang, Wenwei Zhang, Shijia Liu, Yunlu Tian, Xi Liu, Haigang Yan, Yue Cai, et al. 2021. "Rice FLOURY SHRUNKEN ENDOSPERM 5 Encodes a Putative Plant Organelle RNA Recognition Protein That Is Required for Cis-Splicing of Mitochondrial Nad4 Intron 1." Rice 14 (1): 1–13. https://doi.org/10.1186/s12284-021-00463-2.

Wang, Yan, Oliver Berkowitz, Jennifer Selinski, Yue Xu, Andreas Hartmann, and James Whelan. 2018. "Stress Responsive Mitochondrial Proteins in Arabidopsis Thaliana." Free Radical Biology and Medicine 122 (July): 28–39. https://doi.org/10.1016/J.FREERADBIOMED.2018.03.031.

Warren, Jessica M., and Daniel B. Sloan. 2020. "Interchangeable Parts: The Evolutionarily Dynamic TRNA Population in Plant Mitochondria." Mitochondrion 52 (May): 144–56. https://doi.org/10.1016/J.MITO.2020.03.007.

Wartmann, Oliver. 2019. "Charakterisierung Der Kernkodierten Maturasen NMat4 Und NMat5 in Physcomitrella Patens Durch Das CRISPR/Cas9-System Und Homologe Rekombination." Bachelor Thesis AG Knoop IZMB.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. "Bandage: Interactive Visualization of de Novo Genome Assemblies." Bioinformatics 31 {20): 3350–52. https://doi.org/10.1093/BIOINFORMATICS/BTV383.

Wicke, Susann, Gerald M. Schneeweiss, Claude W. dePamphilis, Kai F. Müller, and Dietmar Quandt. 2011. "The Evolution of the Plastid Chromosome in Land Plants: Gene Content, Gene Order, Gene Function." Plant Molecular Biology. Plant Mol Biol. https://doi.org/10.1007/s11103-011-9762-4.

Wickell, David A., and Fay Wei Li. 2020. "On the Evolutionary Significance of Horizontal Gene Transfers in Plants." New Phytologist 225 {1): 113–17. https://doi.org/10.1111/nph.16022.

Wickett, Norman J., Siavash Mirarab, Nam Nguyen, Tandy Warnow, Eric Carpenter, Naim Matasci, Saravanaraj Ayyampalayam, et al. 2014. "Phylotranscriptomic Analysis of the Origin and Early Diversification of Land Plants." Proceedings of the National Academy of Sciences 111 {45): E4859–68. https://doi.org/10.1073/PNAS.1323926111.

Wikström, Niklas, and Kathleen M. Pryer. 2005. "Incongruence between Primary Sequence Data and the Distribution of a Mitochondrial Atp1 Group II Intron among Ferns and Horsetails." Molecular Phylogenetics and Evolution 36 {3): 484–93. https://doi.org/10.1016/j.ympev.2005.04.008.

Wisnovsky, Simon, Sae Rin Jean, and Shana O. Kelley. 2016. "Mitochondrial DNA Repair and Replication Proteins Revealed by Targeted Chemical Probes." Nature Chemical Biology 12 {7): 567–73. https://doi.org/10.1038/nchembio.2102.

Wolf, Paul, Jessie Roper, and Aaron Duffy. 2010. "The Evolution of Chloroplast Genome Structure in Ferns." Genome 53 (9): 731–38. https://doi.org/10.1139/G10-061.

Wu, Zhiqiang, Jocelyn M. Cuthbert, Douglas R. Taylor, and Daniel B. Sloan. 2015. "The Massive Mitochondrial Genome of the Angiosperm Silene Noctiflora Is Evolving by Gain or Loss of Entire Chromosomes." Proceedings of the National Academy of Sciences of the United States of America 112 (33): 10185–91. https://doi.org/10.1073/pnas.1421397112.

Wyatt, David W., Wanjuan Feng, Michael P. Conlin, Matthew J. Yousefzadeh, Steven A. Roberts, Piotr Mieczkowski, Richard D. Wood, Gaorav P. Gupta, and Dale A. Ramsden. 2016. "Essential Roles for Polymerase e-Mediated End Joining in the Repair of Chromosome Breaks." Molecular Cell 63 {4): 662–73. https://doi.org/10.1016/j.molcel.2016.06.020.

Wynn, Emily L., and Alan C. Christensen. 2019. "Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution." G3: Genes, Genomes, Genetics 9 {2): 549–59. https://doi.org/10.1534/g3.118.200948.

Xie, Shangqian, Amy Wing Sze Leung, Zhenxian Zheng, Dake Zhang, Chuanle Xiao, Ruibang Luo, Ming Luo, and Shoudong Zhang. 2021. "Applications and Potentials of Nanopore Sequencing in the (Epi)Genome and (Epi)Transcriptome Era." The Innovation 2 {4): 100153. https://doi.org/10.1016/J.XINN.2021.100153.

Xu, Zhichao, Tianyi Xin, Dorothea Bartels, Ying Li, Wei Gu, Hui Yao, Sai Liu, et al. 2018. "Genome Analysis of the Ancient Tracheophyte Selaginella Tamariscina Reveals Evolutionary Features Relevant to the Acquisition of Desiccation Tolerance." Molecular Plant 11 {7): 983–94. https://doi.org/10.1016/j.molp.2018.05.003.

Xue, Jia Yu, Yang Liu, Libo Li, Bin Wang, and Yin Long Qiu. 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 {1): 53–61. https://doi.org/10.1007/s00294-009-0279-1.

Yadav, Sunil K., Archana, Rahul Singh, Pradhyumna Kumar Singh, and Prema G. Vasudev. 2019. "Insecticidal Fern Protein Tma12 Is Possibly a Lytic Polysaccharide Monooxygenase." Planta 249 {6). https://doi.org/10.1007/s00425-019-03135-0.

Yan, Huili, Yiwei Gao, Lulu Wu, Luyao Wang, Tian Zhang, Changhua Dai, Wenxiu Xu, et al. 2019. "Potential Use of the Pteris Vittata Arsenic Hyperaccumulation-Regulation Network for Phytoremediation." Journal of Hazardous Materials 368 {April): 386–96. https://doi.org/10.1016/j.jhazmat.2019.01.072.

Yang, Yang, Yan Zhao, Yingqi Zhang, Lihua Niu, Wanyue Li, Wenqin Lu, Jinfang Li, Patrick Schäfer, Yuling Meng, and Weixing Shan. 2022. "A Mitochondrial RNA Processing Protein Mediates Plant Immunity to a Broad Spectrum of Pathogens by Modulating the Mitochondrial Oxidative Burst." The Plant Cell 34 {6): 2343–63. https://doi.org/10.1093/PLCELL/KOAC082.

Yu, Chun-Ying, and Hung-Chih Kuo. 2019. "The Emerging Roles and Functions of Circular RNAs and Their Generation." Journal of Biomedical Science 26 {1): 29. https://doi.org/10.1186/s12929-019-0523-z.

Zancani, Marco, Enrico Braidot, Antonio Filippi, and Giovanna Lippe. 2020. "Structural and Functional Properties of Plant Mitochondrial F-ATP Synthase." Mitochondrion 53 {July): 178–93. https://doi.org/10.1016/J.MITO.2020.06.001.

Zerbino, Daniel R., and Ewan Birney. 2008. "Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs." Genome Research 18 {5): 821–29. https://doi.org/10.1101/GR.074492.107.

Zhao, Chen, Fei Liu, and Anna Marie Pyle. 2018. "An Ultraprocessive, Accurate Reverse Transcriptase Encoded by a Metazoan Group II Intron." RNA 24 {2): 183–85. https://doi.org/10.1261/rna.063479.117.

Zhao, Chen, and Anna Marie Pyle. 2017. "The Group II Intron Maturase: A Reverse Transcriptase and Splicing Factor Go Hand in Hand." Current Opinion in Structural Biology. Elsevier Ltd. https://doi.org/10.1016/j.sbi.2017.05.002.

Zhong, Jin, and Alan M. Lambowitz. 2003. "Group II Intron Mobility Using Nascent Strands at DNA Replication Forks to Prime Reverse Transcription." EMBO Journal 22 (17): 4555–65. https://doi.org/10.1093/emboj/cdg433.

Zhou, Hong, and Klaus von Schwartzenberg. 2020. "Zygnematophyceae: From Living Algae Collections to the Establishment of Future Models." Journal of Experimental Botany 71 (11): 3296–3304. https://doi.org/10.1093/JXB/ERAA091.

Zimmerly, Steven, Huatao Guo, Philip S. Perlman, and Alan M. Lambowltz. 1995. "Group II Intron Mobility Occurs by Target DNA-Primed Reverse Transcription." Cell 82 {4): 545–54. https://doi.org/10.1016/0092-8674{95)90027-6.

Zimmerly, Steven, and Cameron Semper. 2015. "Evolution of Group II Introns." Mobile DNA. BioMed Central Ltd. https://doi.org/10.1186/s13100-015-0037-5.

Zmudjak, Michal, Catherine Colas des Francs-Small, Ido Keren, Felix Shaya, Eduard Belausov, Ian Small, and Oren Ostersetzer-Biran. 2013. "MCSF1, a Nucleus-Encoded CRM Protein Required for the Processing of Many Mitochondrial Introns, Is Involved in the Biogenesis of Respiratory Complexes I and IV in Arabidopsis." New Phytologist 199 {2): 379–94. https://doi.org/10.1111/nph.12282.

Zmudjak, Michal, Sofia Shevtsov, Laure D. Sultan, Ido Keren, and Oren Ostersetzer-Biran. 2017. "Analysis of the Roles of the Arabidopsis NMAT2 and PMH2 Proteins Provided with New Insights into

the Regulation of Group II Intron Splicing in Land-Plant Mitochondria." International Journal of Molecular Sciences 18 {11): 2428. https://doi.org/10.3390/ijms18112428.

Zoschke, Reimo, and Ralph Bock. 2018. "Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation." Plant Cell. American Society of Plant Biologists. https://doi.org/10.1105/tpc.18.00016.

Zumkeller, Simon Maria, Philipp Gerke, and Volker Knoop. 2020. "A Functional Twintron, 'zombie' Twintrons and a Hypermobile Group II Intron Invading Itself in Plant Mitochondria." Nucleic Acids Research 48 {5): 2661–75. https://doi.org/10.1093/nar/gkz1194.

Zumkeller, Simon Maria, and Volker Knoop. 2022. "Categorizing 161 Plant Mitochondrial Group II Introns into 29 Families of Related Paralogues Finds Only Limited Links between Intron Mobility and Intron-Borne Maturases." https://doi.org/10.21203/RS.3.RS-2086450/V1.

Zumkeller, Simon Maria, Volker Knoop, and Nils Knie. 2016. "Convergent Evolution of Fern-Specific Mitochondrial Group II Intron Atp1i361g2 and Its Ancient Source Paralogue Rps3i249g2 and Independent Losses of Intron and RNA Editing among Pteridaceae." Genome Biology and Evolution 8 (8): 2505–19. https://doi.org/10.1093/gbe/evw173.

Zumkeller, Simon Maria, Monika Polsakiewicz, and Volker Knoop. 2022. "Rickettsial DNA Invasions and a Scrambled RRNA Cluster with a Trans-Splicing Group I Intron: The Highly Unorthodox Mitogenome of the Fern Haplopteris Ensiformis." https://doi.org/10.21203/RS.3.RS-2152556/V1.

Zupok, Arkadiusz, Danijela Kozul, Mark Aurel Schöttler, Julia Niehörster, Frauke Garbsch, Karsten Liere, Axel Fischer, et al. 2021. "A Photosynthesis Operon in the Chloroplast Genome Drives Speciation in Evening Primroses." The Plant Cell 33 {8): 2583–2601. https://doi.org/10.1093/PLCELL/KOAB155

7.0 Appendix

Publication 1: A functional twintron, “zombie” twintrons and a hypermobile group II intron invading itself in plant mitochondria

Nucleic Acids Research; 2019 Volume 48, Issue 5, 18 March 2020, Pages 2661–2675,
https://doi.org/10.1093/nar/gkz1194
Simon Zumkeller, Philipp Gerke and Volker Knoop

IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn,
Kirschallee 1, D-53115 Bonn
Corresponding author
Email: volker.knoop@uni-bonn.de

Key words: Group I introns, Group II introns, Twintrons, Maturases, Intron retrotransposition Key message: The discovery of twintrons in land plant mitochondria documents group II intron retromobility in recent evolutionary times.

Published online 9 January 2020

Nucleic Acids Research, 2020, Vol. 48, No. 5 ***2661–2675***
doi: 10.1093/nar/gkz1194

A functional twintron, 'zombie' twintrons and a hypermobile group II intron invading itself in plant mitochondria

Simon Zumkeller, Philipp Gerke and Volker Knoop*

IZMB – Institut fu¨r Zellula¨re und Molekulare Botanik, Abteilung Molekulare Evolution, Universita¨t Bonn, Kirschallee 1, D-53115 Bonn, Germany

Received September 25, 2019; Revised November 26, 2019; Editorial Decision December 06, 2019; Accepted December 11, 2019

ABSTRACT

The occurrence of group II introns in plant mitochondrial genomes is strikingly different between the six major land plant clades, contrasting their highly conserved counterparts in chloroplast DNA. Their present distribution likely reflects numerous ancient intron gains and losses during early plant evolution before the emergence of seed plants. As a novelty for plant organelles, we here report on five cases of twintrons, introns-within-introns, in the mitogenomes of lycophytes and hornworts. An internal group II intron interrupts an intron-borne maturase of an *atp9* intron in Lycopodiaceae, whose splicing precedes splicing of the external intron. An invasive, hypermobile group II intron in *cox1,* has conquered nine further locations including a previously overlooked *sdh3* intron and, most surprisingly, also itself. In those cases, splicing of the external introns does not depend on splicing of the internal introns. Similar cases are identified in the mtDNAs of hornworts. Although disrupting a group I intron-encoded protein in one case, we could not detect splicing of the internal group II intron in this 'mixed' group I/II twintron. We suggest the name 'zombie' twintrons (half-dead, half-alive) for such cases where splicing of external introns does not depend any more on prior splicing of fossilized internal introns.

INTRODUCTION

Group II introns are among the most interesting RNA structures in the living world. Most prominent is their likely role as ancestors of the eukaryotic spliceosome machinery, an evolutionary connection for which ever more convincing biochemical and structural evidence is being identified (1–8). Generally, group II introns are connected with attributes like 'autocatalytic', 'self-splicing', 'mobile genetic elements' or 'ribozymes'. However, these intriguing features have been demonstrated for only few members of the class, notably so for some intensively studied group II introns from yeast mitochondria or from eubacteria. In contrast, and although group II introns are remarkably abundant and diverse in plant mitochondria and chloroplasts, none of them has yet been proven to be autocatalytic, self-splicing or mobile.

In particular with respect to the group II introns contained in their genomes, the two endosymbiotic organelles in plants have evolved in very different ways (9). Whereas the overall stability of the chloroplast genome makeup is also reflected in the conserved introns, the plant mitochondrial intron complements in contrast differ significantly among the six major land plant clades—the liverworts, mosses, hornworts, lycophytes, ferns and seed plants. Notably, the characteristic differences between the sets of their conserved mitochondrial introns had impacts on phylogenetic concepts, particularly of early land plants (10–15). In fact, not a single one of >100 mitochondrial intron insertion sites meantime identified in bryophytes had been found to be shared by all the three divisions – liverworts, mosses and hornworts—at the same time (16). In the absence of convincing sequence similarities between intron paralogues, however, any speculations on lateral intron movements explaining this diversity remained moot. A very ancient history of intron gains by retro-copying, followed by lineage-specific losses, would be difficult to trace after long evolutionary time periods obliterating clear sequence similarities between intron paralogues in extant taxa. Nevertheless, such support for the concept of retro-copying events creating group II intron diversity in plant mitochondria has recently come from the characterization of two fern-specific introns very likely originating from ancestral paralogues in plant evolution (17,18).

The transition from an early gametophyte-dominated, bryophyte-type lifestyle to a sporophyte-dominated, tracheophyte-type lifestyle came along with significant

*To whom correspondence should be addressed. Tel: +49 228 73 6466; Fax: +49 228 73 6467; Email: volker.knoop@uni-bonn.de

2662 Nucleic Acids Research, 2020, Vol. 48, No. 5

changes in plant mitochondrial genomes (19). Extant lycophytes, comprising the orders Lycopodiales (club mosses), Selaginellales (spike mosses) and Isoetales (quillworts), represent the most ancient surviving lineages of vascular plants and show a particularly wide range of variation among their mtDNAs. The mitogenome of *Phlegmariurus squarrosus* (Lycopodiales) has a rich gene complement, maps as a circular, non-recombining DNA molecule like in bryophytes and retains several ancient gene syntenies (20). In contrast, the mtDNAs of *Isoetes engelmannii* (Isoetales) and *Selaginella moellendorffii* (Selaginellales) differ dramatically (21,22). Their gene complements are significantly reduced, transcripts are affected by very abundant RNA editing exchanging pyrimidines and massive DNA recombination has not only obliterated ancient gene syntenies but also disrupted intron continuities in some cases to create *trans*-splicing intron arrangements. Among bryophytes, a comparable reduction of mitochondrial gene contents owing to endosymbiotic gene transfer (EGT) and a rise in RNA editing frequencies has occurred in parallel only among hornworts (23–27).

The above phenomena often complicate proper gene annotations and we occasionally find that organelle genome accessions require re-consideration of annotated gene features. Hence, we systematically re-investigated mitochondrial genome sequences of lycophytes and hornworts, which ultimately led us to identify yet a further molecular idiosyncrasy, the presence of 'twintrons'—introns within introns. The term twintron had originally been coined upon their initial discovery in the chloroplast genome of *Euglena gracilis,* originating from a secondary endosymbiosis in this unicellular alga (28–30). The *E. gracilis* chloroplast DNA twintrons combine group II introns and group III introns among and between each other, the latter being a unique type of 'degenerated' group II introns of phylogenetically restricted occurrence. Importantly, splicing of the internal intron was shown to precede splicing of the external intron, likely to reconstitute relevant structures in the latter as a prerequisite for splicing. Such and other issues of twintron arrangements *sensu lato* versus *sensu stricto* have been summarized in an excellent recent review (31).

Here we report on five cases of twintrons *sensu stricto* as a novum in multicellular photosynthetic organisms, occupying different loci in the mitochondrial genomes of Lycopodiales and hornworts. Notably, the 'invading', internal group II introns share significant sequence similarities with intron paralogues in other locations, likely documenting their evolutionary origins. We find that prior splicing of the internal intron is required when affecting the reading frame of an intron-encoded maturase in the external intron but may alternatively be obsolete in other cases when splicing of the external intron is not affected by the internal intron. The degeneration of the internal intron under retention of active splicing of the external intron can be traced through hornwort evolution. An invasive, 'hypermobile' group II intron (cox1i1149g2) in Lycopodiaceae even gave rise to two twintrons – as an internal intron inserted into itself and into a newly identified *sdh3* intron – and furthermore to seven additional 'intron fossils' in intergenic regions.

MATERIALS AND METHODS

Plant material and sequence data

Lycopodiaceae plant material was obtained from the Bonn University Botanical Garden: *Phlegmariurus hippuris* (accession xx-0-BONN-17383), *Phlegmariurus tardieuae* (accession xx-0-BONN-29472) and *Palhinhaea glaucescens* (accession ZW-0-BONN-16644). Sequence analyses were further complemented with Illumina whole genome sequences obtained in our laboratory for *Phlegmariurus hippuris* and with assemblies made from sequence reads available for *Huperzia selago* under BioProject accession number PRJNA281995. The WGS data of *P. hippuris* was produced by BGI on Illumina platform HiSeq 2500/4000. *De novo* assembly of *P. hippuris* mtDNA was performed with the assembly program MEGAHIT (32). SRA-data of *Huperzia selago* (BioProject accession number PRJNA281995) was downloaded from NCBI using the SRA-toolkit (33) and assembly was conducted with the assembly program Trinity (34). Sequences of interest were extracted from the assemblies using the BLAST+ suite (35). Observed sequence differences for the loci investigated here match the taxonomic differences of Lycopodiaceae species (36). The newly obtained DNA and cDNA sequence data were submitted under accession numbers LR721677–LR721682, LR722602–LR722612, LR722615 and LR722624, respectively.

Molecular cloning

Plant nucleic acids were prepared based on established protocols employing CTAB (cetyl-methyl-ammonium bromide) as a detergent for cell lysis after grinding of frozen plant material (37,38). DNA of *P. hippuris* was isolated using the Qiagen DNeasy Kit for plants for WGS. For twintron-splicing analyses in particular, the NEB Monarch RNA Miniprep Kit was used to obtain RNA of higher amount, quality and purity. The synthesis of cDNA included random hexamer oligonucleotides and target-specific primers in parallel. PCR amplification was done using specific primers and Go-Taq polymerase (Thermo Fisher Scientific) and alternative oligonucleotide combinations with primers targeting flanking exon or intron regions and/or enrichment of splicing intermediates. In parallel, treatment with RNaseR (Lucigen) was performed to reduce concentration of non-circular RNAs and to ideally increase intron lariat concentrations, respectively (see supplementary figure S1). Oligonucleotide sequences are given in supplementary table S1. PCR products were isolated applying the NucleoSpin Extract II Kit (Macherey-Nagel), cloned by ligation into the pGEM-T easy vector system (Promega) and transformed via heat-shock into XL1-blue *E. coli* cells (Agilent). Three replicates were sequenced by Sanger sequencing (Macrogen Europe).

Further bioinformatic analyses

Sequences were handled and aligned using the alignment explorer feature of MEGA 7 (39). Alignments are available from the authors upon request. Alignment shading was performed using the GeneDoc alignment editor version 2.7.0 (Nicholas KB and Nicholas HB, 2006, www.psc.

edu/biomed/genedoc). Phylogenetic trees were constructed with MEGA7 (39) using the Maximum Likelihood method and the GTR+G+I model of sequence evolution. RNA secondary structures of introns were determined manually following group I and group II intron consensus structures (40–42), following recommendations for identification and labeling of group II intron signature sequence elements under http://webapps2.ucalgary.ca/~groupii. RNA secondary structure displays were created making use of the VARNA software (43). Labeling of Maturase domains followed recent structural insights (44,45).

RESULTS

Nomenclature issues

In the following, we will use the previously proposed nomenclature to clearly designate the numerous and diverse plant organelle introns (46,47). Briefly, intron labels combine the host gene name, followed by an 'i' and the position upstream of the intron insertion site (using the liverwort *Marchantia polymorpha* as a reference if not indicated otherwise) and finally a designation to distinguish intron types, i.e. 'g2' for group II and 'g1' for group I introns, respectively. As an example, 'atp9i87g2' designates a group II intron in the *atp9* gene, located behind position 87 of the *atp9* reading frame (hence, in intron phase 0 behind codon number 29).

The here reported discovery of twintrons in plant mitochondria prompted us to extend the nomenclature to the respective internal introns, for which we suggest using the label 'ii'. Accordingly, atp9i87g2ii1114g2 refers to the internal group II intron inserted after nucleotide position 1114 of the external group II intron atp9i87g2. As a designation for a complete twintron arrangement, we suggest adding '-twin' behind the label for the external intron to clearly distinguish the (primary) external intron (atp9i87g2), the (secondary) internal intron (atp9i87g2ii1114g2) and the joint twintron arrangement (atp9i87g2-twin). A systematic nomenclature has also been proposed to label maturases encoded within group II introns according to their host intron (48). We here furthermore suggest adding a label for the intron type and an additional 'c' in cases where the maturase frame is continuous with the respective upstream exon. The case of *mat-atp9i87g2c* discussed below is a typical example.

Group II intron atp9i87g2 is a twintron conserved among Lycopodiales

As a unique case, mitochondrial group II intron atp9i87g2 is now found to be conserved between liverworts, mosses, hornworts and lycophytes (Figure 1A), most parsimoniously explained by an early gain in land plants followed by a loss in the stem lineage of euphyllophytes comprising ferns and seed plants. At least one further independent loss of atp9i87g2 has occurred among hornworts given its absence in the mtDNAs of *Nothoceros aenigmaticus* (49) and *Phaeoceros laevis* (24). The intron-encoded maturase *mat-atp9i87g2c* was found to be conserved between the liverwort *Marchantia polymorpha* and the lycophyte *Isoetes engelmannii* (50). However, *mat-atp9i87g2c* is degenerated to variable degrees into a pseudogene or remaining pseudogene fragments only in the available mtDNAs of mosses and hornworts (Figure 1A).

Re-inspecting mitochondrial genome sequence entries, we initially noted a significant size increase of atp9i87g2 to 5756 bp in the mtDNA of the lycophyte *Phlegmariurus squarrosus* (previously named *Huperzia squarrosa*, 20). Moreover, the *P. squarrosus* atp9i87g2 sequence showed distinct homologies to intron-encoded maturases separated by a sequence insertion. A re-analysis of the sequence ultimately now shows that an ectopic group II intron has invaded atp9i87g2 behind intron position 1114 (Figure 1B). This internal intron in the *P. squarrosus* mtDNA, accordingly labelled atp9i87g2ii1114g2 as suggested above (Figure 2), disrupts the maturase reading frame (Supplementary Figure S2) of the outer 'host' group II intron atp9i87g2, creating a twintron.

To investigate the conservation of the peculiar twintron arrangement, we retrieved homologous loci in related taxa of available Lycopodiaceae. Indeed, we found the *atp9* twintron arrangement conserved in the close sister species *Phlegmariurus tardieuae* and *P. hippuris* also in more distant taxa of other genera in the Lycopodiaceae: *Huperzia selago* and *Palhinhaea glaucescens*. Given the high degree of conservation of both the group II intron structure (Figure 1B) and the maturase reading frame (Supplementary Figure S2) disrupted by the internal group II intron sequence, we assumed that splicing of the internal intron could be a prerequisite for splicing of the external atp9i87g2 intron. Indeed, we could confirm a serial splicing of the internal intron prior to the external intron, which is likely necessary to first recreate the continuous maturase reading frame of the external intron. Concomitantly, we detected four events of C-to-U editing in the external intron of *Palhinhaea glaucescens* (Supplementary Figure S2).

The internal intron atp9i87g2ii1114g2 of Lycopodiaceae (Figure 2) is significantly similar in sequence to two other mitochondrial group II introns, including rps14i114g2 present in the *rps14* gene of the *Phlegmariurus squarrosus* mitogenome (Supplementary Figure S3). Intriguingly, rps14i114g2 is also present in the mtDNAs of liverworts, making it a very good candidate for an ancient intron that may have been the donor that gave rise to the twintron arrangement in *atp9* of Lycopodiaceae. Notably, sequence similarities between rps14i114g2 in *Phlegmariurus* and liverworts are higher than for other shared intron orthologues like cox3i171g2 and nad3i140g2. A third group II intron paralogue related in sequence is nad7i1113g2 (Supplementary Figure S3). Evolutionary relationships remain unclear, however, given that the *nad7* gene is degenerated among most liverworts (51) and lost altogether from the mtDNA in hornworts and *Phlegmariurus* (20).

Jumping into itself: group II intron twintron cox1i1149g2 in Lycopodiales

The confirmation of atp9i87g2 as a functional twintron led us to investigate the Lycopodiaceae mtDNA more closely. Group II intron cox1i1149g2 had been identified exclusively in the mtDNAs of *Phlegmariurus squarrosus* (20) and *Selaginella moellendorffii* (22), but never outside of lycophytes.

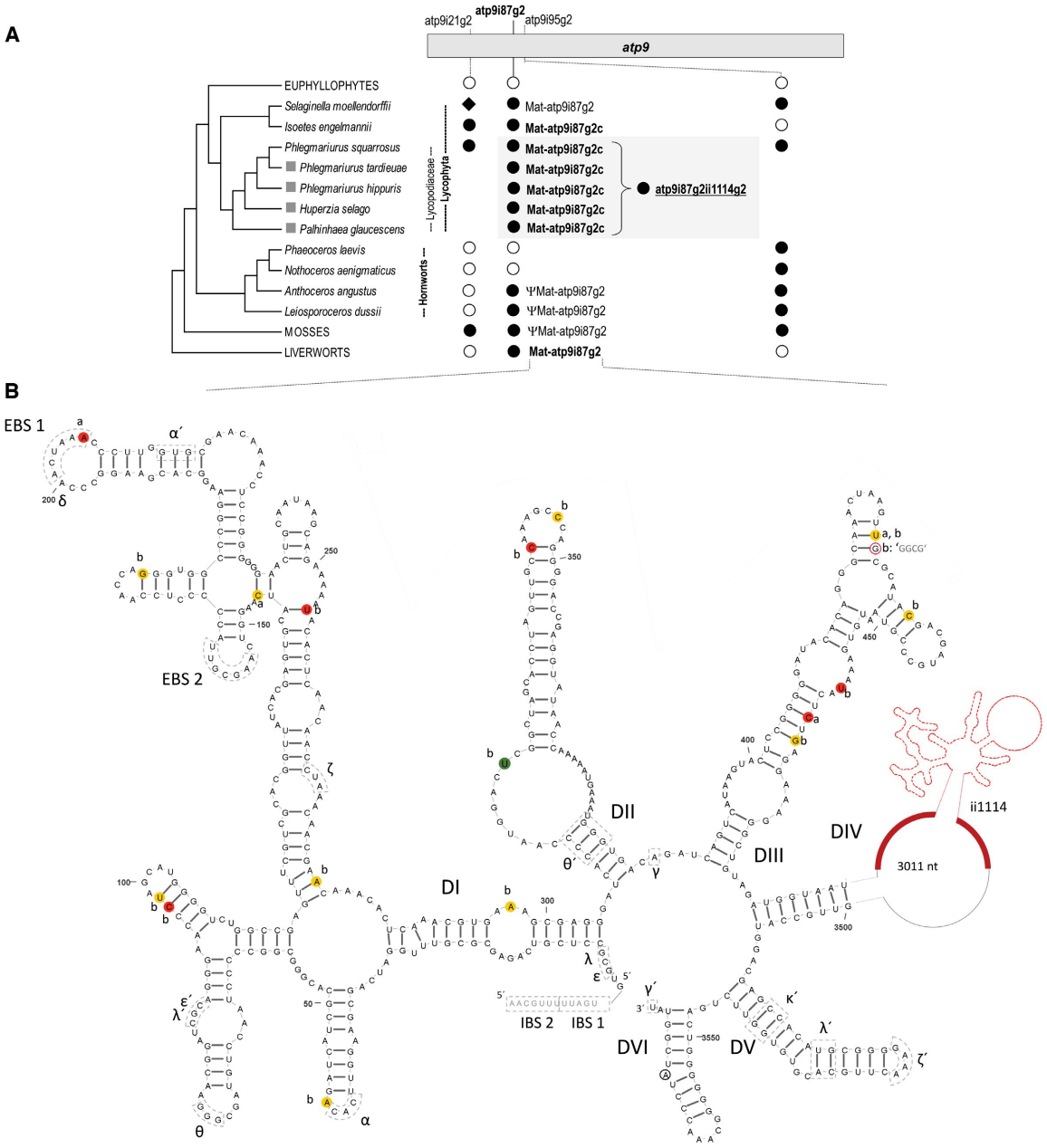


Figure 1. Group II intron atp9i87g2 is a twintron in Lycopodiaceae. (**A**) Presence or absence of group II introns atp9i21g2, atp9i87g2 and atp9i95g2 in the mtDNAs of different plant taxa is indicated by filled or open circles, respectively. Note that *Phlegmariurus squarrosus* had previously been named *Huperzia squarrosa*. Intron atp9i21g2 is disrupted, requiring *trans*-splicing (rhombus) in *Selaginella moellendorffii*. No introns have been identified in the *atp9* genes of ferns or seed plants (euphyllophytes). Intron atp9i87g2 is most widely conserved in the early-emerging embryophyte lineages. It retains an intact intron-encoded maturase (bold) in liverworts and lycophytes that is in frame with *atp9* (mat-atp9i87g2c) except in *Selaginella moellendorffii* (48). Only degenerated pseudogene traces remain in the other taxa (ΨMat). The maturase is now found to be disrupted by internal intron atp9i87g2ii1114g2 (underlined) in *Phlegmariurus* and other Lycopodiaceae investigated here (gray squares next to taxon names). (**B**) Secondary structure of group II intron atp9i87g2 in *Phlegmariurus squarrosus,* now characterized as a serially splicing group II twintron. Roman numerals indicate the six typically conserved domains of group II introns (DI-DVI) and stippled grey lines with Greek letters designate tertiary interactions, respectively (6). EBS1/2 designates exon/intron binding site interactions. The internal group intron (atp9i87g2ii1114g2, see Figure 2) inserted in domain IV (red icon symbolizing a group II intron consensus fold) is found conserved among Lycopodiaceae and disrupts the likewise highly conserved maturase reading frame *mat-atp9i87g2c* (see Supplementary Figure S2), continuous with the upstream coding sequence. Shading highlights transitions (yellow) and transversions (orange) occurring in the closely related species *P. hippuris* (a) and in the distantly related species *Palhinhaea glaucescens* (b) now investigated, respectively. Maturation of the external intron includes four C-to-U edits in the *P. glaucescens* maturase (see Supplementary Figure S2), including a silent exchange in domain II (green shading).

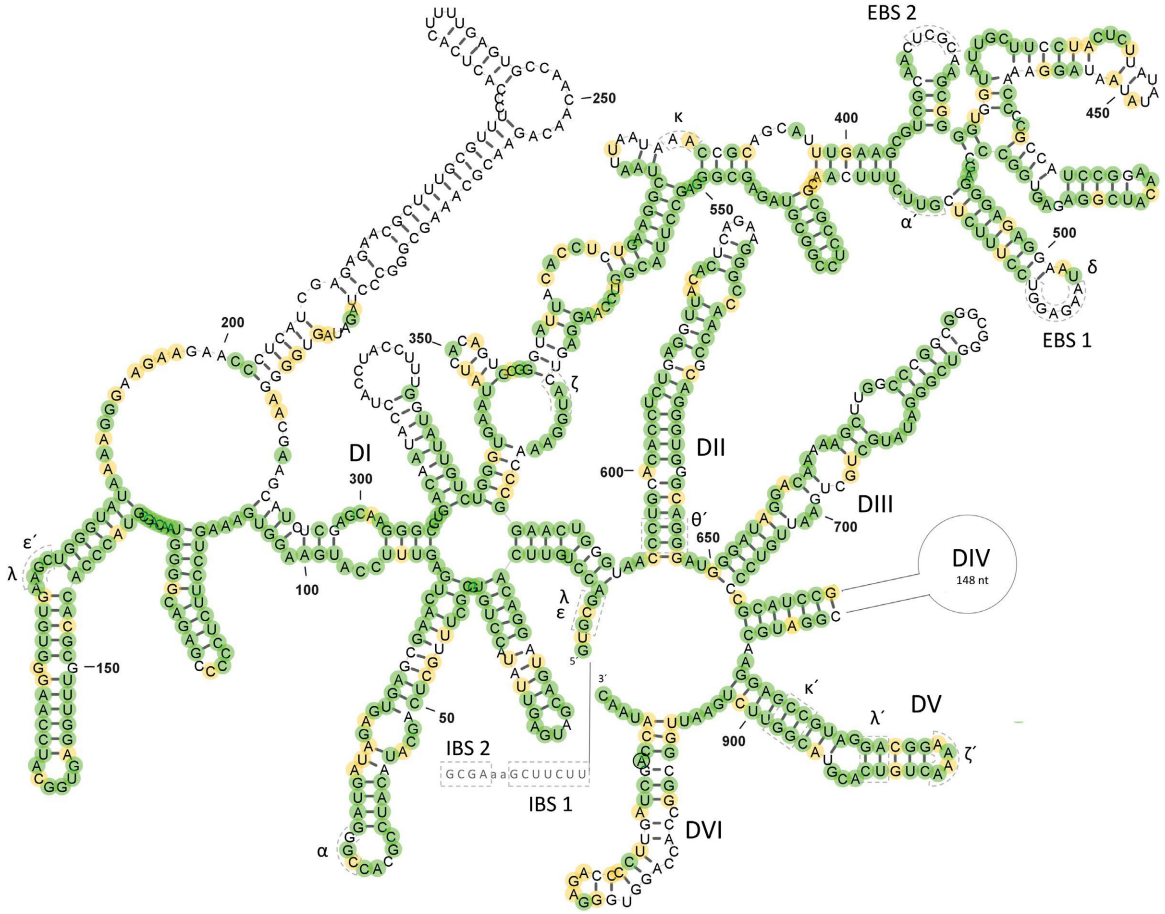


Figure 2. The now identified internal intron atp9i87g2ii1114g2 in Lycopodiaceae shares extensive sequence similarity with intron paralogues rps14i114g2 and nad7i1113g2 (Supplementary Figure S3). Shown is the secondary structure of atp9i87g2ii1114g2 in *Phlegmariurus squarrosus* with nucleotide identities or transitions shared among all three paralogues highlighted in green and yellow, respectively. Annotation of group II intron features is like in figure 1. The IBS2–IBS1 sequence directly upstream of the internal intron structure is located in domain IV of the external atp9i87g2 intron (accordingly ending with nucleotide 1114 of the latter).

Given a significantly larger size of cox1i1149g2 in *Phlegmariurus* compared to *Selaginella* (3594 versus 1941 bp) we carefully re-inspected the *Phlegmariurus* sequence. This revealed cox1i1149g2 in *Phlegmariurus* to be a twintron carrying an internal group II intron located in domain IV, now to be labelled cox1i1149g2ii652g2 (Figure 3). As in the case of atp9i87g2-twin outlined above, this twintron arrangement is likewise conserved among the other Lycopodiaceae species included in our study, again suggesting an early emergence of cox1i1149g2-twin within the family. Most notably, the sequence of the internal intron is highly similar to the external intron (Figure 3), suggesting that in this particular case an intron has been inserted into itself to create this peculiar twintron arrangement.

We found correct splicing of cox1i1149g2 as expected in the Lycopodiaceae. However, different from the case of atp9i87g2-twin we could not detect prior splicing of the internal intron cox1i1149g2ii652g2 (Supplementary Figure S1). Intriguingly, and despite striking overall sequence conservation between the internal and external intron of cox1i1149g2, a sequence change between the functionally essential domains V and VI has likely resulted in a base-pairing shift in the latter (Figure 3). This causes loss of the essential bulged adenosine residue necessary for lariat formation as a possible cause for a defect in splicing of the internal intron. Additionally, the EBS-IBS interactions necessary for splice site determination are weakened by mutations (Figure 3). We suggest the term ‘zombie’ twintron for such cases of half-dead and half-alive twintrons lacking a splicing of the internal intron that does, however, not impede splicing of the external intron. In this case, the internal intron has degenerated into a large and functionally dead sequence insertion in domain IV of the external intron.

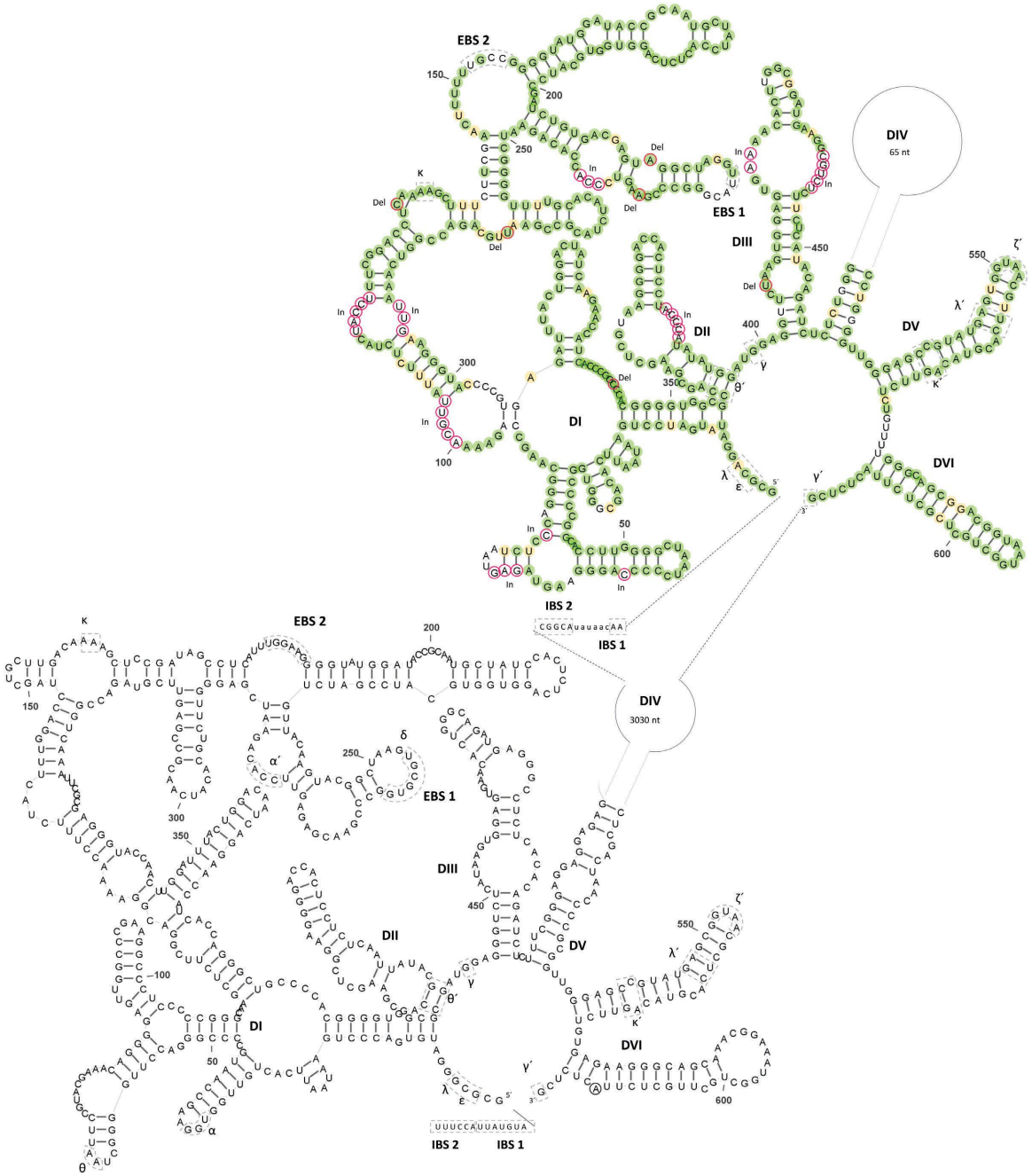


Figure 3. The cox1i1149g2 zombie twintron in lycophytes. Group II intron cox1i1149g2 is unique to lycophytes, previously identified only in the chondromes of *Selaginella moellendorffii* and *Phlegmariurus squarrosus*. Secondary structure modeling now characterizes cox1i1149g2 as a unique twintron in *P. squarrosus* and in the other Lycopodiaceae now investigated, in which the intron has jumped into itself creating internal intron cox1i1149g2ii652g2. Green and yellow shading highlights identical nucleotides and transitions, respectively. Red circles indicate inserted nucleotides (In) or deletions (Del) behind the labeled nucleotide in the internal intron in comparison to the external intron. Designation of intron domains and tertiary interactions is as in figure 1B. Notable differences between the internal and external intron likely rendering the former dysfunctional are absence of a characteristic bulged adenosine in domain VI resulting from a sequence alteration behind domain V and significantly weaker EBS–IBS interactions.

Hypermobility creating multiple cox1i1149g2 insertions

We identified multiple additional sequences with significant similarities to the internal and external intron copies of cox1i1149g2-twin within intergenic sequences (IGS) of the *P. squarrosus* mitogenome (Supplementary Figure S4). We assume that an ancestrally invasive, hypermobile cox1i1149g2 intron gave rise to these paralogous sequence copies (Figure 4), because sequence similarities do not extend beyond the clearly recognizable 5′- and 3′-ends typical for group II introns (Supplementary Figure S4). Identical arrangements with high sequence similarity (>99%) were found in the independent *de novo* assembly of the *P. hippuris* mtDNA in the same eight intergenic regions. These observations strongly argue for ancient intron mobility by retrotransposition rather than DNA recombination, as typically observed for vascular plant mtDNAs, as the source for the multiple sequence copies. Intriguingly, the orientation of intron insertions always coincides with that of the flanking genes with the only exception of the *cox3-trnF*-IGS (Figure 4). We could not detect splicing (mostly not even transcripts) in any of these cases of intron paralogues in intergenic regions, leaving the question open whether they have ever been functional introns before disintegration into molecular fossils. A particularly intriguing case is the cox1i1149g2 copy in the *trnI-trnQ* spacer occupying most of the intergenic region with its splicing being a possible requisite for maturation of the flanking tRNAs. However, also in this case, no intron splicing could be detected.

As in the above case of the *atp9* twintron possibly originating from insertion of a rps14i114g2 copy, we strived to find a possible origin for cox1i1149g2 and its descendants. Most similar among group II intron paralogues is *nad1* intron nad1i669g2, widely conserved among tracheophytes and thus very likely evolutionarily older than the lycophyte-specific intron as a candidate donor locus (Figure 4). A phylogenetic analysis supports the idea that cox1i1149g2 may have originated as a reverse-spliced copy of nad1i669g2, followed by further copying processes into the other nine locations (Supplementary Figure S4B), including the creation of two zombie twintrons. Other than the seven intergenic occurrences, one additional case of cox1i1149g2 sequence similarity suggested to reconsider the previous annotation of the *sdh3* gene resulting in an early stop codon in the *Phlegmariurus squarrosus* mtDNA.

The case of sdh3i349g2-twin in Lycopodiaceae

The conserved carboxyterminal part of the *sdh3* coding sequence appeared to be missing in the mtDNA of *Phlegmariurus squarrosus.* This did not cause suspicion given that *sdh3* and *sdh4* encoding subunits of the succinate dehydrogenase (complex II), are frequently subject to endosymbiotic gene transfer (52). Upon closer inspection of the downstream *sdh3* gene region in *P. squarrosus*, however, we identified a new group II intron to be labeled sdh3i349g2, the splicing of which would provide the previously missing conserved C-terminal coding region of *sdh3*. Intriguingly, the newly identified sdh3i349g2 intron also turned out to be a twintron, created by one additional ectopic insertion of the invasive cox1i1149g2 intron (see Figure 4), here creating the internal intron sdh3i349g2ii42g2 located in domain I of its host intron (Supplementary Figure S5).

Like in the two previous cases, this twintron arrangement is also conserved among the related Lycopodiaceae genera here investigated, again suggesting an early origin within the family. Moreover, not only functional splicing of sdh3i349g2-twin could again be confirmed, but the cDNA sequences also revealed expected events of C-to-U RNA editing, further confirming *sdh3* as a functional mitochondrial gene in the Lycopodiaceae.

Sequence conservation with the functional external intron of cox1i1149g2-twin in this case included the bulged adenosine in domain VI and did also not reveal any other evident degeneration of functionally important intron elements (Supplementary Figure S4A). Interestingly, mutations within the EBS-IBS interaction region are improving complementary binding to the ‘new’ insertion site compared to the likely ancestral paralog cox1i1149g2. However, like in the case of cox1i1149g2-twin, we could again not detect any splicing of the internal intron to occur prior to external intron splicing rendering also this case a zombie twintron.

Intron atp1i1050g2 is a zombie twintron in hornworts, but not in liverworts

Altogether six different group II introns have been identified in the plant mitochondrial *atp1* gene, showing a typical clade-specific distribution (Supplementary Figure S6). Only one group II intron in *atp1* is shared between two major plant clades: atp1i1050g2 is present in hornworts and liverworts. The atp1i1050g2 intron carries an intact maturase reading frame (*mat-atp1i1050g2c*) when present in liverworts. The atp1i1050g2 homologue in hornworts is significantly extended in size, again prompting a closer inspection. Secondary structure modeling revealed a twintron arrangement also in this case (Supplementary Figure S7). The internal intron atp1i1050g2ii1536g2 (*Marchantia polymorpha* reference numbering, see below) shares very high sequence identity with group II intron paralogue cox2i98g2, so far exclusively identified in the hornwort genus *Anthoceros*.

Comparing the *Anthoceros* twintron with its homologues in other hornwort genera reveals significant degeneration of the internal intron with deletion at the 5′-end in *Leiosporoceros* and at the 3′-end in *Phaeoceros* and *Nothoceros* especially within domain I where EBS1 and EBS2 should be located (Supplementary Figure S8). Alignment of the internal intron, its paralogue cox2i98g2 and comparing flanking sequences with the liverwort intron allows to clearly place its insertion site in the middle of a KTI’RGE peptide motif in the ancient maturase, corresponding to an in-frame insertion between amino acid codons 512 and 513 of the *Marchantia polymorpha* mat-atp1i1050g2 homologue (Supplementary Figure S8). Intron atp1i1050g2-twin is highly conserved in the recently studied *A. agrestis* chondrome (53) and splicing is clearly confirmed with our cDNA work. Likely owing to degeneration of several intron motifs, we could not find evidence for splicing of the internal intron before splicing of the external intron in *Anthoceros*, however. Such a splicing event could in any case not reconstitute an

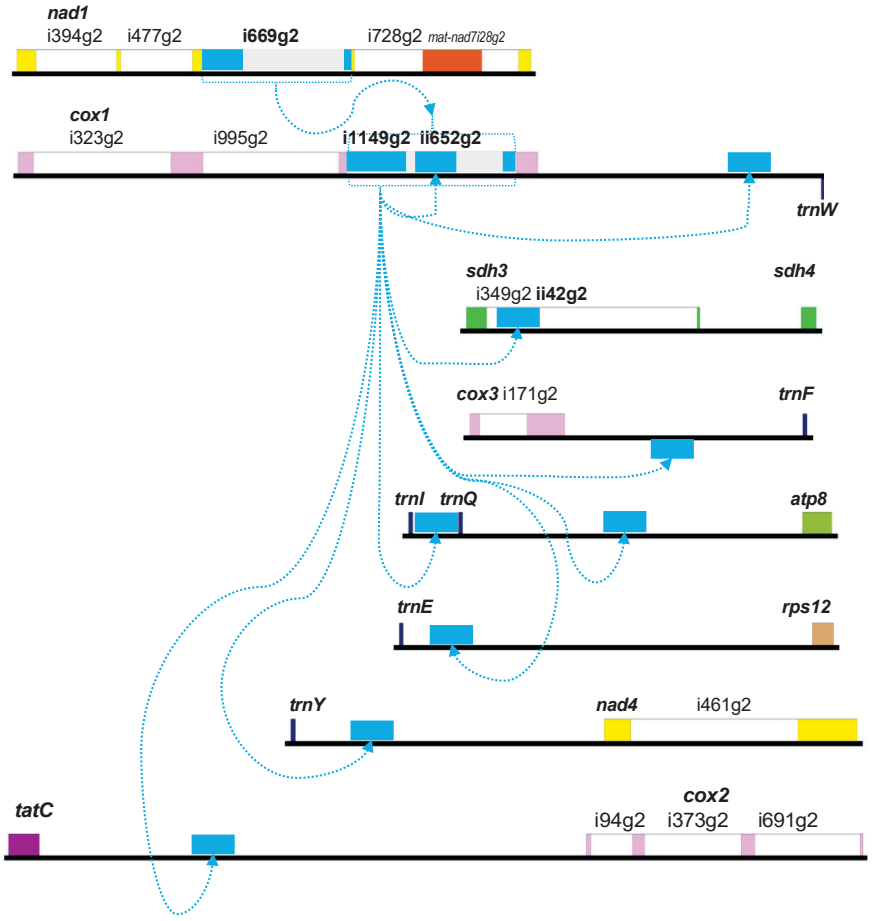


Figure 4. An invasive, hypermobile group II intron. Sequences with significant similarities to the internal and external group II intron of cox1i1149g2-twin (Figure 3) occur multiple times elsewhere in the *Phlegmariurus squarrosus* mtDNA (Supplementary Figure S4). Group II intron nad1i669g2 conserved in vascular plants may have been the ancestral donor initially creating the highly invasive intron paralogue that gave rise to multiple paralogues (light blue boxes and stippled arrow lines). In one case, a second twintron has been created: sdh3i349g2-twin containing sdh3i349g2ii42g2 as an internal intron. The other cases are copies of the donor intron in seven intergenic sequences (IGS): *cox1-trnW, cox3-trnF, trnI-trnQ, trnQ-atp8, trnE-rps12, trnY-nad4* and *tatC-cox2*. Intron orientation is always in the direction of flanking genes except for the case of the *cox3-trnF* IGS copy (blue box below the line). Gene displays have been created with OGDRAW using default coloring for gene categories (*atp* dark green, *cox* pink, *mat* orange, *nad* yellow, *rps* light brown *sdh* light green and *trn* dark blue).

intact maturase in atp1i1050g2 of the hornworts, where the maturase reading frame is further degenerated by multiple frameshifts and larger indels in all taxa.

A mixed group-II-in-group-I twintron in *Anthoceros*

The striking sequence similarities between the internal intron of atp1i1050g2-twin to cox2i98g2 (Supplementary Figure S7) extends to further mitochondrial group II intron paralogues in hornworts, including yet another twintron structure. In this case, a group II intron is located within a group I intron, cox1i1116g1, which is shared between *Anthoceros* and liverworts. Secondary structure modeling reveals that the invading group II intron (to be labeled cox1i1116g1ii209g2) is inserted in domain P5 of the external group I intron (Figure 5). Again, we were able to confirm conservation of this unique mixed twintron arrangement in the mtDNA of the sister taxon *A. agrestis* and to detect splicing of the external intron but no prior splicing of the internal intron, rendering also this case a zombie twintron.

Group I intron cox1i1116g1 in liverworts encodes a protein with similarity to LAGLIDAG-type endonucleases (54,55), whose reading frame continues after codon 372 of the upstream *cox1* reading frame. Significant similarity to the liverwort endonuclease ORF is evident both upstream and downstream of the internal intron in cox1i1116g1-twin of *Anthoceros*. However, even if functionally spliced, the endonuclease ORF could not be re-established in *Anthoceros* owing to reading frame shifts beyond its disruption by the internal group II intron, similar to the above case of the maturase reading frame in atp1i1050g2.

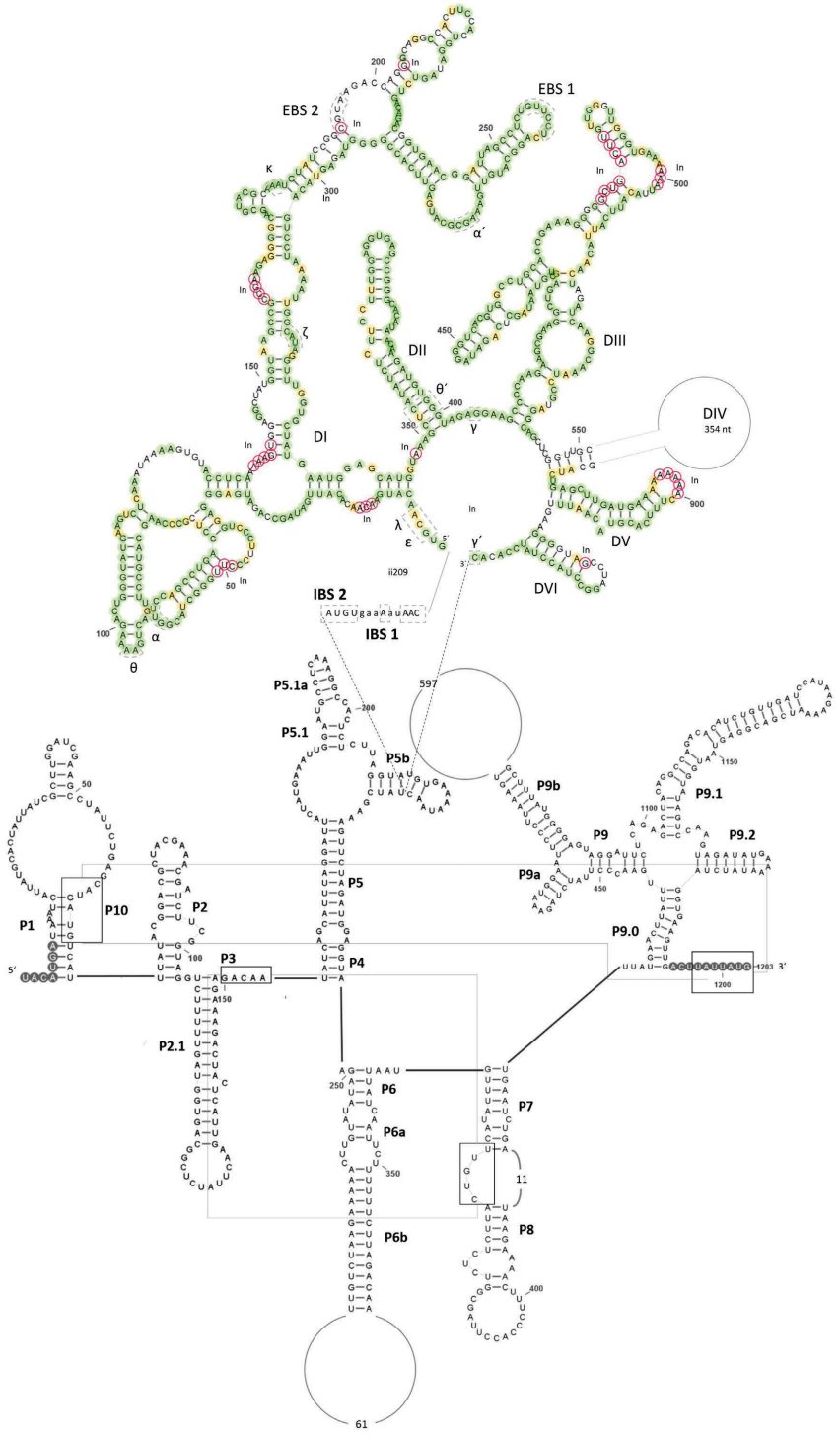


Figure 5. A mixed group I/group II twintron. Secondary structure modeling of group I intron cox1i1116g1 in *Anthoceros angustus* (database accession NC_037476) reveals a twintron arrangement with an internal group II intron (cox1i1116g1ii209g2) inserted in the P5 structure of the external group I intron. Highlighted in green and yellow, respectively, are nucleotide identities and transitions in comparison to cox2i98g2 with red circles indicating indel mutations as previously used in figure 3.

Evolutionary origins of twintrons and other intron paralogues

The exceptionally high similarity of atp1i1050g2ii1536g2 with cox2i98g2 in *Anthoceros* suggests recent copying mechanisms by reverse splicing events. Aside from these two paralogues and the second twintron case now identified in hornworts, cox1i1116g1ii209g2, the sequence similarities additionally extend to three further group II introns, likewise exclusively present in hornworts: nad5i881g2, nad6i444g2 and nad9i502g2. This group of six hornwort paralogues is also related to group II intron nad7i676g2 present in tracheophytes and, more distantly, to cox2i97g2 in liverworts (Figure 6). Notably, the hornwort *cox2* intron paralogue cox2i98g2 is clearly distinct from the latter, being inserted 1 bp further downstream in the *cox2* reading frame. We could not find any evidence for 'intron shifting' as the underlying cause for creation of the two paralogues, as e.g. recently documented elsewhere (56). The disparate occurrence of cox2i97g2 and cox2i98g2 in the two bryophyte clades clearly suggests that they do not share a recent common origin and the phylogenetic analysis supports this idea (Figure 6). The six hornwort group II intron paralogues are placed in one well-supported clade, presumably indicating lateral mitochondrial intron mobility within this plant clade. Intron cox2i98g2, exclusively present in *Anthoceros,* is the likely most recent emerging paralogue nested in a paraphyletic clade of the more widely distributed atp1i1050g2ii1536g2 (Figure 6). The case of the *cox1* twintron likewise restricted in presence to the genus *Anthoceros* is a more ambiguous issue because not only the internal intron but the entire host intron shared with liverworts (cox1i1116g1) is absent in the three other hornwort genera. Deducing a completely resolved time series of the retro-copying events for the related introns is ultimately difficult, given that only nad9i502g2 fully reflects the expected hornwort species phylogeny (57,58). The clustering of *Leiosporoceros* paralogues nad5i881g2 and nad6i444g2 is notable (Figure 6), possibly indicating convergent evolution as recently described for the case of two mitochondrial group II intron paralogues in ferns (18).

DISCUSSION

Group II introns and their splicing factors

The multifarious evolutionary pathways of group II introns are complicated by the interactions of at least two major players: the 'ribozymic' RNA structure itself and the intimately connected 'maturase' present as an intron-encoded protein (IEP), mostly borne within the flexible domain IV of the group II intron and involved both in splicing and retromobility of its host intron (40). This prototypical, and likely ancestral, state of an autocatalytic and mobile group II intron is rare among group II introns in plant mitochondria and chloroplasts where only one such maturase each remains encoded in an intron of the mitochondrial *nad1* gene ('*mat-R*') and in the chloroplast *trnK* gene ('*matK*') among flowering plants (48,59). These two 'remaining' maturases in angiosperms were shown to act promiscuously on multiple introns within each organelle (60,61), similar to other, unrelated splicing factors that were likewise shown to act on multiple introns simultaneously (62–64). Clearly though, this angiosperm-type setup is a derived state of group II intron evolution and earlier emerging plant lineages reflect more ancient and dynamic group II intron scenarios. This is very likely correlated with many more functional maturases present in the mtDNAs of early land plant lineages, e.g. mat-cobi824g2, mat-atp1i989g2, mat-atp1i1050g2, mat-cox1i178g2 and mat-rrnSi1065g2 in liverworts, all of which retain the characteristically conserved YADD Motif in RT domain RT5 known to be crucial for intron mobility. None of the internal introns reported here or the discussed paralogs from which they possibly originate, respectively, carry maturase reading frames that could assist in their splicing.

Mobile group II introns in plant mitochondria

Among the altogether 23 introns (25 counting twintrons twice) present in the six gene structures displayed in figure 6 alone, a full 15 (17) are exclusively present in hornwort mtDNAs. The high diversity of mitochondrial intron occurrence among the six major plant clades (see e.g. Supplementary Figure S6 for the *atp1* gene alone) suggests dozens of intron loss and gain events along the backbone of land plant phylogeny. Establishing evolutionary scenarios for those processes is difficult since the creation of intron paralogues has likely happened in deep phylogenetic time some 400–500 million years ago, obliterating convincing sequence similarities. In two cases, however, the origins of fern-specific mitochondrial introns have recently been traced to more ancestral counterparts shared with other plant lineages (17,18). The here described case of extraordinary sequence similarity of the *Anthoceros*-specific group II intron cox2i98g2 with the internal intron atp1i1050g2ii1536g2 in one of the two hornwort twintrons is a striking novel example for a likely much more recent intron copying event.

The above cases appear to represent rare, one-time copying events by retrotransposition of a group II intron into a new locus (65,66). In stark contrast, cox1i1149g2 in *Phlegmariurus* seems to be a particularly invasive, hypermobile group II intron for which we detected altogether nine additional insertion sites in the mitogenome, including its own domain IV (Figure 4). In fact, the concomitant reinvestigation of the *sdh3* locus has led us to identify *sdh3* as a functional mitochondrial gene conserved among Lycopodiaceae (Supplementary Figure S5). This and the other examples of twintrons reported here have likely been missed in previous analyses given that similarities between repeated sequences are very common in the highly diverse vascular plant mtDNAs and do not attract particular attention. *Vice versa*, we cannot fully exclude that insertions of introns lacking any significant sequence similarity at identical sites have been missed, including taxa outside of the plant lineage. However, the one intriguing example of very likely independent parallel insertions arising from an outer source (very likely fungal, but remaining yet unidentified) concerns a group I intron inserted into the mitochondrial *cox1* gene of disparate angiosperm lineages that do share significant sequence similarities (67–72).

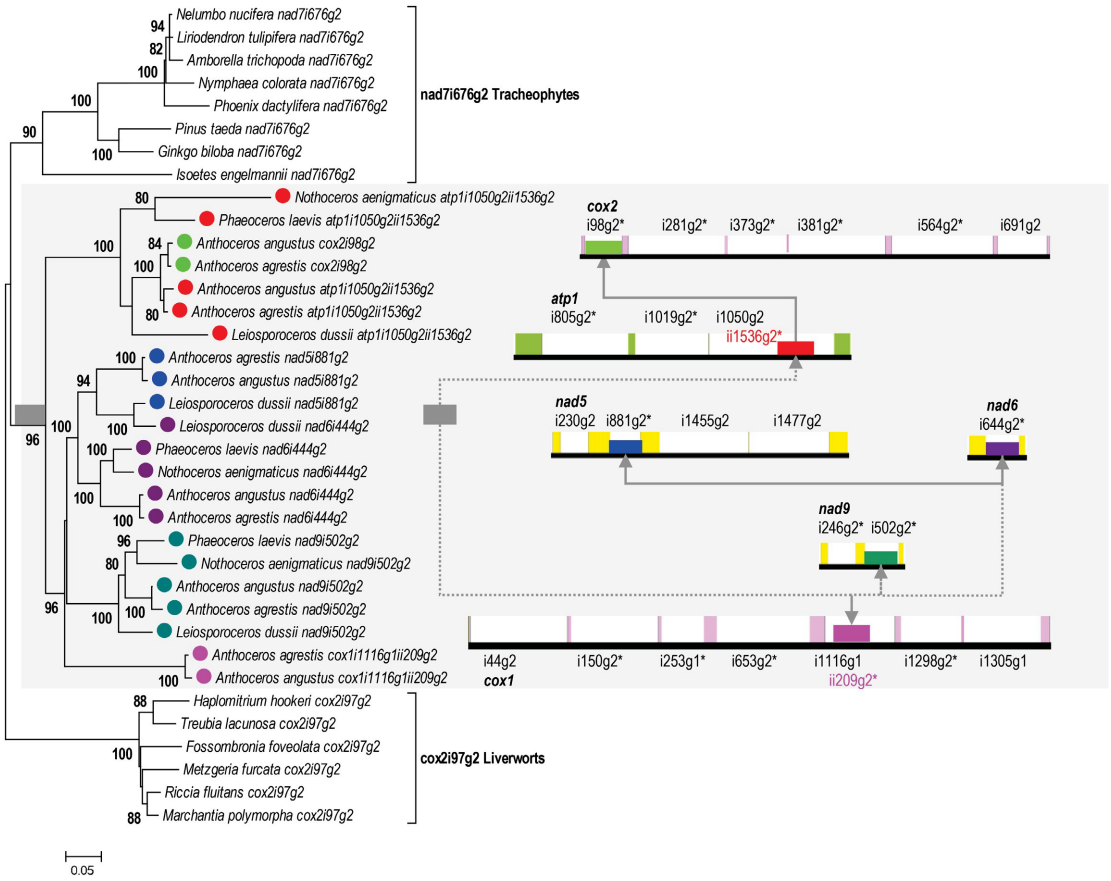


Figure 6. Hornwort twintrons share a common origin with four other group II intron paralogues specific to this clade. The internal introns atp1i1050g2ii1536g2 and cox1i1116g1ii209g2 in the now identified hornwort twintrons share a common origin with four other group II intron paralogues (cox2i98g2, nad5i881g2, nad6i444g2 and nad9i502g2) in hornworts. The clade of hornwort introns (grey background shading) is related to group II intron nad7i676g2 conserved in tracheophytes (top) and, more distantly, to cox2i97g2 in liverworts (bottom). Gapped alignment positions with <80% coverage were eliminated for phylogenetic analysis (737 positions remaining). Node support was determined from 100 bootstrap replicates and is indicated where at least 70%. The six different intron paralogues in hornworts are labeled with dots corresponding in color to the respective introns highlighted in the *Anthoceros* gene structures shown on the right, which were created using OGDRAW (101) with default coloring for gene categories (*nad* yellow, *cox* pink, *atp* green). Asterisks behind intron labels indicate hornwort-specific introns absent in other plant clades. Tilted arrows indicate likely (cox2i98g2 and nad5i881g2/ nad6i444g2) and more hypothetical (stippled lines) copy events to explain the origin of intron paralogues starting from an ancestral donor (gray box) likely related to the nad7i676g2 intron.

Especially the case of zombie twintrons lacking detectable splicing of an evidently degenerated internal intron requires careful alignments with sequences of related taxa to clearly define whether ends of sequence similarities coincide with intron 5′ and 3′ ends. As we have here shown for the hornwort clade and on a shallower phylogenetic level with our investigation of additional Lycopodiaceae taxa, an extended sequence sampling helps to define the nature of repeated sequence similarities. Yet more extended taxonomic sampling could possibly allow to identify points of degeneration of a fully functional twintron (like atp9i87g2-twin) into degenerating zombie twintrons where only splicing of the external intron is retained (like cox1i1149g2-twin and sdh3i349g2-twin).

Serially splicing and progressively degenerating or zombie twintrons right from the start?

The confirmed splicing of the internal intron in atp9i87g2-twin of Lycopodiaceae is a prerequisite to reconstitute the maturase reading frame of the external intron. Similar to the RNA editing events reconstituting important residues of *mat-atp9i87g2* (Supplementary Figure S2), this underlines the important role of this maturase. On the other hand, the ancestral maturase reading frame is degenerated in the moss and hornwort orthologues. Intriguingly, two nuclear genes are present in the moss *Physcomitrella patens* that encode proteins with significant similarity to *mat-atp9i87g2*, possibly acting as necessary splicing factors upon import into mitochondria (48).

Despite different approaches to detect internal intron splicing (RT-PCR primers variably placed in the flanking external intron sequences and in exons in different combinations and amplification of RNaseR-treated RNA preparations to enrich for non-linear RNAs, see Supplementary Figure S1) we could not find evidence for internal intron splicing in the zombie introns reported here. The ectopic insertion of an intron into structurally less relevant parts of an external host intron could remain without consequences even without prior splicing of the novel internal intron. Intron cox1i1149g2-twin in Lycopodiaceae could be an example for such a scenario since the insertion site is located in structurally irrelevant domain IV, here also lacking a functional maturase reading frame (Figure 3). Such speculations seem to be much less likely for sdh3i349g2-twin where the invading hypermobile intron is inserted into the structurally important domain I of the host intron (Supplementary Figure S5). In this case it is actually surprising that we could detect functional splicing of the host intron but not of the internal intron.

We assume that this state has arisen only after evolutionary adaptation of host intron domain I to accommodate the large sequence insertion. Along the same lines, we assume the internal intron inserted into a structurally critical part of the group I host intron cox1i1116g1 to have been functional splicing in its original state (Figure 5). Here, the insertion of the internal intron is located in the compact and structured P5 region rather than e.g. in the P9 region that is known to be variable in size and structure among group I introns. Notably, the cox1i1116g1 counterpart in liverworts carries a highly conserved intron-borne reading frame of the LAGLIDADG type that we find disrupted but retaining conservation in the 5^{1}-part in *Anthoceros*. The lack of cox1i1116g1 in the mitogenomes of the other hornwort genera, either alone or as a twintron, make speculations about its origins moot. Finally, in the case of atp1i1050g2-twin we see strongest support for ancestral functionality coming from the phylogenetic analysis clearly showing that the splicing-competent paralogue cox2i98g2 has evidently just originated recently from the internal intron in *Anthoceros* (Figure 6). Together with the conservation of conserved secondary structure elements and tertiary structure interactions in the internal group II introns of the four zombie twintrons reported here in addition to their serially splicing counterpart atp9i87g2-twin, we assume that most, if not all, have initially been relying on internal splicing to precede external splicing of the host intron.

Independent origins of twintrons in diverse organisms

After the early discovery of twintrons in the peculiar cpDNA of *Euglena gracilis* (28–30,73,74), subsequent publications sporadically demonstrated that twintrons were not restricted to this unique organelle derived from secondary endosymbiosis in the Euglenids (75–77) but exceptionally also existed as group II twintrons in the cpDNA of a cryptomonad (78), in the mtDNA of yeasts (79), in the red alga *Porphyridium purpureum* (80) and in a marine cyanobacterium (81) or as group I twintrons in the nuclear rDNA of non-photosynthetic protists (82,83) and as group I or 'mixed' twintrons in fungal mtDNAs (84–87). Twintrons have also been described as spliceosomal 'stwintrons' in fungi (56,88) and it must also be noted that the term twintron had likewise been used in an extended meaning for a U2-type nuclear spliceosomal intron within a U12-type intron of the '*prospero*' locus in *Drosophila melanogaster* (89), and similar examples in vertebrates (90). Those cases, however, may better be understood as a unique type of alternative splicing rather than a serial process with splicing of an external intron relying on previous splicing of an internal intron. The here described *bona fide* cases of twintrons with internal group II introns in land plant mitochondrial genomes are, to our knowledge, the first examples for the green plant lineage (Viridiplantae).

Plant mitochondria as a playground for group II intron evolution

The mitochondrial genomes of land plants continue to reveal highly interesting pathways of group II intron evolution. Earlier studies have addressed the origin of disrupted group II introns subsequently requiring *trans*-splicing (15,91,92), their occasional losses (93–95) or, more recently, their occupation of novel insertion sites (17,18). The dependence of plant mitochondrial group II introns on (rarely) intron-borne or (mostly) nuclear-encoded matturases and other splicing factors (96) and their co-evolution with such proteins will be of increasing interest in the future. Notably, all available evidence shows that the structural makeup and biochemical features of group II introns are similar, and likely even ancestrally homologous, to those of the spliceosomal introns in the nuclear genomes of eukaryotes (1,3,97,98).

The variable and dynamic plant mitochondrial group II introns will offer a vast field of co-evolution with their (mostly) nuclear encoded protein co-factors. Increasing numbers of high-quality flowering plant genomes have recently allowed to trace the co-evolution between C-to-U RNA editing events in plant organelles and their corresponding nuclear-encoded specificity factors (e.g. 99). However, whereas RNA editing is highly variable even between more closely related plant taxa among angiosperms, intron variability occurs largely between the most ancient land plant lineages arising more than 300 mio. years ago, in an early phase of mitogenome 'intron' evolution preceding a later 'recombination' phase (19). The here discussed case of atp9i87g2-twin with its ancestrally intron-borne maturase, retained in lycophytes, disrupted by an invading internal intron in Lycopodiaceae, degenerating in hornworts and mosses and accompanied by nuclear gene transfer at least in the latter, is an exemplary case. Accordingly, such comparative studies on co-evolution will have to wait for more genome sequences for representatives of early-emerging lineages in the liverworts, mosses, hornworts and early branching vascular plants. Combining such deep diversity studies with the establishment of key taxa like the hornwort *Anthoceros agrestis* as model organisms for reverse genetic approaches (100) will greatly accelerate progress in that direction.

DATA AVAILABILITY

LR721677-LR721682, LR722602-LR722612, LR722615 and LR722624.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge computing time provided by the Paderborn Center for Parallel computing 'PC2' and in particular the kind support of Axel Keller. We wish to thank Tim van Beers very much for suggesting 'zombie' as a label for the degenerated twintrons and the 2018 laboratory course WP04B on plant molecular biology for contributions to the analysis of the Lycopodiaceae twintron arrangements. Also, we are very grateful to our colleagues at the Bonn University Botanic Garden for providing Lycopodiaceae plant material.
Author contributions: S.Z. and P.G. did analysis of sequence data, S.Z. did nucleic acid preparations, PCR amplifications and molecular cloning, V.K. designed and supervised the study and wrote the manuscript, S.Z. and V.K. prepared figures and edited the final manuscript.

FUNDING

Funding for open access charge: Basic university household funding.
Conflict of interest statement. None declared.

REFERENCES

1. Novikova,O. and Belfort,M. (2017) Mobile group II introns as ancestral eukaryotic elements. *Trends Genet.*, **33**, 773–783.
2. Smathers,C.M. and Robart,A.R. (2019) The mechanism of splicing as told by group II introns: ancestors of the spliceosome. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1862**, 194390.
3. Vosseberg,J. and Snel,B. (2017) Domestication of self-splicing introns during eukaryogenesis: the rise of the complex spliceosomal machinery. *Biol. Direct*, **12**, 30.
4. Galej,W.P., Toor,N., Newman,A.J. and Nagai,K. (2018) Molecular mechanism and evolution of nuclear pre-mRNA and group II intron splicing: insights from cryo-electron microscopy structures. *Chem. Rev.*, **118**, 4156–4176.
5. Agrawal,R.K., Wang,H.-W. and Belfort,M. (2016) Forks in the tracks: group II introns, spliceosomes, telomeres and beyond. *RNA Biol.*, **13**, 1218–1222.
6. McNeil,B.A., Semper,C. and Zimmerly,S. (2016) Group II introns: versatile ribozymes and retroelements. *Wiley Interdiscip. Rev. RNA*, **7**, 341–355.
7. Haack,D.B., Yan,X., Zhang,C., Hingey,J., Lyumkis,D., Baker,T.S. and Toor,N. (2019) Cryo-EM structures of a group II intron reverse splicing into DNA. *Cell*, **178**, 612–623.
8. Belfort,M. and Lambowitz,A.M. (2019) Group II Intron RNPs and reverse transcriptases: from retroelements to research tools. *Cold Spring Harb. Perspect. Biol.*, **11**, a032375.
9. Bock,R. and Knoop,V. (2012) In: Bock,R and Knoop,V (eds). *Genomics of Chloroplasts and Mitochondria Auflage: 2*. Springer, Dordrecht.
10. Pruchner,D., Nassal,B., Schindler,M. and Knoop,V. (2001) Mosses share mitochondrial group II introns with flowering plants, not with liverworts. *Mol. Genet. Genomics*, **266**, 608–613.
11. Pruchner,D., Beckert,S., Muhle,H. and Knoop,V. (2002) Divergent intron conservation in the mitochondrial *nad2* gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. *J. Mol. Evol.*, **55**, 265–271.
12. Qiu,Y.L., Cho,Y., Cox,J.C. and Palmer,J.D. (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature*, **394**, 671–674.
13. Qiu,Y.-L., Li,L., Wang,B., Chen,Z., Knoop,V., Groth-Malonek,M., Dombrovska,O., Lee,J., Kent,L., Rest,J. *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15511–15516.
14. Volkmar,U., Groth-Malonek,M., Heinrichs,J., Muhle,H., Polsakiewicz,M. and Knoop,V. (2012) Exclusive conservation of mitochondrial group II intron nad4i548 among liverworts and its use for phylogenetic studies in this ancient plant clade. *Plant Biol. (Stuttg.)*, **14**, 382–391.
15. Groth-Malonek,M., Pruchner,D., Grewe,F. and 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.
16. Knoop,V. (2010) Looking for sense in the nonsense: a short review of non-coding organellar DNA elucidating the phylogeny of bryophytes. *Trop. Biol.*, **31**, 51–60.
17. Knie,N., Grewe,F. and Knoop,V. (2016) Monilophyte mitochondrial *rps1* genes carry a unique group II intron that likely originated from an ancient paralog in *rpl2*. *RNA*, **22**, 1338–1348.
18. Zumkeller,S.M., Knoop,V. and Knie,N. (2016) Convergent evolution of fern-specific mitochondrial group II intron atp1i361g2 and its ancient source paralogue rps3i249g2 and independent losses of intron and RNA editing among Pteridaceae. *Genome Biol. Evol.*, **8**, 2505–2519.
19. Knoop,V. (2013) Plant mitochondrial genome peculiarities evolving in the earliest vascular plant lineages. *J. Syst. Evol.*, **51**, 1–12.
20. Liu,Y., Wang,B., Cui,P., Li,L., Xue,J.-Y., Yu,J. and Qiu,Y.-L. (2012) The mitochondrial genome of the lycophyte *huperzia squarrosa*: the most archaic form in vascular plants. *PLoS One*, **7**, e35168.
21. Grewe,F., Viehoever,P., Weisshaar,B. and 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.
22. Hecht,J., Grewe,F. and 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.
23. Villarreal Aguilar,J.C., Turmel,M., Bourgouin-Couture,M., Laroche,J., Salazar Allen,N., Li,F.-W., Cheng,S., Renzaglia,K. and Lemieux,C. (2018) Genome-wide organellar analyses from the hornwort *Leiosporoceros dussii* show low frequency of RNA editing. *PLoS One*, **13**, e0200491.
24. Xue,J.Y., Liu,Y., Li,L., Wang,B. and Qiu,Y.L. (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.
25. Yoshinaga,K., Iinuma,H., Masuzawa,T. and Uedal,K. (1996) Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res.*, **24**, 1008–1014.
26. Dong,S., Wu,H., Zhang,S., Zhang,L., Liu,Y., Xue,J.Y., Chen,Z. and Goffinet,B. (2018) Complete mitochondrial genome sequence of *Anthoceros angustus*: conservative evolution of the mitogenomes in hornworts. *Bryologist*, **121**, 14–22.
27. Cargill,D.C., Renzaglia,K.S., Villarreal,J.C. and Duff,R.J. (2005) Generic concepts within hornworts: historical review, contemporary insights and future directions. *Aust. Syst. Bot.*, **18**, 7–16.
28. Copertino,D.W. and Hallick,R.B. (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. *EMBO J.*, **10**, 433–442.
29. Copertino,D.W., Christopher,D.A. and Hallick,R.B. (1991) A mixed group II/group III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3 gene: evidence for intron insertion during gene evolution. *Nucleic Acids Res.*, **19**, 6491–6497.
30. Copertino,D.W., Hall,E.T., Van Hook,F.W., Jenkins,K.P. and Hallick,R.B. (1994) A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res.*, **22**, 1029–1036.

31. Hafez,M. and Hausner,G. (2015) Convergent evolution of twintron-like configurations: one is never enough. *RNA Biol.*, **12**, 1275–1288.
32. Li,D., Liu,C.-M., Luo,R., Sadakane,K. and Lam,T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**, 1674–1676.
33. Leinonen,R., Sugawara,H., Shumway,M. and International Nucleotide Sequence Database Collaboration (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
34. Grabherr,M.G., Haas,B.J., Yassour,M., Levin,J.Z., Thompson,D.A., 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.
35. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
36. Field,A.R., Testo,W., Bostock,P.D., Holtum,J.A.M. and Waycott,M. (2015) Molecular phylogenetics and the morphology of the Lycopodiaceae subfamily Huperzioideae supports three genera: *Huperzia, Phlegmariurus* and *Phylloglossum*. *Mol. Phylogenet. Evol.*, **94**, 635–657.
37. Doyle,J.L. and Doyle,J.J. (1990) Isolation of plant DNA from fresh tissue. *Focus (Madison)*, **12**, 13–15.
38. Liao,Z., Chen,M., Guo,L., Gong,Y., Tang,F., Sun,X. and Tang,K. (2004) Rapid isolation of high-quality total RNA from taxus and ginkgo. *Prep. Biochem. Biotechnol.*, **34**, 209–214.
39. Kumar,S., Stecher,G. and Tamura,K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, **33**, 1870–1874.
40. Zimmerly,S. and Semper,C. (2015) Evolution of group II introns. *Mob. DNA*, **6**, 7.
41. Michel,F., Umesono,K. and Ozeki,H. (1989) Comparative and functional anatomy of group II catalytic introns–a review. *Gene*, **82**, 5–30.
42. Fedorova,O. and Zingler,N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.*, **388**, 665–678.
43. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
44. Zhao,C. and Pyle,A.M. (2017) The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand. *Curr. Opin. Struct. Biol.*, **47**, 30–39.
45. Blocker,F.J.H., Mohr,G., Conlan,L.H., Qi,L., Belfort,M. and Lambowitz,A.M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*, **11**, 14–28.
46. Dombrovska,E. and Qiu,Y.-L. (2004) Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Mol. Phylogenet. Evol.*, **32**, 246–263.
47. Knoop,V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.*, **46**, 123–139.
48. Guo,W. and Mower,J.P. (2013) Evolution of plant mitochondrial intron-encoded maturases: frequent lineage-specific loss and recurrent intracellular transfer to the nucleus. *J. Mol. Evol.*, **77**, 43–54.
49. Li,L., Wang,B., Liu,Y. and Qiu,Y.L. (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. *J. Mol. Evol.*, **68**, 665–678.
50. Grewe,F., Herres,S., Viehöver,P., Polsakiewicz,M., Weisshaar,B. and 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 Res.*, **39**, 2890–2902.
51. Groth-Malonek,M., Wahrmund,U., Polsakiewicz,M. and Knoop,V. (2007) Evolution of a pseudogene: exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol. Biol. Evol.*, **24**, 1068–1074.
52. Adams,K.L., Rosenblueth,M., Qiu,Y.L. and Palmer,J.D. (2001) Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics*, **158**, 1289–1300.
53. Gerke,P., Szövényi,P., Neubauer,A., Lenz,H., Gutmann,B., McDowell,R., Small,I., Schallenberg-Rüdinger,M. and Knoop,V. (2019) Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort *Anthoceros agrestis*. *New Phytol.*, doi:10.1111/nph.16297.
54. Stoddard,B.L. (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA*, **5**, 7.
55. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189–5200.
56. Fekete,E., Flipphi,M., Ág,N., Kavalecz,N., Cerqueira,G., Scazzocchio,C. and Karaffa,L. (2017) A mechanism for a single nucleotide intron shift. *Nucleic Acids Res.*, **45**, 9085–9092.
57. Bainard,J.D. and Villarreal,J.C. (2013) Genome size increases in recently diverged hornwort clades. *Genome*, **56**, 431–435.
58. Duff,R.J., Villarreal,J.C., Cargill,D.C. and Renzaglia,K.S. (2007) Progress and challenges toward developing a phylogeny and classification of the hornworts. *Bryologist*, **110**, 214–243.
59. Schmitz-Linneweber,C., Lampe,M.-K.K., Sultan,L.D. and Ostersetzer-Biran,O. (2015) Organellar maturases: a window into the evolution of the spliceosome. *Biochim. Biophys. Acta*, **1847**, 798–808.
60. Zoschke,R., Nakamura,M., Liere,K., Sugiura,M., Börner,T. and Schmitz-Linneweber,C. (2010) An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 3245–3250.
61. Sultan,L.D., Mileshina,D., Grewe,F., Rolle,K., Abudraham,S., Głodowicz,P., Niazi,A.K., 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.
62. Zmudjak,M., Colas des Francs-Small,C., Keren,I., Shaya,F., Belausov,E., Small,I. and Ostersetzer-Biran,O. (2013) mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol.*, **199**, 379–394.
63. Chen,Z., Wang,H.-C., Shen,J., Sun,F., Wang,M., Xu,C. and Tan,B.-C. (2019) PPR-SMR1 is required for the splicing of multiple mitochondrial introns, interacts with Zm-mCSF1, and is essential for seed development in maize. *J. Exp. Bot.*, **70**, 5245–5258.
64. Francs-Small,C.C., Kroeger,T., Zmudjak,M., Ostersetzer-Biran,O., Rahimi,N., Small,I. and Barkan,A. (2012) A PORR domain protein required for *rpl2* and *ccmF(C)* intron splicing and for the biogenesis of c-type cytochromes in *Arabidopsis* mitochondria. *Plant J. Cell Mol. Biol.*, **69**, 996–1005.
65. Dickson,L., Huang,H.-R., Liu,L., Matsuura,M., Lambowitz,A.M. and Perlman,P.S. (2001) Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc. Natl. Acad. Sci.*, **98**, 13207–13212.
66. Eskes,R., Liu,L., Ma,H., Chao,M.Y., Dickson,L., Lambowitz,A.M. and Perlman,P.S. (2000) Multiple homing pathways used by yeast mitochondrial group II introns. *Mol. Cell Biol.*, **20**, 8432–8446.
67. Cusimano,N., Zhang,L.B.L.-B. and Renner,S.S. (2008) Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol. Biol. Evol.*, **25**, 265–276.
68. Sanchez-Puerta,M.V., Abbona,C.C., Zhuo,S., Tepe,E.J., Bohs,L., Olmstead,R.G. and Palmer,J.D. (2011) Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evol. Biol.*, **11**, 277.
69. Cho,Y. and Palmer,J.D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol.*, **16**, 1155–1165.
70. Cho,Y., Qiu,Y.L., Kuhlman,P. and Palmer,J.D. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14244–14249.
71. Sanchez-Puerta,M.V., Cho,Y., Mower,J.P., Alverson,A.J. and Palmer,J.D. (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Mol. Biol. Evol.*, **25**, 1762–1777.
72. Vaughn,J.C., Mason,M.T., Sper-Whitis,G.L., Kuhlman,P. and Palmer,J.D. (1995) Fungal origin by horizontal transfer of a plant

mitochondrial group I intron in the chimeric CoxI gene of *Peperomia*. *J. Mol. Evol.*, **41**, 563–572.
73. Drager,R.G. and Hallick,R.B. (1993) A complex twintron is excised as four individual introns. *Nucleic Acids Res.*, **21**, 2389–2394.
74. Copertino,D.W., Shigeoka,S. and Hallick,R.B. (1992) Chloroplast group III twintron excision utilizing multiple 5^{1}- and 3^{1}-splice sites. *EMBO J.*, **11**, 5041–5050.
75. Pombert,J.-F., James,E.R., Janous˘kovec,J. and Keeling,P.J. (2012) Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorphina aenigmatica* plastid genome. *PLoS One*, **7**, e53433.
76. Thompson,M.D., Copertino,D.W., Thompson,E., Favreau,M.R. and Hallick,R.B. (1995) Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Res.*, **23**, 4745–4752.
77. Thompson,M.D., Zhang,L., Hong,L. and Hallick,R.B. (1997) Two new group-II twintrons in the *Euglena gracilis* chloroplast are absent in basally branching *Euglena* species. *Curr. Genet.*, **31**, 89–95.
78. Maier,U.G., Rensing,S.A., Igloi,G.L. and Maerz,M. (1995) Twintrons are not unique to the euglena chloroplast genome - structure and evolution of a plastome cpn60 gene from a cryptomonad. *Mol. Gen. Genet.*, **246**, 128–131.
79. Schmidt,W.M., Schweyen,R.J., Wolf,K. and Mueller,M.W. (1994) Transposable Group II introns in fission and budding yeast. *J. Mol. Biol.*, **243**, 157–166.
80. Perrineau,M.-M., Price,D.C., Mohr,G. and Bhattacharya,D. (2015) Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. *PeerJ*, **3**, e1017.
81. Pfreundt,U. and Hess,W.R. (2015) Sequential splicing of a group II twintron in the marine cyanobacterium Trichodesmium. *Sci. Rep.*, **5**, 16829.
82. Johansen,S., Elde,M., Vader,A., Haugen,P., Haugli,K. and Haugli,F. (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete Didymium iridis. *Mol. Microbiol.*, **24**, 737–745.
83. Einvik,C., Elde,M. and Johansen,S. (1998) Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. *J. Biotechnol.*, **64**, 63–74.
84. Hafez,M., Majer,A., Sethuraman,J., Rudski,S.M., Michel,F. and Hausner,G. (2013) The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: Twintrons, introns, and intron-encoded proteins. *Fungal Genet. Biol.*, **53**, 71–83.
85. Guha,T.K. and Hausner,G. (2014) A homing endonuclease with a switch: Characterization of a twintron encoded homing endonuclease. *Fungal Genet. Biol.*, **65**, 57–68.
86. Deng,Y., Zhang,Q., Ming,R., Lin,L., Lin,X., Lin,Y., Li,X., Xie,B. and Wen,Z. (2016) Analysis of the mitochondrial genome in Hypomyces aurantius reveals a novel twintron complex in fungi. *Int. J. Mol. Sci.*, **17**, 1049.
87. Deng,Y., Hsiang,T., Li,S., Lin,L., Wang,Q., Chen,Q., Xie,B. and Ming,R. (2018) Comparison of the mitochondrial genome sequences of six annulohypoxylon stygium isolates suggests short fragment insertions as a potential factor leading to larger genomic size. *Front. Miçrobiol.*,**9**, 2079.
88. Kavalecz,N., A ̈g,N., Karaffa,L., Scazzocchio,C., Flipphi,M. and Fekete,E. (2019) A spliceosomal twin intron (stwintron) participates in both exon skipping and evolutionary exon loss. *Sci. Rep.*, **9**, 9940.
89. Scamborova,P., Wong,A. and Steitz,J.A. (2004) An intronic enhancer regulates splicing of the twintron of *Drosophila melanogaster* prospero pre-mRNA by two different spliceosomes. *Mol. Cell Biol.*, **24**, 1855–1869.
90. Janice,J., Ja ̨kalski,M. and Makałowski,W. (2013) Surprisingly high number of Twintrons in vertebrates. *Biol. Direct*, **8**, 4.
91. Malek,O. and 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.
92. Qiu,Y.L. and Palmer,J.D. (2004) Many independent origins of trans splicing of a plant mitochondrial group II intron. *J. Mol. Evol.*, **59**, 80–89.
93. Hepburn,N.J., Schmidt,D.W. and Mower,J.P. (2012) Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol. Biol. Evol.*, **29**, 3111–3120.
94. Grewe,F., Zhu,A. and Mower,J.P. (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.
95. Guo,W., Grewe,F., Fan,W., Young,G.J., Knoop,V., Palmer,J.D. and Mower,J.P. (2016) *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.*, **33**, 1448–1460.
96. Brown,G.G., Colas des Francs-Small,C. and Ostersetzer-Biran,O. (2014) Group II intron splicing factors in plant mitochondria. *Front. Plant Sci.*, **5**, 35.
97. Peters,J.K. and Toor,N. (2015) Group II intron lariat: Structural insights into the spliceosome. *RNA Biol.*, **12**, 913–917.
98. Rogozin,I.B., Carmel,L., Csuros,M. and Koonin,E.V (2012) Origin and evolution of spliceosomal introns. *Biol. Direct*, **7**, 11.
99. Hein,A., Brenner,S. and Knoop,V. (2019) Multifarious evolutionary pathways of a nuclear RNA editing factor: disjunctions in co-evolution of DOT4 and its chloroplast target rpoC1eU488SL. *Genome Biol. Evol.*, **11**, 798–813.
100. Szo ̈ve ́nyi,P., Frangedakis,E., Ricca,M., Quandt,D., Wicke,S. and Langdale,J.A. (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol.*, **15**, 98.
101. Greiner,S., Lehwark,P. and Bock,R. (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.*, **47**, W59–W64.

Publication 2: Rickettsial DNA invasions and a scrambled rRNA cluster with a trans-splicing group I intron: The highly unorthodox mitogenome of the fern Haplopteris ensiformis

Communications Biology; 2022 *submitted*
Available as pre-print on Research Square
https://doi.org/10.21203/rs.3.rs-2152556/v1
Zumkeller S. M. Polsakiewicz M. and Knoop V.

IZMB - Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution,
Universität Bonn,
Kirschallee 1, D-53115 Bonn
Corresponding author
Email: zumkeller@uni-bonn.de

Rickettsial DNA invasions and a scrambled rRNA cluster with a trans-splicing group I intron: The highly unorthodox mitogenome of the fern Haplopteris ensiformis

Simon Zumkeller (zumkellersimon@gmail.com)
University of Bonn https://orcid.org/0000-0002-1630-2439
Monika Polsakiewicz
Universität Bonn
Volker Knoop
University of Bonn https://orcid.org/0000-0002-8485-9423

Article

Keywords: Monilophyte mitogenome, Pteridaceae, Group I introns, Group II introns, Trans-splicing, C-to U and U-to-C RNA editing, Lateral and Horizontal Gene Transfer, Rickettsiales

Posted Date: October 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2152556/v1

Rickettsial DNA invasions and a scrambled rRNA cluster with a *trans*-splicing group I intron: The highly unorthodox mitogenome of the fern *Haplopteris ensiformis*

Simon Zumkeller, Monika Polsakiewicz and Volker Knoop*

IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, D-53115 Bonn

ORCID IDs:
Simon Zumkeller: 0000-0002-1630-2439
Volker Knoop: 0000-0002-8485-9423

*Corresponding author
Email: volker.knoop@uni-bonn.de
Tel +49 228 73 6466, Fax +49 228 73 6467

Key words: Monilophyte mitogenome, Pteridaceae, Group I introns, Group II introns, *Trans*-splicing, C-to U and U-to-C RNA editing, Lateral and Horizontal Gene Transfer, Rickettsiales

Summary

Plant mitochondrial genomes can be complex owing to highly recombinant structures, lack of gene syntenies, heavy RNA editing and invasion of chloroplast, nuclear or even foreign DNA by horizontal gene transfer (HGT). Leptosporangiate ferns remained the last major plant clade without an assembled mitogenome, likely owing to a demanding combination of the above. We here present both organelle genomes of *Haplopteris ensiformis* as a first leptosporangiate fern. More than 1,400 events of C-to-U RNA editing and over 500 events of reverse U-to-C edits affect its organelle transcriptomes. The *Haplopteris* mitogenome has a rich gene complement lacking only the *ccm* gene suite but is highly unorthodox, indicating extraordinary recombinogenic activity. Although eleven group II introns known in disrupted *trans*-splicing states in seed plants exist in conventional *cis*-arrangements, a particularly complex structure is found for the mitochondrial *rrnL* gene, which is split into two parts needing reassembly on RNA level by a novel *trans*-splicing group I intron. Finally, and aside from ca. 80 chloroplast DNA inserts that complicated the mitogenome assembly, the *Haplopteris* mtDNA features as an unparalleled idiosyncrasy 30 variably degenerated protein coding regions from Rickettiales bacteria indicative of heavy bacterial HGT on top of tRNA genes of chlamydial origin.

Introduction

The conserved structure of chloroplast genomes (plastomes) and the stoichiometric dominance of chloroplast (cpDNA) over nuclear and mitochondrial (mtDNA) in total plant nucleic acid preparations have led to a tremendous increase of available complete plastome sequences with the advent of Next Generation Sequencing (NGS) technologies. The number of completed plant mitochondrial genome sequences (mitogenomes) is much lower in contrast, most notably in vascular plants (tracheophytes). This is largely due to the much more variable and complex mitogenome structures in tracheophytes[1–5], which hitherto left complete chondrome assemblies altogether missing for the large clade of leptosporangiate ferns among the monilophytes.

Complex, recombining mitogenomes clustered with repeat sequences and affected by the lateral invasion of chloroplast DNA, or even by horizontal gene transfer (HGT) from other species, have arisen independently in the four large tracheophyte clades: the angiosperms, the gymnosperms, the monilophytes and the lycophytes. The lycophytes as the evolutionary oldest of the four clades of extant vascular plants reflect this most clearly: The mitogenome in the club moss *Phlegmariurus squarrosus* is a circular DNA with a rich gene complement and even retaining several ancestral gene syntenies of the circular mtDNAs in bryophytes[6]. The mtDNAs of the quillwort *Isoetes engelmannii* or the spike moss *Selaginella moellendorffii,* in contrast, are strongly depauperated in gene content and heavily affected by recombination leading to complex and coexisting arrangements of coding islands embedded between repeated sequences[7,8]. Similarly, moderately compact and gene-rich circular mitogenomes are present in the gymnosperms *Cycas taitungensis*[9] and *Ginkgo biloba*[10]. In contrast, the mitogenomes of *Taxus cuspidata*[11] and *Welwitschia mirabilis*[10] are reduced in gene content and this even despite size increase to nearly 1,000 KBp in the latter case. As reflected from a collection of mitochondrial scaffolds and chromosomal assemblies, respectively, yet much larger mitogenomes are present in the conifer I clade in *Picea* species[12,13] or *Larix sibirica*[14].

Among angiosperms, the mtDNA of the magnoliid *Liriodendron tulipifera* represents an ancestral, gene rich state[15]. Other flowering plant species, however, have complex mtDNAs of enormous sizes

exceeding 1,000 KBp, are fragmented into multiple co-existing mitochondrial chromosomes, affected by massive lateral gene transfer (LGT) from the cpDNA or by horizontal gene transfer (HGT) from other species or display combinations of those features to variable degrees. The large mitogenomes exceeding 1 MBp in the cucumber family[16,17], the diverse multipartite mtDNAs in the genus *Silene*[18] or the mtDNA of the isolated flowering plant *Amborella trichopoda* that is heavily affected by HGT from diverse species[19] are prime examples.

Among monilophytes (ferns *sensu lato* including the horsetails), the second largest group of tracheophytes behind the angiosperms, the situation is less clear since only the two complete chondrome sequences of *Ophioglossum californicum* (adder's tongue) and the whisk fern *Psilotum nudum* have been determined[20]. Although the two taxa represent sister clades among the eusporangiate ferns, a paraphyletic grade at the base of extant fern taxa, their mitogenomes reflect differences indicating a dynamic evolution of mtDNAs also among the monilophytes. Aside from slight differences in gene and intron complement, the *O. californicum* mtDNA maps as a single circular mtDNA whereas two separate circular chromosomes exist in *P. nudum*[20].

Most of extant fern diversity with more than 10,000 species resides in the leptosporangiate ferns, however. Studies on selected mitochondrial loci among ferns revealed interesting dynamics of group II introns in their mtDNAs[21,22]. The leptosporangiate fern family of Pteridaceae (Polypodiales) proved to be particularly interesting in a study of mitochondrial group II intron gain, loss and coevolution scenarios[22], also with respect to the concomitant evolution of C-to-U and "reverse" U-to-C RNA editing, which is abundantly present in the endosymbiotic organelles of leptosporangiate ferns[23,24]. The Pteridaceae represent a large fern family comprising some 1,150 species in 45 genera, placed into at least five sub-groups and potential sub-families: the cryptogrammoid, the adiantoid-vittarioid, the cheilanthoid-hemionitidoid, the ceratopteroid-parkerioid, and the pteroid ferns. Habitats occupied by these sub-groups are equally diverse, as they range from terrestrial, including epipetric and epiphytic, to even aquatic lifestyles.

We chose *Haplopteris ensiformis* among the epiphytic, vittarioid "shoestring ferns" for a detailed analysis of its two organelle genomes and transcriptomes. We found a typical, conserved chloroplast genome structure in *H. ensiformis* but identified a highly unorthodox mitogenome characterized by numerous active and inactive repeat sequences and a massive insertion of chloroplast DNA. Transcript maturation of the comparatively rich mitochondrial gene complement, lacking only *ccm* genes for cytochrome maturation, involves splicing of 24 group II and four group I introns and abundant C-to-U and U-to-C RNA editing at nearly 2,000 sites. The most surprising novelties of molecular evolution in a plant mitogenome include a unique novel *trans*-splicing group I intron in the large ribosomal rRNA and, most notably, extended stretches of bacterial, *Rickettsia*-like DNA in the *Haplopteris* mtDNA.

Results

Our choice of the "shoestring fern" *Haplopteris ensiformis* as a first leptosporangiate fern for complete assembly of both organelle genomes was based on pronounced variability in mitochondrial RNA editing and intron (co-)evolution in the monilophyte family Pteridaceae and the sub-family Vittarioideae in particular[22]. We used Next Generation Sequencing (NGS) data to assemble the chloroplast (cpDNA) and mitochondrial genomes (mtDNA) of *H. ensiformis*, accompanied by RNA-seq transcriptome analyses to study RNA processing with a special focus on intron splicing and RNA editing. As expected, chloroplast DNA reads dominated in the NGS data with read coverages of ca 1,800 to 4,200 and allowed the straightforward assembly of the *Haplopteris ensiformis* cpDNA. The mitochondrial DNA, in contrast, revealed overall lower and much more variable read coverages and its assembly was highly complicated by a multitude of repeats, long intergenic stretches and the insertions of foreign DNA and required multiple independent PCR amplifications for verification and complete assembly.

The *Haplopteris ensiformis* plastome and its well-resolved RNA editome

The *Haplopteris ensiformis* chloroplast DNA (148,805 bp) reveals a typical conserved circular plastome structure with a large (80,986 bp LSC) and a small (20,773 bp SSC) single copy region separated by a pair of inverted repeats (IRs of 23,523 bp each). The chloroplast genome carries 116 genes widely

conserved in other land plants, 85 of which encode proteins, including the recently characterized *ycf94* gene, four rRNAs and 27 tRNAs (Fig. 1A). Likewise, the *H. ensiformis* cpDNA contains a set of 20 conserved introns. Our accompanying transcriptome analysis confirmed functional splicing for all of them.

One striking structural difference concerns the presence of two morffo elements, "mobile ORFs in fern organelles"[25] in the chloroplast genome of the related species *Haplopteris elongata* (Fig. 1B). Our data suggest their secondary loss in the now determined cpDNA of *H. ensiformis* rather than an independent gain in *H. elongata* as we will discuss below in the context of the numerous cpDNA inserts that we identified in the now determined *H. ensiformis* mitogenome.

Modern RNA-Seq technologies allow for detection also of low-rate RNA editing sites with reasonable precision and any numbers of reported edits for an organelle transcriptome should nowadays ideally be accompanied by threshold criteria for their detection. Likewise, instead of the frequently used terms "complete" or "partial" editing, the percentage of detected base conversions by RNA editing should be given for the respective editing sites. The generally high coverage of RNA-Seq reads for the chloroplast transcriptome (mostly above 1,000 x and here reaching coverages of up to 250,000 x for the *psbE* gene in the case of the *H. ensiformis* cpDNA) allows for a detailed evaluation of RNA editing events with high precision, allowing for three digits behind the decimal point. Using stringent criteria for DNA and RNA read qualities (see Materials and Methods) we could identify 443 sites of chloroplast RNA editing covered by at least 100 RNA reads and RNA editing frequencies of at least 1.0 % (Suppl. Tab. 1).

We here use our previous nomenclature proposal for unequivocal labeling of RNA editing events [26], indicating the affected locus, the nucleotide resulting from C-to-U or reverse U-to-C editing (eU or eC), the position and, for the majority of edits within protein coding regions, the codon meaning before and after the edit (Suppl. Tab. 1). The careful analysis of RNA editing events contributes significantly to the identification of functional genes in the organelles or the dismissal of others as pseudogenes,

notably plant species like *H. ensiformis* featuring abundant C-to-U and reverse U-to-C RNA editing at the same time. A case in the point is the small reading frame *ycf94/orf51* of hitherto unknown function between *rps16* and *matK*, for which we here find even higher rates of RNA editing at two important sites (ycf94eU2TM and ycf94eU50PL) of 59% and 74%, respectively, than reported previously for other species [27,28]. *Vice versa*, we consider *rps16* a degenerating pseudogene in *H. ensiformis* as we could not confirm the expected removals of stop codons by reverse U-to-C editing.

The range of observed editing efficiencies extends from 99.1 % for the codon sense-changing editing event psbBeU116SL down to 1.0 % for edits accDeU657SS (silent), petBi6g2eU345 (intron) or edit rps2eU349R*, which unexpectedly introduces an early stop codon in the *rps2* coding sequence. Such edits at low frequency are likely "collateral" effects owing to lacking specificity of the chloroplast RNA editing machinery. The same holds true for most codon sense-changing edits that are unexpected (as they do not restore conserved codon identities) and likewise show only inefficient editing with low frequency (Suppl. Tab. 1).

Most of the detected non-silent RNA editing sites in the chloroplast coding sequences, however, confirm expectations for restoring conserved amino acid positions very well and are efficiently edited and, *vice versa*, we find low frequencies of editing nearly exclusively in silent or non-coding positions like 5'- or 3'-UTRs (Suppl. Tab. 1). However, exceptions exist: Prime example for efficient RNA editing events in non-coding regions, which could have been missed altogether in typical RT-PCR-based studies focusing on coding sequences, are cytidine-to-uridine conversions petBi6g2eU478 and rps12i346g2eU80 in the respective introns within *petB* and *rps12* of 94.0 % and 90.2 %, respectively. Conversely, we observe strikingly low efficiencies in many cases of reverse U-to-C editing removing stop codons, for example rpoC2eC232*Q reconstituting a glutamine codon in the *rpoC2* reading frame is edited with only 17.0 % efficiency in the transcript population.

Ten start codons and one stop codon are created by C-to-U editing and altogether 26 stop codons are removed by reverse U-to-C RNA editing in the chloroplast gene transcripts (Fig. 1, Suppl. Tab. 1).

We here use the *accD* gene as a somewhat less conserved protein coding region as an example for discussion (Fig. 2). Codon-changing edits confirm predictions very well with editing frequencies between 61.0% for accDeC730SP and 94.1% for accDeU779SL. Unpredictable edits in the UTRs or silent position accDeU657SS are edited much less efficiently except for accDeU-1 right upstream of the start codon created by editing (Fig. 2B). An intriguing case is editing site accDeU625HY inefficiently edited to 6.4%, for which a histidine or tyrosine is found variably in other taxa. An additional reverse edit is predicted for position accDeC580FL but remained unconfirmed. Note that in most fern cpDNA database entries, just two RNA editing sites are arbitrarily postulated to create an intact *accD* reading frame with a start codon edit in position 2 and to remove a stop codon in position 772.

Given that the organelle transcriptomes of *Haplopteris ensiformis*, and especially the mitochondrial transcriptome (see below), proved to be new examples for abundant C-to-U and U-to-C editing, we here use the opportunity to introduce a nomenclature amendment addressing the complex issue of multiple editings in individual codons. We suggest to indicate the individual and cumulative effects on codon meaning after and before a pipe symbol (|), respectively, and additional silent edits by an underline symbol (Suppl. Fig. S1). As an example we consider edits atpBeU1381PL|PS and atpBeU1382PL|PL changing a proline (P) codon identity in the *atpB* transcript. The first position edit takes place with 97.9% efficiency to expectedly reconstitute a conserved serine (S) codon whereas the unexpected additional 2nd position edit causes a change into a leucine (L) codon, although with only 1.7% efficiency. Yet more complex is the example of a CCC proline codon in the *rpoB* gene edited with different frequencies in all three positions. Edit rpoBeU662PF|PL in the second codon position causes the expected change towards a conserved leucine codon with 66.3% efficiency. However, unexpected edit rpoBeU661PF|PS in the first codon position with only 1.2% efficiency would, considered alone, cause a change to serine. In combination with the edit in the second position the codon is changed to a phenylalanine (F) codon (Suppl. Tab. 1, Suppl. Fig. 1). Moreover, these two non-silent edits are accompanied by a 3rd position edit with 36% efficiency (rpoBeU663PP_FF), which is silent for any of the 4 possible codon identities (P, L, S or F).

Assembly of the highly complex *Haplopteris ensiformis* mitogenome

The assembly of the *Haplopteris ensiformis* mitogenome turned out to be very demanding owing to a combination of several factors, which we will address in separate paragraphs below. The mtDNA sequence reads were not only ca. 10-fold less abundant than those of the cpDNA (on average ca. 150 x) but also much more variable in coverage (ca. 70 x to 600 x). We relied on parallel transcriptome analysis to verify authentic native mitochondrial genes characterized by RNA editing. Evident mtDNA contig assemblies ran into numerous repeated sequences, generally represented with higher read coverage or into extended insertions of laterally transferred chloroplast DNA fragments, a typical feature of vascular plant mitogenomes. Finally, we found numerous surprising similarities with *Rickettsia*-type bacterial genomes indicating multiple horizontal gene transfer (HGT) beyond the previously identified chlamydial tRNA genes in early branching tracheophytes [29].

Multiple recombination breakpoints allow for a huge spectrum of alternative, and likely co-existing, mitogenome arrangements with variable stoichiometries. For clarity, we chose to assemble twelve mtDNA contigs (A to L, ranging in sizes from 2,646 to 77,705 bp) comprising the full mitochondrial sequence complement into nine circular mtDNA chromosomes (chr1 to chr9) as separate GenBank database accessions (OM867545-OM867553). These nine circular chromosomes, however, likely represent only a substoichiometric minority of the truly existing mtDNA molecules owing to the numerous recombination breakpoints (Figs. 3 and 4). In total, 32 recombination breakpoints could be identified. These recombination breakpoints are labeled with ‘r’, the respective contig and consecutive numbers (rA1-rL2). We here display chromosome 5 comprising contigs A-C (Fig. 3 and chromosomes 1-4 and 6-9, variably integrating contigs D-K into the former (Fig. 4). Among these, contig A features prominently as it is flanked at its ends by repeat R9170 in inverted orientations, the largest repeat sequence that we identified in the *H. ensiformis* mitogenome (Fig. 3). Repeat R9170 carries the upstream part of the ribosomal rRNA gene cluster (Fig. 3), hence distantly reminding of the typical IRs in chloroplast genomes (see Fig. 1).

Repeats and Recombination in the *Haplopteris ensiformis* mitogenome

Other than by the many chloroplast DNA inserts that we will discuss below, the assembly of the *Haplopteris ensiformis* mtDNA was complicated by numerous repeated sequences of different sizes. Ancestral mitochondrial gene syntenies are widely eradicated in *H. ensiformis* owing to numerous recombination events – the *rps19-rps3-rpl16* gene continuity on contig F is one remaining exception (Fig. 4). While single-copy sequence contigs with mitochondrial genes had average NGS DNA read coverages around 150-fold (albeit with a broad distribution), identical repeat sequences mostly had coverages exceeding 300-fold. For clarity and discussion, we have labelled and annotated repeats with 'R' followed by the number of nucleotides, also in the corresponding GenBank accessions.

We carefully checked on recombinational activity across repeats with a "template-switch-avoiding" tsa-PCR strategy and examples are shown in figure 5. Repeat R596 (Fig. 5A) is a particularly intriguing example as it is not intergenic but shared as an identical sequence present in domains I of group II introns nad5i1242g2 and rrnLi833g2, located on contigs H and F, respectively (Fig. 4). R596 had a coverage of ca. 500 x whereas the flanking single-copy contigs with *nad5* and *rrnL* coding sequences had coverages of ca. 150 x and 350 x, respectively (Fig. 5A), indicating a non-equilibrium and different stoichiometries of the *nad5* and *rrnL* gene copies. Naturally, active recombination would cause dysfunctional chimeric *nad5*/*rrnL* genes in this case. Exploring potentially active recombination across R596, we confirmed the gene continuities for *nad5* and *rrnL* but found only very weak evidence for active recombination potentially giving rise to the two reciprocal arrangements (Fig. 5A).

However, similar tsa-PCR approaches run across other (intergenic) repeats indicated (stronger) active recombination across the respective repeated sequence resulting in more than only two co-existing, or at least strongly dominating, conformations of the flanking single-copy sequences. Examples are shown for R203 and R513 (Fig. 5B). Whereas all combinations of flanking sequences appear to co-exist for R203, the arrangements A-C and B-D appear to strongly dominate for R513 while arrangement A-D is stoichiometrically under-represented and arrangement B-C even remains undetected (Fig. 5B).

As an additional complication, some recombination points at the end of repeats were in in very close proximity resulting in complex "combined" repeats. We here show results for R180, one copy of which is in group II intron nad2i709g2 of the *nad2* gene (Fig. 5C). In another location, R180 is flanked on one side by consecutive repeats R137 and R259 resulting in alternative pathways to three different endpoints (D, E and F in Fig. 5C). Endpoints D and E result in intact *atp4* and *atp6* gene copies, respectively, whereas the alternative arrangement A-F creates pseudogene fragments for both genes downstream of *nad9*. Only very weak evidence is found for recombination of nad2i709g2 across R180 into any of the three alternative arrangements B-D, B-E and B-F or for the reciprocal arrangement A-C (Fig. 5C).

Other than identical sequence repeats in distant locations, the mitogenome of *Haplopteris* carries copies of variably diverging sequences, sometimes closely spaced. Examples are an imperfect 39 bp IR embedding the *cox1* gene or an inverted sequence repeat of 1.1 Kbp embedding the *nad9* gene with (only) 70% sequence identity between the two copies. Direct repeats of an 85 bp sequence are located behind the *nad1* gene overlapped by a repeated tridecanucleotide motif (CCTCTACTGAGGG) at their ends.

The mitochondrial gene complement in *Haplopteris ensiformis*

Despite its highly complex structure, the *Haplopteris ensiformis* mitogenome has a rich gene complement (Tab. 1). All expected genes for subunits of the respiratory chain complexes I-V (*nad, sdh, cob, cox* and *atp* genes) are present, including *sdh3* and *sdh4* encoding subunits of complex II. Likewise, there is a surprisingly large set of genes for ribosomal proteins, given that many are absent even in the mtDNAs of more ancestral lineages. Notable is the retention of *rps7,* which is lacking in all available hornwort and lycophyte mtDNAs but retained in *H. ensiformis*. Missing from the *Haplopteris* mtDNA is only the suite of *ccm* genes (*ccmB, ccmC, ccmF*) for cytochrome c maturation. Given their absence also in *Ophioglossum californicum* but their conservation in *Psilotum nudum* (Tab. 1) among the eusporangiate ferns, this evidently reflects a further independent loss of the *ccm* gene suite among ferns along with other phylogenetic deep losses in the lycophytes and hornworts.

The tRNA gene complement of the *H. ensiformis* mtDNA is particularly interesting owing to several tRNA genes of xenologous origin. Chloroplast-derived gene copies are present for trnF-GAA, trnMe-CAU, trnN-GUU, trnP-UGG and trnS-GGA (Tab. 1). Moreover, the *H. ensiformis* mtDNA also carries chlamydial-type tRNAs trnN-GUU (Fig. 4) and trnR-UCG (Fig. 3 and Fig. 4) described previously [20,29]. This results in xenologous genes coexisting with their native counterparts for trnF-GAA and trnP-UGG. Genes for native tRNAs are lost for trnL-CAA, trnMe-CAU, trnN-GUU, trnR-ACG. The chlamydial-type trnR-UCG is remarkable since it exists in three slightly differing copies. Similarly notable is the presence of a trnL-UAG in *Haplopteris* that is absent in the eusporangiate ferns (Tab. 1).

RNA Editing in *Haplopteris ensiformis* mitochondria

Given the complex mitogenome structure, the parallel transcriptome and RNA editing analysis was fundamental to identify functional mitochondrial genes in the *Haplopteris ensiformis* mtDNA. All protein-coding genes were found to be affected by RNA editing. Altogether, we identified 1,618 events of mitochondrial RNA editing: 1,091 of the C-to-U type and 527 edits of the reverse U-to-C type (Suppl. Tab. 2). The abundance of RNA editing is highly biased among genes with the *cox1* mRNA being affected by 145 RNA editing sites in contrast to the *atp1* mRNA with only three edits, respectively.

While the very high coverage of RNA-Seq reads in chloroplasts allowed for determination of editing frequencies with high precision, the more than threefold abundance of edits in mitochondria allowed for a better statistical classification of edits categorized by their location and effect (Suppl. Tab. 2). Among the total of 1,618 mitochondrial edits, 1,171 introduce codon changes and of the latter more than 900 are strongly predicted and more than 100 others moderately or weakly expected. We here illustrate the prediction of editing sites by PREPACT with the example of *atp9* finding a perfect match between identified edits and expectations and for *atp6* with only minor deviations from the expected editing pattern (Suppl. Fig. 2). Notably, the strongly expected RNA editing events have an average editing frequency of 83%, much higher than only 53% on average for non-predicted changes of codon identities (Suppl. Tab. 2). Yet lower RNA editing efficiencies are observed for silent sites or the ones in non-coding regions, e.g. only 29% on average in 3'-UTRs (Suppl. Tab. 2). A notable exception

from efficient editing of coding regions is the *matR* maturase encoded in the terminal *nad1* intron (Suppl. Tab. 2). An intriguing observation was made for *matR*, the conserved maturase encoded in nad1i728g2 (Tab. 1). Whereas identification of a start codon for this conserved, and only, mitochondrial group II intron maturase in flowering plants has been puzzling, we now find *matR* in *Haplopteris ensiformis* continuous with upstream *nad1* reading frame, accordingly, to be labelled *mat-nad1i728g2c* following a recent nomenclature proposal[28]. Although only lowly edited, the numerous, and expected, events of RNA editing reconstituting conserved codons and including 15 stop codon removals confirm the functional role of *mat-nad1i728g2c* (Suppl. Tab. 2).

Particularly remarkable is that 44% of the reverse U-to-C edits (233 of 530) serve to re-convert stop codons into arginine or glutamine codons, an important issue to distinguish functional from dysfunctional pseudogenes. For example, within 20 codons upstream of rps3i249g2, six stop codons are removed from the *rps3* coding sequence, including one within the intron binding site (Suppl. Tab. 2).

A yet more dramatic example is the *rpl6-rps13-rps11* cluster with 20 genomic stop codons in a short region (Fig. 6). In fact, the *rps11* gene at DNA level initially appeared to be an amino-terminally truncated pseudogene but turned out to have a proper start codon created by C-to-U editing and a total of seven stops removed by reverse U-to-C editing within the first 20 codons of its reading frame. We wished to test how different RT-PCR-based "classic" approaches would perform in comparison to the RNA-Seq approach to detect RNA editing sites. Towards that end, we used three different strategies for cDNA synthesis using either random hexamers or specific primers targeting the 3'-end of *rps11* in edited or non-edited versions. Here, we made use of two edits in the 3'-UTR closely behind the rps11eU466Q* stop codon editing. Sequencing of an internal amplicon revealed that many editing events were not confirmed in the cDNA sample primed with random hexamers with better performance by the specific primers and notably the one for the edited version of the 3'-end.

A striking bias concerns "silent" edits leaving codon identities unchanged. We observed only 32 silent U-to-C edits but the nine-fold amount (282) of silent C-to-U edits. Interestingly, silent C-to-U edits are frequently found to neighbor non-silent sites ("NESIs") as has previously been observed for the huge editome in the *Selaginella uncinata* chloroplast[30].

Mitochondrial introns in *Haplopteris ensiformis* include a novel *trans*-splicing group I intron.

The *Haplopteris ensiformis* mitogenome shows notable differences to the intron complements in the two eusporangiate fern taxa (Tab. 1). Ancient group II introns nad1i477g2, nad1i669g2, nad5i1477g2 and nad7i917g2 are lost from the *H. ensiformis* mtDNA. *Vice versa*, group I intron cox1i395g1 and group II introns cox2i373g2 and rps14i114g2 in the *Haplopteris* mitogenome lack counterparts in *Ophioglossum* and *Psilotum* (Tab. 1). The group II introns and cox1i395g1 are evidently of ancient origin in the land plant lineage. The latter has previously been identified in liverworts but also in the leptosporangiate ferns and in the horsetail *Equisetum arvense*[31,32].

A striking example documenting the *Haplopteris ensiformis* mitogenome complexity concerns the ribosomal rRNA cluster with a peculiar arrangement featuring a disrupted *rrnL* gene (Fig. 3, Fig. 4). Maturation of the large ribosomal RNA requires *trans*-splicing of a broken group I intron, rrnLi825g1. Secondary structure modelling suggests base-pairing interaction of the two intron parts in the disrupted domain P9.0/P9.1 (Fig. 7). Intriguingly, rrnLi825g1 has a distant positional orthologue as a *cis*-splicing homologue in the charophyte alga *Chara vulgaris*[33]. Despite overall only weak similarity, conserved regions include intron domains P7 and P8 known to contribute to the reactive core of group I introns. Three further introns are located in the downstream part of the *H. ensiformis rrnL* gene. Group II intron rrnLi833g2 is located only 8 bp downstream of the rrnLi825g1 3'-splice site and has distant homologues in liverwort mtDNAs. The two downstream group I introns rrnLi1897g1 and rrnLi1928g1 have hitherto only been identified serendipitously in *rrnL* gene samplings of other Polypodiales species. Notably, none of the four introns are present in the eusporangiate ferns, which feature continuous *rrnL* genes.

Laterally transferred chloroplast DNA fragments in the *Haplopteris* mitogenome

Among other issues, the assembly of the *Haplopteris ensiformis* mitogenome was much complicated by the fact that it contains ca. 80 inserts of chloroplast DNA of variable sizes and with variable degrees of sequence conservation. Similar to the repeat sequences, we annotated these cpDNA inserts with numbers indicating their sizes in base pairs preceded by 'cp' (Suppl. Tab. 3). While the separate chloroplast sequence inserts may have originated from fragmentation after insertion of large stretches of cpDNA, their different degree of sequence conservation rather argue for independent transfer events and likely document independent cpDNA insertions at different time points in evolution (Fig. 8). As an example, an array of seven likely independently acquired cpDNA inserts is present in the intergenic region between *nad5* and *sdh4* (Fig. 8A). This region includes cp4165, the largest continuous stretch of "promiscuous" cpDNA derived from the *ndhH-ndhA-ndhI-ndhG-ndhE* region, sharing 93 % sequence identity with the native chloroplast counterpart. This stretch is directly flanked by cp1271 derived from the chloroplast IR region encoding the *rrnL* gene and sharing even 99 % of identical nucleotides, likely indicating a yet more recent inter-organellar migration into the mitogenome. Chloroplast insert cp1039 downstream of *rps4* (Fig. 8B) is another example for a likely recently acquired lateral sequence transfer. In a phylogenetic analysis, it branches next to the homologous sequence from the *atpA-atpF* region of the now determined *Haplopteris ensiformis* cpDNA as sister to the counterpart in *H. elongata*, even despite the generally high sequence drift among Pteridaceae (Fig. 8B).

At the other end of the spectrum, some cpDNA inserts lack detectable homologous sequences in the *Haplopteris* chloroplast genomes and could only be recognized by similarities with the cpDNA of other taxa. Inserts cp364 and cp749 (Fig. 8A) are examples, which could only be identified by their similarity to sequences in the IR regions in the cpDNAs of genera like *Asplenium* or *Vittaria*. Yet more striking are the cases of morffos, the highly variable "mobile ORFs in fern organelles"[25]. The chloroplast sequence insert cp2126 is an example along those lines (Fig. 8C), which includes a morffo element that

has a top sequence identity of 75% in the cpDNA of *Hemionitis subcordata* in the distant subfamily Cheilanthoideae.

Yet more striking is the case of morffo2, which is intact in the *Haplopteris elongata* cpDNA but truncated to its 5'-terminal 378 bp in the now determined *H. ensiformis* chloroplast genome (Fig. 1B). Intriguingly, chloroplast insert cp1712 in the mtDNA (Suppl. Tab. 3) includes significantly more of the 5'-end of the truncated morffo2 element (628 bp). Taken together, the slow sequence drift in the mitogenome offers examples allowing for a "molecular archeology" for former chloroplast DNA sequences that are not present any more in the recent chloroplast genome. In the latter case, it further supports the point for degeneration of the morffo elements in *H. ensiformis* rather than their independent origin in *H. elongata* (Fig. 1B).

Rickettsial DNA in the *Haplopteris ensiformis* mitogenome

The chlamydial-type *trnN-GUU* gene initially identified in the lycophyte *Phlegmariurus squarrosus*[29] and the chlamydial-type *trnR-UCG* gene found subsequently in the eusporangiate fern mitogenomes[20] are now also identified in the *Haplopteris ensiformis* mtDNA, further corroborating the concept of HGT from bacteria into the mitogenomes of early-branching vascular plants. Most strikingly, we now discovered numerous inserts of "Rickettsia-like" sequences in the *H. ensiformis* mitogenome. Altogether, we identified 30 variably degenerated protein coding regions evidently derived from rickettsial bacteria (see supplementary table 3). Similar to the labels for repeats and cpDNA inserts, we annotated the xenologous bacterial DNA inserts indicating their respective extension in base pairs, in this case preceded by an 'x' (Suppl. Tab. 3, Fig. 9).

We carefully verified the surprising observation of Rickettsia-like DNA inserts in the *Haplopteris ensiformis* mitogenome by PCRs anchoring in the flanking mtDNA regions and consistently corroborated the mitogenome assemblies, as we here exemplarily show for x625 representing a central coding region for the bacterial DNA recombination protein RmuC (Fig. 9A). Towards that end we used independent DNA preparations from two *H. ensiformis* isolates from separate locations and,

385 *vice versa*, included material from *Vittaria lineata*, a closely related species that grows next to one of
386 the *H. ensiformis* populations in the Botanic Garden Bonn. PCR products were consistently obtained
387 for the two independent *H. ensiformis* isolates, but not for the *V. lineata* sample (Fig. 9A). Sequencing
388 of the PCR products fully confirmed the mitogenome assembly, interspersed by Rickettsial-like DNA
389 insertions. Moreover, none of the bacterial insert sequences pointed to contamination by living
390 bacteria as they mostly revealed characteristic degeneration of the protein coding genes to
391 pseudogenes. Finally, we observed continuities of the average read coverages continuing from flanking
392 mitochondrial sequences into the bacterial DNA inserts (Fig. 9). We do not assume functional
393 expression of the xenologous bacterial genes given that we could detect only negligible RNA coverage
394 for some of the regions carrying bacterial DNA inserts.

395 After insertion into the *Haplopteris* mtDNA, the Rickettsia-type coding sequences degenerated by
396 recombination and sequence drift. Some protein sequence similarities, however, remain astonishingly
397 high, likely indicating very recent horizontal gene transfers (Suppl. Fig. 3 and Suppl. Fig. 4). Specifically,
398 top similarities were often observed with the corresponding sequences from *Caedimonas*
399 *varicaedens*[34], followed by slightly lower similarities with homologous loci in *Cand. Paracaedimonas*
400 *acanthamoebae*, other *Caedibacter spp*. or *Cand. Nucleicultrix amoebiphila*. These species belong to
401 the family Holosporaceae among Rickettsiales *sensu lato*. The family Holosporaceae is alternatively
402 considered to be a separate order of its own, the Holosporales.

403 As in the case of the chloroplast sequence inserts, the variable degrees of sequence degeneration
404 suggest independent events of horizontal transfers at different time points in evolution. Alternatively,
405 the insertion of larger xenologous genomic regions followed by later fragmentation in the mitogenome
406 followed by different degrees of sequence degeneration may be possible. One evident example is the
407 bacterial *HscA-RmuC* region (Fig. 9B) located between the *nad1* gene and cp232 (Fig. 9B, see Suppl.
408 Tab. 3). The *HscA* coding region is disrupted by an insert in the coding sequence and the downstream
409 *RmuC* gene is truncated with its central region located as x625 between the *cox2* gene and the *rrnL-*
410 *rrnS* gene cluster, directly flanking cpDNA insert cp416 (Fig. 9A). Despite these discontinuities, the

411 degree of amino acid sequence conservation reaching up to 94 % identity is astonishing and
412 phylogenetic analysis allows a close affiliation of the xenologous RmuC region in the *Haplopteris*
413 mitogenome with *Caedimonas varicaedens* among the Holosporales (Fig. 9). Another xenologous
414 insert x888 (Fig. 9B) carrying parts of the coding region for OMBB, an outer membrane beta-barrel
415 domain containing protein, is much more degenerated and has significantly lower similarities to top-
416 scoring hits with less clearly defined Rickettsiales bacteria (33% identity, 50% similarity).

417 Whereas most of the xenologous bacterial DNA inserts show fragmentation and considerable
418 degeneration and of protein-coding regions, there are also striking counterexamples like x1850, an
419 insert of ca. 1.8 Kbp, containing the *serS-surE-nlpD* region located embedded between cp763 and
420 cp1456 (Suppl. Tab. 3). The coding region of the upstream serine tRNA ligase SerS is N-terminally
421 truncated and carries two stop codons but the reading frame of the downstream coding region for the
422 5'/3'-nucleotidase SurE is perfectly conserved and shares even 97 % sequence similarity with its
423 *Caedimonas varicaedens* counterpart (Suppl. Fig. 3). In this case a similarity of 81% can even be
424 identified at the nucleotide sequence level (Suppl. Fig. 4), clearly suggesting the bacterial donor to be
425 closely related to *C. varicaedens.* Intriguingly, the end of this bacterial sequence insert contributes to
426 repeat R295.

427 Examples for particularly large stretches of Holosporaceae DNA inserts include x2170 derived from
428 the bacterial XpsD-region located behind the mitochondrial *nad4L* gene (Suppl. Tab. 3) and x1920,
429 originating from the bacterial MfdD-RimM-TrmD-RpL19 region inserted upstream of *rps4* (Fig. 9C,
430 Suppl. Tab. 3). In the latter case, the gene for MfdD is severely 5'-truncated, the gene for RimM appears
431 intact, the TrmD reading frame is degenerated by a stop codon and the RpL19 sequence is 3' truncated
432 (Fig. 9C).

433 In some cases, top sequence similarities clearly identify an origin of HGT sequences from
434 Rickettsiales, but not necessarily from within the Holosporaceae. Most notable is an assembly of
435 sequence inserts of altogether more than 7 Kbp downstream of the *atp1* gene (Suppl. Tab. 3) that

carries the suite of genes encoding murA and two CoA-carboxylase subunits in opposite direction, which are nearly unaffected by sequence degeneration except for in-frame stop codons (Suppl. Fig. 3). At nucleotide level, a top similarity of 74 % identical nucleotides is observed with a not further identified Rickettsiaceae bacterium isolate PMG_002 (Suppl. Fig. 4). Intriguingly, the VirB8/B9 coding sequences of two P-type conjugative transfer proteins are located ca. 1.2 Kbp downstream of the murA homology. Top nucleotide sequence similarities are again observed for Rickettsiaceae bacterium isolate PMG_002. However, the virB8/B9 and the MurA coding sequences are not connected in that genome, indicating either a related and yet unidentified Rickettsiales donor, separate gene transfers or subsequent rearrangements after copy-transfer into the *Haplopteris* mtDNA.

Finally, two large stretches of protein coding regions are located upstream of *nad9*, running in opposite direction (labelled Bact-ORF 1 and 2, respectively), but cannot be assigned taxonomically owing to much lower similarities of only 40-50% at protein sequence level: A hypothetical ORF, possibly RNA polymerase, highest similarity with a *Magnetovibrio sp*. database entry (MBM08139) and a DNA-polymerase, highest similarity with a *Zoogloea sp*. sequence entry (KAB2964018).

Discussion

A first leptosporangiate fern mitogenome assembly

The large Pteridaceae family contains about 10 % of extant ferns species[35–37]. Among them, the Adiantoid and Vittarioid sub-families (see Fig. 8) show particularly high levels of substitution rate heterogeneity[38,39]. We became particularly interested in the taxon given the apparently dynamic evolution of mitochondrial introns and RNA editing[22]. Likewise, a striking diversity of chloroplast RNA editing has been reported in the genus *Adiantum* [40].

The genomes of the two endosymbiotic organelles in *Haplopteris ensiformis* here reported are prime examples highlighting the discrepancy between the conservative evolution of chloroplast DNA and the highly dynamic evolution of vascular plant mitochondrial genomes. Among the recently reported flowering plant mitogenomes, the one of the holoparasite *Ombrophytum subterraneum*[41] is

a case in point documenting not only a multi-chromosomal structure but multiple evidence for HGT from its host plants. Like in other cases of multi-chromosomal plant mitogenomes, e. g. the hundreds of different mitochondrial chromosomes in some *Silene* species, no efforts are made any more to come up with the display of a hypothetical, and likely misleading "master-circle"[42,43].

As a first representative for the large clade of leptosporangiate ferns with ca. 10,000 species, the *H. ensiformis* mitogenome again adds to the list of astonishing molecular peculiarities that are hallmarks of vascular plant mtDNAs[44,45]. Given its extraordinary complexity, it comes as no surprise that no other leptosporangiate mitogenome has previously been assembled despite multiple NGS efforts including the water fern genera *Azolla* and *Salvinia*[46,47] or the "flying spider-monkey" tree fern *Alsophila spinulosa*[48] or, most recently, the model ferns *Adiantum capillus-veneris*[49] and *Ceratopteris richardii*[50]. Likewise no mitogenome was reported in a genome assembly effort for the lycophyte *Isoetes taiwanensis*[51], possibly owing to even higher complexity than the ones reported previously for *Isoetes engelmannii* and *Selaginella moellendorffii*[7,8].

We hence speculate that highly complex mitogenomes like the one reported here may be a general feature of leptosporangiate ferns also outside of the Polypodiales. Intriguingly though, and despite its highly dynamic structural evolution, the *H. ensiformis* mtDNA contains a surprisingly rich set of "classic" mitochondrial genes (Tab. 1) when compared to the gene complement of other taxa including the eusporangiate ferns[52].

Mitochondrial intron dynamics in ferns

More notable than the gene complements are the diverging mitochondrial intron complements now identified in *Haplopteris ensiformis* in comparison to the previously analyzed eusporangiate ferns, extending earlier conclusions that much more intron dynamics is present in monilophytes in comparison to their seed plant sister clade[21,22,32,52]. The *Haplopteris* mitogenome reveals retention of evidently ancient introns that have been lost in the eusporangiate ferns like cox1i395g1, cox2i373g2, rps14i114g2 and rrnLi833g2 (Tab. 1).

Surprisingly though, the heavily recombining mtDNA of *Haplopteris ensiformis* has not resulted in disrupted group II introns like in angiosperms or in gymnosperms where ever more transitions to *trans*-splicing have been observed recently[53]. In contrast, most group II introns (11 of 15) found to be *trans*-splicing in at least one seed plant lineage (cox2i373g2, nad1i394g2, nad1i728g2, nad2i542g2, nad2i1282g2, nad4i461g2, nad4i976g2, nad4i1399g2, nad5i1455g2, nad7i209g2 and rpl2i846g2) are present in conventional *cis*-arrangements in the *Haplopteris* mtDNA fully in line with the early evolutionary conclusion that *trans*-splicing introns in seed plants originate from *cis*-arranged ancestors in early-branching plant lineages[54,55]. We speculate that transitions from *cis*- to *trans*-splicing group II introns may rely on co-evolving protein splicing factors that are present in the seed plant but not in the monilophyte lineage. The remaining four group II introns known to exist in a *trans*-splicing state in at least some spermatophytes (cox2i691g2, nad1i669g2, nad5i1477g2 and nad7i917g2) have been lost altogether from the *H. ensiformis* mitogenome (Tab. 1).

In the light of the above it is all the more surprising to find rrnLi825g1 as a "novel" group I intron of yet unclear ancestry in a *trans*-splicing arrangement in the idiosyncratic *rrnL* gene makeup in the *H. ensiformis* mitogenome. In contrast to the numerous examples of *trans*-splicing group II introns in plant organelles alone, *trans*-splicing group I introns appear to exist much more rarely in nature. First reports of *trans*-splicing group I introns in the *Isoetes engelmannii* mitogenome[7] and in the mtDNA of *Trichoplax*[56] have been followed by recognition of *trans*-splicing group I intron cox1i744g1 in *Helicosporidium* mtDNA[57] and of two *trans*-spliced group I introns in *Gigaspora margarita* mtDNA[58]. Remarkably, the *trans*-splicing group I intron cox1i395g1 in the *Isoetes engelmannii* mtDNA[7] exists in a conventional, *cis*-arranged version in the *H. ensiformis* mitogenome.

The mitochondrial *rrnL* gene is entirely devoid of introns not only in seed plants and lycopyhtes but also in hornworts[28] and even in the eusporangiate ferns[20] whereas it features four introns in the *Haplopteris ensiformis* mtDNA: rrnLi825g1, rrnLi833g2, rrnLi1897g1 and rrnLi1928g1 (Fig. 7). The latter two group I introns have already been documented serendipitously in sequence samplings covering parts of the mitochondrial *rrnL* gene[59] and are conserved also outside of Pteridaceae in taxa as least as

512 distant as tree ferns (Cyatheales, e.g. *Plagiogyria stenoptera*, accession DQ647877). The origins of
513 these introns remain unclear as they neither share sequence similarities anywhere else in the plant
514 lineage nor among fungi as is the case for the "rampant invader" group I intron cox1i726g1 sporadically
515 occurring in angiosperms[60,61].

516 ## Verifying native and foreign sequences in the organelle genomes

517 The examples of the *Haplopteris ensiformis* organelle genomes reported here document that parallel
518 analysis of transcriptome along with genome NGS sequencing data is essential to characterize genes
519 as functional or dysfunctional owing to the complex maturation processes including C-to-U and U-to-
520 C RNA editing and *cis*-splicing or *trans*-splicing of split genes and introns. Moreover, the highly complex
521 mitogenome of *H. ensiformis* is a prime case showing that very careful investigations of "alien"
522 sequences in NGS data to tell them apart from native chloroplast DNA or bacterial contaminations may
523 be needed in such complex cases of organelle genome structures. Aside from our experimental
524 verifications, we note that parts of rickettsial DNA inserts in *H. ensiformis* also seem to be present in
525 database entries reporting partial mtDNA sequences of the ferns *Asplenium nidus* (FR669448) and
526 *Dryopteris crassirhizoma* (MW732172). Moreover, we suggest the careful re-evaluation of some fern
527 cpDNA entries (like *Dipteris conjugata* KP136829, *Polypodium vulgare* MT984517, *Cystopteris protrusa*
528 KP136830 or *Selliguea yakushimensis* MN623352) that seem to contain mtDNA stretches as possible
529 artefacts. Certainly, however, once verified and when the likely similarly complex mitogenomes of
530 those taxa would be assembled in the future they may document interesting inter-organellar DNA
531 transfer from mitochondria to chloroplasts.

532 ## Lateral sequence transfers: cpDNA insertions in the *Haplopteris* mitogenome

533 The first evidence for lateral transfer of "promiscuous" chloroplast DNA into a plant mitogenome has
534 already been documented 40 years ago[62] and thereafter found to be merely a standard feature in many
535 seed plant mtDNAs. Most interestingly, among the numerous cpDNA insertions in the mtDNA of
536 *Haplopteris ensiformis* we now find examples evidently documenting ancient cpDNA features that
537 even are not present any more in the recent plastomes, as here seen for the mysterious morffo

elements. The origin and dynamics of these only recently described "mobile ORFs in fern organelles"[25] is presently still enigmatic. Here, we find that apparently intact morffos in the cpDNA of *Haplopteris elongata* have evidently disintegrated in the plastome of *H. ensiformis* (Fig. 1B) but that their counterparts and evidence for yet other morffos is present in its mitogenome as "evolution's mis-placed witnesses".

Origins of bacterial sequences in the *Haplopteris* mtDNA

The role of Horizontal Gene Transfers (HGT) is increasingly appreciated not only for bacterial evolution but also in the evolution of eukaryotic genomes including ferns[63,64]. In several cases, host-parasite interactions are key to the events of HGT[41,65] and such interactions in nature may also be responsible for gene transfer into fern mtDNA[66]. In particular, after first reports on horizontal plant-to-plant transfer of mtDNAs[67,68], it is meantime well understood that HGT has contributed to many seed plant mitogenomes[69]. The most outstanding example is the case of the early-branching flowering plant *Amborella trichopoda* having integrated into its mitogenome not only numerous stretches of mtDNA from other angiosperms but also the near-complete mitogenomes of two mosses[19,70]. A new dimension of HGT into plant mitogenomes opened up with the discovery of tRNA genes from chlamydial origins into very early tracheophyte lineages[29]. Yet more recently, it was found that sequences of fungal origin have been horizontally transferred early into mitogenomes of the Orchid family[71].

Here, we now report on a multitude of Rickettsia-like genome insertions as one of the most astonishing findings emerging from the assembly and analysis of the *Haplopteris ensiformis* mitogenome. Intriguingly, Rickettsiales are known to be obligate intracellular parasites and at the same time assumed to be the extant alpha-proteobacterial lineage most closely related to the progenitor of the eukaryotic mitochondrion. The bacterial DNA inserts in the *H. ensiformis* mtDNA are most likely derived from species closely related to (Cand.) *Caedibacter acanthamoebae*[34] of the Holosporales (or, alternatively, Holosporaceae among Rickettsiales). *Caedibacter* (or *Caedimonas*) endosymbionts transfer the "killer trait" to their *Paramecium* hosts[72,73]. Other Rickettsia are known to be associated

with arthropods, leeches and protists and Rickettsia-like organisms (RLOs) and are associated not only with human or animal diseases but also with numerous plant diseases, for example the Rickettsia endosymbiont of the tobacco whitefly *Bemisia tabaci*. Moreover, Rickettsia have also been associated with a papaya disease[74] and have been identified eustigmatophyte algae[75] and in the green alga *Mesostigma viride*[76]. It will be very interesting to see whether Rickettsial or related bacterial DNA insertions will also be identified in further leptosporangiate mitogenomes and to ultimately identify the exact donor species and the biological mechanisms of the HGT processes. Notably, the xenologous bacterial DNA regions identified in the *H. ensiformis* mitogenome include similarities to the IS481 family transposase and virB8 and virB9 homologs (Suppl. Tab. 3). The latter genes are commonly found among mobile IS elements and are associated with conjugative gene transfer amongst Rickettsia species[77].

Methods

Plant material

The Bonn University Botanic Garden kindly provided plant material for *Haplopteris ensiformis* (xx-0-BONN-24687) and *Vittaria lineata* (xx-0-BONN-17295). Species identities were independently verified by PCR amplification and sequencing of the *rbcL* and *atpA* locus revealing complete sequence identities with independent sequence accessions (*Haplopteris ensiformis* KX164999 and MH359250 and *Vittaria lineata* EF473712 and KU744782).

Organelle genome sequencing and assembly

The Qiagen DNeasy Plant Mini Kit was used for DNA isolation and the Sigma-Aldrich plant RNA isolation Kit for RNA preparation followed by ribosomal RNA depletion using the RiboMinus Plant Kit for RNA-Seq (Thermo Fisher Scientific). RNA quality was checked with a Qubit Fluorometer for RIN-values of at least 0.9. Genome and transcriptome sequencing (paired-end whole genome sequencing and RNAseq) was done commercially at the BGI on an Illumina platform. Raw read data were evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). No adapter artifacts could be detected. The MEGAHIT software[78,79] was used for *de novo* whole genome assembly with independent

runs using three different settings ("strict", "default" and "relaxed"). The "strict" assembly was run with the minimum k-mer option available. The third assembly was run with parameters set for a high ultra-complex metagenomics dataset). Complete assembly of cpDNA was performed with NOVOPlasty[80] using contigs from the MEGAHIT assembly with conserved chloroplast genes as seeds.

RNA reads were assembled with the Trinity software[81,82]. The BLAST[83] suite was used to initially identify contigs with chloroplast or mitochondrial gene content using available lycophyte and fern organelle genomes. Raw read data and assembly data have been deposited under BioProject accession no. PRJNA862965. Whole genome assembly contigs could be clustered into to two sets based on k-mer coverage (mt/cp 1:10). After connections of contigs were verified by PCRs (see below), contigs of the mitochondrial genome were connected by hand.

Verification of mtDNA arrangements

PCRs were used to independently verify the highly complex arrangements of the *Haplopteris ensiformis* mitochondrial DNA resulting from multiple repetitive sequences and insertions of chloroplast and xenologous DNA. PCR amplicons were designed with primers anchoring in neighboring sequence regions of evident mitochondrial identity, preferably coding regions. Special care was taken to investigate repeated sequence for potential recombination creating alternative arrangements of flanking sequences. To best avoid false positives suggesting active recombination resulting from artificial template switches we used a strategy of template-switch avoiding "tsa"-PCRs. To that end, a mix of gel-eluted PCR fragments containing a repeat sequence in different sequence environments (AB and CD) was used to obtain products reflecting a reciprocal exchange of flanking regions (AD and CB). A series of template dilutions (1:10, 1:20, 1:30, 1:40, 1:50 and 1:60) and numbers of PCR cycles (15, 20, 25 and 30) were tested and adjusted to determine the threshold for artificial production of template-switch products.

Identification of mitochondrial genes avoided routine pipelines but started from homologues in the mtDNA of diverse taxa, including the liverwort *Marchantia polymorpha*, the lycophyte *Phlegmariurus*

squarrosus and the gymnosperm *Ginkgo biloba*. Gene identities were verified by the parallel transcriptome studies to confirm intron splicing and C-to-U and U-to-C RNA editing. Identification of tRNA genes combined the use of tRNAscan-SE[84] and sensitive BLASTN searches using a tRNA query set including the recently identified chlamydial tRNA xenologues in early tracheophytes[20,29]. To identify DNA similarities including repeats, chloroplast DNA or xenologous bacterial DNA insertions, we used sensitive BLASTN or XBLAST similarity searches (word sizes = 7 or 3, respectively) and strict random expectancy threshold cutoffs of 1e-10. On nucleotide level, this translates approximately into identification of identical repeated sequences of ca. 40 bp (i.e. slightly larger than the conserved domain V of group II introns) or respective larger, but less similar, regions.

Transcriptome studies and determination of RNA editing sites

Transcriptome studies were used to determine all intron splicing sites. The identification of RNA editing sites was done as previously described[28]. Briefly, DNA and RNA reads were mapped against the organelle contig sequences using GSNAP[85] and JACUSA[86] was used to determine RNA-DNA differences. Thresholds were set to minimally 30 reads and RNA editing efficiencies of at least 1% for chloroplast and at least 5% for mitochondrial transcripts for a strict determination of C-to-U and U-to-C RNA editing sites, respectively. Selected loci were analyzed independently by RT-PCR-based cDNA analyses and sequencing for various reasons like unexpectedly inefficient RNA editing at certain sites or because of coexisting pseudogene copies as discussed under results (chloroplast *rpoC1* and mitochondrial genes *atp1*, *atp8, nad5, nad7, rrnL* and the *rpl6-rps13-rps11* co-transcript).

Data availability and Sequence accessions

Haplopteris ensiformis primary nucleotide sequence reads are submitted to the sequence read archive (SRA) under BioProject accession number PRJNA862965. The assembled chloroplast genome is deposited under accession number OM867544 and the assembled mitogenome chromosomes are available under accession numbers OM867545 to OM867553.

Acknowledgements

The authors wish to thank the Molecular Evolution group for stimulating discussion and productive interaction, especially Philipp Gerke for his contributions to establish pipelines for the analysis of RNA editing in NGS data. The authors gratefully acknowledge the support for this project by computing time provided by the Paderborn Center for Parallel Computing (PC^2).

Author contributions

SZ did wet lab work and established bioinformatic pipelines. MP helped with nucleic acid preparations and molecular cloning. SZ and VK analyzed data and prepared figures. VK wrote the manuscript and all authors edited and approved the final manuscript version.

Table 1. The mitochondrial gene and intron complement of *Haplopteris ensiformis*.

Gene / intron	*Psilotum nudum*	*Ophioglossum californicum*	*Haplopteris ensiformis*	Gene / intron	*Psilotum nudum*	*Ophioglossum californicum*	*Haplopteris ensiformis*
atp1	+	+	+[a]	*rps1*	+	+	+
atp4	+	+	+	rps1i25g2	+	0	+
atp6	+	+	+	*rps2*	+	+	+
atp8	+	+	+[a]	*rps3*	+	+	+
atp9	+	+	+	rps3i74g2	+	+	+
ccmB	+	0	0	rps3i249g2	+	0	+
ccmC	+	0	0	*rps4*	+	+	+
ccmF	+	0	0	*rps7*	+	+	+
ccmFCi829g2	+	0	0	*rps10*	+	+	+[a]
cob	+	+	+[a]	*rps11*	+	+	+
cox1	+	+	+	*rps12*	+	+	+
cox1i395g1	0	0	+	*rps13*	+	+	+
cox1i624g2	+	0	0	*rps14*	+	+	+
cox2	+	+	+	rps14i114g2	0	0	+
cox2i373g2	0	0	+	*rps19*	+	+	+
cox3	+	+	+	*rrn5*	+	+	+
matR	+	+	+	*rrnL*	+	+	+
nad1	+	+	+	rrnLi825g1	0	0	T
nad1i258g2	+	+	+	rrnLi833g2	0	0	+
nad1i394g2	+	0	+	rrnLi1897g1	0	0	+
nad1i477g2	+	+	0	rrnLi1928g1	0	0	+
nad1i669g2	+	+	0	*rrnS*	+	+	+
nad1i728g2	+	+	+	*sdh3*	+	+	+
nad2	+	+	+	*sdh4*	+	+	+
nad2i156g2	+	+	+	*tatC*	+	+	+[a]
nad2i542g2	+	+	+	*trnA-ugc*	+	+	+[b]
nad2i709g2	+	+	+	*trnC-gca*	+	+	+
nad2i1282g2	+	+	+	*trnD-guc*	+	+	+
nad3	+	+	+	*trnE-uuc*	+	+	+
nad4	+	+	+	*trnF-gaa*	+	+	+
nad4i461g2	+	+	+	*trnF-gaa cp*	0	0	+[c]
nad4i976g2	+	+	+	*trnG-gcc*	+	+	+
nad4i1399g2	+	+	+	*trnG-ucc*	+	+	+
nad4L	+	+	+	*trnH-gug*	+	+	+[b]
nad5	+	+	+	*trnI-cau*	+	+	+
nad5i230g2	+	+	+	*trnK-uuu*	+	+	+
nad5i1242g2	+	0	+	*trnL-caa*	+	-	+
nad5i1455g2	+	+	+	*trnL-uaa*[b,c]	+	+	+
nad5i1477g2	+	+	0	*trnL-uag*	0	0	+
nad5i1872g2	+	+	+	*trnMe-cau cp*	+	+	+[c]
nad6	+	+	+	*trnMi-cau*	+	Ψ	+
nad7	+	+	+	*trnN-guu cm*	0	0	+
nad7i140g2	+	+	+	*trnN-guu cp*	+	+	+[c]
nad7i209g2	+	0	+	*trnP-ugg*	+	+	+
nad7i676g2	+	+	+	*trnP-ugg cp*	0	0	+[c]
nad7i917g2	+	+	0	*trnQ-uug*	+	+	+
nad9	+	+	+	*trnR-acg*	+	+	0
rpl2	+	+	+	*trnR-ucg cm*	+	+	+[b]
rpl2i846g2	+	+	+	*trnR-ucu*	+	+	+
rpl5	+	+	+	*trnS-gcu*	+	+	+
rpl6	+	+	+	*trnS-gcu cp*	0	0	+[c]
rpl16	+	+	+	*trnS-gga cp*	0	0	+
				trnS-uga	0	+	+
				trnV-uac	0	+	+
				trnW-cca	+	+	+
				trnY-gua	+	+	+[b]

Table 1. List of mitochondrial genes and group I (g1) and group II (g2) introns in the *Haplopteris ensiformis* mitogenome in comparison to the ones of the eusporangiate ferns *Psilotum nudum* and *Ophioglossum californicum*. The added 'cm' or 'cp' indicate chloroplast-derived or chlamydial-type tRNA genes, respectively. Superscript addendums: 'a' indicates co-existing large pseudogene copies, 'b' indicates co-existing functional copies with only minor sequence differences, 'c' indicates that

chloroplast tRNA genes are part of extended cpDNA inserts. Features distinguishing the *Haplopteris ensiformis* mtDNA from both eusporangiate ferns are highlighted in red, i.e. the presence of a chlamydial *trnN-guu* gene, of chloroplast-derived *trnS-gcu* and *trnS-gga* genes, of introns cox1i395g1, rps14i114g2 and of four introns in the *rrnL* gene including the *trans*-splicing intron rrnLi825g1 vs. the absence of a *trnR-acg* gene and four other group II introns. The maturase "*matR*" in the terminal *nad1* intron, systematically labelled mat-nad1i728g2c, is in-frame with the upstream *nad1* coding region in *H. ensiformis*.

Figure Legends

Figure 1. The *Haplopteris ensiformis* cpDNA.

A. *Haplopteris ensiformis* reveals a typical plant circular plastome structure consisting of a large (LSC) and a small (SSC) single-copy region separated by a pair of inverted repeats (IR) and an expectedly conserved, ancestral gene and intron complement. The genome map was created using OGDRAW [87]. Gene categories are indicated in the legend. Numbers in parentheses indicate the amount of C-to-U (blue) and U-to-C (red) RNA editing for the respective genes. Creations of start or stop codons by C-to-U editing are indicated by symbols '>' and '*' and the removal of stop codons by U-to-C editing is indicated by the exclamation marks, respectively. **B**. The cpDNA of *Haplopteris elongata* (accession MH173086) features two "morffo" elements ("mobile ORFs in fern organelles") in the IR region between *rrn5* and the 3'-part of the *trans*-splicing *rps12* gene. Recognizable sequence homologues of morffo2 can presently only be identified in *Cyclosorus interruptus* (accession MN599066, Thelypteridaceae) and *Histiopteris incisa* (accession MH319942, Dennstaedtiaceae) and a homologue of morffo1 (orange) can presently only be found in the distant fern *Hymenophyllum holochilum* (accession MH265124, Hymenophyllales). Only the upstream part of morffo2 (378 bp) is present in the *H. ensiformis* plastome, while a cpDNA insert in its mitogenome contains an extended region of 628 bp.

Figure 2. The *accD* gene example for chloroplast RNA editing in *Haplopteris ensiformis*.

A. Sequence alignment of the *H. ensiformis accD* gene below its homologue in *Psilotum nudum* shown as one selected example reference out of 22 used for prediction of RNA editing. Alignment was created by PREPACT [88] with identical nucleotides and amino acid shown in grey font and predicted C-to-U RNA editing in blue and reverse U-to-C editing in red. Codons framed by boxes were confirmed as editing sites. The stippled rectangle highlights potential reverse edit accDeC580FL remaining unconfirmed but strongly suggested by the chloroplast editome references in PREPACT 3.0 with the exception of *Ginkgo biloba*. The remaining cases are weak predictions only that are not supported by the majority of other editome references. **B**. The list of expected and observed *accD* edits including those in the 5' and 3'-UTRs and the respective editing frequencies observed with additional remarks.

Figure 3. The *Haplopteris ensiformis* mtDNA: contigs A-C and chromosome 5

Haplopteris ensiformis mtDNA contigs A (hatched), B (grey) and C (dotted) can be connected into a circular chromosome of 217,669 bp. Recombination breakpoints are numbered for each contig and preceded by a small 'r' with radial lines in the circular map indicating transitions into other contigs, allowing for numerous alternative mtDNA arrangements. Eight further circular chromosomes, as listed on top, connect sequences of chromosome V with nine further mtDNA contigs D-L as shown in figure

4. Contig gene maps were created using the SnapGene Viewer software. Native mitochondrial protein coding and tRNA gene sequence are given in lighter and darker blue, respectively, and introns are indicated with black arrows. Numbers next to genes indicate C-to-U (blue) and U-to-C RNA edits (red) with additional symbols indicating removal (!) or creation of stop (*) or start (>) codons, respectively. Introns are indicated with stippled lines and additional black arrows and their standardized labels. Ribosomal RNA genes (here on contig A) are shown in red with the *rrnL* gene featuring a complex gene structure requiring *trans*-splicing via the disrupted group I intron rrnLi825g1 and PSX labels indicate pseudogene fragments shown in grey. The peculiar case of the *trans*-spliced group I intron rrnLi825g1 is highlighted in yellow. Genes for tRNAs of chloroplast or bacterial origin are indicated in green or purple, respectively. For clarity, no other chloroplast or bacterial DNA insertions are shown here. The latter are listed together with annotated features in supplementary table 3.

Figure 4. The *Haplopteris ensiformis* mtDNA: contigs D-L and chromosomes 1-4 and 6-9.

The *Haplopteris ensiformis* mtDNA contigs D-L are linked to recombination points in chromosome 5 (Fig. 3) or within themselves, creating further and/or alternative mtDNA arrangements. Recombination endpoints are labeled and the display of contigs and labels for genes and RNA editing events is as in figure 3. Possible, circular chromosomal structures chr1 to chr4 and chr6 to chr9 are shown.

Figure 5. Repeats and recombination in the *Haplopteris ensiformis* mitogenome.

Recombinations across repeats (in orange) R596 (A), repeats R203 and R 513 (B) and R180 (C) was investigated by tsa-PCR ("template-switch-avoiding") strategies. **A**. R596 is identically present in domains IV of group II introns nad5i1242g2 and rrnLi833g2. Average read coverage of the flanking single copy regions were ca. 150 x for *nad5* (arrangement A-C) and ca. 350 x for *rrnL* (arrangement B-D), apparently adding up to ca. 500 x for R596. PCR products are obtained for the expected gene continuities (AC, BD) with only minor evidence for reciprocal exchanges (AD, BC). **B**. All combinations of flanking sequences (AC-BD) are identified for recombination across R203 whereas a clear bias is seen for recombination across R513 where a product for primer combination B-C remains undetected. **C**. One copy of repeat R180 is located in intron nad2i709g2, another one downstream of the *nad9* gene. The region between R180 and *nad9* contains additional repeats R137 and R259 and all consecutive recombination products can be found whereas there is only very weak evidence for recombination across the R180 copy located in nad2i709g2.

Figure 6. Mitochondrial RNA editing in *Haplopteris ensiformis:* the *rpl6-rps13-rps11* case.

The alignment exemplarily shows RNA editing heavily affecting the *rpl6-rps13-rps11* co-transcript including the removal of seven stop codons within the first 20 codons of the *rps11* reading frame. An internal PCR amplicon covers 46 editing sites from rpl6eC106*R to rps11eC136*Q. Synthesis of cDNA

was primed either with random hexamers (n6) or with specific primers covering the end of *rps11* including the stop codon generation and two edits in the 3'-UTR in an edited (P+) or unedited (P-) version. The alignment displays edits (C-to-U in blue and U-to-C in red) clearly revealed in the RT-PCRs primed by the three different approaches and the results from the RNA-Seq data (RRM). Silent codon edits shown below the protein alignment were exclusively identified in the latter.

Figure 7. A *trans*-splicing group I intron in the *Haplopteris ensiformis* mitochondrial *rrnL* gene.

A. Maturation of *rrnL* includes *trans*-splicing of the disrupted group I intron rrnLi825g1. Three additional introns are removed from the downstream part of *rrnL*. Group II intron rrnLi833g2 has a homologue conserved in the mtDNAs of liverworts and the downstream group I introns rrnLi1897g1 and rrnLi1928g1 are conserved in other Polypodiales species. Splice sites of rrnLi825g1 and rrnLii833g2 (black and grey triangles, respectively) frame the tiny second rrnL exon of only eight nucleotides. **B**. Secondary structure model of the disrupted, *trans*-splicing group I intron rrnLi825g1. The figure was generated using the VARNA software [89]. Regions for base-pairing between the upstream and downstream parts of rrnLi825g1 are found in the group I intron secondary structure P9.1. Intron rrnLi825g1 intron has a positional orthologue in the sweet-water alga *Chara vulgaris.* Green and yellow shading indicates identical nucleotides and transitions, respectively.

Figure 8. A multitude of chloroplast DNA inserts in the *Haplopteris ensiformis* mitogenome.

Selected examples for altogether approximately 80 inserts of chloroplast DNA populating the *Haplomitrium ensiformis* mitogenome (see supplementary table 3). Maximum likelihood trees were conducted with IQ-TREE [90] after automatic model selection of TIM+F+I+G4 or GTR+F+I+G4 and trees were rooted with the Lindsaeaceae family or the Eupolypod II clade, respectively, for cp1039 and cp2126. Bootstrap support is derived from 500 replicates. **A**. The intergenic region between *nad5* and *sdh4* contains the largest collection of likely independently acquired cpDNA inserts including the largest individual insert cp4165 with 93% similarity to the native chloroplast *ndhH-ndhE* region. The other inserts share variable sequence identities with the native *H. ensiformis* cpDNA ranging from 73% for cp686 to 99% for cp1271. Inserts cp364 and cp749 lack evident homologies in the *H. ensiformis* plastome, but are identified by sequence similarities to cpDNAs in other fern genera like *Asplenium* or *Vittaria*, highlighted in red. **B**. Chloroplast DNA insert cp1039 derived from the chloroplast *atpA-atpF* region is an example for a likely very recently acquired insert as evident from its well-supported sister placement to the newly assembled *H. ensiformis* cpDNA (Fig. 1) in a phylogeny including the homologous plastome regions from diverse polypod ferns. **C**. Chloroplast insert cp2128 embedded in other cp inserts of variable sequence conservation carries a unique morffo element, identified by sequence similarity only in the cpDNA of *Hemionitis subcordata*.

Figure 9. Rickettsial-like bacterial inserts in the *Haplopteris ensiformis* mitogenome.

Numerous inserts of bacterial, mostly rickettsial, protein-coding sequences were found to be integrated into the *Haplopteris ensiformis* mitogenome. Xenologous bacterial inserts were annotated to indicate their sizes in base pairs, preceded by 'x' (see supplementary table 3), here showing examples for x625 (A), x888 and x1623 (B) and x1920 (c). Average read coverages are shown on top and PCR amplicons used to verify the mitogenome assemblies are indicated, with PCR results exemplarily shown for x625. **A**. Bacterial insert x625 is located next to cpDNA insert cp416 between *cox2* and the downstream part of *rrnL*. Two overlapping PCRs confirm the mitogenome assembly with linkages into both genes for two independent samples from distantly grown *Haplopteris ensiformis* plant isolations but failed to find products for a *Vittaria lineata* sample growing near *H. ensiformis* isolate 1. PCR products of expected sizes (framed with stippled boxes) were cut out and sequenced and confirmed sequence identity with the mitogenome assembly. The graph on top shows a continuity for the average read coverages of ca. 170 x for *cox2*, cp416 and x625 with an increase to ca. 440 x for the downstream *rrnL* region. Insert x625 carries the central region for DNA recombination protein RmuC with the upstream part located on insert x1623. **B**. Xenologous inserts x888 and x1623 are located between *nad1* and cpDNA insert cp232. Coding regions for HscA, an FeS-protein assembly chaperone and for RmuC borne on x1623 (see supplementary table 3 and supplementary figure S3) are truncated, but highly conserved in primary sequence as exemplarily shown in the phylogenetic analysis for RmuC documenting a close association with *Caedimonas varicaedens*. Severely degenerated coding sequences for OMBB, an outer membrane beta-barrel domain containing protein borne on x888 are highly degenerated and do not allow a clear affiliation with a specific Rickettsiales bacterium. **C**. Xenologous insert x1920 represents a continuous stretch of Rickettsia DNA with top similarities to four coding sequences in *Caedimonas* spp. Coding sequences (see supplementary figure S3) are full-length for ribosome maturation factor RimM and the tRNA guanine-methyltransferase TrmD but amino-terminally truncated for MfdD, a transcription-repair-coupling factor and carboxy-terminally truncated for RpL19 encoding protein 19 of the large ribosomal subunit, respectively.

Supplementary Data

Supplementary Table 1. Chloroplast RNA editing in *Haplopteris ensiformis*.

The table lists RNA editing sites in chloroplast transcripts of *Haplopteris ensiformis.* The first column indicates positions in the newly assembled cpDNA (accession OM867544), the third column indicates labels for editing sites following a previous nomenclature proposal (Lenz et al., 2010) and the fourth column indicates edits other than those changing codon identities. The second column lists the distance to the respective following editing site with immediately neighboring sites highlighted by

green shading and those only two nucleotides apart with blue shading. Editing site labels are now suggested to be amended for multiple edits affecting single codons (highlighted with yellow shading, see supplementary figure 1). Altogether 443 sites of RNA editing were detected in the *Haplopteris ensiformis* chloroplast transcriptome applying strict criteria for detection. The table also includes cpDNA positions for which RNA editing would have been strongly expected from predictions but for which we could not identify a base conversion. A classification of edits according to genes and quality of codon changes is given under B and C, respectively.

Supplementary Table 2. Mitochondrial RNA editing in *Haplopteris ensiformis*.

Altogether 1618 sites of RNA editing (1091 sites of C-to-U and 527 of U-to-C editing) were detected in the *Haplopteris ensiformis* mitochondrial transcriptome. Makeup of listing is as in supplementary table 1. Owing to a co-existing pseudogene copy, RNA editing in *atp8* was analyzed by RT-PCR (indicated with "cDNA" in column efficiency).

Supplementary Table 3. Annotated features in the *Haplopteris ensiformis* mitogenome chromosomes.

List of annotated features (genes, repeat regions, recombination points, promiscuous chloroplast DNA insertions 'cp' and bacterial sequence insertions 'x') for mtDNA contigs A through L variably integrated into mitogenome chromosomes 1 through 9 as displayed in figures 3 and 4, respectively.

Supplementary Figure S1. Nomenclature extension for multiple edits affecting single codons.

The pipe symbol (|) is added for editing site labels when multiple non-silent C-to-U or U-to-C edits affecting single codons. The respective codon change considering the individual edit alone is given, as usual, at the end. The ultimate codon change outcome when taking also the neighboring non-silent change is additionally indicated before the pipe symbol. **A**. All possible ways of converting YYN (Pro, Leu, Ser and Phe) codons are shown with individual changes in the first or second codon position in the top or bottom lines, respectively. C-to-U editing is shown in blue and U-to-C editing in red. First position edits of CUR or UUR Leucine codons are silent when considered individually but are factually not when accompanied by editing in second codon position (LS|LL and LP|LL, blue shading). We name these events "primary silents". *Vice versa*, apparent non-silent 1st position edits of CCR proline or UCR serine codons (PL|PS and SL|SP, green shading) may ultimately appear silent when the 2nd position is edited, too. We name these events "secondary silents". For silent edits in codons also affected by non-silent edits, the codon identities before and after the non-silent edit(s) are shown before and after the underline. **B**. The example shows the exceptional case of five RNA editing events in a row causing Leu-to-Phe exchanges in two successive *nad1* codons. The respective RNA editing efficiencies are indicated. Silent sites are mostly edited with low efficiencies, as here exemplarily seen for nad1eU702LL_SS

edited to only 15%. Notable exceptions are found for efficiently edited silent sites downstream of a Thr-to-Ile codon conversion in *nad4* (**C**) or for three closely spaced silent edits in the amino-terminal part of the *nad5* coding sequence (**D**).

Supplementary Figure S2. Prediction of mitochondrial RNA editing.

RNA editing sites were predicted using the "commons" function of PREPACT 3.0[88] with the selection of reference editomes shown on top (the alga *Chara vulgaris*, the liverwort *Marchantia paleacea*, the moss *Physcomitrella patens*, the lycophytes *Isoetes engelmannii* and *Selaginella moellendorffii*, the eusporangiate ferns *Ophioglossum californicum* and *Psilotum nudum* and the angiosperms *Cocos nucifera* and *Liriodendron tulipifera*). Black font indicates prediction from genomically encoded conserved codons, red font indicates known editing events in the respective reference and single letters indicate a deviating amino acid in a given reference editome. Examples are shown for the *atp9* gene (A) and the *atp6* gene (B).

Supplementary figure S3. Bacterial protein sequence similarities in the *Haplopteris ensiformis* mitogenome.

Bacterial protein sequence similarities identified in the *Haplopteris ensiformis* mitogenome. Shown are the top similar alignments indicating length of the native protein entry, the numbers of identical and similar residues and total alignment length separated by slashes followed by percentages of identical and similar residues, respectively.

Supplementary figure S4. Bacterial nucleotide sequence similarities in the *Haplopteris ensiformis* mitogenome.

Alignments exemplarily displaying bacterial nucleotide sequence similarities in the *Haplopteris ensiformis* mitogenome.

References

1. Gualberto, J. M. & Newton, K. J. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu Rev Plant Biol* **68**, 225–252 (2017).
2. Arrieta-Montiel, M. P. & Mackenzie, S. A. Plant mitochondrial genomes and recombination. in (ed. Kempken, F.) 65–82 (Springer, 2011).
3. Cole, L. W., Guo, W., Mower, J. P. & Palmer, J. D. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. *Mol Biol Evol* **35**, 2773–2785 (2018).
4. Chevigny, N., Schatz-Daas, D., Lotfi, F. & Gualberto, J. M. DNA Repair and the Stability of the Plant Mitochondrial Genome. *Int J Mol Sci* **21**, 328 (2020).

5. Christensen, A. C. Plant Mitochondria are a Riddle Wrapped in a Mystery Inside an Enigma. *Journal of Molecular Evolution* vol. 89 151–156 Preprint at https://doi.org/10.1007/s00239-020-09980-y (2021).

6. Liu, Y. *et al.* The Mitochondrial Genome of the Lycophyte *Huperzia squarrosa*: The Most Archaic Form in Vascular Plants. *PLoS One* **7**, e35168 (2012).

7. Grewe, F., Viehoever, P., Weisshaar, B. & Knoop, V. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. *Nucleic Acids Res* **37**, 5093–5104 (2009).

8. Hecht, J., Grewe, F. & Knoop, V. 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 (2011).

9. Chaw, S.-M. S. M. *et al.* 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 (2008).

10. Guo, W. *et al. Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol* **33**, 1448–1460 (2016).

11. Kan, S.-L., Shen, T.-T., Gong, P., Ran, J.-H. & Wang, X.-Q. The complete mitochondrial genome of Taxus cuspidata (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. *BMC Evol Biol* **20**, 10 (2020).

12. Sullivan, A. R. *et al.* The Mitogenome of Norway Spruce and a Reappraisal of Mitochondrial Recombination in Plants. *Genome Biol Evol* **12**, 3586–3598 (2020).

13. Jackman, S. D. *et al.* Organellar Genomes of White Spruce (Picea glauca): Assembly and Annotation. *Genome Biol Evol* **8**, 29–41 (2015).

14. Putintseva, Y. A. *et al.* Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics* **21**, 654 (2020).

15. Richardson, A. O., Rice, D. W., Young, G. J., Alverson, A. J. & Palmer, J. D. The 'fossilized' mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol* **11**, 29 (2013).

16. Alverson, A. J. *et al.* Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). *Mol Biol Evol* **27**, 1436–1448 (2010).

17. Rodríguez-Moreno, L. *et al.* Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* **12**, 424 (2011).

18. Sloan, D. B. *et al.* Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biol* **10**, e1001241 (2012).

19. Rice, D. W. *et al.* Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**, 1468–1473 (2013).

20. Guo, W., Zhu, A., Fan, W. & Mower, J. P. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytologist* **213**, 391–403 (2017).

21. Knie, N., Grewe, F. & Knoop, V. Monilophyte mitochondrial rps1 genes carry a unique group II intron that likely originated from an ancient paralog in rpl2. *RNA* **22**, 1338–1348 (2016).

22. Zumkeller, S. M., Knoop, V. & Knie, N. Convergent Evolution of Fern-Specific Mitochondrial Group II Intron atp1i361g2 and Its Ancient Source Paralogue rps3i249g2 and Independent Losses of Intron and RNA Editing among Pteridaceae. *Genome Biol Evol* **8**, 2505–19 (2016).

23. Knie, N., Grewe, F., Fischer, S. & Knoop, V. Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns – a monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. *BMC Evol Biol* **16**, 134 (2016).

24. Guo, W., Grewe, F. & Mower, J. P. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. *PLoS One* **10**, e0117075 (2015).

25. Robison, T. A. *et al.* Mobile Elements Shape Plastome Evolution in Ferns. *Genome Biol Evol* **10**, 2558–2571 (2018).

26. Lenz, H. *et al.* Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Curr Genet* **56**, 189–201 (2010).

27. Song, M. *et al.* A novel chloroplast gene reported for flagellate plants. *Am J Bot* **105**, 117–121 (2018).

28. Gerke, P. *et al.* Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort Anthoceros agrestis. *New Phytologist* **225**, 1974–1992 (2020).

29. Knie, N., Polsakiewicz, M. & Knoop, V. Horizontal Gene Transfer of Chlamydial-Like tRNA Genes into Early Vascular Plant Mitochondria. *Mol Biol Evol* **32**, 629–34 (2014).

30. Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N. & Knoop, V. Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte Selaginella uncinata. *RNA* **20**, 1499–1506 (2014).

31. Bégu, D., Castandet, B. & Araya, A. RNA editing restores critical domains of a group I intron in fern mitochondria. *Curr Genet* **57**, 317–325 (2011).

32. Bégu, D. & Araya, A. The horsetail Equisetum arvense</i> mitochondria share two group I introns with the liverwort Marchantia, acquired a novel group II intron but lost intron-encoded ORFs. *Curr Genet* **55**, 69–79 (2009).

33. Turmel, M., Otis, C. & Lemieux, C. The mitochondrial genome of Chara vulgaris: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell* **15**, 1888–1903 (2003).

34. Suzuki, H. *et al.* Draft genome sequence of *Caedibacter varicaedens*, a kappa killer endosymbiont bacterium of the ciliate *Paramecium biaurelia*. *Genome Announc* **3**, 23–25 (2015).

35. Schuettpelz, E., Schneider, H., Huiet, L., Windham, M. D. & Pryer, K. M. A molecular phylogeny of the fern family Pteridaceae: assessing overall relationships and the affinities of previously unsampled genera. *Mol Phylogenet Evol* **44**, 1172–1185 (2007).

36. Schuettpelz, E. & Pryer, K. M. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon* **56**, 1037–1050 (2007).

37. Schuettpelz, E. & Pryer, K. M. Fern phylogeny. in *Biology and Evolution of Ferns and Lycophytes* 395–416 (2008).

38. Grusz, A. L., Rothfels, C. J. & Schuettpelz, E. Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. *BMC Genomics* **17**, 692 (2016).

39. Rothfels, C. J. & Schuettpelz, E. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. *Syst Biol* **63**, 31–54 (2014).

40. Fauskee, B. D., Sigel, E. M., Pryer, K. M. & Grusz, A. L. Variation in frequency of plastid RNA editing within Adiantum implies rapid evolution in fern plastomes. *Am J Bot* **108**, 820–827 (2021).

41. Roulet, M. E. *et al.* Multichromosomal structure and foreign tracts in the Ombrophytum subterraneum (Balanophoraceae) mitochondrial genome. *Plant Mol Biol* **103**, 623–638 (2020).

42. Kozik, A. *et al.* The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet* **15**, 1–30 (2019).

43. Sloan, D. B. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytologist* **200**, 978–985 (2013).

44. Knoop, V. Plant mitochondrial genome peculiarities evolving in the earliest vascular plant lineages. *J Syst Evol* **51**, 1–12 (2013).

45. Mower, J. P., Sloan, D. B. & Alverson, A. J. Plant Mitochondrial Genome Diversity: The Genomics Revolution. in *Plant Genome Diversity Volume 1* vol. 1 123–144 (Springer Vienna, 2012).

46. Sessa, E. B. *et al.* Between two fern genomes. *Gigascience* **3**, 15 (2014).

47. Li, F.-W. *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature* **4**, 460–472 (2018).

48. Huang, X. *et al.* The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants* **8**, 500–512 (2022).

49. Fang, Y. *et al.* The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nature Plants 2022* 1–14 (2022) doi:10.1038/s41477-022-01222-x.

50. Marchant, D. B. *et al.* Dynamic genome evolution in a model fern. doi:10.1038/s41477-022-01226-7.

51. Wickell, D. *et al.* Underwater CAM photosynthesis elucidated by Isoetes genome. *Nat Commun* **12**, 1–13 (2021).

52. Mower, J. P. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion* **53**, 203–213 (2020).

53. Guo, W., Zhu, A., Fan, W., Adams, R. P. & Mower, J. P. Extensive shifts from cis to trans splicing of gymnosperm mitochondrial introns. *Mol Biol Evol* **37**, 1615–1620 (2020).

54. Malek, O., Brennicke, A. & Knoop, V. Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proc Natl Acad Sci U S A* **94**, 553–558 (1997).

55. Malek, O. & Knoop, V. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA* **4**, 1599–609 (1998).

56. Burger, G., Yan, Y., Javadi, P. & Lang, B. F. Group I intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends in Genetics* **25**, 381–386 (2009).

57. Pombert, J. F. & Keeling, P. J. The mitochondrial genome of the entomoparasitic green alga Helicosporidium. *PLoS.One.* **5**, e8954 (2010).

58. Pelin, A. *et al.* The mitochondrial genome of the arbuscular mycorrhizal fungus Gigaspora margarita reveals two unsuspected trans -splicing events of group I introns. *New Phytologist* **194**, 836–845 (2012).

59. Qiu, Y.-L. *et al.* The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A* **103**, 15511–15516 (2006).

60. Cho, Y., Qiu, Y. L., Kuhlman, P. & Palmer, J. D. Explosive invasion of plant mitochondria by a group I intron. *Proc Natl Acad Sci U S A* **95**, 14244–14249 (1998).

61. Cusimano, N., Zhang, L. B. L.-B. & Renner, S. S. Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol Biol Evol* **25**, 265–276 (2008).

62. Stern, D. B. & Lonsdale, D. M. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* **299**, 698–702 (1982).

63. Wickell, D. A. & Li, F. On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist* **225**, 113–117 (2020).

64. Gao, J.-G. Tracking the evolutionary innovations of plant terrestrialization. *Gene* 145203 (2020) doi:10.1016/j.gene.2020.145203.

65. Mower, J. P., Stefanović, S., Young, G. J. & Palmer, J. D. Plant genetics: gene transfer from parasitic to host plants. *Nature* **432**, 165–166 (2004).

66. Davis, C. C., Anderson, W. R. & Wurdack, K. J. Gene transfer from a parasitic flowering plant to a fern. *Proceedings of the Royal Society B: Biological Sciences* **272**, 2237–2242 (2005).

67. Bergthorsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**, 197–201 (2003).

68. Won, H. & Renner, S. S. Horizontal gene transfer from flowering plants to Gnetum. *Proc Natl Acad Sci U S A* **100**, 10824–10829 (2003).

69. Mower, J. P., Jain, K. & Hepburn, N. J. The Role of Horizontal Transfer in Shaping the Plant Mitochondrial Genome. *Adv Bot Res* **63**, 41–69 (2012).

70. Taylor, Z. N., Rice, D. W. & Palmer, J. D. The Complete Moss Mitochondrial Genome in the Angiosperm Amborella Is a Chimera Derived from Two Moss Whole-Genome Transfers. *PLoS One* **10**, e0137532 (2015).

71. Sinn, B. T. & Barrett, C. F. Ancient Mitochondrial Gene Transfer between Fungi and the Orchids. *Mol Biol Evol* **37**, 44–57 (2020).

72. Schrallhammer, M., Castelli, M. & Petroni, G. Phylogenetic relationships among endosymbiotic R-body producer: Bacteria providing their host the killer trait. *Syst Appl Microbiol* **41**, 213–220 (2018).

73. Grosser, K. *et al.* More than the 'killer Trait': Infection with the Bacterial Endosymbiont Caedibacter taeniospiralis Causes Transcriptomic Modulation in Paramecium Host. *Genome Biol Evol* **10**, 646–656 (2018).

74. Davis, M., Ying, Z., Brunner, B., Pantoja, A. & Ferwerda, F. Rickettsial relative associated with papaya bunchy top disease. *Curr Microbiol* **36**, 80–84 (1998).

75. Yurchenko, T. *et al.* A gene transfer event suggests a long-term partnership between eustigmatophyte algae and a novel lineage of endosymbiotic bacteria. *ISME J* **12**, 2163–2175 (2018).

76. Yang, A., Narechania, A. & Kim, E. Rickettsial endosymbiont in the "early-diverging" streptophyte green alga Mesostigma viride. *J Phycol* **52**, 219–229 (2016).

77. Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiol Rev* **38**, 865–891 (2014).

78. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).

79. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

80. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* **9**, gkw955 (2016).

81. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–52 (2011).

82. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).

83. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).

84. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **44**, (2016).

85. Wu, T. D., Reeder, J., Lawrence, M., Becker, G. & Brauer, M. J. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. in *Statistical Genomics. Methods in Molecular Biology* (eds. Mathé, E. & Davis, S.) vol. 1418 283–334 (Humana Press, 2016).

86. Piechotta, M., Wyler, E., Ohler, U., Landthaler, M. & Dieterich, C. JACUSA: site-specific identification of RNA editing events from replicate sequencing data. *BMC Bioinformatics* **18**, 7 (2017).

87. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* **47**, W59–W64 (2019).

88. Lenz, H., Hein, A. & Knoop, V. Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinformatics* **19**, 255 (2018).

89. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–5 (2009).

90. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

Figures

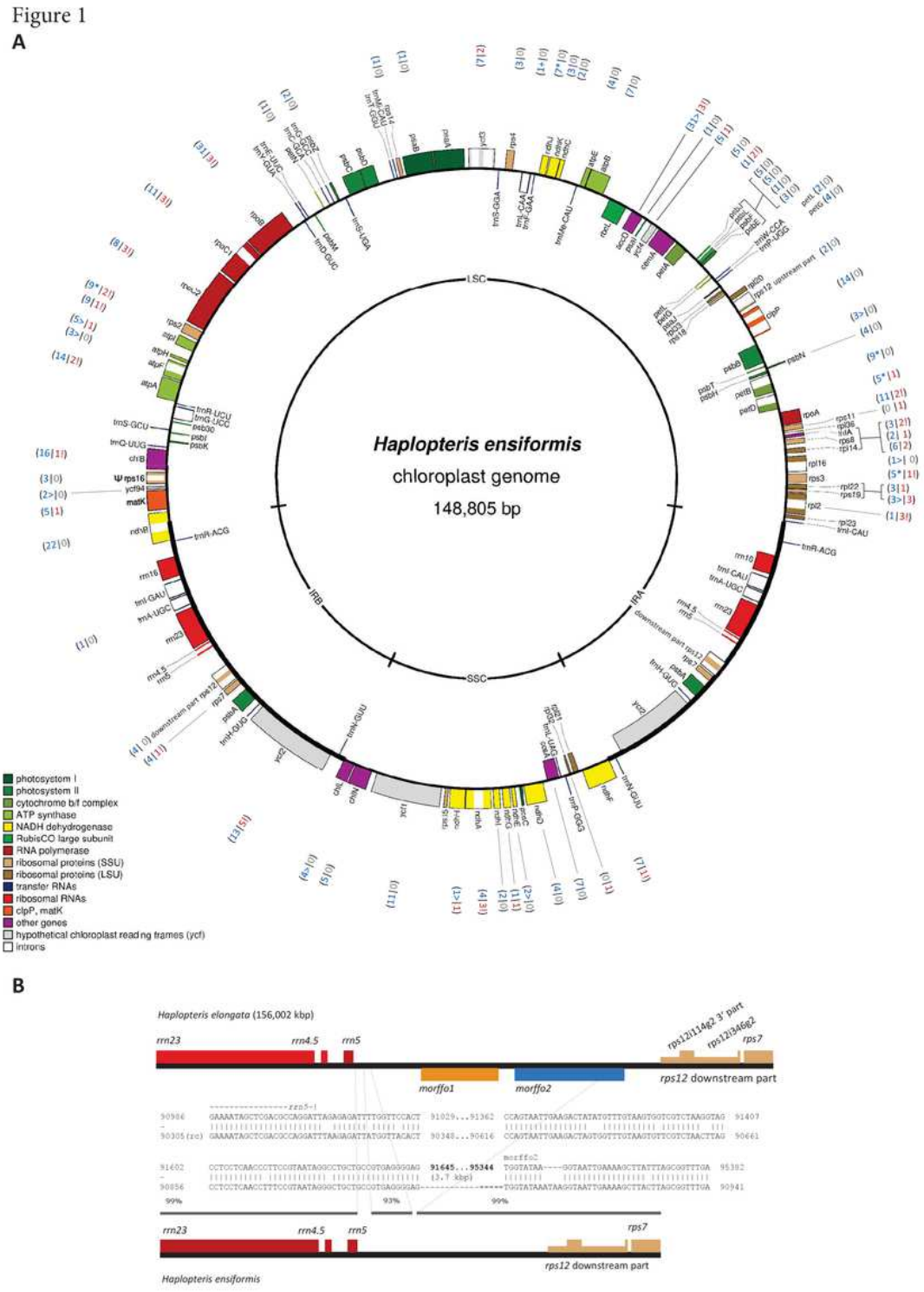


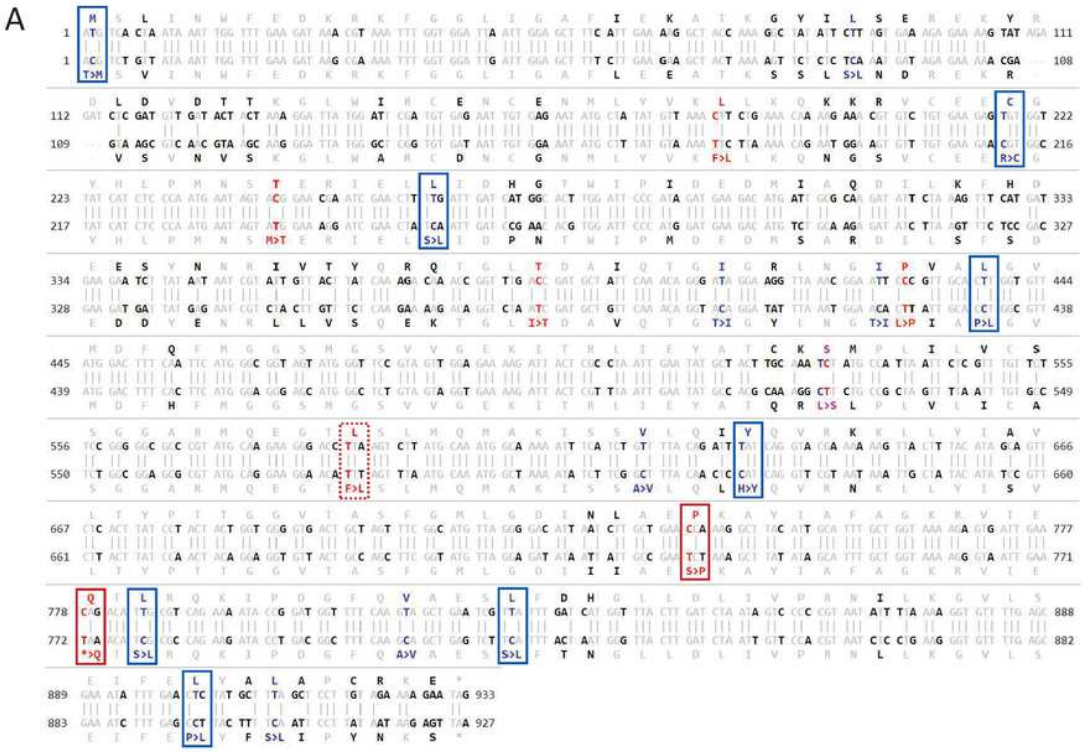
Figure 1

The Haplopteris ensiformis cpDNA.

A. Haplopteris ensiformis reveals a typical plant circular plastome structure consisting of a large (LSC) and a small (SSC) single copy region separated by a pair of inverted repeats (IR) and an expectedly

conserved, ancestral gene and intron complement. The genome map was created using OGDRAW 87. Gene categories are indicated in the legend. Numbers in parentheses indicate the amount of C-to-U (blue) and U-to-C (red) RNA editing for the respective genes. Creations of start or stop codons by C-to-U editing are indicated by symbols '>' and '*' and the removal of stop codons by U-to-C editing is indicated by the exclamation marks, respectively. B. The cpDNA of Haplopteris elongata (accession MH173086) features two "morffo" elements ("mobile ORFs in fern organelles") in the IR region between rrn5 and the 3'-part of the trans-splicing rps12 gene. Recognizable sequence homologues of morffo2 can presently only be identified in Cyclosorus interruptus (accession MN599066, Thelypteridaceae) and Histiopteris incisa (accession MH319942, Dennstaedtiaceae) and a homologue of morffo1 (orange) can presently only be found in the distant fern Hymenophyllum holochilum (accession MH265124, Hymenophyllales). Only the upstream part of morffo2 (378 bp) is present in the H. ensiformis plastome, while a cpDNA insert in its mitogenome contains an extended region of 628 bp.

Figure 2



Edit	Percentage	Comment
accDeU-186	*21.5 %*	*5'-UTR*
accDeU-1	*54.3 %*	*5'-UTR*
accDeU2TM	89.3 %	Start codon creation
accDeU211RC	91.3 %	Also in *Adiantum capillus-veneris*
accDeU257SL	93.5 %	Also in several other taxa (I in seed plants)
accDeU431PL	93.6 %	Also in *A. capillus-veneris* and *Anthoceros angustus*
accDeC580FL	0 %	Strongly predicted by 21 of 22 references
accDeU625HY	6.4%	Variably H or Y elsewhere
accDeU657SS	1.0 %	Silent
accDeC730SP	61.0 %	Also in *A. capillus-veneris* and *Anthoceros angustus*
accDeC772*Q (UAA)	74.2 %	Also in *A. capillus-veneris* and *Anthoceros angustus*
accDeU779SL	94.1 %	Not yet reported
accDeU821SL	91.9 %	Also in *Amborella trichopoda* and *Ophioglossum vulgatum*
accDeU895PL|PS	1.2 %	Likely collateral upstream edit
accDeU896PL|PL	88.6 %	Not yet reported
accDeC+36	*8.5 %*	*3'-UTR*
accDeU+69	*4.9 %*	*3'-UTR*

Figure 2

The accD gene example for chloroplast RNA editing in Haplopteris ensiformis.

A. Sequence alignment of the H. ensiformis accD gene below its homologue in Psilotum nudum shown as one selected example reference out of 22 used for prediction of RNA editing. Alignment was created by PREPACT 88 with identical nucleotides and amino acid shown in grey font and predicted C-to-U RNA editing in blue and reverse U-to-C editing in red. Codons framed by boxes were confirmed as editing sites. The stippled rectangle highlights potential reverse edit accDeC580FL remaining unconfirmed but strongly suggested by the chloroplast editome references in PREPACT 3.0 with the exception of Ginkgo biloba.

The remaining cases are weak predictions only that are not supported by the majority of other editome references. B. The list of expected and observed accD edits including those in the 5’ and 3’-UTRs and the respective editing frequencies observed with additional remarks.

Figure 3

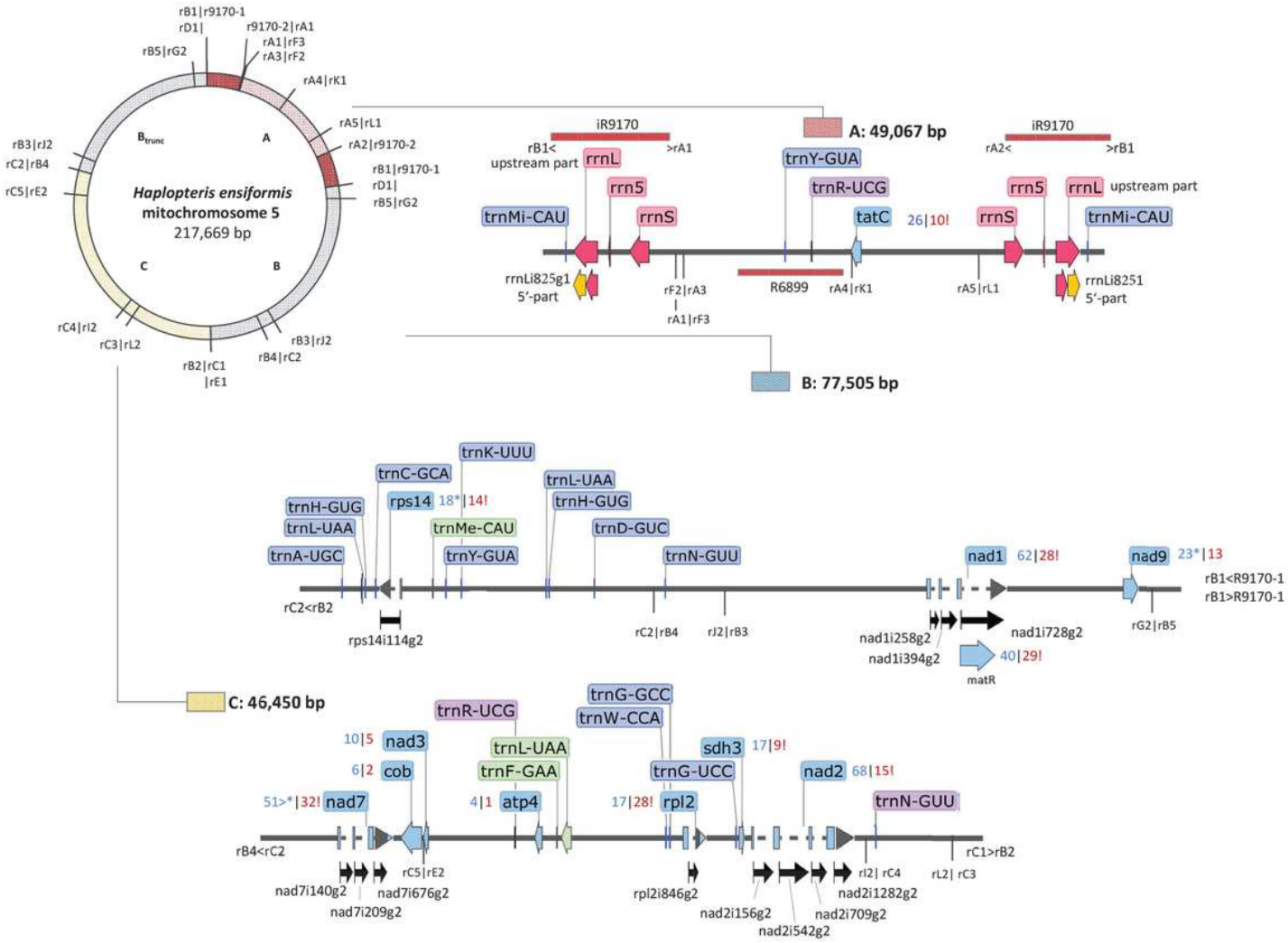


Figure 3

The Haplopteris ensiformis mtDNA: contigs A-C and chromosome 5

Haplopteris ensiformis mtDNA contigs A (hatched), B (grey) and C (dotted) can be connected into a circular chromosome of 217,669 bp. Recombination breakpoints are numbered for each contig and preceded by a small ‘r’ with radial lines in the circular map indicating transitions into other contigs, allowing for numerous alternative mtDNA arrangements. Eight further circular chromosomes, as listed on top, connect sequences of chromosome V with nine further mtDNA contigs D-L as shown in figure 4. Contig gene maps were created using the SnapGene Viewer software. Native mitochondrial protein

coding and tRNA gene sequence are given in lighter and darker blue, respectively, and introns are indicated with black arrows. Numbers next to genes indicate C-to-U (blue) and U-to-C RNA edits (red) with additional symbols indicating removal (!) or creation of stop (*) or start (>) codons, respectively. Introns are indicated with stippled lines and additional black arrows and their standardized labels. Ribosomal RNA genes (here on contig A) are shown in red with the rrnL gene featuring a complex gene structure requiring trans-splicing via the disrupted group I intron rrnLi825g1 and PSX labels indicate pseudogene fragments shown in grey. The peculiar case of the trans-spliced group I intron rrnLi825g1 is highlighted in yellow. Genes for tRNAs of chloroplast or bacterial origin are indicated in green or purple, respectively. For clarity, no other chloroplast or bacterial DNA insertions are shown here. The latter are listed together with annotated features in supplementary table 3.

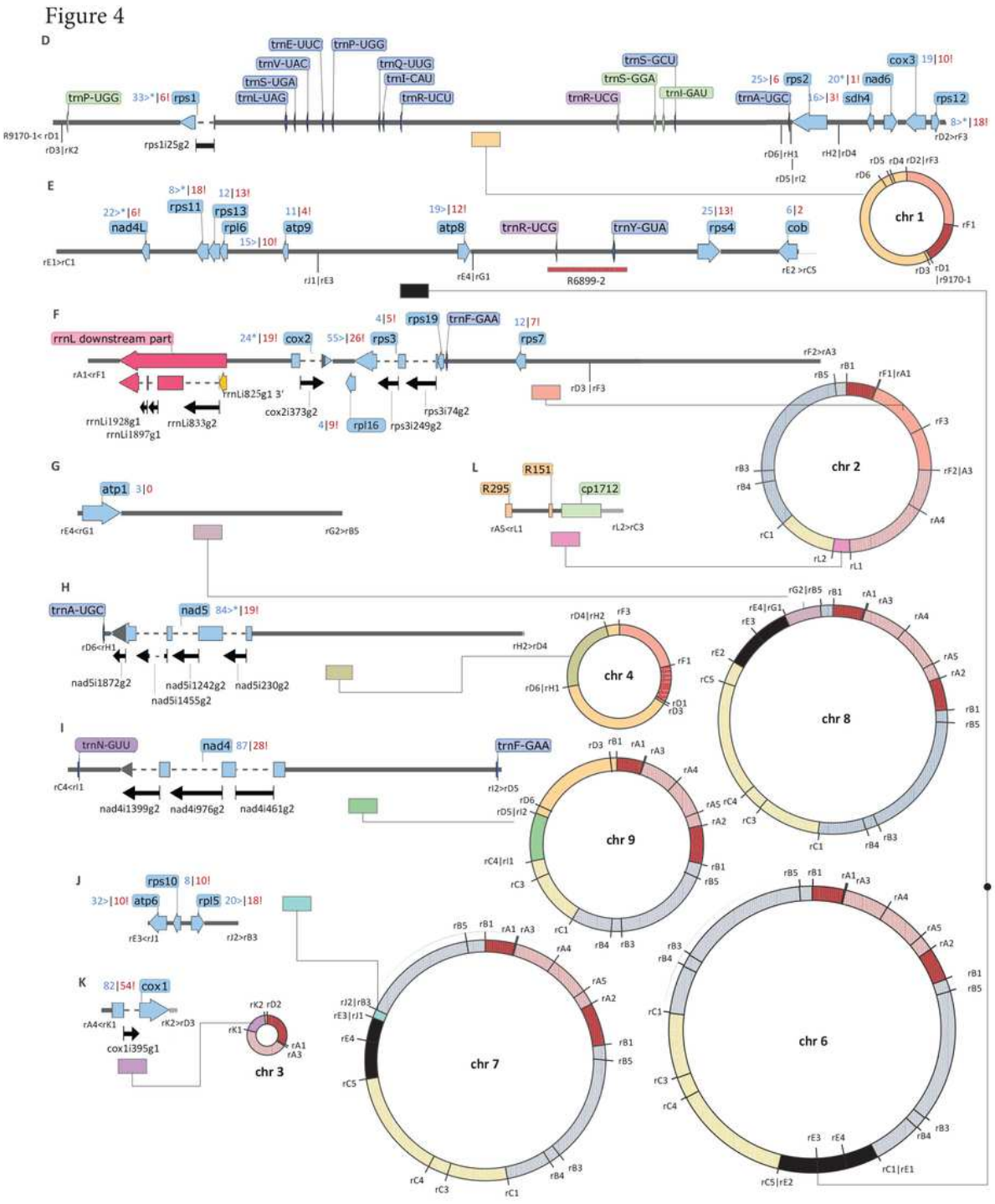


Figure 4

The Haplopteris ensiformis mtDNA: contigs D-L and chromosomes 1-4 and 6-9.

The Haplopteris ensiformis mtDNA contigs D-L are linked to recombination points in chromosome 5 (Fig. 3) or within themselves, creating further and/or alternative mtDNA arrangements. Recombination

endpoints are labeled and the display of contigs and labels for genes and RNA editing events is as in figure 3. Possible, circular chromosomal structures chr1 to chr4 and chr6 to chr9 are shown.

Figure 5

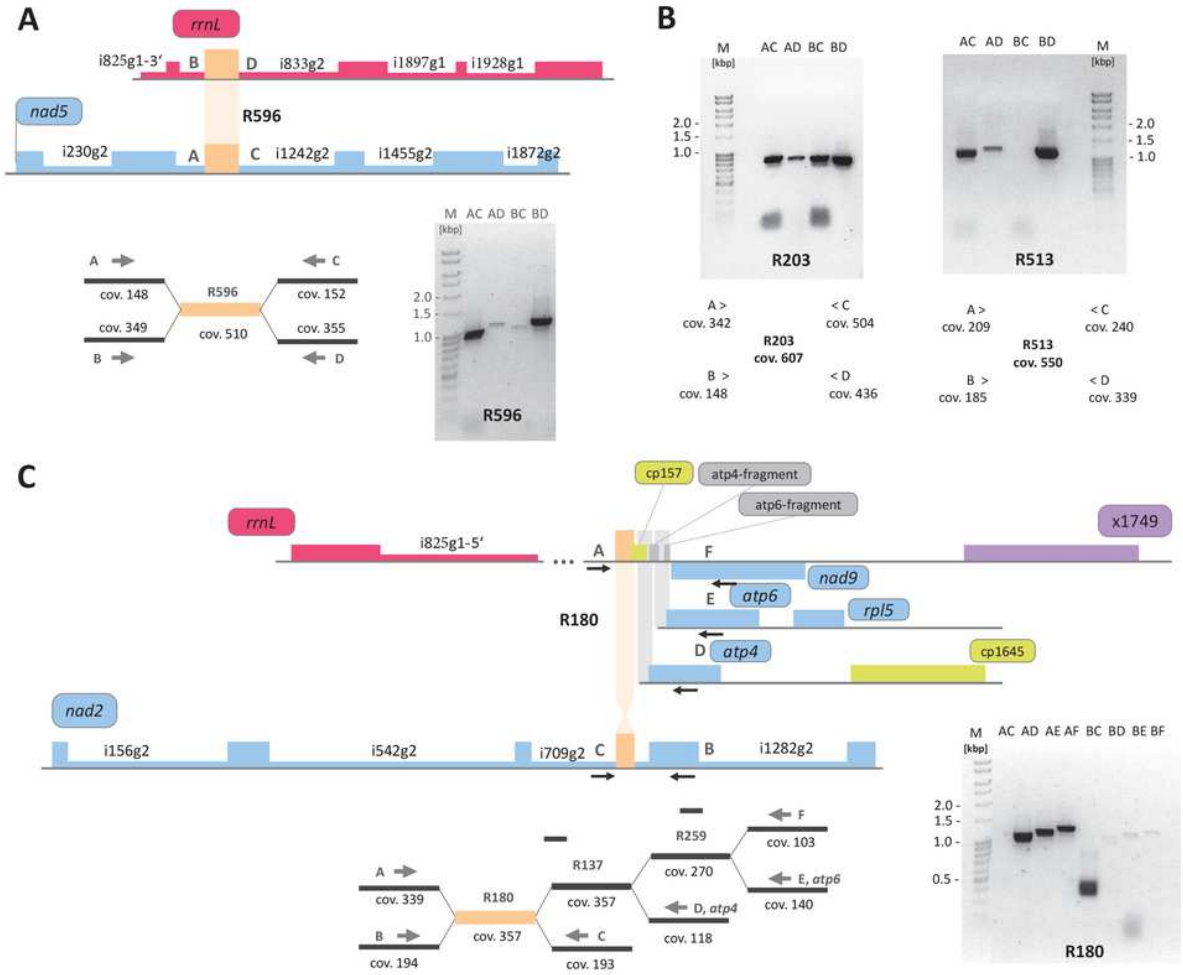


Figure 5

Repeats and recombination in the Haplopteris ensiformis mitogenome.

Recombinations across repeats (in orange) R596 (A), repeats R203 and R 513 (B) and R180 (C) was investigated by tsa-PCR (”template-switch-avoiding”) strategies. A. R596 is identically present in domains

IV of group II introns nad5i1242g2 and rrnLi833g2. Average read coverage of the flanking single copy regions were ca. 150 x for nad5 (arrangement A-C) and ca. 350 x for rrnL (arrangement B-D), apparently adding up to ca. 500 x for R596. PCR products are obtained for the expected gene continuities (AC, BD) with only minor evidence for reciprocal exchanges (AD, BC). B. All combinations of flanking sequences (AC-BD) are identified for recombination across R203 whereas a clear bias is seen for recombination across R513 where a product for primer combination B-C remains undetected. C. One copy of repeat R180 is located in intron nad2i709g2, another one downstream of the nad9 gene. The region between R180 and nad9 contains additional repeats R137 and R259 and all consecutive recombination products can be found whereas there is only very weak evidence for recombination across the R180 copy located in nad2i709g2.

Figure 6

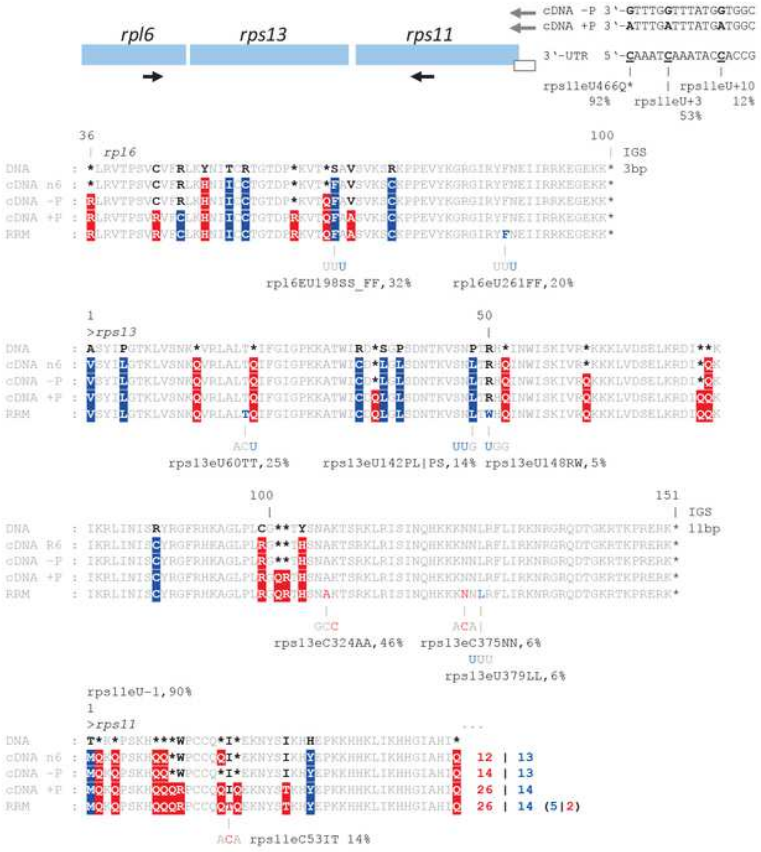


Figure 6

Mitochondrial RNA editing in Haplopteris ensiformis: the rpl6-rps13-rps11 case.

The alignment exemplarily shows RNA editing heavily affecting the rpl6-rps13-rps11 co-transcript including the removal of seven stop codons within the first 20 codons of the rps11 reading frame. An internal PCR amplicon covers 46 editing sites from rpl6eC106*R to rps11eC136*Q. Synthesis of cDNA was primed either with random hexamers (n6) or with specific primers covering the end of rps11 including the stop codon generation and two edits in the 3'-UTR in an edited (P+) or unedited (P-) version. The alignment displays edits (C-to-U in blue and U-to-C in red) clearly revealed in the RT-PCRs primed by the three different approaches and the results from the RNA-Seq data (RRM). Silent codon edits shown below the protein alignment were exclusively identified in the latter.

Figure 7

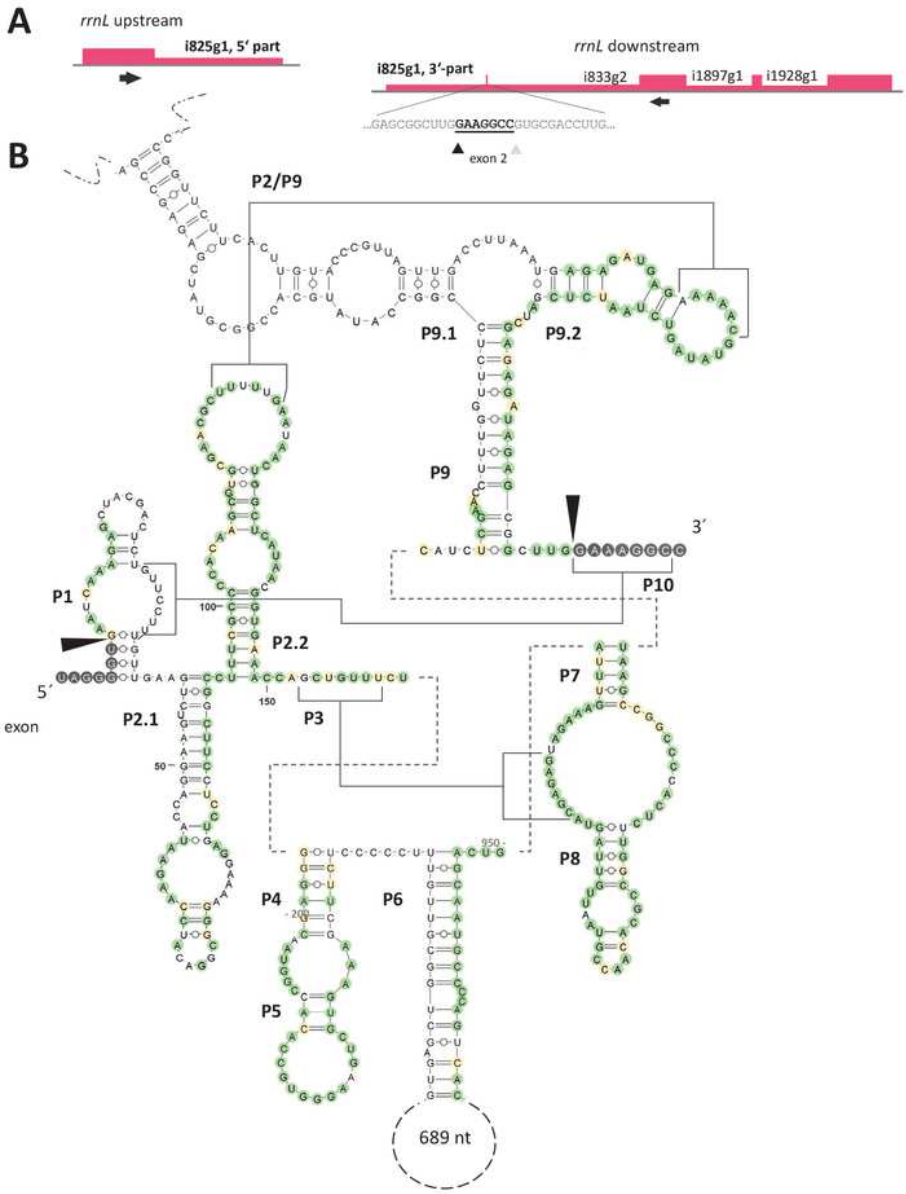


Figure 7

A trans-splicing group I intron in the Haplopteris ensiformis mitochondrial rrnL gene.

A. Maturation of rrnL includes trans-splicing of the disrupted group I intron rrnLi825g1. Three additional introns are removed from the downstream part of rrnL. Group II intron rrnLi833g2 has a homologue conserved in the mtDNAs of liverworts and the downstream group I introns rrnLi1897g1 and rrnLi1928g1

are conserved in other Polypodiales species. Splice sites of rrnLi825g1 and rrnLii833g2 (black and grey triangles, respectively) frame the tiny second rrnL exon of only eight nucleotides. B. Secondary structure model of the disrupted, trans-splicing group I intron rrnLi825g1. The figure was generated using the VARNA software 89. Regions for base-pairing between the upstream and downstream parts of rrnLi825g1 are found in the group I intron secondary structure P9.1. Intron rrnLi825g1 intron has a positional orthologue in the sweet-water alga Chara vulgaris. Green and yellow shading indicates identical nucleotides and transitions, respectively.

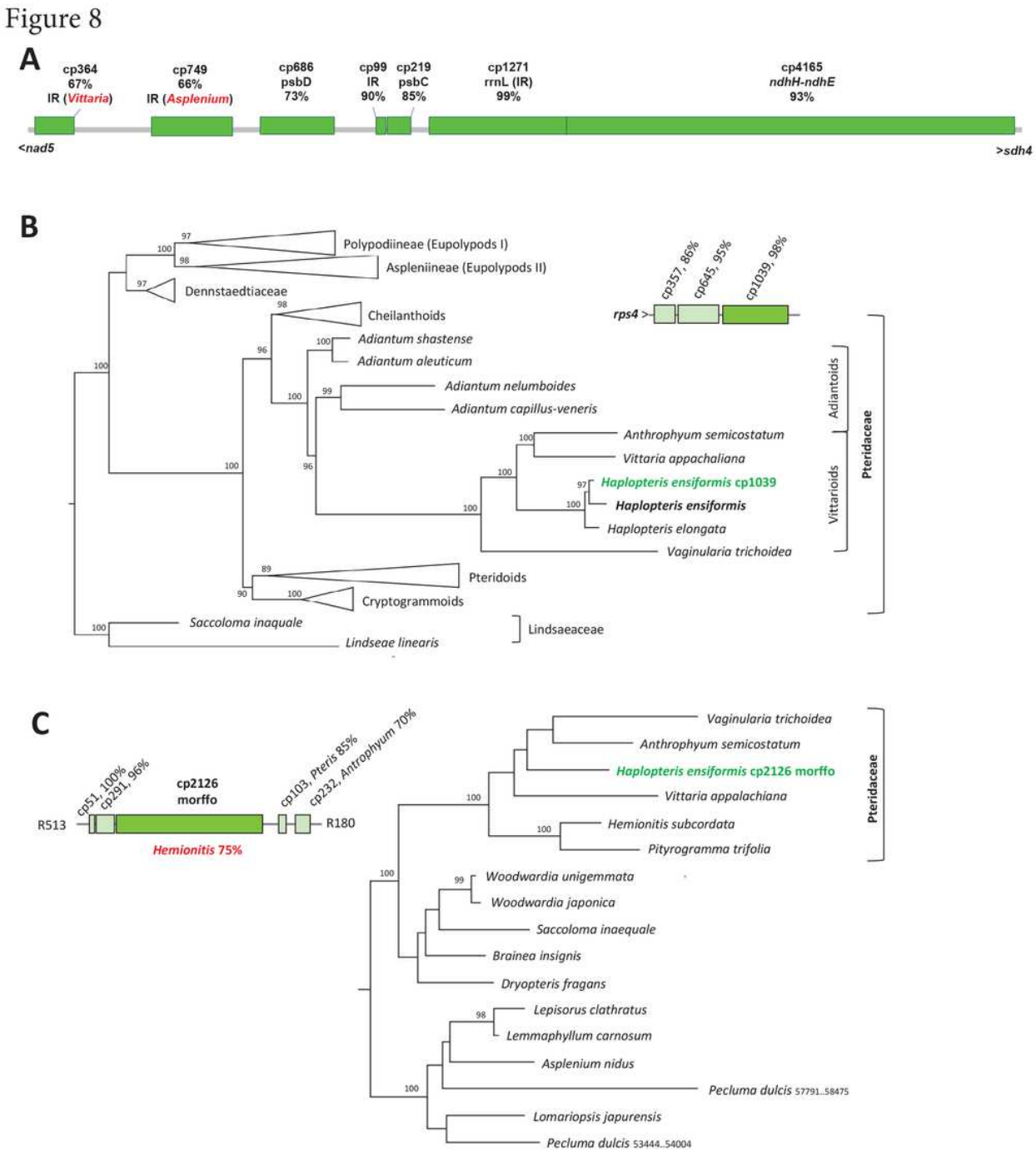


Figure 8

A multitude of chloroplast DNA inserts in the Haplopteris ensiformis mitogenome.

Selected examples for altogether approximately 80 inserts of chloroplast DNA populating the Haplomitrium ensiformis mitogenome (see supplementary table 3). Maximum likelihood trees were conducted with IQ-TREE 90 after automatic model selection of TIM+F+I+G4 or GTR+F+I+G4 and trees were rooted with the Lindsaeaceae family or the Eupolypod II clade, respectively, for cp1039 and cp2126. Bootstrap support is derived from 500 replicates. A. The intergenic region between nad5 and sdh4 contains the largest collection of likely independently acquired cpDNA inserts including the largest individual insert cp4165 with 93% similarity to the native chloroplast ndhH-ndhE region. The other inserts share variable sequence identities with the native H. ensiformis cpDNA ranging from 73% for cp686 to 99% for cp1271. Inserts cp364 and cp749 lack evident homologies in the H. ensiformis plastome, but are identified by sequence similarities to cpDNAs in other fern genera like Asplenium or Vittaria, highlighted in red. B. Chloroplast DNA insert cp1039 derived from the chloroplast atpA-atpF region is an example for a likely very recently acquired insert as evident from its well-supported sister placement to the newly assembled H. ensiformis cpDNA (Fig. 1) in a phylogeny including the homologous plastome regions from diverse polypod ferns. C. Chloroplast insert cp2128 embedded in other cp inserts of variable sequence conservation carries a unique morffo element, identified by sequence similarity only in the cpDNA of Hemionitis subcordata.

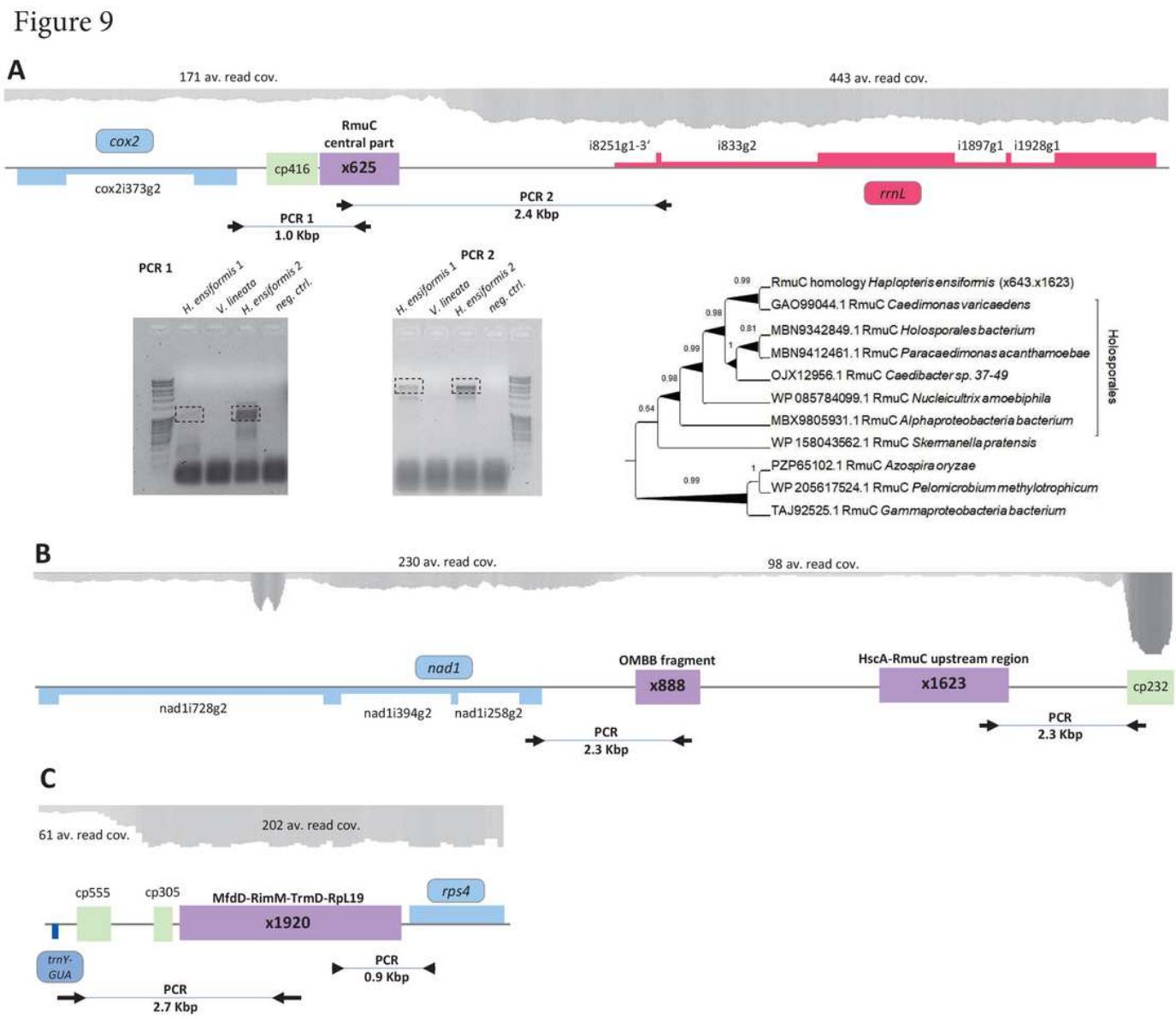


Figure 9

Rickettsial-like bacterial inserts in the Haplopteris ensiformis mitogenome.

Numerous inserts of bacterial, mostly rickettsial, protein-coding sequences were found to be integrated into the Haplopteris ensiformis mitogenome. Xenologous bacterial inserts were annotated to indicate their sizes in base pairs, preceded by 'x' (see supplementary table 3), here showing examples for x625 (A), x888 and x1623 (B) and x1920 (c). Average read coverages are shown on top and PCR amplicons used to verify the mitogenome assemblies are indicated, with PCR results exemplarily shown for x625. A. Bacterial insert x625 is located next to cpDNA insert cp416 between cox2 and the downstream part of rrnL. Two overlapping PCRs confirm the mitogenome assembly with linkages into both genes for two independent samples from distantly grown Haplopteris ensiformis plant isolations but failed to find products for a Vittaria lineata sample growing near H. ensiformis isolate 1. PCR products of expected

sizes (framed with stippled boxes) were cut out and sequenced and confirmed sequence identity with the mitogenome assembly. The graph on top shows a continuity for the average read coverages of ca. 170 x for cox2, cp416 and x625 with an increase to ca. 440 x for the downstream rrnL region. Insert x625 carries the central region for DNA recombination protein RmuC with the upstream part located on insert x1623. B. Xenologous inserts x888 and x1623 are located between nad1 and cpDNA insert cp232. Coding regions for HscA, an FeS-protein assembly chaperone and for RmuC borne on x1623 (see supplementary table 3 and supplementary figure S3) are truncated, but highly conserved in primary sequence as exemplarily shown in the phylogenetic analysis for RmuC documenting a close association with Caedimonas varicaedens. Severely degenerated coding sequences for OMBB, an outer membrane beta-barrel domain containing protein borne on x888 are highly degenerated and do not allow a clear affiliation with a specific Rickettsiales bacterium. C. Xenologous insert x1920 represents a continuous stretch of Rickettsia DNA with top similarities to four coding sequences in Caedimonas spp. Coding sequences (see supplementary figure S3) are full-length for ribosome maturation factor RimM and the tRNA guanine-methyltransferase TrmD but amino-terminally truncated for MfdD, a transcription-repair-coupling factor and carboxy-terminally truncated for RpL19 encoding protein 19 of the large ribosomal subunit, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplFigure1ZumkelleretalHaplopterismtDNANC20220919final2.pdf
- SupplFigure2ZumkelleretalHaplopterismtDNANC20220919final2.pdf
- SupplFigure3ZumkelleretalHaplopterismtDNANC20220919final2.pdf
- SupplFigure4ZumkelleretalHaplopterismtDNANC20220919final2.pdf
- SupplTab1ZumkelleretalCPEditingfinalp1.pdf
- ZumkelleretalSupplTab2MTEditingcomplete.pdf
- SupplTab3Zumkelleretal2022insertsfeaturesSZ20220727.pdf

Publication 3: Categorizing 161 plant mitochondrial group II introns into 29 families of related paralogues finds only limited links between intron mobility and intron-borne maturases

BMC-Ecology and Evolution; 2022 *Under Review*
Available as pre-print on Research Square
https://doi.org/10.21203/rs.3.rs-2086450/v1

Zumkeller S. M. and Knoop V.

IZMB - Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution,
Universität Bonn,
Kirschallee 1, D-53115 Bonn
Corresponding author
Email: volker.knoop@uni-bonn.de

Categorizing 161 plant mitochondrial group II introns into 29 families of related paralogues finds only limited links between intron mobility and intron-borne maturases

Simon Zumkeller
Universität Bonn
Volker Knoop (volker.knoop@uni-bonn.de)
Universität Bonn

Research Article

Keywords: Group II introns, Retrotransposition, Maturases, Group I introns, mtDNA, Mitogenomes, Horizontal Gene Transfer

Posted Date: October 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2086450/v1

Abstract

Group II introns are common in the two endosymbiotic organelle genomes of plants. Chloroplasts harbor 22 positionally conserved group II introns whereas their occurrence in land plant mitogenomes is highly variable and specific for the seven major embryophyte clades: liverworts, mosses, hornworts, lycophytes, ferns, gymnosperms and flowering plants. Each plant group features “signature selections” of ca. 20–30 paralogues from a superset of altogether 105 group II introns meantime identified in embryophyte mtDNAs, suggesting massive intron gains and losses along the backbone of plant phylogeny. We report on systematically categorizing plant mitochondrial group II introns into “families”, comprising evidently related paralogues at different insertion sites, which may even be more similar than their respective orthologues in phylogenetically distant taxa. Including streptophyte algae extends our sampling to 161 and we sort 104 streptophyte mitochondrial group II introns into 25 core families of related paralogues evidently arising from retrotransposition events. Adding to discoveries of only recently created intron paralogues, hypermobile introns and twintrons, our survey led to further discoveries including previously overlooked “fossil” introns in spacer regions or e.g., in the *rps8* pseudogene of lycophytes. Initially excluding intron-borne maturase sequences for family categorization, we added an independent analysis of maturase phylogenies and find a surprising incongruence between intron mobility and the presence of intron-borne maturases. Intriguingly, however, we find that several examples of nuclear splicing factors meantime characterized simultaneously facilitate splicing of independent paralogues now placed into the same intron families. Altogether this suggests that plant group II intron mobility, in contrast to their bacterial counterparts, is not intimately linked to intron-encoded maturases.

Introduction

At first glance, the two endosymbiont genomes in the plant cell seem to have followed very similar evolutionary trajectories with massive gene transfers into the nucleus, strongly reducing the ancestral genomes of an a-proteobacterium and a cyanobacterium that gave rise to mitochondria and chloroplasts, respectively. Even later emerging molecular mechanisms such as the site-specific C-to-U RNA editing characteristic for land plants affect the transcriptomes of both plant organelles equally [1]. In contrast, another characteristic feature of the two endosymbiont genomes in plants shows a striking discrepancy: the overwhelming stasis in the occurrence of group II introns in chloroplast DNAs for more than 500 million years, even extending into the streptophyte algal ancestors vs. the strikingly dynamic evolution of their counterparts in plant mitochondrial genomes. Examples for the latter are documented with the multiple splicing pathways of plant mitochondrial group II introns [2–4], their many transitions from *cis*- into *trans*-splicing arrangements [e.g. 5–12], the recently discovered functional or degenerated twintrons [13] or the evident degeneration of intron-borne maturase reading frames [5, 14] and the creation of nuclear-encoded maturase genes [15–20] in the multifarious pathways of evolution (Fig. 1A-E). Here, we focus on the occupation of new loci by mobile plant mitochondrial group II introns, including examples of introns “fossilized” in pseudogenes or intergenic regions (Fig. 1F-G).

The peculiar clade-specific patterns in the occurrence of plant mitochondrial group II introns [21–23] and the examples for evolutionary recent occupations of new intron insertion sites [24, 25] suggest that plant mitogenomes could offer a particularly rich and attractive data source to elucidate the mobility of group II introns on a broad evolutionary scale. Previous research on group II intron mobility has so far largely focused on functional studies with selected group II model introns in fungi or bacteria that have helped to elucidate the biochemical pathways of intron propagation and invasion of new loci [26–28]. Moreover, many features make group II introns particularly interesting ribozyme RNAs in general [29–38]: Firstly, the biochemical mechanism of their splicing, involving two transesterifications using a 5'-terminal guanosine and a looped out adenosine upstream of the 3'-splice acceptor site, is homologous to that of spliceosomal introns in the nuclear genomes of eukaryotes. In fact, all available evidence convincingly suggests group II introns to be the evolutionary ancestors of the eukaryotic spliceosome machinery [39–42]. Secondly, some group II introns have been demonstrated to have self-splicing activity *in vitro*, a notable case being the second intron in the mitochondrial *rrnL* gene of the brown alga *Pylaiella littoralis* [43] that splices efficiently even at low Mg^{2+} ion concentrations. Thirdly, in their likely evolutionary ancestral state, group II introns carry reading frames for maturases, intron-encoded proteins (IEPs) that are crucial to assist both in splicing but also in the mobility, allowing their host introns to occupy new insertion sites. In their most complete forms maturases consist of DNA-binding and endonucleolytic cleavage (D/En) and reverse transcriptase (RT) domains allowing the conversion of the RNA spliced into a novel DNA location, thereby creating a new intron paralogue.

However, none of the particularly abundant group II introns in plant organelle genomes have been experimentally demonstrated to be mobile and reports on self-splicing or *in vitro* splicing in the presence of chloroplast extracts, respectively, have only recently been published [44, 45]. The lack of plant mitochondrial group II intron self-splicing activity is likely associated with the diverse set of nuclear-encoded proteins that act as either intron-specific or more promiscuous splicing factors [46, 47] and which continue to be characterized in ever-increasing numbers by reverse genetic studies, as it is easily documented with only some most recently published examples [48–52].

While studies of group II intron mobility in plant mitogenomes are hampered by lacking experimental approaches to manipulate plant mtDNAs, the extraordinary diversity of introns in plant mitogenomes offers an intriguing evolutionary perspective. Sequence similarities between plant mitochondrial intron paralogues had become evident soon after the first complete plant mitogenome sequences became available [53]. For example, the overall sequence similarity between the first intron in the *nad2* gene (nad2i156g2) and the second intron of the *nad1* gene (nad1i477g2) in flowering plant mitogenomes has been recognized early [54]. Similarly, the case of nad5i392g2 is striking, an intron that is phylogenetically very restricted to the family Lycopodiaceae among the lycophytes but clearly related to downstream intron nad5i1242g2 in the *nad5* gene, which has a much wider phylogenetic conservation among lycophytes and ferns [55]. More recently emerging examples for evidently related group II intron paralogue pairs are two cases in ferns: atp1i361g2 and rps3i249g2 [25] and rps1i25g2 and rpl2i846g2 [24]. The hitherto most striking example of extraordinary sequence similarity between two intron paralogues has

only very recently been discovered in hornwort mtDNAs accompanying the discovery of twintrons [13]. Intron paralogue cox2i98g2 is exclusively present in the hornwort genus *Anthoceros* and shares more than 98% sequence similarity with the internal group II intron of a twintron in the *atp1* gene (atp1i1050g2ii1536g2), which is widely distributed among the hornworts. This example immediately suggests a recent copying event from *atp1* into *cox2* late in the recent phylogeny of hornworts.

Here, we systematically scanned available mitogenomes of streptophytes (i.e., land plants and the phylogenetically most closely related "charophyte" green algae) and scored the relationships between their group II introns as an effort to elucidate how intron migrations and losses may have contributed to shape the mitogenome makeups in extant plants. We define 25 core "families" F01 - F25 of clearly sequence-related group II intron paralogues, which indicate ancient and recent intron copying events. With the extraordinary slow sequence evolution in mitogenomes of the plant lineage combined with an increasingly well-understood phylogeny of land plants and related streptophyte algae our approach will offer a new perspective on an evolutionary timescale covering more than 500 million years. Altogether, we were able to assign 104 from a total of 161 streptophyte mitochondrial group II introns to one of the now defined 25 families whereas the 63 remaining introns presently lack significant similarity to any other paralogues. We discuss possible intron migration scenarios, considering the likely important role of intron-encoded or free-standing maturases.

Results

The streptophyte mitochondrial group II intron data set

We scanned plant mitochondrial genomes for the occurrence of group II introns, including the complete phylogenetic diversity representing the seven major land plant clades: Flowering plants (Angiosperms), Gymnosperms, Ferns (Monilophytes), Lycophytes, Hornworts (Anthocerotophyta), Mosses (Bryophyta) and Liverworts (Marchantiophyta). We additionally included the available mitogenomes of streptophyte algae ("Charophytes") representing six classes: Zygnematophyceae (Desmidiales: *Closterium bailyanum, Gonatozygon brebissonii*; Zygnematales: *Entransia fimbriata, Roya anglica, R. obtusa* and *Zygnema circumcarinatum*) recently considered to be most closely related to the plant lineage as well as Charophyceae (*Chara vulgaris, C. braunii, Nitella hyalina*), Coleochaetophyceae (*Chaetosphaeridium globosum, Coleochaete scutata*), Chlorokybophyceae (*Chlorokybus atmophyticus*), Klebsormidiophyceae (*Klebsormidium flaccidum*) and the early-branching Mesostigmatophyceae (*Mesostigma viride*). For a clear designation of introns we use the previously suggested intron nomenclature based on the intron insertion site in a given gene behind the reference position in the respective homologue of the liverwort *Marchantia polymorpha* [56, 57].

The example of the *cox2* gene (Fig. 2) is used to introduce into important issues of our analyses addressing the huge diversity of mitochondrial introns in the plant lineage. Altogether twelve different group II insertion sites are presently identified in the *cox2* genes of land plants and streptophyte algae. Their phylogenetic distributions vary widely from introns present in most land plant lineages excluding

liverworts (cox2i373g2 and cox2i691g2) to others presently only identified in the hornwort genus *Anthoceros* (cox2i98g2) or in the streptophyte alga *Coleochaete scutata* (cox2i550g2). The assignments of introns to core families, which comprise significantly similar intron paralogues as introduced in this work, are given below the intron names: cox2i381g2 in family F01, cox2i550g2 in F03, cox2i97g2 and cox2i98g2 in F04, cox2i564g2 in F16 and cox2i94g2 in F19 (see below). Including intron-based maturases allows for definition of maturase-based families and superfamilies, here cox2i373g2, in mF26 and cox2i127g2 in mF27. Remaining "solitary" introns lacking significant similarity to other paralogues are labeled "S". Insertion sites must be considered very carefully and precisely, notably when introns occur differently in a small gene region such as introns cox2i94g2, cox2i97g2, cox2i98g2 and cox2i104g2, an issue occasionally overlooked and leading to mis-annotation in database entries. Introns in the same position were considered as orthologues also when lacking significant sequence similarities across large phylogenetic distances if they were not assigned to different families.

Our re-evaluation of database entries during intron sampling has made us re-consider several intron and splice site annotations and allowed us to suggest yet further, and very likely functionally splicing, introns that were previously unnoticed but would reconstitute important and conserved parts of their host genes (e.g. cox3i34g2 in *Gonatozygon*, nad2i81g2 in *Zygnema* or nad9i89g2 in *Closterium* and *Gonatozygon*) as well as dysfunctional, "fossil" introns (e.g. rps8i52g2f in *Phlegmariurus*), as we will discuss below. Our final total mitochondrial group II intron sampling in streptophyte mitogenomes comprised 161 group II intron paralogues defined by their unique insertion sites.

The occurrence of streptophyte mitochondrial group II intron paralogues in the major clades is displayed with Euler diagrams in Fig. 3. The striking discrepancy of mitochondrial vs. chloroplast group II introns is immediately apparent with 22 land plant chloroplast introns all of which have counterparts in streptophyte algae vs. altogether 161 mitochondrial group II introns, of which only 13 are shared between embryophytes and streptophyte algae (Fig. 3A). Differentiating among the embryophytes, larger intersections are found between bryophytes and tracheophytes (Fig. 3B) than with either group and the outgroup algae and among the latter between hornworts and mosses (Fig. 3C) and between hornworts, mosses and tracheophytes (Fig. 3D), respectively. Notably, of 101 group II introns identified in embryophyte mtDNAs, only one (atp9i87g2) is shared between all three bryophyte clades (Fig. 3C).

The group II intron family concept

Despite their highly conserved six-domain structures, group II intron paralogues *a priori* share no significant sequence similarities aside from mostly conserved sequence motifs at the 5'-end (mostly GUGCG), the 3'- termini (mostly AY) and their characteristic domain V comprising 34 nucleotides, which mostly folds into two base-paired regions of 9 and 5 base pairs, respectively, with a dinucleotide bulge and a terminal tetraloop (most often GNRA) in the overwhelming majority of group II introns. Only a consensus sequence representing domain V may in fact be used as an initial query to scan for group II intron candidates in sequence databases [58].

Our criteria for considering introns occupying different insertion sites for inclusion into families of related paralogues are detailed under methods. In brief, we used several rounds of identifying sequence-related intron paralogues that share significant sequence similarities (beyond short similarities of domain V and the immediate flanking regions), which can exceed those of evidently orthologous introns in the same positions in distant plant taxa. Naturally, some rare borderline cases are represented with introns occupying the same insertion site in phylogenetically very distant taxa. An independent insertion of a given position cannot be excluded as an alternative explanation to vertical transmission followed by sequence divergences obliterating recognizable similarities. Below, we will discuss such borderline cases in the context of our consideration of the now defined group II intron families. A cladogram of group II introns sorted into families is shown in Fig. 4. Subsequently, we will consider, where present, similarities of intron-borne maturases to independently verify family assignments based on the nucleotide sequence similarities alone. Furthermore, maturase similarities will occasionally allow for the inclusion of some previously “solitary” introns lacking significant nucleotide sequence similarities to other paralogues into extended families and for some fusions of primary intron families into “superfamilies” (Fig. 5).

Family F01

Among the here defined families of mitochondrial group II introns in the streptophyte lineage, family F01 is the largest one, comprising altogether 13 intron paralogues (Fig. 4, Fig. 6). It includes four isolated introns, each presently identified in only one streptophyte algae clade (atp9i209g2 and cox3i745g2 in *Chlorokybus*, atp9i214g2 in Charophyceae and cox1i1039g2 in *Coleochaete*, respectively). The “moss-specific” introns cox3i507g2 and nad9i283g2 of F01 are universally conserved in that plant clade. Phylogenetically more broadly distributed introns atp6i439g2, nad4i1399g2, nad5i230g2, nad7i209g2 and rps3i74g2 occur in up to five embryophyte clades. In two cases, the respective introns remain recognizable despite pseudogene degeneration of *nad7* and *rps3* in hornwort mitogenomes. Three introns in family F01 are present both in at least one embryophyte clade and in up to three classes of streptophyte algae: rps3i74g2, atp6i439g2 and cox2i381g2 (Fig. 4, Fig. 6). Notably though, family F01 does not include an intron paralogue present in liverworts. Only four of the intron paralogues in F01 carry maturases: atp9i209g2, cox1i1039g2, cox2i381g2 and nad3i211g2 (Fig. 4, Fig. 6).

Although outgroup rooting is necessarily difficult for the individual clades of intron families, the unrooted phylogeny suggests particular close relations of certain introns in F01, in this case of the “moss” intron cox3i507g2 with the “tracheophyte” intron nad4i1399g2 and of the phylogenetically wider distributed paralogues nad5i230g2 with rps3i74g2 (Fig. 6). Interestingly, the latter two and nad7i209g2 as a third intron paralogue of the F01 family have recently been demonstrated to be affected by the same nuclear-encoded splicing factor in the model angiosperm *Arabidopsis thaliana*, OZ2 [50]. Also quite notably, the maize NCS2 mutant (non-chromosomal stripe mutant) described long ago [59] is the result of a mitogenome recombination between two mitochondrial introns that we have now assigned to family F01: nad4i1399g2 and nad7i209g2.

The phylogeny of intron paralogues is in accordance with known species phylogenies for the well-supported nodes, while apparent phylogenetic “mis-placements” e.g., of the diverging lycophyte sequences, remain without support (Fig. 6). However, and notably so with respect to phenomena like shared splicing factors, the possibility of concerted and converging evolution must be considered. Whereas the case of atp6i439g2 is an example where even the distant algal orthologue in *Zygnema* is included with its counterparts in lycophytes and hornworts in one subclade, this is not the case for rps3i74g2 where the Charales orthologues are attracted to the isolated atp9i214g2 paralogue in *Chara vulgaris*. Our later additional considerations of maturase protein similarities that we will discuss below includes family F01 into a “superfamily” also including the small family F22 and two “solitary” introns (Fig. 5).

Family F02

As in family F01, some intron paralogues in family F02 likewise have a very restricted occurrence: cox3i625g2 is exclusively present in liverworts and nad9i246g2 is only identified in hornworts (Fig. 4). However, despite their phylogenetically disjunct occurrence, cox3i625g2 and nad9i246g2 are closely related paralogues in F02 (Fig. 7). Yet more phylogenetically restricted, nad5i392g2 is only present in the lycophyte order Lycopodiales, here represented by *Phlegmariurus squarrosus*. The extreme diversity of mitogenome evolution within the lycophyte clade with retained genes and introns and a conserved genetic synteny in the *P. squarrosus* mtDNA versus highly recombining mitogenomes with reduced gene sets in Isoetales and Selaginellales is also fully in line with intron rps10i235g2 being retained in *Phlegmariurus* and shared with the seed plants. The similarity of angiosperm rps10i235g2 to liverwort introns rrnLi833g2 and cox3i625g2 had already been noted right along with its initial discovery [60]. Other than rps10i235g2, *P. squarrosus* intron nad5i1242g2 is shared with ferns, similarly indicating an early vascular plant ancestry. Notably, we recently found intron rrnLi833g2 universally conserved among liverworts now to be also present in the mitogenome of the leptosporangiate fern *Haplopteris ensiformis* [61] and in a preliminary mtDNA assembly of the fern *Dryopteris crassirhizoma* (database accession MW732172). Intriguingly, however, the fern rrnLi833g2 introns cluster with their nad5i1242g2 paralogues, possibly indicating concerted evolution or a loss-and-regain scenario.

Intron family F02 neither includes an extant intron paralogue present in mosses nor one carrying a maturase ORF (Fig. 4). Notably though, we were able to identify several additional fossilized group II intron sequences “F02g2f” clearly tracing back to family F02 intron paralogues both in *Phlegmariurus squarrosus* and in hornwort mitogenomes (Suppl. Figure 1A & B). The series of intron copying events leading to the five F02 paralogues now recognized in the *Phlegmariurus* mitogenome (Fig. 7B) remains unclear except for nad5i392g2 present only in Lycopodiales and likely emerging late from the more ancestral nad5i1242g2 paralogue. The newly identified fossil intron rps8i51g2f receives the ‘f’ behind the g2 intron label because we could not confirm functional splicing despite significant similarity to its functional counterparts in *nad5* and *rps10* extending up to the very intron 5’- and 3’-ends as opposed to another intergenic F02 intron fossil in the *tatC-cox2* spacer (Fig. 7C). Altogether five intergenic F02 intron fossils could be discovered in the hornwort mitogenomes, likely originating from nad9i246g2 or a

common ancestor (Fig. 7D). In this case, the two intergenic intron fossils between *rps4* and *tatC* and between *atp6* and *nad6* appear more closely related to each other than the remaining three. The extension of the intron sequence similarities ending with the respective 5'- or 3'-intron termini and their identification in the conservatively evolving mitogenomes of early land plants devoid of heavy recombination as in seed plant mitogenomes strongly argues for their origin by retrotransposition rather than mtDNA recombination.

Family F03

Intron family F03 contains two intron paralogues, nad4i976g2 and nad5i1477g2, shared between hornworts and all four tracheophyte clades, potentially supporting a hornwort-tracheophyte (HT) clade (Fig. 4). However, both nad4i976g2 and another intron paralogue in F03, nad3i140g2, also reveal counterparts in Charophyceae algae. No losses are identified for nad3i140g2 among available liverwort, hornwort or lycophyte mitogenomes. Similarly, two further intron paralogues are universally conserved in mosses and shared with either lycophytes (atp9i21g2) or hornworts (sdh3i100g2). Finally, three F03 introns have a restricted occurrence among streptophyte algae, of which only one is shared with a land plant clade, cox2i250g2 present in *Closterium* (Zygnematophyceae) and in the liverworts. This intron is an interesting exceptional case with a maturase carried in the liverwort orthologues, but not in the algal counterpart of cox2i250g2. Among the eight intron paralogues in F03 (Fig. 4), two pairs of introns show particularly high sequence similarities: nad3i140g2 with nad5i1477g2 and sdh3i100g2 with nad4i976g2. In this case, the intron paralogues occur with a phylogenetic overlap in the hornworts, a clade characterized by high intron mobility as also reflected below with the example of family F04. Again, we were able to identify fossilized group II intron sequences ("F03g2f") in intergenic sequences of the lycophyte *Phlegmariurus squarrosus* clearly tracing back to family F03 (Suppl. Figure 1C).

Family F04

Group II intron family F04 is dominated by introns exclusively occurring in hornworts, indicating a pronounced intron mobility in that land plant clade: cox2i98g2, nad5i881g2, nad6i444g2 and nad9i502g2 (Fig. 4). This is further enhanced by the observation that the respective internal introns of two recently discovered twintrons [13], atp1i1050g2ii1536g2 and cox1i1116g1ii207g2 in the hornwort genus *Anthoceros*, likewise belong to the here defined F04 family. In fact, the extraordinary sequence similarity of 98% between atp1i1050g2ii1536g2 to cox2i98g2 is the most striking case of closely related paralogues in our entire intron sampling, indicating a very recent copying event. Considering that the *atp1* twintron is present in all hornworts whereas cox2i98g2 exists only in the genus *Anthoceros* immediately suggests a copying from the former to the latter insertion site. Notably, however, internal splicing of zombie-twintron atp1i1050g2ii1536g2 could not be detected.

At the same time, family F04 is a prime example for the necessity of careful analyses to distinguish cox2i98g2 from its paralogue cox2i97g2 in liverworts that is inserted one nucleotide upstream in the *cox2* gene (Fig. 2). Intron nad7i676g2 is conserved in all tracheophyte clades. Its former presence in hornworts, and accordingly in a possible HT stem lineage, is elusive owing to the pseudogene degeneration of the

downstream part of *nad7* in all hornwort mtDNAs. Notably, and despite the very clear evidence of recently active retrotransposition, none of the F04 members carries a maturase-ORF (Fig. 4).

Family F05

In contrast to F04 mainly comprising intron paralogues in hornworts, group II intron family F05 (Fig. 4) is dominated by intron paralogues exclusively occurring in liverworts: nad4i548g2, nad4Li100g2, nad7i336g2 and rpl2i28g2. Liverwort-specific intron nad4i548g2 had been introduced to elucidate the liverwort phylogeny [62] and is now found to be notably similar to rpl2i28g2. Other than the four liverwort paralogues, F05 also includes lycophyte introns recently discovered and characterized as the external intron of a “zombie” twintron upon closer reinspection of the *Phlegmariurus squarrosus sdh3* gene [13].

Family F06

Apart from rps3i249g2 that is shared between lycophytes, ferns and gymnosperms, the group II intron paralogues in family F06 are phylogenetically very restricted in occurrence: atp1i361g2 exclusively in ferns, cobi783g2 only in liverworts and cox1i150g2 exclusively in hornworts (Fig. 4). A detailed study of atp1i361g2 concluded that this fern-specific intron has originated from the, likely more ancestral, paralogue rps3i249g2 and also found evidence for convergent evolution of specific intron structures, mainly group II intron domain III, in the two paralogues in later emerging fern lineages [25].

Interestingly, a truncated copy of cobi783g2 had been identified earlier in the spacer between *nad5* and *nad4* in liverworts in an early sampling of this intergenic region among bryophytes [63]. Our extended consideration of intron families now adds support to the idea that also this case of an intron fossil has arisen through copying, likely by a retrotransposition event. Similarity of the fossil sequence starts exactly from the intron 5’-end and extends for nearly 800 bp with 99% identity in the case of *Treubia lacunosa* representing an early liverwort branch (Suppl. Figure 1D) whereas a higher degree of degeneration is observed in derived taxa. Yet other intron fossils have now been discovered in the intergenic region between *rrn5* and *trnM*-CAU of the hornwort *Anthoceros agrestis* and the moss *Sphagnum palustre* mitogenomes (Suppl. Figure 1E). The only evidence for a maturase among the four members of intron family F06 are traces of a former maturase-ORF in cox1i150g2.

Family F07

Intron family F07 contains three paralogues of phylogenetically wide distribution and only one member with isolated occurrence in the liverworts alone: cobi372g2 (Fig. 4). The similarity between introns nad1i477g2 and nad2i156g2, now found to be widely conserved among vascular plants (Fig. 4), has been recognized early after the complex structures of *nad1* and *nad2* in flowering plants had been elucidated [54]. Quite interestingly, the nuclear encoded splicing factor ODB1 has meantime been found to promote splicing of both of these two paralogues [64]. The fourth paralogue in family F07, intron atp9i87g2, is particularly noteworthy for (i) being an intron now identified to be shared between all three bryophyte clades, (ii) carrying an ancestral maturase that has independently degenerated in all four plant clades where it is present and (iii) carrying the internal intron atp9i87g2ii1114g2 of family F08 (see

below) turning it into a twintron in the lycophyte *Phlegmariurus squarrosus*. A history of intron paralogues is immediately suggestive for the family F07 paralogues (Fig. 8). The ancestral maturase-carrying atp9i87g2 independently likely gave rise to cobi372g2 in liverworts and to nad2i156g2 in a possible non-liverwort lineage. After loss of nad2i156g2 in hornworts, it gave rise to nad1i477g2 only in the tracheophyte lineage.

Family F08

Two introns in family F08 have a phylogenetically striking distribution being conserved in liverworts but also occurring in lycophytes: nad7i1113g2 and rps14i114g2 (Fig. 4). Given that the host genes, *nad7* and *rps14*, are frequently subject to EGT, nad7i1113g2 is at present only determined in *Isoetes engelmannii* and rps14i114g2 only in *Phlegmariurus squarrosus* among the lycophytes. However, we now found rps14i114g2 equally conserved in the mitogenome of the leptosporangiate fern *Haplopteris ensiformis* [61]. Yet more notably, the recently characterized inner intron of a twintron in the *atp9* gene, atp9i87g2ii1114g2 [13] as a third member in family F08 is characteristically more similar to its rps14i114g2 counterpart in liverworts than in *P. squarrosus*.

Family F09

Intron ccmFCi829g2 in F09 has a phylogenetically wide distribution in bryophytes and tracheophytes (Fig. 4). Its former presence in lycophytes remains unclear, however, owing to the loss of the entire *ccm* gene suite for cytochrome c maturation in this clade. In contrast, pseudogene traces of *ccmFC* including ccmFCi829g2 are clearly detected in hornworts [65]. Intron rpl2i846g2 is evidently a gain in the tracheophyte stem lineage. This ancestral intron obviously gave rise to its closely related paralogue rps1i25g2 exclusively present in ferns [24]. As in family F08, no traces of maturase-ORFs are recognizable in any of the F09 intron paralogues. Highly interesting, however, the splicing of both F09 paralogues present in *Arabidopsis thaliana*, ccmFCi829g2 and rpl2i846g2, was found to be affected by the same nuclear-encoded splicing factor, WTF9 [66].

Family F10

Intron family F10 comprises group II introns with a particularly complex history (Fig. 4). It contains two intron paralogues with a phylogenetically disjunct distribution: cobi824g2 exclusively present in liverworts and nad1i669g2 previously assumed to be restricted to tracheophytes. Intron nad1i669g2 has received attention as being conserved in a *trans*-splicing arrangement in seed plants. A *cis*-arranged orthologue was initially identified in the fern *Osmunda regalis*, also noticing traces of a degenerated maturase [5]. We now found that intron nad1i669g2 has a clear orthologue also in the mtDNA of the alga *Coleochaete scutata* [67] with a maturase-ORF annotated in the corresponding database entry (MN613583). Intron family F10 also contains the “hypermobile invader” intron cox1i1149g2 in the lycophyte *Phlegmariurus squarrosus*, which gave rise to internal introns of two twintrons (in the newly determined sdh3i349g2 and in itself) and to seven further intron fossils in intergenic regions [13] (Suppl. Figure 1F).

Family F11

Intron family F11 contains introns atp6i80g2 and atp9i95g2 of phylogenetically wider distribution, shared between mosses, hornworts and lycophytes (Fig. 4). Like the above cases of the neighboring introns cox2i97g2 and cox2i98g2 in F04, also atp9i95g2 requires careful inspection given the closely neighboring intron atp9i87g2 in family F07 [65]. Despite the evident dynamics of intron insertions in *atp9* of bryophytes and lycophytes, no evidence is ever found for an intron in the *atp9* gene among ferns, gymnosperms or angiosperms. Intron cox1i1064g2 is exclusively present in mosses and cox1i653g2 is only present in hornworts. Again, the latter needs particular attention to distinguish it from an unrelated "solitary" intron cox1i652g2 inserted one nucleotide upstream in the *Coleochaete* mitogenome.

Family F12

The two intron paralogues in family F12 (Fig. 4) have a strikingly divergent distribution with cobi274g2 presently only identified in Charales algae and nad2i1282g2 present in hornworts and euphyllophytes (i.e., ferns and seed plants). No traces of former maturase reading frames can be detected in nad2i1282g2 or cobi274g2.

Family F13

Group II intron family F13 contains two intron paralogues conserved in liverworts: cox1i178g2 and cox3i171g2 (Fig. 4). The former has counterparts in the algae *Klebsormidium flaccidum* and *Coleochaete scutata*. However, only cox1i178g2 in *Coleochaete* shares significant sequence similarity with the liverwort homologues, clearly warranting family inclusion according to our criteria. The "positional homologue" in *Klebsormidium* neither shares similarity with the liverwort nor with the *Coleochaete* counterpart, leaving its status as a true orthologue vs. a possible "analogue" occupying the same insertion site open at present. The cox1i178g2 introns carry maturase reading frames both in algae and the liverworts, no intron-borne ORFs are present in the much smaller cox3i171g2 paralogues shared between liverworts and the lycophyte *Phlegmariurus squarrosus*.

Family F14

As in family F12, the intron members in family F14 are also very disjunct in occurrence with nad1i258g2 being restricted to the monilophyte (fern) clade and nad1i517g2 so far only identified in the alga *Zygnema circumcarinatum* (Fig. 4). The third paralogue, cox1i511g2, however, is present in liverworts, mosses, lycopyhtes and the algae *Coleochaete* and *Zygnema*. Like most group II introns in the algal mitogenomes also cox1i511g2 carries maturase reading frames. The counterparts in the land plant lineage, however, are frameshifted ORFs in liverworts and degenerated or unrecognizable in the cox1i511g2 orthologues of mosses or in the lycophyte *Selaginella moellendorffii*. We identified an F14-type intron fossil (F14g2f) in the *cox1-rrnS* spacer of the *Coleochaete* mitogenome (Suppl. Figure 1G).

Family F15

As in F14, the intron paralogues in family F15 are likewise phylogenetically disjunct (Fig. 4). Intron cox1i835g2 is presently only identified in the algal genus *Chara* (and absent in the Characeae genera *Nitella* and *Nitellopsis*). Intron nad1i287g2 had initially been identified serendipitously in a screen for ancestors of *trans*-splicing intron nad1i394g2 [5] and is meantime found to be universally conserved both in mosses and in hornworts [56, 65]. As in the above cases, a maturase reading frame is only present in the algal cox1i835g2 but not in the much smaller nad1i287g2 paralogues in the bryophytes of less than 800 bp. The subsequent additional analyses of intron-borne maturases (see below) add S-type intron atp9i145g2 in Charales to an extended family eF15, which is ultimately linked to the superfamily SF13-14 (Fig. 5).

Family F16

Family F16 contains two intron paralogues that occur exclusively in hornworts: cox2i564g2 and nad1i348g2 (Fig. 4). The latter intron is of exceptionally small size of less than 600 nt. and conserved in all hornwort genera. In contrast, cox2i564g2 is ca. five times larger and lost together with the upstream and downstream neighboring introns cox2i381g2 and cox2i691g2 in *Nothoceros aenigmaticus*. Despite their extended sizes of more than 2.5 kb no maturase traces are discernible in the hornwort cox2i564g2 copies.

Family F17

Intron family F17 contains two intron paralogues with a phylogenetic distribution that could have been taken as further support for an NLE (“Non-Liverwort Embryophyte”) clade: nad1i728g2 and nad4i461g2 (Fig. 4). Both introns are absent in liverworts but particularly well conserved in mosses, hornworts, lycophytes and vascular plants with only rare exceptions including the absence of nad4i461g2 in the hornwort *Leiosporoceros dussii* and of nad1i728g2 in the lycophyte *Isoetes engelmannii*. Intron nad4i461g2 counterparts in the algae *Coleochaete scutata* and *Zygnema circumcarinatum* are large introns of 3.1 and 5.7 kb with long maturase reading frames that are continuous with the upstream *nad4* coding sequence. Matching the general observations, only small traces of former maturases remain in the land plant counterparts where the sizes of nad4i461g copies in the mosses are reduced to less than 800 bp.

Intron nad1i728g2 is a particularly interesting case, being the only mitochondrial intron carrying a maturase reading frame in flowering plants, widely known as “matR”, now systematically labelled *mat-nad1i728g2*. Moreover, intron nad1i728g2 is also known for multiple independent transitions from *cis*- to *trans*-splicing with intron disruptions either up- or downstream of the maturase ORF in flowering plants [6]. Intriguingly, a gene transfer of the *matR/mat-nad1i728g2* reading frame into the nuclear genome has been identified in *Pelargonium* [68]. Most interestingly, three different nuclear-encoded splicing factors have already been identified, which affect the two closely related F17 angiosperm paralogues nad1i728g2 and nad4i461g2 simultaneously: EMP8 [69], DEK55 [70] and SMK3 [71]. The third paralogue in F17, cobi399g2, is so far only identified in the *Coleochaete scutata* mitogenome and, as in most cases in the algal mitogenomes, also carries a maturase reading frame. The intergenic region between *trnM-*

CAU and *trnA-UGC* contains an F17-type intron fossil (F17g2f) in the *Phlegmariurus squarrosus* mitogenome (Suppl. Figure 1H).

Family F18

Family F18 contains two intron paralogues presently only identified in algae of the Zygnematophyceae (Fig. 4): nad5i537g2 is present exclusively in *Zygnema circumcarinatum* and nad7i777g2 is present in *Z. circumcarinatum* and in *Closterium baillyanum*, of which only the latter homologue is equipped with an intron-borne maturase ORF.

Family F19

Group II intron family F19 (Fig. 4) comprises two introns present in lycophytes where sequence similarities are blurred by the highly divergent mitogenomes in genera *Isoetes* and *Selaginella* and largely rely on the conserved mitogenome of *Phlegmariurus squarrosus*. While intron cox2i94g2 is exclusively present in all three orders of lycophytes, intron paralogue cox1i323g2 is also present in the mitogenome of *Sphagnum*, representing a very early branch in the phylogeny of mosses. In both, *Phlegmariurus* and *Sphagnum* cox1i323g2 contains maturase remnants. Notably, cox1i323g2 has extensive similarity with the extended intergenic region between *nad9* and *trnI-CAU* in hornwort mtDNAs, representing yet another example for traces of an intron fossil (Suppl. Figure 1I).

Family F20

Intron family F20 comprises two intron paralogues so far exclusively identified in the mitogenome of the alga *Closterium baillyanum* of the Zygnematophyceae: cox3i641g2 and nad9i275g2 (Fig. 4). Neither of the two intron paralogues carries a maturase reading frame.

Family F21

Intron family F21 contains introns presently only found in *Coleochaete scutata*: atp1i66g2, nad5i1725g2, rrnSi435g2 and trnH-GUGi32g2 (Fig. 4). Only the nad5i1725g2 paralogue carries a maturase. The size increase of the *Coleochaete scutata* mitogenome in comparison to that of *Chaetosphaeridium globosum*, also in the Coleochaetales, had in part been ascribed to the presence of 57 vs. only 11 introns [67]. We now find the here defined family F21 particularly interesting, because additional intron homologies are also present in multiple intergenic regions of the *Coleochaete* mitogenome: *trnMf-nad9, mttB-trnL* and *trnV-trnD*. In all three cases, homologies sharply coincide with the intron 5'-end and extend for minimally 600 bp indicating retrotransposition rather than events of DNA recombination at their origins (Suppl. Figure 1J).

Family F22

As in the case of F21, intron family F22 also includes members presently only identified in the mitogenome of *Coleochaete scutata*: atp1i850g2 and nad5i362g2 (Fig. 4). Of these two, only atp1i850g2 carries a maturase reading frame.

Family F23

Intron family F23 comprises two intron paralogues that are hitherto likewise only identified in streptophyte algae: nad7i925g2 in *Closterium baillyanum* and rrnLi2032g2 identified in *Coleochaete scutata* and *Nitella hyalina* (Fig. 4). Only the latter carries a maturase reading frame.

Family F24

Intron family F24 comprises rrnSi1148g2 present in *Coleochaete scutata*, rrnLi1747g2 presently only identified in *Entransia fimbriata*, nad7i250g2 present in the Zygnematophyceae algae *Closterium baillyanum* and *Gonatozygon brebissonii*, rrnLi629g2 present in *Entransia*, *Coleochaete* and *Nitella* and trnS-GCUi43g2, the only F24 paralogue shared with embryophytes (Fig. 4). The latter is present in *Chlorokybus* and the Zygnematophyceae genera *Closterium, Gonatozygon* and *Roya* and highly conserved among liverworts. Moreover, trnS-GCUi43g2 is evidently present as a degenerated copy in the *trnS-GCU* pseudogene retained in the conserved intergenic space between *trnA* and *trnD* in the mitogenomes of mosses and the lycophyte *Phlegmariurus squarrosus*. An independent degeneration has occurred among hornworts with functional copies present in *Leiosporoceros* and *Anthoceros* but pseudogenes in *Nothoceros* and *Phaeoceros*. Hence three independent degenerations of *trnS-GCU* and its intron are evident in land plants in mosses, among hornworts and in the tracheophytes (Suppl. Figure 1K).

Family F25

Family F25 comprises two intron paralogues of phylogenetically disjunct distribution: cobi537g2 present in the Charales algae and nad5i1455g2, present in all embryophytes except the liverworts (Fig. 4).

Maturases in the plant mitochondrial lineage and extended maturase-based intron families

Our categorization of 100 group II introns into the 25 “core” families F01 - F25 outlined above is based on their primary nucleotide sequence similarities alone. To cross-check for independent confirmation and to explore further and deeper relationships, also outside of the streptophytes, we independently compiled the intron-borne maturase ORFs present in 43 streptophyte mitochondrial group II introns as seeds for identifying protein homologs. This seed query data set also included several maturases that remained hitherto unnoticed or not annotated in database entries (e.g., the spliced variants of *mat-atp9i87g2* and *mat-atp9i95g2* in *Phlegmariurus*). The search for homologs ultimately resulted in a large protein data set that also contained related maturases of distant chlorophyte algae as well several maturase proteins in red algae, stramenopiles, fungi, animals and bacteria. The independent phylogenetic analysis of the large maturase protein sequence collection (Suppl. Figure 2) fully confirmed the identified core families F01, F03, F10, F11, F17 and F25, all of which contain at least two intron paralogues carrying maturases (Fig. 5). Moreover, the maturase similarities identified four additional, “maturase-based” group II intron families mF26-mF29 and helped to define superfamilies (SF) of higher order that combine the core intron families and include additional, previously solitary, introns.

For space limitations, we here focus on the examples of the large superfamily SF01-22 comprising families F01, F03, F22 and the two previously solitary introns cox1i652g2 and cox1i769g2 (Fig. 9A) and

on superfamily SF10-28 comprising families F10, F17 and mF28 (Fig. 9B). The independent protein analysis fully confirms monophyly of the maturases in F03 and a well-supported clade of maturases in F01, now extended to include to mat-cox1i769g2, mat-atp1i850g2, the free-standing maturases in the mitogenomes of liverworts and a nuclear maturase copy (mat-nuc1) in the moss *Physcomitrium* (Fig. 9A). Notably, the extended SF01-22 superfamily also includes homologs in fungi having identical insertion sites and clustering with *mat-cox1i652g2* in *Coleochaete* with high support. *Vice versa*, the extended F03 maturase clade likewise includes fungal mitochondrial maturases and a rhodophyte plastid maturase and, maybe even more notable, a cluster of nuclear maturases in tracheophytes (Fig. 9A).

As in the above case, the independent maturase phylogeny perfectly confirms the intron assignments to families F10 and F17 and adds the maturase-based family mF28 for a joint inclusion in superfamily SF10-28 (Fig. 9B). Particularly intriguing further cases for introns giving rise to fossil paralogues are “liverwort” introns atp1i989g2 and atp1i1050g2 (also present in hornworts), which are now jointly placed in mF28 (Fig. 4, Fig. 9B). An extended and significant sequence similarity (with perfect intron domain V and VI ends) of atp1i989g2 is present in the mitogenome of the lycophyte *Phlegmariurus squarrosus* embedding the *trnI-rps11* region with all nine genes in the same direction fitting the intron orientation (Suppl. Figure 1L). Hence, it appears that a huge block of genes was inserted into an intergenic intron fossil paralogue of atp1i989g2. Interestingly, intron atp1i989g2 is absent in the early branching liverwort genera *Treubia* and *Haplomitrium*, which could have indicated a gain within the liverworts only after split of the Haplomitriopsida, but this now seems unlikely given the unorthodox *Phlegmariurus* fossil paralogue. Along the same lines, intron atp1i1050g2 has fossil intron paralogues not only in liverwort mitogenomes but also in the *Phlegmariurus* mtDNA behind *cox1* in the spacer towards the *trnW* gene running in the opposite direction (Suppl. Figure 1M).

A further, very notable insight emerges from the maturase phylogeny: Mitochondrial group II introns with intron-borne maturases at the same insertion sites appear distributed across very distantly related lineages of eukaryotes. The most striking example concerns group II introns inserted into position 1147 of the *cox1* gene. Solitary type intron cox1i1147g2 inhabits mitogenomes of the streptophyte alga *Coleochaete*, but also in chlorophytes, rhodophytes, fungi and metazoa. The associated *mat-cox1i1147g2* RT-domains, the X-domain and the D/En domain are highly conserved. Similarly, the peculiar case of cox1i748g2 in *Equisetum arvense* [72], but not *E. diffusum*, having no similarity to its *Chlorokybus* counterpart is also found in the brown alga *Pylaiella littoralis* and the red alga *Pyropia fucicola.* A third intriguing example along those lines is cox2i373g2 of *Coleochaete* (now placed in mF26) that has maturase-free orthologs in mosses, hornworts and tracheophytes but forms a well-supported maturase-based clade with *mat-cox2i373g2* in ascomycetous fungi, e.g., the endophytic symbiont *Epichloe.*

The remaining “solitary” introns lacking significantly similar paralogs

Altogether 61 streptophyte group II introns lacked paralogues with significant nucleotide sequence similarity precluding their assignment into our 25 core families. Extending the analysis to characteristic

maturase similarities placed 15 of them in the four additional families mF26-mF29 and included another four previously solitary introns into superfamilies (cox1i44g2, cox1i245g2, cox1i769g2 and cox1i652g2). This leaves 42 streptophyte mitochondrial group II intron solitary, lacking both primary nucleotide similarity to paralogues and an intron-borne maturase of significant similarity to protein homologues (Fig. 5).

Many of the core solitary group II introns are presently very restricted in occurrence, most notably many introns presently identified in only one algal genus. This category, for example, also includes the unique *trans*-splicing introns nad3i84g2 and nad3i301g2 in the *nad3* gene in *Mesostigma viride*. Among land plants this includes solitary introns exclusively restricted to mosses (atp1i1127g2 and cox1i1200g2), to hornworts (atp1i805g2, atp1i1019g2, cobi838g2, cox1i1298g2, cox2i281g2 and cox3i109g2) and to lycophytes (cobi693g2, cox1i227g2, cox1i266g2 and cox1i995g2), respectively, whereas no solitary introns are identified that are restricted to either ferns, gymnosperms or angiosperms.

Other solitary introns, however, are shared between at least two major plant clades (Fig. 5) e.g., nad4Li283g2 in liverworts and mosses or nad3i52g2 in hornworts and lycophytes. Yet others are shared among all euphyllophytes (nad5i1872g2), all tracheophytes (nad1i394g2, nad2i542g2 and nad7i917g2) or tracheophytes and at least one bryophyte clade (cox2i691g2, nad2i709g2 and nad7i140g2). An interesting further case is trnN-GUUi38g2 which is present in three algal classes likely close to the land plant lineage, absent in liverworts but clearly recognizable in pseudogenized form in mosses, hornworts and *Phlegmariurus* among the lycophytes (Suppl. Figure 1N).

Discussion

Group II intron mobility outside of land plants

It is reasonable to assume that an ancestral state of group II introns with fully equipped intron-borne maturases is at the origin of their mobility and diversity (Fig. 1A). After the first discoveries of group II introns in bacteria [58, 73–75], their collection has grown immensely [76]. Bacterial group II introns are frequently associated with mobile genetic elements and/or present on plasmids [77] whereas their presence in essential genes is quite rare in bacteria [78]. Accordingly, their routes of dispersal are somewhat hard to trace and additionally complicated by the generally abundant horizontal gene transfer (HGT) among prokaryotes. Notably, prokaryotic group II introns lacking maturases are rare with “only a handful of ORF-less introns in bacteria” [79] and the few examples suggest a (likely quick) degeneration from maturase-bearing counterparts (Fig. 1B).

It should be noted that a “family” terminology for introns has already been used earlier in scoring bacterial group II intron occurrences and their retroelement behavior [79–81]. Among the research on group II introns in the bacterial world [82], the studies of mobile group II introns in *Wolbachia* endosymbionts [83] or of the highly mobile RmInt1 intron shaping the genome of *Sinorhizobium meliloti* and related α-proteobacteria [84] are typical cases in point on low taxonomic levels.

Several reports on variability, and likely mobility, of group II introns have also been published for eukaryotic organelle genomes, e.g. for the chloroplast genome of *Euglena* species [85], for mitochondria of different isolates of the brown alga *Pylaiella littoralis* [86] or in chloroplasts and mitochondria of the red algal genus *Porphyridium* [87, 88]. On somewhat higher taxonomic levels, significant group II intron diversities have been reported for the plastid genomes of cryptophytes [89] or the mitogenomes of diatoms [90, 91]. Extraordinary similarity of cyanobacterial introns in *Porphyra* [92] or the chloroplast of *Chlamydomonas* [93] or *Euglena* [94] have been discussed as horizontal transfer events. Likewise, horizontal transfers were also invoked as the likely cause in other cases of striking intron similarities, e.g. between those of diatoms and a haptophyte [95], for diatoms as likely group II intron donors into a raphidophycean flagellate genus *Chattonella* [96] or to explain the discovery of a first group II intron in a bilaterian mitogenome [97].

The plant mitochondrial group II intron family concept

The extraordinary dynamics of plant mitochondrial introns is in stark contrast not only to the overall stability of the chloroplast intron complements [98, 99], but also to the stasis of nuclear introns since 500 million years and more [100]. Many review articles [21, e.g. 101] and reports on newly completed plant mitogenomes, respectively, often contain comparative summaries and updates on the striking diversity of plant mitogenome makeups with respect to their gene and intron complements [11, 12, 65, 102].

Here, we have added on the previous comparative compilations by providing the hitherto most extensive overview on mitochondrial group II introns by including mitogenome analyses of all embryophyte and streptophyte algae lineages, also taking care of pseudogenization events and the existence of fossil introns in intergenic spacer regions. More importantly, we provide a concept for categorizing group II introns into families of related paralogs as a foundation to explore intron “copying” retrotransposition events during plant evolution. Of altogether 161 streptophyte mitochondrial group II introns, we have assigned 100 into 25 “core” families of minimally two and up to 13 intron paralogues based on their nucleotide sequences alone (Fig. 4). The family assignments may certainly be subject to further changes with newly discovered mitochondrial group II introns in streptophytes added to the 25 defined families or with new families or superfamilies (Fig. 5) being created by newly identified introns linking existing families. While we do not expect many more introns to be discovered in the land plant (embryophyte) lineage, the highly diverse streptophyte algae are presently still underrepresented in the sampling for mitogenomes with only twelve genera representing the five classically distinguished classes and will likely reveal many more intron paralogues in the future.

Somewhat surprisingly, our intron inventory does not reveal a particular strong affiliation of land plants with Zygnematophyceae algae, currently considered to be the sister lineage of embryophytes despite this class being best sampled for mitogenomes with five genera representing four families. Similarly, mitochondrial group II introns have contributed to a model phylogeny of land plants assuming liverworts as sister to all other embryophytes and hornworts as sister to tracheophytes [8, 103, 104]. Alternative

datasets, however, now favor the concept of monophyletic bryophytes with a sister group relationship of mosses and liverworts [105–108]. Both phylogenies must postulate massive gains and losses of mitochondrial group II introns in the early embryophyte lineages (Fig. 10). Testing the alternative phylogenetic concepts by Maximum Parsimony searches for the 101 mitochondrial group II introns present in embryophytes, we still find weak support for the NLE/HT topology requiring nine steps less than the monophyletic Bryophyte topology (Fig. 10).

Similarities of intron paralogues in the absence of intron-borne maturases

The dynamics of group II introns in plant mitochondrial vs. chloroplast genomes is puzzling at first sight. Key to an explanation for this observation may be the co-evolution of group II introns with concomitant splicing factors and the ancestrally intron-borne maturases certainly play an initial key role. In flowering plants, only one maturase each remains present in the organelle DNAs – “MatK” in the group II intron disrupting the *trnK* gene (mat-trnKi37g2) and “MatR” in the terminal group II intron of the *nad1* gene of mitochondria (mat-nad1i728g2). Accordingly, the lack of evidently recent group II intron mobility in angiosperm organelles, and in fact even spermatophytes, is likely no coincidence. Notably, whereas MatK is also the only chloroplast maturase conserved in the other land plant clades, many more mitochondrial maturases exist in early land plant lineages (Fig. 4 and Fig. 5), which have likely contributed to intron retrotranspositions during establishment of the ancestral plant lineages. Additionally, four nuclear-encoded maturases (Fig. 1C) have been shown to functionally affect diverse mitochondrial, but not chloroplast, group II introns in *Arabidopsis* [15–20].

In some cases, it appears indeed suggestive that a maturase-containing intron paralogue is the likely source of the other paralogues in a family. Key examples are the large intron families F01 and F03 where five out of 13 and four out of eight paralogues, respectively, carry maturases at least in taxa representing early branching lineages (Fig. 4). Other examples are family F07 with its likely ancestral “mother intron” atp9i87g2, family F10 with cobi824g2 or nad1i669g2, family F13 with cox1i178g2 and family F14 with cox1i511g2 carrying a maturase in the alga *Coleochaete scutata*.

Vice versa, however, no traces of maturases are discernible in many other families (Fig. 1F) and even despite the existence of related paralogues in the same plant lineages such as the “liverwort-lycophyte” family F05, the “liverwort-lycophyte” family F08, the “moss-hornwort” family F11, the “hornwort” family F16 or the large and diverse family F02 (Fig. 4). Other striking examples include family F06 where it is highly likely that atp1i361g2 has arisen from rps3i249g2 exclusively in ferns, family F09 where the same holds true for rps1i25g2 likely originating from rpl2i846g2 and family F04 dominated by hornwort paralogues including recently emerged twintron arrangements (Fig. 4). The latter represents the most intriguing case of 98% sequence similarity between the maturase-free paralogues atp1i1050g2ii1536g2 and cox2i98g2, indicating a very recent origin of cox2i98g2 in the genus *Anthoceros* originating from the internal intron of the *atp1* twintron present in all hornwort mitogenomes.

Similarly, the lack of intron-borne maturases holds true for most of the solitary, S-type introns lacking any evident similarity to other introns (Fig. 5). Only 14 of 61 S-type introns show up in more than one plant

clade and any evidence for an external origin via HGT is missing. Like the above cases of maturase-less, but clearly related, paralogues their origin from maturase-bearing ancestors would necessitate a surprisingly quick degeneration of intron-borne reading frames that is incompatible with the overall slow sequence evolution in plant mitogenomes, which reveal the retention of pseudogene traces including former maturases in many other cases [14, 109, 110]. The numerous examples of group II intron fossils reported here are a further case in point (Fig. 1G). After an early discovery of a degenerate group II intron in *Chlamydomonas* [111] this issue may have received less attention than deserved in the exploration of newly determined organelle genomes.

The participation of external protein factors rather than only intron-borne maturases promoting retrotransposition, possibly including nuclear-encoded maturases in the early plant lineages [112], seems a more likely explanation for early group II intron propagations in plant mitochondria (Fig. 1E).

Nuclear splicing factors acting on paralogues in the same family and an outlook

Quite interestingly, we found that in several cases nuclear-encoded splicing factors have already been identified (Fig. 1E), which simultaneously act on related intron paralogues that we have now grouped into families. The nuclear-encoded splicing factor ODB1 (“Organelle DNA Binding 1”) containing a RAD52-like domain has been found to promote splicing of the two angiosperm intron paralogues in family F07, nad1i477g2 and nad2i156g2 [64]. Likewise, the two F09 introns present in angiosperms, ccmFci829g2 and rpl2i846g2, depend on WTF9, a splicing factor containing a PORR (Plant Organelle RNA Recognition) domain [66]. In family F01, even three different intron paralogues present in angiosperms (nad5i230g2, nad7i209g2 and rps3i74g2) rely on splicing factor OZ2, a RanBP2-type zinc finger protein [50]. Finally, the two F17 introns present in angiosperms (nad1i728g2 and nad4i461g2 have also been shown to both depend on three different splicing factors: the two pentatricopeptide repeat proteins DEK55 [70] and EMP8 [69] and the mTERF-type (Mitochondrial Transcription Termination Factor) protein SMK3 [71]. It will be highly interesting to see whether the splicing factor functionalities demonstrated in the model angiosperm species *Arabidopsis thaliana* or *Zea mays* are conserved deep in flowering plant evolution or even beyond angiosperms once adequate model taxa are established among gymnosperms, ferns and lycophytes.

Possibly even more interesting for the future developments will be whether these factors, aside from their evident function in facilitating proper RNA structures for forward-splicing, may also have contributed to the ancient retrotranspositions creating the related paralogues. The rapidly growing amount of data on introns in plant organelle genomes and on the functional characterization of nuclear splicing factors calls makes review compilations [46, 47] quickly outdated and rather calls for expert database systems in the future. Towards that end, we will soon establish a relational database system for plant group organelle introns (‘G12Base’) and their functionally characterized splicing factors (‘SpliciFacts’). In the long run, such an endeavor should clearly not remain restricted to group II introns in streptophyte mitogenomes but will certainly be extended in multiple ways. Firstly, it will certainly consider deeper phylogenetic branches in chlorophytes and beyond, already now justified for example by the tremendous variability of introns

already identified in Ulvophyceae, including a likely mitochondrion-to-chloroplast intron transfer [113] or the examples of intron mobility in red algae [87]. Along the same lines, and in spite their striking conservation of insertion sites among land plants (and despite their lack of primary sequence conservation, which precludes any family assignments compared to their mitochondrial counterparts), chloroplast introns will not go unnoticed. Chloroplast intron dynamics on deeper taxonomic levels may be highly interesting. The gain of introns discovered early to define an early split among algae is but one interesting example [114], the proliferation of group II introns in the algal genus *Oedocladium* (Chlorophyceae) another [115].

Finally, and although group I introns rely on entirely different mechanisms of splicing and propagation and are rare in flowering plants, the example of the “rampant invader” group I intron cox1i726g1 in angiosperms of likely fungal origin [116–118] and the generally much more frequent occurrence of group I introns in early branching green (and non-green) lineages clearly warrants the future database inclusion of reports on their occurrence, invasion of new loci and likely mechanisms of splicing, too. In any case, we consider the diversity of organelle introns in the plant mitochondrial lineage a highly interesting evolutionary playground clearly worth of further investigation.

Material And Methods

Primary nucleotide sequence data sampling

Streptophyte group II intron sequences were extracted from GenBank accessions available at the National Center for Biotechnology Information (see Table 1 for the primary selection). The sampling also included the recently determined mitogenome of the leptosporangiate fern *Haplopteris ensiformis*, which revealed introns that are not present in the two available mitogenomes of eusporangiate ferns [119]. Introns were labeled using the standard nomenclature proposal indicating the respective gene and the homologous nucleotide position in the *Marchantia polymorpha* mitogenome reference [56, 57]. Introns were manually re-checked for conserved sequence motifs at their terminal 5’- and 3’-splicing sites and corrections were applied where necessary (see examples under results). Mitogenome accessions were also checked for potentially overlooked group II introns by taking advantage of searching for their characteristic and mostly conserved domains V of 34 bp [58]. Individual introns were used as queries in sensitive BLASTN searches to identify significant similarities also in intergenic sequences (IGS) of mitogenomes to identify cases of “intron fossils” (see main text).

Table 1

List of Species-data used for different analyses List of taxa and mitogenomes for detailed initial scoring of mitochondrial group II intron presence. The respective numbers of group II introns are indicated. Taxon selection was aimed to maximize phylogenetic diversity with the respective clades. Care was taken to extract intron sequences with proper 5'- and 3'-termini. Asterisks indicate species not further considered for detailed group II intron sequence analyses owing to redundancy with other species within the respective group. A plus sign (+) in the IGS analysis column indicated careful inspection of intergenic sequences for the presence of intron fossils

Major lineage		Species	Acc. number	Mt- group II introns
ALG	Coleochaetophyceae	*Chaetosphaeridium globosum*	NC_004118.1	2 +
	Coleochaetophyceae	*Coleochaete scutata*	NC_045180.1	26 +
	Mesostigmatophyceae	*Mesostigma viride*	AF353999	3
	Chlorokybophyceae	*Chlorokybus atmophyticus*	EF463011.1	14 +
	Charophyceae	*Nitella hyalina*	NC_017598	8
	Charophyceae	*Chara vulgaris*	NC_005255.1	13 +
	Klebsormidiophyceae	*Klebsormidium flaccidum*	KP165386	4
	Zygnematophyceae	*Closterium baillyanum*	NC_022860.1	12 +
	Zygnematophyceae	*Roya obtusa*	NC_22863.1	2
	Zygnematophyceae	*Entransia fimbriata*	NC_22861.1	2
	Zygnematophyceae	*Zygnema circumcarinatum*	MT040698.1	10 +
	Zygnematophyceae	*Gonatozygon brebissonii*	NC_046951.1	6
LIV	Haplomitriopsida	*Treubia lacunosa*	NC_016122.1	21 +
	Marchantiopsida	*Marchantia polymorpha*	MK202951.1	25 +
	Jungermanniopsida	*Calypogeia fissa*	NC_035980	19 +
	Jungermanniopsida	*Pleurozia purpurea*	NC_013444	22
MOS	Sphagnopsida	*Flatbergium novo-caledoniae**	KU725492	26 +
	Sphagnopsida	*Sphagnum palustre*	NC_024521.1	26 +
	Bryopsida	*Physcomitrium patens*	NC_007945.1	24 +

Major lineage		Species	Acc. number	Mt- group II introns
	Bryopsida	*Anomodon rugelii*	NC_016121	24
	Bryopsida	*Ulota crispa*	NC_031393	24
HOR	Leiosporocerotopsida	*Leiosporoceros dussii*	NC_039751.1	34 +
	Anthocerotopsida	*Anthoceros agrestis*	NC_049004.1	39 +
	Anthocerotopsida	*Nothoceros aenigmaticus*	EU660574	28 +
	Anthocerotopsida	*Phaeoceros laevis*	NC_013765	31
LYC	Lycopodiopsida	*Phlegmariurus squarrosus*	NC_017755.1	37 +
	Lycopodiopsida	*Isoetes engelmannii*	FJ390841.1, FJ176330.1, FJ010859.1, FJ628360.1	23
	Lycopodiopsida	*Selaginella moellendorffii*	JF338143.1 – JF338147.1	33
FER	Polypodiopsida	*Ophioglossum californicum*	NC_030900.1	20
	Polypodiopsida	*Psilotum nudum*	NC_030952 & KX171639.1	24 +
	Polypodiopsida	*Haplopteris ensiformis*	OM867545-OM867553	24 +
GYM	Ginkgoopsida	*Ginkgo biloba*	NC_027976.1	25 +
	Cycadopsida	*Cycas taitungensis*	NC_010303	25 +
	Gnetopsida	*Welwitschia mirabilis*	NC_029130	10
ANG	Magnoliopsida	*Amborella trichopoda*	KF754799-KF754803	24
	Magnoliopsida	*Nelumbo nucifera*	NC_030753	23 +
	Magnoliopsida	*Zea mays*	NC_007982.1	22 +
	Magnoliopsida	*Arabidopsis thaliana*	NC_037304.1	23 +

Maturase sequence sampling

Mitochondrial and nuclear encoded maturases were collected independently of the primary group II intron sequences, extending an earlier sampling [112] with sequences from streptophyte algae, also including evident maturases that have not been annotated in sequence entries (see Suppl. Figure 2). The conservation and disintegration of the respective maturases was evaluated with respect to subdomains RT0 - RT7, including signature peptide motifs [120] as well as the conservation of the X-domain and signature peptide motifs for DNA-endonucleases [121]. The ultimate set of streptophyte mitochondrial maturases was used for BLASTP searches for homologs with significant similarities also outside of the plant mitochondrial lineage in other eukaryotic lineages: Metazoa, Fungi, Stramenopiles, Oomycetes, Cryptophytes, Rhodophytes and Chlorophytes. The top ten significant hits for each query were retained after removal of duplicates for alignment and phylogenetic tree construction (see Suppl. Figure 3).

The “most-distant ortholog” concept to define core primary group II intron families

Starting from occasional observations that streptophyte mitochondrial group II intron paralogs in different insertion sites may share more sequence similarities than orthologs in phylogenetically distant clades, we developed a “most-distant-ortholog”-concept (MDO) to cluster the collection of 161 introns (781 representatives in the initial collection of organelle genomes, see Table 1) into families of related paralogs. The full collection of group II intron sequences was initially compared against itself using the stand-alone BLAST + tool [122]. Sensitive BLASTN search parameters were set as follows: word size of 7, match and mismatch values of 2 and − 3 and penalty values of 5 and 2 for gap opening and extension and a random expectancy cutoff value of 1e-5, respectively. Results were evaluated using the RStudio package “dplyr” (tidyverse library) [123] of the. Initial clustering following the MDO criterion used the BLASTN search bit score values for most distant orthologs as cut-off to consider paralog hits with higher bit scores for inclusion into a given family.

Refinement of the group II intron family concepts

Intron homologs present at evidently identical insertion sites that failed to be included in the initial orthologous clustering (“false negatives”) owing to large, “unbridged” phylogenetic distances owing to (multiple) losses (e.g., rps3i74g2 present in charophyceaen algae, hornworts and tracheophytes) were included among the respective family sampling if not showing higher similarity to members of another family. Taxonomically isolated introns receiving only low bit scores of less than 200 were manually re-inspected and only considered for family inclusion when independent sequence similarities were identified in multiple intron regions and/or clearly responsible for conserved RNA secondary structures (see next paragraph). *Vice versa*, we wished to exclude false positive family assignments for paralogs with arbitrarily high bit scores owing to taxon-specific sequence evolution e.g., due to independently arising repeat motifs.

Group II intron alignments, phylogenetic analyses and RNA structure modeling

The clustering of intron orthologs and related paralogs into families was manually re-evaluated and re-checked by careful inspection of the respective full alignments. Intron sequences were automatically

aligned with MAFFT [124] followed by manual adjustment where necessary. Sequence conservation was visualized with GeneDoc 2.7 (https://genedoc.software.informer.com/). Alignment positions with 60% site coverage identified in MEGA 7.0 [125] were extracted for further phylogenetic analyses and modelling of RNA-secondary structures. The filtered group II intron family alignments were used for Maximum Likelihood phylogenetic analyses with IQ-TREE 2 [126]. Substitution model testing was done automatically, and node support was evaluated with 1,000 ultra-fast bootstrap and SH-aLRT repetitions, respectively. Phylogenetic Trees were visualized and edited with TreeGraph2 [127]. RNA secondary structures were modelled manually following established group II intron structure modeling routines [76] with predictions for substructure generated by the mxfold2 application [128]. Nucleotide positions from group II intron family alignments with 60% site coverage sites were preferably used for folding compared to non-conserved sites. RNA secondary structures were annotated in Dot-Bracket annotation and visualized with VARNA [129].

Evaluation of streptophyte mitochondrial group II intron presence and absence

Presence and absence of the full collection of streptophyte mitochondrial group II intron paralogues were coded as ‘1’ and ‘0’, respectively. The R4.1.2 packages “dplyr” and “eulerr” were used for evaluation and plotting of group II intron distributions in the major land plant lineages to create Venn/Euler diagrams [130, 131]. Intron distribution data was converted into the FASTA format to model gain and loss scenarios empirically (gain/loss costs of 4 and 1, respectively) for group II introns for different plant phylogeny scenarios with GLOOME [132].

Declarations

Consent of participation

Not applicable

Consent of publication

Not applicable

Availability of data and materials

Nucleotide sequence data including accessions used in this study are listed in manuscript Table 1. Nucleotide sequence data is available at the National Center for Biotechnology Information (NCBI) cited 2022 September 26. Available from https://www.ncbi.nlm.nih.gov/nucleotide/.

Competing interests

The authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper. The results/data/figures in this manuscript have not been published elsewhere, nor are they under consideration (from you or one of your Contributing Authors) by another publisher.

Funding

No funding

Author contribution

SZ established bioinformatic pipelines and collected data. VK and SZ analyzed data and prepared figures. VK wrote the main manuscript text and all authors reviewed the manuscript.

Acknowledgements

We are grateful for inspiring discussions with members of the IZMB Bonn Molecular Evolution group and to Dietrich Dmitriev for comments on the manuscript.

References

1. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101:1040–56.
2. Li-Pook-Than J, Bonen L. Multiple physical forms of excised group II intron RNAs in wheat mitochondria. Nucleic Acids Res. 2006;34:2782–90.
3. Ngu M, Massel K, Bonen L. Group II introns in wheat mitochondria have degenerate structural features and varied splicing pathways. Int J Biochem Cell Biol. 2017;91:156–67.
4. Massel K, Silke JR, Bonen L. Multiple splicing pathways of group II trans-splicing introns in wheat mitochondria. Mitochondrion. 2016. https://doi.org/10.1016/j.mito.2016.03.002.
5. Malek O, Knoop V. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA. 1998;4:1599–609.
6. Qiu YL, Palmer JD. Many independent origins of trans splicing of a plant mitochondrial group II intron. J Mol Evol. 2004;59:80–9.
7. Guo W, Zhu A, Fan W, Adams RP, Mower JP. Extensive shifts from cis to trans splicing of gymnosperm mitochondrial introns. Mol Biol Evol. 2020. https://doi.org/10.1093/molbev/msaa029.
8. Groth-Malonek M, Pruchner D, Grewe F, Knoop V. Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol. 2005;22:117–25.

9. Knoop V, Altwasser M, Brennicke A. A tripartite group II intron in mitochondria of an angiosperm plant. Mol Gen Genet MGG. 1997;255:269–76.
10. Kück U, Schmitt O. The Chloroplast Trans -Splicing RNA–Protein Supercomplex from the Green Alga Chlamydomonas reinhardtii. Cells (MDPI). 2021;10:290.
11. Yu R, Sun C, Liu Y, Zhou R. Shifts from *cis*- to *trans*-splicing of five mitochondrial introns in *Tolypanthus maclurei*. PeerJ. 2021;9:e12260.
12. Kan S-L, Shen T-T, Gong P, Ran J-H, Wang X-Q. The complete mitochondrial genome of Taxus cuspidata (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. BMC Evol Biol. 2020;20:10.
13. Zumkeller SM, Gerke P, Knoop V. A functional twintron, "zombie" twintrons and a hypermobile group II intron invading itself in plant mitochondria. Nucleic Acids Res. 2020;48:2661–75.
14. Ahlert D, Piepenburg K, Kudla J, Bock R. Evolutionary origin of a plant mitochondrial group II intron from a reverse transcriptase/maturase-encoding ancestor. J Plant Res. 2006;119:363–71.
15. Keren I, Tal L, Des Francs-Small CC, Araújo WL, Shevtsov S, Shaya F, et al. NMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex i assembly and function. Plant J. 2012;71:413–26.
16. Keren I, Bezawork-Geleta A, Kolton M, Maayan I, Belausov E, Levy M, et al. AtnMat2, a nuclear-encoded maturase required for splicing of group II introns in Arabidopsis mitochondria. RNA. 2009;15:2299–311.
17. Shevtsov-Tal S, Best C, Matan R, Aldrin Chandran S, Brown GG, Ostersetzer-Biran O. nMAT3 is an essential maturase splicing factor required for holo-complex I biogenesis and embryo-development in *Arabidopsis thaliana* plants. Plant J. 2021;:tpj.15225.
18. Cohen S, Zmudjak M, Colas Des Francs-Small C, Malik S, Shaya F, Keren I, et al. NMAT4, A maturase factor required for nad1 pre-mRNA processing and maturation, is essential for holocomplex i biogenesis in Arabidopsis mitochondria. Plant J. 2014;78:253–68.
19. Schmitz-Linneweber C, Lampe M-KK, Sultan LD, Ostersetzer-Biran O. Organellar maturases: A window into the evolution of the spliceosome. Biochim Biophys Acta. 2015;1847:798–808.
20. Mohr G, Lambowitz AM. Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucleic Acids Res. 2003;31:647–52.
21. Mower JP. Variation in protein gene and intron content among land plant mitogenomes. Mitochondrion. 2020;53:203–13.
22. Pruchner D, Nassal B, Schindler M, Knoop V. Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol Genet Genomics. 2001;266:608–13.
23. Pruchner D, Beckert S, Muhle H, Knoop V. Divergent intron conservation in the mitochondrial nad2 gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. J Mol Evol. 2002;55:265–71.

24. Knie N, Grewe F, Knoop V. Monilophyte mitochondrial rps1 genes carry a unique group II intron that likely originated from an ancient paralog in rpl2. RNA. 2016;22:1338–48.
25. Zumkeller SM, Knoop V, Knie N. Convergent Evolution of Fern-Specific Mitochondrial Group II Intron atp1i361g2 and Its Ancient Source Paralogue rps3i249g2 and Independent Losses of Intron and RNA Editing among Pteridaceae. Genome Biol Evol. 2016;8:2505–19.
26. Mukhopadhyay J, Hausner G. Organellar introns in fungi, algae and plants. Cells (MDPI). 2021;10:1–35.
27. Lambowitz AMAM, Zimmerly S. Mobile group II introns. Annu Rev Genet. 2004;38:1–35.
28. Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. Biotechnological applications of mobile group II introns and their reverse transcriptases: Gene targeting, RNA-seq, and non-coding RNA analysis. Mob DNA. 2014;5:1–19.
29. Robart AR, Zimmerly S. Group II intron retroelements: function and diversity. Cytogenet Res. 2005;110:589–97.
30. Bonen L, Vogel J. The ins and outs of group II introns. Trends Genet. 2001;17:322–31.
31. McNeil BA, Semper C, Zimmerly S. Group II introns: versatile ribozymes and retroelements. Wiley Interdiscip Rev RNA. 2016;7:341–55.
32. Zimmerly S, Semper C. Evolution of group II introns. Mob DNA. 2015;6:7.
33. Lambowitz AM, Zimmerly S. Group II Introns: Mobile Ribozymes that Invade DNA. Cold Spring Harb Perspect Biol. 2011;3:a003616–6.
34. Lambowitz AM, Belfort M. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. Microbiol Spectr. 2015;3.
35. Pyle AM. Group II Intron Self-Splicing. Annu Rev Biophys. 2016;45:183–205.
36. Pyle AM, Fedorova O, Waldsich C. Folding of group II introns: a model system for large, multidomain RNAs? Trends Biochem Sci. 2007;32:138–45.
37. Patel S, Fedorova O. Ribozyme Structural Elements: Group II Introns. In: Reference Module in Life Sciences. Elsevier; 2021.
38. Fedorova O, Zingler N. Group II introns: structure, folding and splicing mechanism. Biol Chem. 2007;388:665–78.
39. Peters JK, Toor N. Group II intron lariat: Structural insights into the spliceosome. RNA Biol. 2015;12:913–7.
40. Chalamcharla VR, Curcio MJ, Belfort M. Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. Genes Dev. 2010;24:827–36.
41. Smathers CM, Robart AR. The mechanism of splicing as told by group II introns: Ancestors of the spliceosome. Biochim Biophys Acta - Gene Regul Mech. 2019. https://doi.org/10.1016/j.bbagrm.2019.06.001.
42. Novikova O, Belfort M. Mobile Group II Introns as Ancestral Eukaryotic Elements. Trends Genet. 2017;33:773–83.

43. Costa M, Fontaine JM, Loiseaux-De Goër S, Michel F. A group II self-splicing intron from the brown alga Pylaiella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. J Mol Biol. 1997;274:353–64.
44. Barthet MM, Pierpont CL, Tavernier EK. Unraveling the role of the enigmatic MatK maturase in chloroplast group IIA intron excision. Plant Direct. 2020;4:e00208.
45. Inaba-Hasegawa K, Ohmura A, Nomura M, Sugiura M. Development of an In Vitro Chloroplast Splicing System: Sequences Required for Correct pre-mRNA Splicing. Plant Cell Physiol. 2021;62:1311–20.
46. Brown GG, Colas des Francs-Small C, Ostersetzer-Biran O. Group II intron splicing factors in plant mitochondria. Front Plant Sci. 2014;5 February:35.
47. de Longevialle AF, Small ID, Lurin C. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. MolPlant. 2010;3:691–705.
48. Qian J, Li M, Zheng M, Hsu Y-F. Arabidopsis SSB1, a Mitochondrial Single-Stranded DNA-Binding Protein, is Involved in ABA Response And Mitochondrial RNA Splicing. Plant Cell Physiol. 2021. https://doi.org/10.1093/PCP/PCAB097.
49. Cao S-K, Liu R, Sayyed A, Sun F, Song R, Wang X, et al. Regulator of Chromosome Condensation 1-Domain Protein DEK47 Functions on the Intron Splicing of Mitochondrial Nad2 and Seed Development in Maize. Front Plant Sci. 2021;0:1612.
50. Bentolila S, Gipson AB, Kehl AJ, Hamm LN, Hayes ML, Mulligan RM, et al. A RanBP2-type zinc finger protein functions in intron splicing in Arabidopsis mitochondria and is involved in the biogenesis of respiratory complex I St ephane. Nucleic Acids Res. 2021;:1–17.
51. Lee K, Leister D, Kleine T. Arabidopsis Mitochondrial Transcription Termination Factor mTERF2 Promotes Splicing of Group IIB Introns. Cells. 2021;10:315.
52. Yang Y-Z, Ding S, Wang Y, Wang H-C, Liu X-Y, Sun F, et al. PPR20 Is Required for the cis-Splicing of Mitochondrial nad2 Intron 3 and Seed Development in Maize. Plant Cell Physiol. 2020;61:370–80.
53. Ohyama K, Oda K, Ohta E, Takemura M. Gene organization and evolution of introns of a liverwort, Marchantia polymorpha, mitochondrial genome. In: Brennicke A, Kück U, editors. Weinheim: VCH Verlagsgesellschaft; 1993. pp. 115–29.
54. Lippok B, Brennicke A, Wissinger B. Differential RNA editing in closely related introns in *Oenothera* mitochondria. Mol Gen Genet. 1994;243:39–46.
55. Vangerow S, Teerkorn T, Knoop V. Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. Plant Biol. 1999;1:235–43.
56. Dombrovska E, Qiu Y-L. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol. 2004;32:246–63.
57. Knoop V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 2004;46:123–39.

58. Knoop V, Kloska S, Brennicke A. On the identification of group II introns in nucleotide sequence data. J Mol Biol. 1994;242:389–96.
59. Marienfeld JR, Newton KJ. The Maize Ncs2 Abnormal Growth Mutant Has A Chimeric Nad4-Nad7 Mitochondrial Gene and Is Associated with Reduced Complex-I Function. Genetics. 1994;138:855–63.
60. Knoop V, Ehrhardt T, Lättig K, Brennicke A. The gene for ribosomal protein S10 is present in mitochondria of pea and potato but absent from those of Arabidopsis and Oenothera. Curr Genet. 1995;27:559–64.
61. Zumkeller SM, Polsakiewicz M, Knoop V. Rickettsial DNA invasions and a scrambled rRNA cluster with a trans-splicing group I intron: The highly unorthodox mitogenome of the fern Haplopteris ensiformis. Submitted. 2022.
62. Volkmar U, Groth-Malonek M, Heinrichs J, Muhle H, Polsakiewicz M, Knoop V. Exclusive conservation of mitochondrial group II intron nad4i548 among liverworts and its use for phylogenetic studies in this ancient plant clade. Plant Biol (Stuttg). 2012;14:382–91.
63. Groth-Malonek M, Rein T, Wilson R, Groth H, Heinrichs J, Knoop V. Different fates of two mitochondrial gene spacers in early land plant evolution. Int J Plant Sci. 2007;168:709–17.
64. Gualberto JM, Le Ret M, Beator B, Kühn K. The RAD52-like protein ODB1 is required for the efficient excision of two mitochondrial introns spliced via first-step hydrolysis. Nucleic Acids Res. 2015;43:6500–10.
65. Gerke P, Szövényi P, Neubauer A, Lenz H, Gutmann B, McDowell R, et al. Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort Anthoceros agrestis. New Phytol. 2020;225:1974–92.
66. Francs-Small CC des, Kroeger T, Zmudjak M, Ostersetzer-Biran O, Rahimi N, Small I, et al. A PORR domain protein required for rpl2 and ccmF(C) intron splicing and for the biogenesis of c-type cytochromes in Arabidopsis mitochondria. Plant J Cell Mol Biol. 2012;69:996–1005.
67. Turmel M, Otis C, Lemieux C. Complete mitogenome of the streptophyte green alga *Coleochaete scutata* (Coleochaetophyceae). Mitochondrial DNA Part B Resour. 2019;4:4209–10.
68. Grewe F, Zhu A, Mower JP. Loss of a *Trans*-Splicing *nad1* Intron from Geraniaceae and Transfer of the Maturase Gene *matR* to the Nucleus in *Pelargonium*. Genome Biol Evol. 2016;8:3193–201.
69. Sun F, Zhang X, Shen Y, Wang H, Liu R, Wang X, et al. The pentatricopeptide repeat protein EMPTY PERICARP8 is required for the splicing of three mitochondrial introns and seed development in maize. Plant J. 2018. https://doi.org/10.1111/tpj.14030.
70. Ren RC, Yan XW, Zhao YJ, Wei YM, Lu X, Zang J, et al. The novel E-subgroup pentatricopeptide repeat protein DEK55 is responsible for RNA editing at multiple sites and for the splicing of nad1 and nad4 in maize. BMC Plant Biol. 2020;20:1–15.
71. Pan Z, Ren X, Zhao H, Liu L, Tan Z, Qiu F. A mitochondrial transcription termination factor, ZmSmk3, is required for nad1 Intron4 and nad4 Intron1 splicing and kernel development in maize. G3 Genes, Genomes. Genet. 2019;9:2677–86.

72. Bégu D, Araya A. The horsetail Equisetum arvense</i > mitochondria share two group I introns with the liverwort Marchantia, acquired a novel group II intron but lost intron-encoded ORFs. Curr Genet. 2009;55:69–79.
73. Ferat JL, Michel F. Group II self-splicing introns in bacteria. Nature. 1993;364:358–61.
74. Ferat JL, Legouar M, Michel F. Multiple Group-Ii Self-Splicing Introns in Mobile Dna from Escherichia-Coli. Comptes Rendus I Acad des Sci Ser Iii-Sciences la Vie-Life Sci. 1994;317:141–8.
75. Knoop V, Brennicke A. Evidence for a group II intron in Escherichia coli inserted into a highly conserved reading frame associated with mobile DNA sequences. Nucleic Acids Res. 1994;22:1167–71.
76. Candales MA, Duong A, Hood KS, Li T, Neufeld RAE, Sun R, et al. Database for bacterial group II introns. Nucleic Acids Res. 2012;40.
77. Dai L, Zimmerly S. The dispersal of five group II introns among natural populations of Escherichia coli. RNA. 2002;8:1294–307.
78. Ferat J-L, Le Gouar M, Michel F. A group II intron has invaded the genus Azotobacter and is inserted within the termination codon of the essential groEL gene. Mol Microbiol. 2003;49:1407–23.
79. Simon DM, Clarke NAC, McNeil BA, Johnson I, Pantuso D, Dai L, et al. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA. 2008;14:1704–13.
80. Toor N, Zimmerly S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA. 2002;8:1373–7.
81. Dai L, Zimmerly S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. Nucleic Acids Res. 2002;30:1091–102.
82. Martínez-Abarca F, Toro N. Group II introns in the bacterial world. Mol Microbiol. 2000;38:917–26.
83. Leclercq S, Giraud I, Cordaux R. Remarkable Abundance and Evolution of Mobile Group II Introns in Wolbachia Bacterial Endosymbionts. Mol Biol Evol. 2011;28:685–97.
84. Toro N, Martínez-Abarca F, Molina-Sánchez MD, García-Rodríguez FM, Nisa-Martínez R. Contribution of Mobile Group II Introns to Sinorhizobium meliloti Genome Evolution. Front Microbiol. 2018;9:627.
85. Thompson MD, Copertino DW, Thompson E, Favreau MR, Hallick RB. Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus Euglena. Nucleic Acids Res. 1995;23:4745–52.
86. Ikuta K, Kawai H, Müller DG, Ohama T. Recurrent invasion of mitochondrial group II introns in specimens of Pylaiella littoralis (brown alga), collected worldwide. Curr Genet. 2008;53:207–16.
87. Perrineau M-M, Price DC, Mohr G, Bhattacharya D. Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. PeerJ. 2015;3:e1017.
88. Kim D, Lee JM, Cho CH, Kim EJ, Bhattacharya D, Yoon HS. Group II intron and repeat-rich red algal mitochondrial genomes demonstrate the dynamic recent history of autocatalytic RNAs. BMC Biol. 2022;20:1–16.

89. Khan H, Archibald JM. Lateral transfer of introns in the cryptophyte plastid genome. Nucleic Acids Res. 2008;36:3043–53.
90. Pogoda CS, Keepers KG, Hamsher SE, Stepanek JG, Kane NC, Kociolek JP. Comparative analysis of the mitochondrial genomes of six newly sequenced diatoms reveals group II introns in the barcoding region of cox1. Mitochondrial DNA Part A, DNA mapping. Seq Anal. 2019;30:43–51.
91. Guillory WX, Onyshchenko A, Ruck EC, Parks M, Nakov T, Wickett NJ, et al. Recurrent Loss, Horizontal Transfer, and the Obscure Origins of Mitochondrial Introns in Diatoms (Bacillariophyta). Genome Biol Evol. 2018;10:1504–15.
92. Burger G, Saint-Louis D, Gray MW, Lang BF. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: Cyanobacterial introns and shared ancestry of red and green algae. Plant Cell. 1999;11:1675–94.
93. Odom OW, Shenkenberg DL, Garcia JA, Herrin DL. A horizontally acquired group II intron in the chloroplast psbA gene of a psychrophilic Chlamydomonas: In vitro self-splicing and genetic evidence for maturase activity. RNA. 2004;10:1097–107.
94. Sheveleva EV, Hallick RB. Recent horizontal intron transfer to a chloroplast genome. Nucleic Acids Res. 2004;32:803–10.
95. Ehara M, Watanabe KI, Ohama T. Distribution of cognates of group II introns detected in mitochondrial cox1 genes of a diatom and a haptophyte. Gene. 2000;256:157–67.
96. Kamikawa R, Masuda I, Demura M, Oyama K, Yoshimatsu S, Kawachi M, et al. Mitochondrial Group II Introns in the Raphidophycean Flagellate Chattonella spp. suggests a Diatom-to-Chattonella Lateral Group II Intron Transfer. Protist. 2009;160:364–75.
97. Vallès Y, Halanych KM, Boore JL. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS ONE. 2008;3:e1488.
98. Kwon E-C, Kim J-H, Kim N-S. Comprehensive genomic analyses with 115 plastomes from algae to seed plants: structure, gene contents, GC contents, and introns. Genes Genomics. 2020;42:553–70.
99. Lemieux C, Otis C, Turmel M. Comparative Chloroplast Genome Analyses of Streptophyte Green Algae Uncover Major Structural Alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. Front Plant Sci. 2016;7:697.
100. Teich R, Grauvogel C, Petersen J. Intron distribution in Plantae: 500 million years of stasis during land plant evolution. Gene. 2007;394:96–104.
101. Bonen L. Cis- and trans-splicing of group II introns in plant mitochondria. Mitochondrion. 2008;8:26–34.
102. Orton LM, Fitzek E, Fitzek E, Feng X, Grayburn WS, Mower JP, et al. Zygnema circumcarinatum UTEX 1559 chloroplast and mitochondrial genomes provide insight into land plant evolution. J Exp Bot. 2020;71:3361–73.
103. Qiu YL, Cho Y, Cox JC, Palmer JD. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature. 1998;394:671–4.

104. Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, et al. The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci U S A. 2006;103:15511–6.
105. Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, et al. The interrelationships of land plants and the nature of the ancestral embryophyte. Curr Biol. 2018;28:733–45.
106. Renzaglia KS, Villarreal Aguilar JC, Garbary DJ. Morphology supports the setaphyte hypothesis: mosses plus liverworts form a natural group. Bryophyt Divers Evol. 2018;40:11–7.
107. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, et al. The timescale of early land plant evolution. Proc Natl Acad Sci U S A. 2018;115:E2274–83.
108. Su D, Yang L, Shi X, Ma X, Zhou X, Hedges SB, et al. Large-Scale Phylogenomic Analyses Reveal the Monophyly of Bryophytes and Neoproterozoic Origin of Land Plants. Mol Biol Evol. 2021;38:3332–44.
109. Steinhauser S, Malek O, Schock I, Beckert S, Knoop V. Group II introns in plant mitochondria: Trans-splicing and transgenic approaches. In: Moller IM, Gardeström P, Glimelius K, Glaser E, editors. Leiden: Backhuys Publishers; 1998. pp. 201–6.
110. Groth-Malonek M, Wahrmund U, Polsakiewicz M, Knoop V. Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial *nad7* Gene Supports *Haplomitrium* as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae. Mol Biol Evol. 2007;24:1068–74.
111. Nedelcu AM, Lee RW. A degenerate group II intron in the intronless mitochondrial genome of Chlamydomonas reinhardtii: Evolutionary implications. Mol Biol Evol. 1998;15:918–22.
112. Guo W, Mower JP. Evolution of plant mitochondrial intron-encoded maturases: frequent lineage-specific loss and recurrent intracellular transfer to the nucleus. J Mol Evol. 2013;77:43–54.
113. Turmel M, Otis C, Lemieux C. Mitochondrion-to-Chloroplast DNA. Transfers and Intragenomic Proliferation of Chloroplast Group II Introns in Gloeotilopsis Green Algae (Ulotrichales, Ulvophyceae). Genome Biol Evol. 2016;8:2789–805.
114. Manhart JR, Palmer JD. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. Nature. 1990;345:268–70.
115. Brouard J-S, Turmel M, Otis C, Lemieux C. Proliferation of group II introns in the chloroplast genome of the green alga Oedocladium carolinianum (Chlorophyceae). PeerJ. 2016;4:e2627.
116. Cho Y, Palmer JD. Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. Mol Biol Evol. 1999;16:1155–65.
117. Cusimano N, Zhang LBL-B, Renner SS. Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. Mol Biol Evol. 2008;25:265–76.
118. Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of Peperomia. J Mol Evol. 1995;41:563–72.

119. Guo W, Zhu A, Fan W, Mower JP. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. New Phytol. 2017;213:391–403.
120. Blocker FJH, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. RNA. 2005;11:14–28.
121. San Filippo J, Lambowitz AM. Characterization of the C-Terminal DNA-binding/DNA Endonuclease Region of a Group II Intron-encoded Protein. J Mol Biol. 2002;324:933–51.
122. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
123. Hadley Wickham R, François L, Henry, Kirill Müller. A Grammar of Data Manipulation [R package dplyr version 1.0.0]. Media. 2020.
124. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
125. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.
126. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4.
127. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 2010;11 SUPPL. 6.
128. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12:1–9.
129. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009;25:1974–5.
130. Larsson J, Gustafsson P, {eulerr}: Area-Proportional Euler and Venn Diagrams with Ellipses. Proceedings of {{International Workshop}} on {{Set Visualization}} and {{Reasoning}}. 2018;2116:84–91.
131. Wickman H, François R, Henry L, Muller K. dlpyr: A Grammar of Data Manipulation. CRAN Repository. 2021;:1–2.
132. Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. Bioinformatics. 2010;26:2914–5.
133. Marchetti F, Cainzos M, Shevtsov S, Córdoba JP, Sultan LD, Brennicke A, et al. Mitochondrial Pentatricopeptide Repeat Protein, EMB2794, Plays a Pivotal Role in NADH Dehydrogenase Subunit nad2 mRNA Maturation in Arabidopsis thaliana. Plant Cell Physiol. 2020;61:1080–94.

Figures

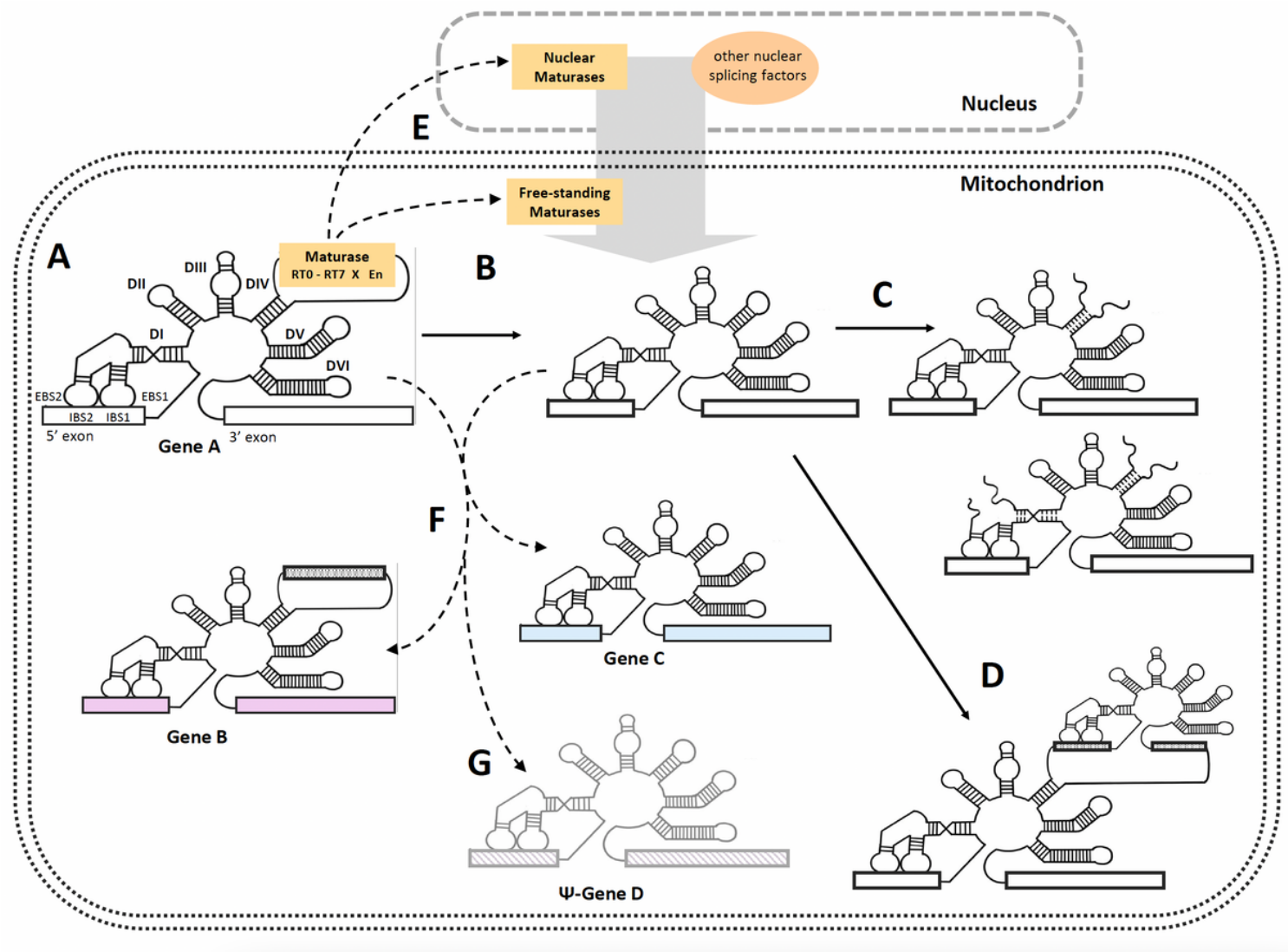


Figure 1

Plant mitochondrial group II introns evolving

A. In its likely evolutionary ancestral state, a group II intron encodes a multifunctional maturase in domain IV of its characteristic six-domain structure (DI through DVI). A fully equipped maturase features reverse transcriptase domains RT0-RT7 (the finger and palm domains), followed by a maturase-specific 'X' domain and a DNA endonuclease 'En' domain with relevance for intron mobility. Interaction of exon binding sites (EBS 1 and 2) in domain DI with corresponding intron binding sites at the end of the upstream exon are equally important for splicing and retromobility **B**. Intron-borne maturases frequently degenerate and get lost during plant mitogenome evolution. **C**. Multiple cases of *cis*-to-*trans*-splicing transitions are evident for plant mitochondrial group II introns, creating bipartite introns mostly broken in domain DIV or even tripartite group II introns [9]. **D**. Plant mitochondrial group II introns may invade other introns, creating twintrons [13]. **E**. The contribution of group II intron-encoded maturases, their free-standing paralogues in plant mitochondrial or nuclear genomes and those of multiple phylogenetically unrelated splicing factors [18, 19, 46, 133] to all the evolutionary processes summarized in the figure is not fully understood at present. **F**. Of particular relevance in this survey is the capability of plant

mitochondrial group II introns to occupy new gene locations, creating sequence-related group II paralogues, here grouped into families. **G**. The complete survey of plant mitochondrial group II introns presented includes paralogues that degenerated together with their host genes, or which can be identified as “fossil” introns in intergenic locations. Group II intron illustration have been made manually with ProCreate®.

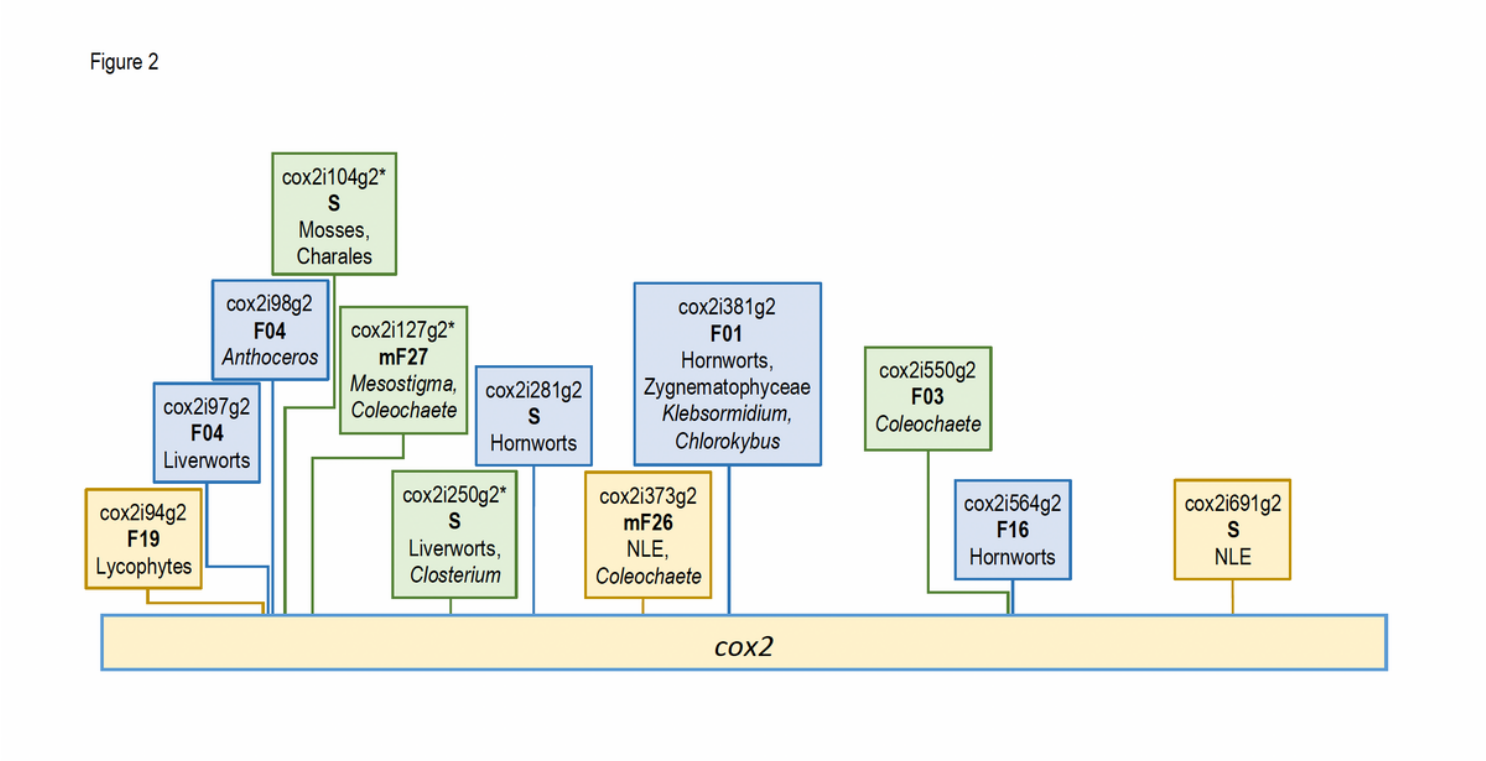


Figure 2

The *cox2* gene.

The example of the *cox2* gene illustrates important issues of streptophyte mitochondrial group II intron diversity. Altogether twelve different group II insertion sites are presently identified in different plant lineages and streptophyte algae with colors distinguishing introns in bryophytes (blue), those shared with or present only in algae (green) and those shared with or present only in vascular plants (yellow). The latter include cox2i373g2 and cox2i691g labeled with NLE for “Non-Liverwort Embryophytes” given their presence in all land plant lineages except the liverworts. Asterisks indicate cases of introns occupying the same position in algae and land plants that were considered orthologues despite lack of significant sequence similarity.

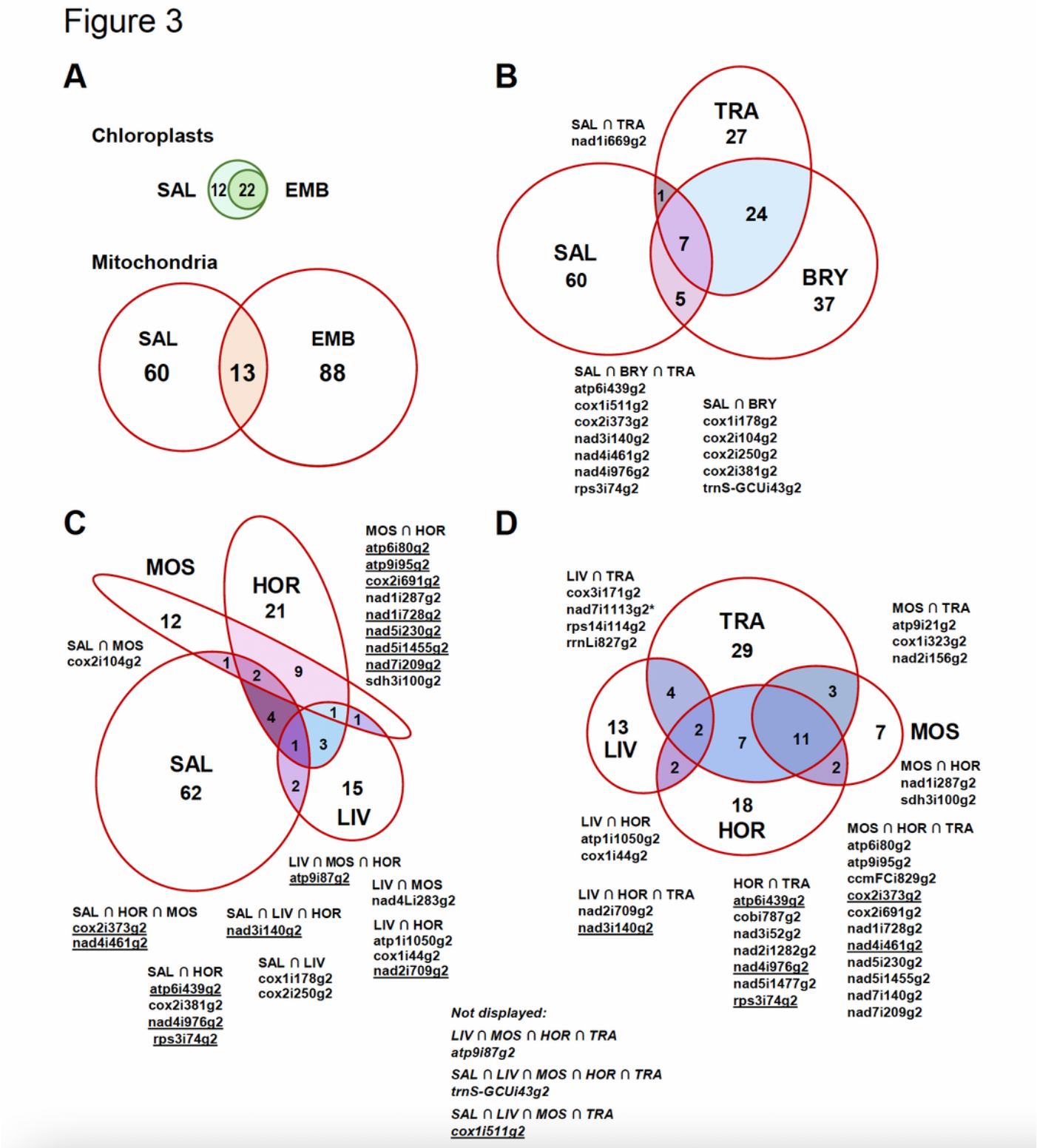


Figure 3

Euler diagrams of streptophyte mitochondrial group II intron distributions

Euler diagrams displaying the occurrence of organelle group II introns. **A**. 22 chloroplast group II introns (green circles) of embryophytes (EMB) are a highly conserved subset of 34 homologues in streptophyte algae (SAL) whereas of 161 mitochondrial group II introns in streptophyte mitogenomes (red circles) only

13 are shared between streptophyte algae and land plants. **B**. Mitochondrial group II intron paralogues present in the specific intersections are listed for Streptophyte algae (SAL), Tracheophytes (TRA) and Bryophytes (BRY) and, in panels **C** and **D**, more specifically for the three bryophyte clades of Liverworts (LIV), Mosses (MOS) and Hornworts (HOR). Underlined in panel C are introns also occurring in tracheophytes and, *vice versa*, in panel D those also present in streptophyte algae. An asterisk indicates that the status of intron nad7i1113g2 is unclear owing to pseudogene degeneration. Highlighted in red in panel C is intron atp9i87g2 as the only intron shared between the three bryophyte classes, which is not displayed in panel D owing to limitations of Euler displays. Likewise, intron cox1i511g2, highlighted below in italics, is not displayed in panels C and D.

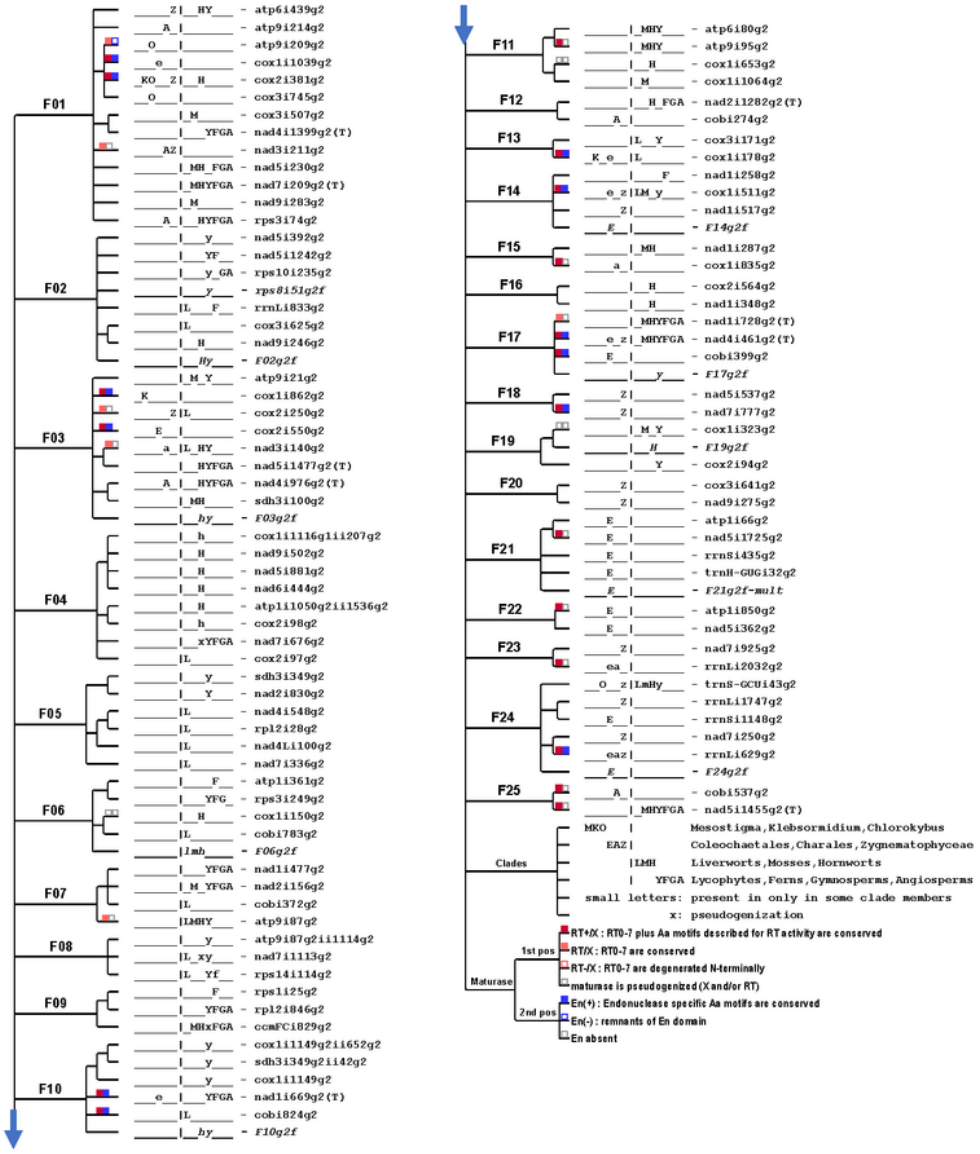


Figure 4

Core families F01 - F25 of streptophyte mitochondrial group II introns.

Cladogram of streptophyte mitochondrial group II introns categorized into core families F01-F10 (left panel) and F11-F25 (right panel) as described in the text. Occurrence of introns in the different streptophyte clades is given with a clade code before the respective intron label with underscores

indicating absence in a given clade as shown in the cladogram legend (bottom right). Small letters indicate cases of possibly late intron gain within a given clade and ‘x’ indicates pseudogenization or complete loss of the respective host gene, respectively. Presence of maturases is indicated with square symbols distinguishing variable degrees of conservation of the maturase RT and X domains and the En domains as shown in the legend (bottom right). Evidently defect intron fossils are labeled with ”g2f” and indicated in italics, the identification of such fossil intron inserts in intergenic sequences of mitogenomes is indicated with the respective family number.

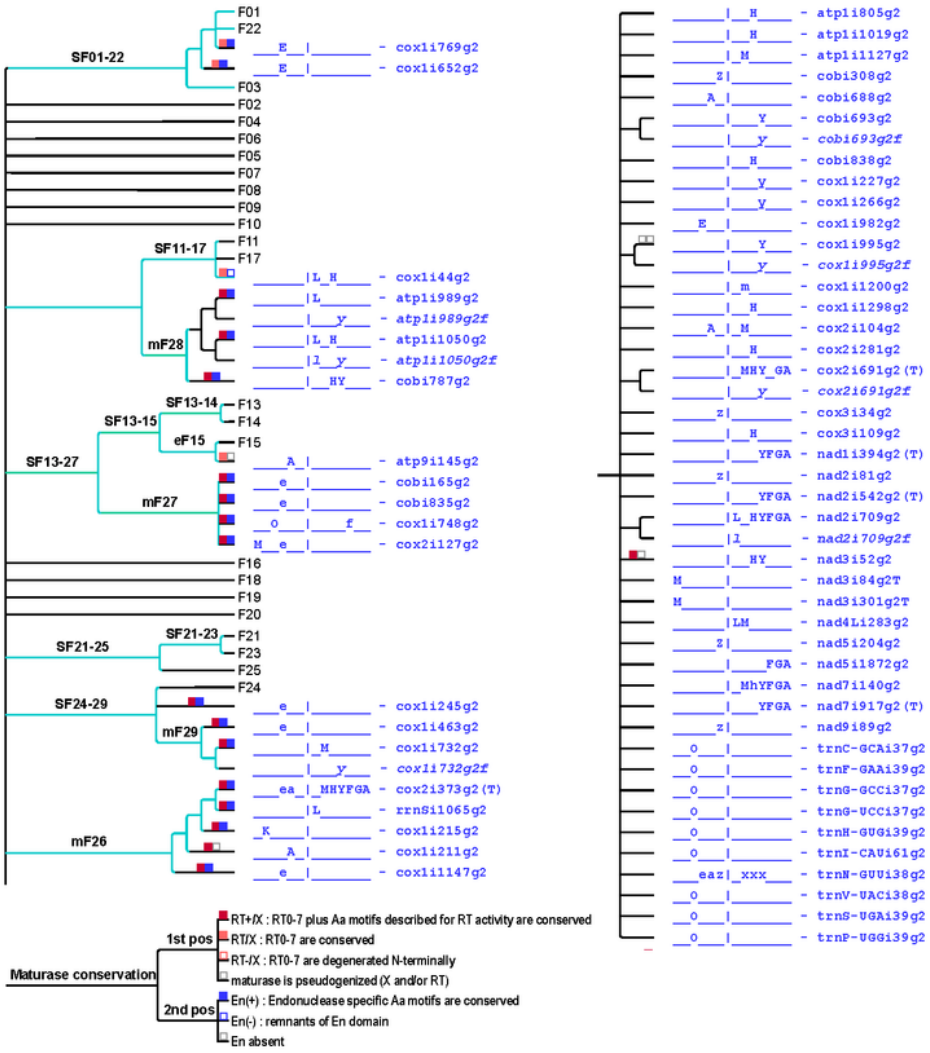


Figure 5

Group II intron superfamily cladogram and solitary group II introns.

“Solitary” plant mitochondrial group II introns having no significant nucleotide similarities to other paralogues that could warrant inclusion in one of the core group II intron families F01 - F25 (Fig. 4) are indicated in blue font. Independent phylogenetic analyses of intron-borne maturase protein sequences resulted in identification of “maturase-based” intron families mF26 - m29 and superfamilies (SF) extending the core families (light blue branch lines, left panel). Symbols for maturase conservation and clade codes are as in figure 4.

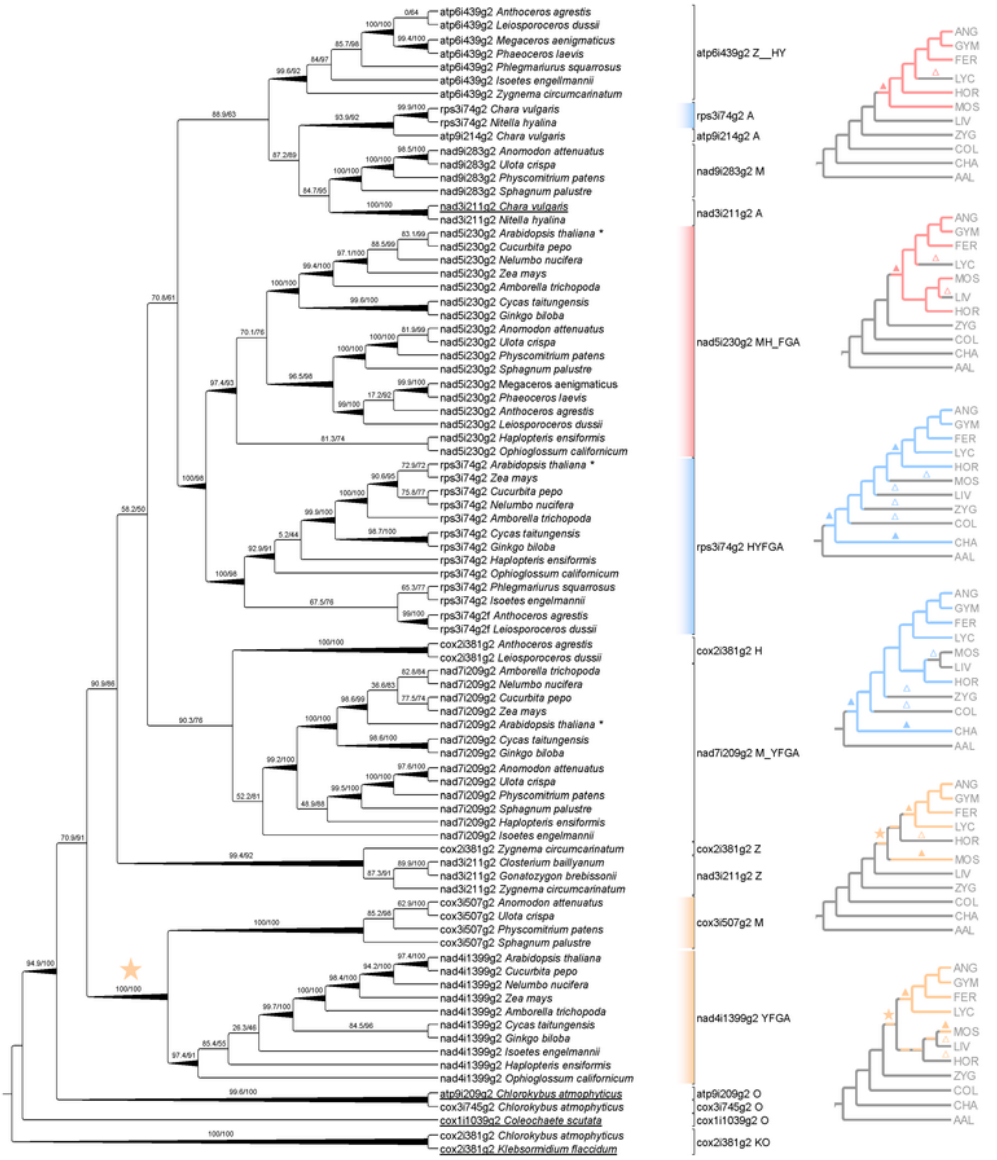


Figure 6

Family F01 of streptophyte mitochondrial group II introns.

Detailed phylogeny of group II intron paralogues in family F01. The tree display is rooted arbitrarily with F01 intron paralogs cox2i381g2 in the early-branching streptophyte algae *Klebsormidium flaccidum* and *Chlorokybus atmophyticus* (ML-based phylogeny obtained with IQ-Tree, based on the intron alignment

positions with minimally 60 % coverage and GTR+F+R4 identified as best-fitting substitution model). Node support is given both for ultrafast bootstrapping and the SH-aLR test before and after the slash, respectively, and internal branches are highlighted as rectangles where both values exceed 90%. Underlining indicates introns carrying a maturase reading frame. No conflicts between reliably determined nodes within the paralogue subclades and known species phylogenies are observed. Capital letters behind intron names indicate presence in mosses (M), hornworts (H), lycophytes (Y), ferns (F), gymnosperms (G) and angiosperms (A). The tree suggests two closely related paralog pairs of embryophyte introns cox3i507g2 and nad4i1399g2 (light orange label) and of nad5i230g2 (red label) and rps3i74g2 (blue label). Asterisks indicate three intron in *Arabidopsis thaliana*, for which OZ2 has been identified as a common splicing factor [50]. Simplified cladograms on the right show alternative gain/loss (filled/empty triangles) scenarios for debated “Non-Liverwort-Embryophyte” (NLE) and “monophyletic Bryophytes” (mB) topologies for nad5i230 (top) and rps3i74g2 (middle). The nature of their ancestor (orange start symbol) is unclear given the phylogenetically disjunct distribution of cox3i507g2 and nad4i1399g2, but the mB topology shown would postulate an additional intron loss.

Figure 7

Family F02 of streptophyte mitochondrial group II introns.

A. Detailed phylogeny of group II intron paralogues in family F02. Phylogenetic resolution for F02 paralogues is lower than in the case of F01 introns (Fig. 6). For simplicity, only a simple ML consensus tree for nodes >50% obtained with MEGA7.0 is shown. Paralogues cox3i625g2 in liverworts and nad9i246g2 in hornworts and introns nad5i1242g2 and rrnLi833g2 in ferns, respectively, emerge as closely related paralogues **B.** Presence of nad5i1242g2 also in ferns and of rps10i235g2 also in seed plants leaves the series of intron copying events including creation of two intron fossils in *rps8* and in the *tatC-cox2*spacer unclear except for the likely late emergence of nad5i392g2 from nad5i1242g2 only in Lycopodiales. **C.** Both F02 intron fossils in *Phlegmariurus squarrosus* have characteristic sequence conservation at the 3’ intron end including terminal domains V and VI (indicated by dot-bracket annotation) compared to the functional paralogs, but only rps8i52g2f features additional sequence conservation also at the 5’ intron end. **D.** A F02 group II intron paralog, likely related to nad9i246g2, created multiple intron fossils (F02g2f) in several intergenic regions in hornwort mitogenomes. These intron fossils are widely conserved within the intergenic regions *rpl10..trnH-GUG* and *atp6..nad6* but more restricted for another fossil between pseudogenes for *rps8* and *rpl6* in *Megaceros aenigmaticus*.

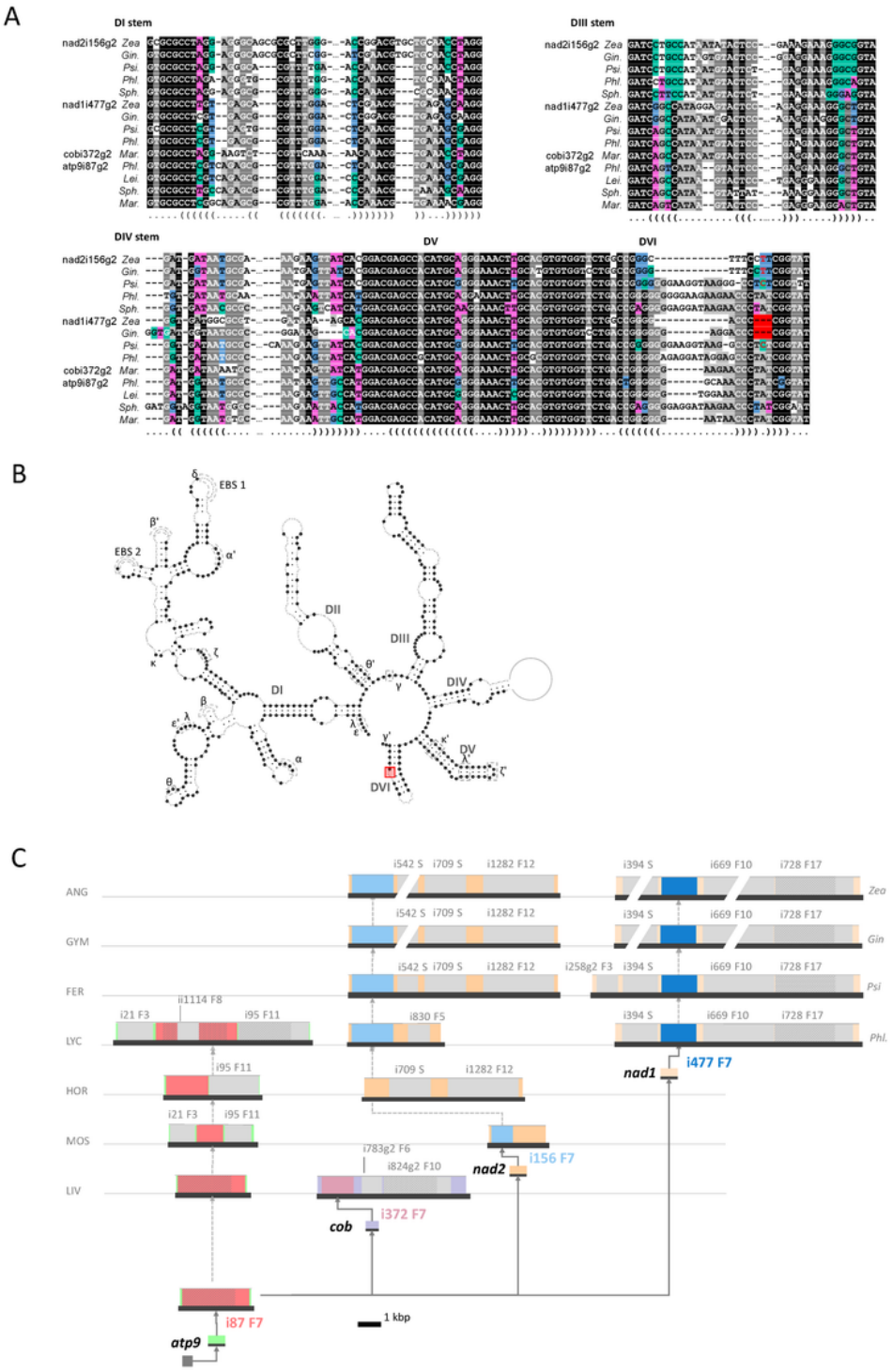


Figure 8

Family F07 of streptophyte mitochondrial group II introns.

Group II intron paralogs atp9i87g2, cobi372g2, nad2i156g2 and nad1i477g2 in family F07 feature structurally relevant sequence conservations across the entire intron sequences. **A.** Exemplary alignments of domain DI and DIII regions and complete domains DV and VI together of selected species *Zea mays*

(Zea), *Ginkgo biloba* (Gin), *Psilotum nudum* (Psi), *Phlegmariurus squarrosus* (Phl) *Sphagnum palustre* (Sph), *Marchantia polymorpha* (Mar), *Leiosporoceros dussii* (Lei) with dot bracket annotations indicating base pairings. Co-conversions that do affect sequence conservation but not RNA secondary structures are colored based on base-pairing (A-T in pink, G-C in green; G-T in blue). Red boxes and shading indicate degeneration of the conserved bulged A in domain VI, usually involved in lariat formation during splicing reaction. **B**. Secondary structure of F07-type group II introns based on overall conserved regions of the paralogs. Roman numerals indicate the six typically conserved domains of group II introns (DI-DVI) with EBS 1 & 2 in DI and Greek letters designate tertiary interactions sites, respectively [32]. Black shading of nucleotides shown in panel A corresponds to equally shaded nucleotides in the secondary structure model. **C.** Intron atp9i87g2 (red box) is the only group II intron conserved in all three bryophyte lineages and additionally present in lycophytes. It encodes maturase mat-atp9i87g2c in liverworts (pattern fill) but this is absent in the moss and hornwort orthologues. In the lycophyte *P. squarrosus* mat-atp9i87g2 is interrupted by internal intron ii1114g2. Ancestral intron atp9i87g2 likely gave rise to cobi372g2 (pink box) in liverworts (LIV), to nad2i156g2 (light blue box) in mosses (MOS) and to nad1i477g2 (blue box) in the tracheophyte lineage, first arising in lycophytes (LYC). Additional group II introns (grey) and their family assignments are indicated. Transitions from *cis*- to *trans*-splicing in seed plants are indicated with diagonal interruptions.

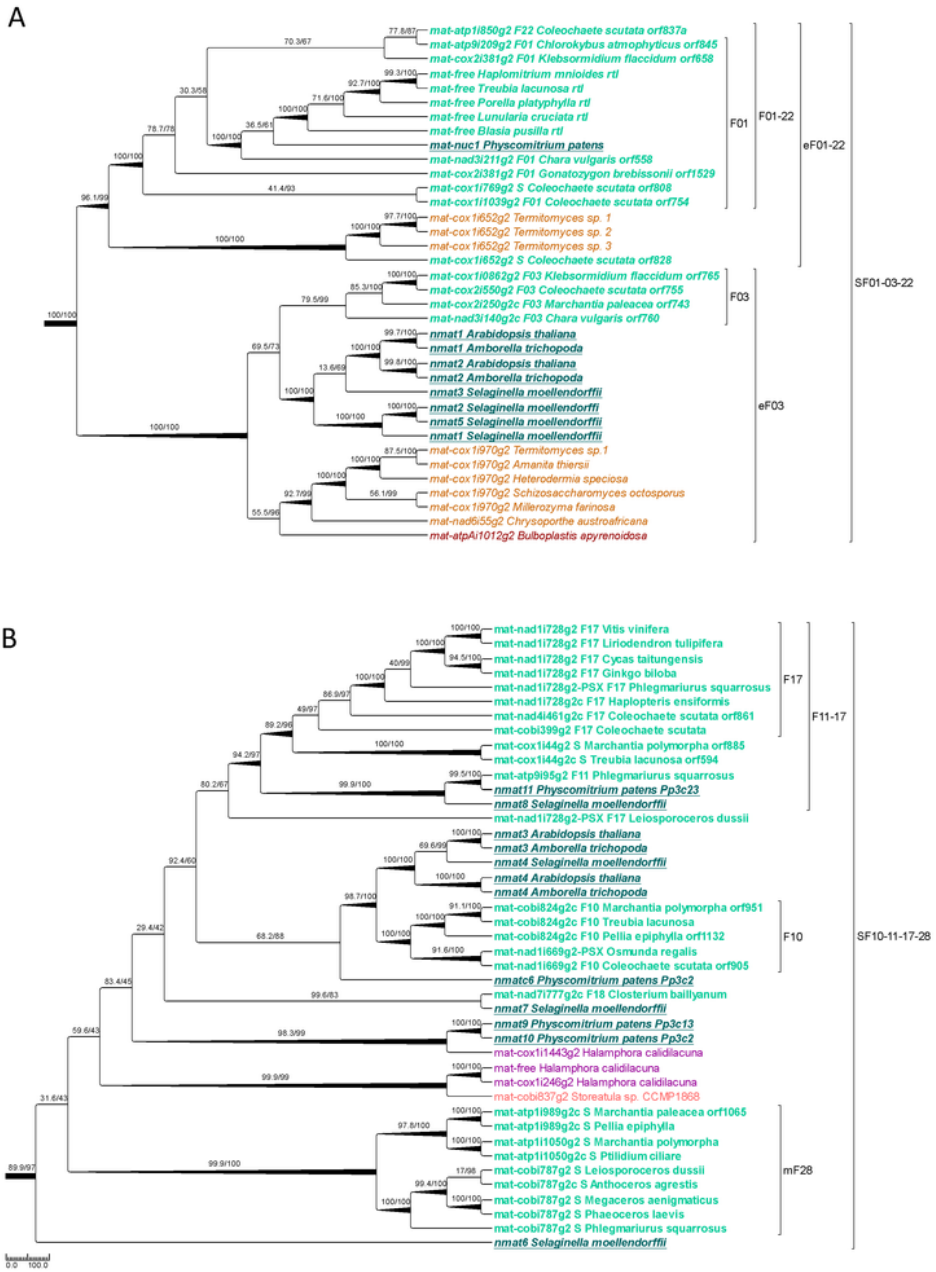


Figure 9

Group II intron superfamilies defined by maturase similarities.

Two selected subclades of the comprehensive maturase phylogeny shown in supplementary figure 2 for group II intron superfamilies SF01-03-22 (**A**) and SF10-11-17-28 (**B**). The eukaryote maturase phylogeny contains samples from nuclear encoded maturases of embryophytes (dark green with underline),

streptophyte mitochondria (green and bold), fungal mitochondria (orange), metazoan mitochondria (blue) and organelles of chlorophytes (light green), stramenopiles (purple), cryptophytes (beige), rhodophytes (red) and oomycetes (black). **A**. SF01-03-22 contains solitary-type streptophyte mitochondrial maturase *mat-cox1i769g2* and *mat-cox1i652g2*, which forms a joint clade with orthologs from fungal *Termitomyces* mitochondria. The free-standing mitochondrial maturases of liverworts ("mat-free") cluster with the nuclear encoded mat-nuc1 of *P. patens* and with maturases of algal F01 paralogs while no F01 intron paralogues are present in liverwort mitogenomes. Extended family eF03 supports a close relation of F03 paralogs and the well-characterized nuclear-encoded maturases nmat1 and nmat2 of angiosperms together with nuclear maturase paralogs in the lycophyte *S. moellendorffii* and, more distantly, with mitochondrial maturase paralogs in fungi and a chloroplast paralogue in a rhodophyte. **B.** Superfamily SF10-11-17-28 contains F17 *mat-nad1i728g2(*or *matR),* the only mitochondrial encoded maturase in euphyllophytes together with the maturases in the S-type intron cox1i44g2 and *mat-atp9i95g2*. Angiosperm nuclear encoded nmat3 and nmat4 form a joint clade with maturases of F10 paralogs including the pseudogenized (PSX) *mat-nad1i699g2* in the fern *Osmunda regalis* indicating the ancestral presence of a maturase in this intron paralogue conserved also in angiosperms. Maturases of S-type introns *mat-atp1i1050g2* and *mat-atp1i989g2* form a well-supported "maturase family" clade mF28 with *mat-cobi787g2* as sister group.

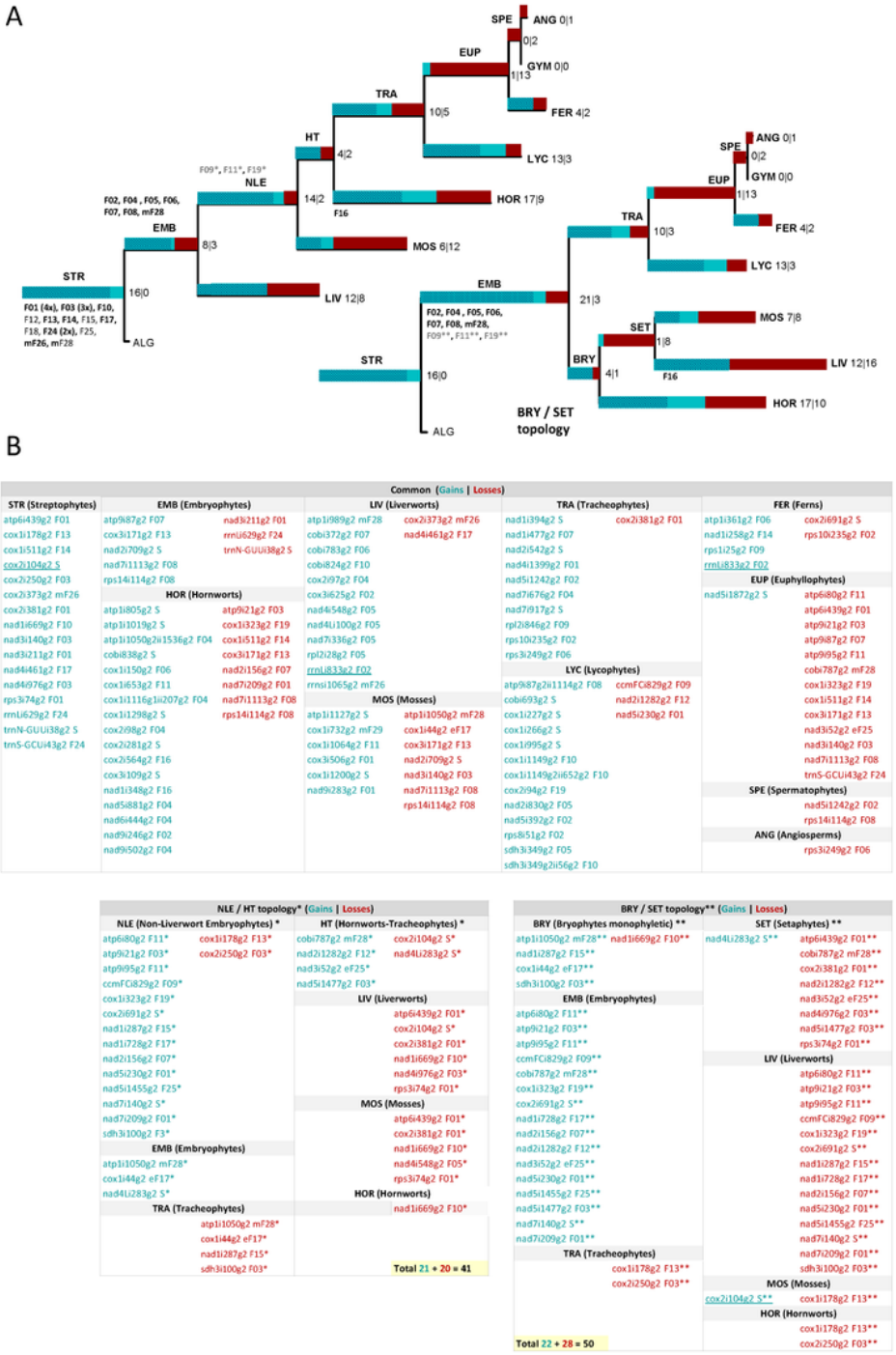


Figure 10

Possible gain-loss-scenarios for plant mitochondrial group II introns.

A. Cladograms for alternative phylogenies “NLE/HT” (Non-Liverwort Embryophyte and Hornwort-Tracheophyte clades, left) and alternative “BRY/SET” (monophyletic Bryophyte and Setaphyte clades, right). Further labels indicate Streptophytes (STR), Embryophytes (EMB), Liverworts (LIV), Mosses (MOS),

Hornworts (HOR), Lycophytes (LYC), Tracheophytes (TRA), Euphyllophytes (EUP), Ferns (FER), Gymnosperms (GYM), Angiosperms (ANG), Spermatophytes (SPE), and Streptophyte algae (ALG). Bars indicate plant mitochondrial group II intron gains (blue) and losses (red) after search for maximum parsimony evolution for the distribution of 101 embryophyte mitochondrial group II introns assuming weights of 4 to 1 for gains over losses. Underlining of rrnLi833g2 and cox1i104g2 highlights likely independent gain events as discussed in the text. Numbers before and after the pipes indicate intron gain and loss events. Hatched blue areas indicate introns grouped into families and the origin of the respective families are indicated on the respective branches. **B.** Detailed listing for gains and losses of the embryophyte group II introns with those common to both evolutionary scenarios on top and those that need to be specifically assumed for the NLE/HT scenario (asterisks, total 102 gain and 190 loss events) shown below to the left and those for the BRY/SET scenario (double asterisks, total 103 gain and 199 loss events) shown below to the right.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1mtg2familiesSZ20220926.pdf
- Supplementaryfigure2mtg2ZumkellerKnoop2022.pdf

7.1 Supporting material

Supplementary material is provided on a USB flash drive for

the printed copies and as a ZIP archive in the online version of this thesis.

Supplementary_Table_1_Maturase_sub-domain_analyses.pdf

7.1.1 Supplementary material Publication 1

Suppl-Tab-S1-Primer-Zumkeller-etal-2019.xlsx

Zumkeller-etal-Suppl-Figs-S1-S8.pdf

7.1.2 Supplementary material Publication 2

SupplFigure1ZumkelleretalHaplopterismtDNANC20220919 nal2.pdf

SupplFigure2ZumkelleretalHaplopterismtDNANC20220919 nal2.pdf

SupplFigure3ZumkelleretalHaplopterismtDNANC20220919 nal2.pdf

SupplFigure4ZumkelleretalHaplopterismtDNANC20220919 nal2.pdf

SupplTab1ZumkelleretalCPEditing nalp1.pdf

ZumkelleretalSupplTab2MTEditingcomplete.pdf

SupplTab3Zumkelleretal2022insertsfeaturesSZ20220727.pdf

7.1.2 Supplementary material Publication 3

SupplementaryFigure1mtg2familiesSZ20220926.pdf

Supplementaryfigure2mtg2ZumkellerKnoop2022.pdf