Molecular Phylogenetics and Evolution xxx (xxxx) 107236

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

In and out: Evolution of viral sequences in the mitochondrial genomes of legumes (Fabaceae)

In-Su Choi [a, b, *], Martin F. Wojciechowski [b], Tracey A. Ruhlman [a], Robert K. Jansen [a, c]

[a] Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA
[b] School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
[c] Centre of Excellence in Bionanoscience Research, Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Comparative genomics
Group II intron
Horizontal gene transfer
Intracellular gene transfer
Target primed reverse transcription

ABSTRACT

Plant specific mitoviruses in the 'genus' *Mitovirus* (*Narnaviridae*) and their integrated sequences (non-retroviral endogenous RNA viral elements or NERVEs) have been recently identified in various plant lineages. However, the sparse phylogenetic coverage of complete plant mitochondrial genome (mitogenome) sequences and the non-conserved nature of mitochondrial intergenic regions have hindered comparative studies on mitovirus NERVEs in plants. In this study, 10 new mitogenomes were sequenced from legumes (Fabaceae). Based on comparative genomic analysis of 27 total mitogenomes, we identified mitovirus NERVEs and transposable elements across the family. All legume mitogenomes included NERVEs and total NERVE length varied from ca. 2 kb in the papilionoid *Trifolium* to 35 kb in the mimosoid *Acacia*. Most of the NERVE integration sites were in highly variable intergenic regions, however, some were positioned in six *cis*-spliced mitochondrial introns. In the *Acacia* mitogenome, there were L1-like transposons including an almost full-length copy with target site duplications (TSDs). The integration sites of NERVEs in four introns showed evidence of L1-like retrotransposition events. Phylogenetic analysis revealed that there were multiple instances of precise deletion of NERVEs between TSDs. This study provides clear evidence that a L1-like retrotransposition mechanism has a long history of contributing to the integration of viral RNA into plant mitogenomes while microhomology-mediated deletion can restore the integration site.

1. Introduction

Plant mitochondrial genomes (mitogenomes) display a markedly increased size compared to their counterparts in most animals (Knoop, 2004). Except for some extreme cases, mitogenomes of seed plants exhibit a highly variable size ranging from ca. 200 to 1,000 kb (Choi et al., 2019). Regardless of their size fluctuations, most lineages show a decreased number of mitochondrial protein coding genes compared to ancestral angiosperms (Richardson et al., 2013) due to frequent losses of ribosomal proteins and succinate dehydrogenase genes (Adams and Palmer, 2003). Mitogenome expansion can be import-driven [e.g., intracellular gene transfer (IGT) and horizontal gene transfer (HGT)] (Bergthorsson et al., 2003; Richardson and Palmer, 2006; Goremykin et al., 2012) and can also result from the duplication of native mitochondrial DNA (mtDNA) (Negruk, 2013). The mutation pattern observed in plant mitogenomes is sometimes characterized as "Jekyll and Hyde" because of the striking contrast in levels of sequence conservation among different genomic regions (Christensen, 2018; Smith, 2020). Synonymous substitution rates in mitochondrial protein coding sequences (CDS) are the lowest among the three genomes of a plant cell (Wolfe et al., 1987; Drouin et al., 2008). In contrast, mitochondrial intergenic regions are regarded as highly unstable in angiosperms based on rapid decline of shared DNAs between taxa (Guo et al., 2016). In Fabaceae only the core sequences (~100 kb), comprising genic regions and intergenic

Abbreviations: CDS, coding sequences; EN, endonuclease; HGT, horizontal gene transfer; IGT, intracellular gene transfer; L1 or Line1, long interspersed elements-1; LTR, long terminal repeat; MITOGENOME, mitochondrial genome; ML, maximum likelihood; MMEJ, microhomology-mediated end-joining; MTDNA, mitochondrial DNA; MTP, mitochondrial transit peptide; NERVE, non-retroviral endogenous RNA viral element; ORF, open reading frame; RDRP, RNA-dependent RNA polymerase; RNH, ribonuclease H; RT, reverse transcriptase; TE, transposable element; TPRT, target primed reverse transcription; TSD, target site duplication; ZF-RVT, zinc-binding in reverse transcriptase

* Corresponding author.
E-mail addresses: 86ischoi@gmail.com (I-S Choi), mfwojciechowski@asu.edu (M.F. Wojciechowski), truhlman@austin.utexas.edu (T.A. Ruhlman), jansen@austin.utexas.edu (R.K. Jansen).

https://doi.org/10.1016/j.ympev.2021.107236
Received 11 February 2021; Received in revised form 11 June 2021; Accepted 14 June 2021

regions in close proximity to genes, can be aligned between highly diverged taxa regardless of their mitogenome size (Choi et al., 2019). This contradictory evolutionary pattern has been explained by a combination of different mitochondrial DNA repair systems (accurate and error-prone) and subsequent selection pressure on mitogenome molecules without deleterious mutations in genic regions (Christensen, 2013, 2014, 2018). A recent study (Wu et al., 2020) based on intraspecific variation of 1,135 individuals of *Arabidopsis thaliana* (Brassicaceae) reinforced this idea. Synonymous substitution rates in CDS were similar to substitution rates in intergenic regions, however in CDS regions numbers of indels and nonsynonymous substitutions were greatly reduced (Wu, et al. 2020).

Members of virus family *Narnaviridae* are simple, unencapsidated RNA viruses (2.3–3.6 kb in length) that encode a single RNA-dependent RNA polymerase (RdRp) required for their replication (Hillman and Cai, 2013). This family includes two genera, *Narnavirus* and *Mitovirus*, that are restricted to different subcellular locations, the cytosol and mitochondria, respectively (Hillman and Esteban, 2011). It was thought that fungi were the exclusive hosts for mitoviruses until Nibert et al. (2018) provided evidence for a monophyletic group of plant specific mitoviruses derived from fungal mitoviruses as early as 430 million years ago. The non-retroviral endogenous RNA viral elements (NERVEs) of mitoviruses have long been identified in various plant mitochondrial and nuclear genomes even before the discovery of contemporary plant mitoviruses (Bruenn et al., 2015). Reverse transcriptase (RT) activity of transposable elements (TEs) has been suggested as a mechanism for mitovirus NERVE integration (Bruenn et al., 2015; Nibert et al., 2018). However, verification of the mechanism is still lacking. The NERVEs of diverse RNA viruses present in eukaryotic genomes are usually shorter than their related viral genomes, potentially the result of post-integration mutational decay (Nibert et al., 2018 and references therein). Mutational decay may partly explain why the integration mechanism has not been clearly characterized. In Fabaceae, a mitovirus NERVE was identified from a group II intron of *rps10* (rps10i235) in the mitogenome of soybean, *Glycine* max (Chang et al., 2013). Subsequent studies of Fabaceae mitogenomes revealed that the NERVE in rps10i235 may have been established early in the diversification of subfamily Papilionoideae and subsequently deleted in association with flanking short direct repeats (Shi et al., 2018; Zhang et al., 2019).

Some group II introns include an open reading frame (ORF) and are considered to have mobility based on their intron-encoded protein (Bonen and Vogel, 2001; Lehmann and Schmidt, 2003). Most angiosperm mitochondrial introns are group II and lack an ORF (Bonen, 2012), therefore they lack autonomous mobility (Zimmerly and Semper, 2015). The only remaining functional group II intronic ORF in angiosperm mitogenomes is *matR* (in nad1i728), which encodes a protein with a truncated RT domain and a conserved domain of intron maturase (Brown et al., 2014). Mobile group II introns are also the putative ancestors of a group of non-long terminal repeat (non-LTR) retrotransposons [e.g., long interspersed elements-1 (L1)] (Zimmerly et al., 1995; Bonen, 2012; Brown et al., 2014; Zimmerly and Semper, 2015). Group II introns and L1s utilize the same retrotransposition mechanism, known as target primed reverse transcription (TPRT), which employs target DNAs as the primer for reverse transcription (Zimmerly et al., 1995). However, the RT domain structure of group II introns is more closely related to viral RdRp than to retroviral RT (Stamos et al., 2017). Hence, the presence of NERVEs (largely comprising RdRp) within an intron of Papilionoideae is very intriguing because it illustrates not only an insertion of foreign sequence but also cryptic structural similarity to group II intron-encoded RT. A similar example may be the twintrons, introns-within-introns in mitogenomes of lycophytes and hornworts (Zumkeller et al., 2020).

Generally TEs are grouped as either class I, retrotransposons, or class II, DNA transposons (Finnegan, 1989). Transposable elements constitute a large proportion of nearly all sequenced nuclear genomes and have high diversity and complexity (Piégu et al., 2015). Autonomous TEs, which encode a protein with RT activity, are able to mediate retrotransposition of their mRNAs as well as other RNA molecules in the cell (Kubiak and Makałowska, 2017). Transposable elements are also present in mitogenomes of angiosperms but as fragmented copies (Mower et al., 2012). Intracellular gene transfers from the mitochondrion to the nucleus are frequent in plant cells and DNA (including complementary DNA, synthesized from mRNA) appears to be the primary material transferred, rather than mRNA (Henze and Martin, 2001; Kleine et al., 2009). In contrast, IGT in the reverse direction (nucleus to mitochondrion) is rare (Kleine et al., 2009; Mower et al., 2012), and there is little evidence to indicate the source of the transferred genetic material (i.e., DNA versus RNA).

Evidence of IGT and interspecific HGT can be readily lost due to the bizarre mutation patterns of the plant mitogenome (Christensen, 2018). At the same time, evidence may be well preserved if the integration occurred proximal to genes. Such proximity may facilitate 'piggybacking' in that signatures of IGT and HGT can pass through the filter of selection and share the benefit of accurate (long homology based) repair (Christensen, 2013, 2014) over evolutionary time. For example, mitochondrial intergenic regions in close proximity to ribosomal protein genes (*rps1, rps4*, *rps7*) include ancient fungal sequences in orchids (Sinn and Barrett, 2020). In legumes, pseudogenized *rps1* proximal to *nad5* exon1 (ca. 200 bp apart) persisted for a long time with only a few mutations (Choi et al., 2020). Mitochondrial intron sequences also have the potential to include molecular fossils of ancestral gene transfer events. Most importantly, introns can be aligned even between highly diverged plants by aligning neighboring exons except in cases of *trans*-splicing introns (Qiu and Palmer, 2004; Guo et al., 2020).

Fabaceae (legumes) includes six subfamilies (Cercidoideae, Detarioideae, Duparquetioideae, Dialioideae, Caesalpinioideae, Papilionoideae), 770 genera and 20,000 species and is an economically and ecologically important plant family (Lewis et al., 2005; Yahara et al., 2013; LPWG, 2017). The family is ideally suited for a comparative study of mitovirus NERVEs in plant mitogenomes given that previous studies have identified differential evolution of the rps10i235 NERVE in Papilionoideae. Furthermore, recent advances in the understanding of legume evolution (Cardoso et al., 2012, 2013; Cannon et al., 2015; Bauchet et al., 2019; Egan and Vatanparast, 2019; Ren et al., 2019; Stai et al., 2019; Koenen et al., 2020a, 2020b; Zhang et al., 2020) provide the framework for interpretations of mitogenome evolution in a broader systematic context. Here, 10 new mitogenomes of Fabaceae were sequenced, which together with 17 published mitogenomes, enabled a family-wide comprehensive analysis of plant mitovirus NERVEs.

2. Materials and Methods

2.1. Genome assembly

Ten species of Fabaceae representing 10 genera in the two subfamilies Caesalpinioideae and Papilionoideae were selected for mitogenome assembly (Table 1). Raw Illumina (San Diego, CA) reads (100 bp paired-end) originally generated in studies by Sabir et al. (Sabir et al., 2014) and Schwarz et al. (Schwarz et al., 2015) were used. *De novo* genome assembly, read mapping and gap closing were conducted in Geneious Prime 2019.2.5 (https://www.geneious.com) using the Geneious assembler, mapper and aligner. The methods described in Choi et al. (2019) were used for mitogenome assembly based on plastome-filtered reads. Mitochondrial contigs were identified by BLAST searches with available Fabaceae mitogenome sequences at NCBI (Table S1) as queries using BLASTN 2.8.0+ with default options (Altschul et al., 1997). Overlapping regions and repeat motifs among mitochondrial contigs were detected by Geneious read mapping (low

Table 1
Mitogenome sequencing statistics for 10 legumes.

Subfamily	Species	Total reads	Plastome-filtered reads	# of contigs	Total assembled length (bp)	Mean coverage	GC (%)	NCBI accessions
Caesalpinioideae	*Ceratonia siliqua*	88,817,956	81,033,800	3	475,642	421	45.3	MW448447 - MW448449
Caesalpinioideae	*Prosopis glandulosa*	68,668,968	64,974,930	6	758,210	315	44.8	MW448450 - MW448455
Papilionoideae	*Arachis hypogaea*	72,013,986	71,184,106	1	592,341	146	44.7	MW448460
Papilionoideae	*Lupinus albus*	61,118,168	53,702,036	1	405,579	288	44.7	MW448461
Papilionoideae	*Indigofera tinctoria*	61,665,674	58,979,300	1	546,392	251	44.7	MW448462
Papilionoideae	*Apios americana*	69,129,894	67,960,448	1	434,145	108	45	MW448463
Papilionoideae	*Pachyrhizus erosus*	97,730,832	87,658,468	4	308,122	492	45.4	MW448456 - MW448459
Papilionoideae	*Vigna unguiculata*	56,094,996	52,562,044	1	383,314	303	45.1	MW448464
Papilionoideae	*Robinia pseudoacacia*	60,118,944	58,360,256	1	396,058	321	45.2	MW448465
Papilionoideae	*Glycyrrhiza glabra*	77,536,200	69,943,206	1	440,064	201	45.2	MW448466

sensitivity) and repeat finder (minimum repeat length: 30 bp; maximum mismatches: 20%). Contigs were manually merged based on detected overlaps and repeat motifs. Remaining gaps were closed by polymerase chain reaction and subsequent Sanger sequencing of amplicons at the University of Texas Genomic Sequencing and Analysis Facility in Austin. Finally, contigs were refined by read mapping of total plastome-filtered reads with custom sensitivity options (maximum gaps: 5%; maximum mismatches: 5%; only map paired reads that match nearby).

2.2. Annotation

In addition to the 10 new mitogenomes (Table 1), 17 previously published Fabaceae mitogenomes were reannotated (Table S1) for comparative analyses. Annotations of Fabaceae mitogenomes were conducted using Geseq (Tillich et al., 2017) based on the mitogenomes of *Psilotum nudum* (KX171638 and KX171639), *Ginkgo biloba* (KM672373) and *Liriodendron tulipifera* (NC_021152) as BLAT reference sequences. The tRNAs were validated by tRNAscan-SE v2.0 (Lowe and Chan, 2016). Annotations were assessed and manually modified in Geneious Prime. All 10 newly assembled mitogenome sequences were deposited in GenBank (MW448447-MW448466).

2.3. Phylogenomic analysis

To infer phylogenetic relationships among the 27 legume taxa based on mitogenome sequences, two other taxa of the rosid nitrogen-fixing clade, *Cucurbita pepo* (Cucurbitaceae) and *Rosa chinensis* (Rosaceae), and two taxa of Salicaceae, *Populus tremula* and *Salix suchowensis*, were included as outgroups (Choi et al., 2019). Twenty-six shared mitochondrial CDS were extracted from the mitogenomes of the 31 taxa. Orthologous sequences from the 31 taxa were aligned with MAFFT v.7.017 (Katoh et al., 2002) using default options. Poorly aligned regions were deleted or manually adjusted. Aligned sequences were concatenated for each of the species. The Akaike information criterion (AIC) was used to select an appropriate nucleotide substitution model in jModelTest v.2.1.6 (Darriba et al., 2012). Maximum likelihood (ML) analyses were conducted using RAxML v.8 (Stamatakis, 2014) using a GTR + I + τ model and 1000 bootstrap replications. jModelTest and ML analyses were performed in the CIPRES Science Gateway (Miller et al., 2010).

2.4. Content analysis

Based on annotations of the 27 legume mitogenomes, the status of mitochondrial protein genes and *cis*-spliced introns was assessed. Full length ORFs lacking out of frame indels and premature stop codons were designated as intact genes and sequences with these features were designated pseudogenes or missing genes. Missing genes were determined by the lack of a recognizable segment of >100 bp of the gene sequence. Lengths of each intron across taxa were estimated and illustrated as box plots using BoxPlotR (Spitzer et al., 2014). The presence or absence of 19 *cis*-spliced introns was determined and accumulative lengths were calculated for each species. The intron nomenclature used by Guo et al. (Guo et al., 2017) was followed, which includes the host gene name followed by 'i' (intron) and the nucleotide position of the intron insertion site, for example "nad1i728".

Mitovirus NERVEs were identified from BLASTX searches using each of the legume mitogenomes as queries and translated RdRp of reported mitoviruses (BK010422–BK010442) in Nibert et al. (2018) as subjects with a word size of 3 and an e-value of $1e^{-4}$. To avoid overestimation by inclusion of large repeats, all but one copy of repeats >1 kb were manually deleted from all mitogenomes as described in Choi et al. (2019). All BLAST hit coordinate information was transferred to each mitogenome as an annotation in Geneious Prime. Overlapping regions between NERVEs were excluded from estimation. The length of each NERVE was estimated for *cis*-intronic regions and other regions. Introns, containing NERVEs, were aligned with orthologous sequences from related genera in their respective subfamilies (Papilionoideae and Caesalpinioideae) with MAFFT (Katoh et al., 2002) using default options. Alignments were manually adjusted to minimize gaps and maximize apparent homologous regions. For the case of rps3i74, *Senna occidentalis* (NC_038221) was compared with its congener *S. tora* (NC_038053).

Putative loci with long terminal repeat (LTR) and non-LTR transposons were initially inspected using the CENSOR webserver (Kohany et al., 2006) with default parameters and Viridiplantae selected as the sequence source. Among the initial results, sequence fragments shorter than 500 bp were excluded. Subsequently, ±4 kb regions of TEs were used as queries for BLASTX using non-redundant protein sequences in NCBI as subjects with a word size of 6 and an e-value of $1e^{-4}$. If the region either approximated to TEs or included a large ORF (>1,000 bp), BLASTP was conducted using translated ORFs as queries and non-redundant protein sequences in NCBI as subjects with a word size of 6 and an e-value of $1e^{-4}$. Finally, regions that had an almost full-length conserved RT domain in BLASTX or BLASTP searches were selected.

2.5. Phylogenetic analysis of rps13 and rps10i235

Phylogenetic analysis of *rps13* gene sequences was performed using sequences of representative angiosperm species and three legume species (Table S1). *Liriodendron tulipifera* was used as an outgroup and ML analysis was conducted as described above.

Sequences of rps10i235 among the 27 legumes were aligned with MAFFT using default options. Mitovirus NERVE sequences were then excluded from the alignment. Phylogenetic relationships of rps10i235 without NERVEs were analyzed as described above using *Cercis canadensis* as an outgroup.

2.6. Screening of mitochondrial targeting protein with RT and EN domains

Open reading frames harboring reverse transcriptase (RT) and endonuclease (EN) domains in mitogenomes of Caesalpinioideae were translated into amino acids sequences and used as queries in BLASTP. The subject database comprised Caesalpinioideae (taxid:3804) sequences from the non-redundant protein sequence database at NCBI

with a word size of 6 and an e-value of $1e^{-4}$. Retrieved sequences with only one domain were excluded. Mitochondrial transit peptides were predicted in the retrieved sequences using Localizer (Sperschneider et al., 2017), TargetP-2.0 (Armenteros et al., 2019), Mitoprot II (Claros and Vincens, 1996) and Predotar (Small et al., 2004). Putative mitochondria-targeted protein sequences that were supported by at least two programs were selected.

3. Results

3.1. Mitogenome assembly

Assemblies produced a single mitochondrial contig for seven species and gaps were closed by PCR and Sanger sequencing. Due to their complexity, which included sequences of plastid origin and repeat accumulation, draft mitogenomes for three species (*Ceratonia siliqua*, *Prosopis glandulosa* and *Pachyrhizus erosus*) were represented by concatenations of three to six assembled contigs in analyses along with the completed mitogenomes. Assembled mitogenomes ranged from 308,122 to 758,210 bp in length with depth of coverage averaging from 108X to 421X (Table 1).

3.2. Phylogenetic relationships of mitogenomes

Maximum likelihood (ML) analyses were conducted using 26 shared protein coding genes (26 CDS; 25,115 bp aligned length) extracted from the mitogenome sequences of 27 Fabaceae species (including 10 new and 17 previously published). The phylogeny (Fig. 1), which was largely congruent with family-wide phylogenies based on other molecular data (Cardoso et al., 2013; Schwarz et al., 2015; LPWG, 2017), showed that subfamilies Cercidoideae and Detarioideae diverged earlier than Caesalpinioideae and Papilionoideae. Taxa from Caesalpinioideae and Papilionoideae formed monophyletic groups that are sisters to each other. All but five nodes of Fabaceae were well supported (> 90% bootstrap values). The three mimosoid taxa (*Leucaena trichandra*, *Acacia ligulata*, *Prosopis glandulosa*) formed a monophyletic group that was highly supported (100% bootstrap), however, phylogenetic relationships among these taxa were weakly supported.

3.3. Loss and gain of mitochondrial protein coding genes

The number of intact genes identified in the 27 Fabaceae mitogenomes varied from 31 to 37 (Fig. S1). This variation was attributable to nine genes (*cox2*, *rpl2*, *rpl10*, *rps1*, *rps13*, *rps14*, *rps19*, *sdh3* and

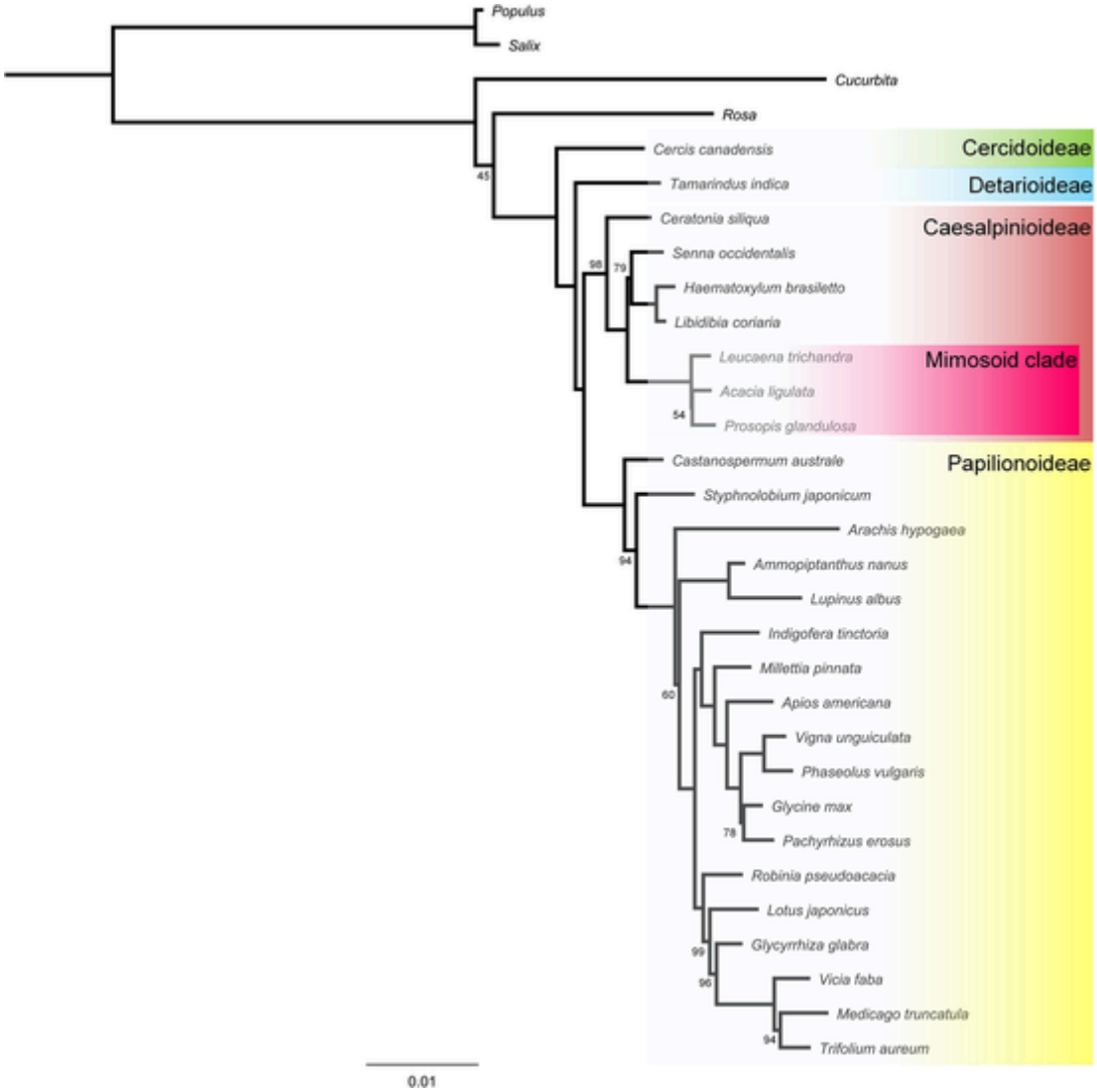

Fig. 1. Maximum likelihood tree of Fabaceae based on mitogenome gene sequences. The mitogenome protein coding regions of 26 genes were used to infer the phylogeny. Bootstrap values < 100 are indicated at the nodes. Scale indicates number of nucleotide substitutions per site. Subfamilies are named (right) and indicated by shaded boxes. In this study, specific epithet is omitted in other figures if the taxa are same as in Fig. 1. Otherwise, it is specified.

sdh4). While nine of the 10 newly sequenced mitogenomes showed identical content of intact genes to previously published related taxa in their subfamilies (Caesalpinioideae and Papilionoideae), *Indigofera tinctoria* (Papilionoideae) was distinct. In *I. tinctoria*, *rps14* appeared pseudogenized while all other legume taxa have an intact copy. In contrast, an intact copy of *rps13* (360 bp) was only found in *I. tinctoria*. Two species of Caesalpinioideae, *Haematoxylum brasiletto* and *Acacia ligulata*, had pseudogenized copies (333 and 332 bp, respectively), and the remaining 24 species did not contain recognizable (>100 bp) *rps13* sequences.

Maximum likelihood analysis of the intact and pseudogenized *rps13* genes from Fabaceae along with a selection of sequences from across angiosperms showed a foreign origin of the gene in a tree (Fig. S2) in which several nodes were not well resolved. Two nodes, one with two Fabaceae species (*H. brasiletto* and *A. ligulata*) and the other with one Fabaceae species (*I. tinctoria*) and two species of Convolvulaceae (*Cuscuta gronovii* and *Ipomoea nil*), were supported with high bootstrap values (92% and 94%, respectively).

3.4. Variation in intron content among Fabaceae

Acacia ligulata showed the highest number of mitogenomic introns while the papilionoid *Trifolium aureum* had the lowest (19 versus 15). Likewise, *A. ligulata* had the greatest accumulative intron length, *T. aureum* the least (41,116 bp versus 22,758 bp) (Fig. 2a). Variation in the number of *cis*-spliced introns (15–19) was due to differential loss of ccmFci829, cox2i691, rps3i74 and of the *rpl2* gene. The intron of *rpl2* (rpl2i846) was unique to Caesalpinioideae (Fig. S1). While the loss of cox2i691 was shared by all Papilionoideae, losses of ccmFci829 and rps3i74 were specific to the papilionoid genus *Trifolium*.

The length of individual introns varied from 826 to 4,898 bp (Fig. S3). The size range of 11 of these introns (nad1i477, nad1i728, nad2i1282, nad4i461, nad4i976, nad5i230, nad5i1872, nad7i140, nad7i209, nad7i676 nad7i917) was narrow (<1,000 bp), while the range for the remaining eight introns (ccmFci829, cox2i691, nad2i156, nad2i709, nad4i1399, rpl2i846, rps3i74, rps10i235) was wider

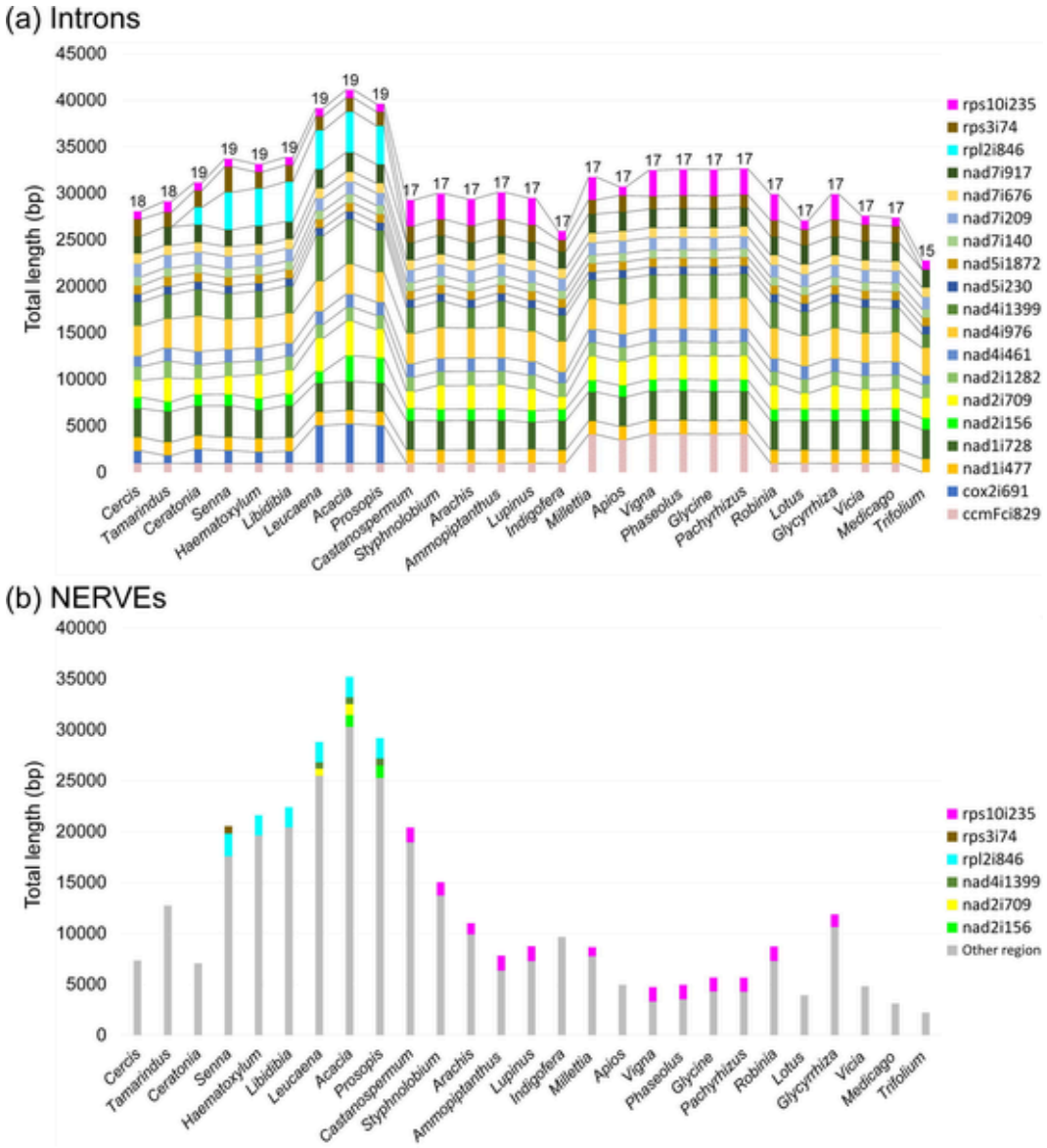

Fig. 2. Number and length variation of *cis*-spliced introns and accumulated NERVEs in legume mitogenomes. (a) The values above each bar represent the number of *cis*-spliced introns in each taxon. Dotted lines illustrated the length variation of each intron. (b) The position and length of NERVEs are presented. Grey bars indicating 'other regions' refer to noncoding, non-intronic regions of the mitogenome. In both (a) and (b), taxon names are listed below each histogram according to the phylogeny in Fig. 1. The color key for each of the introns is presented on the right.

(1,511–3,436 bp) across legumes. Except for ccmFci829 and cox2i691, the enlarged introns were associated with mitovirus NERVEs (Fig. 2b).

3.5. Mitovirus NERVE accumulation in introns and intergenic regions

Similar to intron variation, *Acacia ligulata* showed the highest number of intronic NERVEs while *T. aureum* had the lowest (4 versus 0). Values for accumulated NERVE sequence length followed the same trend; *A. ligulata* contained more than 15 times as much sequence compared to *Trifolium aureum* (35,209 bp versus 2,218 bp) (Fig. 2b). There was no clear preference for intronic compared to intergenic regions as NERVE integration sites. The number of intronic NERVEs tended to parallel genome-wide accumulation patterns with caesalpinioid legumes having higher amounts of NERVE sequences compared to other subfamilies, except in *Ceratonia siliqua*. The amount of NERVE sequences in *C. siliqua* (7,058 bp) was similar to the early diverging taxon *Cercis canadensis* (Cercidoideae) (7,339 bp). Compared to *C. siliqua*, other caesalpinioid legumes accumulated a substantially higher amount of NERVE sequences (20,540–35,209 bp), acquiring five intronic NERVEs in nad2i156, nad2i709, nad41399, rpl2i846 and rps3i74. The NERVE in rps10i235 was unique to papilionoids and shared by 12 of 18 taxa. Across the subfamily Papilionoideae, variation in the amount of total NERVEs (Fig. 2b) was driven by differential accumulation in intergenic regions rather than introns.

3.6. Transposable elements in Caesalpinioideae mitogenomes

Three loci (Fig. 3 and Fig. S4) that met the criteria for transposon-related sequences were detected (see Section 2.4. *Content analysis* in Materials and Methods). One locus included a domain of LTR-like RT while the other two loci were non-LTR-like RT (L1). All three were identified in mitogenomes of caesalpinioid legumes. The LTR-like RT domain located between *atp4* and *ccmC* in the mitogenomes of *Ceratonia siliqua*, *Haematoxylum brasiletto* and *Libidibia coriaria* did not reside within an ORF (Fig. S4). The other two loci were not in close proximity to any mitochondrial protein-coding regions. To discriminate between these two regions, they were designated L1FMT1 (Fig. 3a, see legend for locus information) and L1FMT2 (Fig. 3b). Three mimosoid legumes, *Leucaena trichandra*, *Acacia ligulata* and *Prosopis glandulosa,* shared L1FMT1, for which the integration site was ambiguous. The RT domain was not located within an ORF in *L. trichandra* and *P. glandulosa*, however, it was situated within an ORF in *A. ligulata*. *Acacia* also contained L1FMT2 as a 6,583 bp element flanked by TSDs of 15 bp with a single mismatch (Fig. 3b,c). L1FMT2 included several large ORFs (> 1,000 bp). Submission of the L1FMT2 sequence to the NCBI conserved domain database (Marchler-Bauer et al., 2017) identified domains of DUF4283 (pfam14111; unknown function), Zf-CCHC_4 (pfam14392; zinc knuckle), PTZ00121 (PTZ00121; a provisional MAEBL), Exo_endo_phos (pfam03372; endonuclease/exonuclease/phosphatase family), RT_nLTR_like (cd01650; RT_nLTR: non-LTR retrotransposon and non-LTR retrovirus reverse transcriptase), Zf-RVT (pfam13966; zinc-binding in reverse transcriptase) and RNase_H-like (cd06222; ribonuclease H-like superfamily). Domains of RT and ribonuclease H (RNH) did not reside within ORFs. *Acacia ligulata* is the only species that contained ORFs together with RT and a putative endonuclease (EN) domain in the mitogenome among the legume taxa analyzed. The Exo_endo_phos, RT_nLTR_like and RNase_H_like are considered domains of EN, RT, and RNH, respectively.

3.7. Putative mitochondrial targeted protein with RT and EN domains

A single uncharacterized, nuclear-encoded protein (Fig. S5) that met the criteria (see section *2.6. Screening of mitochondrial targeting protein with RT and EN domains* of Materials and Methods) was detected in

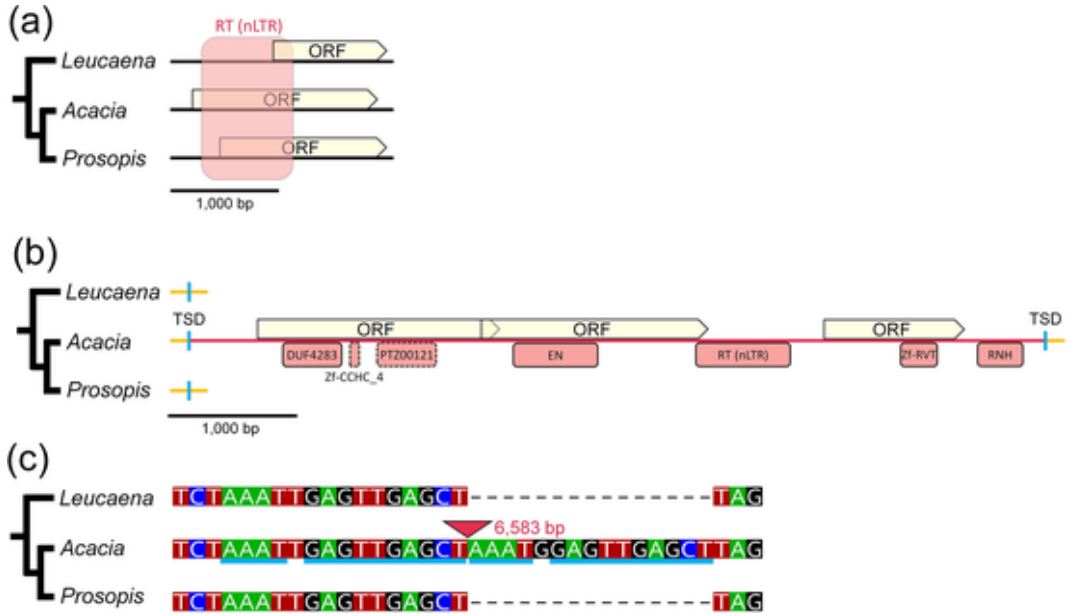

Fig. 3. Identified L1 elements in the three mimosoid genera, *Leucaena trichandra*, *Acacia ligulata*, and *Prosopis glandulosa*. Cladograms (left) are based on the phylogeny in Fig. 1. (a) Schematic representation of the L1FMT1 open reading frames (ORFs). Only the largest ORFs are illustrated. The region that matched the RT (nLTR) domain based on BLASTX or BLASTP is shaded by a red box. The illustrated regions of L1FMT1 are corresponding to positions of *Leucaena trichandra* (MH717173; 379,073 to 381,113), *Acacia ligulata* (NC_040998; 44,811 to 46,856), and *Prosopis glandulosa* (MW448450; 5,904 to 3,858). (b) Schematic representation of integration site of L1FMT2. The mitochondrial regions prior to integration are marked with yellow lines, while the region of L1FMT2 is indicated with a red line. The blue, vertical bars indicate target site duplications (TSDs). The ORFs (>1,000 bp) are illustrated. The regions that matched to known domains based on BLASTX or BLASTP are marked with red boxes, and those with solid lines represent conserved domains while the dotted lines indicate truncated domains. Sizes are indicated by the scale bar below each region in base pairs (bp). The illustrated regions of L1FMT2 are corresponding to positions of *Leucaena trichandra* (MH717173; 549,337 to 549,618), *Acacia ligulata* (NC_040998; 100,539 to 107,418), and *Prosopis glandulosa* (MW448450; 17,229 to 16,948). (c) Alignments around target site duplications are enlarged. Target site duplications are underlined with blue bars except for mismatches. Insertion site of L1FMT2 is indicated by the red triangle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Prosopis alba. The domains were identical to the 3′ portion of L1FMT2 (Fig. 3b). However, it was not possible to conclusively determine if the protein targets to mitochondria since this prediction was supported by two programs (Localizer and Mitoprot II) but not by two others (TargetP-2.0 and Predotar) (Fig. S5).

3.8. Characteristics of intronic NERVEs

The integration sites of six intronic NERVEs were identified (Figs. 4, 5, and S6). Four of them, nad2i156, rpl2i846, rps3i74 (Fig. 4) and rps10i235 (Fig. 5), were delimited by 10–16 bp TSDs while two others were not. In nad2i709 and nad4i1399 of caesalpinioid legumes, NERVE sequences were found together with other non-homologous sequences relative to *Ceratonia siliqua*, making it difficult to ascertain their origin. The nad4i1399 in *Leucaena trichandra* included an intact copy of the *rps19* gene close to the mitovirus NERVE, which probably resulted from a mitochondrial DNA rearrangement between intergenic and intronic regions. In *Senna*, only one of two species possessed an intronic NERVE in rps3i74 (Fig. 4c), and it was the shortest among all intronic NERVEs with TSDs. The majority of this NERVE corresponded to a portion of RdRp in the *Petunia exserta* mitovirus 1 genomic RNA (BK010432) but was greatly truncated in the 3′ region (Fig. S7).

In Caesalpinioideae, intronic NERVEs with TSDs were detected in all members of a clade or found in single species (Fig. 4a,b). However, the pattern was more complicated for rps10i235 in Papilionoideae (Fig. 5). Papilionoids both with (12 taxa) and without (6 taxa) intronic NERVEs were not contained in monophyletic groups (Fig. 5). The sequence of rps10i235 (excluding its NERVE) was analyzed (Fig. S8) to test if the phylogeny inferred was congruent with that generated from the concatenated CDS data set (Fig. 1). The phylogenetic tree of rps10i235 (without NERVE; fig. S8 Material online) was poorly supported at most nodes. The tree toology was congruent with the CDS tree (Fig. 1) except

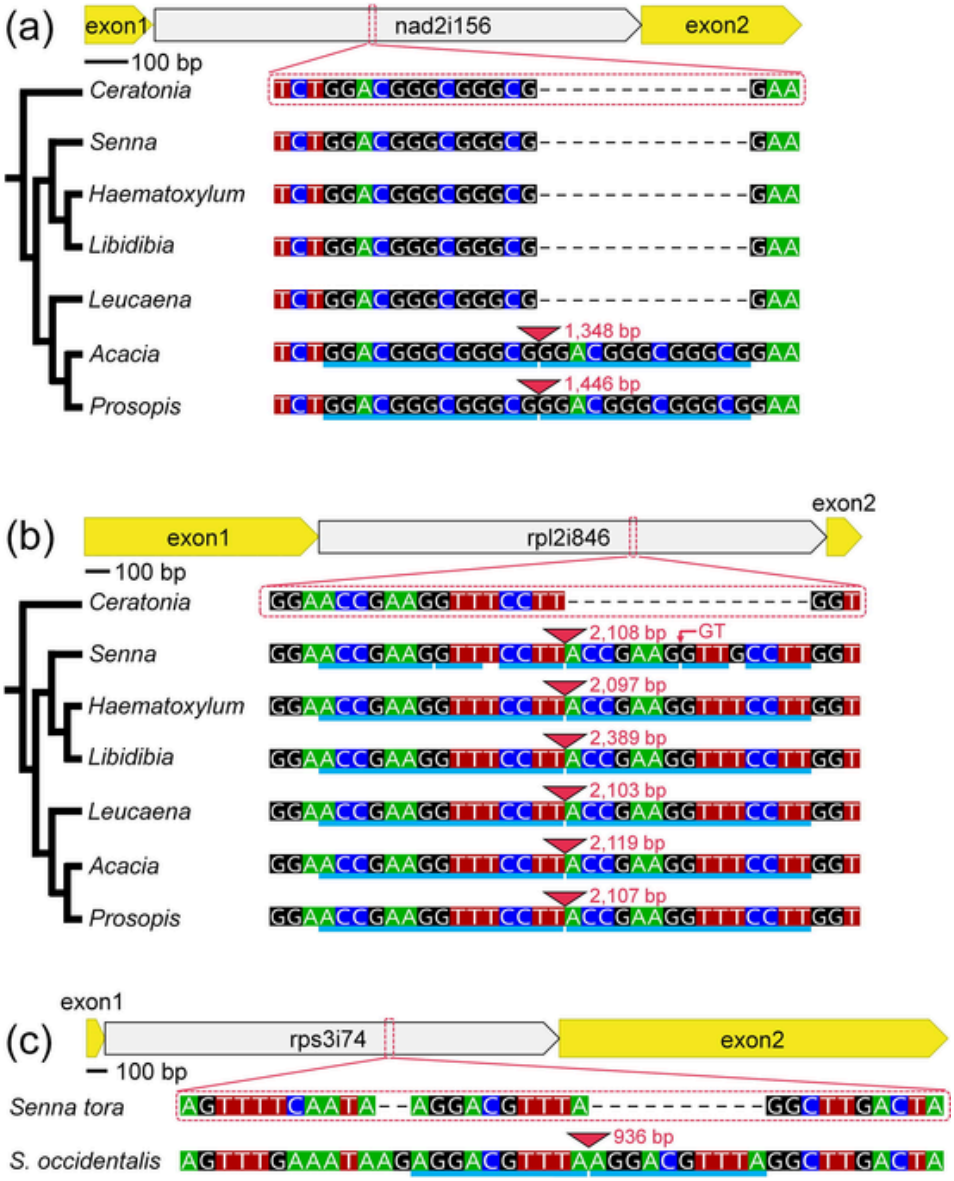

Fig. 4. Structure of intronic NERVEs in Caesalpinioideae. Gene regions of *Ceratonia siliqua* are presented above each sequence alignment of nad2i156 (a) and rpl2i846 (b). For rps3i74 (c), relevant region from *Senna tora* is presented. Sizes are indicated by the scale bar below each region in basepairs (bp). Alignments around target site duplications are enlarged and corresponding regions in *C. siliqua* or *S. tora* are marked with red lines. Topology shown is based on the phylogeny in Fig. 1. Target site duplications are underlined with blue bars except for mismatches. Small insertions, unique to a single taxon, are shown in red with red arrows. NERVEs are indicated with red triangles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

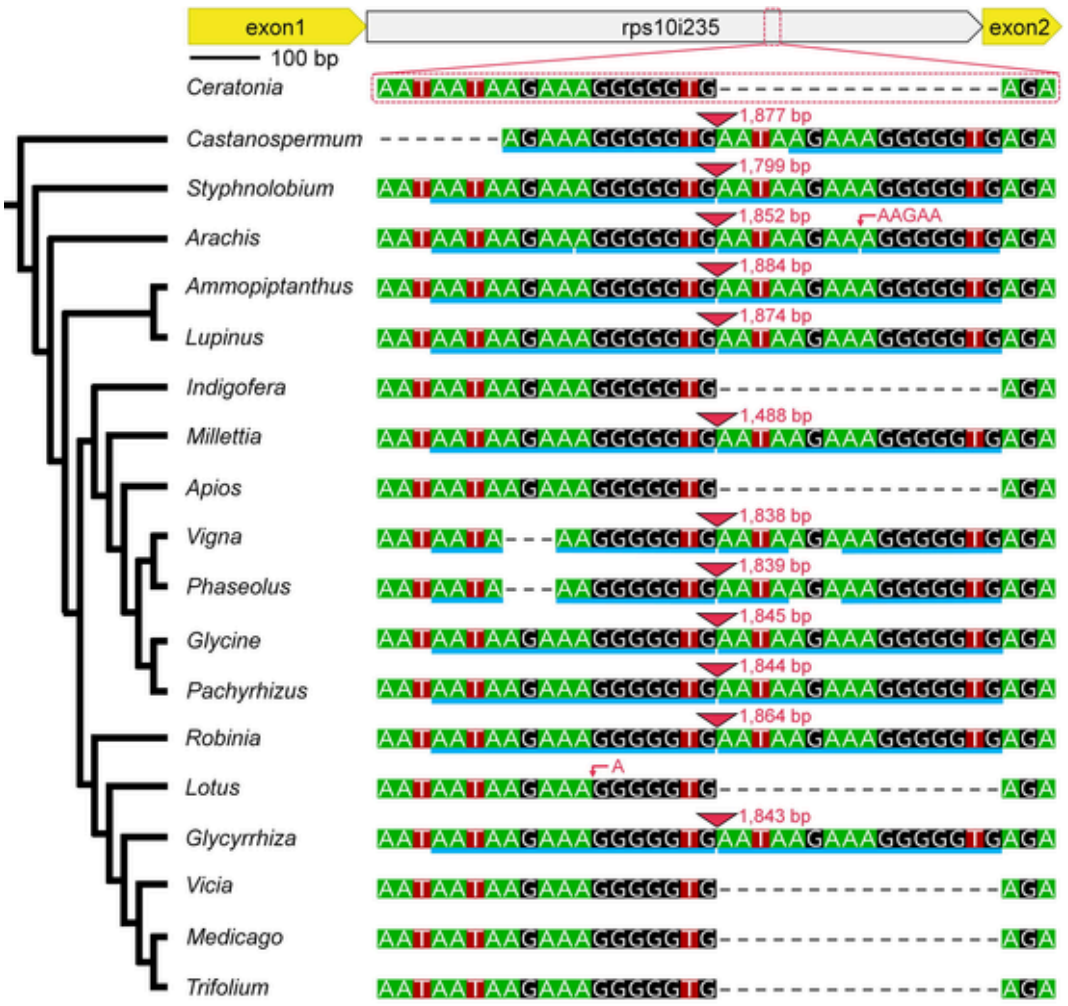

Fig. 5. Structure of the Papilionoideae intronic NERVE in rps10i235. A schematic representation of *rps10* in *Ceratonia siliqua*, which lacks a NERVE in this region, is presented above the sequence alignment. The size of the *C. siliqua* gene is indicated by the bar below exon, in base pairs (bp). A portion of the alignment around target site duplications is enlarged and corresponding regions in *C. siliqua* are indicated with red lines. The cladogram (left) is based on the mitochondrial CDS phylogram (Fig. 1). Target site duplications are underlined with blue bars except mismatches. Small insertions, unique to single taxa, are shown in red with red arrows. NERVEs are indicated with red triangles along with their length. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for the positions of *Lotus/Glycyrrhiza*, *Styphnolobium*, and *Apios/Millettia* but in all cases bootstrap support was very low in the rps10i235 tree.

4. Discussion

4.1. Gene content evolution in legumes and recapture of rps13 in Indigofera tinctoria

Among 27 legume taxa, several lineages have lost at least one of eight mitochondrial protein coding genes (*cox2*, *rpl2*, *rpl10*, *rps1*, *rps14*, *rps19*, *sdh3*, and *sdh4*) (Fig. S1). Among them, the loss of the *rps14* has not been reported in legumes, however, it was frequently lost across angiosperms (Adams et al., 2002). The presence of intact *rpl2* and *rps19* genes, which are not present in the other subfamilies, is a distinct feature of Caesalpinioideae. These genes are regarded as remnants of an ancient legume mitogenome since intact copies formed a monophyletic group with a pseudogenized copy from *Cercis canadensis* (Cercidoideae) in previous phylogenetic analyses (Choi et al., 2019). However, an intact *rps13* gene, discovered in *I. tinctoria*, may not be regarded as a remnant of an ancestral copy as it does not form a monophyletic group with pseudogenized copies from Caesalpinioideae but instead is nested within a clade of Convolvulaceae (Fig. S2). Therefore, the intact copy of *rps13* in *I. tinctoria* has likely been horizontally transferred from taxa of Convolvulaceae. Since there is no native intact copy of *rps13* in any Fabaceae, the transfer of *rps13* can be classified as re-capture horizontal gene transfer (HGT), which is a rare event (Sánchez Puerta, 2014). Legumes have been primarily designated as a donor of mitochondrial DNA in HGT between other parasitic plants (Barkman et al., 2007; Sanchez-Puerta et al., 2017; Kovar et al., 2018; Choi et al., 2019; Sanchez-Puerta et al., 2019). The recapture HGT of *rps13* suggests that legumes can also serve as a recipient of mitochondrial DNA from other plant species. Additional cases may be discovered in legumes as more mitogenome sequences become available.

4.2. Possible relationship between integration of mitovirus genomic RNA and L1-like mechanisms in plant mitochondria

Widespread distribution of mitovirus NERVEs was discovered in legume mitogenomes (Fig. 2). In some lineages, the total amount of NERVE sequences as well as the number of intronic NERVEs are noticeably higher than in related taxa. For example, two monophyletic groups in Caesalpinioideae [(*Leucaena*, *Acacia*, *Prosopis*) and (*Senna*, *Haematoxylum*, *Libidibia*, *Leucaena*, *Acacia*, *Prosopis*)] have a substantial amount of NERVE sequences compared to sister lineages of those

groups. A recent study of a plant mitovirus infecting *Chenopodium quinoa* (Amaranthaceae) revealed that mitoviruses could be vertically transmitted to all progeny of infected plants (Nerva et al., 2019), implying that a single initial infection can persist for many generations. This idea is also supported by the fact that mitovirus-related sequence reads have been observed in higher frequency from floral tissues (10-fold) than from other tissues in *Petunia exserta* (Nibert et al., 2018). It is conceivable that accumulated NERVEs in a monophyletic group represent a relic of long-lasting infection of mitovirus in the common ancestor.

Another factor that may account for differences in the accumulation of NERVEs is the variable ability of plant mitochondria to take up mitovirus genomic RNA via reverse transcriptase (RT) activity. Some plant mitochondria may have RT activity based on the presence of a domain in group II intron-encoded protein genes (e.g., *matR* in nad1i728) or other ORFs in the mitogenome (Wahleithner et al., 1990; Fassbender et al., 1994; Moenne et al., 1996). However, the RT domain in *matR* was truncated in angiosperms (Brown et al., 2014).

In many angiosperms mtDNA of nuclear origin, which are commonly identified by comparison to nuclear TEs, comprise 1–5% of mitogenomes (Mower et al., 2012) but most exist as small fragments (Knoop et al., 1996; Alverson et al., 2010; Mower et al., 2012 Choi et al., 2019;). It is noteworthy that the mitogenome of *Acacia ligulata* includes an ORF (L1FMT1) with a full-length RT domain derived from long interspersed elements-1 (L1) belonging to a group of non-LTR retrotransposons (Fig. 3a). The level of intactness and the presence of 15 bp TSDs suggest that *A. ligulata* integrated the other L1 (L1FMT2) more recently than *Leucaena* or *Prosopis* even though the RT domain was not identified in an ORF (Fig. 3b,c). It is likely that L1s in the mitogenome of *A. ligulata* are active or were previously active in mediating the integration of mitovirus genomic RNA since this species has accumulated the highest amount of mitovirus NERVEs among the legumes examined (Fig. 2). The acquisition of an ORF with the RT domain in the ancestor of *A. ligulata* may have been repeated over the evolutionary history of legumes. If this is the case, integration of mitovirus genomic RNA would be accelerated while the ORF for RT is intact and actively transcribed in the mitogenome.

Transposable element transfer to mitochondria via an RNA-intermediate has not been reported in angiosperms. The putative full-length L1 (L1FMT2; 6,583 bp) with flanking TSDs (Fig. 3b,c) suggests import as an mRNA followed by integration in the mitogenome by a L1-like retrotransposition mechanism. LINE-1 retrotransposons are 6–8 kb elements that encode two proteins (ORF1p and ORF2p) (Furano, 2000). Seven domains, DUF4283, Zf-CCHC_4, PTZ00121 (a provisional MAEBL), EN, RT, Zf-RVT and RNH, are present in L1FMT2. The first two (DUF4283 and Zf-CCHC_4) are dominant domains of ORF1p and latter three (EN, RT, and RNH) are usually part of ORF2p in plant L1s (Ivancevic et al., 2016). The Zf-RVT is closely related to the RT (non-LTR) domain in *Arabidopsis* (Galván-Gordillo et al., 2016). The odd domain in L1FMT2 is PTZ00121 (a provisional MAEBL). MAEBL is a domain of erythrocyte-binding proteins present in malarial parasites, so its presence in L1FMT2 seems highly unlikely. However, the domain of PTZ00121 has been identified from transcriptomes of a fungus *Phakopsora pachyrhizi*, causing Asian soybean rust (Rincão et al., 2018) and *Acyrthosiphon pisum* (pea aphid) (Burke and Moran, 2011). The PTZ00121 domain in L1FMT2 may reflect a complex domain recruitment history of L1s in plants (Heitkam et al., 2014). Identification of variable domain content led to a search for an L1-like protein with a mitochondrial transit peptide (mTP). An uncharacterized protein (LOC114733531) in *Prosopis alba* (Fig. S5) contained the very same domains found in the 3′ regions of L1FMT2. However, localization to mitochondria was equivocal since it is supported by only two of the four utilized targeting software programs.

The L1s were not the major constituent among TEs in the nuclear genomes of Papilionoideae (Kreplak et al., 2019), and whether or not L1s in Caesalpinioideae are currently active remains unknown. If L1s are highly active in nuclear genomes of Caesalpinioideae, they could influence mitogenomes by producing N-terminal truncated L1 protein with mTP.

The exact integration sites of four intronic NERVEs and a putative L1 (L1FMT2) flanked by TSDs (10–16 bp) were identified (Figs. 3–5). The similar integration site structures across legumes and other lines of evidence from *A. ligulata* suggest that integration of viral elements into introns is mainly mediated by target-primed reverse transcription (TPRT) of L1, which has both RT and EN domains (Kojima, 2010). Endonuclease-independent retrotransposition is also possible however this mechanism does not create TSDs (Morrish et al., 2002). The exact integration site of NERVEs in non-intronic regions is difficult to determine due to the non-conserved nature of intergenic regions of mitogenomes. Alternatively, NERVEs in non-intronic regions may integrate by a different mechanism.

The integration process of intronic NERVEs in both nad2i709 and nad4i1399 is unclear since the sequence of those introns is highly variable, which may be the result of multiple insertions and deletions after integration (Fig. S6). The presence of an intact *rps19* gene proximal to mitovirus NERVEs in nad4i1399 of *Leucaena trichandra* suggests that the integration sites of NERVEs have been disturbed by mitochondrial DNA rearrangements. Similarly, the introns ccmFci829 and cox2i691 show drastic size variation in legumes (Fig. S3) due to insertion of sequence of unknown origin (Choi et al., 2019). Comparative analyses with more phylogenetically dense sampling of mitogenomes are required to elucidate the evolutionary history of these introns.

The length of plant mitovirus RNA genomes have been estimated at 2,701–2,944 nucleotides (Nibert et al., 2018). However, the size range of intronic NERVEs with TSDs was much broader (936–2,389 bp) and the longest copy identified was shorter than known plant mitoviruses (Figs. 4 and 5). Lengths of older NERVEs can be modified via insertions and deletions. The shortest intronic NERVE, in rps3i74, appeared to be the youngest because it was only detected in one of the two species of *Senna*. Compared to RdRp of a plant mitovirus (DAB41750), the intronic NERVE in rps3i74 is severely truncated in 3′ sites (Fig. S7). Truncation of the 5′ region is commonly observed and can be explained in L1 retrotransposition by microhomology-mediated or non-homologous end-joining between the target genome and the point of the truncation during the processing of TPRT (Zingler et al., 2005; Suzuki et al., 2009; Kojima, 2010). Together with the presence of TSDs and 5′ truncation, 3′ poly(A) tails are a hallmark of L1 retrotransposition but a 3′ truncated retrocopy with TSDs is uncommon (Morrish et al., 2002). A 3′ truncation was reported for *retro*-pseudogenes lacking poly(A) tails in mammals (Schmitz et al., 2004). Two pathways that include RNA degradation before the integration and internal (instead of 3′ end) priming during TPRP have been suggested to explain this phenomenon (Noll et al., 2015). In plant mitochondria both pathways are possible. The 3′ poly (A) tail allows L1 ORF2p to bind RNA molecules in humans (Doucet et al., 2015). However, in plant mitochondria mRNAs usually do not include a poly(A) tail and polyadenylation marks the mRNA for degradation following primary cleavage by endoribonuclease (Schuster and Stern, 2009; Hammani and Giegé, 2014). Therefore, truncation can occur even before the integration if L1-like retrotransposition requires a poly(A) tail in plant mitochondria. However, the cases reported in this study are not sufficient to suggest preference in 3′-end recognition. In some land plants, relaxed 3′-end recognition of L1 has been reported (Ohshima, 2012). If this is the case for L1-like retrotransposition of mitovirus NERVEs, the 3′end truncation could be attributed to internal priming during TPRP rather than RNA degradation. Since retrotransposition activity in plant mitochondria is an unexplored field of research, additional empirical evidence is essential to elucidate the underlying molecular mechanisms regarding the genesis of mitovirus NERVEs.

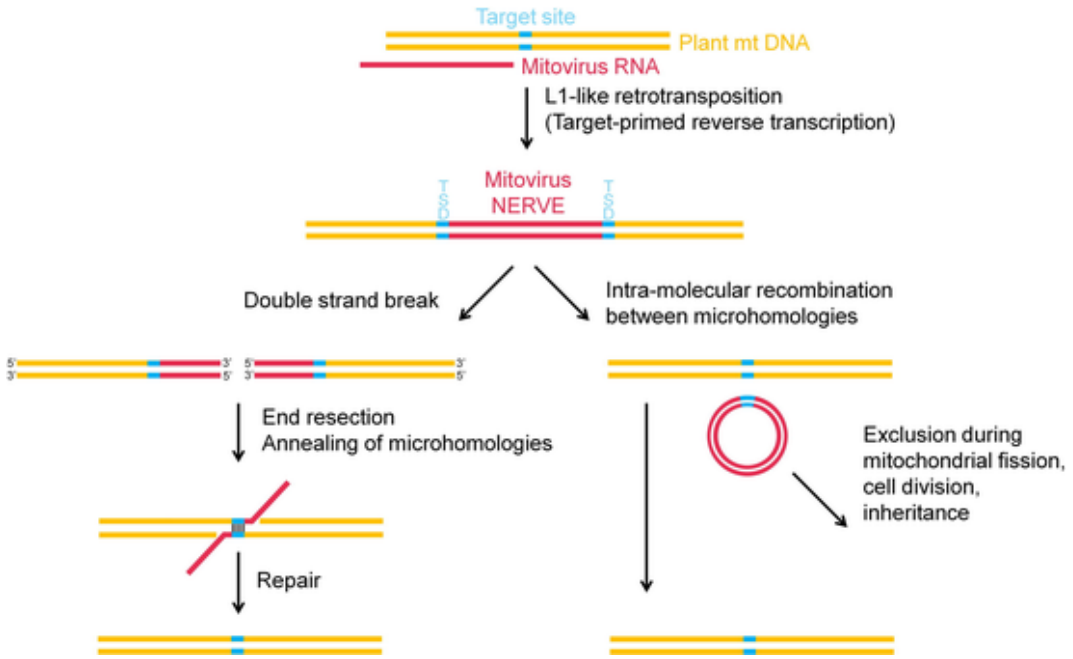

Fig. 6. Possible molecular mechanisms for integration and precise deletion of mitovirus NERVE. Plant mitochondrial DNA, mitovirus-related elements, target site are shown in yellow, red, and blue, respectively. Before integration, mitovirus RNA can be a full-length genomic RNA or truncated copy with poly(A) tail. The L1-like retrotransposition with target site duplications (TSDs) requires the activity of reverse transcriptase to generate complementary DNA while endonuclease (EN) cleaves the target site and provides a primer for reverse transcription. How the EN recognizes the target site in plant mitochondria and its site preference is currently unknown. In the process of L1-like retrotransposition, the mitovirus sequence can be truncated during end joining to the target site and internal priming. After integration, precise deletion can occur via microhomology-mediated end joining (MMEJ) as well as through intra-molecular recombination based on sequence homology between TSDs. Following a double-strand break in the NERVE, the process of MMEJ can be initiated. The end resection requires activity of various enzymes (including exonuclease) and results in exposure of microhomologies of single strand DNAs for annealing. After annealing, removal of non-homologous tails, gap filling by DNA synthesis, and ligation can result in reversion of the NERVE integration. Intra-molecular recombination is able to generate two subgenomic molecules. The smaller circular molecule only includes mitovirus NERVE and one copy of the target site. Barring re-integration based on intermolecular recombination between TSDs, the larger genomic molecule, with all essential mitochondrial genes, returns to its pre-integration status and is perpetuated through mitochondrial fission, cell division, and inheritance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.3. Multiple and precise deletion of NERVEs in rps10i235 and its implications for mitogenome evolution

The most prominent feature of intronic NERVEs in this study was the precise deletion of the viral element as well as one copy of the TSD in rps10i235. The NERVE in rps10i235 was detected in 12 of 18 papilionoid taxa (Fig. 5). The integration site in six of the 18 papilionoid species has reverted to the pre-integration state and is identical to the caesalpinioid *Ceratonia siliqua*. *Castanospermum australe*, which contains the rps10i235 NERVE, is the earliest diverging papilionoid in the mitochondrial CDS phylogeny (Fig. 1) and the genus belongs to the earliest diverging ADA clade in Papilionoideae (Cardoso et al., 2012, 2013). This suggests that the integration occurred in a common ancestor of Papilionoideae making the NERVE in rps10i235 at least 58.6 million years old, the divergence time of the crown node of Papilionoideae (Lavin et al., 2005), and was then deleted multiple times. An alternative is that the reversion of the integration site is due to multiple horizontal transfers of the intron (without NERVE) from species of other subfamilies of Fabaceae. To test alternative explanations, a phylogenetic tree using only the intron sequences was constructed (Fig. S8). None of the intron sequences from papilionoid taxa were more closely related to taxa in other subfamilies, and the tree topology of Papilionoideae was similar to that of the concatenated mitochondrial CDS phylogeny (Fig. 1), with some minor exceptions. Therefore, an HGT scenario can be rejected. This kind of precise deletion of a retrocopy has not been reported in plants. However, in primate genomes precise removal of a retrocopy as well as one copy of TSD (10–20 bp) was rarely observed from L1 protein-mediated insertion sites, (0.5–1%) (van de Lagemaat et al., 2005).

Two possible mechanisms for precise deletion of intronic NERVE sequences as well as their TSDs in rps10i235 are proposed (Fig. 6). The first is based on an error-prone repair mechanism known as microhomology-mediated end-joining (MMEJ) while the other is based on intramolecular recombination between direct repeats. Both mechanisms rely on sequence identity between the two 16 bp TSDs.

In *Castanospermum australe*, the two copies of the target site have diverged (Fig. 5), however relatives of the taxa where precise deletions occurred (e.g., *Millettia pinnata*, *Robinia pseudoacacia*, *Glycyrrhiza glabra*) have maintained sequence identity between the duplicates over ca. 59 million years. The minimal length of identity required for this MMEJ is only six nucleotides with some mismatches allowed (García-Medel et al., 2019). Moderate sequence divergence between TSDs may not greatly affect the probability of precise deletion, however the length of microhomologies is likely to affect the stability of the annealed intermediates during MMEJ (Seol et al., 2018).

Intramolecular recombination between direct repeats in the master mitogenome can result in the formation of two smaller subgenomic molecules. If one of the two subgenomic molecules lacks mitochondrial genes and does not re-integrate into the other subgenome, it can result in deletion of sequences between the two direct repeats (Small et al., 1989). It was suggested that the frequency of this kind of recombination is positively correlated with the size of the repeat unit (Arrieta-Montiel et al., 2009). Recombination between large direct repeats (>1,000 bp) can easily generate subgenomic molecules even in a single individual (Maréchal and Brisson, 2010) but this is considered a reversible event by recombination between the two subgenomic molecules (Andre et al., 1992). In contrast, the frequency of recombination between short direct repeats is low and regarded as non-reversible. In *Arabidopsis thaliana*,

recombination between 10 and 18 bp direct repeats occurs only rarely in wild-types but plants lacking expression of AtWhy1 and AtWhy3, which suppress short-homology recombination, are prone to generate circular molecules and/or head-tail concatemers of subgenomic regions of plastid DNA (Maréchal et al., 2009). Likewise, disruption of recombination surveillance machinery (e.g., *OSB1*, *RecA1*, *RecA3*, and *Msh1*) increased the rate of recombination between short direct repeats in plant mitogenomes (Maréchal and Brisson, 2010).

Recombination between short direct repeats such as the TSDs of intronic NERVEs in rps10i235 is less likely since the frequency of this kind of recombination is positively correlated with the size of the repeat unit (Arrieta-Montiel et al., 2009). However, the long evolutionary history of this intronic NERVE (58.6 million years, Lavin et al., 2005) may have allowed precise deletion multiple times. Microhomology-associated repair and recombination mechanisms have been suggested as a factor resulting in increased mitogenome complexity (Christensen, 2013, 2014, 2018; Gualberto and Newton, 2017; Xia et al., 2020). As shown in this study, however, microhomology-mediated deletion mechanisms may have contributed to mitogenome maintenance by allowing a dynamic equilibrium between the influx of viral sequences and the preservation of native sequences.

5. Conclusions

A L1-like retrotransposition mechanism has contributed to the formation of mitovirus NERVEs in the mitogenomes of legumes. Furthermore, a possible role for microhomology-mediated deletion mechanisms in NERVEs elimination is suggested. The identification of short direct repeats and the analysis of associated sequences have the potential to aid in the discovery of more retrotransposition in plant mitogenomes. Lastly, ORFs of transposable elements in the mitogenome that may encode proteins can mediate mitovirus integration but the possible contribution of nuclear-encoded protein factors cannot be ruled out. It is clear that plant mitoviruses have infected legume mitogenomes. However, the existence of contemporary mitoviruses infecting legumes has not been confirmed and how these are related to fossilized intronic mitovirus NERVEs remains elusive. More transcriptome and genome sequencing would enhance future understanding of the evolutionary history of mitoviruses and mitogenomes.

CRediT authorship contribution statement

In-Su Choi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Martin F. Wojciechowski:** Conceptualization, Funding acquisition, Validation, Supervision, Writing - review & editing. **Robert K. Jansen:** Conceptualization, Funding acquisition, Validation, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from the National Science Foundation (DEB-1853024) to R.K.J., M.F.W. and T.A.R., and the Sidney F. and Doris Blake Professorship in Systematic Botany to R.K.J.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107236.

References

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380–395.
Adams, K.L., Qiu, Y.-L., Stoutemyer, M., Palmer, J.D., 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. PNAS 99, 9905–9912.
Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K., Palmer, J.D., 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol. Biol. Evol. 27, 1436–1448.
Andre, C., Levy, A., Walbot, V., 1992. Small repeated sequences and the structure of plant mitochondrial genomes. Trends Genet. 8, 128–132.
Armenteros, J.J.A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., Nielsen, H., 2019. Detecting sequence signals in targeting peptides using deep learning. Life Sci. Alliance. 2, e201900429.
Arrieta-Montiel, M.P., Shedge, V., Davila, J., Christensen, A.C., Mackenzie, S.A., 2009. Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics 183, 1261–1268.
Barkman, T.J., McNeal, J.R., Lim, S.-H., Coat, G., Croom, H.B., Young, N.D., Claude, W.d., 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol. Biol. 7, 248.
Bauchet, G.J., Bett, K.E., Cameron, C.T., Campbell, J.D., Cannon, E.K., Cannon, S.B., Carlson, J.W., Chan, A., Cleary, A., Close, T.J., 2019. The future of legume genetic data resources: Challenges, opportunities, and priorities. Legume Sci. 1, e16.
Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D., 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424, 197–201.
Bonen, L., 2012. Evolution of mitochondrial introns in plants and photosynthetic microbes. Adv. Bot. Res. 63, 155–186.
Bonen, L., Vogel, J., 2001. The ins and outs of group II introns. Trends Genet. 17, 322–331.
Brown, G.G., Colas des Francs-Small, C., Ostersetzer, O., 2014. Group II intron splicing factors in plant mitochondria. Front. Plant Sci. 5, 35.
Bruenn, J.A., Warner, B.E., Yerramsetty, P., 2015. Widespread mitovirus sequences in plant genomes. PeerJ 3, e876.
Burke, G., Moran, N., 2011. Responses of the pea aphid transcriptome to infection by facultative symbionts. Insect Mol. Biol. 20, 357–365.
Cannon, S.B., McKain, M.R., Harkess, A., Nelson, M.N., Dash, S., Deyholos, M.K., Peng, Y., Joyce, B., Stewart, Jr., C.N., Rolf, M., 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol. Biol. Evol. 32, 193–210.
Cardoso, D., de Queiroz, L.P., Pennington, R.T., de Lima, H.C., Fonty, É., Wojciechowski, M.F., Lavin, M., 2012. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. Am. J. Bot. 99, 1991–2013.
Cardoso, D., Pennington, R., de Queiroz, L., Boatwright, J., Van Wyk, B.-E., Wojciechowski, M., Lavin, M., 2013. Reconstructing the deep-branching relationships of the papilionoid legumes. S. Afr. J. Bot. 89, 58–75.
Chang, S., Wang, Y., Lu, J., Gai, J., Li, J., Chu, P., Guan, R., Zhao, T., 2013. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE 8, e56502.
Choi, I.-S., Ruhlman, T.A., Jansen, R.K., 2020. Comparative mitogenome analysis of the genus *Trifolium* reveals independent gene fission of *ccmFn* and intracellular gene transfers in Fabaceae. Int. J. Mol. Sci. 21, 1959.
Choi, I.-S., Schwarz, E.N., Ruhlman, T.A., Khiyami, M.A., Sabir, J.S., Hajarah, N.H., Sabir, M.J., Rabah, S.O., Jansen, R.K., 2019. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol. 19, 448.
Christensen, A.C., 2013. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol. Evol. 5, 1079–1086.
Christensen, A.C., 2014. Genes and junk in plant mitochondria—repair mechanisms and selection. Genome Biol. Evol. 6, 1448–1453.
Christensen, A.C., 2018. Mitochondrial DNA repair and genome evolution. In: Logan, D.C. (Ed.), Plant Mitochondria. Wiley-Blackwell, New York, pp. 11–32.
Claros, M.G., Vincens, P., 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur. J. Biochem. 241, 779–786.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.
Doucet, A.J., Wilusz, J.E., Miyoshi, T., Liu, Y., Moran, J.V., 2015. A 3′ poly (A) tract is required for LINE-1 retrotransposition. Mol. Cell 60, 728–741.
Drouin, G., Daoud, H., Xia, J., 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol. Phylogenet. Evol. 49, 827–831.
Egan, A.N., Vatanparast, M., 2019. Advances in legume research in the genomics era. Aust. Syst. Bot. 32, 459–483.
Fassbender, S., Brühl, K., Ciriacy, M., Kück, U., 1994. Reverse transcriptase activity of an intron encoded polypeptide. EMBO J. 13, 2075–2083.
Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. 5, 103–107.
Furano, A.V., 2000. The biological properties and evolutionary dynamics of mammalian

LINE-1 retrotransposons. Prog. Nucl. Res. Molec. Biol. 64, 255–294.
Galván-Gordillo, S.V., Martínez-Navarro, A.C., Xoconostle-Cázares, B., Ruiz-Medrano, R., 2016. Bioinformatic analysis of *Arabidopsis* reverse transcriptases with a zinc-finger domain. Biologia 71, 1223–1229.
García-Medel, P.L., Baruch-Torres, N., Peralta-Castro, A., Trasviña-Arenas, C.H., Torres-Larios, A., Brieba, L.G., 2019. Plant organellar DNA polymerases repair double-stranded breaks by microhomology-mediated end-joining. Nucleic Acids Res. 47, 3028–3044.
Goremykin, V.V., Lockhart, P.J., Viola, R., Velasco, R., 2012. The mitochondrial genome of Malus domestica and the import-driven hypothesis of mitochondrial genome expansion in seed plants. Plant J. 71, 615–626.
Gualberto, J.M., Newton, K.J., 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu. Rev. Plant Biol. 68, 225–252.
Guo, W., Grewe, F., Fan, W., Young, G.J., Knoop, V., Palmer, J.D., Mower, J.P., 2016. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. Mol. Biol. Evol. 33, 1448–1460.
Guo, W., Zhu, A., Fan, W., Adams, R.P., Mower, J.P., 2020. Extensive shifts from *cis* to *trans* splicing of gymnosperm mitochondrial introns. Mol. Biol. Evol. 37, 1615–1620.
Guo, W., Zhu, A., Fan, W., Mower, J.P., 2017. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. New Phytol. 213, 391–403.
Hammani, K., Giegé, P., 2014. RNA metabolism in plant mitochondria. Trends Plant Sci. 19, 380–389.
Heitkam, T., Holtgräwe, D., Dohm, J.C., Minoche, A.E., Himmelbauer, H., Weisshaar, B., Schmidt, T., 2014. Profiling of extensively diversified plant LINEs reveals distinct plant-specific subclades. Plant J. 79, 385–397.
Henze, K., Martin, W., 2001. How do mitochondrial genes get into the nucleus?. Trends Genet. 17, 383–387.
Hillman, B.I., Cai, G., 2013. The family Narnaviridae: simplest of RNA viruses. In: Advances in virus research. Elsevier, pp. 149–176.
Hillman, B.I., Esteban, R., 2011. Family Narnaviridae. In: King, A.M., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus taxonomy: Ninth Report of the International Committee on taxonomy of viruses. Elsevier, New York, pp. 1055–1060.
Ivancevic, A.M., Kortschak, R.D., Bertozzi, T., Adelson, D.L., 2016. LINEs between species: evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. Genome Biol. Evol. 8, 3301–3322.
Katoh, K., Misawa, K., Kuma, K.i., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.
Kleine, T., Maier, U.G., Leister, D., 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu. Rev. Plant Biol. 60, 115–138.
Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr. Genet. 46, 123–139.
Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sünkel, S., Ullrich, H., Brennicke, A., 1996. copia-, gypsy-and LINE-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. Genetics 142, 579–585.
Koenen, E.J., Ojeda, D.I., Bakker, F.T., Wieringa, J.J., Kidner, C., Hardy, O.J., Pennington, R.T., Herendeen, P.S., Bruneau, A., Hughes, C.E., 2020a. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) mass extinction event. Syst. Biol. syaa 41.
Koenen, E.J., Ojeda, D.I., Steeves, R., Migliore, J., Bakker, F.T., Wieringa, J.J., Kidner, C., Hardy, O.J., Pennington, R.T., Bruneau, A., 2020b. Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. New Phytol. 225, 1355–1369.
Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinf. 7, 474.
Kojima, K.K., 2010. Different integration site structures between L1 protein-mediated retrotransposition in cis and retrotransposition in trans. Mob. DNA. 1, 17.
Kovar, L., Nageswara-Rao, M., Ortega-Rodriguez, S., Dugas, D.V., Straub, S., Cronn, R., Strickler, S.R., Hughes, C.E., Hanley, K.A., Rodriguez, D.N., 2018. PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. Genome Biol. Evol. 10, 2501–2517.
Kreplak, J., Madoui, M.-A., Cápal, P., Novák, P., Labadie, K., Aubert, G., Bayer, P.E., Gali, K.K., Syme, R.A., Main, D., 2019. A reference genome for pea provides insight into legume genome evolution. Nature Genet. 51, 1411–1422.
Kubiak, M.R., Makałowska, I., 2017. Protein-coding genes' retrocopies and their functions. Viruses 9, 80.
Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst. Biol. 54, 575–594.
Legume Phylogeny Working Group (LPWG), 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66, 44–77.
Lehmann, K., Schmidt, U., 2003. Group II introns: structure and catalytic versatility of large natural ribozymes. Crit. Rev. Biochem. Mol. Biol. 38, 249–303.
Lewis, G.P., Schrire, B., Mackinder, B., Lock, M., 2005. Legumes of the World. Royal Botanic Gardens, Kew Richmond, UK.
Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44, W54–W57.
Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45, D200–D203.
Maréchal, A., Brisson, N., 2010. Recombination and the maintenance of plant organelle genome stability. New Phytol. 186, 299–317.
Maréchal, A., Parent, J.-S., Véronneau-Lafortune, F., Joyeux, A., Lang, B.F., Brisson, N., 2009. Whirly proteins maintain plastid genome stability in *Arabidopsis*. PNAS 106, 14693–14698.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 gateway computing environments workshop (GCE). IEEE, pp. 1–8.
Moenne, A., Bégu, D., Jordana, X., 1996. A reverse transcriptase activity in potato mitochondria. Plant Mol. Biol. 31, 365–372.
Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., Moran, J.V., 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nature Genet. 31, 159–165.
Mower, J.P., Jain, K., Hepburn, N.J., 2012. The role of horizontal transfer in shaping the plant mitochondrial genome. Adv. Bot. Res. 63, 41–69.
Negruk, V., 2013. Mitochondrial genome sequence of the legume *Vicia faba*. Front. Plant Sci. 4, 128.
Nerva, L., Vigani, G., Di Silvestre, D., Ciuffo, M., Forgia, M., Chitarra, W., Turina, M., 2019. Biological and molecular characterization of *Chenopodium quinoa* mitovirus 1 reveals a distinct small RNA response compared to those of cytoplasmic RNA viruses. J. Virol. 93, e01998-1918.
Nibert, M.L., Vong, M., Fugate, K.K., Debat, H.J., 2018. Evidence for contemporary plant mitoviruses. Virology 518, 14–24.
Noll, A., Raabe, C.A., Churakov, G., Brosius, J., Schmitz, J., 2015. Ancient traces of tailless retropseudogenes in therian genomes. Genome Biol. Evol. 7, 889–900.
Ohshima, K., 2012. Parallel relaxation of stringent RNA recognition in plant and mammalian L1 retrotransposons. Mol. Biol. Evol. 29, 3255–3259.
Piégu, B., Bire, S., Arensburger, P., Bigot, Y., 2015. A survey of transposable element classification systems–a call for a fundamental update to meet the challenge of their diversity and complexity. Mol. Phylogenet. Evol. 86, 90–109.
Qiu, Y.-L., Palmer, J.D., 2004. Many independent origins of trans splicing of a plant mitochondrial group II intron. J. Mol. Evol. 59, 80–89.
Ren, L., Huang, W., Cannon, S.B., 2019. Reconstruction of ancestral genome reveals chromosome evolution history for selected legume species. New Phytol. 223, 2090–2103.
Richardson, A.O., Palmer, J.D., 2006. Horizontal gene transfer in plants. J. Exp. Bot. 58, 1–9.
Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J., Palmer, J.D., 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol. 11, 29.
Rincão, M.P., Carvalho, M.C. d. C.G. d., Nascimento, L.C., Lopes-Caitar, V.S., Carvalho, K. d., Darben, L.M., Yokoyama, A., Carazzolle, M.F., Abdelnoor, R.V., Marcelino-Guimarães, F.C., 2018. New insights into Phakopsora pachyrhizi infection based on transcriptome analysis in planta. Genet. Mol. Biol. 41, 671–691.
Sabir, J., Schwarz, E., Ellison, N., Zhang, J., Baeshen, N.A., Mutwakil, M., Jansen, R., Ruhlman, T., 2014. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol. J. 12, 743–754.
Sanchez-Puerta, M.V., Edera, A., Gandini, C.L., Williams, A.V., Howell, K.A., Nevill, P.G., Small, I., 2019. Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). Mol. Phylogenet. Evol. 132, 243–250.
Sanchez-Puerta, M., García, L.E., Wohlfeiler, J., Ceriotti, L.F., 2017. Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. New Phytol. 214, 376–387.
Sánchez Puerta, M.V., 2014. Involvement of plastid, mitochondrial and nuclear genomes in plant-to-plant horizontal gene transfer. Acta. Soc. Bot. Pol. 83, 317–323.
Schmitz, J., Churakov, G., Zischler, H., Brosius, J., 2004. A novel class of mammalian-specific tailless retropseudogenes. Genome Res. 14, 1911–1915.
Schuster, G., Stern, D., 2009. RNA polyadenylation and decay in mitochondria and chloroplasts. Prog. Molec. Biol. Transl. Sci. 85, 393–422.
Schwarz, E.N., Ruhlman, T.A., Sabir, J.S., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Bailey, C.D., Jansen, R.K., 2015. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. J. Syst. Evol. 53, 458–468.
Seol, J.-H., Shim, E.Y., Lee, S.E., 2018. Microhomology-mediated end joining: Good, bad and ugly. Mutat. Res-Fundam. Mol. Mech. Mutagen. 809, 81–87.
Shi, Y., Liu, Y., Zhang, S., Zou, R., Tang, J., Mu, W., Peng, Y., Dong, S., 2018. Assembly and comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* 'JinhuaiJ2'. PLoS ONE 13, e0202485.
Sinn, B.T., Barrett, C.F., 2020. Ancient mitochondrial gene transfer between fungi and the orchids. Mol. Biol. Evol. 37, 44–57.
Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4, 1581–1590.
Small, I., Suffolk, R., Leaver, C., 1989. Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell 58, 69–76.
Smith, D.R., 2020. Can green algal plastid genome size be explained by DNA repair mechanisms?. Genome Biol. Evol. 12, 3797–3802.
Sperschneider, J., Catanzariti, A.-M., DeBoer, K., Petre, B., Gardiner, D.M., Singh, K.B., Dodds, P.N., Taylor, J.M., 2017. LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Sci. Rep. 7, 44598.
Spitzer, M., Wildenhain, J., Rappsilber, J., Tyers, M., 2014. BoxPlotR: a web tool for generation of box plots. Nat. Methods 11, 121.
Stai, J.S., Yadav, A., Sinou, C., Bruneau, A., Doyle, J.J., Fernández-Baca, D., Cannon, S.B., 2019. *Cercis*: a non-polyploid genomic relic within the generally polyploid legume family. Front. Plant Sci. 10, 345.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stamos, J.L., Lentzsch, A.M., Lambowitz, A.M., 2017. Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. Mol. Cell 68, 926–939.

Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichiyanagi, K., Adachi, N., Koyama, H., Takeda, S., Okada, N., 2009. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. PLoS Genet. 5, e1000461.
Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., Greiner, S., 2017. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45, W6–W11.
van de Lagemaat, L.N., Gagnier, L., Medstrand, P., Mager, D.L., 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15, 1243–1249.
Wahleithner, J.A., MacFarlane, J.L., Wolstenholme, D.R., 1990. A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial *nad1* gene. PNAS 87, 548–552.
Wolfe, K.H., Li, W.-H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. PNAS 84, 9054–9058.
Wu, Z., Waneka, G., Sloan, D.B., 2020. The tempo and mode of angiosperm mitochondrial genome divergence inferred from intraspecific variation in *Arabidopsis thaliana*. G3-Genes Genomes Genet. 10, 1077–1086.
Xia, H., Zhao, W., Shi, Y., Wang, X.-R., Wang, B., 2020. Microhomologies are associated with tandem duplications and structural variation in plant mitochondrial genomes. Genome Biol. Evol. 12, 1965–1974.
Yahara, T., Javadi, F., Onoda, Y., de Queiroz, L.P., Faith, D.P., Prado, D.E., Akasaka, M., Kadoya, T., Ishihama, F., Davies, S., 2013. Global legume diversity assessment: concepts, key indicators, and strategies. Taxon 62, 249–266.
Zhang, R., Jin, J.-J., Moore, M.J., Yi, T.-S., 2019. Assembly and comparative analyses of the mitochondrial genome of *Castanospermum australe* (Papilionoideae, Leguminosae). Aust. Syst. Bot. 32, 484–494.
Zhang, R., Wang, Y.-H., Jin, J.-J., Stull, G.W., Bruneau, A., Cardoso, D., De Queiroz, L.P., Moore, M.J., Zhang, S.-D., Chen, S.-Y., 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. Syst. Biol. 69, 613–622.
Zimmerly, S., Guo, H., Perlman, P.S., Lambowltz, A.M., 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82, 545–554.
Zimmerly, S., Semper, C., 2015. Evolution of group II introns. Mob. DNA. 6, 7.
Zingler, N., Willhoeft, U., Brose, H.-P., Schoder, V., Jahns, T., Hanschmann, K.-M.O., Morrish, T.A., Löwer, J., Schumann, G.G., 2005. Analysis of 5′ junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5′-end attachment requiring microhomology-mediated end-joining. Genome Res. 15, 780–789.
Zumkeller, S., Gerke, P., Knoop, V., 2020. A functional twintron, 'zombie' twintrons and a hypermobile group II intron invading itself in plant mitochondria. Nucleic Acids Res. 48, 2661–2675.