

RNA splicing of a long *cox2i691* and the evolution of *cox2* introns in angiosperm mitochondria

Alejandro A. Edera (✉ aedera@sinc.unl.edu.ar)

Universidad Nacional del Litoral <https://orcid.org/0000-0001-9864-3159>

Anna V. Williams

The University of Western Australia

Katharine A. Howell

The University of Western Australia

Paul G. Nevill

The University of Western Australia

Ian Small

ARC Centre of Excellence in Plant Energy Biology

M. Virginia Sanchez-Puerta

Instituto de Biología Agrícola de Mendoza

Research Article

Keywords: NA splicing, intron elongation, localized retroprocessing, *cox2i691*, C-to-U RNA editing, RNAseq

Posted Date: November 29th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2273528/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

Table 1 is available in the Supplementary Files section.

RNA splicing of a long *cox2i691* and the evolution of *cox2* introns in angiosperm mitochondria

Alejandro A. Edera^{1*}, Anna V. Williams^{2,3,4}, Katharine A. Howell², Paul G. Nevill^{3,4}, Ian Small^{2,5} and M. Virginia Sanchez-Puerta^{6,7}

¹ Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL, CONICET, Ciudad Universitaria UNL, 3000, Santa Fe, Argentina.

² Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, Western Australia, Australia.

³ Botanic Gardens and Parks Authority, Kings Park and Botanic Garden, Fraser Avenue, Kings Park, Western Australia, Australia.

⁴ School of Plant Biology, The University of Western Australia, Crawley, Western Australia, Australia.

⁵ Centre of Excellence in Computational Systems Biology, The University of Western Australia, Crawley, Western Australia, Australia.

⁶ IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, M5528AHB Chacras de Coria, Argentina.

⁷ Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, 5500 Mendoza, Argentina.

* corresponding author: aedera@sinc.unl.edu.ar

AAE: <https://orcid.org/0000-0001-9864-3159>

AVW: <https://orcid.org/0000-0003-0349-4248>

KAH: <https://orcid.org/0000-0001-6498-0472>

PGN: <https://orcid.org/0000-0001-8238-0534>

IS: <https://orcid.org/0000-0001-5300-1216>

MVSP: <https://orcid.org/0000-0003-2511-5093>

Word counts: 6987

Number of figures and tables: 6 and 1, respectively.

Supporting information: 6 figures, 5 tables and 3 datasets.

Abstract

In angiosperms, the mitochondrial *cox2* gene harbors up to two introns, commonly referred to as *cox2i373* and *cox2i691*. Although each of these introns is ~1.5 kb in length, a *cox2i691* of 4.2 kb has been reported in *Acacia ligulata* (Fabaceae). It is still unclear whether its unusual length is due to a *trans*-splicing arrangement, the loss of functionality of the interrupted *cox2*, or a rare elongation of a *cis*-spliced intron. Here, we tried to shed light on these hypotheses by thoroughly analyzing short-read RNA sequencing with a multi-step computational strategy. The results revealed that the *Acacia cox2* is very likely functional and its long intron is spliced in *cis* in a very efficient manner despite its length. To better characterize how conserved *cox2i691* is among angiosperms, we studied 222 fully-sequenced mitogenomes from 30 plant orders. Unlike *cox2i373*, our results point out that *cox2i691* shows a distribution among plants that is shaped by frequent intron loss events driven by localized retroprocessing. In addition, *cox2i691* undergoes episodic elongations that are taxonomically scattered and poorly related to genome-wide repeat content. Such elongations were found to occur within domain IV and seem to be due to the incorporation of exogenous DNA. Surprisingly, we found that *cox2i691* is erroneously annotated as absent in 30 mitogenomes deposited in public databases.

Keywords: RNA splicing, intron elongation, localized retroprocessing, *cox2i691*, C-to-U RNA editing, RNAseq.

Key message

RNA splicing was found to be efficient for a long group II intron whose length is due to an insertion that is frequently reported within domain IV in diverse angiosperm orders.

53 Introduction

54 Mitochondrial genomes carry a small number of essential genes well known for their roles in energy
 55 production through oxidative phosphorylation (Møller *et al.*, 2021). Some of these genes can be interrupted by
 56 highly structured non-coding DNA called introns. To yield translatable, mature mRNAs, the coding regions (or
 57 exons) interrupted by introns within precursor mRNAs (pre-mRNA) must be ligated after the removal of the
 58 introns in a post-transcriptional process called RNA splicing (Bonen, 2012). According to their structural
 59 features, plant organellar introns can be classified as either group I or group II (Bonen, 2008), where the latter
 60 group is distinguished by the presence of six helical domains (Bonen, 2012). Group II introns are predominant
 61 in plant mitochondria (Brown *et al.*, 2014) and can be either spliced in a *cis* or *trans* manner (Glanz & Kück,
 62 2009). Unlike *cis* introns, *trans* introns are encoded in DNA as independent fragments whose hybridization after
 63 transcription restores their full-length for enabling their splicing..

64 In plant mitochondria, intron content usually varies among plant species (Alverson *et al.*, 2011; Guo *et al.*,
 65 2020) mainly due to intron or gene loss (Skippington *et al.*, 2015; Cuenca *et al.*, 2016). In addition, shifts from
 66 *cis* to *trans* splicing have been reported for several introns (Qiu & Palmer, 2004; Kim & Yoon, 2010; Hecht *et al.*,
 67 2011; Petersen *et al.*, 2015; Skippington *et al.*, 2015; Guo *et al.*, 2016; Yu *et al.*, 2021). Particularly,
 68 angiosperms show variable intron content in the *cox2* gene (Park *et al.*, 2015; Cuenca *et al.*, 2016), which
 69 encodes subunit 2 of cytochrome c oxidase, a highly conserved enzyme involved in energy metabolism in all
 70 respiring eukaryotes (Soto *et al.*, 2012). This gene can have up to two non-homologous group II introns (Lippok
 71 *et al.*, 1992), which are located at two well-conserved positions (splicing sites), commonly referred to as
 72 *cox2i373* and *cox2i691* (Qiu *et al.*, 1998). Whereas *cox2i373* has been relatively well studied in angiosperms
 73 (Qiu *et al.*, 1998; Joly *et al.*, 2001; Kudla *et al.*, 2002), information about *cox2i691* is still fragmentary (Rabbi &
 74 Wilson, 1993; Qiu *et al.*, 1998; Sloan *et al.*, 2010; Hepburn *et al.*, 2012; Cuenca *et al.*, 2016; Mower, 2020).

75 Angiosperm *cox2* introns also show variability in length generally due to insertion/deletions within domain
 76 IV (Bonen, 2008). On average, both *cox2* introns show a length of around 1.5 kb, but some striking exceptions
 77 have been reported. For example, the 11.4 kb *cox2i373* intron of *Nymphaea colorata* (Dong *et al.*, 2018)
 78 represents the longest intron reported to date in angiosperm organelles. Nevertheless, such a peculiar length
 79 deviation is not exclusive to *cox2i373*. For example, a 4.2 kb *cox2i691* intron has been reported for *Acacia*
 80 *ligulata* (Sanchez-Puerta *et al.*, 2019), a member of the subfamily Caesalpinioideae within the Fabaceae (Azani
 81 *et al.*, 2017). Since such introns have not yet been subject to thorough transcriptomic analyses, it is not clear
 82 whether their unusual lengths are due to a *trans*-splicing arrangement or a functionality loss of the interrupted
 83 genes.

84 To shed light on the splicing activity of long introns, we investigated the splicing of *Acacia ligulata*
 85 *cox2i691* and the functionality of *cox2* by analyzing short-read RNA sequencing data with a multi-step
 86 computational strategy. In addition, we collected 225 *cox2* sequences from 30 orders of angiosperms to study
 87 the evolution of *cox2i691*. During this process, we found that *cox2i691* was erroneously annotated as absent in
 88 30 mitogenomes obtained from public databases.

Materials and methods

RNA sequencing

Total RNA was isolated from leaves of a seedling of *Acacia ligulata* grown in Kings Park, Perth, Australia, as previously described (Williams *et al.*, 2015). Extracted RNA was depleted of rRNA and sequenced with Illumina technology. Library preparation was performed with 300 ng of total RNA using the TruSeq Stranded Total RNA LT sample preparation kit and Plant rRNA Removal Mix (Illumina, San Diego, USA), following the manufacturer's directions. The library was clustered on a Rapid Flow Cell v2 (Illumina), using the HiSeq Rapid PE Cluster Kit v2 (Illumina) and on-instrument cluster generation on the HiSeq1500 platform, and sequenced using the HiSeq Rapid SBS Kit v2. A total of 148,319,698 100 nt read pairs (forward & reverse reads) were obtained. Raw RNA-seq data were deposited in the Sequence Read Archive under accession number PRJNA869960. Raw data was processed to trim adapters and low-quality bases using fastp with default options (Chen *et al.*, 2018). A total of 140,710,359 read pairs were obtained with a median per-base sequencing quality of Q38 (QUAL field, SAM specification v1).

Multi-step computational strategy and splicing efficiency

To identify *Acacia* RNA read pairs derived from intron-containing or intronless mRNAs, the RNA sequencing data were aligned progressively through three consecutive steps, each performed using Bowtie v2.3.4.3 with the following options: --end-to-end --very-sensitive --no-discordant --no-mixed --fr --nofw -D 30 -R 10 -L 18. The first step was designed to exclude plastid contaminants, by discarding those read pairs that were successfully mapped to the whole *Acacia* plastid genome (NC_026134) or to its spliced gene sequences. The unmapped reads obtained from the previous step were then used by the second step to perform a new read alignment using the whole *Acacia* mitogenome (NC_040998) as reference. The read pairs that were successfully aligned by this second step were further processed to categorize them as derived from intron-containing mRNAs, i.e., when they were mapped to intron regions or exon-intron/intron-exon junctions. However, since reads are short, the latter step is unable to align read pairs to consecutive exons or exon-exon junctions. This prompted us to design a third step in which the read pairs that were unmapped by the second step were used to perform a new read alignment using the spliced versions of the intron-containing genes as reference. The read pairs that were successfully aligned by this third step were categorized as derived from intronless mRNAs. Finally, to reduce miscategorizations, the reads aligned by the second step close to exon-intron or intron-exon junctions were re-aligned to the same spliced references used by the third step. If these re-aligned read pairs obtained better mapping scores (MAPQ field in the SAM file) in comparison to the scores assigned by the second step, they were finally categorized as derived from intronless mRNAs.

The RNA splicing efficiency of each intron was calculated as: $\text{spliced} / (\text{spliced} + (\text{unspliced } 5' + \text{unspliced } 3') / 2)$, where unspliced and spliced were the average number of reads categorized as intronless and intron-containing, respectively. Moreover, we calculated the conservation of the *Acacia* introns by performing BLASTn 2.5.0+ searches (Altschul *et al.*, 1997) against the non-redundant nucleotide database at NCBI GenBank.

C-to-U RNA editing identification and protein domain predictions

To identify C-to-U editing sites, the reads aligned by the second and third steps were pooled and used to infer the transcriptomic composition of each cytidine within each mitochondrial gene. By following a previous protocol (Wu *et al.*, 2015), a cytidine was identified as edited if its transcriptomic composition was both supported by at least 30 reads and showed >10% of uridines. For each identified site, its editing efficiency (editing extent) was defined as the relative number of uridines in its transcriptomic composition. Identified sites were assessed by scores obtained from both PREP-Mt (Mower, 2009) and evolutionary conservation estimations previously reported that were based on sequencing data (Edera *et al.*, 2018). In addition, to infer the functional role of the residues affected by the editing of the sites identified within *cox2*, we used the amino acids derived from the *cox2* coding regions to predict protein domains and protein topology with HMMER and TMHMM (Krogh *et al.*, 2001; Finn *et al.*, 2011), respectively.

Identification of transcripts from *cox2* and *rpl2* nuclear copies

Since mitochondrial genes can be functionally transferred to the nucleus (Covello & Gray, 1992; Adams *et al.*, 1999), we looked for nuclear transcripts from these genes in the Acacia transcriptome. To this aim, a *de novo* RNA assembly was performed with SPAdes v3.11.1 with default parameters (Prjibelski *et al.*, 2020). The resulting assembly was screened through BLASTn and tBLASTx 2.5.0+ searches, using as queries Acacia *cox2* and *rpl2* as well as soybean nuclear *cox2*. As a control for the presence of nuclear transcripts in the transcriptome, we also screened the assembly using nuclear housekeeping genes.

Nucleotide sequence alignments and phylogenetic trees of *cox2*

For evolutionary analyses, *cox2* gene sequences were collected from fully-sequenced mitogenomes of 30 angiosperm orders obtained from NCBI GenBank. Genomic coordinates of *cox2* genes were inferred by performing BLASTn 2.5.0+ searches (task=short and word_size=7). Duplicated *cox2* genes inferred within the same genome were excluded when showing no differences. *cox2* annotation errors were identified when the inferred gene coordinates differ from those publicly reported at NCBI GenBank. The coding regions and introns of the collected *cox2* gene sequences were aligned with MAFFT v.7.017 (Katoh *et al.*, 2019), using default options. Poorly aligned regions were manually adjusted and poorly conserved intronic regions were removed with trimAl v1.2 (Capella-Gutiérrez *et al.*, 2009). From the resulting alignments (Dataset S1-3), phylogenetic analyses were performed using RAxML v8.2 with maximum likelihood, a GTR+gamma model and 1,000 bootstrapping replicates (Stamatakis, 2014). For the coding regions of *cox2*, phylogenetic inferences excluded editing sites. In addition, the repeat content of each intron was estimated from self-BLASTn hits with sequence identities of at least 90%.

Inferring ancestral state of *cox2* introns and intron loss through retroprocessing

To study the evolution of *cox2i373* and *cox2i691*, a binary character was defined for each intron describing its absence (0) or presence (1) in each collected *cox2* sequence. These binary characters were then used to reconstruct ancestral intron states with the R package Ape (Paradis & Schliep, 2019). Reconstructions used a phylogenetic tree of angiosperms, obtained with V.PhyloMaker2 (Jin & Qian, 2019), and a discrete model with different transition rates (ARD) estimated by a joint likelihood function.

Since intron loss is frequently accompanied by cytidine-to-thymidine substitutions at editing sites (Cuenca *et al.*, 2016), we identified such substitutions or losses within the collected *cox2* sequences. To this aim, we used *cox2* positions confidently predicted (score>0.6) by PREP-Mt as editable. They were then mapped to the *cox2* alignment. A total of 38 positions (columns in the alignment) showed at least one predicted editing site. Loss events were identified as thymidines present in those columns. To avoid overestimation, evolutionarily recent losses were only considered, excluding vertically-inherited thymidines that were inferred through ancestral state reconstructions. Following the same protocol employed for introns, these reconstructions treated each column as a binary character (C or non-C), where the C state was assumed as potentially editable. A thymidine was inferred as vertically inherited when it showed a non-C as an ancestral state in: basal angiosperms, monocots, basal eudicots, superrosids or superasterids. Otherwise, a thymidine was inferred as an evolutionarily recent loss. Finally, for each *cox2* sequence, the number of inferred loss events was normalized by dividing it by 38. The 85th percentile calculated from these normalized numbers was used as a threshold to identify *cox2* sequences with significant losses of editing sites.

cox2i691 structure prediction

To predict the secondary structure of *cox2i691*, we used the shortest sequence of *cox2i691* that corresponds to *Cercis canadensis*, which is found among the collected *cox2* sequences. This sequence was folded using RNALfold v2.5 with default options. The stability of the predicted helical domains was assessed by performing domain-specific predictions using sequences spanning each domain and its flanking sequences with increasing lengths. This assessment showed that the domain III-VI were predicted stably regardless of how long the

flanking sequences were. Later, the *Cercis canadensis* *cox2i691* sequence of the predicted structure was replaced by the consensus sequence calculated from the nucleotide alignment of *cox2i691*. Finally, the level of conservation of each base in the consensus sequence, ranging from 0 (min) to 1 (max), was calculated from the same nucleotide alignment.

Results

The *cox2* gene of *Acacia* is efficiently transcribed and presents a *cis*-splicing intron

Approximately 3.2 million, high quality, paired-end RNA reads were aligned to the mitochondrial genome of *Acacia*. Of these, 87% aligned to the 38 protein-coding genes, 22 tRNAs, and three rRNAs. In comparison to the intergenic mitochondrial regions, the average read depth of each coding region was substantially higher (Table S1). Both *cox2* exons exhibited elevated RNA levels (Fig. 1A) and a low but continuous coverage on the 4.2 kb *cox2* intron, which is expected due to the fast degradation of introns (Grimes *et al.*, 2014).

Because introns can shift from *cis*- to *trans*-splicing (Kim & Yoon, 2010; Skippington *et al.*, 2015; Guo *et al.*, 2016; Yu *et al.*, 2021), we evaluated whether it explains the unusual length of the *Acacia cox2* intron. To this aim, we analyzed the read depth and read coverage on all *Acacia* introns. We found that the lengthy *Acacia cox2i691* is *cis*-splicing, because it is contiguously covered by RNA-seq reads (Fig. 1A) as the other 18 *cis*-splicing *Acacia* introns (Fig. S1), contrasting with the interrupted coverages observed on the *trans*-splicing introns (Fig. S2). In addition, we estimated the length of each *trans*-spliced intron fragment based on two features: RNA read depth and intron sequence conservation across angiosperms. For 60% of the *trans*-spliced introns, both features agreed on the estimated lengths. These estimations showed that, in contrast to the *cis*-splicing introns, the sequence contiguous to each *trans*-spliced intron end displayed long regions with none or very low read depths. These results provide support for the *cis*-splicing status of the *cox2i691* in *Acacia*.

The *Acacia cox2* intron is spliced efficiently

To evaluate whether the *Acacia cox2* gene is spliced, we categorized the read pairs aligned to it as derived from either intron-containing or intronless mRNAs according to where they were mapped to. Read pairs were categorized as intron-containing read pairs when one of the mates was mapped to one of the exons while the other mate was mapped to the intron. In contrast, read pairs were categorized as intronless when both mates were not mapped to the same exon. Read pairs were left as uncategorized when they did not meet the previous criteria. This categorization revealed that the two categories of read pairs showed differential read depths (Fig. 1B). In contrast to the intron-containing read pairs (red filled curve in Fig. 1B), the read depth of the intronless pairs was relatively higher and formed a Gaussian-like shape over the exon-exon junction (blue filled curve in Fig. 1B). In addition, uncategorized read pairs (gray filled curve) were relatively underrepresented at the exon-exon junction, supporting the effectiveness of the proposed read pair categorization for studying intron splicing. Altogether, these results demonstrate that the *Acacia cox2* is being spliced.

By employing the read pair categorization mentioned above, we were also able to calculate the splicing efficiency for the *cox2* intron as well as for the other *Acacia* mitochondrial introns. The obtained splicing efficiencies were in agreement with previous estimations in *Arabidopsis thaliana* (Garcia & Sanchez-Puerta, 2021). A comparative analysis showed that the *cox2* intron was ranked among the most efficient splicings in *Acacia* (Table 1). This suggests that RNA splicing is still catalyzed efficiently even when the intron exhibits an unusual length, which is in line with previous findings in nuclear-encoded introns in plants (Chang *et al.*, 2017). In addition, we observed that the *rpl2* intron had the lowest splicing efficiency in *Acacia*. The read depth of the *rpl2* intron showed three peaks close to the 5' end (Fig. S1), but a further evaluation showed that they are not mapping artifacts nor multi-mapping read pairs produced by repetitive or homologous regions. Moreover, the *Acacia* transcriptome showed no transcripts derived from other *rpl2* copies, which rules out the possibility that the observed low splicing efficiency of *rpl2* is due to the expression of *rpl2* copies functionally transferred to the nucleus (Adams *et al.*, 2001). Despite displaying a poor splicing efficiency, we hypothesize that *rpl2* is very likely functional, because the number of intronless *rpl2* mRNAs is similar to those derived from

the *rpl2* exon 1 (Figure S3), and a similar splicing efficiency was observed in the mitochondrial genome of *A. thaliana* (Garcia & Sanchez-Puerta, 2021).

The categorization of read pairs also showed that, unlike the intronless read pairs, the intron-containing read pairs were asymmetrically mapped around the exon-exon junction of *cox2* (red-filled curves in Fig. 1B). This asymmetry shows that the intron-exon junction is overrepresented in comparison to the exon-intron junction (Table 1). We observed a similar pattern in the other *Acacia* introns (Fig. S3). We interpreted such a pattern as favoring a model where the 5' exon is released first (Li-Pook-Than & Bonen, 2006). However, we found that such intron release is not the only one operating in the *Acacia* mitochondria since other introns showed their exon-intron junctions overrepresented (i.e., *nad1-i1*, *nad4-i1*, *nad4-i2* and *nad7-i3*). This latter pattern fits better with a model where the 3' exon is cleaved before the 5' exon. This intron release model has been previously reported. It is a minor pathway in flowering plant mitochondria (Li-Pook-Than & Bonen, 2006) and is known as the spliced exon reopening pathway (Jarrell *et al.*, 1988; Murray *et al.*, 2001; Laroche *et al.*, 2021).

The *Acacia cox2* gene is functional

In plant mitochondria, almost all mRNAs are post-transcriptionally C-to-U edited to enable the production of functional proteins (Small *et al.*, 2020). Therefore, one common approach to assess the functionality of mitochondrial genes is by analyzing how efficient they are C-to-U edited (Aubert *et al.*, 1992; Adams *et al.*, 1999; Handa, 2003; Mower & Palmer, 2006). We used such an approach to verify the functionality of the *Acacia cox2* gene. To this aim, we first conducted a genome-wide identification of C-to-U editing sites and then we studied the sites that were identified within *cox2*. We identified a total of 637 C-to-U editing sites within the 38 protein-coding genes of the *Acacia* mitogenome (Table S2). We assessed that this identification was reliable by corroborating that the identified sites showed the typical codon preference (Giegé & Brennicke, 1999), in which the editing of 90% of the sites results in amino acid changes. In addition, we evaluated the conservation of the sites: 90% of the identifications were moderately or strongly conserved in other plants, as indicated by PREP-Mt and conservation scores > 0.4.

Within the *cox2* gene, we identified 12 non-synonymous editing sites that were efficiently edited as indicated by their editing extents (Table S2). To understand the impact of these sites on rendering functional protein products, we mapped them to the amino acid sequence deduced from the *cox2* coding regions and analyzed their positions with respect to two protein domains that we found through HMM-based predictions (Fig. S4A). Four out of the 12 editing sites were within a predicted transmembrane domain (PF02790). Since RNA editing increases protein hydrophobicity (Jobson & Qiu, 2008), these four sites might be contributing to the conformation of the transmembrane helices and thus they might be modulating COX2 anchoring. The other eight sites were within a domain predicted as periplasmic (PF00116), which is exposed to the mitochondrial intermembrane space and is believed to play a crucial role in binding soluble cytochrome c (Timón-Gómez *et al.*, 2018). Notably, experimental evidence found by previous reports indicates that both protein domains are relevant for protein functionality, since the disruption of four of the 12 identified sites (164, 256, 547 and 701) leads to protein function loss (Kim *et al.*, 2009; Dahan *et al.*, 2014; Qi *et al.*, 2017). Furthermore, a comparative analysis revealed that the editing of eight of the 12 identified sites restores residues that are well conserved among phylogenetically distant species (Fig. S4B). Altogether, these results point out that the *cox2* is functional since a complex biological process such as RNA editing is acting on *cox2* transcripts very precisely to enable the production of functionally competent *cox2* proteins.

Later, we evaluated whether *Acacia* has nuclear-encoded *cox2* copies, whose presence could be reducing selection pressure on the mitochondrial *cox2*. Such nuclear copies may be the result of the mitochondrial *cox2* gene transferred to the nucleus, which has been frequently observed in other legumes (Covello & Gray, 1992). Analysis of the *Acacia* transcriptome found no evidence for transcripts derived from nuclear-encoded *cox2* copies. This indicates that, if a nuclear-encoded *cox2* copy exists, it is not transcribed and thus the mitochondrial-encoded *cox2* gene is under the pressure of supplying functional proteins involved in plant energy production.

RNA editing occurs before intron splicing in *Acacia cox2* transcripts

To additionally assess the role of RNA editing in protein functionality, we explored how fast RNA editing was in comparison to RNA splicing in the *Acacia cox2* transcripts. To this aim, we calculated the editing extents from the intronless read pairs mapped to the *Acacia cox2* gene and compared them with their counterparts calculated from the intron-containing read pairs.

Due to the RNAseq library size, we were able to calculate both editing extents in a reliable way for only two editing sites at positions: 701 and 724, which were four and 21 nucleotides downstream and upstream, respectively, from the splicing point. Both sites were found to be efficiently edited (editing odd=0.04-0.13) in both intron-containing and intronless read pairs (Table S3). Thus, this result shows that RNA editing is carried out efficiently before the splicing of *cox2*, preventing the synthesis of functionally incompetent proteins. Similar results have been previously observed in other plants (Sutton *et al.*, 1991; Yang *et al.*, 1991; Hepburn *et al.*, 2012). Despite observing a similar pattern in the timing between both post-transcriptional processes in the editing sites of other *Acacia* intron-containing genes, we found five non-synonymous editing sites where RNA editing was more efficient after RNA splicing. Interestingly, three of these five sites have been previously reported for wheat mitochondria (Li-Pook-Than *et al.*, 2007), suggesting that such timing between the two post-transcriptional processes might be a conserved feature for some sites across plants.

cox2i691 is frequently overlooked in the gene annotations of angiosperm mitogenomes

To study the distribution of *cox2i373* and *cox2i691* across angiosperms, we collected a total of 222 fully-sequenced mitogenomes from 30 plant orders. We assessed the correctness of their *cox2* annotations since it has been reported that the *cox2* gene from *Leucaena trichandra* (Kovar *et al.*, 2018) was erroneously annotated as lacking *cox2i691* (Choi *et al.*, 2019) despite being virtually identical to the *Acacia* counterpart (Fig. S5). Surprisingly, we found that 30 of the collected genomes had their *cox2* genes annotated as lacking *cox2i691* despite the fact that the intron and exon sequences were present in the genomes (Table S4). This recurrent annotation error very likely responds to the fact that the presence of *cox2i691* results in a *cox2* gene containing a short 3' exon (~80bp) that is easily missed by default BLASTn searches. We corrected the *cox2* annotations by performing BLASTn searches and then aligning the resulting *cox2* gene sequences (Dataset S1-3). Except for two highly divergent sequences from the genus *Viscum* (Petersen *et al.*, 2015; Skippington *et al.*, 2015), we observed that the *cox2* sequences were remarkably conserved not only along the full-length of the coding regions but also at both extremes of *cox2i691* (Fig. 2), which supports the presence of the intron.

Multiple independent losses of both *cox2* introns during angiosperm evolution

To gain further insight into the evolution of the *cox2i373* and *cox2i691* introns among angiosperms, we analyzed their distribution across the collected *cox2* sequences (Table S5). Notably, we observed that the distribution of *cox2i691* was substantially different from that of *cox2i373* (p -value=0.37), as indicated by Pagel's binary character correlation test (Pagel, 1994). We next evaluated whether the distribution of each intron was shaped by multiple intron gains through horizontal transfers (Cho *et al.*, 1998; Sanchez-Puerta *et al.*, 2008). To evaluate this, we estimated phylogenetic trees from the whole coding sequence of *cox2* as well as from each of its three potential exonic regions, and from each of the introns *cox2i373* and *cox2i691*. The estimated trees showed no conclusive evidence for intron gains through horizontal transfers (Fig. S6), as the estimated relationships were congruent with the known organismal ones or showed very weak bootstrap support. Therefore, our data support that both *cox2* introns were present in the ancestor of the angiosperms (Qiu *et al.*, 1998; Bonen, 2012) and their distributions were differentially shaped during plant evolution by independent events where one or both introns were lost independently.

To characterize such intron loss events, we inferred the ancestral states of both introns and then we mapped them to the angiosperm phylogeny (Fig. 3). Except for the losses previously reported in the order Alismatales (Cuenca *et al.*, 2016), the resulting mappings showed that *cox2i373* was present in all the basal angiosperms and monocots analyzed. Although *cox2i373* is present in all the sampled Asparagales, previous reports have shown that *cox2i373* is absent in some plants within the family Ruscaceae (Kudla *et al.*, 2002), none of which were included in our sampling. Within the superrosids, *cox2i373* was lost in the ancestor of the rosoid group, which is

in agreement with previous findings (Joly *et al.*, 2001). In contrast, *cox2i373* was frequently present in the superasterids, but absent in the orders of Ericales and Lamiales as well as in the Campanulaceae. Our data allowed us to infer that, during the evolution of the order Caryophyllales, *cox2i373* was lost three times: in the lineages leading to *Fallopia*, *Sesuvium* and the tribe Sileneae. Overall, 10 losses of the *cox2i373* were identified among the >200 species belonging to the 30 angiosperm orders.

Next, we analyzed the inferences for *cox2i691*. First of all, they highlight the impact that the misannotations errors could have had if they were not properly identified (blue circles in Fig. 5). In particular, such errors could have seriously misled the inferences in Caesalpinioideae and Campanulids. Therefore, the curated data collected in this study allowed us to study the loss events of *cox2i691* with great precision. Our results showed that *cox2i691* is present in all basal angiosperms but absent in some monocots (Fig. 5, Table S5), i.e., *Allium cepa* (Amaryllidaceae) and the Poales (Sloan *et al.*, 2010). The intron is also present in all the basal eudicots as well as the early-branching superrosids. In contrast to *cox2i373*, the superrosids present *cox2i691* but the intron was lost in Salicaceae, Papilionoideae (Fabales) and Rosales. Some of these losses have been previously reported (Sloan *et al.*, 2010; Choi *et al.*, 2019). We also observed a recent loss in *Cucumis* (Cucurbitales), as reported previously (Alverson *et al.*, 2011). Our data also showed that *cox2i691* was lost in *Geranium*, which is in agreement with the substantial reduction of genome content that has been observed in the family Geraniaceae (Parkinson *et al.*, 2005; Park *et al.*, 2015). Our data indicated that two copies of *cox2* were present in *Corchorus olitorius* (Malvales), *Prosopis glandulosa* (Fabales) and *Ombrophytum subterraneum* (Balanophoraceae, Santalales). Interestingly, each of these three independent gene duplications included one intronless copy. Our data also allowed us to determine the origin of the horizontally transferred intronless *cox2* gene of *Lophophytum mirabile* (Balanophoraceae) (Sanchez-Puerta *et al.*, 2019). Since all the Caesalpinioideae but *Prosopis glandulosa* present *cox2i691*, this suggests that the *L. mirabile cox2* was horizontally transferred from an ancestor of *Prosopis*. Within the Myrtales, no intron losses were detected but an intronless *cox2* in *Oenothera berteriana* was previously reported (Hiesel & Brennicke, 1983). Within the superasterids, ancestral losses were observed in the Ericales and in the ancestor of Gentianales and Solanales. The order Caryophyllales showed intron losses within the genera *Fallopia* and *Beta* as well as within the tribe Sileneae, whose introns have been very likely lost due to substantial chromosomal variation (Sloan *et al.*, 2010). Recent losses of the *cox2i691* were inferred in *Codonopsis lanceolata* (Campanulaceae) and in *Ajuga reptans* (Zhu *et al.*, 2014). Overall, we inferred that *cox2i691* was lost a total of 16 times during angiosperm evolution.

***cox2* intron loss through retroprocessing**

We tried to determine the molecular mechanisms behind the observed intron losses. Although introns can be removed through multiple mechanisms, such as genomic deletion in which frameshift mutations are generally introduced after introns are removed resulting in potentially non-viable proteins (Pruitt & Hanson, 1989), intron loss by retroprocessing is one of the most frequent in plants (Geiss *et al.*, 1994; Itchoda *et al.*, 2002; Cuenca *et al.*, 2016). Through this mechanism, genes lose their introns through gene conversion with intronless alleles (Geiss *et al.*, 1994; Itchoda *et al.*, 2002; Cuenca *et al.*, 2016). The origin of these intronless alleles can be exogenous, via horizontal gene transfer, or endogenous, via cDNA products obtained from reverse transcription of mature transcripts (Hepburn *et al.*, 2012). Since the phylogenetic trees that we previously estimated show no conclusive evidence for horizontal gene transfers (Fig. S6), we identified intron deletion mediated by endogenous products. For such identification, we used the fact that retroprocessed genes usually show C-to-T substitutions at multiple editing sites (Cuenca *et al.*, 2016). We focused the analysis at 38 *cox2* positions that are commonly edited in angiosperms.

The analysis revealed that the *cox2* genes of at least 30 taxa participated in retroprocessing events (Fig. 4; Table S5). Of them, only nine cases showed a substantial number of C-to-T substitutions but no intron loss. Among them, one of the most striking cases is observed in Rosales. There, *Hemiptelea davidii*, *Cannabis sativa* and *Morus notabilis* showed elevated numbers of C-to-T substitutions (76-88%) radically contrasting with the much lower levels of such substitutions observed in the remaining Rosales (12-24% on average). This indicates that an event of retroprocessing occurred in the common ancestor of the three aforementioned plants after the divergence of *Ziziphus* (Rhamnaceae). A similar event can be also observed in *Senna occidentalis* (59%) and *S. tora* (26%).

On the other hand, the remaining 21 taxa identified as being affected by retroprocessing also exhibited intron deletions (black arrow heads in Fig. 4), eight of which showed to be ancestral events (blue arrows). Notably, in 14 taxa belonging to the families Salicaceae and Caryophyllaceae, intron losses were accompanied by a considerable number of C-to-T substitutions (78-94%), some of which have been previously observed (Sloan *et al.*, 2010; Edera *et al.*, 2018; Brenner *et al.*, 2019). A similar pattern is observed in *Geranium* (Park *et al.*, 2015). Exceptionally, we found that both introns were lost in a single event in the tribe Sileneae (Edera *et al.*, 2018) as well as the genus *Fallopia*. Another interesting case is observed in *Corchorus olitorius*. It has two copies of *cox2* that differ in intron content: one containing *cox2i691*, and the other not only lacking both introns but also displaying a substantial number of C-to-T substitutions. This strongly indicates a case where a spliced mRNA was retroprocessed and then inserted back to the genome such that its intron-containing counterpart remained completely intact, resulting in two gene copies co-existing in the same genome but differing in intron content (Cuenca *et al.*, 2012).

Episodic elongation of *cox2i691* through expansions within domain IV

We also investigated the length distribution of *cox2i373* and *cox2i691* by using the intron sequences contained in the collected *cox2* sequences. This revealed that both introns exhibit contrasting length patterns (Table S5). In general, the length of *cox2i373* was very stable across plants (Fig. 5), except for few exceptions such as the 794 bp and 11.4 kb introns in *Tripsacum* (Poales) and *Nymphaea colorata* (Nymphaeales), respectively. It has been postulated that the long length of the *N. colorata* *cox2i373* is due to the high repetitive content of the mitogenome (Dong *et al.*, 2018), which was also observed in our analyses. In contrast to *cox2i373*, the lengths of *cox2i691* exhibited higher variability, ranging from 732 bp in *Cercis* (Fabaceae) to 5.2 kb in *Scutellaria* (Lamiaceae). Particularly, *cox2i691* showed the highest variability within the Fabales (0.7-4.2 kb), where *Acacia* displays the longest intron. Nevertheless, our results also show some plants harboring *cox2i691* that are even longer than that of *Acacia*: *Scutellaria tsinyunensis* (5.1 kb), *Rothea serrata* (5.1 kb), *Corchorus capsularis* (4.9 kb) and *Corchorus olitorius* (4.4 kb). Note that *R. serrata*, *C. capsularis* and *C. olitorius* introns were identified here in *cox2* genes erroneously annotated as intronless. Along with these extreme cases of intron elongation, the results show that intron length variability is mostly limited to four plant orders: Malpighiales, Fabales, Malvales and Lamiales.

We aimed to identify the location of the intron elongations among the six helical domains commonly found in group II introns. Unlike *cox2i231* (Michel *et al.*, 1989; Bonen, L., 2008), the secondary structure of *cox2i691* has not yet been well studied. Consequently, to determine the domain localization, we predicted the consensus secondary structure of *cox2i691* that shows a central hub from which six helical domains extend (Fig. 6). The two domains at the 3' end were in agreement with the structural features previously reported for domains V and VI of group II introns (Michel *et al.*, 1989). Moreover, the sequences comprising all the predicted domains show a high level of conservation across angiosperms (yellow bases), suggesting that the folding of such sequences is functionally relevant for the splicing of *cox2i691*. Based on the positioning of the predicted domains, we determined where the largest episodic intron elongations occur. All the elongations were observed within domain IV, which is known to be less conserved among plants and is thought to have a minor role in the splicing reaction (Michel *et al.*, 1989; Malek *et al.*, 1997). Our data also suggested the minor functional role of domain IV since it was virtually absent in the *cox2i691* of *Cercis canadensis*, which was the shortest *cox2i691* sequence found in this study.

Although intron elongation could be due to repeat propagation (Guo *et al.*, 2017), we found that repeat content could not explain most of the observed intron elongations (Table S5). Alternatively, the incorporation of exogenous DNA material could be responsible for the increase in intron length (Choi *et al.*, 2019). We tried to corroborate this by performing BLASTn searches against the nucleotide NCBI database using the sequences inserted within domains IV as queries. However, the searches returned no significant hits, preventing us from inferring the origin of such intron insertions.

Discussion

Long introns can be spliced efficiently in plant mitochondria

Several studies have reported the presence of unusually long introns in angiosperm mitochondria (Alverson *et al.*, 2011; Dong *et al.*, 2018), which generally reside within essential genes involved in respiratory pathways. By massively sequencing RNA, we studied a 4.2-kb group II intron within the *cox2* gene (*cox2i691*) from the mitogenome of *Acacia ligulata*, a member of Caesalpinoideae (Fabales). The resulting sequencing data showed to be sufficiently deep and even along *Acacia* introns to identify *trans*-spliced introns in a reliable manner (Dong *et al.*, 2018; Lin *et al.*, 2022). We inferred that the *Acacia cox2i691* is spliced in *cis* and found that the splicing is very efficient despite its length. Therefore, RNA splicing in plant mitochondria can be catalyzed even in the presence of long introns, enabling introns to enlarge without interfering with the expression of the interrupted genes. This capability of RNA splicing explains why there are a number of angiosperms with lengthy mitochondrial introns, such as the disproportionally large ones found in *Cucumis sativus* and *Nymphaea colorata* whose lengths are of ~4.7 kb and 11 kb (Alverson *et al.*, 2011; Dong *et al.*, 2018), respectively.

Acacia has a functional mitochondrial *cox2* gene

Previous studies have shown that the *cox2* gene has been functionally transferred to the nuclear genome in certain legumes (Covello & Gray, 1992) and, in some cases, its mitochondrial counterpart has been retained, as in soybean (Adams *et al.*, 1999). The *Acacia* transcriptome showed no hints of transcripts from nuclear *cox2* copies. So, if *Acacia* has *cox2* nuclear copies, they remain silent or inactive according to our data, and thus the mitochondrial copy is uniquely responsible for the production of *cox2* protein products. This hypothetical scenario indeed has also been reported in pea (Nugent & Palmer, 1991), where the *cox2* gene resides in both the nuclear and mitochondrial compartments but only the mitochondrial one is functional. Regardless of whether *Acacia* has or not nuclear *cox2* copies, our data show that the mitochondrial *cox2* copy is functional, and thus the splicing of its long *cox2i691* is needed for the production of mature mRNAs.

In addition, we found that the *Acacia cox2* transcripts were efficiently C-to-U edited at very well conserved positions that encode for amino acids within protein domains that are relevant for the functioning of COX2. Moreover, such C-to-U RNA editing was found to occur efficiently even before *Acacia cox2* mRNAs were spliced. This efficiency of RNA editing is relevant from a functional perspective. It might be preventing the production of potentially defective proteins translated from mature but partially edited transcripts (Phreaner *et al.*, 1996) since the translation of respiratory mitochondrial genes (e.g., *cox2*) is significantly more efficient in comparison to the other mitochondrial genes (Planchard *et al.*, 2018). Taken together, these lines of evidence support the functionality of the *Acacia cox2* gene. This contrasts with what is observed in bacteria where group II introns can have detrimental effects on gene expression (Qu *et al.*, 2018).

cox2i691 is frequently lost through localized retroprocessing events

Our analyses showed that *cox2i373* and *cox2i691* have been lost multiple times during angiosperm evolution with a total of 26 intron losses inferred based on ancestral reconstruction. This is in line with previous reports where intron loss almost exclusively occurs in genes containing *cis*-spliced introns, with two exceptions (Grewé *et al.*, 2016; Guo *et al.*, 2020). For example, massive loss of mitochondrial introns has been reported in *Welswithschia* (Guo *et al.*, 2016), Cupressophytes (Guo *et al.*, 2020), the angiosperms *Geranium* (Park *et al.*, 2015) and Alismataceae (Cuenca *et al.*, 2016). Our analyses revealed that retroprocessing was unambiguously the mechanism behind 8 out of the 26 observed intron losses. Interestingly, some of the genes affected by retroprocessing showed dramatic changes in which not only all the introns were lost but also virtually all the editable cytosines were replaced by thymidines, as observed previously (Sloan *et al.*, 2010; Edera *et al.*, 2018; Brenner *et al.*, 2019). Most of the retroprocessing events analyzed here were localized, i.e., affecting particular

gene regions (Hao *et al.*, 2010; Cuenca *et al.*, 2016; Edera *et al.*, 2018). In *Ajuga reptans*, for example, retroprocessing leads to the loss of *cox2i691* along with its nearest editing sites (Zhu *et al.*, 2014).

We also found frequent cases where genes impacted by retroprocessing exhibited substantial loss of editing sites but not of introns. Besides the localized effect, we hypothesize that this retroprocessing without intron loss is a very likely scenario in *cox2* because *Acacia* intron-containing *cox2* mRNAs were found to be very well edited before intron release, as similarly reported for other flowering plants (Sutton *et al.*, 1991; Yang *et al.*, 1991; Hepburn *et al.*, 2012). As a result, mitochondrial transcriptomes can contain species of mRNA intermediates that are partially or fully edited and their introns are still retained, which would lead to the elimination mediated by retroprocessing of editing sites but not introns.

Cis to trans-splicing shifts of *cox2* introns and episodic elongations of *cox2i691*

We found contrasting patterns of size variation on the *cox2i373* and *cox2i691* introns, where *cox2i691* exhibited a relatively higher variability given by the presence of episodic elongations that were mostly limited to four plant orders (Malpighiales, Fabales, Malvales and Lamiales). Interestingly, we found that the most extreme elongations of *cox2i691* occur within an intron region that exhibits poor conservation across angiosperms and is flanked by very well conserved regions identified as domain III (upstream) and domain V (downstream). Therefore, intron elongates through sequence expansions taking place within domain IV. This is in line with previous studies indicating that intron extensions are generally confined to domain IV (Malek & Knoop, 1998; Glanz & Kück, 2009), which is thought to have a minor role in the splicing reaction (Michel *et al.*, 1989; Malek *et al.*, 1997). The elongations that we observed in *cox2i691* are not rare events in plant mitochondria since other introns have shown a similar behavior (Guo *et al.*, 2017; Shi *et al.*, 2018). We hypothesize that such elongations might make more probable that *cox2i691* shifts from *cis*- to *trans*-splicing (Malek & Knoop, 1998; Laroche & Bousquet, 1999; Groth-Malonek *et al.*, 2005). This shifting might be mediated by disruption of the elongated domains IV through DNA recombination (Glanz & Kück, 2009). Since diverse studies have linked long introns in plants with the abundance of repeats in the mitogenomes (Alverson *et al.*, 2011; Guo *et al.*, 2017; Dong *et al.*, 2018), we investigated whether repeat content is a factor behind the observed elongations of *cox2i691*. However, we found that such a factor was not responsible for the majority of the *cox2i691* elongations, particularly the most extreme cases. Alternatively, although such intron elongations may be due to the insertion of exogenous genetic material (Choi *et al.*, 2019), our analyses were unable to determine the origin of the novel sequences within the elongated *cox2i691* sequences, probably due to a lack of availability of the donor DNA in the public databases.

Author contribution

AE and MVS-P designed the study. AW, KH and PN identified and collected *Acacia ligulata*. AW and KH extracted and purified RNA. AE collected angiosperm data. AE designed, coded and performed pipelines for data analyses. AE, IS and MVS-P analyzed results. AE, IS and MVS-P wrote the manuscript.

Data availability

The data that support the findings of this study are openly available in SRA at <https://www.ncbi.nlm.nih.gov/sra>, reference number PRJNA869960.

Supplementary information

Data S1. Alignment of *cox2* coding regions.

Data S2. Alignment of *cox2i373* introns.

Data S3. Alignment of *cox2i691* introns.

Figure S1. RNA sequencing data aligned to genes containing *cis*-splicing introns within the Acacia mitogenome. Filled curves show the RNA read depth. Black rectangles indicate coding regions.

Figure S2. Estimated lengths of the *trans*-splicing intron fragments within the Acacia mitogenome. The blue curve depicts the read depth for Acacia *cox2i691*, included for reference. The orange curves show read pairs aligned to upstream and downstream regions of *trans*-splicing introns. The darker orange curves indicate intron regions conserved across angiosperms.

Figure S3. RNA read depth of the intron-containing genes within the Acacia mitogenome. The filled curves depict RNA reads aligned to the spliced version of *cox2*. The blue-filled and red-filled curves show aligned RNA reads containing the exon-exon and the exon-intron/intron-exon junctions, respectively, where the intronic regions are clipped.

Figure S4 Role of RNA editing on the functionality of protein products of Acacia *cox2*. **A)** Predicted protein domains are indicated as horizontal filled rectangles. Filled areas plot the probability of finding certain topological elements along the protein sequence. Vertical black lines indicate residues affected by non-synonymous C-to-U editing. Dots indicate sites whose lack of editing leads to protein disruption. **B)** Amino acid sequence alignment. Yellow boxes indicate amino acids yielded by C-to-U edited cytidines.

Figure S5. Sequence alignment between Acacia and *Leucaena cox2* genes. Colored boxes indicate Acacia annotations for exon 1, *cox2i691* and exon 2 (blue, orange and red, respectively). The black box indicates the erroneously stop codon as annotated in the *Leucaena* mitogenome. Gaps are represented as dashes.

Figure S6. Phylogenetic trees estimated from *cox2* gene sequences of angiosperms. Trees were estimated from the whole gene sequences excluding editing sites, each of the three possible exonic regions delimited by both introns and *cox2i373* and *cox2i691* sequences. To improve readability, trees are visualized without estimated branch lengths.

Table S1. Average read depth of coding regions from the Acacia mitogenome.

Table S2. C-to-U editing sites experimentally identified in the *Acacia ligulata* mitogenome using sequenced RNA.

Table S3. Timing of RNA editing of non-synonymous sites and splicing in the Acacia intron-containing mitochondrial genes.

Table S4. Curated genomic coordinates of the *cox2* genes from full-sequenced mitogenomes of diverse angiosperms.

Table S5. Intron content and statistics of the *cox2* gene in angiosperms.

References

Adams, K.L., Song, K., Roessler, P.G., Nugent, J.M., Doyle, J.L., Doyle, J.J. and Palmer, J.D., 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proceedings of the National Academy of Sciences*, **96**: 13863-13868.

Adams, K.L., Ong, H.C. and Palmer, J.D., 2001. Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Molecular Biology and Evolution*, **18**: 2289-2297.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, **25**: 3389-3402.

Alverson, A.J., Rice, D.W., Dickinson, S., Barry, K. and Palmer, J.D., 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *The Plant Cell*, **23**: 2499-2513.

Aubert, D., Bisanz-Seyer, C. and Herzog, M., 1992. Mitochondrial *rps14* is a transcribed and edited pseudogene in *Arabidopsis thaliana*. *Plant molecular biology*, **20**: 1169-1174.

Azani, N., Babineau, M., Bailey, C.D., Banks, H., Barbosa, A.R., Pinto, R.B., Boatwright, J.S., Borges, L.M., Brown, G.K., Bruneau, A. and Candido, E., 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *taxon*, **66**: 44-77.

Bell, M.V., Cowper, A.E., Lefranc, M.P., Bell, J.I. and Sreaton, G.R., 1998. Influence of intron length on alternative splicing of CD44. *Molecular and Cellular Biology*, **18**: 5930-5941.

- 558 **Bonen, L., 2008.** Cis-and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*, **8**: 26-34.
- 559 **Bonen, L., 2012.** Evolution of mitochondrial introns in plants and photosynthetic microbes. In *Advances in*
- 560 *Botanical Research*, **63**: 155-186.
- 561 **Brenner, W.G., Mader, M., Müller, N.A., Hoenicka, H., Schroeder, H., Zorn, I., Fladung, M. and Kersten,**
- 562 **B., 2019.** High level of conservation of mitochondrial RNA editing sites among four populus species. *G3:*
- 563 *Genes, Genomes, Genetics*, **9**: 709-717.
- 564 **Brown, G.G., Colas des Francs-Small, C. and Ostersetzer-Biran, O., 2014.** Group II intron splicing factors
- 565 in plant mitochondria. *Frontiers in plant science*, **5**: 35.
- 566 **Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T., 2009.** trimAl: a tool for automated alignment
- 567 trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**: 1972-1973.
- 568 **Chang, N., Sun, Q., Hu, J., An, C. and Gao, H., 2017.** Large introns of 5 to 10 kilo base pairs can be spliced
- 569 out in *Arabidopsis*. *Genes*, **8**: 200.
- 570 **Chen, S., Zhou, Y., Chen, Y. and Gu, J., 2018.** fastp: an ultra-fast all-in-one FASTQ preprocessor.
- 571 *Bioinformatics*, **34**: i884-i890.
- 572 **Cho, Y., Qiu, Y.L., Kuhlman, P. and Palmer, J.D., 1998.** Explosive invasion of plant mitochondria by a
- 573 group I intron. *Proceedings of the National Academy of Sciences*, **95**: 14244-14249.
- 574 **Choi, I.S., Schwarz, E.N., Ruhlman, T.A., Khiyami, M.A., Sabir, J.S., Hajarrah, N.H., Sabir, M.J., Rabah,**
- 575 **S.O. and Jansen, R.K., 2019.** Fluctuations in Fabaceae mitochondrial genome size and content are both ancient
- 576 and recent. *BMC plant biology*, **19**: 1-15.
- 577 **Covello, P.S. and Gray, M.W., 1992.** Silent mitochondrial and active nuclear genes for subunit 2 of
- 578 cytochrome c oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. *The EMBO journal*, **11**:
- 579 3815-3820.
- 580 **Cuenca, A., Petersen, G., Seberg, O. and Jähren, A.H., 2012.** Genes and processed paralogs co-exist in plant
- 581 mitochondria. *Journal of molecular evolution*, **74**: 158-169.
- 582 **Cuenca, A., Ross, T.G., Graham, S.W., Barrett, C.F., Davis, J.I., Seberg, O. and Petersen, G., 2016.**
- 583 Localized retroprocessing as a model of intron loss in the plant mitochondrial genome. *Genome Biology and*
- 584 *Evolution*, **8**: 2176-2189.
- 585 **Dahan, J., Tcherkez, G., Macherel, D., Benamar, A., Belcram, K., Quadrado, M., Arnal, N. and Mireau,**
- 586 **H., 2014.** DEFICIENT1 Gene Leads to Cytochrome c Oxidase Depletion and Reorchestrated Respiratory
- 587 Metabolism in *Arabidopsis*. *Plant Physiology*, **166**: 1788-1802.
- 588 **Dalby, S.J. and Bonen, L., 2013.** Impact of low temperature on splicing of atypical group II introns in wheat
- 589 mitochondria. *Mitochondrion*, **13**: 647-655.
- 590 **Dong, S., Zhao, C., Chen, F., Liu, Y., Zhang, S., Wu, H., Zhang, L. and Liu, Y., 2018.** The complete
- 591 mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low
- 592 recombination. *BMC genomics*, **19**: 1-12.
- 593 **Edera, A.A., Gandini, C.L. and Sanchez-Puerta, M., 2018.** Towards a comprehensive picture of C-to-U RNA
- 594 editing sites in angiosperm mitochondria. *Plant Molecular Biology*, **97**: 215-231.
- 595 **Finn, R.D., Clements, J. and Eddy, S.R., 2011.** HMMER web server: interactive sequence similarity
- 596 searching. *Nucleic acids research*, **39**: W29-W37.
- 597 **Garcia, L.E. and Sanchez-Puerta, M.V., 2021.** Transcriptional Landscape and Splicing Efficiency in
- 598 *Arabidopsis* Mitochondria. *Cells*, **10**: 2054.
- 599 **Geiss, K.T., Abbas, G.M. and Makaroff, C.A., 1994.** Intron loss from the NADH dehydrogenase subunit 4
- 600 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate.
- 601 *Molecular and General Genetics MGG*, **243**: 97-105.
- 602 **Giegé, P. and Brennicke, A., 1999.** RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in
- 603 ORFs. *Proceedings of the National Academy of Sciences*, **96**: 15324-15329.
- 604 **Glanz, S. and Kück, U., 2009.** Trans-splicing of organelle introns—a detour to continuous RNAs. *Bioessays*, **31**:
- 605 921-934.
- 606 **Grewe, F., Zhu, A. and Mower, J.P., 2016.** Loss of a trans-splicing *nad1* intron from Geraniaceae and transfer
- 607 of the maturase gene *matR* to the nucleus in *Pelargonium*. *Genome Biology and Evolution*, **8**: 3193-3201.

- 608 **Grimes, B.T., Sisay, A.K., Carroll, H.D. and Cahoon, A.B., 2014.** Deep sequencing of the tobacco
609 mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC*
610 *genomics*, **15**: 1-10.
- 611 **Groth-Malonek, M., Pruchner, D., Grewe, F. and Knoop, V., 2005.** Ancestors of *trans*-splicing
612 mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants.
613 *Molecular Biology and Evolution*, **22**: 117-125.
- 614 **Guo, W., Grewe, F., Fan, W., Young, G.J., Knoop, V., Palmer, J.D. and Mower, J.P., 2016.** *Ginkgo* and
615 *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Molecular Biology*
616 *and Evolution*, **33**: 1448-1460.
- 617 **Guo, H., Liu, J., Luo, L., Wei, X., Zhang, J., Qi, Y., Zhang, B., Liu, H. and Xiao, P., 2017.** Complete
618 chloroplast genome sequences of *Schisandra chinensis*: Genome structure, comparative analysis, and
619 phylogenetic relationship of basal angiosperms. *Sci. China Life Sci*, **60**: 1286-1290.
- 620 **Guo, W., Zhu, A., Fan, W., Adams, R.P. and Mower, J.P., 2020.** Extensive shifts from *cis*-to *trans*-splicing
621 of gymnosperm mitochondrial introns. *Molecular Biology and Evolution*, **37**: 1615-1620.
- 622 **Handa, H., 2003.** The complete nucleotide sequence and RNA editing content of the mitochondrial genome of
623 rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis*
624 *thaliana*. *Nucleic acids research*, **31**: 5907-5916.
- 625 **Hao, W., Richardson, A.O., Zheng, Y. and Palmer, J.D., 2010.** Gorgeous mosaic of mitochondrial genes
626 created by horizontal transfer and gene conversion. *Proceedings of the National Academy of Sciences*, **107**:
627 21576-21581.
- 628 **Hecht, J., Grewe, F. and Knoop, V., 2011.** Extreme RNA editing in coding islands and abundant
629 microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant
630 mtDNA recombination in early tracheophytes. *Genome Biology and Evolution*, **3**: 344-358.
- 631 **Hepburn, N.J., Schmidt, D.W. and Mower, J.P., 2012.** Loss of two introns from the *Magnolia tripetala*
632 mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of
633 intron loss. *Molecular Biology and Evolution*, **29**: 3111-3120.
- 634 **Hiesel, R. and Brennicke, A., 1983.** Cytochrome oxidase subunit II gene in mitochondria of *Oenothera* has no
635 intron. *The EMBO Journal*, **2**: 2173-2178.
- 636 **Itchoda, N., Nishizawa, S., Nagano, H., Kubo, T. and Mikami, T., 2002.** The sugar beet mitochondrial *nad4*
637 gene: an intron loss and its phylogenetic implication in the Caryophyllales. *Theoretical and Applied Genetics*,
638 **104**: 209-213.
- 639 **Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L. and Perlman, P.S., 1988.** Group II intron self-
640 splicing. Alternative reaction conditions yield novel products. *Journal of Biological Chemistry*, **263**: 3432-3439.
- 641 **Jin, Y. and Qian, H., 2019.** V. PhyloMaker: an R package that can generate very large phylogenies for vascular
642 plants. *Ecography*, **42**: 1353-1359.
- 643 **Jobson, R.W. and Qiu, Y.L., 2008.** Did RNA editing in plant organellar genomes originate under natural
644 selection or through genetic drift?. *Biology Direct*, **3**: 1-22.
- 645 **Joly, S., Brouillet, L. and Bruneau, A., 2001.** Phylogenetic implications of the multiple losses of the
646 mitochondrial *coxII.i3* intron in the angiosperms. *International Journal of Plant Sciences*, **162**: 359-373.
- 647 **Katoh, K., Rozewicki, J. and Yamada, K.D., 2019.** MAFFT online service: multiple sequence alignment,
648 interactive sequence choice and visualization. *Briefings in bioinformatics*, **20**: 1160-1166.
- 649 **Kim, S.R., Yang, J.I., Moon, S., Ryu, C.H., An, K., Kim, K.M., Yim, J. and An, G., 2009.** Rice OGR1
650 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *The Plant*
651 *Journal*, **59**: 738-749.
- 652 **Kim, S. and Yoon, M.K., 2010.** Comparison of mitochondrial and chloroplast genome segments from three
653 onion (*Allium cepa* L.) cytoplasm types and identification of a *trans*-splicing intron of *cox2*. *Current genetics*,
654 **56**: 177-188.
- 655 **Kovar, L., Nageswara-Rao, M., Ortega-Rodriguez, S., Dugas, D.V., Straub, S., Cronn, R., Strickler, S.R.,
656 Hughes, C.E., Hanley, K.A., Rodriguez, D.N. and Langhorst, B.W., 2018.** PacBio-based mitochondrial
657 genome assembly of *Leucaena trichandra* (Leguminosae) and an intragenomic assessment of mitochondrial RNA
658 editing. *Genome biology and evolution*, **10**: 2501-2517.

- 659 **Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L., 2001.** Predicting transmembrane protein
660 topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, **305**:
661 567-580.
- 662 **Kudla, J., Albertazzi, F., Blazević, D., Hermann, M. and Bock, R., 2002.** Loss of the mitochondrial *cox2*
663 intron 1 in a family of monocotyledonous plants and utilization of mitochondrial intron sequences for the
664 construction of a nuclear intron. *Molecular Genetics and Genomics*, **267**: 223-230.
- 665 **Kurihara-Yonemoto, S. and Handa, H., 2001.** Low temperature affects the processing pattern and RNA
666 editing status of the mitochondrial *cox2* transcripts in wheat. *Current Genetics*, **40**: 203-208.
- 667 **Kurihara-Yonemoto, S. and Kubo, T., 2010.** Increased accumulation of intron-containing transcripts in rice
668 mitochondria caused by low temperature: is cold-sensitive RNA editing implicated?. *Current Genetics*, **56**: 529-
669 541.
- 670 **Laroche, J. and Bousquet, J., 1999.** Evolution of the mitochondrial *rps3* intron in perennial and annual
671 angiosperms and homology to *nad5* intron 1. *Molecular biology and evolution*, **16**: 441-452.
- 672 **Laroche, J., F., Monat, C., Verreault, E. and Cousineau, B., 2021.** Molecular characterization of both
673 transesterification reactions of the group II intron circularization pathway. *Nucleic acids research*, **49**: 6996-
674 7010.
- 675 **Li-Pook-Than, J. and Bonen, L., 2006.** Multiple physical forms of excised group II intron RNAs in wheat
676 mitochondria. *Nucleic acids research*, **34**: 2782-2790.
- 677 **Li-Pook-Than, J., Carrillo, C., Niknejad, N., Calixte, S., Crosthwait, J. and Bonen, L., 2007.** Relationship
678 between RNA splicing and exon editing near intron junctions in wheat mitochondria. *Physiologia Plantarum*,
679 **129**: 23-33.
- 680 **Lin, Y., Li, P., Zhang, Y., Akhter, D., Pan, R., Fu, Z., Huang, M., Li, X. and Feng, Y., 2022.** Unprecedented
681 organelle genomic variations in morning glories reveal independent evolutionary scenarios of parasitic plants
682 and the diversification of plant mitochondrial complexes. *BMC biology*, **20**: 1-16.
- 683 **Lippok, B., Brennicke, A. and Wissinger, B., 1992.** The *coxII* gene in carrot mitochondria contains two
684 introns. *Molecular and General Genetics MGG*, **232**: 322-327.
- 685 **Malek, O., Brennicke, A. and Knoop, V., 1997.** Evolution of trans-splicing plant mitochondrial introns in pre-
686 Permian times. *Proceedings of the National Academy of Sciences*, **94**: 553-558.
- 687 **Malek, O. and Knoop, V., 1998.** Trans-splicing group II introns in plant mitochondria: the complete set of cis-
688 arranged homologs in ferns, fern allies, and a hornwort. *RNA*, **4**: 1599-1609.
- 689 **Michel, F., Kazuhiko, U. and Haruo, O., 1989.** Comparative and functional anatomy of group II catalytic
690 introns—a review. *Gene*, **82**: 5-30.
- 691 **Møller, I.M., Rasmusson, A.G. and Van Aken, O., 2021.** Plant mitochondria—past, present and future. *The*
692 *Plant Journal*, **108**: 912-959.
- 693 **Mower, J.P. and Palmer, J.D., 2006.** Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*.
694 *Molecular Genetics and Genomics*, **276**: 285-293.
- 695 **Mower, J.P., 2009.** The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes
696 and user-defined alignments. *Nucleic acids research*, **37**: W253-W259.
- 697 **Mower, J.P., 2020.** Variation in protein gene and intron content among land plant mitogenomes.
698 *Mitochondrion*, **53**: 203-213.
- 699 **Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. and Jarrell,**
700 **K.A., 2001.** Excision of group II introns as circles. *Molecular cell*, **8**: 201-211.
- 701 **Nugent, J.M. and Palmer, J.D., 1991.** RNA-mediated transfer of the gene *coxII* from the mitochondrion to the
702 nucleus during flowering plant evolution. *Cell*, **66**: 473-481.
- 703 **Pagel, M., 1994.** Detecting correlated evolution on phylogenies: a general method for the comparative analysis
704 of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **255**: 37-45.
- 705 **Paradis, E. and Schliep, K., 2019.** ape 5.0: an environment for modern phylogenetics and evolutionary
706 analyses in R. *Bioinformatics*, **35**: 526-528.
- 707 **Park, S., Grewe, F., Zhu, A., Ruhlman, T.A., Sabir, J., Mower, J.P. and Jansen, R.K., 2015.** Dynamic
708 evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New*
709 *Phytologist*, **208**: 570-583.

- 710 **Parkinson, C.L., Mower, J.P., Qiu, Y.L., Shirk, A.J., Song, K., Young, N.D., DePamphilis, C.W. and**
 711 **Palmer, J.D., 2005.** Multiple major increases and decreases in mitochondrial substitution rates in the plant
 712 family Geraniaceae. *BMC evolutionary biology*, **5**: 1-12.
- 713 **Petersen, G., Cuenca, A., Møller, I.M. and Seberg, O., 2015.** Massive gene loss in mistletoe (*Viscum*,
 714 Viscaceae) mitochondria. *Scientific reports*, **5**: 1-7.
- 715 **Phreaner, C.G., Williams, M.A. and Mulligan, R.M., 1996.** Incomplete editing of *rps12* transcripts results in
 716 the synthesis of polymorphic polypeptides in plant mitochondria. *The Plant Cell*, **8**: 107-117.
- 717 **Planchard, N., Bertin, P., Quadrado, M., Dargel-Graffin, C., Hatin, I., Namy, O. and Mireau, H., 2018.**
 718 The translational landscape of *Arabidopsis* mitochondria. *Nucleic Acids Research*, **46**: 6218-6228.
- 719 **Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. and Korobeynikov, A., 2020.** Using SPAdes de novo
 720 assembler. *Current protocols in bioinformatics*, **70**: 102.
- 721 **Pruitt, K.D. and Hanson, M.R., 1989.** Cytochrome oxidase subunit II sequences in *Petunia* mitochondria: two
 722 intron-containing genes and an intron-less pseudogene associated with cytoplasmic male sterility. *Current*
 723 *genetics*, **16**: 281-291.
- 724 **Qi, W., Tian, Z., Lu, L., Chen, X., Chen, X., Zhang, W. and Song, R., 2017.** Editing of mitochondrial
 725 transcripts *nad3* and *cox2* by Dek10 is essential for mitochondrial function and maize plant development.
 726 *Genetics*, **205**: 1489-1501.
- 727 **Qiu, Y.L., Cho, Y., Cox, J.C. and Palmer, J.D., 1998.** The gain of three mitochondrial introns identifies
 728 liverworts as the earliest land plants. *Nature*, **394**: 671-674.
- 729 **Qiu, Y.L. and Palmer, J.D., 2004.** Many independent origins of trans splicing of a plant mitochondrial group II
 730 intron. *Journal of Molecular Evolution*, **59**: 80-89.
- 731 **Qu, G., Piazza, C.L., Smith, D. and Belfort, M., 2018.** Group II intron inhibits conjugative relaxase
 732 expression in bacteria by mRNA targeting. *Elife*, **7**: e34268.
- 733 **Rabbi, M.F. and Wilson, K.G., 1993.** The mitochondrial *coxII* intron has been lost in two different lineages of
 734 dicots and altered in others. *American journal of botany*, **80**: 1216-1223.
- 735 **Sanchez-Puerta, M.V., Cho, Y., Mower, J.P., Alverson, A.J. and Palmer, J.D., 2008.** Frequent,
 736 phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Molecular*
 737 *biology and evolution*, **25**: 1762-1777.
- 738 **Sanchez-Puerta, M.V., Edera, A., Gandini, C.L., Williams, A.V., Howell, K.A., Nevill, P.G. and Small, I.,**
 739 **2019.** Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum*
 740 *mirabile* (Balanophoraceae). *Molecular phylogenetics and evolution*, **132**: 243-250.
- 741 **Shi, Y., Liu, Y., Zhang, S., Zou, R., Tang, J., Mu, W., Peng, Y. and Dong, S., 2018.** Assembly and
 742 comparative analysis of the complete mitochondrial genome sequence of *Sophora japonica* 'JinhuaiJ2'. *PloS*
 743 *one*, **13**: e0202485.
- 744 **Skippington, E., Barkman, T.J., Rice, D.W. and Palmer, J.D., 2015.** Miniaturized mitogenome of the
 745 parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes.
 746 *Proceedings of the National Academy of Sciences*, **112**: E3515-E3524.
- 747 **Sloan, D.B., MacQueen, A.H., Alverson, A.J., Palmer, J.D. and Taylor, D.R., 2010.** Extensive loss of RNA
 748 editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving
 749 force. *Genetics*, **185**: 1369-1380.
- 750 **Small, I.D., Schallenberg-Rüdinger, M., Takenaka, M., Mireau, H. and Ostersetter-Biran, O., 2020.** Plant
 751 organellar RNA editing: what 30 years of research has revealed. *The Plant Journal*, **101**: 1040-1056.
- 752 **Soto, I.C., Fontanesi, F., Liu, J. and Barrientos, A., 2012.** Biogenesis and assembly of eukaryotic cytochrome
 753 c oxidase catalytic core. *Biochimica Et Biophysica Acta (BBA)-Bioenergetics*, **1817**: 883-897.
- 754 **Stamatakis, A., 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
 755 phylogenies. *Bioinformatics*, **30**: 1312-1313.
- 756 **Sutton, C.A., Conklin, P.L., Pruitt, K.D. and Hanson, M.R., 1991.** Editing of pre-mRNAs can occur before
 757 *cis*- and *trans*-splicing in *Petunia* mitochondria. *Molecular and cellular biology*, **11**: 4274-4277.
- 758 **Timón-Gómez, A., Nývltová, E., Abriata, L.A., Vila, A.J., Hosler, J. and Barrientos, A., 2018.**
 759 Mitochondrial cytochrome c oxidase biogenesis: Recent developments. In *Seminars in cell & developmental*
 760 *biology*, **76**: 163-178.

761 **Williams, A.V., Boykin, L.M., Howell, K.A., Nevill, P.G. and Small, I., 2015.** The complete sequence of the
762 *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PloS one*, **10**: e0125768.
763 **Wu, Z., Stone, J.D., Štorchová, H. and Sloan, D.B., 2015.** High transcript abundance, RNA editing, and small
764 RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC*
765 *genomics*, **16**: 1-13.
766 **Yang, A.J. and Mulligan, R.M., 1991.** RNA editing intermediates of *cox2* transcripts in maize mitochondria.
767 *Molecular and cellular biology*, **11**: 4278-4281.
768 **Yu, R., Sun, C., Liu, Y. and Zhou, R., 2021.** Shifts from *cis*-to *trans*-splicing of five mitochondrial introns in
769 *Tolypanthus maclurei*. *PeerJ*, **9**: e12260.
770 **Zhu, A., Guo, W., Jain, K. and Mower, J.P., 2014.** Unprecedented heterogeneity in the synonymous
771 substitution rate within a plant genome. *Molecular biology and evolution*, **31**: 1228-1236.
772
773
774
775
776

Legends

Fig. 1. RNA read depth of the *cox2* gene from *Acacia*. **A)** The filled curve depicts aligned reads. **B)** The filled curves depict reads aligned to the spliced version of *cox2*. The blue-filled and red-filled curves show aligned RNA reads containing the exon-exon and the exon-intron/intron-exon junctions, respectively, where the intronic regions are clipped.

Fig. 2. Sequence conservation of the *cox2* gene in angiosperms. The heatmap plots the bases of three fragments extracted from the collected *cox2* gene sequences. The left, middle and right fragments correspond to the 5' and 3' ends of *cox2i691* and from the 3' exon, respectively. Filled squares on the leftmost indicate misannotated *cox2* gene sequences found in this study..

Fig. 3. Distribution of *cox2i373* and *cox2i691* introns across angiosperms. Blue circles indicate introns identified in this study but were misannotated as intronless in Genbank. Ancestral state reconstructions of introns are shown in internal nodes.

Fig. 4. RNA editing site and intron losses in *cox2* through retroprocessing. The heatmap columns depict the nucleotides in 38 homologous positions of *cox2* which are commonly edited. Horizontal bars plot the relative number of C-to-T substitutions. The vertical line shows the threshold (85th percentile) used to determine a significant number of C-to-T substitutions. Arrowheads indicate a significant number of C-to-T substitutions accompanied by intron losses.

Fig. 5. Intron lengths in angiosperm *cox2* genes. Intron absence is represented as zero-length horizontal bars. The 11.4 kb *cox2i373* intron of *Nymphaea colorata* is cropped to improve visualization. Species names are shown for the five longest *cox2i691* introns. Hatched bars indicate intron misannotations found in this study.

Fig. 6. Consensus secondary structure of *cox2691*. The six predicted helical domains are labeled as DI-VI. Sequence conservation is color-encoded. The localization of the longest intron elongations is depicted as a dashed line. Small sequence portions lacking structure, indicated by ellipses, were clipped to improve structure visualization.

Figures

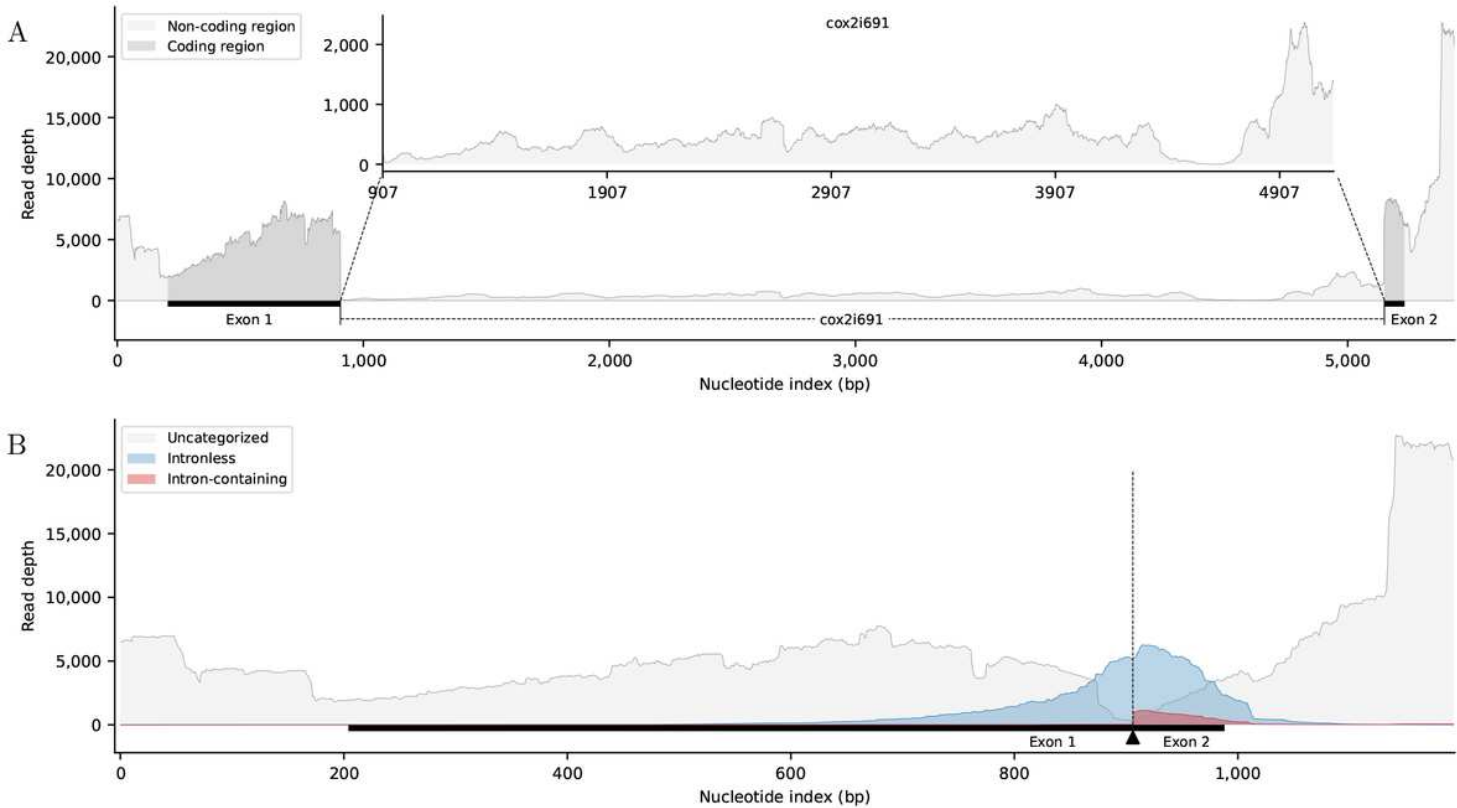


Figure 1

RNA read depth of the *cox2* gene from *Acacia*. A) The filled curve depicts aligned reads. B) The filled curves depict reads aligned to the spliced version of *cox2*. The blue-filled and red-filled curves show aligned RNA reads containing the exon-exon and the exon-intron/intron-exon junctions, respectively, where the intronic regions are clipped.

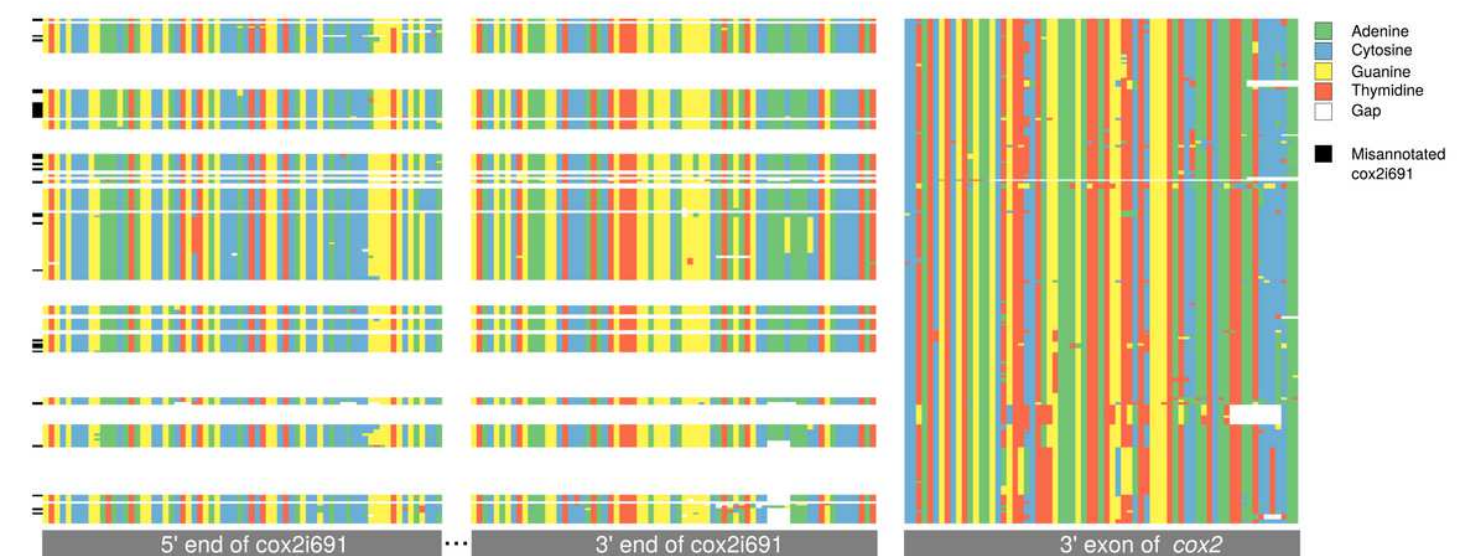


Figure 2

Sequence conservation of the *cox2* gene in angiosperms. The heatmap plots the bases of three fragments extracted from the collected *cox2* gene sequences. The left, middle and right fragments correspond to the 5' and 3' ends of *cox2i691* and from the 3' exon, respectively. Filled squares on the leftmost indicate misannotated *cox2* gene sequences found in this study.

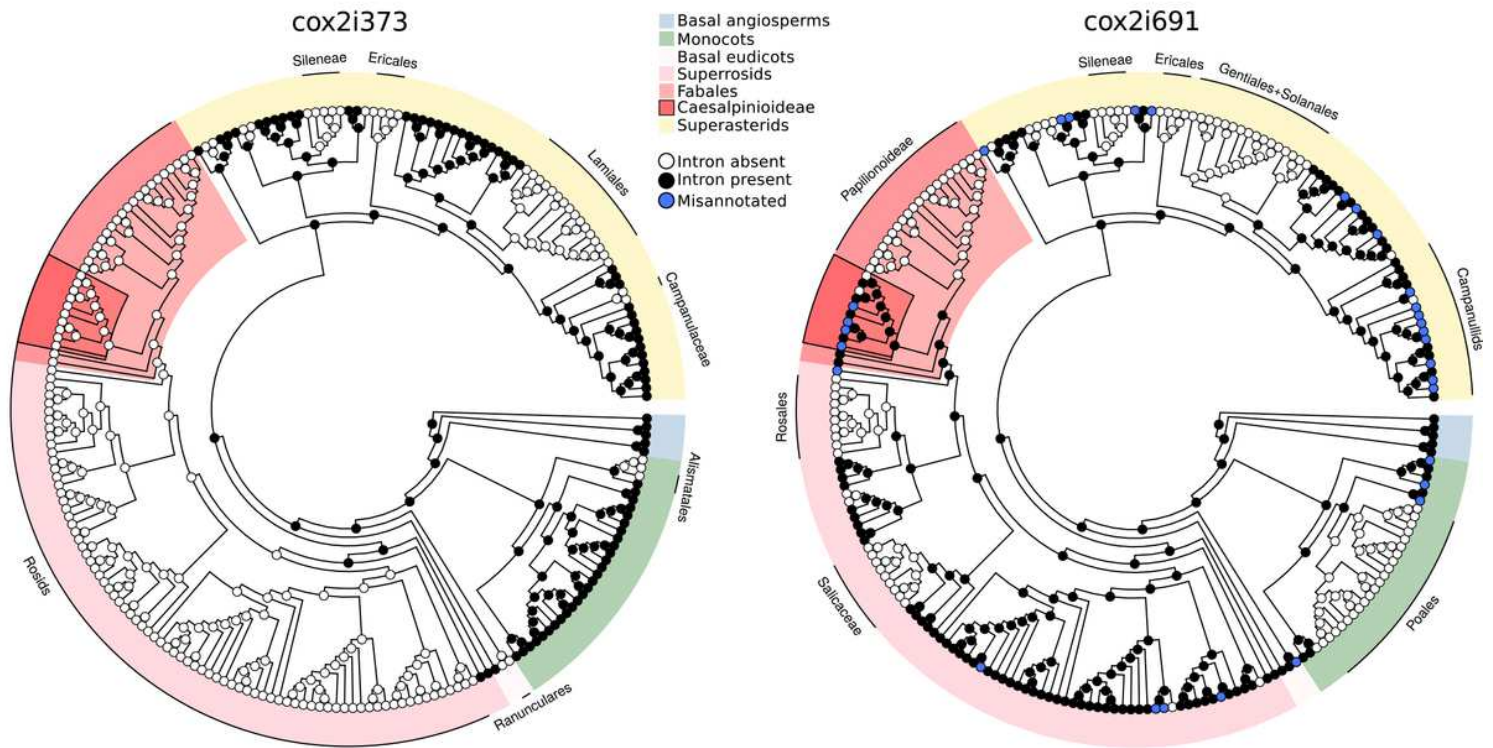


Figure 3

Distribution of *cox2i373* and *cox2i691* introns across angiosperms. Blue circles indicate introns identified in this study but were misannotated as intronless in Genbank. Ancestral state reconstructions of introns are shown in internal nodes.

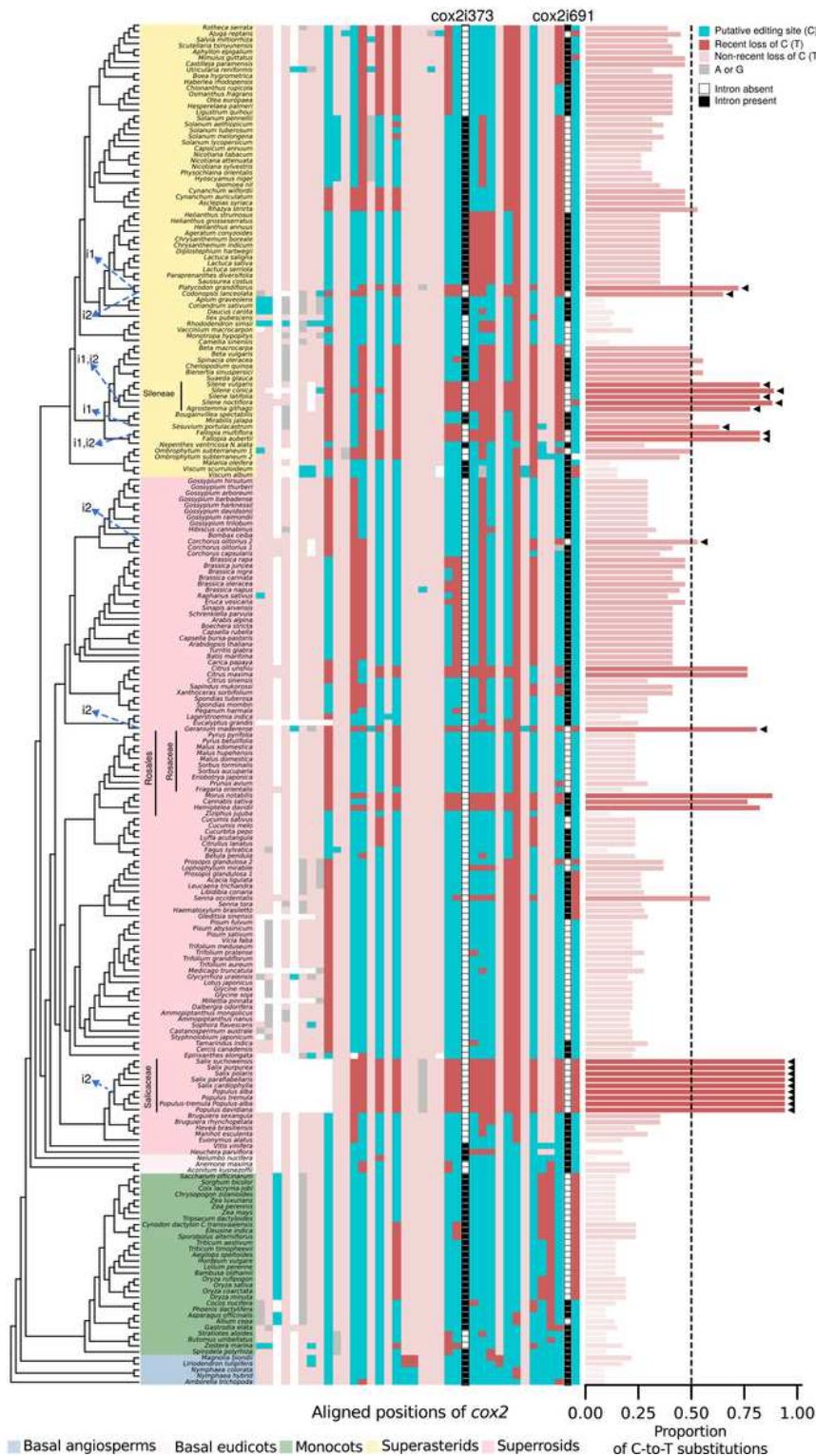


Figure 4

RNA editing site and intron losses in *cox2* through retroprocessing. The heatmap columns depict the nucleotides in 38 homologous positions of *cox2* which are commonly edited. Horizontal bars plot the relative number of C-to-T substitutions. The vertical line shows the threshold (85th percentile) used to determine a significant number of C-to-T substitutions. Arrowheads indicate a significant number of C-to-T substitutions accompanied by intron losses.

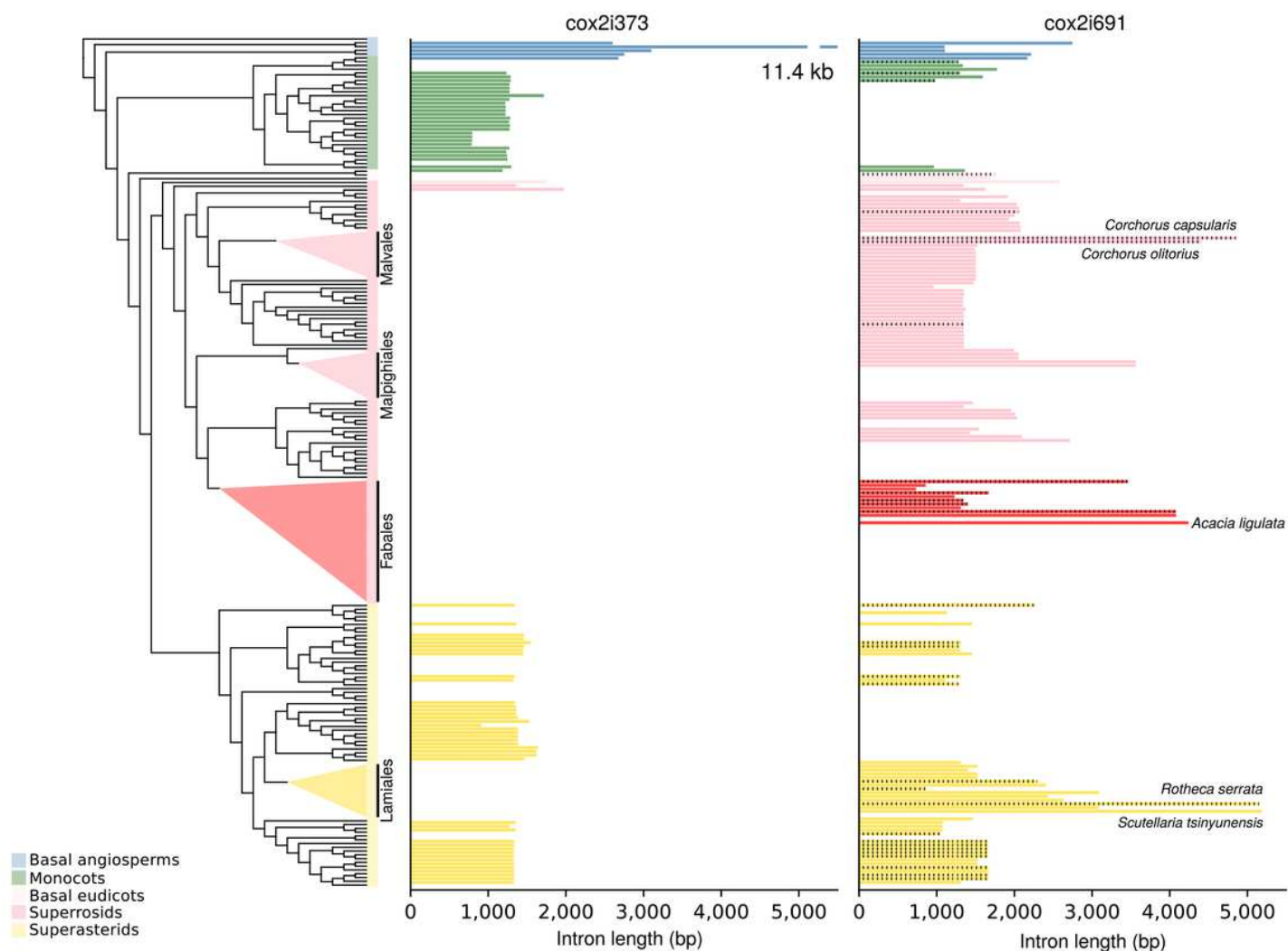


Figure 5

Intron lengths in angiosperm cox2 genes. Intron absence is represented as zero-length horizontal bars. The 11.4 kb cox2i373 intron of *Nymphaea colorata* is cropped to improve visualization. Species names are shown for the five longest cox2i691 introns. Hatched bars indicate intron misannotations found in this study.

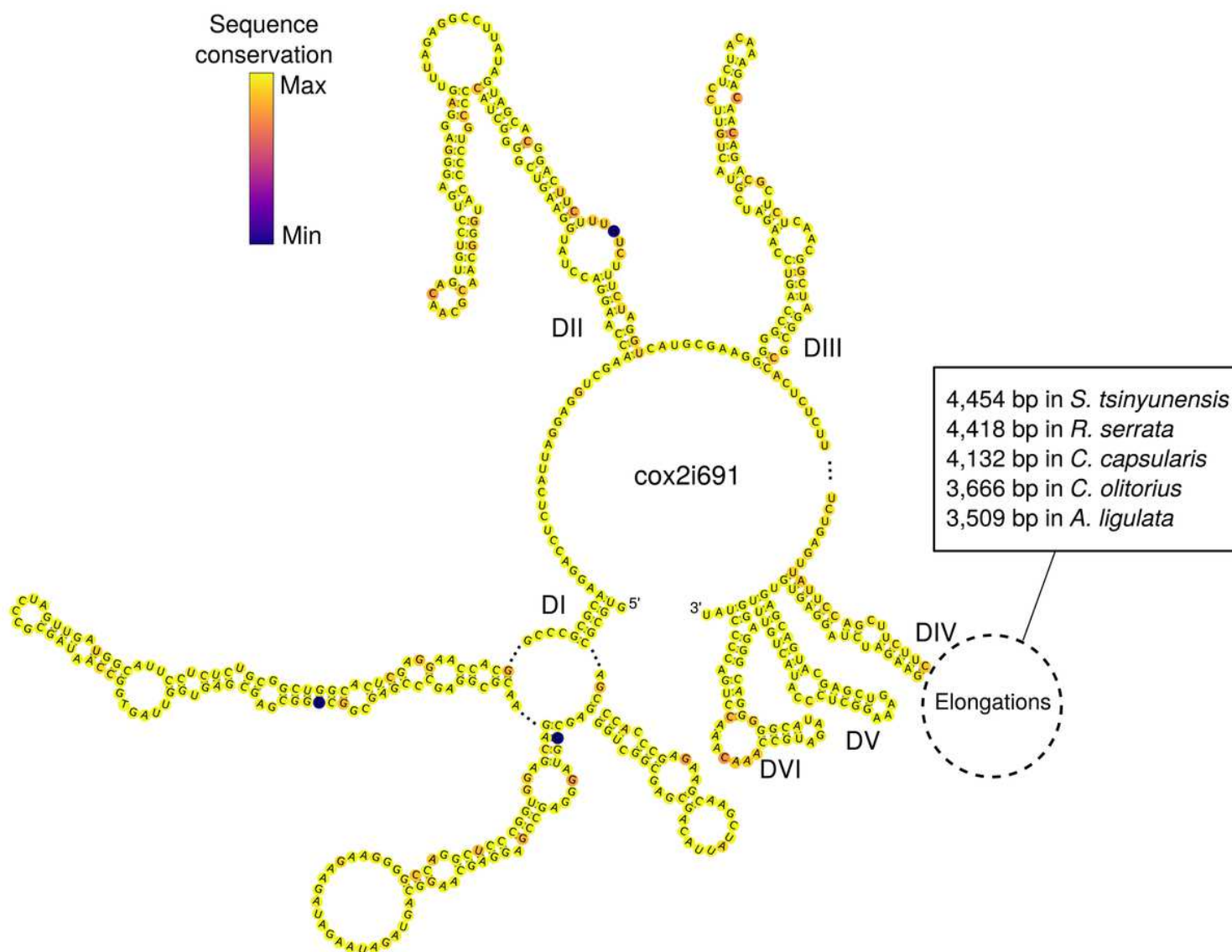


Figure 6

Consensus secondary structure of cox2691. The six predicted helical domains are labeled as DI-VI. Sequence conservation is color-encoded. The localization of the longest intron elongations is depicted as a dashed line. Small sequence portions lacking structure, indicated by ellipses, were clipped to improve structure visualization.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1splicingefficiency.xlsx](#)
- [DataS1cox2.fas](#)
- [DataS2cox2i373.fas](#)

- [DataS3cox2i691.fas](#)
- [FigS1genereaddepths.pdf](#)
- [FigS2lengthofintronfragments.pdf](#)
- [FigS3maturereaddepths.pdf](#)
- [FigS4cox2functionality.pdf](#)
- [FigS5leucaenamisannotation.pdf](#)
- [FigS6phylotrees.pdf](#)
- [TableS1averagereaddepths.xlsx](#)
- [TableS2identifiedsites.xlsx](#)
- [TableS3upstreamdownstreamesitesfromsplicingpoints.xlsx](#)
- [TableS4cox2coordinates.xlsx](#)
- [TableS5intronpresenceabsence.xlsx](#)