Published online 9 January 2020

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

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

Simon Zumkeller, Philipp Gerke and Volker Knoop*

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

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

ABSTRACT

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

INTRODUCTION

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

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

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

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

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

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

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

MATERIALS AND METHODS

Plant material and sequence data

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

Molecular cloning

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

Further bioinformatic analyses

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

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

RESULTS

Nomenclature issues

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

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

Group II intron atp9i87g2 is a twintron conserved among Lycopodiales

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

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

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

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

Jumping into itself: group II intron twintron cox1i1149g2 in Lycopodiales

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

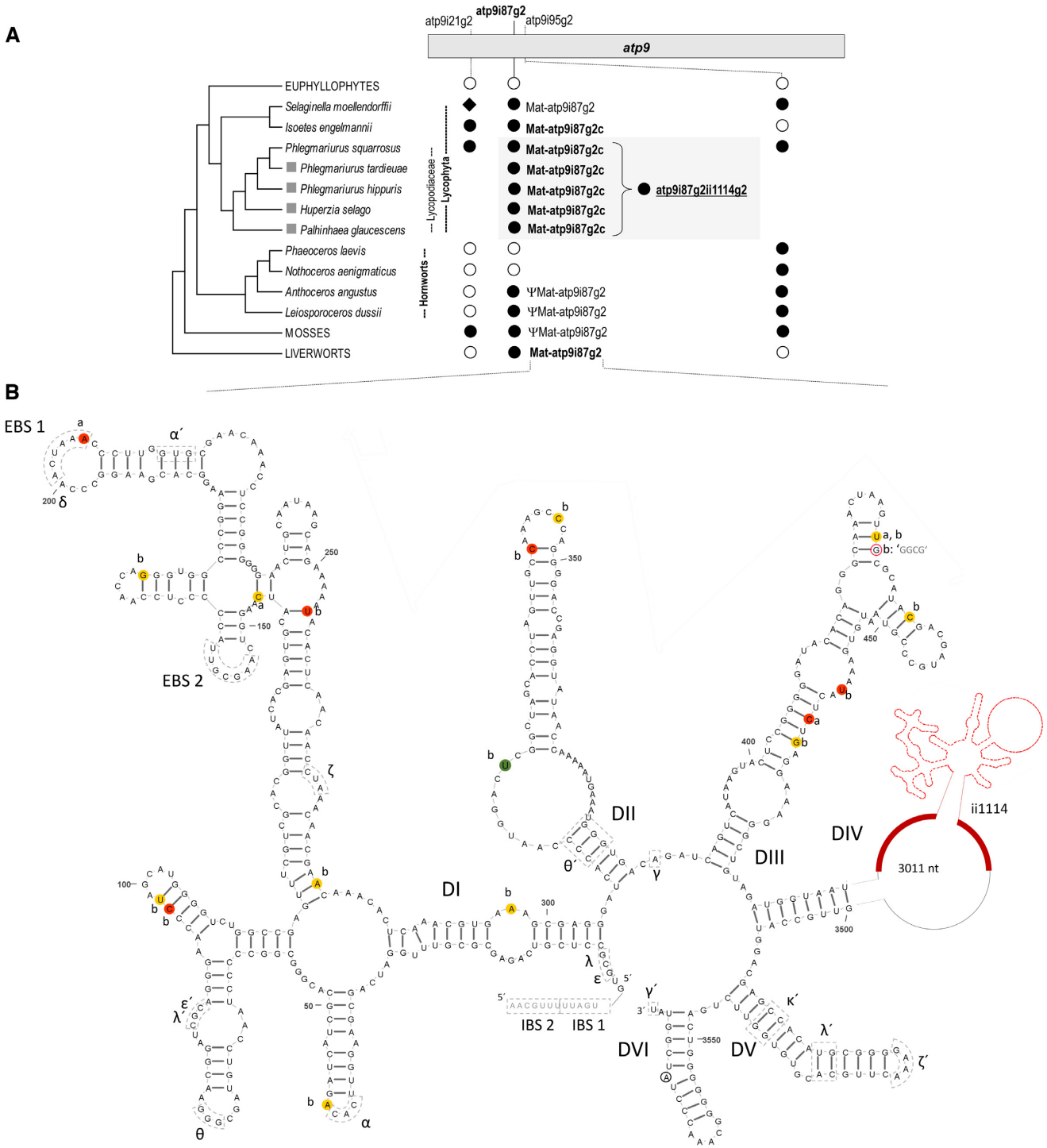

Figure 1. Group II intron atp9i87g2 is a twintron in Lycopodiaceae. (**A**) Presence or absence of group II introns atp9i21g2, atp9i87g2 and atp9i95g2 in the mtDNAs of different plant taxa is indicated by filled or open circles, respectively. Note that *Phlegmariurus squarrosus* had previously been named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

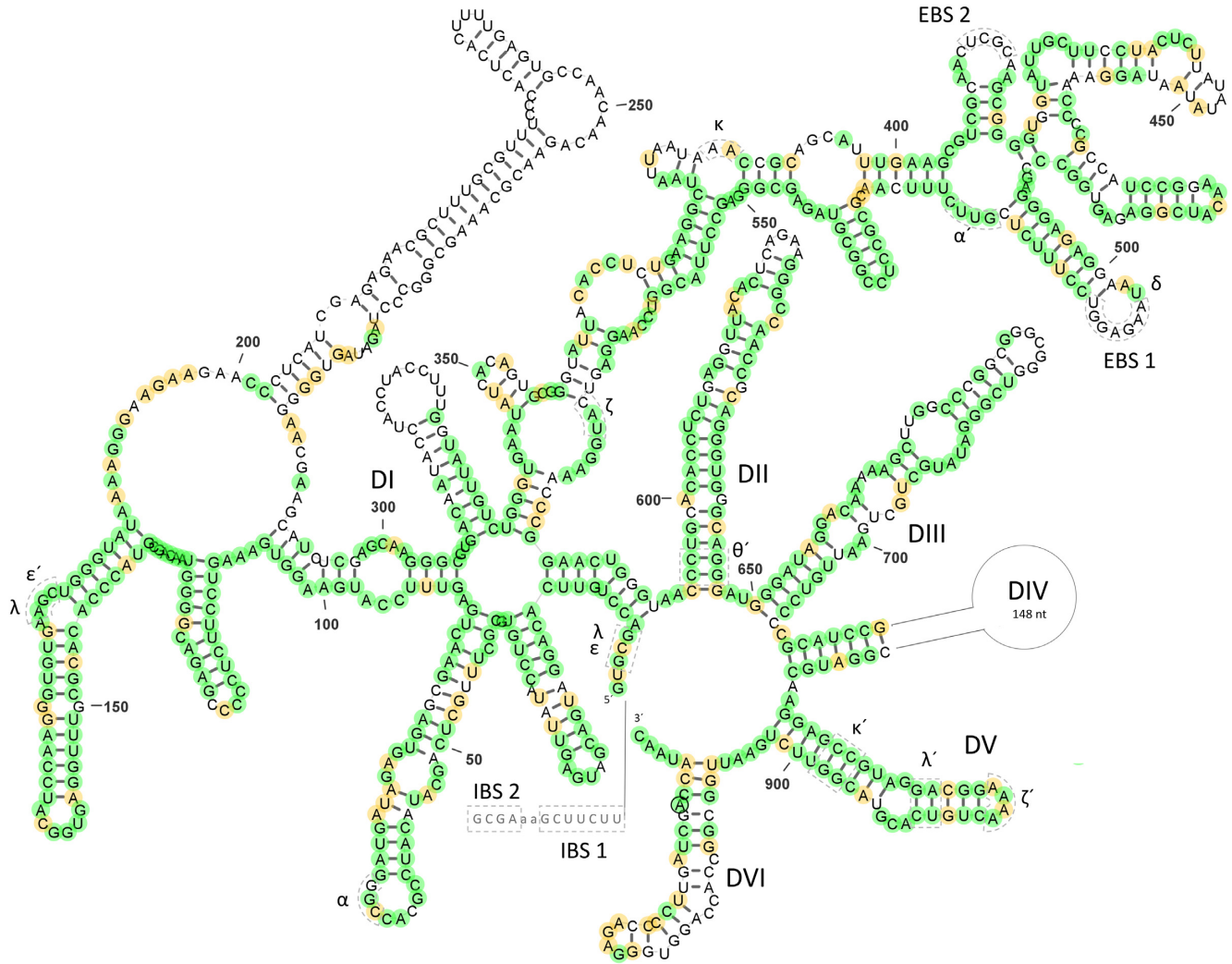

Figure 2. The now identified internal intron atp9i87g2ii1114g2 in Lycopodiaceae shares extensive sequence similarity with intron paralogues rps14i114g2 and nad7i1113g2 (Supplementary Figure S3). Shown is the secondary structure of atp9i87g2ii1114g2 in *Phlegmariurus squarrosus* with nucleotide identities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

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

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

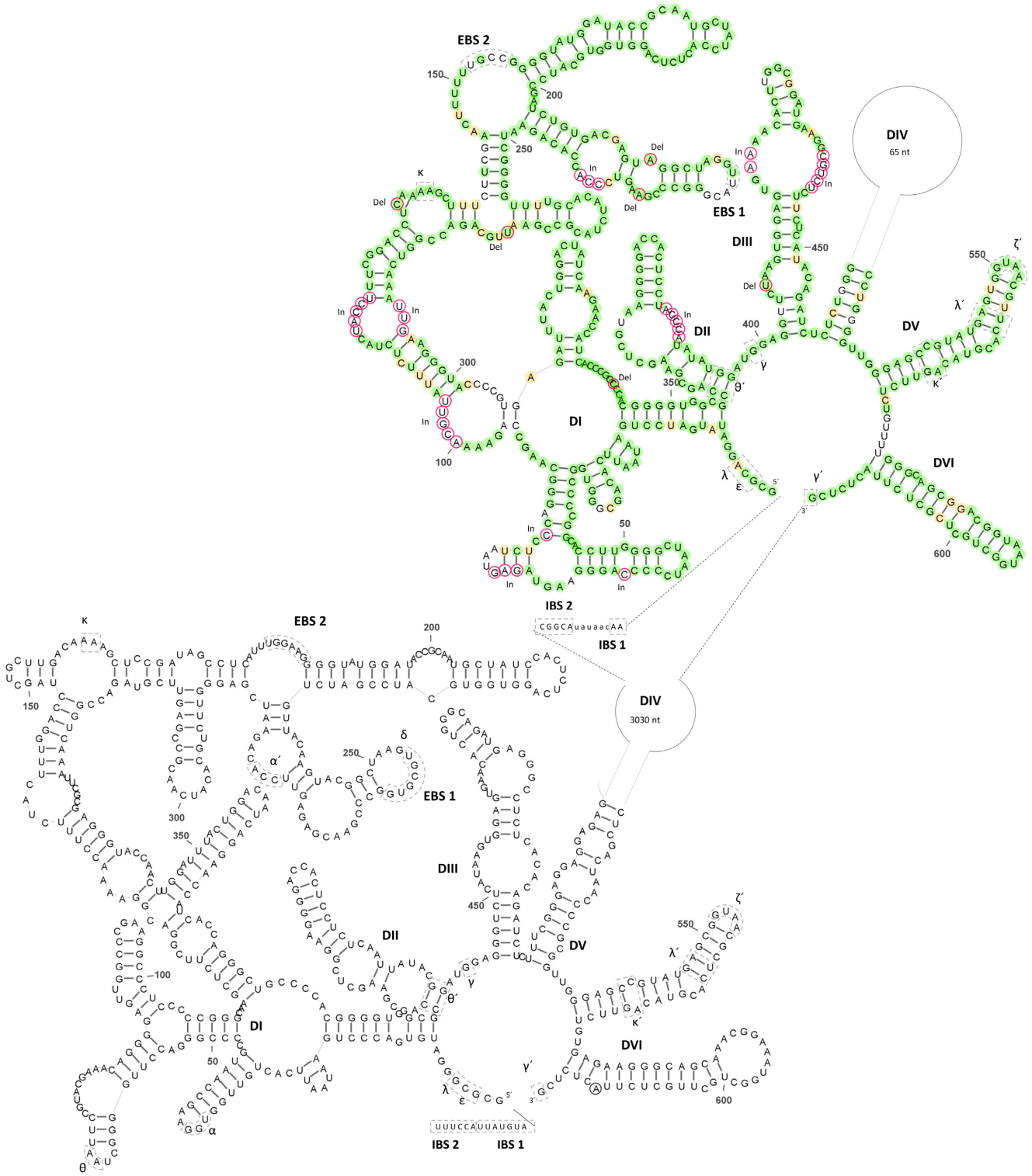

Figure 3. The cox1i1149g2 zombie twintron in lycophytes. Group II intron cox1i1149g2 is unique to lycophytes, previously identified only in the chondromes of *Selaginella moellendorffii* and *Phlegmariurus squarrosus*. Secondary structure modeling now characterizes cox1i1149g2 as a unique twintron in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

Hypermobility creating multiple cox1i1149g2 insertions

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

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

The case of sdh3i349g2-twin in Lycopodiaceae

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

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

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

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

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

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

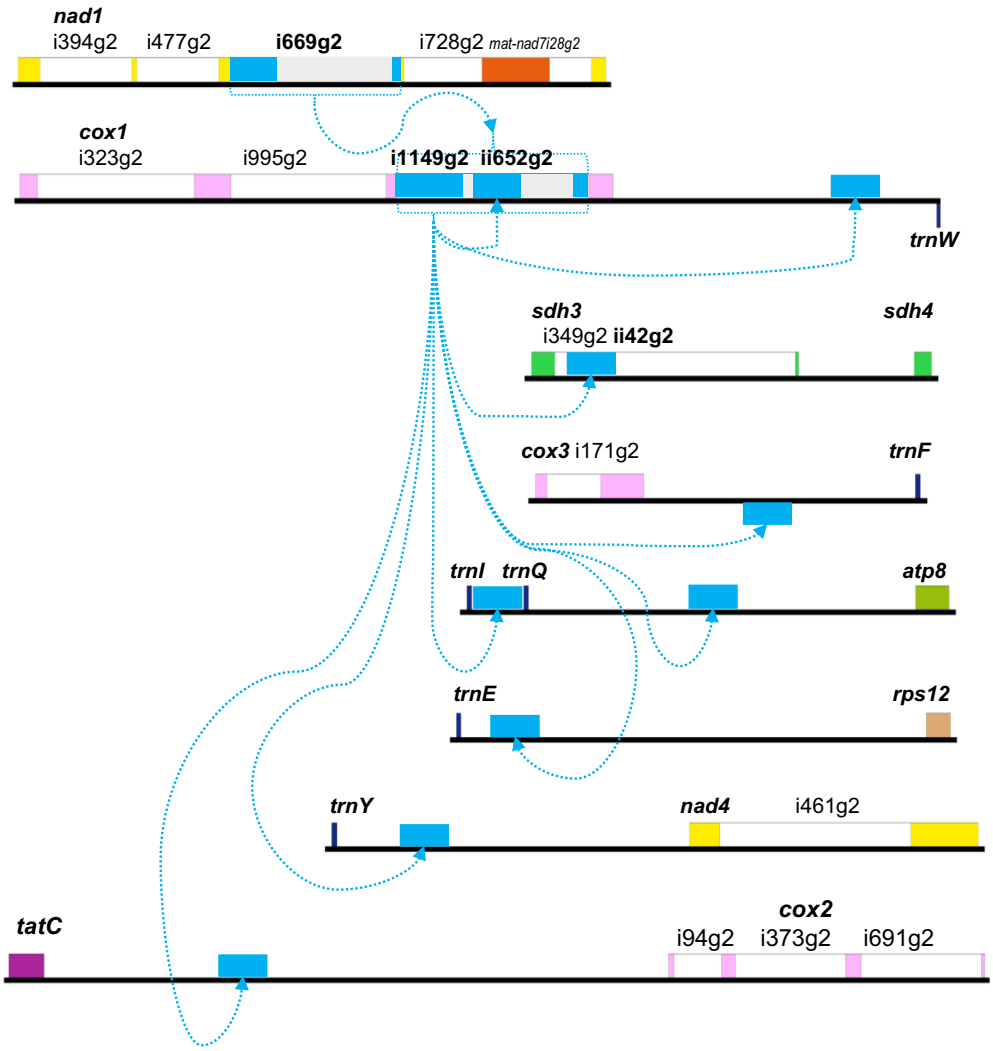

Figure 4. An invasive, hypermobile group II intron. Sequences with significant similarities to the internal and external group II intron of cox1i1149g2-twin (Figure 3) occur multiple times elsewhere in the *Phlegmariurus squarrosus* mtDNA (Supplementary Figure S4). Group II intron nad1i669g2 conserved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

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

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

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

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

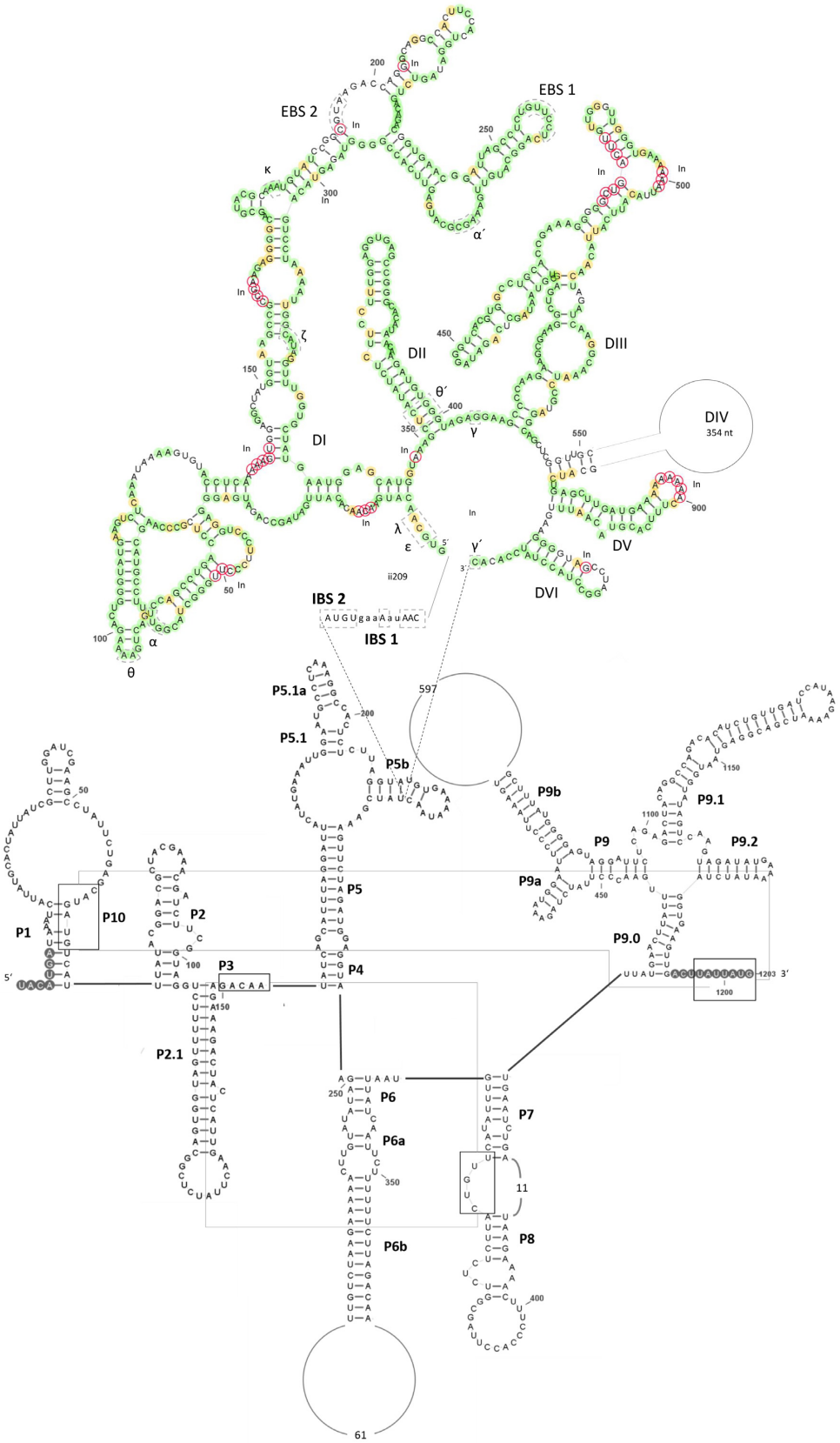

Figure 5. A mixed group I/group II twintron. Secondary structure modeling of group I intron cox1i1116g1 in *Anthoceros angustus* (database accession NC_037476) reveals a twintron arrangement with an internal group II intron (cox1i1116g1ii209g2) inserted in the P5 structure of the external group I intron.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

Evolutionary origins of twintrons and other intron paralogues

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

DISCUSSION

Group II introns and their splicing factors

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

Mobile group II introns in plant mitochondria

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

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

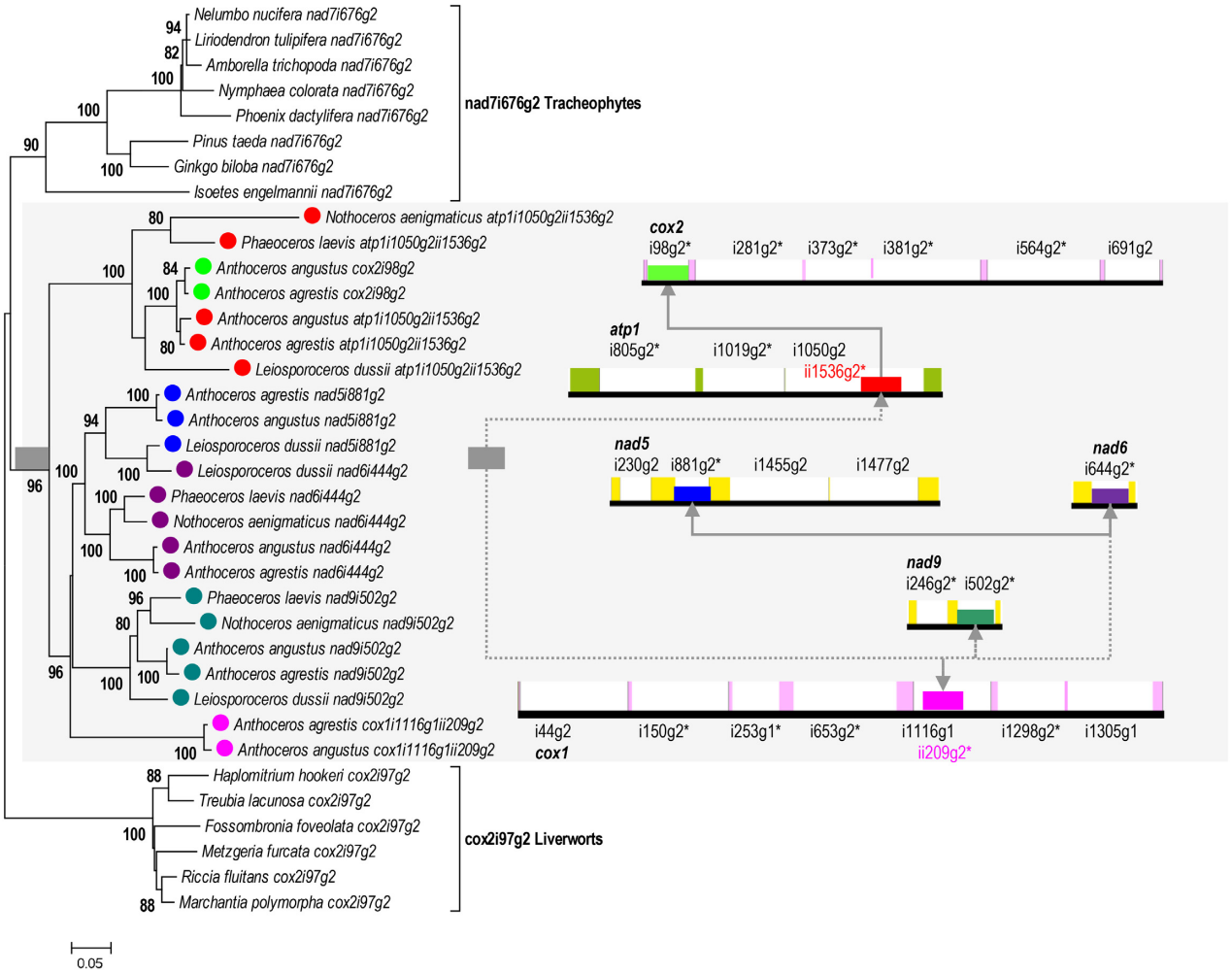

Figure 6. Hornwort twintrons share a common origin with four other group II intron paralogues specific to this clade. The internal introns atp1i1050g2ii1536g2 and cox1i1116g1ii209g2 in the now identified hornwort twintrons share a common origin with four other group II intron paralogues (cox2i98g2, nad5i881g2,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

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

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

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

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

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

Independent origins of twintrons in diverse organisms

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

Plant mitochondria as a playground for group II intron evolution

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

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

DATA AVAILABILITY

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge computing time provided by the Paderborn Center for Parallel computing 'PC2' and in particular the kind support of Axel Keller. We wish to thank Tim van Beers very much for suggesting 'zombie' as a label for the degenerated twintrons and the 2018 laboratory course WP04B on plant molecular biology for contributions to the analysis of the Lycopodiaceae twintron arrangements. Also, we are very grateful to our colleagues at the Bonn University Botanic Garden for providing Lycopodiaceae plant material.

Author contributions: S.Z. and P.G. did analysis of sequence data, S.Z. did nucleic acid preparations, PCR amplifications and molecular cloning, V.K. designed and supervised the study and wrote the manuscript, S.Z. and V.K. prepared figures and edited the final manuscript.

FUNDING

Funding for open access charge: Basic university household funding.

Conflict of interest statement. None declared.

REFERENCES

1. Novikova,O. and Belfort,M. (2017) Mobile group II introns as ancestral eukaryotic elements. *Trends Genet.*, **33**, 773–783.
2. Smathers,C.M. and Robart,A.R. (2019) The mechanism of splicing as told by group II introns: ancestors of the spliceosome. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1862**, 194390.
3. Vosseberg,J. and Snel,B. (2017) Domestication of self-splicing introns during eukaryogenesis: the rise of the complex spliceosomal machinery. *Biol. Direct*, **12**, 30.
4. Galej,W.P., Toor,N., Newman,A.J. and Nagai,K. (2018) Molecular mechanism and evolution of nuclear pre-mRNA and group II intron splicing: insights from cryo-electron microscopy structures. *Chem. Rev.*, **118**, 4156–4176.
5. Agrawal,R.K., Wang,H.-W. and Belfort,M. (2016) Forks in the tracks: group II introns, spliceosomes, telomeres and beyond. *RNA Biol.*, **13**, 1218–1222.
6. McNeil,B.A., Semper,C. and Zimmerly,S. (2016) Group II introns: versatile ribozymes and retroelements. *Wiley Interdiscip. Rev. RNA*, **7**, 341–355.
7. Haack,D.B., Yan,X., Zhang,C., Hingey,J., Lyumkis,D., Baker,T.S. and Toor,N. (2019) Cryo-EM structures of a group II intron reverse splicing into DNA. *Cell*, **178**, 612–623.
8. Belfort,M. and Lambowitz,A.M. (2019) Group II Intron RNPs and reverse transcriptases: from retroelements to research tools. *Cold Spring Harb. Perspect. Biol.*, **11**, a032375.
9. Bock,R. and Knoop,V. (2012) In: Bock,R and Knoop,V (eds). *Genomics of Chloroplasts and Mitochondria Auflage: 2*. Springer, Dordrecht.
10. Pruchner,D., Nassal,B., Schindler,M. and Knoop,V. (2001) Mosses share mitochondrial group II introns with flowering plants, not with liverworts. *Mol. Genet. Genomics*, **266**, 608–613.
11. Pruchner,D., Beckert,S., Muhle,H. and Knoop,V. (2002) Divergent intron conservation in the mitochondrial *nad2* gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. *J. Mol. Evol.*, **55**, 265–271.
12. Qiu,Y.L., Cho,Y., Cox,J.C. and Palmer,J.D. (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature*, **394**, 671–674.
13. Qiu,Y.-L., Li,L., Wang,B., Chen,Z., Knoop,V., Groth-Malonek,M., Dombrovska,O., Lee,J., Kent,L., Rest,J. *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15511–15516.
14. Volkmar,U., Groth-Malonek,M., Heinrichs,J., Muhle,H., Polsakiewicz,M. and Knoop,V. (2012) Exclusive conservation of mitochondrial group II intron nad4i548 among liverworts and its use for phylogenetic studies in this ancient plant clade. *Plant Biol. (Stuttg.)*, **14**, 382–391.
15. Groth-Malonek,M., Pruchner,D., Grewe,F. and Knoop,V. (2005) Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol. Biol. Evol.*, **22**, 117–125.
16. Knoop,V. (2010) Looking for sense in the nonsense: a short review of non-coding organellar DNA elucidating the phylogeny of bryophytes. *Trop. Biol.*, **31**, 51–60.
17. Knie,N., Grewe,F. and Knoop,V. (2016) Monilophyte mitochondrial *rps1* genes carry a unique group II intron that likely originated from an ancient paralog in *rpl2*. *RNA*, **22**, 1338–1348.
18. Zumkeller,S.M., Knoop,V. and Knie,N. (2016) Convergent evolution of fern-specific mitochondrial group II intron atp1i361g2 and its ancient source paralogue rps3i249g2 and independent losses of intron and RNA editing among Pteridaceae. *Genome Biol. Evol.*, **8**, 2505–2519.
19. Knoop,V. (2013) Plant mitochondrial genome peculiarities evolving in the earliest vascular plant lineages. *J. Syst. Evol.*, **51**, 1–12.
20. Liu,Y., Wang,B., Cui,P., Li,L., Xue,J.-Y., Yu,J. and Qiu,Y.-L. (2012) The mitochondrial genome of the lycophyte *huperzia squarrosa*: the most archaic form in vascular plants. *PLoS One*, **7**, e35168.
21. Grewe,F., Viehoever,P., Weisshaar,B. and Knoop,V. (2009) A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.*, **37**, 5093–5104.
22. Hecht,J., Grewe,F. and Knoop,V. (2011) Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol. Evol.*, **3**, 344–358.
23. Villarreal Aguilar,J.C., Turmel,M., Bourgouin-Couture,M., Laroche,J., Salazar Allen,N., Li,F.-W., Cheng,S., Renzaglia,K. and Lemieux,C. (2018) Genome-wide organellar analyses from the hornwort *Leiosporoceros dussii* show low frequency of RNA editing. *PLoS One*, **13**, e0200491.
24. Xue,J.Y., Liu,Y., Li,L., Wang,B. and Qiu,Y.L. (2010) The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr. Genet.*, **56**, 53–61.
25. Yoshinaga,K., Iinuma,H., Masuzawa,T. and Uedal,K. (1996) Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res.*, **24**, 1008–1014.
26. Dong,S., Wu,H., Zhang,S., Zhang,L., Liu,Y., Xue,J.Y., Chen,Z. and Goffinet,B. (2018) Complete mitochondrial genome sequence of *Anthoceros angustus*: conservative evolution of the mitogenomes in hornworts. *Bryologist*, **121**, 14–22.
27. Cargill,D.C., Renzaglia,K.S., Villarreal,J.C. and Duff,R.J. (2005) Generic concepts within hornworts: historical review, contemporary insights and future directions. *Aust. Syst. Bot.*, **18**, 7–16.
28. Copertino,D.W. and Hallick,R.B. (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. *EMBO J.*, **10**, 433–442.
29. Copertino,D.W., Christopher,D.A. and Hallick,R.B. (1991) A mixed group II/group III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3 gene: evidence for intron insertion during gene evolution. *Nucleic Acids Res.*, **19**, 6491–6497.
30. Copertino,D.W., Hall,E.T., Van Hook,F.W., Jenkins,K.P. and Hallick,R.B. (1994) A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res.*, **22**, 1029–1036.

31. Hafez,M. and Hausner,G. (2015) Convergent evolution of twintron-like configurations: one is never enough. *RNA Biol.*, **12**, 1275–1288.
32. Li,D., Liu,C.-M., Luo,R., Sadakane,K. and Lam,T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**, 1674–1676.
33. Leinonen,R., Sugawara,H., Shumway,M. and International Nucleotide Sequence Database Collaboration (2011) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
34. Grabherr,M.G., Haas,B.J., Yassour,M., Levin,J.Z., Thompson,D.A., Amit,I., Adiconis,X., Fan,L., Raychowdhury,R., Zeng,Q. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, **29**, 644–652.
35. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
36. Field,A.R., Testo,W., Bostock,P.D., Holtum,J.A.M. and Waycott,M. (2015) Molecular phylogenetics and the morphology of the Lycopodiaceae subfamily Huperzioideae supports three genera: *Huperzia, Phlegmariurus* and *Phylloglossum*. *Mol. Phylogenet. Evol.*, **94**, 635–657.
37. Doyle,J.L. and Doyle,J.J. (1990) Isolation of plant DNA from fresh tissue. *Focus (Madison)*, **12**, 13–15.
38. Liao,Z., Chen,M., Guo,L., Gong,Y., Tang,F., Sun,X. and Tang,K. (2004) Rapid isolation of high-quality total RNA from taxus and ginkgo. *Prep. Biochem. Biotechnol.*, **34**, 209–214.
39. Kumar,S., Stecher,G. and Tamura,K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, **33**, 1870–1874.
40. Zimmerly,S. and Semper,C. (2015) Evolution of group II introns. *Mob. DNA*, **6**, 7.
41. Michel,F., Umesono,K. and Ozeki,H. (1989) Comparative and functional anatomy of group II catalytic introns–a review. *Gene*, **82**, 5–30.
42. Fedorova,O. and Zingler,N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.*, **388**, 665–678.
43. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
44. Zhao,C. and Pyle,A.M. (2017) The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand. *Curr. Opin. Struct. Biol.*, **47**, 30–39.
45. Blocker,F.J.H., Mohr,G., Conlan,L.H., Qi,L., Belfort,M. and Lambowitz,A.M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*, **11**, 14–28.
46. Dombrovska,E. and Qiu,Y.-L. (2004) Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Mol. Phylogenet. Evol.*, **32**, 246–263.
47. Knoop,V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.*, **46**, 123–139.
48. Guo,W. and Mower,J.P. (2013) Evolution of plant mitochondrial intron-encoded maturases: frequent lineage-specific loss and recurrent intracellular transfer to the nucleus. *J. Mol. Evol.*, **77**, 43–54.
49. Li,L., Wang,B., Liu,Y. and Qiu,Y.L. (2009) The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *J. Mol. Evol.*, **68**, 665–678.
50. Grewe,F., Herres,S., Viehöver,P., Polsakiewicz,M., Weisshaar,B. and Knoop,V. (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.*, **39**, 2890–2902.
51. Groth-Malonek,M., Wahrmund,U., Polsakiewicz,M. and Knoop,V. (2007) Evolution of a pseudogene: exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol. Biol. Evol.*, **24**, 1068–1074.
52. Adams,K.L., Rosenblueth,M., Qiu,Y.L. and Palmer,J.D. (2001) Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics*, **158**, 1289–1300.
53. Gerke,P., Szövényi,P., Neubauer,A., Lenz,H., Gutmann,B., McDowell,R., Small,I., Schallenberg-Rüdinger,M. and Knoop,V. (2019) Towards a plant model for enigmatic U-to-C RNA editing: the organelle genomes, transcriptomes, editomes and candidate RNA editing factors in the hornwort *Anthoceros agrestis*. *New Phytol.*, doi:10.1111/nph.16297.
54. Stoddard,B.L. (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA*, **5**, 7.
55. Taylor,G.K. and Stoddard,B.L. (2012) Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. *Nucleic Acids Res.*, **40**, 5189–5200.
56. Fekete,E., Flipphi,M., Ág,N., Kavalecz,N., Cerqueira,G., Scazzocchio,C. and Karaffa,L. (2017) A mechanism for a single nucleotide intron shift. *Nucleic Acids Res.*, **45**, 9085–9092.
57. Bainard,J.D. and Villarreal,J.C. (2013) Genome size increases in recently diverged hornwort clades. *Genome*, **56**, 431–435.
58. Duff,R.J., Villarreal,J.C., Cargill,D.C. and Renzaglia,K.S. (2007) Progress and challenges toward developing a phylogeny and classification of the hornworts. *Bryologist*, **110**, 214–243.
59. Schmitz-Linneweber,C., Lampe,M.-K.K., Sultan,L.D. and Ostersetzer-Biran,O. (2015) Organellar maturases: a window into the evolution of the spliceosome. *Biochim. Biophys. Acta*, **1847**, 798–808.
60. Zoschke,R., Nakamura,M., Liere,K., Sugiura,M., Börner,T. and Schmitz-Linneweber,C. (2010) An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 3245–3250.
61. Sultan,L.D., Mileshina,D., Grewe,F., Rolle,K., Abudraham,S., Głodowicz,P., Niazi,A.K., Keren,I., Shevtsov,S., Klipcan,L. *et al.* (2016) The reverse transcriptase/rna maturase protein matr is required for the splicing of various group II introns in Brassicaceae mitochondria. *Plant Cell*, **28**, 2805–2829.
62. Zmudjak,M., Colas des Francs-Small,C., Keren,I., Shaya,F., Belausov,E., Small,I. and Ostersetzer-Biran,O. (2013) mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in *Arabidopsis*. *New Phytol.*, **199**, 379–394.
63. Chen,Z., Wang,H.-C., Shen,J., Sun,F., Wang,M., Xu,C. and Tan,B.-C. (2019) PPR-SMR1 is required for the splicing of multiple mitochondrial introns, interacts with Zm-mCSF1, and is essential for seed development in maize. *J. Exp. Bot.*, **70**, 5245–5258.
64. Francs-Small,C.C., Kroeger,T., Zmudjak,M., Ostersetzer-Biran,O., Rahimi,N., Small,I. and Barkan,A. (2012) A PORR domain protein required for *rpl2* and *ccmF(C)* intron splicing and for the biogenesis of c-type cytochromes in *Arabidopsis* mitochondria. *Plant J. Cell Mol. Biol.*, **69**, 996–1005.
65. Dickson,L., Huang,H.-R., Liu,L., Matsuura,M., Lambowitz,A.M. and Perlman,P.S. (2001) Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc. Natl. Acad. Sci.*, **98**, 13207–13212.
66. Eskes,R., Liu,L., Ma,H., Chao,M.Y., Dickson,L., Lambowitz,A.M. and Perlman,P.S. (2000) Multiple homing pathways used by yeast mitochondrial group II introns. *Mol. Cell Biol.*, **20**, 8432–8446.
67. Cusimano,N., Zhang,L.B.L.-B. and Renner,S.S. (2008) Reevaluation of the cox1 group I intron in Araceae and angiosperms indicates a history dominated by loss rather than horizontal transfer. *Mol. Biol. Evol.*, **25**, 265–276.
68. Sanchez-Puerta,M.V., Abbona,C.C., Zhuo,S., Tepe,E.J., Bohs,L., Olmstead,R.G. and Palmer,J.D. (2011) Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evol. Biol.*, **11**, 277.
69. Cho,Y. and Palmer,J.D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol.*, **16**, 1155–1165.
70. Cho,Y., Qiu,Y.L., Kuhlman,P. and Palmer,J.D. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14244–14249.
71. Sanchez-Puerta,M.V., Cho,Y., Mower,J.P., Alverson,A.J. and Palmer,J.D. (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Mol. Biol. Evol.*, **25**, 1762–1777.
72. Vaughn,J.C., Mason,M.T., Sper-Whitis,G.L., Kuhlman,P. and Palmer,J.D. (1995) Fungal origin by horizontal transfer of a plant

mitochondrial group I intron in the chimeric CoxI gene of *Peperomia*. *J. Mol. Evol.*, **41**, 563–572.
73. Drager,R.G. and Hallick,R.B. (1993) A complex twintron is excised as four individual introns. *Nucleic Acids Res.*, **21**, 2389–2394.
74. Copertino,D.W., Shigeoka,S. and Hallick,R.B. (1992) Chloroplast group III twintron excision utilizing multiple 5′- and 3′-splice sites. *EMBO J.*, **11**, 5041–5050.
75. Pombert,J.-F., James,E.R., Janouškovec,J. and Keeling,P.J. (2012) Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorphina aenigmatica* plastid genome. *PLoS One*, **7**, e53433.
76. Thompson,M.D., Copertino,D.W., Thompson,E., Favreau,M.R. and Hallick,R.B. (1995) Evidence for the late origin of introns in chloroplast genes from an evolutionary analysis of the genus *Euglena*. *Nucleic Acids Res.*, **23**, 4745–4752.
77. Thompson,M.D., Zhang,L., Hong,L. and Hallick,R.B. (1997) Two new group-II twintrons in the *Euglena gracilis* chloroplast are absent in basally branching *Euglena* species. *Curr. Genet.*, **31**, 89–95.
78. Maier,U.G., Rensing,S.A., Igloi,G.L. and Maerz,M. (1995) Twintrons are not unique to the euglena chloroplast genome - structure and evolution of a plastome cpn60 gene from a cryptomonad. *Mol. Gen. Genet.*, **246**, 128–131.
79. Schmidt,W.M., Schweyen,R.J., Wolf,K. and Mueller,M.W. (1994) Transposable Group II introns in fission and budding yeast. *J. Mol. Biol.*, **243**, 157–166.
80. Perrineau,M.-M., Price,D.C., Mohr,G. and Bhattacharya,D. (2015) Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. *PeerJ*, **3**, e1017.
81. Pfreundt,U. and Hess,W.R. (2015) Sequential splicing of a group II twintron in the marine cyanobacterium Trichodesmium. *Sci. Rep.*, **5**, 16829.
82. Johansen,S., Elde,M., Vader,A., Haugen,P., Haugli,K. and Haugli,F. (1997) In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete Didymium iridis. *Mol. Microbiol.*, **24**, 737–745.
83. Einvik,C., Elde,M. and Johansen,S. (1998) Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. *J. Biotechnol.*, **64**, 63–74.
84. Hafez,M., Majer,A., Sethuraman,J., Rudski,S.M., Michel,F. and Hausner,G. (2013) The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: Twintrons, introns, and intron-encoded proteins. *Fungal Genet. Biol.*, **53**, 71–83.
85. Guha,T.K. and Hausner,G. (2014) A homing endonuclease with a switch: Characterization of a twintron encoded homing endonuclease. *Fungal Genet. Biol.*, **65**, 57–68.
86. Deng,Y., Zhang,Q., Ming,R., Lin,L., Lin,X., Lin,Y., Li,X., Xie,B. and Wen,Z. (2016) Analysis of the mitochondrial genome in Hypomyces aurantius reveals a novel twintron complex in fungi. *Int. J. Mol. Sci.*, **17**, 1049.
87. Deng,Y., Hsiang,T., Li,S., Lin,L., Wang,Q., Chen,Q., Xie,B. and Ming,R. (2018) Comparison of the mitochondrial genome sequences of six annulohypoxylon stygium isolates suggests short fragment insertions as a potential factor leading to larger genomic size. *Front. Microbiol.*,**9**, 2079.
88. Kavalecz,N., Ág,N., Karaffa,L., Scazzocchio,C., Flipphi,M. and Fekete,E. (2019) A spliceosomal twin intron (stwintron) participates in both exon skipping and evolutionary exon loss. *Sci. Rep.*, **9**, 9940.
89. Scamborova,P., Wong,A. and Steitz,J.A. (2004) An intronic enhancer regulates splicing of the twintron of *Drosophila melanogaster* prospero pre-mRNA by two different spliceosomes. *Mol. Cell Biol.*, **24**, 1855–1869.
90. Janice,J., Jąkalski,M. and Makałowski,W. (2013) Surprisingly high number of Twintrons in vertebrates. *Biol. Direct*, **8**, 4.
91. Malek,O. and Knoop,V. (1998) Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA*, **4**, 1599–1609.
92. Qiu,Y.L. and Palmer,J.D. (2004) Many independent origins of trans splicing of a plant mitochondrial group II intron. *J. Mol. Evol.*, **59**, 80–89.
93. Hepburn,N.J., Schmidt,D.W. and Mower,J.P. (2012) Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol. Biol. Evol.*, **29**, 3111–3120.
94. Grewe,F., Zhu,A. and Mower,J.P. (2016) Loss of a *trans*-splicing *nad1* intron from Geraniaceae and transfer of the maturase gene *matr* to the nucleus in *Pelargonium*. *Genome Biol. Evol.*, **8**, 3193–3201.
95. Guo,W., Grewe,F., Fan,W., Young,G.J., Knoop,V., Palmer,J.D. and Mower,J.P. (2016) *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.*, **33**, 1448–1460.
96. Brown,G.G., Colas des Francs-Small,C. and Ostersetzer-Biran,O. (2014) Group II intron splicing factors in plant mitochondria. *Front. Plant Sci.*, **5**, 35.
97. Peters,J.K. and Toor,N. (2015) Group II intron lariat: Structural insights into the spliceosome. *RNA Biol.*, **12**, 913–917.
98. Rogozin,I.B., Carmel,L., Csuros,M. and Koonin,E.V (2012) Origin and evolution of spliceosomal introns. *Biol. Direct*, **7**, 11.
99. Hein,A., Brenner,S. and Knoop,V. (2019) Multifarious evolutionary pathways of a nuclear RNA editing factor: disjunctions in co-evolution of DOT4 and its chloroplast target rpoC1eU488SL. *Genome Biol. Evol.*, **11**, 798–813.
100. Szövényi,P., Frangedakis,E., Ricca,M., Quandt,D., Wicke,S. and Langdale,J.A. (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biol.*, **15**, 98.
101. Greiner,S., Lehwark,P. and Bock,R. (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.*, **47**, W59–W64.