

# 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

Julia Hecht, Felix Grewe, and Volker Knoop\*

Abteilung Molekulare Evolution, Institut für Zelluläre und Molekulare Botanik, Universität Bonn, D-53115 Bonn, Germany

\*Corresponding author: E-mail: volker.knoop@uni-bonn.de.

Hecht and Grewe contributed equally to this work.

**Accepted:** 17 March 2011

## Abstract

Using an independent fosmid cloning approach and comprehensive transcriptome analysis to complement data from the *Selaginella moellendorffii* genome project, we determined the complete mitochondrial genome structure of this spikemoss. Numerous recombination events mediated mainly via long sequence repeats extending up to 7kbp result in a complex mtDNA network structure. Peculiar features associated with the repeat sequences are more than 80 different microsatellite sites (predominantly trinucleotide motifs). The *S. moellendorffii* mtDNA encodes a plant-typical core set of a twin-arginine translocase (*tatC*), 17 respiratory chain subunits, and 2 rRNAs but lacks *atp4* and any tRNA genes. As a further novelty among plant chondromes, the *nad4L* gene is encoded within an intron of the *nad1* gene. A total of 37 introns occupying the 20 mitochondrial genes (four of which are disrupted into *trans*-splicing arrangements including two novel instances of *trans*-splicing introns) make the *S. moellendorffii* chondrome the intron-richest and gene-poorest plant mtDNA known. Our parallel transcriptome analyses demonstrates functional splicing of all 37 introns and reveals a new record amount of plant organelle RNA editing with a total of 2,139 sites in mRNAs and 13 sites in the two rRNAs, all of which are exclusively of the C-to-U type.

**Key words:** lycophytes, RNA editing, *trans*-splicing, microsatellites, endosymbiotic gene transfer.

## Introduction

Plant mitochondrial genomes have become legendary for the ever-growing list of molecular peculiarities discovered over the last 3 decades. The mitochondrial genomes (chondromes) of plants exceed the ones of animals in size by two orders of magnitude and exceptionally even more, occasionally resulting in plant chondromes of bacterial genome sizes (Ward et al. 1981; Alverson et al. 2010). Most of the “extra” DNA in plant chondromes is intergenic, in parts derived from foreign sequences acquired through transfer from the nucleus or the chloroplast or even horizontally from mtDNA of other species (Bergthorsson et al. 2003). Other factors increasing plant mitochondrial genome sizes are additional genes not present in the animal lineage and organellar introns belonging to the group I or group II classes (Knoop et al. 2010). Moreover, frequent DNA recombination shuffling gene orders continuously

and producing coexisting alternative genomic arrangements are typical of flowering plant genomes. Such recombination events have also led to rearrangements not only between but also within genes and have produced *trans*-splicing introns in the course of plant evolution (Malek and Knoop 1998). Finally, the plant organelle type of RNA editing exchanges the pyrimidine nucleotides cytidine and uridine at specific transcript sites and thus alters the encoded information at the RNA level (Knoop 2010).

Many of the peculiar plant chondrome features—including RNA editing, genomic recombination, *trans*-splicing, and insertions of foreign DNA—are absent, however, in green algae and early branching land plant lineages. For example, neither of these four mtDNA features exists in the mtDNA of the liverwort *Marchantia polymorpha* (Oda et al. 1992). RNA editing, however, is a very early gain in

land plant evolution, present in other bryophyte clades and apparently only secondarily lost in the marchantiid liverworts (Malek et al. 1996; Steinhauser et al. 1999; Groth-Maloney et al. 2007). No evidence for insertions of chloroplast, nuclear or foreign DNA, for active DNA recombination or for *trans*-splicing introns has so far been identified, however, in any bryophyte chondrome including the completely determined mtDNA sequences of a moss (Terasawa et al. 2006), a jungermanniid liverwort (Wang et al. 2009) and two hornworts (Li et al. 2009; Xue et al. 2010). In contrast, all of these genomic peculiarities were recently shown to exist in the mitochondrial DNA (mtDNA) of a lycophyte, the quillwort *Isoetes engelmannii* (Grewe et al. 2009).

Lycophytes are the sister clade to all other vascular plants (tracheophytes), that is, the euphyllophytes, which comprise the spermatophytes (seed plants) and the monilophytes (true ferns, whisk ferns, and horsetails). Hence, lycophytes represent the most ancient surviving lineage of tracheophytes that dominated the earth's flora since Devonian times, that is, much more than 200 My before the rise of flowering plants in the cretaceous some 140 My ago. Highly active chondrome DNA recombination, *trans*-splicing introns, and the incorporation of foreign DNA all seem to arise with the earliest tracheophytes as reflected with the existence of all these phenomena in the mtDNA of *Isoetes*.

Extant lycophytes comprise three well-defined orders, the Isoetales (quillworts), the Selaginellales (spike mosses), and the Lycopodiales (club mosses). Given their pivotal position in the evolution of land plants and in the light of the chondrome peculiarities in *I. engelmannii* (Grewe et al. 2009), we wished to elucidate the mtDNA structure of a sister lycophyte. To this end, we made initial use of available mitochondrial sequence data from the *Selaginella moellendorffii* genome sequencing project. With support from an independent fosmid cloning approach, we ultimately determined the *S. moellendorffii* mitochondrial genome structure, determined its complete mtDNA sequence and complemented this by exhaustive cDNA analyses.

To our surprise, we found that the *Selaginella* mtDNA reveals many differences to its sister lycophyte *Isoetes* showing that quite different pathways have been followed in mitochondrial genome evolution in the two ancient tracheophyte lineages. Both taxa are characterized by abundant RNA editing and, with more than 2,000 sites of C-to-U RNA editing, *Selaginella* actually even breaks the record of editing abundance hitherto set by *Isoetes* (Grewe et al. 2011). However, whereas lots of U-to-C editing sites were found in *Isoetes*, no single such event was detected in *Selaginella*. Both taxa are characterized by frequent DNA recombination but as a novelty, the sequence repeats involved in recombination are associated with microsatellite repeats in *Selaginella*, which have no counterparts in *Isoetes*. Introns in *Isoetes* mtDNA are tiny, whereas they are huge in *Selaginella* and while promiscuous DNA of nuclear and

chloroplast origin had been identified in the *Isoetes* mtDNA, no such sequences exist in the *Selaginella* mtDNA. Furthermore, and as a complete novelty for a plant chondrome, the *S. moellendorffii* mtDNA is devoid of any tRNA encoding gene.

## Materials and Methods

### Molecular Cloning

*Selaginella moellendorffii* plant material was obtained commercially (Shady Plants Fern Nursery, Coolbooa, Clashmore, Co, Waterford, Ireland) and further cultivated in the laboratory. Total genomic DNA was isolated using a CTAB protocol. After size fractionation into approximately 38kbp fragments, DNA was blunt ended and cloned into the fosmid vector pCC1FOS using the CopyControl Fosmid Library Production Kit (EPICENTRE). A library of 5,000 fosmid clones was obtained, of which 1,500 were manually sorted, filter spotted, and hybridized with P<sup>32</sup>-labelled gene probes (reverse transcriptase–polymerase chain reaction [RT-PCR] derived for *nad2*, *nad9*, *cob*, *cox2*, and *atp9*). Identity of four positive fosmid clones was verified through terminal insert sequencing, showing homology with two available *S. moellendorffii* scaffold sequences (M162 and 213). Fosmid clone 3 was used for sublibrary production after revealing a PCR product for the *S. moellendorffii* mitochondrial 26S rRNA gene present in database entry GQ246802 but lacking from scaffold sequences M162 and 213. Fosmid DNA was isolated using NucleoBond Xtra Midi EF Kit (Macherey Nagel), sheared by Nebulizers (Invitrogen), blunted using an End-It DNA End-Repair Kit (EPICENTRE), A-tailed with Taq-Polymerase (Genaxxon), and fractionated by preparative electrophoresis in 0.8% agarose. Fragments of 2–2.5 kb in size were recovered using the NucleoSpin Extract II Kit (Macherey Nagel) and cloned into pGEM-T Easy vector (Promega). Ninety-six plasmid subclones were end sequenced and assembled, and primer walking was used to complete the sequence gap including the 26S rRNA.

Total *S. moellendorffii* RNA was isolated using the NucleoSpin RNA Plant Kit (Macherey Nagel), cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas) in the presence of random hexamer primers as recommended by the manufacturer. Oligonucleotides for RT-PCR were designed to ideally target 5'- and 3'-flanking untranslated regions (UTRs) to allow for amplification of full-coding regions of all genes, which was successful except for the 5'-UTRs of *atp6* and *atp9* and the 3'-UTR of *cox1*. Oligonucleotide pairs (all sequence information available from the authors upon request) were used for PCR amplification according to the standard protocol of GoTaq DNA Polymerase (Promega) in a GeneAmp PCR System 2700 (Applied Biosystems) with annealing temperatures between 50 °C and 55 °C. Amplicons were recovered from agarose gels as described and cloned into pGEM T Easy vector (Promega).

## Sequence Analysis

Mitochondrial protein coding regions were initially identified with sensitive (minimum word size) TblastN and BlastN similarity searches (Altschul et al. 1997) using an artificial concatenation of encoded proteins in the *M. polymorpha* and *Chara vulgaris* mitochondrial genomes and of structural RNA genes (tRNAs + rRNAs) of *Physcomitrella patens* mtDNA against the available sequence data of the *S. moellendorffii* genome project at <http://genome.jgi-psf.org/Selmo1/>. The tRNAscan-SE server was additionally used to scan for tRNA genes and pseudogenes (Lowe and Eddy 1997). Further sequence handling was essentially done using the alignment explorer of the MEGA software (Kumar et al. 2008). Identification and refinement of gene structures for annotation was aided by parallel cDNA analysis. Repeat sequences and microsatellites were analyzed using REPUTER at <http://bibiserv.techfak.uni-bielefeld.de/reputer/> and MSATFINDER at <http://www.genomics.ceh.ac.uk/msatfinder/>. Identification of RNA candidate sites in DNA and analyses and annotation of RNA editing on cDNA sequences was done with the help of PREPACT (Lenz et al. 2009). Assignment of group I and group II introns to their respective classes was based on inspection for conserved sequence signatures and identification of orthologues in the respective insertion sites in other plant taxa. Splice sites were determined by comparison with cDNA sequences.

## Sequence Assembly, Annotation, and Submission

The mitochondrial nature of *S. moellendorffii* Joint Genome Institute genome scaffolds M162 and 213 was verified by cDNA analyses of candidate gene-coding regions showing sequence colinearity except for *cis*- and *trans*-splicing events and perfect sequence identity except for numerous sites of C-to-U RNA editing. Likewise, terminal sequences of four fosmid clones independently retrieved from cloning DNA from a different biological source of *S. moellendorffii*, and the insert sequence from one of those (fosmid #3) were identical to the two scaffold sequences except for microsatellite repeat numbers as outlined under results. Despite an overall complex, *S. moellendorffii* mitochondrial genome structure (fig. 1) resulting from ten recombining sequence repeats R1–R10, the fosmid cloning suggested that 1) mtDNA molecules in a size range of at least 40kbp should indeed physically exist in nature and 2) that scaffold assemblies 213 and M162 were correct, not considering two extended gaps in the contig assembly of the latter. Ultimately, mtDNA sequences were submitted and annotated as five separate database entries (excluding M162 assembly gaps), each ending in active repeat sequences to limit sequence redundancies and to clearly suggest alternative coexisting mtDNA arrangements in nature (see figs. 1–3) while avoiding any suggestion of hitherto unproven large sequence continuities in a hypothetical “master circle”: Part 1: R9–R6a–*atp8*–*nad6*–*nad3*–R1–*cox3*–R2–*nad2*–R4 (42,231 bp,

accession no. JF338143, from scaffold 213), Part 2: R4–*atp9e2*–R8–*nad4*–R3a–R10–*cox1*–R1–*cox2e12*–*nad9*–R5–*rrnS*–R6–R9–*nad1*–R9–*cobe3*–*tatC*–R4–*nad5*–*atp1*–*cobe12*–R3 (143,606 bp, accession no. JF338144, from upstream sequence continuity in scaffold M162), Part 3: R6a–*nad7*–R7b–*atp9e34*–R8–*atp9e2*–R4b–*atp6*–R7 (35,649 bp, accession no. JF338145, from central sequence continuity in scaffold M162), Part 4: R6–*cox2e34*–R5 (20,004 bp, accession no. JF338146 from terminal sequence continuity in scaffold M162), and Part 5: R5–*rrnL*–R10 (accession no. JF338147, 19,723 bp from sequence insert in fosmid #3). Notably, the mitochondrial sequence gap closed with fosmid #3 insert (part 5) comes from a different biological source of *S. moellendorffii* and makes the total mtDNA complexity a “hybrid” in a strict sense, reflected by the recognizable microsatellite repeat variability (see Results). Database entry annotations include features on coding sequences, introns, microsatellites, and repeat regions. At present, scaffold 213 is also deposited in GenBank as entry NW\_003314473, however, unidentified as mitochondrial in nature and suggesting several hypothetical but very unlikely gene models. Seven separate mtDNA regions had earlier been submitted to the database without detailed annotation (accessions GQ264802–08) as *nad9*-like, *atp1*-like, *cox1*-like, *nad2*-like, *nad7*-like, *atp9*-like, and *nad4*-like, respectively, as a result of a previous *S. moellendorffii* study focusing on chloroplast DNA (Smith 2009). All cDNA sequences retrieved in our study, which confirm splice sites and reveal RNA editing positions, are submitted separately for each gene (database accessions JF276233–JF276250).

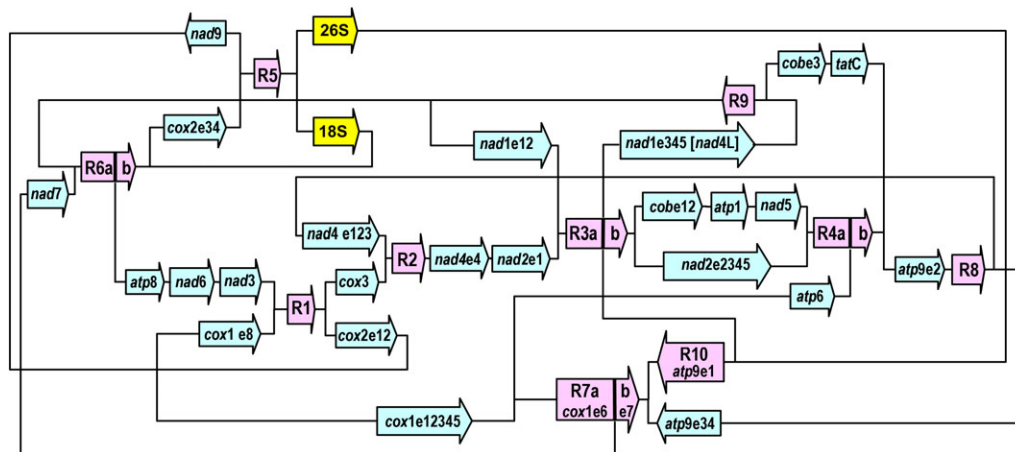
## Data Deposition

The *S. moellendorffii* mitochondrial genome sequences have been deposited in GenBank under accession numbers JF338143–JF338147 (parts 1–5). Accompanying cDNA sequences have been deposited under accession numbers JF276233–JF276250.

## Results

### Elucidating the Selaginella Chondrome Complexity

We started our experimental approach to determine the *S. moellendorffii* mitochondrial genome by creating an artificial protein sequence concatenation, which comprises all known plant mitochondrial protein sequences (essentially the *M. polymorpha* mitochondrial gene complement plus *nad7* and *rpl14* missing there that were taken from *P. patens*). Using this artificial concatenated query sequence, we scanned the publicly available *S. moellendorffii* genome sequence data by TblastN and identified scaffolds “213” of 46.7kbp and “M162” of 227kbp, which carried significant mitochondrial sequence homologies. Given that 1) scaffold M162 contained two gaps in the contig assembly (of 1,951 bp and 3,568 bp, respectively), that 2) no physical linkage between



**FIG. 1.**—The *Selaginella moellendorffii* mitochondrial genome structure is determined by ten recombinationally active sequence repeats (R1–R10, see fig. 2) linking the single-copy regions (a–r, see fig. 3) through a multitude of recombinational pathways. Repeats R3, R4, R6, and R7 possess additional recombinational breakpoints (a/b), further increasing mtDNA complexity. Repeated sequences are mostly intergenic with the exception of R7 extending over exons 6 and 7 of the *cox1* gene, R10 containing exon 1 of the *atp1* gene, and R9 extending into the first exon of *nad1*. Exon (e) numbers are indicated only where gene continuities are disrupted through rearrangements, introns are also present in *atp6*, *nad3*, *nad5*, *nad7*, and the 18S rRNA. Arrows indicate the directions of transcription for coding regions, which are mostly conserved across sequence repeats with the exception of *nad9*/R5, *cobe3*/R9, *tatC*/R4, and *cox1*/R7.

the two scaffolds was evident, and that 3) no evidence for a mitochondrial 26S rRNA was found in the sequence data (although obviously existing, as documented by database accession GQ246802, Smith 2009), we decided to complement the sequence analysis with an independent fosmid cloning approach of *S. moellendorffii* DNA. A small fosmid library of approximately 5,000 fosmid clones was obtained, of which 1,536 were manually sorted and screened with a mix of mitochondrial gene probes. Initial RT-PCR analyses performed in parallel had verified transcription (and RNA editing) in initial “bona fide” mitochondrial gene sequence candidates, and cDNA probes of five mitochondrial genes (*nad2*, *nad9*, *cob*, *cox2*, and *atp9*) were used for library screening and revealed four positive fosmid clones. Their eight end sequences perfectly mapped to the initial scaffold sequences with the exception of one end sequence of fosmid 2. End sequences of the other fosmids in fact mapped to both scaffolds and hence suggested that physical connections should exist in at least some recombined mtDNA arrangements. Continuous, full-length colinearity of fosmid inserts (around ca. 39kbp each) and the scaffold assemblies, however, was obviously absent in all four cases. This indicated frequent recombination events and suggested that the initial scaffold assemblies would reflect only one of several possible coexisting mitochondrial genome arrangements.

Attempts to identify 26S rRNA gene homologies on the fosmids via PCR revealed products of expected size in two of them and one (fosmid 3) was selected for sequencing. Fosmid 3 end sequences (36.3kbp insert size) were colinear with scaffold M162 but ran in opposite directions up to two recombination sites. Recombination points embedded 7.5kbp of novel sequence in fosmid 3 not represented in the

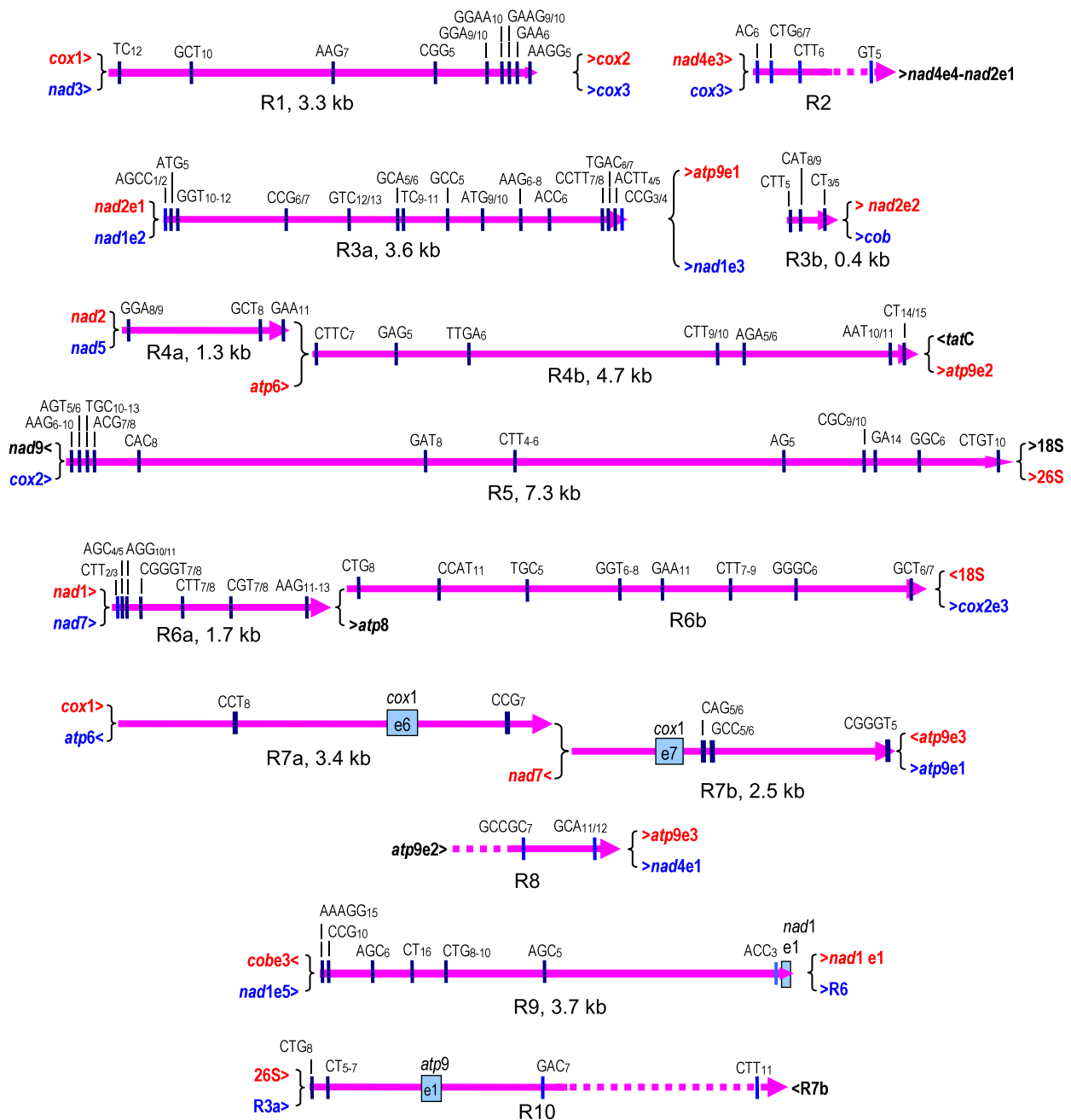
*S. moellendorffii* genome scaffold sequences and this included both the so far elusive 26S rRNA gene and the missing insert terminal sequences of fosmid 2.

We concluded 1) that sequence assembly of *S. moellendorffii* scaffolds was impeded by frequent recombination events in its mtDNA, resulting in apparent gaps in the M162 scaffold and 2) that unique sequences in the two separate scaffolds were in fact physically linked through recombination across sequence repeats in at least some coexisting topological alternatives.

### Repeats and Microsatellites

Inspecting all available sequence data, we found several large sequence repeats flanking different single-copy sequence regions. By combining the existing scaffold sequences and the new fosmid sequences, we were ultimately able to assemble a recombinational mitochondrial genome map (fig. 1). It turned out that the *S. moellendorffii* mtDNA complexity is essentially explained by ten large sequence repeats, labeled R1 through R10 (fig. 2), which provide alternative pathways through the flanking single-copy genomic regions (fig. 3). These large repeated sequences in *S. moellendorffii* are mostly intergenic (R1, R4, R5, and R6) or located in introns (R2, R3, and R8). The exceptions are R7 affecting exons 6 and 7 of the *cox1* gene, R9 extending into *nad1* exon 1, and R10 including exon 1 of the *atp1* gene. Four of the large repeated sequences (R3, R4, R6, and R7) feature additional internal recombination breakpoints and are hence in fact composed of two recombinationally active parts (designated as a/b in figs. 1 and 2).

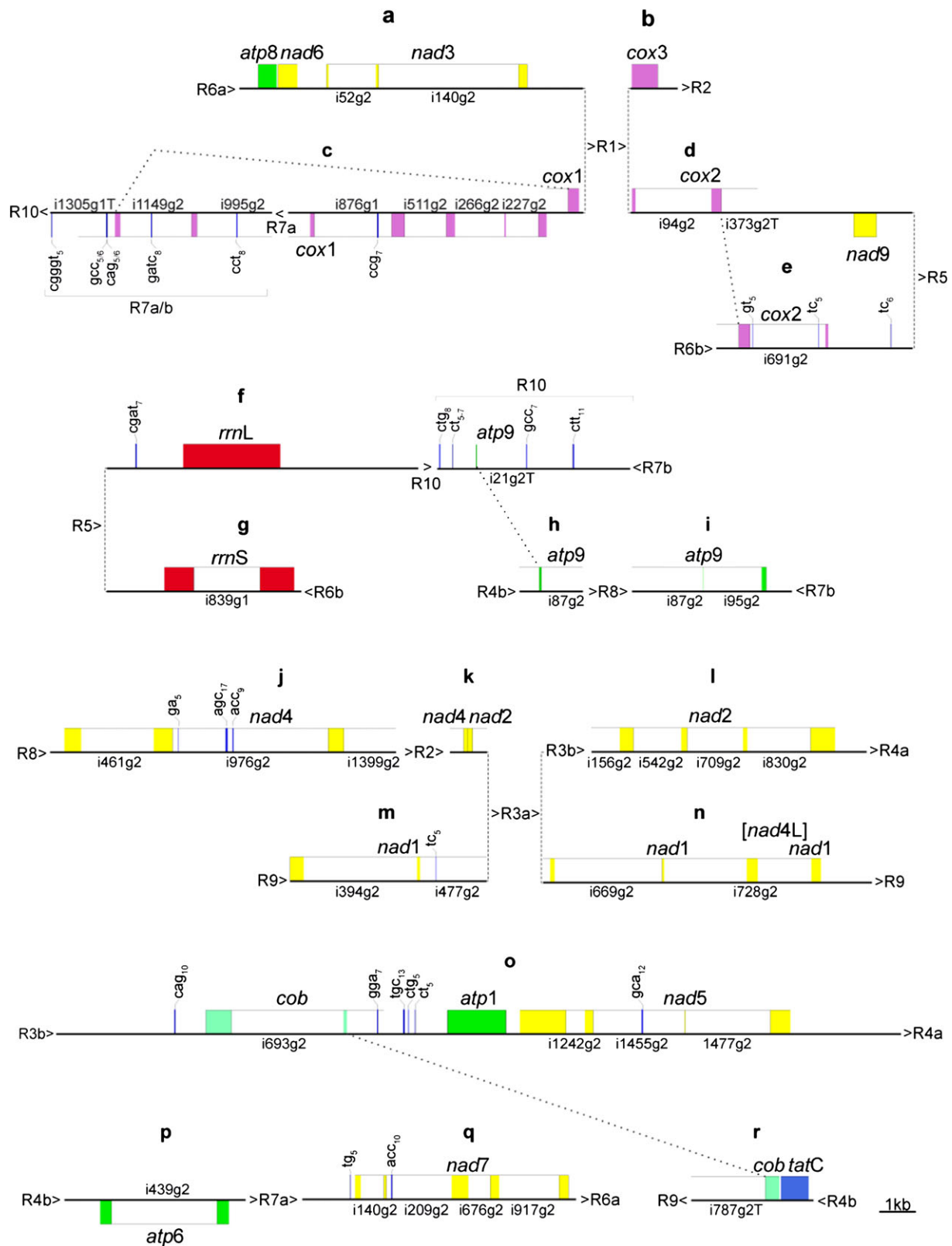
The long sequence repeats immediately remind of those previously found in flowering plant mitochondria, which are



**FIG. 2.**—The ten large repeated sequences R1–R10 in *Selaginella moellendorffii* mtDNA extend to up to 7.3 kb (R7). Repeats R3, R4, R6, and R7 are composite repeats (a/b) with internal sites of recombination. The exact sizes of R2, R8, and R10 are undetermined given that only one terminating recombination point is identified in these three cases. All repeats are characterized by numerous microsatellite sequences which often vary in repeat number as indicated.

involved in chondrome rearrangements and can result in stoichiometric shifting of subgenomic mtDNA topologies, occasionally beyond detection level for one (or more) of alternative DNA topologies (Arrieta-Montiel et al. 2001). In the cases of R1 and R9, no evidence for recombinational activity was found. For example, the pathways *nad3*-R1-*cox3* and *cox1*-R1-*cox2* were identified for R1 in the available sequences but not the reciprocal exchanges (fig. 2), also

leaving open the possibility that some repeated sequences may be recombinationally inactive. In three cases (R2, R8, and R10), recombination breakpoints were only identified for one end leaving the actual extension of these repeated sequences undetermined. Alternatively, these may be regarded as part of a composite repeat with neighboring repeats, which additionally include coding regions, that is: [R2-*nad4e2*-*nad4e1*-R3], [R8-*atp9e2*-R4], and [R10-R7].



**FIG. 3.**—A total of 18 single-copy sequence islands (a–r), flanked by sequence repeats R1–R10 (see fig. 2), mainly most of the coding regions of the *Selaginella moellendorffii* mtDNA. To display full gene complement, repeats R7 and R9 are additionally included, which contain parts of *cox1* and *atp9*, respectively. Gene displays were created using OGDRAW (Lohse et al. 2007). Color coding of exons indicates gene for complex I (yellow), III (light

Frequent recombination in *S. moellendorffii* mtDNA left only a single continuity of three genes undisturbed by recombination: *atp8-nad6-nad3* (figs. 1 and 3). In one additional case of three linked, nonrecombining coding sequences, *cob* exons 1 and 2 are located upstream of *atp1* and *nad5*, but the missing *cob* exon 3 is added by a novel *trans*-splicing event now discovered (fig. 3), which will be discussed below. Interestingly though, directions of transcription are retained not only for these two gene continuities but very frequently also across recombination events. In fact, we could easily arrange the orientations of repeated sequences according to the dominating direction of transcription of flanking gene sequences (fig. 1), and only five cases are found, where individual genes run against the dominating trend for transcriptional orientation (fig. 1): *nad9* at R5, the 18S rRNA at R6, *cob* exon 3 at R9, *tatC* at R4, and finally the first seven *cox1* exons running in opposite direction to the other genes flanking R7. Notably, *cox1* exon 8 is located upstream of exon 1 in inverted orientation but in the same transcriptional direction as the other three genes flanking R1 (*nad3*, *cox3*, and *cox2*). Exon 8 of *cox1* is added via a *trans*-splicing event during transcript maturation (see below).

Striking features of the repeated sequence units in the *S. moellendorffii* mtDNA are numerous short sequence repeat microsatellite motifs (fig. 2). We observed a total of 82 such microsatellites with at least 5 repeat units. Dominating in abundance are trinucleotide motif repeats (57), followed by fewer tetranucleotide (11), dinucleotide (10), and pentanucleotide (4) microsatellites. In comparison, only 16 such microsatellites occur in the single-copy genomic regions (fig. 3). Here, they are completely absent from coding sequences and mainly occur in intron sequences, most significant examples being AGC<sub>17</sub> and ACC<sub>9</sub> in *nad4i976* (fig. 3j), GCA<sub>12</sub> in *nad5i1455* (fig. 3o), and ACC<sub>10</sub> in *nad7i209* (fig. 3q).

The largest among all microsatellites is an AAAGG<sub>15</sub> microsatellite motif at the beginning of R9 (fig. 2). Occasional minor variability in microsatellite repeat numbers are the only sequence differences between repeat copies in different sequence environments (fig. 2). For statistical significance, we initially only considered microsatellites with repeat numbers of minimally five. However, upon closer inspection of repeats, we found that microsatellites with lower and variable repeat numbers were present close to recombinational breakpoints, for example, AGCC<sub>1/2</sub> and CCG<sub>3/4</sub> close to the termini of R3a (fig. 2). Most surprising, however, were insights from comparison of homologous sequences in the genomic scaffolds with those derived from our fosmid cloning approach: Numerous sites of additional low copy number repeat variability (mostly 1/2, occasionally 2/3, 0/1, or 0/2) were seen

also outside of repeat regions, with trinucleotide motifs again strongly dominating in abundance. One such example of increased repeat variability is shown for the upstream terminus of R5 (fig. 4). These differences of yet higher microsatellite variability most likely reflect the different *S. moellendorffii* isolates used for DNA preparations in our independent fosmid cloning approach versus the previous genomic sequencing approach. We conclude that the peculiar variability of very low copy number motifs (mostly trinucleotide) reflect the tendency of *S. moellendorffii* mtDNA toward evolving novel microsatellite sites, which subsequently pave the way to evolve recombinationally active sequences.

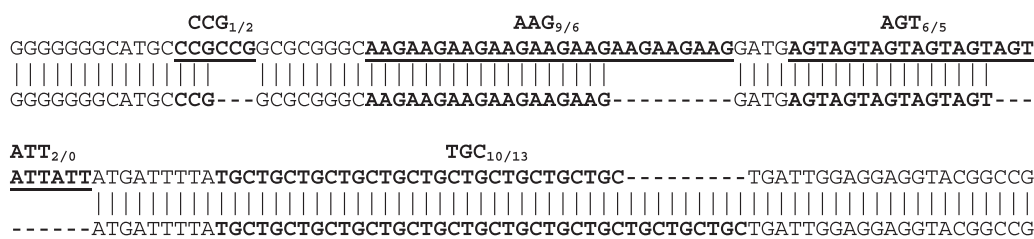
### The Gene and Intron Complement

The *S. moellendorffii* mtDNA encodes a full, plant-typical core set of genes for subunits of the respiratory chain complexes I (*nad1–7* and 9), III (*cob*), IV (*cox1–3*), and V (*atp1*, 6, 8, and 9) and the *tatC* gene encoding a twin-arginine translocase (table 1). Absent from the chondrome, however, are an *atp4* gene and genes for subunits of complex II (*sdh*), cytochrome *c* biogenesis and maturation (*ccm*), and a 5S rRNA. Highly striking and a novelty among the complete plant chondrome sequences determined so far is the absence of any genes encoding tRNAs. Whereas *ccm* genes are absent from the *I. engelmannii* chondrome as well, *atp4*, *sdh3*, four ribosomal proteins, a 5S rRNA and 13 tRNAs are encoded there (table 1).

A contrasting picture is observed for the complement of 35 group II (g2) and two group I (g1) introns: Seven group II introns having homologues in seed plants are present in *Selaginella* but absent in *Isoetes* (*nad1i477g2*, *nad1i669g2*, *nad1i728g2*, *nad4i976g2*, *nad7i140g2*, *cox2i373g2*, and *cox2i691g2*). Reciprocally, introns *nad7i1113g2* and *cox1i395g1* conserved in liverworts are present in *Isoetes* but not in *Selaginella*. The intron-rich *cox1* gene in particular shows further variability with two more introns in each taxon not present in the respective other (table 1). The recently discovered *trans*-splicing group I intron *cox1i1305g1* in *Isoetes*, however, is conserved in *Selaginella* with the same physical discontinuity.

Most interestingly, all introns known as *trans*-splicing in seed plants (*nad1i394g2*, *nad1i669g2*, *nad1i728g2*, *nad2i542g2*, *nad5i1455g2*, and *nad5i1477g2*) have *cis*-splicing counterparts in *S. moellendorffii* without exception, whereas the spikemoss features *trans*-splicing group II introns in three other locations (table 1). The case of a disrupted, *trans*-splicing *cox2i373g2T* intron has very recently also been reported for *Allium* (Kim and Yoon 2010), whereas *cob1i787g2T* and *atp9i21g2T* represent two novel cases of *trans*-splicing introns in nature.

← green), IV (purple), V (dark green), ribosomal RNAs (red), and *tatC* (blue). Dotted lines connect exons linked via *trans*-splicing introns *cox1i1305g1T* (c), *cox2i373g2T* (d and e), *atp9i21g2T* (R10-h), and *cob1i787g2T* (o-r). The *atp8-nad6-nad3* arrangement (a) is the only case of a complete three-gene-continuity unaffected by recombination in *S. moellendorffii* mtDNA.



**Fig. 4.**—An example of microsatellite variability at the beginning of repeat R9, proximal to the *nad9* gene. The upper sequence comes from *Selaginella moellendorffii* scaffold M162 and continues into the 18S rRNA, the lower sequence comes from fosmid 3 identified in an independent cloning approach and continues toward the 26S rRNA.

In comparison to other plant mtDNAs, the *S. moellendorffii* mtDNA features an extraordinarily high average GC content of 68.1% that exceeds by far those of other land plant mtDNAs hitherto sequenced (in all of which GC content is below 50%). It is interesting to see that the chloroplast DNA of *S. moellendorffii* similarly has an elevated GC content of 51% (Smith 2009).

Intergenic regions are comparatively large in *S. moellendorffii* mtDNA with two exceptions: *cob-tatC* with only 50 bp and *nad2-nad4*, which actually overlap by 17 bp. In contrast to the *I. engelmannii* mtDNA (Grewe et al. 2009), no insertions of chloroplast or nuclear DNA are discernible. Finally, one further peculiarity previously not observed in other plant mtDNAs (not considering intron-borne maturases or endonucleases) is the insertion of one gene in the intron of another: The *S. moellendorffii nad4L* gene is located within group II intron *nad1i728g2* of the *nad1* gene in the same direction of transcription (fig. 3n).

### RNA Editing in mRNAs and rRNAs

In parallel to determining the complete *S. moellendorffii* mtDNA, we performed an exhaustive complementary cDNA analysis covering all coding regions (18 protein coding and 2 rRNAs). We found that all 37 intron sequences (table 1) are correctly spliced as can be predicted from their respective group I or group II secondary structures. Moreover, the cDNA analyses showed that RNA editing in *S. moellendorffii* is in fact yet more abundant than in the sister lycophyte *I. engelmannii*. In total, we observed 2,139 RNA editing sites in the 18 mRNAs (table 1). Of these, 424 (i.e., ca. 20%) are silent, whereas the others introduce 1,488 codon changes. A complete list of RNA editing positions discovered in our study using the recently proposed nomenclature (Lenz et al. 2009; Rüdinger et al. 2009) is given in [supplementary table 1](#) (Supplementary Material online). In two cases, more than 200 RNA editing events affect single mRNAs: 249 sites in *nad2* and 237 sites in *nad4*. In fact, these two genes and *cox1* in *S. moellendorffii* now even exceed the *I. engelmannii nad5* gene previously setting the record for editing site abundance among land plant mitochondrial genes (table 2).

The highest density of editing sites, however, was found in the small *nad3* reading frame (fig. 5). Here, 77 edits change

more than every fifth nucleotide in the only 369 nucleotide long *nad3* reading frame and introduce 45 codon changes. This includes the alteration of seven codons in a row by 14 edits changing the genomically encoded PTSPPTP protein sequence into SILFIIF. This sequence of codons is located at the end of a stretch of 45 nucleotides, 22 of which are edited (fig. 5). The alteration of even eight codons in a row (PPTPPTPT to LFILLMSI) in the *nad4L* gene is another example of extreme RNA editing density ([supplementary table 1, Supplementary Material online](#)). Only 1 of the 18 protein encoding genes (*nad7*) has start and stop codons on DNA level, whereas either the one or the other or even both (in *atp1*, *cox3*, and *nad4*) have to be introduced by RNA editing ([supplementary table 1, Supplementary Material online](#)).

Most strikingly, all events of RNA editing in *S. moellendorffii* are exclusively C-to-U exchanges, a very surprising finding given that RNA editing frequently also operates in reverse in hornworts, ferns, and lycophytes including *Selaginella*'s sister genus *Isoetes* (Grewe et al. 2011). Likewise, striking is a particular bias in the codon changes observed. The 392 proline (CCN) to leucine (CUN) codon changes exceed all others by far, followed by 140 proline (CCY) to phenylalanine (UUY) changes, which require two simultaneous editings per codon. Surprisingly, these are followed in abundance by 137 alanine (GCN) to valine (GUN) changes. This latter type of RNA editing is generally rare in other plants due to the guanidine nucleotide preceding the editing position but in the case of *Selaginella* they even exceed the proline (CCN) to serine (UCN) edits normally observed in very high proportion in other plant mtDNAs.

Finally, we could identify several editing sites in the two rRNAs. Three C-to-U editing events were found in the 26S rRNA, 10 in the 18S rRNA. In the latter case, all edits are located in the first 18S rRNA exon where nine of them cluster in an upstream sequence stretch of only 200 nucleotides (fig. 6A). With one single exception (editing site *rrnSeU295*), all editing events reintroduce uridines conserved in rRNAs of other taxa as exemplarily shown in the alignment with the *Pinus* homologue (fig. 6A). Notably, the 10th editing identified in the *Selaginella* 18S rRNA affects the last nucleotide of the 18S rRNA 5'-exon and may directly influence splicing of group I intron *rrn18i839g1*, as it likely influences the base pairing



**Table 1**

Gene and Intron Complement in the mtDNA of *Selaginella moellendorffii* (Smoe, This Work) in Comparison to the Ones of the Liverwort *Marchantia polymorpha* (Mpol, Oda et al. 1992), the Moss *Physcomitrella patens* (Ppat, Terasawa et al. 2006), the Hornwort *Megaceros aenigmaticus* (Maen, Li et al. 2009), the Lycophyte *Isoetes engelmannii* (leng, Grewe et al. 2009), and the Gymnosperm *Cycas taitungensis* (Ctai, Chaw et al. 2008)

Genes/Introns	Mpol	Ppat	Maen	leng	Smoe	Ctai	Genes/Introns	Mpol	Ppat	Maen	leng	Smoe	Ctai
<i>atp1</i>	+	+	+	+	+	+	nad5i753g1	+	+				
atp1i805g2			+				nad5i1242g2				+	+	
atp1i989g2	++						nad5i1455g2		++	++	+	+	<i>trans</i>
atp1i1019g2			++				nad5i1477g2			++	+	+	<i>trans</i>
atp1i1050g2	++		++				nad5i1872g2						+
atp1i1128g2		+					nad6	+	+	+	+	+	+
<i>atp4</i>	+	+	+	+		+	nad6i444g2			+			
<i>atp6</i>	+	+	+	+	+	+	nad7	Ψ	+		+	+	+
atp6i80g2		+	+				nad7i140g2		+			+	+
atp6i439g2			+	+	+		nad7i209g2		+		+	+	+
<i>atp8</i>	+	+	Ψ	+	+	+	nad7i336g2	+					
<i>atp9</i>	+	+	+	+	+	+	nad7i676g2				+	+	+
atp9i21g2		+		+	<i>trans</i>		nad7i917g2				+	+	+
atp9i87g2	++	+		++	+		nad7i1113g2	+			+		
atp9i95g2		+	++		+		nad9	+	+	+	+	+	+
<i>ccmB</i>	+	+				+	nad9i246g2			+			
<i>ccmC</i>	+	+				+	nad9i283g2		+				
<i>ccmF</i>	+	+	Ψ			+	nad9i502g2			+			
ccmFCi829g2		+	+			+	rpl2	+	+				+
<i>cob</i>	+	+	+	+	+	+	rpl2i28g2	+					
cobi372g2	+						rpl2i917g2						+
cobi420g1		+					rpl5	+	+	Ψ	+		+
cobi693g2				+	+		rpl6	+	+	Ψ			
cobi783g2	+						rpl10	+	+	+			
cobi787g2			++	+	<i>trans</i>		rpl16	+	+				+
cobi824g2	++						rps1	+	+	Ψ	Ψ		+
cobi838g2			++				rps2	+	+		+		+
<i>cox1</i>	+	+	+	+	+	+	rps3	+	+		+		+
cox1i44g2	++		++				rps3i74g2				+		+
cox1i150g2			++				rps3i257g2						++
cox1i178g2	++						rps4	+	+	Ψ	+		+
cox1i227g2				+	+		rps7	+	+	Ψ			+
cox1i266g2				+	+		rps8	+		Ψ			
cox1i323g2				+			rps10	+					+
cox1i375g1	+						rps10i235g2						+
cox1i395g1	++			+			rps11	+	+	Ψ			+
cox1i511g2	++	+			+		rps12	+	+	Ψ			+
cox1i624g1	+	+					rps13	+	+	+			+
cox1i730g1	++						rps14	+	+	+			+
cox1i732g2		++					rps14i114g2	+					+
cox1i876g1					+		rps19	+	+				+
cox1i995g2				+	+		rrn5	+	+	+	+		+
cox1i1064g2		+					rrnL	+	+	+	+	+	+
cox1i1116g1	++						rrnLi827g2	+					
cox1i1149g2					+		rrnS	+	+	+	+	+	+
cox1i1298g2			++				rrnSi839g1				+	+	
cox1i1305g1	+			<i>trans</i>	<i>trans</i>		rrnSi1065g2	++					
<i>cox2</i>	+	+	+	+	+	+	sdh	+	+	Ψ	+		+
cox2i94g2				+	+		sdh3i100g2		+	+			
cox2i97g2	+						sdh4	+	+	+			
cox2i104g2		+					tatC	+	+	+	+	+	+
cox2i250g2	++						trnA(ugc)	+	+	+			
cox2i281g2			++				trnC(gca)	+	+	+	+		+
cox2i373g2		+	++		<i>trans</i>	+	trnD(guc)	+	+	+			+
cox2i691g2		+				+	trnE(uuc)	+	+	+	+		+

**Table 1**  
**Continued**

Genes/Introns	Mpol	Ppat	Maen	leng	Smoe	Ctai	Genes/Introns	Mpol	Ppat	Maen	leng	Smoe	Ctai
<i>cox3</i>	+	+	+	+	+	+	<i>trnF(gaa)</i>	+	+	+	+		+
cox3i171g2	+						<i>trnG(gcc)</i>	+	+	+	+		+
cox3i506g2		+					<i>trnG(ucc)</i>	+	+				
cox3i625g2	+						<i>trnH(gug)</i>	+	+	+			
<i>nad1</i>	+	+	+	+	+	+	<i>trnI(cau)</i>	+	+	+	+		+
nad1i287g2		+	+				<i>trnK(uuu)</i>	+	+	+	+		+
nad1i348g2			+				<i>trnL(caa)</i>	+	+	+			+
nad1i394g2				+	+	trans	<i>trnL(uag)</i>	+	+		Ψ		+
nad1i477g2					+	+	<i>trnL(uaa)</i>	+	+	+			
nad1i669g2					+	trans	<i>trnM(cau)</i>	+	+	+	+		+
nad1i728g2		+	++		+	++	<i>trnMf(cau)</i>	+	+	+	+		+
<i>nad2</i>	+	+	+	+	+	+	<i>trnN(guu)</i>	+					+
nad2i156g2		+		+	+	+	<i>trnP(ugg)</i>	+	+	+	+		+
nad2i542g2				+	+	trans	<i>trnQ(uug)</i>	+	+	+	+		+
nad2i709g2	+		+	+	+	+	<i>trnR(acg)</i>	+	+				
nad2i830g2				+	+		<i>trnR(ucg)</i>	+					
nad2i1282g2			+			+	<i>trnR(ucu)</i>	+	+				+
<i>nad3</i>	+	+	+	+	+	+	<i>trnS(gcu)</i>	+					+
nad3i52g2			++	+	+		<i>trnSi43g2</i>	+					
nad3i140g2	+		++	+	+		<i>trnS(uga)</i>	+	+		+		+
<i>nad4</i>	+	+	+	+	+	+	<i>trnS(gga) cp</i>						+
nad4i461g2		+	++	+	+	+	<i>trnT(ggu)</i>	+	+	+			
nad4i548g2	+						<i>trnV(uac)</i>	+	+				
nad4i976g2			++		+	+	<i>trnW(cca)</i>	+	+	+	+		+
nad4i1399g2				+	+	+	<i>trnY(gua)</i>	+	+	+	+		+
<i>nad4L</i>	+	+	+	+	+	+	∑genes	70	66	41	40	20	62
nad4Li100g2	+						∑introns	32	27	30	30	37	27
nad4Li283g2	+	+											
<i>nad5</i>	+	+	+	+	+	+							
nad5i230g2		+	+			+							

NOTE.—Omitted from the list were chloroplast-derived tRNA genes *trnH*, *trnM*, and *trnV* exclusively present in the *Cycas* mtDNA. Intron nomenclature relies on labeling according to the preceding nucleotide in the mature RNA (Dombrowska and Qiu 2004) amended by the labels g1/g2 for group I and group II introns (Knoop 2004), respectively. Double plus signs indicate presence of intron-borne ORFs; Ψ indicates recognizable pseudogene remnants and “trans” indicates trans-splicing intron configurations.

needed for the conserved paired region P1 (fig. 6B). A further potential editing event may be located 6-nt downstream in the intron region contributing to the P1 pairing and could similarly be a prerequisite for splicing of *rrn18i839g1*.

## Discussion

The number of features distinguishing the mtDNA of *S. moellendorffii* from the one of the sister lycophyte *I. engelmannii*, which has previously been determined (Grewe et al. 2009), is astounding (table 3). A yet higher number and density of RNA editing sites than previously observed in *Isoetes* is impressive as such, but yet more significant is the complete absence of U-to-C editing in the presence of 2,139 sites of C-to-U editing in *Selaginella* contrasting the corresponding numbers of 222 and 1,560 in *Isoetes*. The concentration of editing sites (fig. 5a) actually reminds of the extreme “pan-editing” of the uridine insertion type occasionally observed in trypanosomes (Feagin et al. 1988; for a recent review, see Knoop 2010). The one obvious explanation for the high degree of C-to-U editing observed is the unprecedented high GC

content of 68.1% in a plant mtDNA now found in *S. moellendorffii* (table 3), which even exceeds the recently reported case of 57% GC in the alga *Polytomella capuana* by far (Smith and Lee 2008). This observation in fact revitalizes an early suggestion that RNA editing frequencies may shift in response to GC content drift in organelle genomes (Malek et al. 1996).

The extension of introns to sizes of more than 6 kb in *S. moellendorffii* likewise stands in complete contrast to the tiny introns in the mtDNA of *I. engelmannii* (table 3). Orthologous introns occupying identical sites may be up to 8-fold (*rrnSi839g1*) or even 9-fold (*nad2i156g2*) larger in *Selaginella* than in *Isoetes*. Consequently, it is all the more difficult to explain why several typical plant mitochondrial genes and promiscuous DNA sequences originating from the chloroplast or nucleus are present in the compact and gene-dense chondrome of *Isoetes* but are absent in *Selaginella* (table 3).

Along the same lines, the complete absence of any tRNA-encoding genes in *S. moellendorffii* mtDNA is a first example for plants. It could be envisaged that tRNA genes may be overlooked when subject to massive RNA editing. Testing the tRNAscan-SE program (Lowe and Eddy 1997) on the

**Table 2**

RNA Editing Sites Discovered in the *Selaginella moellendorffii* (Smoe) Transcriptome Are Listed for Comparison with Data from Similarly Comprehensive Transcriptome Studies in the Moss *Physcomitrella patens* (Ppat, Rüdinger et al. 2009), the Lycophyte *Isoetes engelmannii* (Ieng, Grewe et al. 2011), the Flowering Plants *Beta vulgaris* (Bvul, Mower and Palmer 2006), *Silene noctiflora* and *Silene latifolia* (Snoc and Slat, Sloan, MacQueen, et al. 2010), *Arabidopsis thaliana* (Atha, Giegé and Brennicke 1999), *Brassica napus* (Bnap, Handa 2003), and *Oryza sativa* (Osat, Notsu et al. 2002)

Genes	Ppat	Ieng	Smoe	Bvul	Slat	Snoc	Bnap	Atha	Osat
Numbers of editing sites									
<i>nad1</i>	0	57	137	20	19	11	23	24	23
<i>nad2</i>	0	100	<b>249</b>	24	21	18	25	32	30
<i>nad3</i>	1	57	77	12	8	5	10	12	15
<i>nad4</i>	1	156	<b>237</b>	19	16	11	35	32	20
<i>nad4L</i>	0	27	47	10	9	6	9	9	10
<i>nad5</i>	2	<b>166</b>	143	17	18	15	29	27	11
<i>nad6</i>	0	64	89	11	10	6	11	10	18
<i>nad7</i>	0	115	103	20	19	9	28	27	32
<i>nad9</i>	0	50	72	5	5	1	8	7	12
<i>sdh3</i>	0	23							
<i>sdh4</i>	0			4					
<i>cob</i>	0	121	122	13	9	6	8	7	19
<i>cox1</i>	1	110	<b>181</b>	0	0	0	1	0	4
<i>cox2</i>	1	14	97	9	3	2	13	15	19
<i>cox3</i>	1	101	133	4	1	1	7	8	1
<i>atp1</i>	0	131	160	3	3	0	5	5	5
<i>atp4</i>	0	6		12	11	6	8	8	9
<i>atp6</i>	0	95	80	12	11	7	1	1	17
<i>atp8</i>	0	35	35	2	2	2	3	0	4
<i>atp9</i>	1	34	44	5	4	1	4	4	8
<i>ccmB</i>	0			30	27	19	39	39	35
<i>ccmC</i>	0			28	23	15	25	28	36
<i>ccmFN(1)</i>	0			23	22	16	15	22	31
<i>ccmFN(2)</i>	0						10	12	
<i>ccmFC</i>	2			13	12	8	13	16	27
<i>rpl2</i>	0						2	1	1
<i>rpl5</i>	0	30		5	6	4	9	10	1
<i>rpl6</i>	0								
<i>rpl10</i>	0				n.i.	n.i.			
<i>rpl16</i>	0						6	5	12
<i>rps1</i>	0								3
<i>rps2</i>	0	28							10
<i>rps3</i>	0	62		6	4	3	8	13	10
<i>rps4</i>	0	55		11			19	15	15
<i>rps7</i>	0			3			1	0	2
<i>rps8</i>	0								
<i>rps10</i>	0								
<i>rps11</i>	0								4
<i>rps12</i>	0			6			7	8	0
<i>rps13</i>	0			2	1	0			8
<i>rps14</i>	1						0		0
<i>rps19</i>	0								
<i>tatC</i>	0	68	133	19	15	11	27	24	33
<i>matR</i>	0			9	8	6	8	9	n.i.
<b>Total</b>	<b>11</b>	<b>1,705</b>	<b>2,139</b>	<b>357</b>	<b>287</b>	<b>189</b>	<b>417</b>	<b>430</b>	<b>491</b>

NOTE.—Dark gray shading indicates that a given gene is absent from the respective mtDNA, light shading indicates gene overlaps where editing has been counted once only (*rpl16*, *rps3* in *Arabidopsis*) or pseudogenes (*sdh4* in *Beta*, *rps11* and *rps14* in *Oryza*), *matR* is the *nad1i728g2* intron-borne maturase present in seed plants

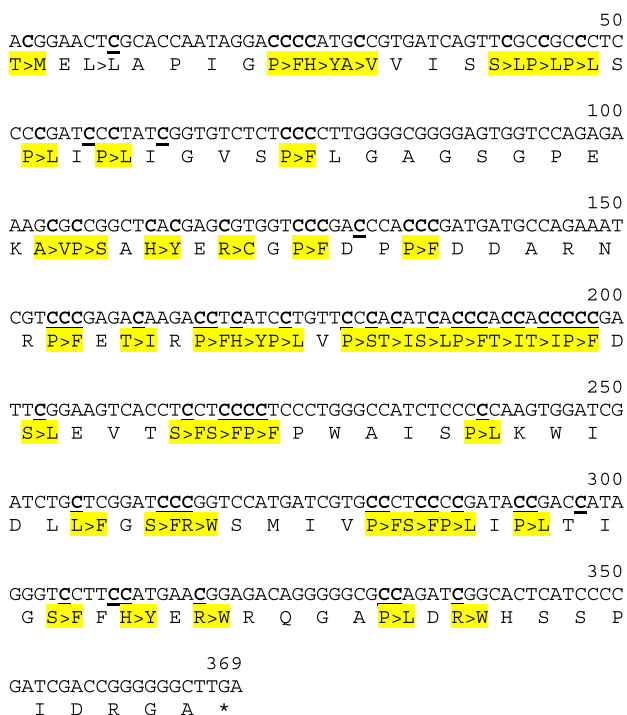
*I. engelmannii* mtDNA, however, gave evidence that even the heavily edited tRNAs in this taxon (with up to 18 sites in a single tRNA) are faithfully identified, at least as tRNA-pseudo-gene candidates.

Such massive loss of mitochondrial tRNA genes has previously been reported for animal mtDNA, for example, in the phylum Cnidaria (Haen et al. 2010) and very recently also for the angiosperm *Silene latifolia* where only six native mitochondrial tRNAs are retained in the chondrome (Sloan, Al- verson, et al. 2010). In contrast to endosymbiotic gene transfer (EGT) of protein genes which acquire mitochondrial targeting signals in their new nuclear location, tRNA genes lost from mtDNA are known to be functionally replaced by cytosolic tRNA counterparts (Glover et al. 2001; Duchene et al. 2009, 2011). Hence, we could expectedly not identify any bona fide mitochondrial tRNA genes in the nuclear *S. moellendorffii* assemblies.

Genes for 5S rRNAs are a generally conserved feature in land plant mitochondrial genomes, and the absence of a 5S rRNA gene in the *S. moellendorffii* mtDNA is surprising. Outside of the plant lineage, however, 5S rRNA genes are frequently absent from mitochondrial genomes (Lang et al. 1996) or so highly diverged that their identification on DNA level alone was impossible (Bullerwell et al. 2003). An ultimate answer for *S. moellendorffii* mitochondria relies on purification and analyses of mitochondrial ribosomes or rRNAs, respectively.

The absence of any ribosomal protein (*rpl*, *rps*) genes and of *sdh* genes encoding subunits of complex II in *S. moellendorffii* mtDNA seems to evolutionarily anticipate the frequent independent EGT of these genes into the nucleus, which has been demonstrated among angiosperms (Adams et al. 2002). Although EGT was not observed for *ccm* genes encoding proteins of cytochrome-c-biogenesis in that angiosperm survey, their absence from *Selaginella* mtDNA is also unsurprising given that they also lack from *I. engelmannii* and hornwort mtDNAs (Grewe et al. 2009; Li et al. 2009; Xue et al. 2010). However, these observations postulate at least two independent losses of *ccm* genes from mitochondria in the lycophyte and hornwort stem lineages. A novelty with respect to gene complement is the loss of a mitochondrially encoded *atp4* gene from the mtDNA of *S. moellendorffii*, which has previously not been reported for the land plant lineage. Interestingly, the *atp4* gene although retained in the *I. engelmannii* mtDNA is misedited on transcript level to introduce a stop codon (Grewe et al. 2011), possibly indicating an emerging pseudogene that may be functionally replaced by a functional nuclear copy. Moreover, an absence of *atp4* but simultaneous retention of *atp1*, *atp6*, and *atp9* (as well as *sdh3* and several *rpl* and *rps*

(n.i. = not investigated). The four top-scoring genes with more than 160 editing sites each in the respective mRNAs are indicated in bold.



**Fig. 5.**—RNA editing sites in the *nad3* gene are highlighted with underlining where cytidines are converted into uridines in the mRNA. Resulting codon changes are highlighted by shading.

genes) in mtDNA has previously been described for the cryptophyte *Hemiselmis andersenii* (Kim et al. 2008). Given the high degree of recombinational activity in the mtDNAs of *Selaginella* and *Isoetes*, it is not surprising that nearly no gene arrangements are conserved between the two taxa. In fact, *atp8-nad6* and *nad4-nad2* are the only two conserved gene continuities.

The actual physical structures of plant mitochondrial genomes have been debated since long. Populations of smaller, overlapping, and/or linear molecules may actually be stoichiometrically dominant over circular DNAs of genome size, if existing at all, even in the absence of recombination repeats in the circular-mapping genomes of the bryophytes (Bendich 1993; Manchekar et al. 2006). The situation is significantly complicated with the larger and recombinationally active flowering plant mtDNAs, in which subgenomic molecules may shift in stoichiometry (Abdelnoor et al. 2003; Arrieta-Montiel and Mackenzie 2011) and which may differ significantly in appearance even between closely related isolates of the same species (e.g., Ullrich et al. 1997; Allen et al. 2007). Obviously, this is also true for *Selaginella* mtDNA where elucidating the true physical structure of mtDNA molecule populations will necessarily rely on purification of large amounts of mtDNA to be analyzed by methods such as pulsed-field gel electrophoresis.

Being the extant representatives of the most ancient surviving lineage of vascular plants, lycophytes are a crucial clade to understand plant evolution. The complete sequence of the *S. moellendorffii* nuclear genome will be a major milestone in this regard (Banks 2009). As evolutionarily “early” vascular plants, lycophytes may be considered a transitionally evolutionary state in the development from a bryophyte-type lifestyle to the more advanced tracheophytes, the euphyllophytes. On the other hand, it is noteworthy that common characters may also evolve surprisingly convergently and independently, and the recent studies on evolution of syringyl lignin biosynthesis in tracheophytes are an important case in point (Weng et al. 2010).

So far, it appears that *Isoetes* has had a tendency to retain more ancient mtDNA features from the bryophyte-like



**Fig. 6.**—Ten events of RNA editing were identified in the mitochondrial 18S rRNA of *Selaginella moellendorffii*. **A:** Nine sites of RNA editing converting cytidines into uridines are located in the region between nucleotides 294 and 488 of the upstream 18S rRNA exon. All but the second of these events (editing site *rrnSeU295*) introduce uridines conserved in the corresponding 18S rRNA sequence of the gymnosperm *Pinus* shown for comparison. **B:** A 10th editing event affects the nucleotide upstream of the *rrnSi869g1* splice site (arrow) and may be crucial to the splicing event by influencing base pair formation in P1.

**Table 3**

Comparative Overview About mtDNA Features Deviating in the Mitochondrial Genomes of *Isoetes engelmannii* (Grewe et al. 2009 and 2011) and *Selaginella moellendorffii* (This Work)

Chondrome Features	<i>Isoetes engelmannii</i>	<i>Selaginella moellendorffii</i>
RNA editing		
mRNA, C-to-U sites	1,560	2,139
mRNA, U-to-C sites	222	0
In rRNAs	0	13
Genes and introns		
Proteins	24	18
tRNAs	13	0
rRNAs	3	2
Group I introns ( <i>trans</i> )	2 (1)	2 (1)
Group II introns ( <i>trans</i> )	27 (0)	34 (3)
Intron size range	327–2,072	516–6,932
Average intron size <sup>a</sup>	526	2,455
Sequence characteristics		
GC content	48.7%	68.1%
cp inserts	1	0
nuc inserts	2	0
Microsatellites <sup>b</sup>	6	90

<sup>a</sup>Considering *cis*-splicing introns only.

<sup>b</sup>With motif repeats >4.

ancestors (reverse U-to-C editing shared with hornworts, introns shared with liverworts), whereas *Selaginella* features more common characteristics with seed plants (introns, extended intergenic repeat regions active in recombination). However, given the unequivocal monophyly of lycophytes as a whole, it must be assumed that all these characteristics were jointly present in their last common ancestor with euphyllophytes. It will be highly interesting in this regard to have the mitochondrial genome sequence of a member of the remaining third order of lycophytes, the Lycopodiales (genus *Huperzia*, Qiu YL, personal communication), available soon.

The largest benefit to understand the peculiarities of the *S. moellendorffii* mtDNA now outlined here will come from the completed nuclear genome analysis (Banks 2009). One highly interesting aspect will be the investigation of homologues for nuclear factors recently determined to take part in mitochondrial genome rearrangements in flowering plants (Abdelnoor et al. 2003; Shedge et al. 2007). Possibly even more exciting will be the investigation of nuclear-encoded factors for recognizing mitochondrial RNA editing positions, among which the large plant-specific gene families of RNA-binding pentatricopeptide repeat (PPR) proteins with PPR repeats of variable length (PLS-type) hold a very crucial position (Lurin et al. 2004; Salone et al. 2007; Zehrmann et al. 2009; Knoop 2010; Knoop and Rüdinger 2010; Tasaki and Sugita 2010). The extreme degree of editing in its mitochondrial transcriptome described here obviously correlates well with a particularly large PPR gene family in *S. moellendorffii* ([http://wiki.genomics.purdue.edu/index.php/PPR\\_gene\\_family](http://wiki.genomics.purdue.edu/index.php/PPR_gene_family)), and this could make the lycophyte an interesting object of study to understand the RNA sequence-recognition code of PLS-type PPR proteins.

## Supplementary Material

Supplementary table 1 is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

The *Selaginella moellendorffii* complete genome sequencing initiative coordinated by Dr Jo Ann Banks (Purdue University West Lafayette, IN) and communication on this project with Dr Mitsuyasu Hasebe (National Institute for Basic Biology, Okazaki, Japan) is very gratefully acknowledged. We also wish to thank Monika Polsakiewicz for excellent technical assistance and to acknowledge earlier work on lycophyte mtDNAs based on cosmid cloning attempts by Karolin Kalmbach, Yesim Kümetepe, Julia Neuwirt, and Patrick Johner in Ulm, Germany. We are very grateful for grant Kn411/6-1 by the Deutsche Forschungsgemeinschaft (to D.F.G.) supporting the work on early vascular plant mtDNAs in the authors' laboratory.

## Literature Cited

- Abdelnoor RV, et al. 2003. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc Natl Acad Sci U S A*. 100(10):5968–5973.
- Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc Natl Acad Sci U S A*. 99(15):9905–9912.
- Allen JO, et al. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177(2):1173–1192.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.
- Alverson AJ, et al. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus*

- and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 27(6): 1436–1448.
- Arrieta-Montiel M, et al. 2001. Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 158(2):851–864.
- Arrieta-Montiel MP, Mackenzie SA. 2011. Plant mitochondrial genomes and recombination. In: Kempken F, editor. *Plant mitochondria*. New York: Springer. p 65–82.
- Banks JA. 2009. Selaginella and 400 million years of separation. *Annu Rev Plant Biol.* 60:223–238.
- Bendich AJ. 1993. Reaching for the ring: the study of mitochondrial genome structure. *Curr Genet.* 24:279–290.
- Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424(6945):197–201.
- Bullerwell CE, Schnare MN, Gray MW. 2003. Discovery and characterization of *Acanthamoeba castellanii* mitochondrial 5S rRNA. *RNA* 9(3):287–292.
- Chaw SM, et al. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol Biol Evol.* 25(3):603–615.
- Dombrowska E, Qiu YL. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Mol Phylogenet Evol.* 32(1):246–263.
- Duchene AM, El Farouk-Ameqrane S, Sieber F, Maréchal-Drouard L. 2011. Import of RNAs into plant mitochondria. In: Kempken F, editor. *Plant mitochondria*. New York: Springer. p 241–260.
- Duchene AM, Pujol C, Marechal-Drouard L. 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Curr Genet.* 55(1):1–18.
- Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53(3):413–422.
- Giegé P, Brennicke A. 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A.* 96(26):15324–15329.
- Glover KE, Spencer DF, Gray MW. 2001. Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria. *J Biol Chem.* 276(1):639–648.
- Grewe F, et al. Forthcoming. 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(7):2890–2902.
- Grewe F, Viehoveer P, Weisshaar B, Knoop V. 2009. A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 37(15):5093–5104.
- Groth-Malonek M, Wahrmond U, Polsakiewicz M, 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(4):1068–1074.
- Haen KM, Pett W, Lavrov DV. 2010. Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA synthetases and mtDNA-encoded tRNAs in Cnidaria. *Mol Biol Evol.* 27(10):2216–2219.
- Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* 31(20): 5907–5916.
- Kim E, et al. 2008. Complete sequence and analysis of the mitochondrial genome of *Hemiselmis andersenii* CCMP644 (Cryptophyceae). *BMC Genomics.* 9:215.
- Kim S, Yoon MK. 2010. Comparison of mitochondrial and chloroplast genome segments from three onion (*Allium cepa* L.) cytoplasm types and identification of a *trans*-splicing intron of *cox2*. *Curr Genet.* 56(2):177–188.
- Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet.* 46(3):123–139.
- Knoop V. 2010. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci.* 68:567–586.
- Knoop V, Rüdinger M. 2010. DYW-type PPR proteins in a heterolobosean protist: plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett.* 584(20):4287–4291.
- Knoop V, Volkmar U, Hecht J, Grewe F. 2010. Mitochondrial genome evolution in the plant lineage. In: Kempken F, editor. *Plant mitochondria*. New York: Springer. p 3–29.
- Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 9(4):299–306.
- Lang BF, Goff LJ, Gray MW. 1996. A 5 S rRNA gene is present in the mitochondrial genome of the protist *Reclinomonas americana* but is absent from red algal mitochondrial DNA. *J Mol Biol.* 261(5): 407–413.
- Lenz H, et al. 2009. Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Curr Genet.* 56(2):189–201.
- Li L, Wang B, Liu Y, Qiu YL. 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(6):665–678.
- Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet.* 52(5–6):267–274.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964.
- Lurin C, et al. 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell.* 16(8):2089–2103.
- Malek O, Knoop V. 1998. *Trans*-splicing group II introns in plant mitochondria: the complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. *RNA* 4(12):1599–1609.
- Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* 15(6):1403–1411.
- Manchekar M, et al. 2006. DNA recombination activity in soybean mitochondria. *J Mol Biol.* 356(2):288–299.
- Mower JP, Palmer JD. 2006. Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. *Mol Genet Genomics.* 276(3):285–293.
- Notsu Y, et al. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics.* 268(4):434–445.
- Oda K, et al. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol.* 223(1):1–7.
- Rüdinger M, Funk HT, Rensing SA, Maier UG, Knoop V. 2009. RNA editing: 11 sites only in the *Physcomitrella patens* mitochondrial

- transcriptome and a universal nomenclature proposal. *Mol Genet Genomics*. 281:473–481.
- Salone V, et al. 2007. A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett*. 581(22):4132–4138.
- Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell*. 19(4):1251–1264.
- Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evol Biol*. 10:274.
- Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. 2010. Extensive loss of RNA editing sites in rapidly evolving silene mitochondrial genomes: selection vs. retroprocessing as the driving force. *Genetics* 185(4):1369–1380.
- Smith DR. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. *Plant Mol Biol*. 71(6):627–639.
- Smith DR, Lee RW. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol Biol Evol*. 25(3):487–496.
- Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? *J Mol Evol*. 48(3):303–312.
- Tasaki E, Sugita M. 2010. The moss *Physcomitrella patens*, a model plant for the study of RNA editing in plant organelles. *Plant Signal Behav*. 5(6):1–3.
- Terasawa K, et al. 2006. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol Biol Evol*. 24(3):699–709.
- Ullrich H, Lüttig K, Brennicke A, Knoop V. 1997. Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. *Plant Mol Biol*. 33(1):37–45.
- Wang B, Xue J, Li L, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Curr Genet*. 55(6):601–609.
- Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell* 25(3):793–803.
- Weng JK, et al. 2010. Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. *Plant Cell*. 22(4):1033–1045.
- Xue JY, Liu Y, Li L, Wang B, Qiu YL. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr Genet*. 56(1):53–61.
- Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M. 2009. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell*. 21(2):558–567.

**Associate editor:** John Archibald