Group II Intron-Mediated Trans-Splicing in the Gene-Rich Mitochondrial Genome of an Enigmatic Eukaryote, *Diphyllatia rotans*

Ryoma Kamikawa1,2,*, Takashi Shiratori3, Ken-Ichiro Ishida4, Hideaki Miyashita1,2, and Andrew J. Roger5,6

1Graduate School of Human and Environmental Studies, Kyoto University, Japan  
2Graduate School of Global Environmental Studies, Kyoto University, Japan  
3Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan  
4Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan  
5Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada  
6Program in Integrated Microbial Biodiversity, Canadian Institute for Advanced Research, Halifax, Nova Scotia, Canada  

*Corresponding author: E-mail: kamikawa.ryoma.7v@kyoto-u.ac.jp.  
Associate editor: Martin Embley  
Accepted: January 21, 2016  
Data deposition: This project has been deposited at GenBank/EMBL/DDBJ under the accession AP015014 (complete mitochondrial genome sequence of *Diphyllatia rotans*).

Abstract

Although mitochondria have evolved from a single endosymbiotic event, present day mitochondria of diverse eukaryotes display a great range of genome structures, content and features. Group I and group II introns are two features that are distributed broadly but patchily in mitochondrial genomes across branches of the tree of eukaryotes. While group I intron-mediated trans-splicing has been reported from some lineages distantly related to each other, findings of group II intron-mediated trans-splicing has been restricted to members of the Chloroplastida. In this study, we found the mitochondrial genome of the unicellular eukaryote *Diphyllatia rotans* possesses currently the second largest gene repertoire. On the basis of a probable phylogenetic position of *Diphyllatia*, which is located within Amorphea, current mosaic gene distribution in Amorphea must invoke parallel gene losses from mitochondrial genomes during evolution. Most notably, although the cytochrome c oxidase subunit (cox) 1 gene was split into four pieces which located at a distance to each other, we confirmed that a single mature mRNA that covered the entire coding region could be generated by group II intron-mediated trans-splicing. This is the first example of group II intron-mediated trans-splicing outside Chloroplastida. Similar trans-splicing mechanisms likely work for bipartitely split *cox2* and *nad3* genes to generate single mature mRNAs. We finally discuss origin and evolution of this type of trans-splicing in *D. rotans* as well as in eukaryotes.

Key words: Diphyllatia, introns, mitochondria, inverted repeats, trans-splicing.

Main Text

Mitochondria in extant eukaryotic cells are direct descendants of endosymbiotic alpha-proteobacteria which were already integrated as organelles in the last common ancestor of extant eukaryotes (LECA) (Gray 2012). Mitochondrial genomes are streamlined in size and gene number when compared with current alpha-proteobacterial genomes (Gray et al. 2004). Nonessential genes in the ancestral endosymbiont genome were presumably lost prior to LECA, and, once a protein targeting system evolved, other genes were transferred to host nuclear genomes (Gray 1999; Adams and Palmer 2003). Comparative analysis of mitochondrial genomes have revealed that even closely related strains/species/lineages can have different gene repertoires, suggesting...
that gene loss from mitochondrial genomes is still ongoing (e.g., Gray 1999; Adams and Palmer 2003; Hancock et al. 2010; Masuda et al. 2011; Kamikawa et al. 2014). The most gene-rich mitochondrial genomes are those of the excavate group Jakobida with 65–100 kb in length, which carry 61–66 protein genes and 30–34 RNA genes (Burger et al. 2013). The most gene-rich mitochondrial genome outside Jakobida so far was found in two heterolobosean Naegleria species of which 49 kb-long mitochondrial genomes carried 42 protein genes and 23 RNA genes (Herman et al. 2013). The other extreme, mitochondrial genomes of dinoflagellates, apicomplexan parasites, and their close relatives, that is, Chromera vela and Vitrella brassicaformis, retain only two or three protein genes and fragmented rRNA genes (Feagin 2000; Kamikawa et al. 2007; Kamikawa, Nishimura, et al. 2009; Plegontov et al. 2015). The mitochondrial genomes of apicomplexan parasites are also the smallest in size (6–7 kb; Feagin 2000).

In addition to gene repertoires and size, the diversity of mitochondrial genomes is also reflected in types of introns and splicing. Introns found in mitochondrial genomes are group I and group II introns with extensive RNA secondary structures. While all eukaryotes with few exceptions possess spliceosomal introns in their nuclear genomes, group I and group II introns in mitochondrial genomes are sparsely distributed in the tree of eukaryotes; many mitochondrial genomes completely lack them. It has been argued that this patchy distribution of group I and II introns in organelar genomes is the mainly the product of homing/retrohoming and transposition/retrotransposition mechanisms facilitated by endonucleases (e.g., Hardy and Clark-Walker 1991; Goddard and Burt 1999; Gogarten and Hilario 2006; Kamikawa, Masuda, et al. 2009; Nishimura et al. 2012, 2014; Zimmerley and Semper 2015). During splicing, introns are removed from a precursor RNA, and the concomitant ligation of exons results in the formation of a mature transcript. When this process of intramolecular ligation involves only a single RNA molecule it is called cis-splicing (Glanz and Kuck 2009). In cases that involve more than one primary transcript in an intermolecular ligation, the RNA is processed by trans-splicing (Glanz and Kuck 2009). This latter splicing reaction is a known variant mechanism of spliceosomal, group I and group II intron splicing (Dorn et al. 2001; Hastings 2005; Fischer et al. 2008; Glanz and Kuck 2009; Nilsen and Graveley 2005). Group I intron trans-splicing was first discovered in mitochondria of placozoan animals (Burger et al. 2009) and a lycophyte plant (Grewe et al. 2009), but was later found to be also distributed in mitochondria of some green algae and fungi (Pombert and Keeling 2010; Nadimi et al. 2012; Pelin et al. 2012; Pombert et al. 2013). Group II intron trans-splicing was discovered more than 30 years ago from mitochondria of the liverwort Marchantia polymorpha (Fukuzawa et al. 1986) and plastids of the green alga Chlamydomonas reinhardtii (Kück and Kuck 1987). To date, group II intron-mediated trans-splicing has only been found in the Chloroplastida assemblage composed of green algae and land plants (e.g., Bonen 2008; Glanz and Kuck 2009).

Diphylleia rotans is a bactivorous, unicellular eukaryote with two flagella. Electron microscopic and molecular phylogenetic analyses suggested close relationship between Diphylleia and the early branching eukaryote Colleolyta (Bruggerole et al. 2002; Zhao et al. 2012), resulting in a group that has been named Diphylleia (Cavalier-Smith 2003). The precise phylogenetic position of Diphylleia in the tree of eukaryotes remains unclear (Zhao et al. 2012). Recently, it was proposed that eukaryotes can be divided into three large clades called Amorphea, Excavata, and Diaphoretickes and the root of the eukaryotic tree of life lies between Amorphea on the one hand, and Excavata and Diaphoretickes on the other (Adl et al. 2012; Brown et al. 2013; Erne et al. 2014; Derelle et al. 2015). In this scheme, Diphylleia is supposed to be more likely a member of Amorphea (Derelle et al. 2015).

We completely sequenced the mitochondrial genome sequence of the “deeply branching” protist D. rotans to gain insight into the early evolution of mitochondrial genomes in eukaryotes. We find that, consistent with an early-branching position of Diphylleia, this organism has a relatively rich gene content, the second most “bacterial-like” mitochondrial gene content after the jakobids. Unexpectedly, we also found the first examples of group II intron-mediated trans-splicing from a eukaryote outside of the Chloroplastida.

The Structure and Content of the Mitochondrial Genome of D. rotans

The complete sequence of the mitochondrial genome of D. rotans can be mapped as a circular molecule that is 62,563 bp in length (fig. 1). The overall A+T content is 65.6%. The coding region occupies 41,240 bp which is approximately 65.9% of the entire sequence. This genome encodes three rRNA genes and 27 tRNA genes, which is sufficient to translate codons corresponding to all the amino acids (supplementary table S1, Supplementary Material online). We identified 49 protein-coding genes by their sequence similarity to orthologues and 7 functionally uncharacterized open reading frames (orf98, orf136, orf139, orf151, orf152, orf154, and orf211). Two cis-spliced group I introns exist in cox1 and nad5, and two cis-spliced group II introns occur in cox11 and nad7. We confirmed these introns were spliced in vivo by RT-PCR and sequencing of the resulting amplicons (supplementary fig. S1, Supplementary Material online). In addition to the conventional introns, we also found three group II introns in multiple pieces in cox1, cox2, and nad3 (discussed below in more detail). Almost all the D. rotans introns lack intronic open reading frames (ORFs). Of these, only cox2 intron carries an intronic ORF that contains a maturation domain and lacks other domains typical of group II intron ORFs such as the reverse transcriptase domain, maturation domain, and DNA-binding domain, respectively.

Genome Biol. Evol. 8(2):458–466. doi:10.1093/gbe/evw011 Advance Access publication February 1, 2016 459
The complete mitochondrial genome of *D. rotans*. Functionally identifiable protein-coding, SSU rRNA, and LSU rRNA genes are depicted as closed boxes whereas unidentified ORFs are as open boxes. 5S rRNA and tRNA genes are shown as lines. Intron sequences are colored in gray. Split genes are highlighted in magenta. Three types of conserved palindromic sequences and nonconserved palindromic sequences are indicated by different colors: type A in red, type B in blue, type C in yellow, and nonconserved type in black.

**Fig. 1.**—The complete mitochondrial genome of *D. rotans*. Functionally identifiable protein-coding, SSU rRNA, and LSU rRNA genes are depicted as closed boxes whereas unidentified ORFs are as open boxes. 5S rRNA and tRNA genes are shown as lines. Intron sequences are colored in gray. Split genes are highlighted in magenta. Three types of conserved palindromic sequences and nonconserved palindromic sequences are indicated by different colors: type A in red, type B in blue, type C in yellow, and nonconserved type in black.
domain, and endonuclease domain (e.g., Zimmerly and Semper 2015). Unfortunately, the introns sequences in this genome are divergent, and therefore, their evolutionary origins are impossible to investigate using molecular phylogenetic analyses. Nevertheless, it is noteworthy that the two group I introns and one group II intron in D. rotans are located in homologous sites to those of other eukaryotes (supplementary table S2, Supplementary Material online).

**Abundant Palindromic Repeats in the D. rotans Mitochondrial Genome**

Closer inspection of the D. rotans mitochondrial genome revealed a series of short palindromic repeats; the consensus sequences, complementary bases, and copy numbers of the palindromic elements are outlined in figure 1. The short palindromic repeats in the D. rotans mitochondrial genome are of 21–45 nt in length and restricted to intergenic and intronic regions, with the exception of the palindromic elements in unidentified orf139, orf151, and orf154. When contained within introns, the palindromic repeats are confined to the non-ORF portions of the group I and group II introns (fig. 1). In total, we detected 83 short palindromic repeats that could be divided into three general types (type A to C; fig. 1 and supplementary fig. S2, Supplementary Material online) except for 12 “nonconserved” elements. Within each repeat type, the nucleotide sequences are almost identical to each other (supplementary fig. S2, Supplementary Material online). Palindromic repeat elements in other mitochondrial genomes have been thought to be selfish, mobile DNA elements (Nakazono et al. 1994; Paquin et al. 2000). This hypothesis can simply explain our observation of many palindromes with similar nucleotide sequences over the mitochondrial genome. Some short palindromic repeats have been thought to contribute to regulation of transcription (Ohta et al. 1998) and they could be involved in similar functions in D. rotans mtDNA. In any case, a notable consequence of such repeats is that they contribute to genome rearrangements (Nedelcu and Lee 2015), and scrambling of genes (Smith and Lee 2009); Hopefully, gathering more mitochondrial genome sequences from related diphyllatians will reveal whether they undergo rapid genome rearrangements mediated by the palindromes and whether such rearrangements resulted in generation of the split introns.

**A Large Set of Proteins Encoded on Mitochondrial DNA**

Although the mitochondrial gene repertoire of D. rotans is nearly a subset of that of the Jakobida species (fig. 2), it is conspicuously larger than that of Naegleria spp. (containing 42 protein genes; fig. 2). Thus, D. rotans has the most gene-rich mitochondrial genome outside Jakobida. In addition to the relatively common genes encoding subunits of complex I (nad1-4,4L,5-7,9), complex III (cob), complex IV (cox1-3), and complex V (atp6,8,9), D. rotans mitochondrial DNA encodes rarely found genes for complex I (nad8,10,11), complex II (sdh2-4), complex V (atp4), cytochrome c maturase (ccmC and ccmF), cytochrome c oxidase assembling protein (cox11), and ribosomal proteins (rps10, rpl11,20,23, 31,32,34; fig. 2). Among the ribosomal protein genes, only D. rotans and Jakobida possess rpl32, and rpl34 encoded on their mitochondrial genomes; rpl23, by contrast, is a gene which has been identified in no other mitochondrial genome to date (fig. 2).

To gain a better understanding of the mitochondrial genome content of the last eukaryotic common ancestor (LECA) and evolution of the gene repertoire in Diphylloptera and its close relatives, we mapped the mitochondrial genome content of diverse lineages across the tree of eukaryotes. Assuming only events of gene loss or gene transfer to the nucleus, the mitochondrial genome of LECA must have encoded at least 70 proteins (fig. 2; black balloon a). If the eukaryote root falls between Amorphea (A) on the one hand, and Excavata (E) and Diaphoretickes (D) on the other (Derelle et al. 2015), the common ancestor E + D likely had 67 genes after losing only three (balloon c in fig. 2). The lineage leading to the ancestor of Amorphea appears to have lost at least 13 genes resulting in 57 genes (balloon b in fig. 2). Curiously, the D. rotans mitochondrial genome possesses genes for rpl23, rpl32, rpl34, nad8, nad10, sdh2-4, and cox11 whereas all other amorphean mitochondrial genomes lack them (fig. 2). There are other mitochondrial genomes within the Amorphea whose unique gene content implies massive parallel gene loss. For example, the mitochondrial genome of Malawimonas jakobiformis exclusively carries genes for rpl1, rpl18, rpl36, and ccmB (although M. jakobiformis branching within Amorphea remains controversial). Similarly, some, but not all, amoebozoon species possess rps16, tufA, and atp1. The requirement for massive parallel loss events in these cases does not significantly change if the root of eukaryotes is located in several alternative positions such as the branch leading to Jakobids (He et al. 2014) or that leading to Euglenozoa (Cavalier-Smith 2010). However, if the precise phylogenetic position of D. rotans is found to be outside Amorphea in future analyses, the gene loss events predicted here would require reevaluation.

**Group II Intron-Mediated Trans-Splicing**

The cox1 gene in the D. rotans mitochondrial genome is broken up into four fragments, that are scattered over the genome (fig. 1) that will henceforth be referred to as cox1-E1 to cox1-E4. Two of the fragments, cox1-E2 and cox1-E3,
are interrupted by a cis-spliced group I intron as mentioned above and located between rpl2 and rpl23 (fig. 1, supplementary figs. S3 and S4, Supplementary Material online). Cox1-E4 is located on the same strand as cox1-E2/cox1-E3 but is separated by approximately 5 kb of DNA that encodes a total of four genes: cox1-E4 is located downstream of nad1 (fig. 1, supplementary figs. S3 and S4, Supplementary Material online). In contrast, cox1-E1 is located upstream of nad2 on the opposite strand from the other cox1 fragments (fig. 1, supplementary figs. S3 and S4, Supplementary Material online). The curious fragmentation patterns and distribution over the genome are suggestive that a mature cox1 transcript might be assembled via trans-splicing reactions. To provide evidence that the mitochondrial cox1 pieces in D. rotans are processed into single mature mRNA in vivo, we conducted RT-PCR experiments, using primers located in the respective flanking exons (supplementary fig. S3, Supplementary Material online). Sequencing of the resulting PCR products confirmed that the exons of cox1 were accurately ligated in vivo (fig. 3A and B), probably through trans-splicing. Sequencing of the PCR products confirmed 100% sequence identity to the corresponding exons. No PCR product was obtained using total genomic DNA as the template (fig. 3A and B), ruling out the possibility that contiguous homologues exist in the nuclear genome.

Some trans-splicing phenomena in mitochondria are associated by split group I or group II introns (Glanz and Kuck 2009). In other cases, such as trans-splicing in Diplonema (Vlcek et al. 2011) and dinoflagellate mitochondrial mRNAs (Jackson and Waller 2013), the mechanisms are unknown. In both group I and group II intron-mediated trans-splicing, fragmented intron pieces flanking to a coding region of transcripts form a precise secondary structure, and are then excised to generate contiguous mRNA. In silico scanning of the D. rotans mitochondrial genome allowed us to identify partial group II intron fragments flanking the split cox1 gene pieces; RNAweasal and Mfannot identified domain V, which is a highly conserved structural domain of group II introns (Zimmerly and Semper 2015), in the flanking regions of cox1 fragments (data not shown). We found that 3′-flanking
region of cox1-E1 and 5'-flanking region of cox1-E2 had the potential to form an intermolecular stem structure (supplementary fig. S5A, Supplementary Material online). We also predicted a group II intron-like RNA secondary structure, that is, six stem-loop structures radiating from a central wheel (Zimmerly and Semper 2015; supplementary fig. S5A, Supplementary Material online), of the 30-flanking region of cox1-E1 and the 50-flanking region of cox1-E2 as a whole. In this prediction, the intermolecular stem structure is tentatively identified as Domain IV. Similarly, we found that 3′-flanking region of cox1-E3 and 5′-flanking region of cox1-E4 also had the potential to form a group II intron RNA secondary structure with the intermolecular stem structure again within Domain IV (supplementary fig. S5B, Supplementary Material online). The RT-PCR results and predicted secondary structures of flanking regions strongly suggest that cox1 of D. rotans uses group II intron-mediated trans-splicing for maturation. In what follows, we propose a model of the maturation process of the cox1 transcripts (fig. 3C). First, each set of cox1-E1/flanking group II intron fragment and cox1-E4/flanking group II intron fragment is transcribed independently whereas cox1-E2 and cox1-E3, separated by a cis-spliced group I intron, are transcribed as a single transcript with group II intron fragments at each end. The three transcripts then associate by forming intermolecular stem structures. Finally, each intron or intron fragment is spliced via either the group I (cis-spliced intron within the cox1-E2/cox1-E3 transcript) or group II intron splicing mechanisms, resulting in a single mature cox1 mRNA. Again, to date, group II intron-mediated trans-splicing has been only found in organellar genomes of green algae and land plants (Bonen 2008) and thus this is the first report of group II intron-mediated trans-splicing in a species outside Chloroplastida.

The cox2, nad3, and ccmF genes of the Diphyleia mitochondrial genome are also broken up into two fragments, and the fragments are scattered over the genome. Although the bipartite fragments of ccmF (ccmF-N and ccmF-C) and nad3 (nad3-E1 and nad3-E2) are encoded on the same strand of the mitochondrial genome, they are located at a distance of more than 15 kb from each other. The bipartite fragments of cox2

Fig. 3.—Trans-splicing in cox1 transcripts. (A) Reverse transcriptase PCR for transcripts of cox1-E1 and cox1-E2. Genomic DNA (left lane), cDNA (middle lane), and distilled water (right lane) were used as templates. (B) Reverse transcriptase PCR for transcripts of cox1-E3 and cox1-E4. Genomic DNA (left lane), cDNA (middle lane), and distilled water (right lane) were used as templates. (C) A model of cox1 mRNA maturation. Mitochondrial DNA, cox1 coding regions, and intron regions are depicted as thin lines, boxes, and thick lines, respectively. Independent gene fragments and their transcripts are distinguished by different colors, whereas the group I intron splitting cox1-E2 and cox1-E3 is colored in black. Four cox1 gene fragments are located on the mitochondrial DNA separately. Cox1-E2 and cox1-E3 are separated by insertion of a cis-spliced group I intron. In this model, premature RNAs from separately distributed gene fragments for cox1 are transcribed with flanking intron regions. Flanking intron regions form intermolecular stem structures that associated to form group II intron secondary structures.
are located on different strands from each other. We confirmed cox2 and nad3 were not pseudogenes but posttranscriptionally processed, without sequence modification, by sequencing of an RT-PCR product (supplementary figs. S5C–SSF, Supplementary Material online). We inferred a group II intron-like secondary structure with intermolecular stems between the 3′-flanking region of N-terminal coding region and 5′-flanking region of C-terminal coding region in cox2 and nad3 (supplementary figs. SSC–SSF, Supplementary Material online), strongly suggesting that the two genes are also spliced in trans by group II intron splicing mechanisms. However, the cox2 and nad3 introns are disrupted at sites within Domain V and Domain I, respectively, different from the Domain IV split sites that have been previously reported for trans-spliced introns (Bonen 2008) and found in the cox1 split introns of D. rotans (supplementary fig. S5, Supplementary Material online). To verify our prediction that the cox2 and nad3 introns are truly split within Domain V and Domain VI, respectively, sequencing mitochondrial genomes of close relatives of D. rotans may reveal intact close homologs of these introns. Because the intermolecular stem structures we predict in these group II introns involve only a few dozen base-paired residues, we expect that D. rotans may possess proteins that stabilize these structures, as proposed for plant organellar trans-splicing (Bonen 2008).

In contrast, we find no evidence of mature mRNAs that cover both ccmF-N and ccmF-C in both PCR reactions with gDNA and cDNA (data not shown) and neither show intermolecular secondary structures flanking them, suggesting 1) the proteins may be truly fragmented into two functional peptides like the CcmF of the land plant Arabidopsis thaliana (Rayapuram et al. 2008) or 2) the ccmF pieces we detect are pseudogenes.

Understanding the Evolutionary Origins of Trans-Splicing in Eukaryotes

In Chloroplastida, trans-spliced group II introns are thought to have been generated by physical splits of cis-spliced group II introns caused by genomic rearrangements in these mitochondrial genomes (Qiu and Palmer 2004). Similarly, it is possible that the palindromic repeats that are widely distributed over the D. rotans mitochondrial genome have mediated genomic rearrangements (e.g., Nedelcu and Lee 1998) in an ancestral diphyllyatian mitochondrial genome and this generated the gene fragments and trans-splicing introns we have observed in D. rotans. In support of this scenario, five of the split intron fragments and two cis-spliced introns are located in close proximity to and include short palindromic repeats, respectively (fig. 1). If the split introns in D. rotans recently evolved, the mitochondrial genomes of close relative of D. rotans could have contiguous introns that are homologous to the D. rotans split introns. Future determination of complete mitochondrial genomes of other diphyllyatian may help clarify the origins of split introns in D. rotans.

Diphylezia rotans has the first mitochondrial genome outside the Chloroplastida where trans-splicing of introns is known to occur. We note that the apusozoan Thecamonas trehans, a sister lineage of Opisthokonta (Derelle et al. 2015), has a candidate split group II intron in its mitochondrial genome: its nad3 is annotated as a bipartite split gene, the two fragments are not encoded next to each other, and one of the fragments is flanked by a region that has the potential to form a group II intron-like structure (NC_026452; Valach et al. 2014). Clearly, group II intron-mediated trans-splicing is more widespread in the tree of eukaryotes than previously thought.

Materials and Methods

Culturing of D. rotans, DNA Extraction, and Mitochondrial DNA Sequencing

Diphylezia rotans NIES-3764 was cultivated with the cyanobacterial strain Microcystis aeruginosa NIES-298 as a food source, in C-Si medium (http://mcc.nies.go.jp/02medium.html; last accessed February 9, 2016) at 20 °C under 10–50 micromole photons/m²/s with the 14 h light/10 h dark cycle. DNA was extracted from cells of D. rotans NIES-3764 together with the cyanobacterial cells using the Plant DNA extraction kit (Jena BioSciences) following the manufacturer’s instructions. Total DNA was sent to Hokkaido System Science Co., Ltd (Japan) for 100-bp paired-end sequencing by the Illumina HiSeq2000 platform using the 350-bp library constructed with the Truseq Nano DNA LT Sample Prep Kit (illumina) following the manufacturer’s instructions. The sequenced reads were filtered on the basis of fluorescence purity by Chastity (Chastity = Highest intensity/(Highest intensity + Next highest intensity)): reads were passed if reads had no more than one cycle of a chastity below 0.6 within the first 25 cycles. Removal of adapter sequences, all the reads containing “N,” and all the reads containing adapter sequences at the 3′-ends were performed using cutadapt ver.1.1 (https://cutadapt.readthedocs.org/en/stable/, last accessed February 9, 2016), resulting in 21 million reads yielding 21 Gb of data where 90% of nucleotides had Q30 scores and of which the mean quality score was greater than 35. Subsequent assembly was performed with Velvet ver.1.2.08 (Zerbino et al. 2009) with hash length 65. Seven contigs derived from the mitochondrial genome were detected by homology search using the protein and rRNA sequences from the jakobid Andalucia godoyi as queries. The mean coverage of these detected contigs was more than 60. We also performed an assembly with the same program with hash length 95 and there was no significant conflict detected between the contigs assembled with the different hash length parameters. After prediction of adjacent contigs...
with paired-end information, gaps between the contigs generated with a hash length 65 were closed by PCR amplification (primers reported in supplementary table S3, Supplementary Material online) and subsequent Sanger sequencing of amplimers. Genes encoding proteins and rRNAs were identified by BLASTx and BLASTN searches against the nonredundant databases at the National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/blast.cgi, last accessed February 9, 2016; Altschul et al. 1990), which also confirmed the identified genes were not derived from misassemblies of cyanobacteria-derived reads. Transfer RNA-encoding genes were found by using tRNAscan-SE (Lowe et al. 1997). The mitochondrial genome sequence was deposited in the DNA Data Bank of Japan (GenBank/EMBL/DDBJ; accession no. AP015014).

Annotation was also performed by MFANNOT (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface, last accessed February 9, 2016) and RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface, last accessed February 9, 2016) with the standard genetic code. Intramolecular interaction of split group I intron RNA precursors were predicted by Mfold (Zuker 2003) with the default settings, followed by manual modifications according to the model structure of these types of organellar introns. Palindromic sequence elements were detected by EMBOSS explorer (http://emboss.bioinformatics.nl/cgi-bin/emboss/palindrome, last accessed February 9, 2016) with emboss/palindrome, last accessed February 9, 2016) with RNAweaselInterface.pl, last accessed February 9, 2016) and RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl, last accessed February 9, 2016) with MFANNOT online and subsequent Sanger sequencing of amplimers. Some detected sequences were manually excluded if a loop region was longer than a stem region.

To confirm that cox1, cox2, nad3, and ccmF were split genes as assembled, we performed PCR assays for the regions between the split gene fragments and their adjacent protein-coding genes (supplementary figs. S3 and S4, Supplementary Material online). In order to confirm in vivo splicing reactions, we conducted reverse transcriptase PCR assays. RNA was extracted from cultures using Trizol (Invitrogen) following the manufacturer’s instruction. cDNA was synthesized with random hexamers, total RNA, and 3’-rapid-amplification-of-cDNA-ends kit (Invitrogen). PCR was conducted with either gDNA, cDNA, or distilled water as the template and primers were designed for each exon (supplementary fig. S3, Supplementary Material online), and PCR products were sequenced by the Sanger method.

Supplementary Material

Supplementary figures S1–S5 and tables S1–S3 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was performed, in part by grants from the Japan Society for Promotion of Sciences (15H05606 and 15H14591 awarded to R.K.). This work was also supported by a grant from the Institute for Fermentation, Osaka, Japan (awarded to R.K.) and by a Natural Sciences and Engineering Research Council Discovery grant (awarded to A.J.R.).

Literature Cited

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 29:380–395.

Adl SM, et al. 2012. The revised classification of eukaryotes. J Eukaryot Microbiol. 59:429–493.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Beaudet D, et al. 2013. Mitochondrial genome rearrangements in Glomus species triggered by homologous recombination between distinct mtDNA haplotypes. Genome Biol Evol. 5:1628–1643.

Brown MW, et al. 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc Biol Sci. 280:20131755.

Bonen L. 2008. cis- and trans-splicing of group II introns in plant mitochondria. Mitochondrion 8:26–34.

Brugerolle G, Bricheux G, Philippe H, Coffe G. 2002. Collobdicyton tricillatum and Diphyleia rotans (=Aulacornonas submarina) form a new family of flagellates (Collobdicytonidae) with tubular mitochondrial cristae that is phylogenetically distant from other flagellate families. Protoplast 153:59–70.

Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. Genome Biol Evol. 5:418–438.

Burger G, Yan Y, Javadi P, Lang BF. 2009. Group-I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. Trends Genet. 25:381–386.

Cavalier-Smith T. 2003. The excavate protozoan phyla Metamonada Grasse emend. (Aeromonadom, Parabasalas, Carpediemonas, Ephyrangia) and Loukozoa emend. (Jakobaea, Malawimonas); their evolutionary affinities and new higher taxa. Int J Syst Evol Microbiol. 53:1741–1758.

Cavalier-Smith T. 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett. 6:342–345.

Derelle R, et al. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc Natl Acad Sci U S A. 112:E693–E699.

Dorn R, Reuter G, Loewendorf A. 2001. Transgene analysis proves mRNA trans-splicing at the complex mod/mdg4 locus in Drosophila. Proc Natl Acad Sci U S A. 98:9724–9729.

Eme L, Sharpe SC, Brown MW, Roger AJ. 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb Perspect Biol. 6:a016139

Feagin JE. 2000. Mitochondrial genome diversity in parasites. Int J Parasitol. 30:371–390.

Fischer SE, Butler MD, Pan Q, Ruvkun G. 2008. Trans splicing in C elegans generates the negative RNAI regulator ERI-6/7. Nature 455:491–496.

Flegontov P, et al. 2015. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. Mol Biol Evol. 32:1115–1131.

Fukuzawa H, et al. 1986. Coding sequences for chloroplast ribosomal protein S12 from the liverwort Marchantia polymorpha are separated far apart on the different DNA strands. FEBS Lett. 198:11–15.

Glantz S, Kuck U. 2009. Trans-splicing of organelle introns—a detour to continuous RNAs. Bioessays 31:921–934.
Grewe F, Viehoever P, Weisshaar B, Knoop V. 2009. A
Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Ann Rev
Gray MW. 2012. Mitochondrial evolution. Cold Spring Harb Perspect Biol.
Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish
Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish
Genome Biol. Evol. 24:465–470.
Herman EK, et al. 2013. The mitochondrial genome and a 60-kb nuclear
Hancock L, Goff L, Lane C. 2010. Red algae lose key mitochondrial genes
Hastings KE. 2005. SL
Hardy CM, Clark-Walker GD. 1991. Nucleotide sequence of the COX1
gene. Proc Natl Acad Sci U S A. 96:13880–13885.
Kamikawa et al. 2011. Mitochondrial genomes from two red tide-forming
Kamikawa R, Inagaki Y, Sako Y. 2009. Analysis of the mitochondrial
Ku¨ ck U, Choquet Y, Schneider M, Dron M, Bennoun P. 1987. Structural and
transcription analysis of two homologous genes for the P700 chlorophyll alpha apoproteins in Chlamydomonas reinhardtii. Evidence for in vivo trans-splicing. EMBO J. 6:2185–2196.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection
Nakazono M, Kanno A, Tsutsumi N, Hirai A. 1994. Palindromic repeated
sequences (PRSSs) in the mitochondrial genome of rice: evidence for their insertion after divergence of the genus Oryza from the other Gramineae. Plant Mol Biol. 24:273–281.
Nedelcu AM, Lee RW. 1998. Short repetitive sequences in green algal
mitochondrial genomes: potential roles in mitochondrial genome evo-
lution. Mol Biol Evol. 15:690–701.
Nilsen TW, Gravelle BR. 2010. Expansion of the eukaryotic proteome by
alternative splicing. Nature 463:457–463.
Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. 2012. Two independ-
ent lateral group I intron transfers in the katablepharid Leucocryptos
marina. PLoS One 7:e37307.
Nishimura Y, Kamikawa R, Hashimoto T, Inagaki Y. 2014. An intronic open
reading frame was released from one of group II introns in the mito-
ochondrial genome of the haptophyte Chrysochromulina sp. NIES-
1333. Mob Genet Elements. 4:e29384.
Ohta N, Sato N, Kurowa T. 1998. Structure and organization of the mito-
ochondrial genome of the unicellular red algae Cyanidioschyzon mer-
olae: evidence from the complete nucleotide sequence. Nucleic Acids Res.
26:5190–5198.
Paquin B, Laforest MJ, Lang BF. 2000. Double-hairpin elements in the mito-
ochondrial DNA of Allozymes: evidence for mobility. Mol Biol Evol.
17:1760–1768.
Pelín A, et al. 2012. The mitochondrial genome of the arbucular mycor-
rhizal fungus Gigaspora margarita reveals two unsuspected trans-splic-
ing events of group I introns. New Phytol. 194:836–845.
Pombert JF, Keeling PJ. 2010. The mitochondrial genome of the entomo-
parasitic green alga Helicopsoridium. PLoS One 5:e89594.
Pombert JF, Ols C, Turmel M, Lemieux C. 2013. The mitochondrial
genome of the prasinophyte Prasinodermia coloniale reveals two trans-spliced group I introns in the large subunit rRNA gene. PLoS One 8:e84325.
Qui Y-L, Palmer JD. 2004. Mary independent origin of trans splicing of a plant mitochondrial group II intron. J Mol Evol. 59:80–89.
Rayaparum N, Hagenmuller J, Grienenger JM, Bonnard G, Giege´ P. 2008. The three mitochondrial encoded CcmF proteins form a com-
plex that interacts with CCMH and c-type apocytochromes in Arabidopsis. J Biol Chem. 283:25200–25208.
Smith DR, Lee RW. 2009. The mitochondrial and plastid genomes of Valvax carteri: bloated molecules rich in repetitive DNA. BMC Genomes 10:132.
Valach C, Laforest MJ, Lang BF. 2014. Widespread occurrence of organelle genome-encoded 55 rRNAs including permuted molecules. Nucleic Acids Res. 42:13764–13777.
Vlcek C, Marande W, Teijeiro S, Lukes J, Burger G. 2011. Systematically
fragmented genes in a multipartite mitochondrial genome. Nucleic Acids Res. 39:979–988.
Zerboin DR, McEwen GK, Margules EH, Birney E. 2009. Pebble and rock
band: heuristic resolution of repeats and scaffolding in the velvet short-
read de novo assembler. PLoS One 4:e8407.
Zhao S, et al. 2012. Colloclidicyon—an ancient lineage in the tree of eu-
karyotes. Mol Biol Evol. 29:1557–1568.
Zimmerly S, Semper C. 2015. Evolution of group II introns. Mob DNA. 6:7.
Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization
prediction. Nucleic Acids Res. 31:3406–3415.