Evolutionarily Conserved cox1 Trans-Splicing Without cis-Motifs

Georgette N. Kiethega,1 Marcel Turcotte,2 and Gertraud Burger*1,3

1Department of Biochemistry, Université de Montréal, Montreal, Canada
2School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada
3Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montreal, Canada

*Corresponding author: E-mail: gertraud.burger@umontreal.ca.
Associate editor: Andrew Roger

Abstract

In the protist Diplonema papillatum (Diplonemea, Euglenozoa), mitochondrial genes are systematically fragmented with each nonoverlapping piece (module) encoded individually on a distinct circular chromosome. Gene modules are transcribed separately, and precursor transcripts are assembled to mature mRNA by a trans-splicing process of yet unknown mechanism. Expression of the cox1 gene that consists of nine modules, also involves RNA editing by which six uridines are added between Modules 4 and 5. Here, we investigate whether the unusual features of cox1 are shared by all Diplonemea and what the mechanism of trans-splicing might be. We examine three additional species representing both Diplonemea genera, namely D. papillatum described before, and D. ambulator, Diplonema sp.2, and Rhynchopus euleeides and discover that in all Diplonemea, the cox1 gene is discontinuous and split up into nine modules that each reside on a distinct chromosome. Positions of gene breakpoints vary by up to two nucleotides. Further, all taxa have six nonencoded uridines inserted in cox1 mRNA at exactly the same position as D. papillatum. In silico searches do not detect signatures of introns known to engage in trans-splicing, in particular Group I, Group II, spliceosomal, and transfer RNA introns. Nor did we find statistically significant reverse-complementary motifs between adjacent modules and their flanking regions, or residues conserved within or across species. This provides compelling evidence that trans-splicing in Diplonemea mitochondria does not rely on sequence elements in cis but rather proceeds by a mechanism employing matchmaking trans factors, such as RNAs or proteins.

Key words: multichromosome mtDNA, gene fragmentation, U-insertion RNA editing, diplonemids (Euglenozoa), phylogeny, pattern search.

The Diplonemea (or “diplonemid”) Diplonema papillatum possesses a most unconventional mitochondrial genome (Maslov et al. 1999; Marande et al. 2005). Instead of a single chromosome type, D. papillatum mitochondrial DNA (mtDNA) is composed of a hundred or so distinct circular molecules of two sizes (Class A, 6 kbp; Class B, 7 kbp). Intriguingly, these chromosomes do not encode multiple genes but rather single nonoverlapping gene pieces (modules) of ~60–340 bp that are transcribed individually and then assembled to mRNA (Marande and Burger 2007; Vlcek et al. 2010). The cox1 transcript seems to be the only RNA undergoing editing. Here, we address two questions. First, are unorthodox genome architecture, gene structure, and gene expression peculiarities of D. papillatum or rather shared by all diplonemids? Second, does trans-splicing rely on a known intron-splicing machinery and are any conserved sequence motifs involved?

We characterized the cox1 gene and its transcript from three additional diplonemid species, D. ambulator, Diplonema sp. 2, and Rhynchopus euleeides (Roy, Faktorova, Benada, et al. 2007) For methods, see Methods section (including references), and for primers used in RT-PCR, see supplementary table S6, Supplementary Material online. Sequences were deposited in GenBank under accession numbers JF698650-80 (supplementary table S5, Supplementary Material online). The cox1 cDNA sequences obtained are contiguous and align perfectly with that from D. papillatum (supplementary fig. S1, Supplementary Material online). Although the translation code is the same in all diplonemids (including UGA = W), codon frequencies and A + T-content differ drastically (supplementary table S1A–D, Supplementary Material online).

Mitochondrial chromosomes of the diplonemids studied here are circular and of two size classes as in D. papillatum (Marande and Burger 2007), but sizes range from 4.5–9 kbp (supplementary table S2, Supplementary Material online). For each species, we sequenced the chromosomes that carry the cox1 portions corresponding to Modules 4, 5, and 9 in D. papillatum and found that these coding regions are each encoded separately by a distinct mitochondrial chromosome as well (supplementary table S3, Supplementary Material online). Noncoding chromosome regions share substantial sequence-similar stretches of up to 4 kbp (supplementary fig. S2, Supplementary Material online). Multichromosome mtDNAs seem to predominate in all euglenozoan groups: kinetoplastids (Lukeš et al. 2002), euglenids (Euglena gracilis, Talen et al. 1974; Yasuhira and Simpson 1997; Gray et al. 2004; Spencer and Gray 2011 and Peranema cantuscyni, Roy, Faktorova, Lukeš, et al. 2007), and diplonemids described here.

We pinpointed the cox1 gene modules by sequencing all genomic coding regions plus adjacent noncoding stretches to compare these with the cDNA. Always, cox1 is split into nine
Paracoccus denitrificans
Marchantia polymorpha
Cyanophora paradoxa
Porphyra purpurea
Acanthamoeba castellanii
Physithora infestans
Saccharomyces cerevisiae
Bos taurus
Bipalium mannan
Reclinomonas americana
Neogloea gruberi
Euglena gracilis
Leishmania tarentolae
Trypanosoma brucei
Diplonema ambulator
Rhynchopus euleides
Diplonema sp. 2
D. papillatum
D. papillatum-Edited

The Cox1 protein sequence of diplonemids is fairly well conserved within the group, but highly divergent compared with other taxa. Regions most variable between diplomemids coincide with those of general low conservation; further, there is no indication of sequence constraints imposed by module junctions (supplementary fig. S4, Supplementary Material online). RNA editing of the diplonemid Cox1 transcripts specifies V-F-S, I-F-S, and L-F-S (fig. 2). Without the added Us, the diplonemid proteins would lack positions otherwise invariably present, although moderately conserved, in Cox1 of other organisms. In the protein's tertiary structure (available for Bos taurus), the corresponding tripeptide is located in a loop that interact with a second more C-terminal loop. Interactions involve positively charged residues in Loop 1 and hydrophobic and/or small residues in Loop 2. Remarkably, the situation is inverse in the diplonemid proteins (Marande 2007). We speculate that the ancestral diplonemid Cox1 gene lost nucleotides at the boundary of Module 4 and/or 5, which was patched secondarily through filling-in bases by a preexisting uridylyl transferase. Because the resulting amino acids were unable to interact with residues of Loop 2, these latter underwent compensatory substitutions to restore the protein's function-critical tertiary structure.

**FIG. 2.** Multiple sequence alignment of Cox1 proteins. The region shown corresponds to residues 251–276 in Bos taurus.

The same position, others slightly shifted (fig. 1B and C; supplementary fig. S3, Supplementary Material online). The Cox1 fragmentation pattern being shared by all examined species allows two extrapolations. First, not only Cox1 but probably all mitochondrial genes are fragmented and trans-spliced in diplonemids as reported for D. papillatum (Vlcek et al. 2011). Second, gene fragmentation and trans-splicing emerged in the common ancestor of diplomemids.

In D. papillatum, U-insertion RNA editing occurs between Cox1 Modules 4 and 5. At that position, six nonencoded Us are also encountered in the other three species (fig. 1D). Surprising is the strict conservation of the editing pattern across diplomemids, whereas in kinetoplastids, the pattern varies considerably, even within the same genus (Feagin 1990). For more details, see Supplementary Discussion, Supplementary Material online.
Known trans-splicing relies on noncontiguous classical introns, notably Group I, Group II, and transfer RNA (tRNA) (or “archaeal”) introns in the case of organelle genes and spliceosomal introns for nuclear genes (Bonen and Vogel 2001). Group I, Group II, and spliceosomal introns possess distinctive nucleotide sequences and secondary and tertiary structures (Breathnach and Chambon 1981; Bonen 1993), which, however, we did not detect for diplomemid cox1. Nor could we uncover any nucleotides conserved across all modules or flanking regions of a given species or across all species for a given module. Instead, multiple sequence alignments of homologous regions show a random distribution of nucleotides (supplementary fig. S5, Supplementary Material online). This eliminates the possibility of classical introns mediating trans-splicing in diplomemid mitochondria.

We investigated whether two neighboring modules or their flanking regions can interact with one another via sequence complementarity to form a helix-loop-helix structure typical for tRNA introns (for more details on in silico analyses, see Supplementary Methods, Supplementary Material online). Alternatively, sequence complementarity without this particular 2D structure could point to a new splicing mechanism relying on sequence elements in cis. In silico search for sequence complementarity assumed a “seed” region of at least six consecutive nucleotides interacting via canonical or wobble base pairing. The dynamic-programming–based algorithm compared for each junction all fixed-length segments from the upstream module with those from the downstream module, and an interaction plot was generated for each junction. Presumably the same trans-splicing mechanism at all junctions, we expected sequence-complementatory segments at the same relative location for all junctions. Yet, no such shared position was found. Therefore, we analyzed the minimum offset to circumscribe a common region of complementarity, and this for the original data set and for simulated data sets of identical dinucleotide composition. Again, the result was negative: There was no significant difference in occurrences of sequence-complementatory regions between the original and simulated data sets (supplementary table S4, Supplementary Material online). These results strongly suggest that cis-elements are not involved in trans-splicing of diplomemid cox1 and corroborate our hypothesis of matchmaking trans-factors. Such factors could also ensure, when required, RNA editing, a task achieved by guide RNAs (gRNAs) in trypanosome mitochondria (Stuart et al. 2005). Work is in progress to test in D. papillatum if the hypothetical trans-acting matchmakers are RNAs. Because preliminary data do not indicate trypanosome-like gRNAs, guiding “proteins” are a sensible alternative to consider.

Finally, we constructed a phylogenetic tree with Cox1 protein sequences from diplomemids, other euglenozoans, Naegleria, and several slow-evolving taxa as outgroups. For more details on phylogenetic analyses, see Supplementary Methods, Supplementary Material online. Figure 3 confirms that diplomemids are the sister group of kinetoplastids and that euglenids diverge basally to the split of the two former groups, as observed in nuclear phylogenies (Busse and Preisfeld 2002; Simpson and Roger 2004; von der Heyden et al. 2004; Breglia et al. 2007). However, the topology within diplomemids differs. In most nucleargene–based trees, the Diplonema genus is monophyletic (with modest statistical support), whereas in the mitochondrial tree, Diplonema embraces R. euleeides (with high support). This suggests that Diplonema consists of three rather than two genera, notably Rhynchopus, the D. papillatum–D. ambulator–clade, and one represented by Diplonema sp. 2 (this topology is shared by one nuclear phylogeny [von der Heyden et al. 2004]). To reconcile the conflicting molecular and morphology-based classification, it would be worthwhile to reexamine the morphological characters that are traditionally used for distinguishing Diplonema and Rhynchopus.

Supplementary Material

Supplementary tables S1–S6, figures S1–S5, Material and Methods, Results, Discussion, and additional references are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank S. Teijeiro (Universite´ de Montre´al) for generating most cDNA data and providing technical advice and assistance, Yifei Yan (Universite´ de Montre´al) for sharing his data on the D. papillatum cox1 Module 8–Module 9 junctions—data he generated in the context of his Master’s thesis—and Pavel Poliak (University of South Bohemia) for generating data on the R. euleeides cox1 Module 8–Module 9 junction during an internship in the laboratory of G.B. B. Fran@t (Universite´ de Montre´al) kindly conducted the phylogenetic analysis and provided helpful comments to the manuscript. This work was supported by grants from the Canadian Institute for Health Research (CIHR, grant MOP-79309; G.B.) and the National Science and Engineering Research Council, Canada (NSERC, grant 250909-2006; M.T.), and a PhD scholarship from the Programme Canadien de Bourses de la Francophonie (PCBF scholarship; G.N.K.).
References


