

# Mitochondrial DNA as a Genomic Jigsaw Puzzle

William Marande and Gertraud Burger\*

Mitochondrial genomes exhibit considerable diversity in chromosome form and number as well as coding content (1), and they have adopted various modes of transcript maturation, such as RNA editing (2) and trans-splicing (3). The mitochondrial DNA (mtDNA) of *Diplonema papillatum*, a member of the free-living diplomonads, the sister group of kinetoplastids (4), is composed of >100 regularly structured chromosomes 6 kbp (class A) or 7 kbp (class B) in size (5). We show that the constant region of each chromosome, which is identical within each size class, covers ~95% of the molecules; a short (<500 bp) variable region (cassette) encloses a distinct gene piece (module) of 60 to 351 bp (Fig. 1A and table S1).

Analysis of genomic (250 kbp) and transcriptome (35 kbp) sequences (6) reveals a conventional complement of mitochondrial genes coding for ribosomal RNAs, apocytochrome b, and subunits of cytochrome oxidase, NADH (reduced form of nicotinamide adenine dinucleotide) dehydrogenase, and ATP (adenosine triphosphate) synthase. Without exception, corresponding transcripts are contiguous, whereas the genomic coding regions are fragmented into at least three pieces. For example, the *cox1* gene (specifying cytochrome oxidase subunit 1) consists of nine

modules, two of which (modules 1 and 4) are encoded on class B and all others on class A chromosomes (table S1).

How are these systematically modularized genes expressed? Polymerase chain reaction (PCR) amplification with total genomic DNA does not reveal concatenated gene modules, which rules out splicing at the DNA level. Likewise, fragmentation of gene products does not occur, because *Diplonema*'s cDNAs are contiguous. What remains is module concatenation at the RNA level, as experimentally confirmed by Northern hybridization and cDNA sequencing (6). In Northern hybridizations of polyadenylated [poly(A)] RNA, a probe containing the C-terminal-most gene module of *cox1* (m9) identifies nine distinct bands: the mature *cox1* mRNA, an RNA species of ~250 nucleotides (nt) (m9 alone), and combined modules 8 plus 9 (m8...m9), 7 plus 8 plus 9 (m7...m9), etc. to m2...m9 (Fig. 1B, lane 1). Similar results are obtained with other *cox1* modules as probes, except that individual transcripts of internal modules show up exclusively in total, not in poly(A), RNA (Fig. 1B, lanes 2 to 5, asterisks). The cDNA library constructed from poly(A) RNA contains the same intermediates seen in Northern experiments (fig. S1). To summarize, (i) gene modules are transcribed in-

dividually, (ii) only transcripts of C-terminal gene modules undergo polyadenylation, and (iii) contiguous mRNAs are generated via concatenation of separate module transcripts.

Transcript processing in *Diplonema* mitochondria is different from known trans-splicing (3) because sequences adjacent to coding regions are neither conserved nor do they display landmarks of group I, II, or spliceosomal introns. In addition, module junctions do not occur at known organellar intron insertion points, which are otherwise found in highly conserved gene regions. This suggests an RNA processing mechanism, whose detailed biochemistry remains to be uncovered, for *Diplonema*'s discontinuous mitochondrion-encoded genes.

Lastly, we detected potential evidence for RNA editing (Fig. 1C). A run of six noncoded uridines in the *cox1* transcript between modules 4 and 5 specify amino acids in the deduced Cox1 protein that align unambiguously with homologs in other species. Three lines of evidence support that this insert results from RNA editing rather than a minimodule: (i) gene modules observed in *Diplonema* mitochondria are >10 times longer, (ii) there is no occurrence of six thymidine or adenines in the available ~250-kbp mtDNA sequence, and (iii) uridine-based (insertion and deletion) RNA editing operates in the sister group of diplomonads, the kinetoplastids (7). We posit that RNA editing in *Diplonema* is directed by (yet undetected) guide RNAs but that their primary role is alignment of cognate module transcripts. Thus, we hypothesize that transcript maturation in *Diplonema* mitochondria requires ligase, nuclease, and helicase—activities present in kinetoplastid editosomes (7).

## References and Notes

1. M. W. Gray, B. F. Lang, G. Burger, *Annu. Rev. Genet.* **38**, 477 (2004).
2. M. W. Gray, *IUBMB Life* **55**, 227 (2003).
3. L. Bonen, *FASEB J.* **7**, 40 (1993).
4. A. G. B. Simpson, A. J. Roger, *Mol. Phylogenet. Evol.* **30**, 201 (2004).
5. W. Marande, J. Lukes, G. Burger, *Eukaryot. Cell* **4**, 1137 (2005).
6. Materials and methods are available on Science Online.
7. K. D. Stuart, A. Schnaufer, N. L. Ernst, A. K. Panigrahi, *Trends Biochem. Sci.* **30**, 97 (2005).
8. This work was supported by the Canadian Institutes of Health Research Institute of Genetics (MOP-79309) and the Canadian Institute for Advanced Research. GenBank accession nos. are AY686226 and EU123536 to EU123538.

## Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5849/415/DC1

Materials and Methods

SOM Text

Fig. S1

Table S1

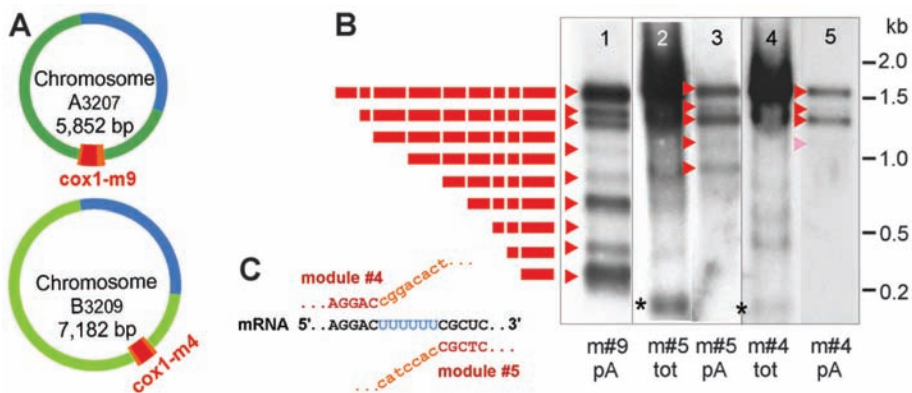
References

19 July 2007; accepted 29 August 2007

10.1126/science.1148033

Department of Biochemistry, Université de Montréal, Montréal, Quebec H3T 1J4, Canada.

\*To whom correspondence should be addressed. E-mail: Gertraud.Burger@UMontreal.ca



**Fig. 1.** (A) Representative mitochondrial chromosomes of *Diplonema*. Green and blue arcs are the constant region; the green portion is identical in chromosomes of class A and B. Cassettes include a gene module (red box) and flanking regions (orange boxes). (B) Northern hybridization with polyadenylated (pA) or total RNA (tot). Probes are *cox1* coding regions, i.e., the C-terminal (m9, lane 1) and two internal modules (m5, lanes 2 and 3; m4, lanes 4 and 5). Triangles point to bands, and asterisks indicate transcripts of individual modules. The presence of bands in lanes 2 and 4 that are not seen in lane 1 indicates ligation of non-poly(A) transcripts and suggests that different pathways of module assembly exist. Rows of red boxes symbolize polyadenylated transcript intermediates. (C) *cox1* mRNA with noncoded uridines (blue) between m4 and m5. Genomic sequences of gene modules and module-flanking regions are indicated in red uppercase and orange lowercase letters, respectively.