

Université de Montréal

Structure et expression des gènes mitochondriaux de *Diplonema papillatum*

par

William Marande

Département de Biochimie,
Faculté de médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph. D.)
en biochimie

Avril, 2007

© William Marande, 2007



W
4
U58
2007
v.109

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.


Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.



Université de Montréal
Faculté des études supérieures


Cette thèse intitulée :

Structure et expression des gènes mitochondriaux de *Diplonema papillatum*

Présentée par :
William Marande

a été évaluée par un jury composé des personnes suivantes :

Pascal Chartrand, président-rapporteur
Gertraud Burger, directeur de recherche
David Morse, membre du jury
Linda Bonen, examinateur externe
....., représentant du doyen de la FES



Résumé

Le génome mitochondrial est habituellement composé d'un seul chromosome. Cependant, des déviations majeures ont été répertoriées, notamment chez les eucaryotes unicellulaires flagellés du groupe des kinétoplastides (phylum Euglenozoa). Le génome des kinétoplastides contient deux types de molécules d'ADN : les maxicercles qui codent pour des gènes mitochondriaux « cryptiques » et les minicercles qui codent pour des ARN guides impliqués dans l'édition des transcrits par insertion et suppression d'uridines.

Concernant les génomes mitochondriaux des autres membres des Euglenozoa, soit les diplonémides et les euglénides, très peu d'information est connue. Mon projet de thèse consistait à analyser le génome mitochondrial de *Diplonema papillatum* qui est un membre du groupe des diplonémides, le groupe frère des kinétoplastides.

L'analyse de *D. papillatum* a mis à jour un génome mitochondrial différent de celui des kinétoplastides. Son ADNmt est composé de centaines de chromosomes circulaires qui font 6 et 7 kbp et qui portent, chacun, un unique fragment de gène (exon) dont la taille varie de 60 à 355 bp. Par exemple, le gène *cox1*, qui code pour la sous-unité 1 de la cytochrome oxidase, est réparti en neuf exons distincts, chacun codé par un chromosome différent. Jusqu'à maintenant, 49 exons correspondants à neuf gènes ont été identifiés via le séquençage de banques génomiques et d'ADN complémentaires. Nous avons montré que l'ARNm est généré par de l'épissage en trans dont le mécanisme est encore inconnu. De plus, nous avons observé un cas d'édition d'ARN, par insertion d'uridines, similaire à celui présent chez les kinétoplastides.

En conclusion, l'analyse de l'ADNmt de *D. papillatum* nous a permis de découvrir un génome mitochondrial inouï. Le mécanisme d'expression des gènes promet de présenter un nouveau genre d'épissage en trans. De plus, nos résultats montrent que le génome mitochondrial multi-chromosomes est un caractère commun aux kinétoplastides et diplonémides.

Mots-clés : génome mitochondrial, épissage en trans, gènes fragmentés, protiste, Euglenozoa, édition d'ARN.

Abstract

Mitochondrial genomes, usually, consist of a single chromosome. However, major deviations have occurred in the unicellular flagellates, the kinetoplastids (Euglenozoa). Their mitochondrial DNA contains two types of DNA molecules: maxicircles bearing mitochondrial encoded genes and minicircles encoding guide RNA that mediate RNA editing via uridine insertion/deletions.

About the mitochondrial genome from others Euglenozoa, little is known. My project was to investigate the mitochondrial genome of *Diplonema papillatum*, a member of the poorly studied diplonemids, the sister group of kinetoplastids.

Our results indicate that the mtDNA of *D. papillatum* is composed of numerous circular chromosomes of 6 and 7 kbp in size, each carrying a single gene fragment of 60-355 bp in size. For example, the *cox1* gene, coding for the subunit 1 of the cytochrome oxidase, occurs in nine separated gene pieces (exons), each encoded on a different chromosome. So far, we have identified 49 exons from a total of nine genes by sequencing genomic and cDNA libraries. We have shown that a contiguous *cox1* transcript is generated via trans-splicing. The mechanism by which trans-splicing proceeds appears to be yet undescribed. We also observed RNA editing similar as in kinetoplastids, but to a much lesser degree.

In conclusion, the analysis of the highly unusual mitochondrial genome and gene expression within *D. papillatum* promises to lead to the discovery of a new type of trans-splicing. In addition, our results show that the multichromosomes mitochondrial genome is common between diplonemids and kinetoplastids.

Keywords : mitochondrial genome, trans-splicing, gene fragmented, protist, Euglenozoa, RNA editing.

Table des matières

Mise en contexte	1
Chapitre 1. Introduction	2
1.1. Le génome mitochondrial	2
1.1.1. Structure de l'ADNmt	2
1.1.2. Contenu en gènes	5
1.1.3. Taille versus contenu en gènes.....	12
1.1.4. Évolution des génomes mitochondriaux	14
1.1.5. Structure de gènes fragmentés	17
1.1.6. Expression des gènes mitochondriaux	24
1.2. Le génome mitochondrial des Euglenozoa	25
1.2.1. Les kinétoplastides	26
1.2.2. Les euglénides	35
1.2.3. Les diplonémides	35
Chapitre 2. Résultats	37
2.1. Article 1 : Structure unique du génome mitochondrial des diplonémides, le groupe frère des kinétoplastides.....	37
2.2. Article 2 : Un seul exon par chromosome chez <i>Diplonema papillatum</i>	49
Chapitre 3. Conclusion	89
3.1. Mécanismes potentiels d'épissage en trans.....	91
3.1.1. Via un intron discontinu.....	91
3.1.2. Via une molécule de type ARNg	91
3.2. ADNmt et expression des gènes au sein des diplonémides	94
3.2.1. ADN mitochondrial.....	94
3.2.2. Expression du gène <i>cox1</i>	94
3.3. Les perspectives	95
3.3.1. Connaissance du contenu en gène.....	95
3.3.2. Cartographie des ARN	95

3.3.3.	Test des hypothèses d'épissage en trans	96
3.3.4.	Vue d'ensemble de l'ADNmt des Euglenozoa	97

Liste des tableaux

Tableau 1: les gènes mitochondriaux et leurs fonctions 7

Liste des figures

Figure 1 : Architecture des génomes mitochondriaux	3
Figure 2: Processi biologiques dans la mitochondrie.....	6
Figure 3 : Arbre phylogénétique schématique montrant la base des animaux, des champignons et des plantes terrestres.	11
Figure 4 : Taille et contenu en gène de génomes mitochondriaux comparés à une α -proteobactérie (<i>Rickettsia</i>).....	13
Figure 5 : Transfert des gènes mitochondriaux vers le noyau	16
Figure 6 : Consensus de la structure secondaire des introns de groupe I mitochondriaux. .	20
Figure 7: Modèle d'une structure secondaire prédite des introns de groupe II conventionnels.....	21
Figure 8 : Mécanisme d'épissage des introns de groupe I et II.....	22
Figure 9 : Microscopie électronique de l'ADNk de <i>C. fasciculata</i>	28
Figure 10 : Images montrant l'ADNk des trypanosomatides et des bodonides.....	29
Figure 11 : Schéma de l'insertion de U dans un ARNm.....	34
Figure 12 : Étapes d'expression des gènes de <i>D. papillatum</i>	90
Figure 13 : Mécanismes d'épissage en trans médié par une molécule de type ARN guide	92
Figure 14 : Parallèle entre l'hypothèse ARNg chez <i>D. papillatum</i> et l'édition d'ARN chez les trypanosomatides.	93

Liste des abréviations

A	adénine
ADN	acide déoxyribonucléique
ADNc	ADN complémentaire
ADNk	ADN kinétoplastique
ADNmt	ADN mitochondrial
ARN	acide ribonucléique
ARNg	ARN guide
ARNm	ARN messenger
ARNt	ARN de transfert
C	cytosine
DAPI	4',6'-Diamidino-2-phenylindole
G	guanine
kpb	kilo paires de bases
MURF	maxicircle unidentified reading frame
NADH	nicotinamide adénine dinucléotide
nt	nucléotide
ORF	open reading frame
pb	paire de base
PCR	polymerase chain reaction
RT-PCR	reverse transcriptase PCR
S	svedberg
T	thymidine
U	uracile
µm	micromètre

*Je dédicace ce mémoire à mon fils, Jonah, et
à ma femme, Emmanuelle.*

Remerciements

Tout d'abord, je tiens à remercier le Dr Gertraud Burger, sans qui ce projet n'aurait pu ni commencer, ni avancer et ni aboutir. Cette collaboration de plus de 5 ans m'aura beaucoup appris sur les nombreux aspects importants pour une carrière scientifique. Gertraud m'a aussi permis d'assister à de nombreux congrès internationaux durant lesquels j'ai rencontré des gens passionnés et passionnant sans qui la recherche n'aurait pas autant d'intérêt. Je pense principalement au Dr Julius Lukes avec qui nous avons étroitement collaboré.

Je tiens à remercier tous les membres présents et passés du laboratoire de Gertraud, de Franz et d'Hervé, avec qui j'ai partagé cette aventure pour le meilleur et pour le pire. Un gros merci à Shona, Lyse et Eliás pour leur aide, leurs conseils et les discussions que nous avons eu. Je n'oublie pas Jung hwa et Yun pour leur gentillesse, leur écoute et pour avoir fait de mes demandes des priorités. Les étudiants ont fait parti de mon quotidien et leurs réflexions, critiques et suggestions m'ont beaucoup apporté autant au niveau scientifique que personnel. Je souhaite dire merci à Naiara pour sa spontanéité, son honnêteté et son intérêt scientifique que j'affectionne particulièrement. Notre laboratoire est multiculturel et ainsi, j'ai pu découvrir, le Québec grâce à Joannie, Jean-françois, Dominic et Nicolas, l'Inde à travers Siva, le Maroc via Rachid, l'Iran avec Pasha et beaucoup d'autres pays encore. Je souhaite bonne chance à Vincent pour la suite du projet sur *Diplonema*.

Je tiens à remercier ma femme avec qui j'ai partagé tous les bons comme les mauvais moments de cette thèse mais surtout avec qui j'ai avancé dans la vie. Nous nous sommes mariés et avons eu notre fils pendant ce temps et ces moments m'ont permis de relativiser et de juger l'importance des évènements qui peuvent arriver au cours d'un doctorat.

Merci au CIEC qui a financé la première année de ma thèse.

Mise en contexte

La mitochondrie est une organelle à double membrane présent dans la majorité des cellules eucaryotes. Cet organite héberge de nombreuses voies métaboliques telles que le cycle de Krebs, le transport d'électron et la phosphorylation oxydative, la synthèse des centres fer-soufre, la biosynthèse des hèmes, pour n'en nommer que quelques unes. De plus, la mitochondrie joue un rôle central dans la mort cellulaire programmée (pour revues, voir Scheffler, 2001; Reichert & Neupert, 2004).

Il est maintenant établi que la mitochondrie a pour origine une α -protéobactérie endosymbiotique. De cet ancêtre, elle a conservé une partie de l'information génétique. Les génomes mitochondriaux des eucaryotes contemporains sont d'une diversité inattendue en taille, en contenu génétique et en mécanismes d'expression (Burger et al., 2003b; Gray et al., 2004).

Un des génomes mitochondriaux les plus remarquable est présent chez les kinétoplastides, des eucaryotes unicellulaires comprenant des agents responsables de maladies humaines telles que les leishmanioses. Dans ces organismes, l'ADNmt est composé de milliers de molécules circulaires formant une structure dense observable au microscope (le kinétoplaste). Les gènes sont dits cryptiques, les transcrits devant subir une modification massive avant d'être traduits.

Les travaux présentés dans cette thèse portent sur le génome mitochondrial de *Diplonema papillatum*, un membre des diplonémides, le groupe frère des kinétoplastides. Ce génome révèle également de nombreux aspects exceptionnels mais différents de ceux décrits pour les kinétoplastides

Ainsi dans l'introduction, je décrirai les génomes mitochondriaux par rapport à leur structure physique, leur taille et leur contenu en gène avec quelques notions sur le contexte évolutif. Je présenterai les connaissances sur les gènes fragmentés et l'expression du génome dans la mitochondrie. Finalement, je parlerai du génome mitochondrial des Euglenozoa qui est le phylum auquel appartient *Diplonema*.

Chapitre 1. Introduction

1.1. Le génome mitochondrial

1.1.1. Structure de l'ADNmt

Il existe principalement trois structures d'ADNmt : réellement circulaire, linéaire concaténé et linéaire monomérique. Il peut y avoir un seul type de chromosome (en plusieurs copies) ou plusieurs (Figure 1).

ADN mitochondrial circulaire ou « circular mapping » ?

Dans les années 60, les premières visualisations de l'ADNmt par microscopie électronique ont été réalisées chez les mammifères. Chez l'humain, 90-95% de l'ADNmt observé en microscopie correspondait à des cercles ayant une taille de 15 à 20 kpb (Borst & Kroon, 1969; Borst, 1972). De plus, chez une grande partie des espèces, l'ADNmt se cartographie comme des cercles par restriction enzymatique et par séquençage. En conséquence, la structure de l'ADN mitochondrial est généralement considérée comme circulaire. Cependant, cette idée changea avec l'arrivée des techniques d'électrophorèse en champs pulsés permettant la séparation de grandes molécules. Les résultats révélaient que la majorité des génomes circulaires étaient, en fait, des génomes linéaires enchaînés les uns à la suite des autres (Bendich, 1996).

Le cas de la levure, *Saccharomyces cerevisiae* illustre bien ce changement de perception. Les analyses de l'ADNmt ont montré que la structure de l'ADN était majoritairement linéaire avec moins de 1% de molécules circulaires (Guerineau *et al.*, 1975). L'ADN linéaire observé était considéré comme un artefact provenant de cercles brisés. Ultérieurement, les résultats d'électrophorèse en champs pulsés et des films de

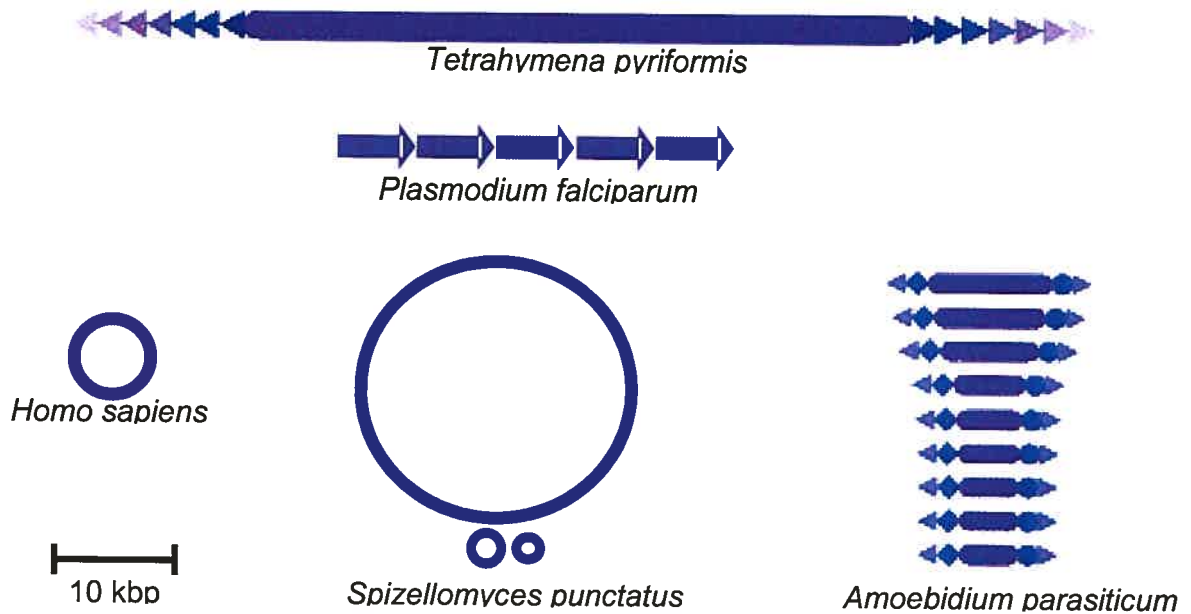


Figure 1 : Architecture des génomes mitochondriaux

Les molécules d'ADN sont principalement circulaires superenroulées chez l'humain (Boore, 1999) et linéaires monomériques chez *T. pyriformis* (Suyama & Miura, 1968). Chez le champignon, *Spizellomyces punctatus* (Burger et al., 2003b), l'ADNmt comprend 3 molécules circulaires, alors que plusieurs centaines de molécules linéaires composent l'ADNmt d'*Amoebidium* (Burger et al., 2003a). L'ADNmt de *Plasmodium* correspond à une succession de chromosomes linéaires en tandem de 6 kbp (Feagin, 1992). Les triangles représentent des répétitions terminales de 32 pb chez *T. pyriformis* et 40 pb chez *A. parasiticum*. Les ronds et les quadrilatères pleins représentent des répétitions sous-terminales, qui sont de 100 pb et 65 pb, respectivement, chez *A. parasiticum*.

déplacement de l'ADNmt ont démontré que les chromosomes de *S. cerevisiae* sont des molécules linéaires de 75 kbp organisées en tandem (Carle & Olson, 1985) (Smith & Bendich, 1990). Les analyses de *Schizosaccharomyces pombe* et *Neurospora crassa* montrent aussi une structure linéaire en tandem (Bendich, 1996).

Nous savons aujourd'hui qu'il existe très peu de génome mitochondriaux avec des chromosomes circulaires tels que les mammifères, les kinétoplastides ou encore *Acanthamoeba castellanii* (Bohnert & Herrmann, 1974).

ADNmt linéaire monomérique

Depuis la découverte de l'ADNmt linéaire du cilié *Tetrahymena pyriformis* (Suyama & Miura, 1968), et ensuite *P. aurelia* (1975), d'autres génomes linéaires monomériques ont été observés chez de nombreuses espèces sans relation phylogénétique telles que les levures *Anseluna mrakii* (Wesolowski & Fukuhara, 1981) et *Candida parapsilosis*, (Kovac *et al.*, 1984), l'algue verte *Chlamydomonas reinhardtii* (Ryan *et al.*, 1978) et certains cnidaires (Boore, 1999) (voir la revue Nosek & Tomaska, 2003). Pour maintenir les extrémités des chromosomes lors des cycles de réplication, chaque espèce arbore une stratégie différente allant de télomères chez *T. pyriformis* à des structures de type « tige-boucle » chez *P. aurelia* ou encore une extrémité simple brin comme chez *C. reinhardtii* (Nosek & Tomaska, 2003).

ADNmt avec des chromosomes multiples

Chez les plantes terrestres, la structure des chromosomes mitochondriaux apparaît plus complexe avec un chromosome maître et d'autres cercles subgénomiques issues de recombinaison entre séquences répétées (Backert *et al.*, 1997). Une exception à ce modèle est *Marchantia polymorpha* qui possède un chromosome linéaire en tandem de 186 kbp qui ne subit pas de recombinaison en raison de l'absence d'éléments répétés (Bendich, 2001).

Alors que l'ADNmt est généralement composé d'un seul type de chromosome, chez certaines espèces le génome arbore de multiples chromosomes. Chez les animaux *Globodera* (Armstrong *et al.*, 2000) et *Dicyema* (Watanabe *et al.*, 1999), le génome est constitué de plusieurs petits chromosomes circulaires (entre 6 et 10 kpb pour *Globodera* et entre 1 et 2 kpb pour *Dicyema*) avec la particularité chez *Dicyema* qu'il n'y a qu'un seul gène par cercle. Chez le champignon *Spizellomyces punctatus* (Burger *et al.*, 2003b), trois chromosomes sont présents, un de 59 kbp et deux de ~1 kbp. Un cas exceptionnel a été identifié chez le protiste *Amoebidium parasiticum* (Burger *et al.*, 2003a). Son ADNmt est formé de plusieurs centaines de chromosomes linéaires différents qui ont une taille allant de 0.3 à 8.3 kpb. Une des questions majeure des génomes composés de nombreuses molécules est comment la ségrégation se fait lors de la division de la mitochondrie puisqu'il n'existe pas de centromère comme dans le noyau.

1.1.2. Contenu en gènes

À travers les eucaryotes, les gènes mitochondriaux sont impliqués dans cinq mécanismes principaux (figure 2). De façon invariable, ils codent pour la respiration et le couplage de la phosphorylation oxydative ainsi que la traduction; moins fréquemment, pour l'importation/maturation des protéines mitochondriales et la maturation des ARN; dans de rares cas, la transcription (voir tableau 1) (Burger *et al.*, 2003b).

Les animaux

L'ensemble des gènes codés par l'ADNmt des espèces animales est relativement constant. Le génome est compact (>90% de région codante), il a une taille de ~16 kbp et code pour la petite et la grande sous unité ARN des ribosomes mitochondriaux, pour 13 protéines

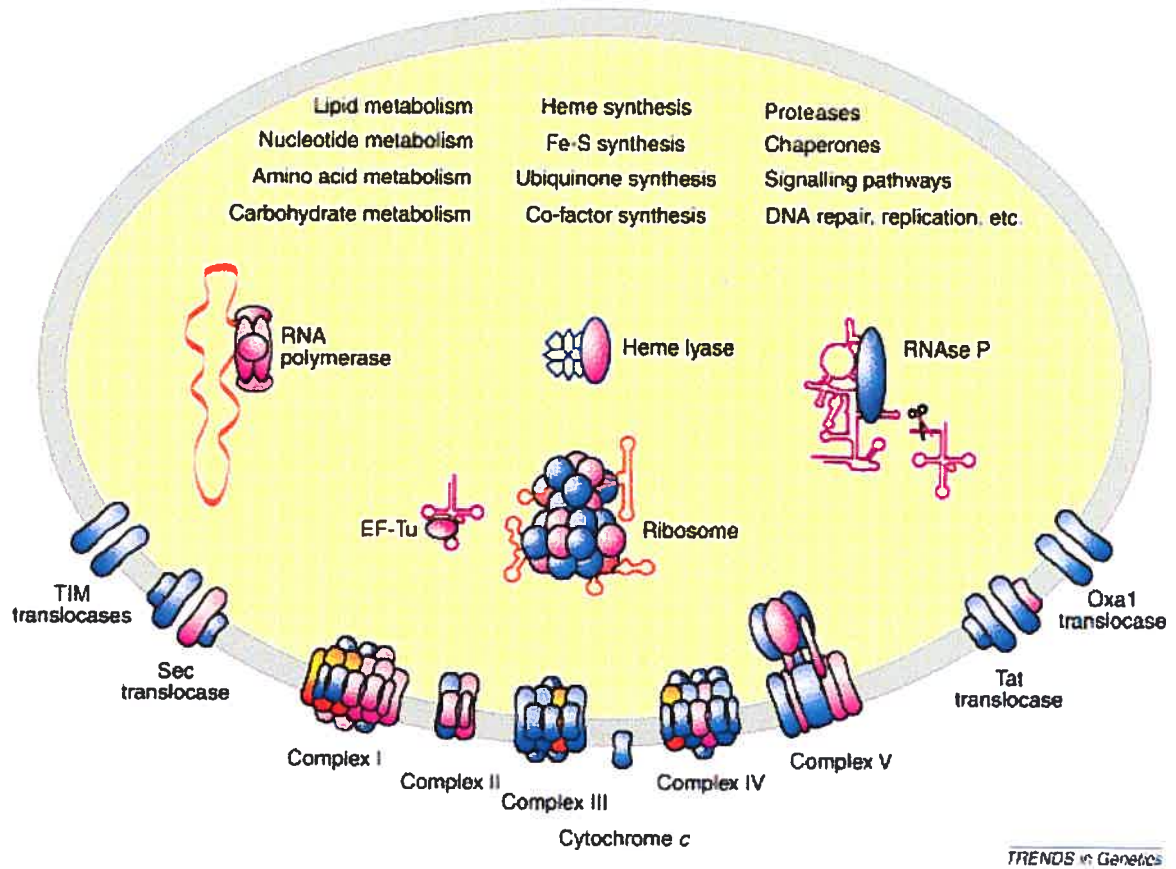


Figure 2: Processi biologici dans la mitochondrie.

La majorité des fonctions mitochondriales est montrée. La plupart des composants mitochondriaux sont codés par le noyau (en bleu), ceux en rose codés par l'ADNmt dans certains eucaryotes alors qu'une petite portion est codée invariablement par l'ADNmt (en orange). Les complexes I-V correspondent à la chaîne de transport d'électrons et à la phosphorylation oxydative. Les TIM-translocases sont impliquées dans l'importation des protéines et leur insertion dans la membrane interne. Tat, Sec et Oxa 1 translocase sont impliquées dans l'exportation de protéines hors de la matrice et leur insertion dans la membrane interne. La RNAse P est un ribozyme qui mature l'extrémité 5' des ARNt.

Cette figure est tirée de (Burger et al., 2003b)

Tableau 1: les gènes mitochondriaux et leurs fonctions

Transport d'électron et synthèse d'ATP	
Complexe I : NADH deshydrogénase	<i>nad 1, 2, 3, 4, 4L, 5, 6, 7, 8, 9, 10, 11</i>
Complexe II : Succinate deshydrogenase	<i>sdh 2, 3, 4</i>
Complexe III : CoQ-cytochrome c oxidase	<i>cob</i>
Complexe IV : Cytochrome oxidase	<i>cox1, 2, 3</i>
Complexe V : ATP synthase	<i>atp1, 6, 8, 9, 3</i>
Traduction	
ARN ribosomiques	<i>rns, rnl, rn5</i>
ARN de transfert	<i>trnA, C, ..., W, Y</i>
Protéines de la petite sous unité ribosomale	<i>rps 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14, 19</i>
Protéines de la grande sous unité ribosomale	<i>rpl 2, 5, 6, 11, 14, 16, 1, 10, 18, 19, 20, 27, 31, 32, 34</i>
Facteur d'élongation	<i>tufA</i>
Transcription	
Corps de l'ARN polymérase	<i>rpo A, B, C</i>
Similarité avec le facteur sigma	<i>rpoD</i>
Import de protéine	
ABC transporteur	<i>ccmB (yej V), ccmA (yej W)</i>
Dégagement de l'hème	<i>ccmC (yejU)</i>
Transporteur de type SecY	<i>secY</i>
Transporteur Sec-indépendant	<i>tatA (mttA), tatC (mttB; ymf16)</i>
Maturation de protéine	
Protéine d'assemblage du cytochrome oxydase	<i>cox 11</i>
Maturation des hèmes c	<i>ccmF (yej R)</i>
Maturation des ARN	
RNAse P (5' des ARNt)	<i>rnpB</i>
Gènes rares d'origine incertaine	
DNA polymérase (dérivé de plasmide)	<i>dpo</i>
RNA polymérase (dérivé de plasmide)	<i>rpo</i>

Transcriptase inverse (dérivé de plasmide)	<i>rtl</i>
Réparation de l'ADN	<i>mutS</i>
Transférerase de méthyl	<i>dam</i>

En gras sont représentés les gènes qui se trouvent uniquement chez les jakobides et les malawimonades. Ce tableau est issue de (Gray *et al.*, 2004)

(*cox1-3, cob, nad1,2,3,4, 4L, 5, 6* et *atp6, 8, 9*) et jusqu'à 22 ARNt (Boore, 1999). Toutefois, des exceptions se retrouvent chez les anémones de mer qui n'ont que 2 ARNt (Beagley *et al.*, 1998), comme les chaetognathes (des vers sagittaires) qui possèdent le plus petit génome mitochondrial (11.5 kbp), dans lesquels il manque les gènes codants pour les sous-unités ATP6 et 8 et 21/22 ARNt (Helfenbein *et al.*, 2004) ou encore chez les placozoaires (organisme marin simple, plusieurs milliers de cellules) qui possèdent le plus grand génome mitochondrial séquencé (40 kbp) avec un pourcentage d'ADN non codant très élevé (45%) pour un animal (Dellaporta *et al.*, 2006).

Les champignons (Fungi)

Chez les espèces des quatre groupes majeurs des champignons (ascomycètes, basidiomycètes, zygomycètes et chytridiomycètes), de grandes variations persistent au sein du clade même si le jeu de gènes ressemble à celui des animaux. Chez la plupart des chytridiomycètes, l'ensemble des ARNt est de 7-9. De plus, les tailles varient énormément allant de 19 kbp chez *S. pombe* à 100 kbp chez *Podospora anserina* dont 75 % du génome est attribuable à des introns (Cummings *et al.*, 1990; Bullerwell *et al.*, 2003b). Chez deux groupes d'ascomycètes (les schizosaccharomycétales et les saccharomycétales), les sept gènes du complexe I sont absents alors qu'ils sont présents chez la plupart des champignons (Foury *et al.*, 1998; Bullerwell *et al.*, 2003a).

Les plantes

La taille de l'ADNmt des plantes est largement supérieure à celle des animaux ou des champignons (> 100 fois pour certains cucurbitacés). Cependant, le nombre de gènes (69 gènes chez *Marchantia*) est à peine deux fois plus grands que celui des animaux (37 gènes chez les mammifères). Les gènes supplémentaires sont impliqués dans la traduction (les protéines rps et rpl), dans la respiration et dans l'importation de protéines. L'augmentation de la taille des génomes est la conséquence d'une accumulation d'introns

(32 chez *Marchantia*), de séquences provenant du plastide et du noyau ainsi que des éléments transposables (Notsu *et al.*, 2002). Il est important de noter que la majorité du génome est d'origine inconnue.

Les protistes

Les protistes comprennent tous les eucaryotes, à l'exception des animaux, des plantes et des champignons. Ils incluent aussi les ancêtres unicellulaires de ces trois derniers groupes (figure 4). Le séquençage systématique de génomes mitochondriaux de protistes, en particulier par le consortium OGMP (Organelle Genome Megasequencing Program) a révélé une grande diversité dans les contenus en gène ainsi que dans l'organisation du génome mitochondrial.

Les protistes sont plus hétérogènes que les animaux, les champignons et les plantes réunis tel que décrit dans la revue « Mitochondria of protists » (Gray *et al.*, 2004). En résumant, les ADNmt des protistes ont une taille moyenne de 50 kpb qui peut varier de 6 kbp chez les apicomplexes à >200 kbp chez *A. parasiticum*. Le contenu en gène varie de cinq chez les apicomplexes à 100 chez les jakobides. Le génome de 69 kpb de *Reclinomonas americana* (un jakobide) est exceptionnel par son contenu. Il comprend le plus grand jeu de gènes incluant la sous unité ribosomique 5S, la composante ARN de la RNase P, et au moins 18 protéines inédites au niveau mitochondrial (tableau 1). Le plus surprenant est la présence de quatre gènes codants pour la sous unité d'une polymérase de type eubactérien. La découverte de ce « génome bactérien en miniature » (Lang *et al.*, 1997) a permis de confirmer, d'une façon convaincante, l'origine unique de la mitochondrie tout en offrant une vision du génome ancestral de la mitochondrie. D'autres études intéressantes sont celles des protistes qui divergent à la base des animaux, des plantes et des champignons (figure 3). Par exemple, *Monosiga brevicollis*, un choanoflagellé, est à la base des animaux, or il possède un génome riche en gène (55 gènes) et a une taille quatre fois supérieurs à celle des animaux (76 kbp) (Burger *et al.*, 2003a). Il représente probablement

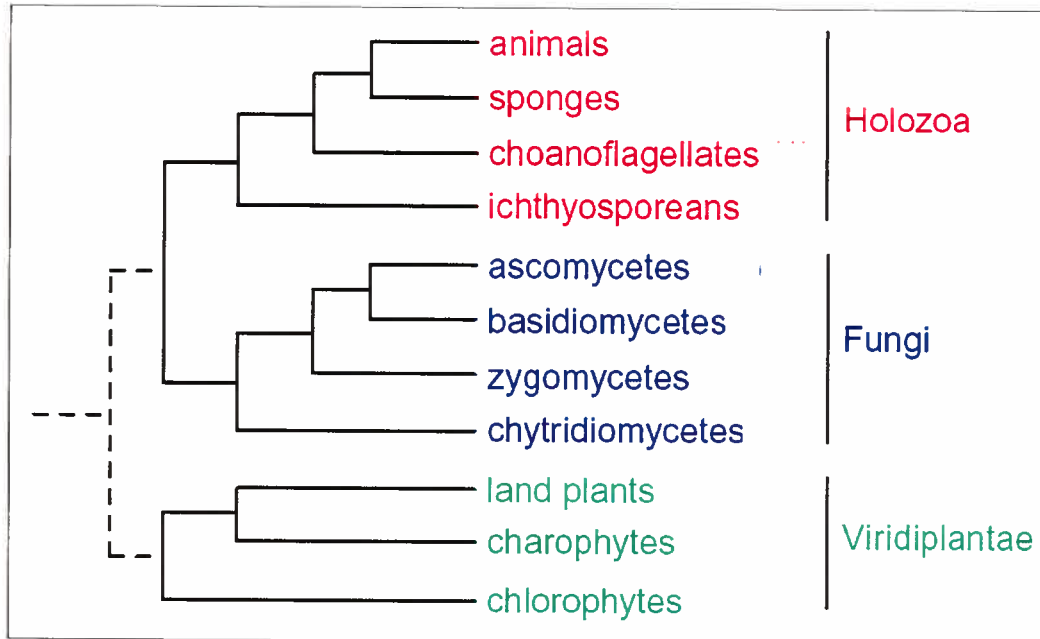


Figure 3 : Arbre phylogénétique schématique montrant la base des animaux, des champignons et des plantes terrestres.

Représentation des Holozoa (en rouge), des Fungi (en bleu) et des Viridiplantae (en vert). Les groupes de protistes sont les choanoflagellés, les ichtyosporés, les chytridiomycétes, les charophytes et les chlorophytes. Cette figure est modifiée de (Bullerwell & Gray, 2004).

l'état ancestral du génome mitochondrial des animaux. La réduction du génome chez les animaux a dû avoir lieu après la séparation entre les choanoflagellés et les animaux.

1.1.3. Taille versus contenu en gènes

Le premier génome mitochondrial séquencé était celui de l'humain en 1981. Maintenant, il y a ~1100 séquences complètes de génomes mitochondriaux publiées (http://megasun.bch.umontreal.ca/ogmp/projects/other/mt_list.html). La grande majorité provient des animaux. Néanmoins, presque tous les grands clades eucaryotes sont représentés. Toutes ces données démontrent une grande gamme de tailles et de capacité codante.

Les génomes mitochondriaux sont majoritairement <50 kpb: ceci représente à peu près 1% de la taille des génomes bactériens dont l'ADNmt est dérivé. Avec seulement 6 kpb, le génome des apicomplexes est le plus petit connu. Il est ~ 100 fois plus petit que celui du maïs qui est le plus grand séquencé (569,630 kpb) (Clifton *et al.*, 2004) et >300 fois plus petit que certains ADNmt de cucurbitacées (~ 2000 kpb) (Ward *et al.*, 1981).

Tout comme la taille, le nombre de gènes codés par la mitochondrie varie grandement, allant de presque 100 chez les jakobides à seulement cinq chez *Plasmodium*. Exception faite des petits génomes qui ont un rapport strict taille/gène, il n'y a pas de corrélation entre l'augmentation de la taille et son contenu en gène (figure 4).

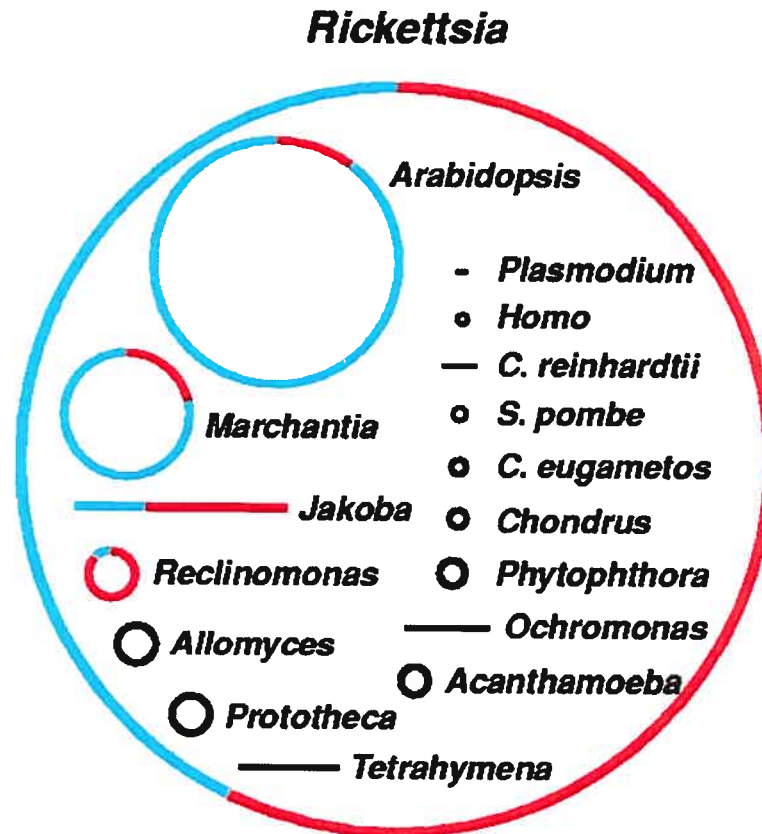


Figure 4 : Taille et contenu en gène de génomes mitochondriaux comparés à une α -proteobactérie (*Rickettsia*).

Les cercles et les lignes représentent, respectivement, la structure circulaire ou linéaire des génomes. Pour les génomes >60 kbp, l'ADN codant pour des gènes de fonction connue (rouge) est distingué de l'ADN incluant les régions intergéniques et les ORFs sans annotation de fonction (vert). Cette figure est tirée de (Lang *et al.*, 1999).

1.1.4. Évolution des génomes mitochondriaux

Si l'on compare la capacité de codage des génomes d'organelles à celle des bactéries avec lesquelles ils partagent leur ancêtre, nous constatons une forte réduction de la quantité de l'information chez les organelles : ainsi le génome de *Rickettsia prowazekii* (le génome eubactérien actuel le plus proche phylogénétiquement des mitochondries) comporte 834 gènes codant des protéines (Andersson *et al.*, 1998; Gray, 1998), tandis que le génome mitochondrial avec le plus grand nombre de gènes, *R. americana*, n'en comporte que 67.

Trois raisons différentes expliquent cette réduction (Lang *et al.*, 1999; Adams & Palmer, 2003; Bonen, 2006). La première correspond à l'élimination de gènes superflus. En effet, la transformation de l'endosymbiote ancêtre en organelle correspond à un changement de niche écologique accompagné d'une spécialisation. Par conséquent, un certain nombre de gènes de l'endosymbiote sont devenus inutiles et ont été progressivement éliminés du génome. C'est le cas pour les gènes impliqués dans le mouvement ou dans la synthèse de la paroi bactérienne. À partir du moment où l'ancêtre bactérien est passé du stade d'organisme libre à celui d'endosymbiote, il s'est retrouvé dans un milieu riche en métabolites intermédiaires issus de l'hôte, rendant ainsi obsolètes les gènes nécessaires à leur synthèse (Itoh *et al.*, 2002). Aucun génome mitochondrial connu ne comporte de gènes codant pour la biosynthèse des nucléotides, lipides ou acides aminés.

Une deuxième raison expliquant la perte de gènes mitochondriaux est le remplacement de leur fonction par un gène nucléaire comme c'est le cas des ARNt mitochondriaux. Chez *T. pyriformis* et *P. aurelia*, l'ADNmt ne code, respectivement, que pour trois et sept ARNt. Pour assurer la traduction de tous les codons des gènes mitochondriaux, les ARNt manquants sont importés du cytoplasme (Chiu *et al.*, 1975; Suyama, 1986).

La troisième explication est le transfert de gènes de l'organelle vers le noyau, une route prise par les gènes de la chaîne de transfert d'électron ainsi que ceux de la traduction. Ce processus est toujours en cours comme l'indique le cas du gène *rpl2* chez les plantes (figure 5). Chez le riz, le gène est entier et codé par l'ADNmt alors que chez le maïs, la copie entière se trouve dans le noyau. Des étapes intermédiaires à ce processus existe puisque chez l'ancêtre commun des plantes eudicotes, l'extrémité 3' du gène a été transférée dans le noyau alors que l'extrémité 5' est resté dans l'ADNmt. Pour finir ce schéma, chez la laitue, l'extrémité 5' a été transférée dans le noyau qui possède ainsi les deux fragments du gène (voir la revue Adams & Palmer, 2003). Le gène *cox2* chez les algues vertes (décrit dans la partie 'Structure de gène fragmenté sans intron' et dans la figure 5) est un exemple similaire de transfert de gène vers le noyau.

Les mitochondries sans génome.

Chez de nombreux parasites ou organismes anaérobiques, la respiration n'est plus essentielle. La perte de cette fonction a entraîné une réduction significative de la mitochondrie et une transformation en hydrogénosome ou mitosome. Les hydrogénosomes jouent un rôle important dans la production d'énergie de la cellule. Les mitosomes semblent avoir retenu uniquement l'assemblage des clusters Fe-S. Très récemment, l'existence d'un génome dans un hydrogénosome a été démontré (Boxma *et al.*, 2005). Ce génome comporte au moins sept gènes codants pour des sous-unités du complexe I (NADH déshydrogénase, *nad2*, *4L*, *5*, *7*), pour des protéines ribosomiques (*rpl2* et *rpl14*) et la petite sous-unité ribosomique mitochondriale. Cette découverte a permis de faire le lien évolutif entre la mitochondrie et les hydrogénosomes sans génome et de confirmer que les hydrogénosomes proviennent de la mitochondrie.

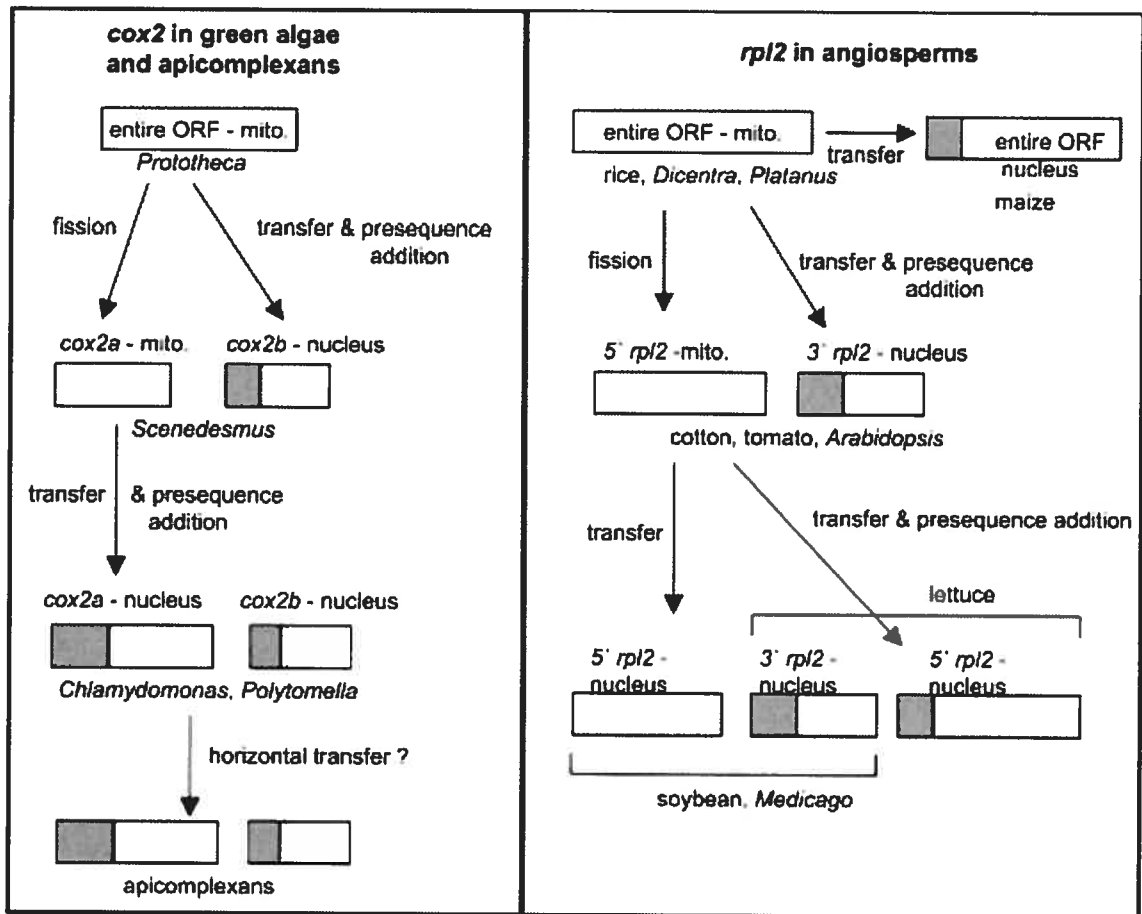


Figure 5 : Transfert des gènes mitochondriaux vers le noyau

Schéma tiré de l'article (Adams & Palmer, 2003)

1.1.5. Structure de gènes fragmentés

... avec intron

L'interruption des gènes par les introns est un phénomène très répandu dans la mitochondrie. Les introns possèdent une structure secondaire et tertiaire qui permet de juxtaposer les séquences exoniques éloignées. Dans la mitochondrie, il existe deux familles d'introns: ceux de groupe I et ceux de groupe II. Ils diffèrent l'un de l'autre par leur structure secondaire, leur mécanisme d'excision et leur manière de propagation.

Introns de groupe I

Découverts dans les années 80, les introns de groupes I ont été les premiers ribozymes (ARN catalytiques) décrits (Kruger *et al.*, 1982; Michel *et al.*, 1982; Waring *et al.*, 1982). En effet, *in vitro*, les introns n'ont besoin que de $MgCl_2$ et de guanosine comme cofacteur pour s'auto-épisser. Toutefois, il est important de signaler que tous les introns de groupe I ne sont pas autocatalytiques *in vitro* (Lang *et al.*, 2007). Ils sont abondants dans le noyau et les organelles des eucaryotes mais plutôt rares dans les génomes bactériens (Haugen *et al.*, 2005). La structure secondaire des introns de groupe I est composée de 10 éléments appariés (nommés P1 à P10) (figure 6) qui sont organisés en 3 domaines dans la structure tertiaire (Adams *et al.*, 2004). Le mécanisme d'excision se fait en deux étapes de trans-estérifications (figure 8). La première étape est le clivage du site 5' d'épissage via l'attaque d'une guanosine exogène. La deuxième étape correspond à l'attaque du site 3' d'épissage par l'extrémité 3' de l'exon libéré précédemment qui permet l'excision de l'intron et la ligation des 2 exons. La taille est petite chez les introns de groupe I nucléaires (250-500 nt) mais elle varie énormément pour les introns des mitochondries (142 à >3000 nt). Les plus grands introns possèdent une ORF codant pour une protéine qui soit assiste l'épissage (les maturases), soit permet la dispersion de l'intron dans le génome via le mécanisme de « homing » (voir la revue Stoddard, 2005).

Introns de groupe II

Généralement, les introns de groupe II ne possèdent pas d'activité autocatalytique. Ils sont présents dans le génome des organelles, des bactéries et des archaebactéries. On les retrouve peu chez les champignons et les algues mais ils sont présents dans les génomes mitochondriaux des plantes terrestres. Le chloroplaste de *Euglena gracilis* représente un cas exceptionnel avec 91 introns de groupe II dans son génome (Hallick *et al.*, 1993). La structure des introns est composée de six tige-boucles nommées domaine 1 à 6 (figure 7). Le mécanisme d'excision se fait aussi en deux étapes de trans-estérifications (figure 8). La première étape est le clivage du site 5' d'épissage via l'attaque nucléophile d'une adénosine non appariée, située proche de l'extrémité 3' de l'intron. La deuxième étape correspond à l'attaque du site 3' d'épissage par l'extrémité 3' de l'exon libéré précédemment qui permet l'excision de l'intron, sous la forme d'un lasso typique des introns de groupe II et la ligation des 2 exons. Toutefois, il existe d'autres mécanismes d'épissage tel que l'excision d'intron sous forme de cercle ou linéaire (Bonen & Vogel, 2001; Murray *et al.*, 2001; Li-Pook-Than & Bonen, 2006). Les introns de groupes II sont des éléments mobiles et leur méthode de propagation se nomme «*retrohoming*». Ce mécanisme complexe fait intervenir le clivage d'ADN suivi d'un épissage inverse qui donne lieu à un hybride ADN/ARN et qui se finit par la transcription inverse de cet hybride via une transcriptase inverse. Cette dernière enzyme est majoritairement codée par l'intron lui-même (Bonen & Vogel, 2001; Lambowitz & Zimmerly, 2004).

Introns de groupe II discontinus : épissage en trans

Dans certains cas, les introns de groupe II sont discontinus avec au moins deux segments placés à des endroits distants du génome. Les transcrits de ces segments se réorganisent en une structure complète, résultant à l'épissage en *trans* de ces morceaux d'intron. Par exemple, chez les plantes terrestres, il existe plusieurs cas où les introns de

groupes II sont discontinus. Chez le blé, le gène *nad1* est fragmenté en cinq morceaux totalisant 40 kbp et il nécessite 3 étapes d'épissage en trans et une étape en *cis* (Chapdelaine & Bonen, 1991). Malek *et al* ont montré que, chez les plantes, les introns discontinus des gènes du complexe I (*nad1*, 2 et 5) étaient le résultat de la fragmentation des introns de groupe II continus (Malek *et al.*, 1997; Malek & Knoop, 1998).

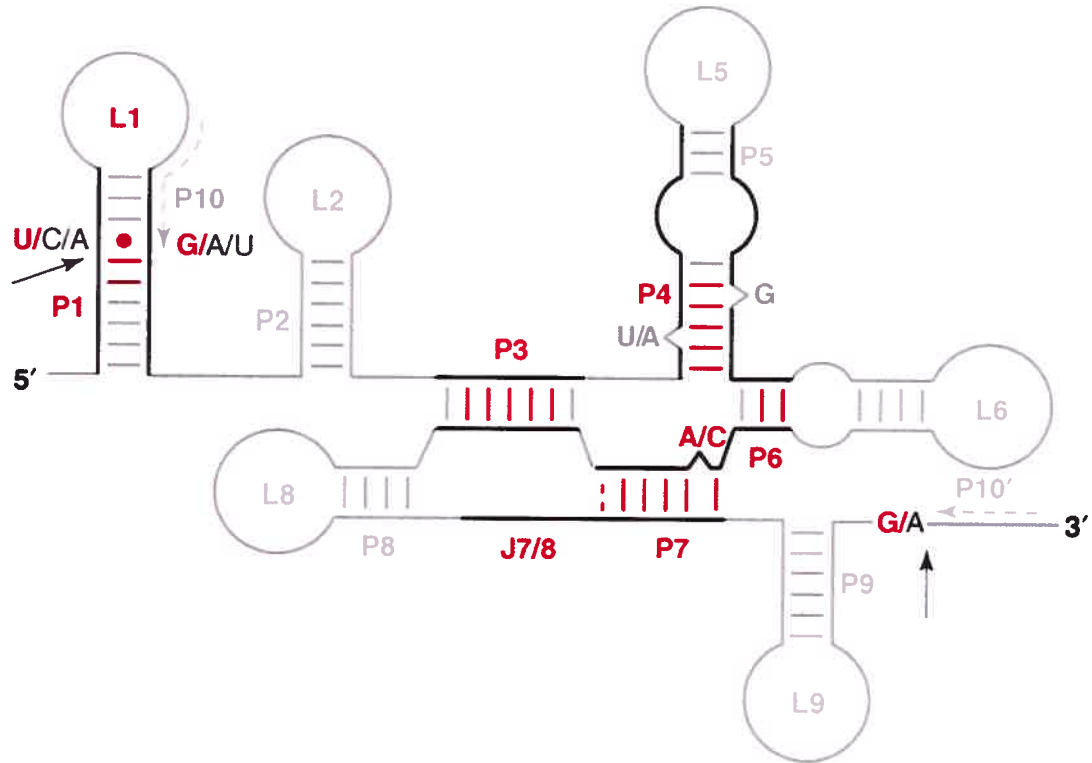


Figure 6 : Consensus de la structure secondaire des introns de groupe I mitochondriaux.

Les 10 éléments appariés sont numérotés P1 à P10.

Les lettres P, L et J signifient 'Pairing', 'Loop' et 'joining'. Le rapprochement des exons se fait via l'interaction entre le P10 et le P10'.

Cette figure est tirée de (Lang et al., 2007).

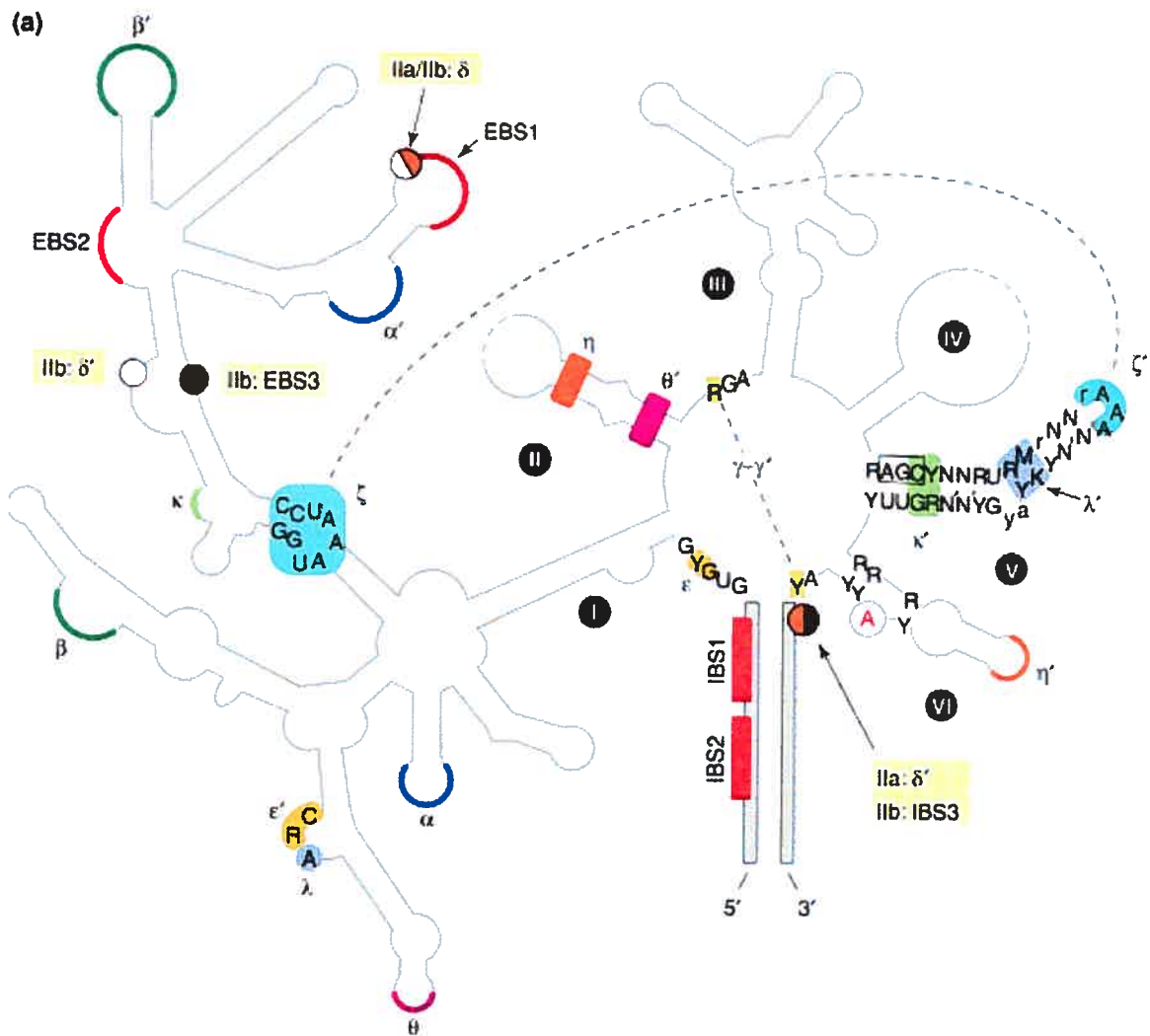


Figure 7: Modèle d'une structure secondaire prédite des introns de groupe II conventionnels.

Les 6 tiges-boucles sont représentés dans des ronds noirs numéroté de I à VI.

Les domaines EBS (Exon Binding Site) et IBS (Intron Binding Site) interagissent et permettent le rapprochement des exons.

Cette figure est tirée de (Bonen & Vogel, 2001)

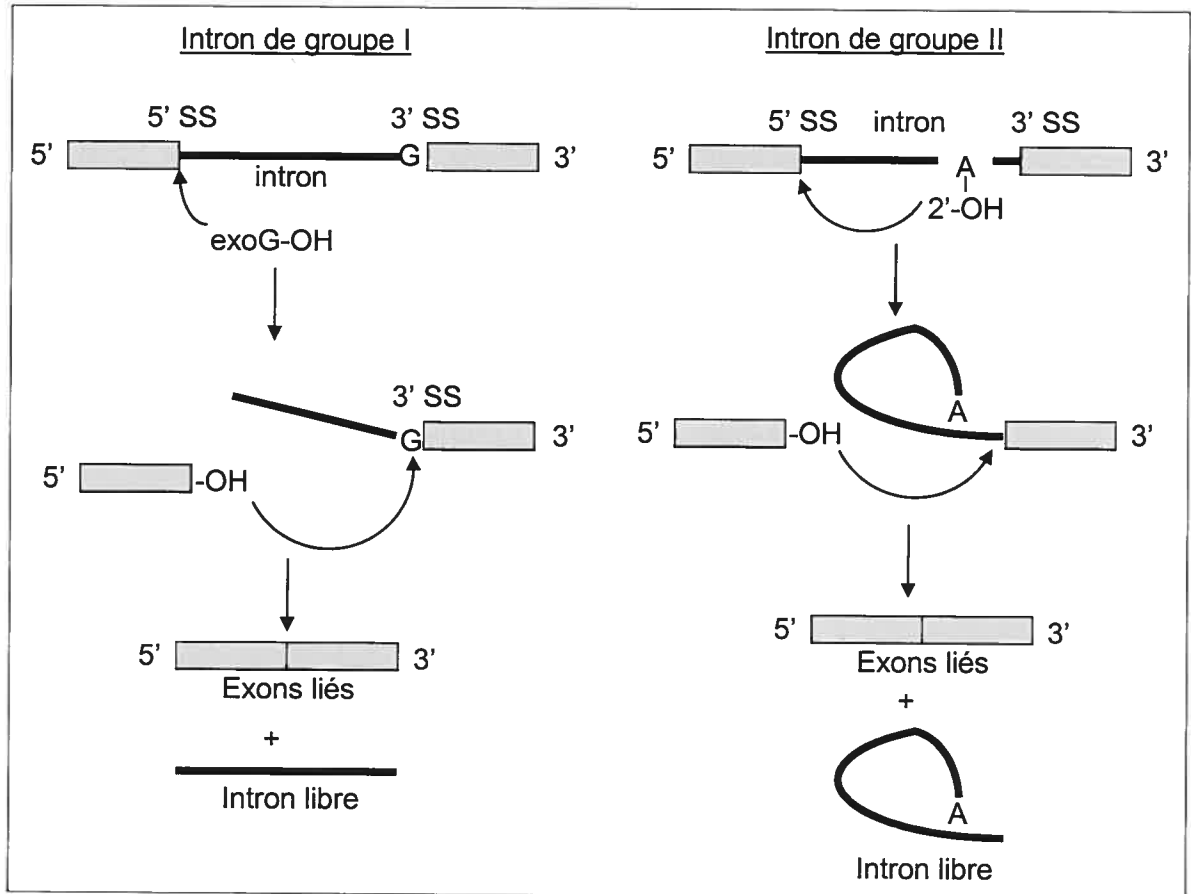


Figure 8 : Mécanisme d'épissage des introns de groupe I et II

... sans intron

Il existe plusieurs cas de fragmentation de gènes n'impliquant pas les introns.

En 1988, Boer et Gray ont découvert que les ARN ribosomiques mitochondriaux de *C. reinhardtii* étaient fragmentés en de nombreux morceaux dispersés de façon désordonnée tout au long du génome. L'ARNr de la petite sous unité est composé de quatre fragments et la grande huit (Boer & Gray, 1988). Ces petits morceaux d'ARN ne sont pas liés covalamment mais plutôt appariés (pont hydrogène) pour former un ribosome fonctionnel (Denovan-Wright & Lee, 1995). Une fragmentation complexe des ARNr mitochondriaux est aussi présente chez les apicomplexes (Vaidya *et al.*, 1989; Feagin *et al.*, 1997) et les dinoflagellés (Kamikawa *et al.*, 2007).

Des gènes en morceaux codant des protéines sont rares dans la mitochondrie. Le gène *nad1* chez *T. pyriformis* et chez *P. aurelia* est séparé en deux parties qui sont transcrits indépendamment sans être liés au niveau ARN. Il est supposé qu'il en résulte deux polypeptides mais la preuve formelle manque (Burger *et al.*, 2003b).

Un cas intrigant d'un gène fragmenté est retrouvé chez les algues vertes. Le génome mitochondrial de *C. reinhardtii* est l'un des seuls à ne pas posséder le gène *cox2*. En fait, le gène *cox2* de *C. reinhardtii* est fragmenté en deux dans l'ADN nucléaire (Perez-Martinez *et al.*, 2001). Chez une algue qui est proche de *C. reinhardtii*, *Scenedesmus obliquus*, son génome mitochondrial ne révèle qu'une moitié du gène *cox2* (Kuck *et al.*, 2000; Nedelcu *et al.*, 2000). Ainsi, on peut penser que la deuxième partie du gène *cox2* de *S. obliquus* se trouve dans le noyau (figure 5).

1.1.6. Expression des gènes mitochondriaux

Les jakobides sont les seuls à posséder, dans leur génome mitochondrial, le complexe de l'ARN polymérase de type bactérienne $\alpha_2\beta\beta'\sigma$ (rpoA-D) mais aucun promoteur de type bactérien n'a pu être identifié (Lang *et al.*, 1997). Chez tous les autres eucaryotes connus, la transcription mitochondriale est effectuée par une polymérase de type phagique T3/T7 codée dans le noyau.

Chez les animaux, la transcription des gènes mitochondriaux est initiée à partir d'une région promotrice située dans la région non-codante du génome résultant en deux précurseurs polycistroniques, un correspondant au brin lourd et l'autre au brin léger. La maturation en monocistrons se fait par l'excision des ARNt qui sont placés après chacun des gènes codant pour des protéines de l'ADNmt. Ce mode de transcription se nomme « ponctuation par ARN de transfert ». Chez les champignons, la transcription est dépendante de l'espèce comme chez *S. pombe* dans laquelle il y a 3 transcrits polycistroniques ou encore *S. cerevisiae* avec 19 sites de transcription donnant des transcrits polycistroniques contenant au moins 2 séquences codantes (Schafer, 2005). Chez les plantes, plusieurs promoteurs peuvent initier la transcription d'un même gène (Kuhn *et al.*, 2005).

Les transcrits mitochondriaux subissent des modifications post-transcriptionnelles qui altèrent leur séquence de différentes manières (voir la revue Gott & Emeson, 2000). Chez les trypanosomes, l'édition consiste à insérer et supprimer des centaines de U du transcrit (le mécanisme est décrit dans la section « Génome mitochondrial des kinétoplastides »). Dans la mitochondrie des plantes, l'édition est caractérisée par une conversion de C en U et occasionnellement de U en C. Il y a 427 sites d'éditions chez *Arabidopsis thaliana* (Giege & Brennicke, 1999), 427 chez le colza (Handa, 2003) et 491 chez le riz (Notsu *et al.*, 2002). Une séquence de vingt nucléotides en aval des sites, et aussi quelques nucléotides en amont, sont impliqués dans l'édition de l'ARN. Récemment, dans

le chloroplaste d'*Arabidopsis thaliana*, dont les transcrits subissent le même type d'édition, il a été montré que la protéine CRR4 se fixe sur une séquence 25 nucléotides en aval du site d'édition et permet la modification (conversion de C en U) de façon spécifique (Okuda et al., 2006). Cette protéine appartient à la famille des protéines PPR qui est composée de plus de 400 membres chez les plantes et qui sont impliquées dans plusieurs fonctions de maturations des ARN dans la mitochondrie et le chloroplaste. Par extrapolation, il est fort probable que certains membres de cette famille soient spécifiques des sites d'édition de la mitochondrie (Shikanai, 2006). Une fois ce facteur fixé sur la séquence, il recruterait la machinerie de conversion qui impliquerait une déaminase qui ferait la conversion de C en U. Cependant, les cofacteurs effectuant la conversion sont toujours inconnus (Brennicke et al., 1999). D'autres phénomènes d'éditions d'ARN ont été observés dans la mitochondrie de *Physarum polycephalum*, où les ARNt, les ARNr et les ARNm subissent principalement des insertions de C mais aussi de U et de dinucléotides (Mahendran et al., 1991). Chez les dinoflagellés, les ARNm sont modifiés par des substitutions de nucléotides (Lin et al., 2002).

1.2. Le génome mitochondrial des Euglenozoa

Le phylum des Euglenozoa (Cavalier-Smith, 1981) est composé de trois groupes majeurs d'unicellulaires flagellés, les kinétoplastides, les euglénides et les diplonémides (Simpson, 1997; Vickerman, 2000). Plusieurs phylogénies moléculaires utilisant la petite sous unité ARN ribosomique confirment la monophylie des Euglenozoa (Maslov et al., 1999; Preisfeld et al., 2001) sans résoudre les relations internes du phylum dû à un jeu de données pauvre en signal (Moreira et al., 2001; Busse & Preisfeld, 2002). La phylogénie utilisant les gènes *hsp70* et *hsp90* montrent un meilleur support statistique avec les euglénides à la base des Euglenozoa suivi de la division des diplonémides et des kinétoplastides (Simpson & Roger, 2004).

De nombreuses excentricités ont été découvertes dans ce phylum. Par exemple, le glycosome est une organelle spécifique aux trypanosomatides dans laquelle s'effectue la glycolyse de la cellule; les introns dans les introns (les « twintrons ») abondent dans le génome chloroplastique d'*Euglena gracilis*; l'épissage en trans nucléaire (« spliced leader ») a été observé chez tous les Euglenozoa; l'édition des ARNm mitochondriaux est massive chez les kinétoplastides. Au niveau de l'ADN mitochondrial, nos connaissances sont restreintes aux kinétoplastides; très peu est connu des euglénides et des diploméides.

1.2.1. Les kinétoplastides

Les kinétoplastides incluent les trypanosomatides qui sont des parasites responsables de maladies humaines telles que la maladie du sommeil (*Trypanosoma brucei*), la maladie de Chagas (*Trypanosoma cruzi*) et les leishmanioses (*Leishmania*). Les membres non-parasites des kinétoplastides sont référencés en tant que bodonides. L'ADNmt compact des trypanosomatides, aussi nommé ADN kinétoplastique est l'une des caractéristiques les plus extravagantes de ces parasites. L'ADNmt forme un gigantesque réseau de milliers de molécules d'ADN interconnectées. Ce réseau contient deux types de molécules circulaires : les maxicercles présents en douzaine de copies et les minicercles présents en plusieurs milliers de copies (figure 9). Les premières indications de son existence datent de la fin du XIX^{ème} siècle et la remarquable complexité moléculaire fut révélée par la microscopie électronique dans la fin des années 60, début des années 70 (Riou & Delain, 1969; Laurent & Steinert, 1970; Renger & Wolstenholme, 1970; Simpson & Da Silva, 1971).

Structure de l'ADNmt : le kinétoplaste

La grande majorité des connaissances sur l'ADNk est dérivée de l'étude du petit groupe des trypanosomatides donnant l'impression que les réseaux étaient tous identiques avec seulement quelques petites variations. Le réseau nommé eukinétoplaste est le mieux

caractérisé. Celui de *Crithidia fasciculata* contient 5000 minicercles (2.5 kbp) et 25 maxicercles (~37 kbp) et chaque minicercle est connecté à 3 voisins créant ainsi une vraie 'cotte de maille'. Les minicercles ont une conformation relâchée (non superenroulée). Chez les trypanosomatides, le réseau est compact et situé à la base du flagelle auquel il est physiquement lié. Chez les bodonides, les minicercles ne sont pas inter-liés. Malgré ceci, le pro-ADNk de *Bodo saltans* ressemble en microscopie électronique et confocal (après la coloration de l'ADN avec du DAPI) à celui des trypanosomatides. Chez *Dimastigella trypaniformis*, *Dimastigella mimosa* (Stolba *et al.*, 2001) et *Cruzella marina*

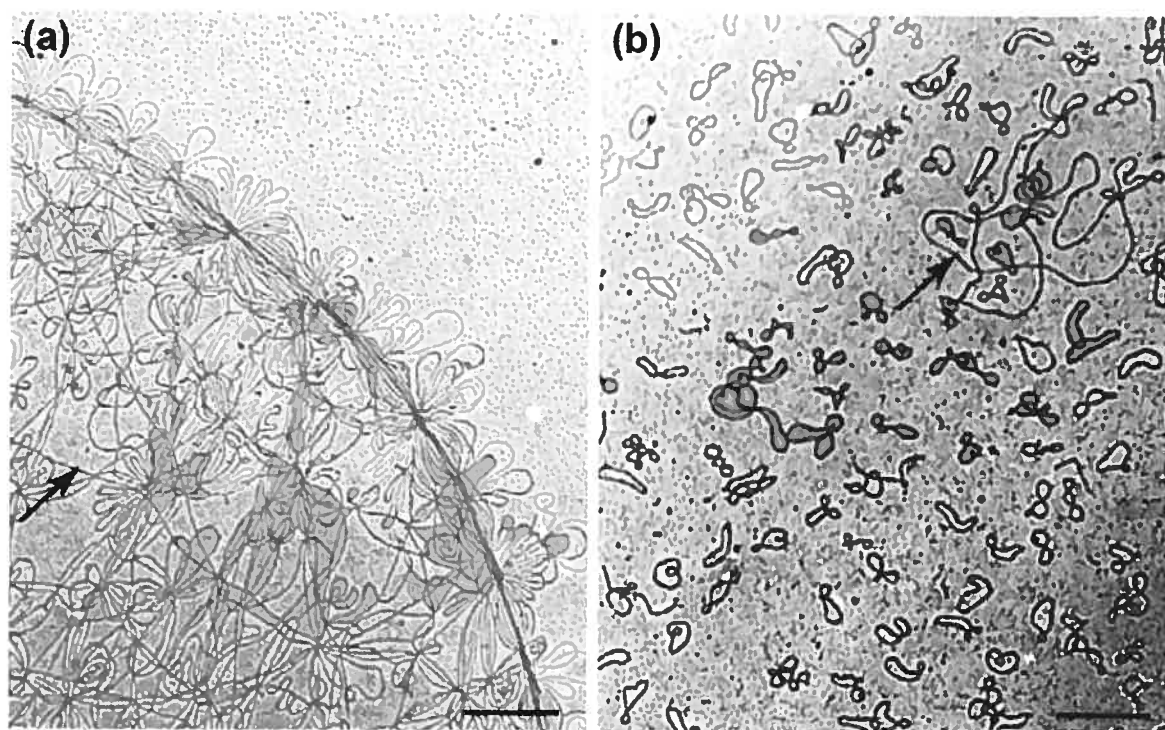


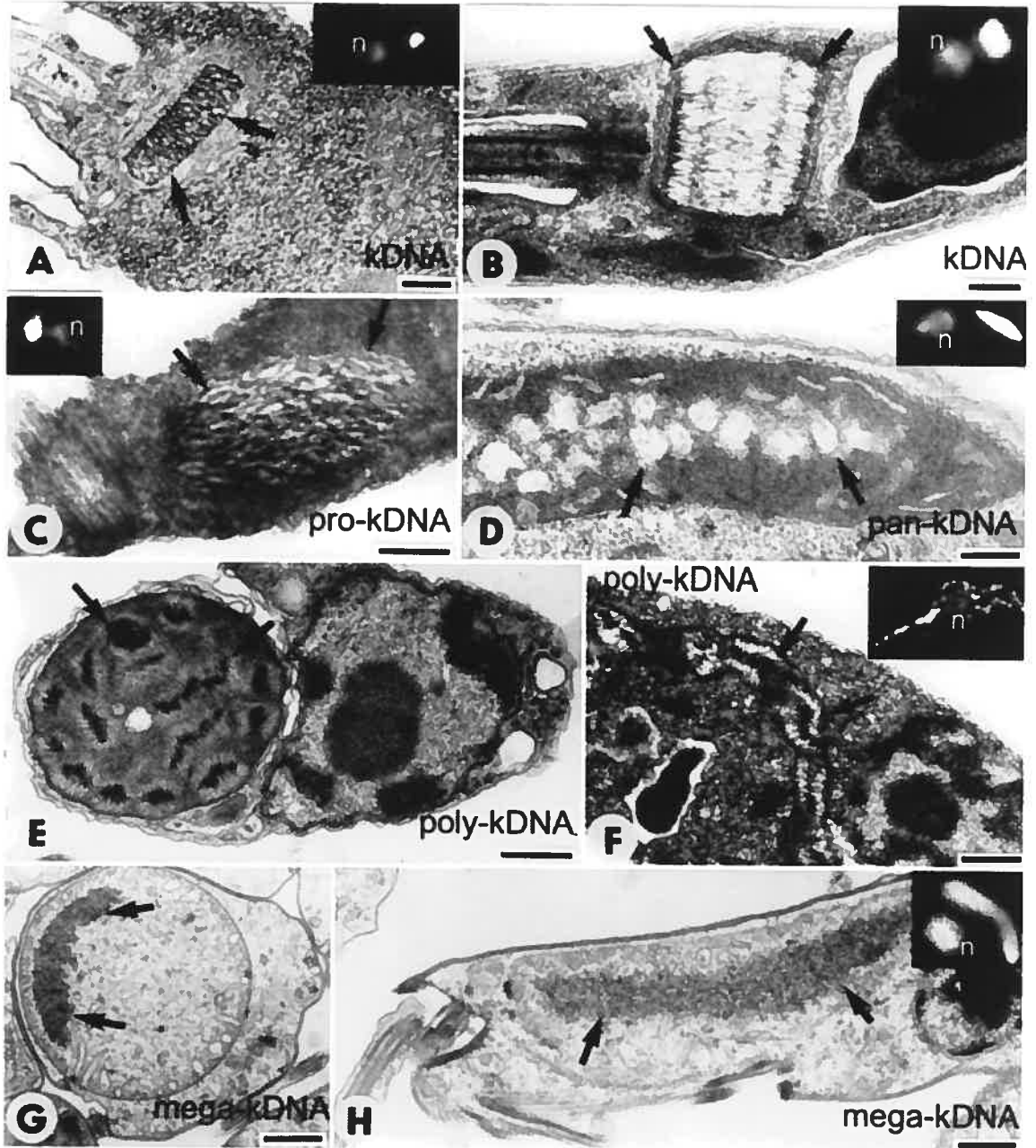
Figure 9 : Microscopie électronique de l'ADNk de *C. fasciculata*.

(a) Segment du réseau intact. Les petits cercles sont les minicercles et les brins plus longs sont des parties des maxicercles (flèche). Le réseau entier est elliptique, plan et $\sim 10 \mu\text{m}$ par $15 \mu\text{m}$. L'échelle est à $0.3 \mu\text{m}$. (b) Le réseau défait par la topoisomérase II montrant des minicercles individuels (2.5kbp) et un seul maxicercle (flèche, 38kbp). L'échelle est à $0.5 \mu\text{m}$. L'image est tirée de l'article (Shapiro *et al.*, 1999)

Figure 10 : Images montrant l'ADNk des trypanosomatides et des bodonides.

Les flèches indiquent l'ADNk. Les photos insérées montrent les cellules colorées au DAPI (n, nucleus). (A) Section longitudinale au travers de l'ADNk à l'allure de disque de *C. fasciculata*. L'épaisseur du disque fait à peu près la moitié de la circonférence d'un minicercle (2.5 kbp). (B) Section longitudinale au travers du disque de l'ADNk de *T. avium*. Le disque apparaît cylindrique à cause de la grande taille des minicercles (10 kbp), mais l'organisation est similaire à celle de *C. fasciculata*. (C) Pro-ADNk de *B. saltans* dans une région dilatée de la mitochondrie, proche du corps de la base du flagelle. (D) Pan-ADNk de *C. helicis*, composé de plusieurs loci sensibles aux électrons dans la matrice mitochondriale. (E) Sections et section longitudinale (F) de la mitochondrie de *D. trypaniformis* montrant plusieurs poly-ADNK. (G et H) Sections transverses (G) et longitudinales (H) de *T. borreli* dans lequel un corps dense dû au mega-ADNK est réparti tout au long de la matrice mitochondriale. Échelle, 200 nm dans les panneaux A à F et 1 μm dans les panneaux G et H. Les cellules dans les encadrés sont à la même échelle.

Cette image est tirée de l'article (Lukes *et al.*, 2002)



(Zikova *et al.*, 2003), la coloration DAPI révèlent plusieurs corpuscules denses dispersées à travers la matrice mitochondriale. Dans ce poly-ADNk, les minicercles sont relâchés. Chez *Cryptobia helicis* (Lukes *et al.*, 1998), la coloration DAPI allume toute la matrice mitochondriale montrant la dispersion des molécules d'ADN superenroulées. Ce type de structure est nommé pan-kinétoplaste. Une dernière forme complètement différente est observée chez *Trypanoplasma borreli* (Maslov & Simpson, 1994). Des séquences de type minicercles sont concaténées les uns à la suite des autres pour ne former qu'une seule et grande molécule de 200 kbp.

Les maxicercles : l'information génétique cryptique

Les maxicercles codent pour des gènes mitochondriaux classiques et ont une taille allant de 19 kpb chez *Bodo caudatus* (Hajduk *et al.*, 1986) à 70 kbp chez *Bodo saltans* (Lukes *et al.*, 2002). La variation de leur taille est principalement due à la variation d'une grande région variable sans gène (8 kbp chez *T. brucei* et 12 kbp chez *L. tarentolae*), composée d'éléments répétés et impliquée dans la réplication. Les maxicercles correspondent à l'équivalent des chromosomes mitochondriaux habituels et comprennent les gènes codant pour les complexes de la chaîne de transfert d'électron soit *cox1*, *2* et *3*, *cob*, de *nad1*, *4*, *5*, *7*, *8*, *9*, *atp6*, *rps12* et au moins deux ORFs de fonctions inconnus nommées MURF (revue dans (Alfonzo *et al.*, 1997; Feagin, 2000)). La petite et la grande sous unité ribosomique sont très petites (650 et 1300 pb) comparativement aux autres espèces (par exemple, chez *Reclinomonas americana*: *rnl* = 2650 pb et *rns* = 1600 pb) (Lang *et al.*, 1997). Chez *L. tarentolae* et *Crithidia fasciculata*, les maxicercles possèdent, respectivement, 15 et 7 ARNg alors que celui de *T. brucei* n'en a que trois. Une des particularités du génome est qu'il ne code pour aucun ARNt.

L'édition des ARNm via l'insertion ou la suppression de U modifie les transcrits afin de les rendre traduisibles en protéines. Le nombre de changements est variable et dépendant de l'espèce et du gène. Le gène *cox2*, chez *T. brucei* et *C. fasciculata*, subit

l'insertion de quatre U remettant en phase la séquence codante (Benne *et al.*, 1986) alors que le gène *cox3* de *T. brucei* subit l'addition de 547 U et la délétion de 41 U (Feagin *et al.*, 1988).

Les minicercles et la machinerie d'édition d'ARN : la sauvegarde de l'information génétique

Les minicercles, dont la taille varie entre 0.5 et 10 kbp, représentent près de 90% de la masse du kinétoplaste. Au sein d'une même espèce, la séquence des minicercles peut être identique (en excluant la région codant pour les ARNg) comme chez *C. fasciculata* ou différente comme chez *T. brucei* avec ~100 minicercles différents (Hong & Simpson, 2003). Toutefois, tous les minicercles possèdent une région conservée de 100 à 200 pb, dans lesquels se trouvent deux motifs universels (un dodecamère et un hexamère) qui sont impliqués dans la réplication et un motif nommé « bent helix » dont la fonction est inconnue.

L'information spécifique de l'édition des ARNm est présente au niveau de la structure primaire des ARNg qui sont composés de trois régions aux fonctions différentes (figure 11). La région d'ancrage a une longueur de 10 nt qui permet de cibler spécifiquement le site d'édition (SE) en s'hybridant avec une séquence complémentaire de l'ARNm juste en amont du SE. Le premier mésappariement marque le début du SE sur l'ARNm et le début de la région informative de l'ARNg qui sert de matrice pour les insertions et suppressions de Us. Le transfert de l'information du guide au messager est basé sur les appariements Watson-Crick ainsi que les G : U. Une fois l'édition complétée, la région guide devient entièrement complémentaire de la région éditée de l'ARNm. Enfin, la région 3' de l'ARNg est la queue poly(U) dont le rôle semble stabiliser le complexe ARNg/ARNm par hybridation avec la région pré-éditée riche en purine.

Les ARNg ne sont pas catalytiques. Ce sont des protéines qui effectuent l'insertion et la suppression de Us. La cascade enzymatique est la suivante : (1) une endonucléase clive le SE de l'ARNm, (2) la suppression se fait par une exonucléase 3'-5' U spécifique et l'insertion par une « Terminale uridylyl transférase » (TUTase) et (3) une ligase ARN joint les deux morceaux une fois le processus fini. Ces protéines font parti du complexe nommé « editosome » qui comprend plus de 20 protéines (Stuart & Panigrahi, 2002).

Les autres modifications des ARNs

La transcription chez les kinétoplastides produit des transcrits polycistroniques pour les gènes des minicercles et des maxicercles. Chez *T. brucei*, les minicercles codent pour plusieurs ARNg qui possèdent tous une initiation de la transcription. L'arrêt de la transcription coïncide avec l'origine de réplication et le motif « bent helix ». L'ARN polycistronique des minicercles est ensuite clivé par le complexe 19S (nommé selon son coefficient de sédimentation) en ARNg monocistroniques (Grams *et al.*, 2000). L'ARNm est coupé aux bouts 5' et 3', cependant beaucoup de gènes se chevauchent et le clivage en 3' d'un gène a pour conséquence de tronquer le transcrit du gène suivant (Koslowsky & Yahampath, 1997).

La queue poly(A) qui est ajoutée suite au clivage en 3' des ARNm des trypanosomatides joue un double rôle. Les ARNm non édités avec une queue poly(A)₂₀ sont dégradés plus rapidement que les ARN non édités sans queue poly(A). A l'inverse, les ARNm édités qui ont une queue poly(A) de 20 A ou plus longue sont plus stables que ceux qui n'en ont pas.

Une autre modification est l'ajout d'une queue polyU aux ARN ribosomiques mitochondriaux. La longueur de la queue ajoutée à l'ARN 9S (11 Us) est constante alors que celle de l'ARN 12S est hétérogène suggérant deux TUTases différentes ou des régulations différentes.

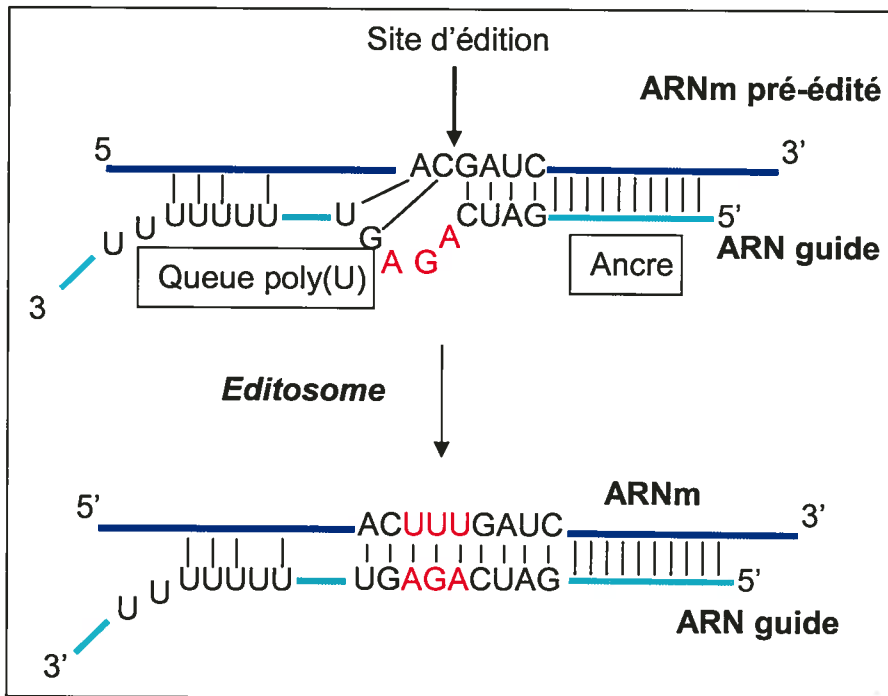


Figure 11 : Schéma de l'insertion de U dans un ARNm

1.2.2. Les euglénides

Le groupe des euglénides est composé de plusieurs centaines d'espèces phagotrophes, phototrophes ou osmotrophes (Leander et al., 2001; Busse & Preisfeld, 2003; Breglia et al., 2007). L'organisme modèle de ce groupe est *Euglena gracilis*, un organisme photosynthétique qui possède un plastide avec 3 membranes issu d'une endosymbiose secondaire. La mitochondrie de *E. gracilis* a été étudiée à plusieurs reprises. Les analyses en microscopie électronique montrent des molécules linéaires de 5 kpb alors que la complexité génétique est proche d'une molécule de 60 kpb (Nass et al., 1974; Fonty et al., 1975; Tessier et al., 1997; Yasuhira & Simpson, 1997). La séquence de l'ADN complémentaire du gène *cox1* a montré que l'ARNm ne possédait pas de queue poly(A) à la différence des trypanosomatides (Tessier et al., 1997; Yasuhira & Simpson, 1997). Toutefois, la structure reste un mystère avec la possibilité d'avoir un ADNmt composé de nombreuses molécules linéaires de petites tailles.

1.2.3. Les diplonémides

Ce groupe comporte deux genres, *Diplonema* et *Rhynchopus*. Le genre *Diplonema* a été décrit en 1913 par Griessmann comme un flagellé marin non coloré avec une forme cylindrique qui s'effile antérieurement, un corps muni de deux petites flagelles qui sortent d'une poche flagellaire et subissant des changements de forme rapide (métaboliques) (Griessman, 1914). *Diplonema ambulator* a été découvert lors de l'analyse des espèces qui détériorent la plante *Cryptocoryne*. La comparaison par microscopie électronique de *D. ambulator* avec *Isonema nigricans* (Schuster et al., 1968) et *Isonema papillatum* (Porter, 1973) a révélé qu'il n'y avait aucune différence significative entre les genres *Isonema* et *Diplonema*, définissant ainsi le genre *Isonema* comme un synonyme junior de *Diplonema* (Triemer & Ott, 1990).

Diplonema, qui ne possède qu'une seule mitochondrie par cellule comme les trypanosomes et les euglènes, n'a pas de kinétoplaste. L'ADNmt de *D. papillatum* a été étudié la première fois par Maslov en 1999 (Maslov et al., 1999). Les résultats de cette étude sont les suivants : (1) l'ADNmt est riche en G+C et la structure est inconnue; (2) la séquence de la moitié 3' de l'ADNc de *cox1* a été déterminée. L'ARNm de *D. papillatum* est polyadénylé et le codon TGA spécifie le tryptophane.

Mon projet

Les connaissances de l'ADNmt des Euglenozoa étaient restreintes aux kinétoplastides. Dans le but d'avoir un aperçu plus large de l'ADNmt de ce clade, nous nous sommes penchés sur le groupe frère des kinétoplastides soit les diplonémides. Plus précisément, nous avons commencé par l'analyse de *Diplonema papillatum* pour trois raisons : (i) Cet organisme n'est pas un parasite et il est non pathogène; (ii) Il est possible de le cultiver de façon axénique; (iii) Les études de Maslov *et al* ont montré qu'il était possible de travailler avec son ADNmt.

Les premières questions auxquelles nous souhaitons répondre étaient :

- Quelle est la structure de l'ADNmt? Existe-t'il plusieurs chromosomes différents comme pour l'ADNmt des kinétoplastides?
-
- Existe-t'il de l'édition d'ARN? Si oui, est-il identique à celui présent chez les kinétoplastides?

Les réponses a ces questions devraient aussi nous fournir des indices sur l'évolution du génome mitochondrial au sein des Euglénozoa.

Chapitre 2. Résultats

Dans le premier article, nous montrons que l'ADNmt est composé de chromosomes circulaires et que le gène *cox1* est fragmenté au minimum en deux morceaux qui sont codés par deux chromosomes distincts.

A partir de ces résultats préliminaires, nous avons plusieurs questions à résoudre telles que :

- Quelle est la composition génétique des chromosomes?
- Est-ce que tous les gènes sont fragmentés? Si oui, en combien de morceaux?
- Quel est le mécanisme qui permet la production d'un produit mature pour le gène *cox1*?

Afin de répondre à ces questions, nous avons séquencé l'ADN mitochondrial ainsi qu'une banque d'ADN complémentaire des ARN poly(A) de *Diplonema papillatum*. Le deuxième article décrit l'analyse de la séquence des chromosomes, la structure des gènes et leur expression.

2.1. Article 1 : Structure unique du génome mitochondrial des diplonémides, le groupe frère des kinétoplastides

par William Marande, Julius Lukes et Gertraud Burger

publié dans Journal of Eukaryotic cell (2005)

Dans cet article, les expériences concernant la microscopie électronique de *Diplonema* et de son ADNmt ont été conduites par le Dr Julius Lukes.

J'ai effectué les autres expériences présentées dans l'article.

Unique Mitochondrial Genome Structure in Diplonemids, the Sister Group of Kinetoplastids

William Marande,¹ Julius Lukeš,² and Gertraud Burger^{1,3*}

Université de Montréal, Robert-Cedergren Centre for Bioinformatics and Genomics, Department of Biochemistry, 2900 Boulevard Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada¹; Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, 37005 České Budějovice, Czech Republic²; Program in Evolutionary Biology, Canadian Institute for Advanced Research, Toronto, Ontario, Canada³

Received 27 January 2005/Accepted 19 April 2005

Kinetoplastid flagellates are characterized by uniquely massed mitochondrial DNAs (mtDNAs), the kinetoplasts. Kinetoplastids of the trypanosomatid group possess two types of mtDNA molecules: maxicircles bearing protein and mitoribosomal genes and minicircles specifying guide RNAs, which mediate uridine insertion/deletion RNA editing. These circles are interlocked with one another to form dense networks. Whether these peculiar mtDNA features are restricted to kinetoplastids or prevail throughout Euglenozoa (euglenids, diplomemids, and kinetoplastids) is unknown. Here, we describe the mitochondrial genome and the mitochondrial ultrastructure of *Diplonema papillatum*, a member of the diplomemid flagellates, the sister group of kinetoplastids. Fluorescence and electron microscopy show a single mitochondrion per cell with an ultrastructure atypical for Euglenozoa. In addition, DNA is evenly distributed throughout the organelle rather than compacted. Molecular and electron microscopy studies distinguish numerous 6- and 7-kbp-sized mitochondrial chromosomes of monomeric circular topology and relaxed conformation *in vivo*. Remarkably, the *coxI* gene (and probably other mitochondrial genes) is fragmented, with separate gene pieces encoded on different chromosomes. Generation of the contiguous *coxI* mRNA requires *trans*-splicing, the precise mechanism of which remains to be determined. Taken together, the mitochondrial gene/genome structure of *Diplonema* is not only different from that of kinetoplastids but unique among eukaryotes as a whole.

Euglenozoa (as defined in reference 13) comprise three morphologically distinct groups, the euglenids, diplomemids, and kinetoplastids. Several types of evidence support the taxonomic cohesion of Euglenozoa. They share discoidal mitochondrial cristae and a characteristic architecture of the flagellar and feeding apparatus (58, 69). Moreover, Euglenozoa add a spliced-leader RNA, a short, capped, noncoding transcript, to all cytoplasmic mRNAs (for reviews, see references 6 and 33). Another feature in common is the modified base “J” (beta-D-glucosyl-hydroxymethyluracil) that occurs in euglenozoan nuclear DNA (16). Most importantly, molecular phylogenies recover the monophyly of Euglenozoa (see, e.g., references 61 and 64).

But within Euglenozoa, phylogenetic relationships have been controversial for decades (30, 32, 40, 42, 43, 70). Only recently published molecular phylogenies provide significant support for euglenids diverging most basally within Euglenozoa, followed by the split into two sister groups, the diplomemids and kinetoplastids (59, 60, 61) (Fig. 1). While Euglenozoa as a whole have been viewed historically as one of the most early branching eukaryotic lineages (64), a critical review of statistical support and robustness of global phylogenetic trees questions the basal position of Euglenozoa (50).

The largest and best-studied euglenozoan group are kinetoplastids, small colorless flagellates characterized by massed

mitochondrial DNAs (mtDNAs) that form stainable structures visible by light microscopy, i.e., the kinetoplasts. Kinetoplastids are subdivided into trypanosomatids and bodonids. Trypanosomatid mtDNAs are composed of interlocked (catenated) networks of “maxicircles” and “minicircles” that are packaged with the aid of proteins into a single, dense disk (74). Maxicircles encode classical mitochondrial genes, most of whose transcripts undergo massive uridine insertion and deletion editing to generate translatable mRNAs. Minicircles encode guide RNAs (gRNAs), the templates for editing pre-mRNAs (reviewed in references 17, 37, and 62). Within bodonids, some taxa have a single kinetoplast similar to trypanosomatids, while others possess multiple mtDNA nodules or DNA distributed diffusely throughout the mitochondrion without visible superstructure. RNA editing activity has been confirmed for two bodonids, *Bodo saltans* and *Trypanoplasma borreli* (for a review, see reference 35).

Euglenids are the second largest group within Euglenozoa. The model organism *Euglena gracilis* has been extensively studied, at the level of both the nuclear and the chloroplast genome, whereas euglenid mtDNA seems to be intractable with present methods. So far, *coxI* (coding for cytochrome oxidase subunit 1) is the only mitochondrial gene cloned and sequenced for *E. gracilis* (68, 76). The structure of the mitochondrial genome as a whole remains unknown.

Diplonemids, the third and smallest group within Euglenozoa, contains a total of two genera, *Rhynchopus* and *Diplonema* (30, 55). First described in 1914 (24), *Diplonema* was rediscovered in the late 1960s (55), designated *Isonema*, and misplaced within euglenids (52). Most diplomemids inhabit marine ecosystems (63) and, similar to their kinetoplastid sister group,

* Corresponding author. Mailing address: Université de Montréal, Robert-Cedergren Centre for Bioinformatics and Genomics, Department of Biochemistry, 2900 Boulevard Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada. Phone: (514) 343-7936. Fax: (514) 343-2210. E-mail: [REDACTED]

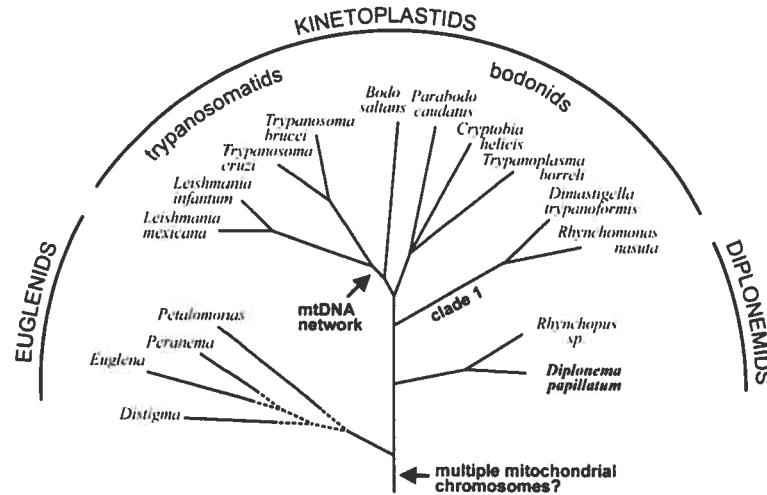


FIG. 1. Schematic phylogenetic tree of Euglenozoa. The topology is taken from studies by Simpson and coworkers (60, 61), where bootstrap support is significant for both the common ancestry of Euglenozoa and the diplomemids and kinetoplastids (but see references 43 and 73). The branching order within euglenids is tentative. Arrows indicate, as proposed in this report, the time points in evolutionary history when multiple mitochondrial chromosomes and mtDNA networks arose in Euglenozoa. Note that *Parabodo caudatus* was previously designated *Bodo caudatus* (43).

they include pathogenic taxa. Some *Rhynchopus* species parasitize diatoms (54) and others lobster (73), while *Diplonema* not only has been reported to infest clams (29) but also has been associated with *Cryptocoryne* disease, which causes the sudden decay of aquarium plants (65).

The mitochondrial genome of diplomemids is virtually uncharacterized. The only molecular data currently available are a partial *cox1* cDNA sequence of *Diplonema papillatum* (40), but the subcellular location of the corresponding transcript or its gene is not known. We initially attempted to characterize *Diplonema* mtDNA using whole genome shotgun sequencing within the context of the Organelle Genome Megasequencing Program (<http://megasun.bch.umontreal.ca/ogmproj.html>). But sequence assembly of the presumptive mtDNA was highly ambiguous, due to extensive sequence repeats, and not a single known mitochondrial gene was discernible in the ~20 kbp that was sequenced (G. Burger, B. F. Lang, and M. W. Gray, unpublished results). This report describes the approach that allowed us to identify and characterize mtDNA for *Diplonema*, thus laying the groundwork for the development of new strategies to sequence its complete mitochondrial genome and for the discovery of an unprecedented mitochondrial gene and genome structure for this organism.

MATERIALS AND METHODS

Strains and protist culture. *D. papillatum* (ATCC 50162) was obtained from the American Type Culture Collection. The organism was cultivated axenically at ~22°C in artificial seawater and enriched with 1% fetal bovine serum (Wisent) and 0.1% tryptone.

Light microscopy. Cells were resuspended in phosphate-buffered saline (PBS) at a concentration of 10^7 cells/ml, and the suspension was spotted onto poly-L-

lysine-coated slides. After the cells were allowed to adhere for 30 min in a humidity chamber, the slides were submerged for 5 min in methanol. Fixation was stopped by washing the slides with PBS, and cells were stained with 0.1 µg/ml DAPI (4',6'-diamidino-2-phenylindole, a DNA-specific fluorescence dye) in PBS for 3 min. For staining with the fluorescence dye YOYO1 (oxazole yellow dimer; Molecular Probes), cells were attached to slides as described above, fixed in PBS with 4% paraformaldehyde for 3 min, washed twice in PBS, and permeabilized in PBS with 0.5% Tween 20 for 30 min. After a single wash in PBS containing 0.05% Tween 20 for 5 min, cells were incubated with 10 nM YOYO1 for 20 min and rinsed twice with PBS. The DAPI- and YOYO1-stained cells were mounted in antifade reagent [0.233 g/l, 4-diazabicyclo-(2,2,2) octane; 1 ml 0.2 M Tris-HCl, pH 8.0, and 9 ml glycerol], and the slides were examined with a Zeiss Axioplan 100 microscope.

Electron microscopy. For transmission electron microscopy, cells were washed twice in PBS and fixed in the same buffer with 2.5% glutaraldehyde for 1 h at 4°C and then washed in PBS, fixed with 2% osmium tetroxide in the same buffer, and finally washed in distilled water. After dehydration in graded series of ethanol, the cells were embedded in Epon-Araldite. Thin sections of embedded cells were stained with lead citrate and uranyl acetate and examined under a JEOL JEM 1010 microscope. The images of adjacent sections (60 to 70 nm thick) were registered interactively by translations and rotations. Contours of structures to be reconstructed were outlined manually in distinct colors. Volumetric models of the structures were generated by plotting the contours in successive planes of the three-dimensional (3D) data sets. To remove irregularities due to deformation of microscopic sections, the data sets were resampled and smoothed. Surface models of the structures were constructed as triangulated isosurfaces of the data by the marching cubes algorithm (28). The final model was generated by combining the colored surfaces. Model images were obtained by rendering the resulting 3D scenes using the OpenGL library.

For electron microscopy studies of the topology and size of DNA molecules, *Diplonema* mtDNA was prepared by the cytochrome c method and spread and recorded as described earlier (36). The contour lengths of the circular chromosomes were measured from prints using a HIPAD digitizing tablet (Houston Instruments). The sizes determined were 2.05 ± 0.05 µm and 2.35 ± 0.05 µm. The precise magnification was determined by replica grating (Balzers). Plasmid pBR322 served as a size standard for the conversion from µm to kbp.

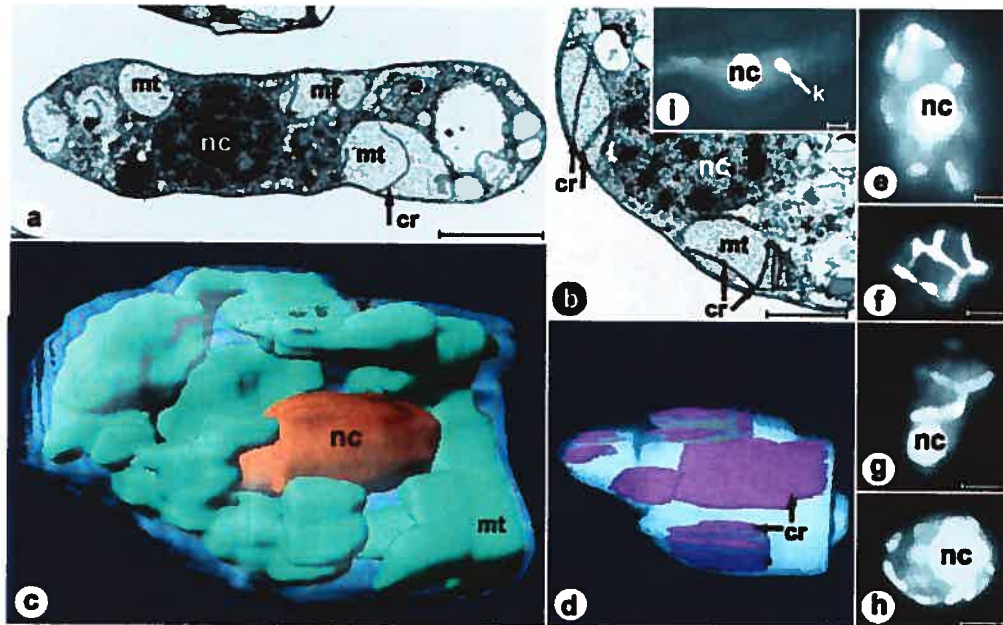


FIG. 2. Ultrastructure of *D. papillatum* mitochondria. Electron microscopy (a and b), light microscopy (e to i), and 3D reconstruction (c and d) based on electron microscopy images. Electron micrographs of longitudinally sectioned (a) and cross-sectioned (b) *Diplonema* cells. Prominent flat cristae are visible in the organellar lumen. (c) 3D reconstruction of the whole cell, based on 48 consecutive sections. The single mitochondrion is shown in cyan, and the posteriorly located oval nucleus is shown in gold. (d) 3D reconstruction of a large compartment of the mitochondrion, based on 25 consecutive sections. Cristae are shown in purple. (e to g) YOYO1-stained *D. papillatum* cells focused on the nucleus (e and g) and the cell periphery (f). (h) DAPI-stained *D. papillatum* cell. (i) Control staining (YOYO1) of the kinetoplast *Trypanosoma brucei*. Bars, 2 μm (a), 0.5 μm (b), and 2.5 μm (e to i). nc, nucleus; mt, mitochondrion; k, kinetoplast; cr, cristae.

Isolation of mitochondria and mtDNA extraction. Cells of *D. papillatum* were resuspended in STE buffer (0.6 M sorbitol, 50 mM Tris-HCl, and 5 mM EDTA [pH 7.4]) and broken mechanically by shaking with glass beads. The mitochondria were collected after three differential centrifugation steps using a Sorvall SS34 rotor, as follows: two times at 8,000 rpm ($7,650 \times g$) for 10 min and once at 18,000 rpm ($38,700 \times g$) for 20 min. We confirmed that this subcellular fraction is highly enriched in mitochondria by electron microscopy, which revealed the typical double-membrane-bounded vesicles (data not shown). The mitochondrion-containing pellet was resuspended in STE buffer and subsequently lysed in the presence of 1% (vol/vol) sodium dodecyl sulfate (SDS) and 100 $\mu\text{g/ml}$ proteinase K. The SDS was eliminated by NaCl precipitation, the lysate was phenol-chloroform extracted, and the DNA was precipitated by ethanol. Nuclear DNA was isolated from the pellet obtained after the first centrifugation step. For a more thorough purification of mitochondria, the mitochondrion-containing pellet was resuspended in 15% sucrose (and 0.1 mM EDTA) and layered on top of a sucrose step gradient (25, 35, 45, and 65% sucrose). Centrifugation was carried out by using a swing-out rotor (Sorvall TH641) at 38,000 rpm ($224,000 \times g$) for 90 min. The mitochondrial band, which formed between the layers of 35% and 45% sucrose, was collected and used directly for mtDNA extraction. The overall mtDNA patterns (monitored by agarose gel electrophoresis) before and after the gradient were identical.

Pulsed-field gel electrophoresis. Whole-chromosome sized DNAs were prepared by using agarose plugs at a concentration of 10^7 cells/ml. Electrophoresis was carried out in a contour-clamped homogeneous electric field electrophoresis apparatus (CHEF-DR III, Bio-Rad) in 1% agarose (SeaKem) and $1 \times$ Tris-acetate-EDTA buffer at 16°C for 11 h, using a linear switch time ramp (1 to 6 s) at an angle of 120° and a voltage of 6 V/cm. After migration, the gel was stained with ethidium bromide.

Isolation and fractionation of total DNA. Cells from a fresh culture were washed in STE buffer and lysed in the presence of 1% SDS and 100 $\mu\text{g/ml}$ proteinase K for 1 h at 50°C . The SDS was eliminated by NaCl precipitation. Total DNA was purified according to two different methods, either (i) the lysate was phenol-chloroform extracted and, subsequently, the DNA was ethanol precipitated, or (ii) DNA was ethanol precipitated and extracted by spooling around a Pasteur pipette. The second method allows extraction without the centrifugation that might break large DNA molecules. Total DNA was fractionated according to A+T content using CsCl-Hoechst 33258 equilibrium density centrifugation in an NVT65 rotor at 55,000 rpm for 16 h.

Southern hybridization. Electrophoresis and blotting were performed by standard procedures. The DNA probes were labeled by random priming in the presence of $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ (Perkin Elmer). Hybridization was performed at 65°C in $1.5 \times$ SSPE (3.6 M NaCl, 0.2 M sodium phosphate, and 20 mM EDTA [pH 7.4]) plus 1% SDS, 0.5% dry milk, and 0.1 mg/ml salmon sperm.

PCR amplification. N-terminal and C-terminal fragments of *cox1* were amplified from total or mtDNA using primer set 1 and 2 and primer set 7 and 8, respectively. Different combinations of primers 3 to 8 were used to amplify internal fragments. Primer sequences (5' to 3') are as follows: primer 1, GCAC TGCTAGTAGGTATCAT; primer 2, CTCTGGGTGCCCGAAGACC; primer 3, CGCTCTACAACATGCTGGGG; primer 4, CCAGGA GAACACCTTGAT GGA; primer 5, CCTGGTAATGCATGACAGCT; primer 6, GGTATCCATC AGGTGCATCT; primer 7, CCTGGTGTACTGGTAGTC; and primer 8, CCTCCAGATGCATGGATGCT. Assuming a gene that is colinear with the cDNA, the expected lengths of the PCR products are 150 bp for primers 1 and 2, 1,006 bp for primers 1 and 8, 171 bp for primers 3 and 4, 543 bp for primers 3 and 6, 771 bp for primers 3 and 8, 500 bp for primers 5 and 8, and 228 bp for primers 7 and 8.

RT-PCR. Total RNA was extracted using TRIzol. First-strand cDNA synthesis by AMV reverse transcriptase (Boehringer) was primed by an anchored dT oligonucleotide, followed by PCR-amplification using *cox1* primers listed above. Alternatively, first strand synthesis was primed by the 3'-distal primer 8, and amplification proceeded by using primers 1 and 8 or 3 and 8. The expected lengths of reverse transcriptase (RT)-PCR products are as listed above.

DNase I digestion. One μg of mtDNA was digested with 1×10^{-3} , 2×10^{-3} , or 4×10^{-3} units of DNase I (Fermentas) for 6 min at 4°C and then stopped by the addition of EDTA (50 mM).

DNA sequencing and sequence analysis. PCR products were purified by spin columns (Millipore). DNA sequencing was performed with an MJ automated sequencer, using the cycle-sequencing/dye termination chemistry (BigDye Terminator 3.1, ABI). Sequence similarity searches were conducted locally (FASTA) (48) against data retrieved from the Organelle Genome Database at <http://megasun.bchumontreal.ca/gobase/> and by remote BLAST searches at the National Center for Biotechnology Information (1).

Nucleotide sequence accession number. The DNA sequence reported in this paper was deposited in GenBank under accession no. AY686226.

RESULTS

Cellular ultrastructure and mtDNA packaging in *Diplonema*. Transmission electron micrographs of both longitudinally and cross-sectioned *D. papillatum* cells show several large mitochondrial structures (Fig. 2a and b), which are, in fact, branches of a single organelle, as demonstrated later. The mitochondrial lumen is homogeneously electron transparent and contains few, exceptionally large and irregularly arranged cristae, which are flat, uniformly thick (~25 nm), unbranched, and only rarely attached to the mitochondrial inner membrane. Although several mitochondria have been cross-sectioned completely, we did not detect kinetoplast-like bodies (35), as seen with the control (*Trypanosoma brucei*) (Fig. 2i).

3D reconstruction of serially sectioned *Diplonema* cells reveals a single, highly branched mitochondrion, anastomosing throughout the cell, but mostly located peripherally (Fig. 2c). Mitochondrial cristae, which protrude irregularly into the lumen, exhibit a variable but overall exceptionally large size (up to 2 μm long and 1 μm wide), lamellar shape, and longitudinal rather than transversal arrangement (Fig. 2d). Both nucleic acid-staining dyes used, YOYO1 (Fig. 2e to g) and DAPI (Fig. 2 h), light up large amounts of mtDNA meandering through the cell. To precisely colocalize DNA and organelle, we attempted to label the mitochondrion with MitoTracker in whole cells. But this experiment failed, possibly due to insufficient membrane permeability. Nevertheless, the pattern of DNA distribution within the cell, as monitored by fluorescence microscopy, is fully consistent with the spatial mitochondrial morphology obtained from 3D reconstruction. This provides clear evidence that mtDNA is uniformly distributed in the mitochondrial branchwork and not compacted in a single kinetoplast, as in trypanosomatids.

Purification and identification of mtDNA. Employing degenerate primers for RT-PCR, Maslov and colleagues isolated a partial *cox1* cDNA of *D. papillatum*, which includes ~1/2 of the expected coding sequence starting at a position corresponding to residue 246 of the *T. brucei* protein (40). This cDNA probe hybridized to a G+C-rich DNA fraction of *Diplonema*, which the authors considered to be mtDNA, because *cox1* is mitochondrion encoded in all species examined to date. However, several considerations suggest the use of caution about such an extrapolation. One is a recent report about the unexpected cytoplasmic localization of typical chloroplast

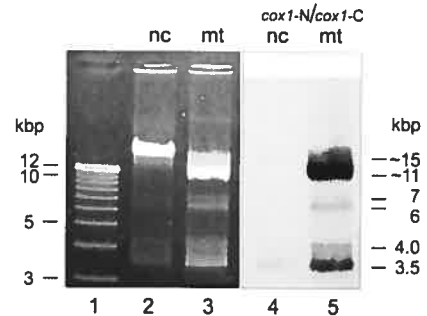


FIG. 3. Identification of *D. papillatum* mtDNA. Agarose gel electrophoresis of DNA from different subcellular fractions. Electrophoresis was performed in the presence of ethidium bromide. Lane 1, size markers (1 kb plus ladder); lane 2, DNA isolated from the nuclear subcellular fraction (nc). The pattern is identical with that of the A+T-rich band from CsCl-bisbenzimidazole equilibrium centrifugation (data not shown). Lane 3, DNA extracted from the mitochondrial subcellular fraction (mt). The pattern is identical with that of the G+C-rich band from CsCl-bisbenzimidazole equilibrium centrifugation (data not shown). Lanes 4 and 5, Southern hybridization of blotted lanes 2 and 3, using as a probe a mixture of *cox1* N-terminal and C-terminal fragments (*cox1*-N and *cox1*-C; see Fig. 7A) generated by PCR. Note that the nuclear fraction is slightly cross-contaminated by mtDNA, as revealed by staining (lane 2) and hybridization (lane 4). Moreover, a minute proportion (<1/1,000) of a DNA species migrating at ~23 kbp is seen in all preparations and most likely corresponds to covalently closed tandem homodimers of mitochondrial chromosomes, molecules well known from other systems with circular mtDNA, such as mammals (14) and *Bodo* species (4, 25). Sizes indicated to the right of lane 5 correspond to the apparent sizes of chromosomes, which are different from their true sizes when molecules are not linear (see text).

genes in dinoflagellates (31). Moreover, high G+C content is unusual for mtDNA but rather typical for nuclear DNA. Finally, the *cox1* cDNA was obtained from a polyadenylated mRNA fraction, while mitochondrial mRNAs of most eukaryotes, except mammals, apicomplexans, and kinetoplastids, do not have A-tails.

To identify the mtDNA of *Diplonema* beyond doubt, we gently disrupted the cells mechanically, separated mitochondria from nuclei and other subcellular structures by kinetic step-gradient centrifugation, and extracted the DNA from each subcellular fraction (see Materials and Methods). In parallel, total cellular DNA was separated by CsCl-bisbenzimidazole equilibrium centrifugation, which yielded two discrete bands, one more A+T rich, the other more G+C rich. Gel electrophoresis shows that the A+T-rich material corresponds to the DNA recovered from nuclei (Fig. 3, lane 2), whereas the G+C-rich material corresponds to mtDNA (Fig. 3, lane 3). In Southern hybridization, labeled PCR-generated fragments of *Diplonema cox1* cDNA indeed hybridize to the DNA extracted from mitochondria (Fig. 3, lane 5). Moreover, digestion with restriction enzymes demonstrates that the G+C-rich DNA species is much less complex than the A+T-rich one (data not shown). Together, these results confirm that in *Diplonema*, the more G+C-rich DNA fraction represents the mitochondrial genome.

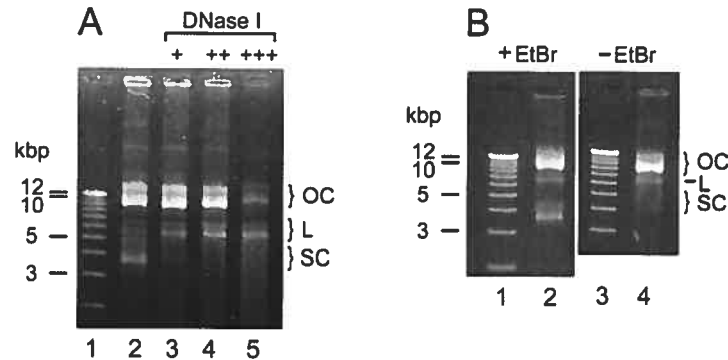


FIG. 4. Conformation of *D. papillatum* mitochondrial chromosomes. OC, relaxed closed circles; L, linear forms; SC, supercoiled, covalently closed circles. (A) incubation of mtDNA in the presence of DNase I. Lane 1, size markers as in Fig. 3, lane 2, control without DNase I treatment; lane 3 to 5, incubation with 1×10^{-3} units (+), 2×10^{-3} units (++), and 4×10^{-3} units (+++) of DNase I for 6 min. (B) Electrophoretic migration in the presence (lane 2) and absence (lane 4) of ethidium bromide. Lanes 1 and 3, size markers as in Fig. 3.

Shape and conformation of mtDNA. From their analyses of restriction enzyme digests, Maskov and colleagues concluded that *D. papillatum* mtDNA "is probably composed of several types of circles, [whose] exact size, number of types and topological arrangements remain to be investigated" (40). We determined the physical properties of this genome by the following experiments.

Electrophoresis of purified *Diplonema* mtDNA on agarose gels (in the presence of ethidium bromide) yielded a complex pattern of three pairs of bands, with apparent sizes of 3.5 and 4 kbp, 6 and 7 kbp, and 11 and 15 kbp, present in unequal stoichiometry (see Fig. 3, lane 3). To test whether these multiple bands are due to two circular molecules occurring in three different conformations or, rather, to six molecules of different sizes, we incubated mtDNA for various periods with the nicking enzyme DNase I. In the case of circular, supercoiled DNA, an increasing number of nicks would first abolish the supercoiled conformation and then linearize the circle, followed by fragmentation and final degradation of this linear form (15). Figure 4A illustrates that treatment of *Diplonema* mtDNA with DNase I eliminates progressively the bands at 3.5 and 4 kbp followed by those at 11 and 15 kbp, proceeding gradually by first enhancing and then diminishing the bands at 6 and 7 kbp. This behavior is consistent with two molecule classes, of 6 kbp and 7 kbp, present in the following three different conformations: (i) linear (migrating at 6 kbp and 7 kbp according to their size), (ii) relaxed circular (migrating at 11 and 15 kbp), and (iii) supercoiled, covalently closed-circular (migrating at 3.5 and 4.0 kbp). In the present paper, the two size classes of molecules are referred to as class A (6-kbp) and class B (7-kbp) chromosomes.

Ethidium bromide induces the supercoiled conformation in relaxed closed DNA circles by intercalation between two successive base pairs. This unwinds the DNA double helix and produces negative supercoils (51). To test whether supercoiling is secondarily induced by staining, we performed electrophoretic separation in the absence of this dye. Indeed, the

fast-migrating bands are absent (Fig. 4B, lane 4), demonstrating that *in vivo*, the major portion of *Diplonema* mtDNA is in monomeric, relaxed circular conformation.

Electron microscopy captures visually the size and topology of *Diplonema* mitochondrial chromosomes (Fig. 5). For this experiment, DNA was isolated in the absence of intercalating agents. Among 61 screened DNA molecules, >95% are circular and >85% are in clearly relaxed conformation; 54% measured 6.0 to 6.3 kbp (class A) and 46% measured 7.0 to 7.3 kbp (class B). These results are in full agreement with the molecular data described above.

With sizes of only 6 and 7 kbp, *Diplonema* mitochondrial chromosomes are unusually small. To investigate the potential presence of large mtDNA molecules that may not have entered

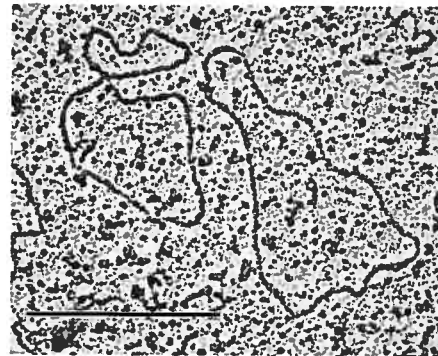


FIG. 5. Electron microscopy analysis of *Diplonema* mtDNA. Circular chromosomes in relaxed conformation. Left, class A chromosome (contour length corresponding to 6.0 to 6.3 kbp); right, class B chromosome (contour length corresponding to 7.0 to 7.3 kbp). Bar, 0.5 μ m.

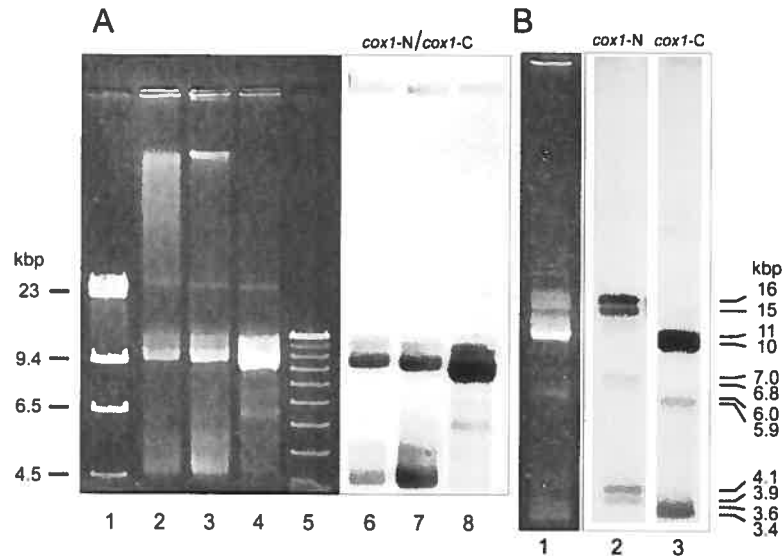


FIG. 6. Structure of the *D. papillatum* *cox1* gene. (A) Pulsed-field separation of DNA. Lane 1, size marker lambda HindIII, lane 2, total DNA purified by spooling; lane 3, total DNA purified by phenol-chloroform extraction; lane 4, mtDNA isolated from the mitochondrial subcellular fraction; lane 5, size marker as in Fig. 3; lanes 6 to 8, blotted lanes 2 to 4, hybridized with a mixture of *cox1-N* and *cox1-C* as a probe (see Fig. 7A). (B) Southern hybridization of mtDNA with two different *cox1* probes. Lane 1, mtDNA separated by agarose gel electrophoresis in the presence of ethidium bromide; lane 2, hybridization of blotted lane 1 with *cox1-N* (N-terminal fragment); lane 3, hybridization of blotted lane 1 with *cox1-C* (C-terminal fragment) (see Fig. 7A). Apparent molecule sizes (see legend to Fig. 3) are indicated to the right of lane 3. For information on the minute band migrating at ~23 kbp, see legend to Fig. 3. Note that due to the particular migration conditions in this experiment, electrophoretic mobilities of DNA species differ from those shown in Fig. 3 and 4.

the gel during conventional electrophoresis, we used pulsed-field gel electrophoresis. Under conditions separating molecules of several hundred kbp, no significant high-molecular band in mtDNA is discernible, either by ethidium bromide-staining or by Southern hybridization (Fig. 6A, lanes 4 and 8). This result shows that *Diplonema* mtDNA does not include detectable amounts of larger (maxicircle-like) molecules.

Similarity of class A and B chromosomes and localization of *cox1*. To test whether the two chromosome classes have sequence in common, we performed Southern hybridizations with three different probes, complete A-class and B-class chromosomes and a 4-kbp genomic *cox1* clone of *Diplonema*, which was isolated and provided to us by D. Maslov. We sequenced this clone and discovered a ~220-bp-long stretch of significant sequence similarity to *cox1*, corresponding to amino acids ~193 to 268 of the *T. brucei* protein. This stretch of *cox1*, referred to below as *cox1-N*, is bounded by highly repetitive sequence (GenBank accession no. AY686226). Each of the above three probes hybridized equally strongly to all *Diplonema* mtDNA bands (data not shown), showing that the two chromosomes share significant sequence similarity.

The two chromosome classes also include distinct regions, as seen in the following Southern hybridization results. Probes for this experiment contained solely *cox1* regions, which were obtained by PCR amplification from mtDNA: *cox1-N* (see above)

and a most C-terminal *cox1* portion designated *cox1-C* (corresponding to residues 477 to 543 of the *T. brucei* protein; for primers, see Fig. 7A). Surprisingly, the *cox1-N* probe recognized exclusively class B molecules (Fig. 6B, lane 2), whereas the *cox1-C* probe specifically bound to class A chromosomes (Fig. 6B, lane 3). In addition, Southern hybridization reproducibly resolved double bands of A and B chromosomes. This implies the presence of two versions each of chromosomes carrying *cox1-N* and *cox1-C* sequences, differing in size by 100 to 200 bp.

With the discovery of N-terminal and C-terminal portions of *Diplonema cox1* residing on separate mitochondrial chromosomes, the question arises whether a low abundant, continuous *cox1* exists as well. Therefore, we conducted PCR experiments with total DNA, using diverse primer pairs to cover different portions of the hypothetical *cox1* coding region (Fig. 7A). Figure 7B shows that only amplification of *cox1* segments shorter than 300 bp was successful. As a positive control, we performed RT-PCR with total RNA using far distant primers, which yielded products of the predicted sizes (Fig. 7B, lanes 7 and 8) and expected sequence (identical to the *cox1* cDNA sequence; GenBank accession no. AF119813 [40]). This demonstrates that instead of a contiguous *cox1* gene, *Diplonema* possesses several ~250-bp *cox1* regions scattered across multiple chromosomes.

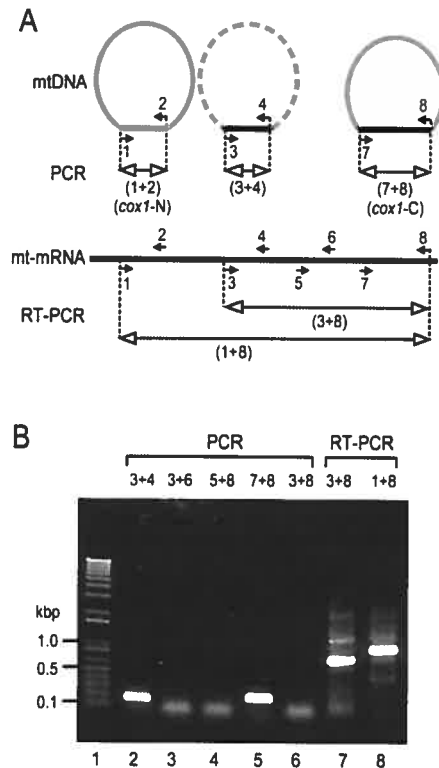


FIG. 7. PCR amplification of the *D. papillatum cox1* gene. (A) Schematic representation of the *cox1* sequence and PCR primers. Semicircular forms indicate mtDNA chromosomes, with the *cox1* coding region in black and the rest of the chromosome in gray. The existence of the corresponding chromosome remains to be demonstrated (symbolized by the dashed arc), but the *cox1* region has been obtained by PCR. The chromosome depicted on the left gave rise to the genomic *cox1* clone, which was isolated by D. Maslov and sequenced by us (this report). Filled small arrows with numbers indicate primers and their orientations. Lines with open arrowheads depict PCR products. The portion of mitochondrial mRNA between primers 3 and 8 corresponds to the published *cox1* cDNA sequence (GenBank accession no. AF119813 [40]). (B) Electrophoretic separation of PCR products. Lane 1, size marker as in Fig. 3; lanes 2 to 6, mtDNA was PCR amplified using different primer pairs as indicated above the lanes. The numbers refer to the primers in Fig. 7A. Lanes 7 to 8, total RNA was reverse transcribed and PCR amplified using primer set 3 and 8 and primer set 1 and 8, respectively. The diffuse bands migrating at ~50 bp in lanes 3, 4, 6, and 7 are most likely primer dimers, as they also occur in control experiments without DNA (not shown). For primer sequences, see Materials and Methods.

DISCUSSION

Atypical ultrastructure of *Diplonema* mitochondria. Our 3D reconstructions show for the first time that *D. papillatum* contains a single mitochondrion per cell, as do kinetoplastids.

Apart from that, *Diplonema* mitochondria appear to be highly atypical for Euglenozoa. Compared to the threadlike mitochondria with numerous tiny and discoidal mitochondrial cristae in kinetoplastids and euglenids (9, 18; see also reference 49), the organelle of *D. papillatum* is amoeboid in structure and encloses scarce, giant, and flat mitochondrial cristae (see also reference 58). Our study provides new evidence for the notion that *Diplonema* is at odds with the distinctive discoidal mitochondrial cristae of "Discicristata" (Euglenozoa and Heterolobosea) (47).

Our fluorescence microscopy studies show that, in contrast to the kinetoplast DNA (kDNA) of trypanosomatids, *Diplonema* mtDNA is distributed diffusely throughout the organelle. A comparable mtDNA distribution occurs sporadically in bodonids (pan-kDNA and mega-kDNA), but the large majority has mtDNA molecules assembled into discrete nodules (poly-kDNA) or a single bundle-like structure (pro-kDNA) (72; see compilation in Table 1).

Unique mitochondrial gene organization in *Diplonema*. DNA sequencing, Southern hybridization, and PCR experiments provide evidence that *D. papillatum* mtDNA contains at least three distinct ~250 bp-long stretches of *cox1* sequence, corresponding approximately to residues 193 to 247, 269 to 326, and 477 to 543 of the *Trypanosoma* protein. Each of these stretches is located on a separate chromosome; a continuous *cox1* coding region was not detected. Two different scenarios could explain this finding. Reminiscent of the situation for kinetoplastids, the observed *cox1* stretches might represent grNA genes involved in editing of heavily encrypted transcripts. However, kinetoplastid grNA genes are 5 to 10 times shorter than the *Diplonema cox1* stretches, and their regions of sequence identity with the mature transcript are only 5 to 15 bp long. Therefore, it is much more likely that the ~250-bp *cox1* stretches are separate coding portions of a discontinuous *cox1* gene.

Discontinuous genes are widespread in organelle genomes outside Euglenozoa and also occur in nuclear genomes. Four types of discontinuous genes can be distinguished. Most abundant are intron-containing genes, whose intervening sequences are removed from pre-mRNA by *cis*-splicing. A second type involves exons that are encoded in distant locations or even on different strands of the chromosome and that are transcribed separately, each with parts of flanking introns. The precise linking of these discrete precursor transcripts to form a mature RNA proceeds by *trans*-splicing. In organelles, best known from plant mitochondria, this step is mediated by the distinct secondary structure folding of group II introns (41), while *trans*-splicing in the nucleus, first described for *T. brucei*, involves spliceosomes (for reviews, see references 6, 33, and 57). Another type of discontinuous genes involves DNA splicing that generates secondarily a contiguous gene version (e.g., macronuclear genes in ciliates [reviewed in reference 27] and mammalian immunoglobulin and T-cell receptor genes [reviewed in reference 22]). The fourth type, previously dubbed "genes in pieces" (5), but here more specifically referred to as "modular genes," is characterized by discontinuous gene products. Modular mitoribosomal RNAs are relatively frequent, with examples in protists (5, 19, 23, 26, 53, 71) and fungi (20). The few cases of modular mitochondrial protein genes include *nad1* (coding for subunit 1 of NADH dehydrogenase) of cili-

TABLE 1. Architecture and gene content of diplomemid and kinetoplastid mtDNAs

Taxon	mtDNA distribution ^a	kDNA network ^b	Chromosome type	Size (kbp)	Shape/conformation	Coding content(s)	RNA editing	Source or reference
Diplonemida								
<i>D. papillatum</i>	Pan-kDNA-like	No	A-class	6	Circular	Genes, gRNAs	ND ^c	This study
			B-class	7	Relaxed	Genes, gRNAs	ND	
Kinetoplastida-trypanosomatids								
<i>Crithidia fasciculata</i>	Eu-kDNA	Yes	Minicircle	2.5	Circular	gRNAs	Yes	75
			Maxicircle	38	Relaxed	Genes, gRNAs		
<i>L. tarentolae</i>	Eu-kDNA	Yes	Minicircle	0.9	Circular	gRNAs	Yes	67
			Maxicircle	21	Relaxed	Genes, gRNAs		
<i>Phytomonas serpens</i>	Eu-kDNA	Yes	Minicircle	1.5	Circular	gRNAs	Yes	38
			Maxicircle	27-36	Relaxed	Genes, gRNAs		
<i>Trypanosoma avium</i>	Eu-kDNA	Yes	Minicircle	10	Circular	ND	ND	77
			Maxicircle	ND	Relaxed	ND		
<i>T. brucei</i>	Eu-kDNA	Yes	Minicircle	1.0	Circular	gRNAs	Yes	44
			Maxicircle	21-27	Relaxed	Genes, gRNAs		
<i>Trypanosoma vivax</i>	Eu-kDNA	Yes	Minicircle	0.5	Circular	ND	Yes?	7
			Maxicircle	23	Relaxed	ND		
Kinetoplastida-bodonids								
<i>P. caudatus</i>	Pan-kDNA?	No	Minicircle	10-12	Supercoiled	ND	Yes	25
			Maxicircle	19	ND	ND		
<i>B. saltans</i>	Pro-kDNA	No	Minicircle	1.4	Circular	gRNAs	Yes	3
			Maxicircle	70	Relaxed	Genes, gRNAs		
<i>Cruzella marina</i>	Poly-kDNA	No	Minicircle	2.0	Circular	gRNAs?	ND	78
						Relaxed		
<i>Cryptobia helicis</i>	Pan-kDNA	No	Maxicircle	ND	ND	ND	ND	36
			Minicircle	4.2	Supercoiled	gRNAs?		
<i>Dimastigella trypaniformis</i>	Poly-kDNA	No	Maxicircle	43	ND	Genes	ND	66
			Minicircle	1.45	Circular	gRNAs?		
<i>Trypanoplasma borreli</i>	Mega-kDNA	No	Maxicircle	ND	ND	ND	Yes	34,39
			Minicircle	200	Linear 1 kbp	gRNAs		
			Maxicircle	80	Tandem rep.	Genes		

^a Determined by DNA staining in whole cells.

^b Determined by electron microscopy and molecular studies of DNA.

^c ND, not determined.

ates (12) and *cox2* (coding for cytochrome oxidase subunit 2) of the green alga *Scenedesmus obliquus*. In all instances, gene modules are encoded on the same DNA molecule, with the notable exception of *Scenedesmus cox2*, one module of which is encoded by mtDNA (45) and the other by nuclear DNA (21; reviewed in reference 11). Here, we report discontinuous genes of *Diplonema* mtDNA that differ from all types described above in that each of the multiple gene fragments is located on a distinct mitochondrial chromosome.

Evidently, gene expression of *Diplonema cox1* does not involve DNA splicing; otherwise, a contiguous gene should have been detected by PCR. The gene fragments do not code for protein pieces, because the *cox1* cDNA is contiguous. Instead, the distinct *cox1* gene fragments must be transcribed separately and joined together posttranscriptionally by some kind of *trans*-splicing. The molecular mechanism by which this process may take place is currently under investigation.

From the existence of at least three separate and dispersed *cox1* gene fragments and the failure to amplify genomic *cox1* regions of >300 bp, we infer that this gene is split up entirely into ~250-bp fragments. At an average length of 1,500 to 1,650 bp for euglenozoan *cox1* coding sequences (34, 40, 76), the predicted number of *cox1* coding fragments in *Diplonema*

should be five or six. With each of these fragments encoded on a separate chromosome and at least two chromosome versions per gene fragment (see Fig. 6B), the genome portion occupied by *cox1* alone would then amount to as much as 60 to 84 kbp. This low portion of coding sequence is fully consistent with the absence of detectable genes in the previously determined ~20 kbp random sequence of *Diplonema* mtDNA.

Conjectures on structure and size of the *Diplonema* mtDNA. *Diplonema* is one of the few taxa with true monomeric circular mtDNA molecules, together with several animals and kinetoplastids. Otherwise, the shape of mitochondrial chromosomes is generally linear, composed of multiple, tandemly arranged copies, yet appearing circular in physical mapping and sequence assembly (2; reviewed in reference 46). Based on the following estimations, *Diplonema* may possess one of the largest numbers of mitochondrial chromosomes. Assuming a gene set of 15 genes coded by ~14 kbp as in *T. brucei* or *Leishmania tarentolae* (its closest neighbors with completely sequenced mtDNA; 17), and a single ~250-bp gene fragment per chromosome, the number of distinct mitochondrial chromosomes in *Diplonema* would amount to as many as ~56. This is in stark contrast to the majority of eukaryotes, which possess only a single type of mitochondrial chromosome; notable exceptions

are *Amoebidium parasiticum* (10) and kinetoplastids (reviewed in references 8 and 56). Finally, *Diplonema* mtDNA might be among the largest mitochondrial genomes. Assuming that there are 56 different chromosomes (see above), each 6 to 7 kbp long, the total size would exceed 360 kbp, which is a size comparable to that of higher plant and ichthyosporean mtDNA (reviewed in reference 23). Obviously, comprehensive mitochondrial genome data will be required to confirm these conjectures.

Evolution of euglenozoan mtDNA structure and expression. The new data presented here enable several evolutionary issues to be revisited. One is the longstanding controversy over whether the mtDNA network of trypanosomatids is an ancestral or a derived trait in Euglenozoa. The "network-late" scenario is gaining increasing support based on several lines of evidence. First, surveys of mtDNA organization show that bodonid mtDNA circles are not interlocked (reviewed in reference 35); second, protein phylogenies imply that trypanosomatids descended from within the bodonids (59, 61); third, molecular studies indicate that catenated mtDNA is absent from euglenozoan groups diverging prior to kinetoplastids, as shown with *E. gracilis* (23) and *Diplonema* (this report; see Fig. 1).

Another contentious question has been which mtDNA conformation is primitive in kinetoplastids, supercoiled as in *Parabodo caudatus* and *Cryptobia helcis*, or relaxed circular as in all other kinetoplastids (35; see compilation in Table 1). Our finding of relaxed-circular mtDNA for *Diplonema* strengthens the notion that this conformation represents the primitive state in Euglenozoa. Finally, the multichromosome mtDNA of kinetoplastids has been commonly considered a recent deviation, but two observations, the numerous mtDNA circles in *Diplonema* described here and the recently reported highly heterogeneous collection of small DNA molecules recovered from *E. gracilis* mtDNA, have brought this view into question (23). Therefore, we posit that a multipartite genome structure emerged as early as in the common ancestor of Euglenozoa (see Fig. 1).

What are the implications of this hypothesis for the ancestral gene structure in this lineage? Did euglenozoan mitochondrial genes originally resemble those of kinetoplastids (clustered contiguous cryptogenes plus gRNA genes located on multiple molecules) or, rather, those of *Diplonema* (discontinuous genes spread over multiple molecules)? Both scenarios are conceivable. The kinetoplastid state could have evolved from a *Diplonema*-like state by integration of all contiguous, reverse-transcribed mRNAs and rRNAs into a single chromosome and subsequent transformation of the initial gene fragments into gRNA genes. Alternatively, *Diplonema*-like gene fragments may have originated from compact kinetoplastid-type genes via recombination events, followed by the loss of the contiguous gene versions. Obviously, molecular mtDNA data from a broad taxon diversity are required to shed light on the evolution of gene organization and genome architecture in Euglenozoa.

Outlook. The gene and genome structure described here are novel and intriguing per se. In addition, these insights enable us to devise new approaches for deciphering the highly repetitive *Diplonema* mtDNA. Work is in progress to generate comprehensive mitochondrial genome and transcriptome data

from *Diplonema*, which will be instrumental for uncovering the enigmatic mechanism of gene expression in this system.

ACKNOWLEDGMENTS

We thank J. Janáček (Institute of Physiology, Prague, Czech Republic), D. Novotná, and P. Masarová (Institute of Parasitology, České Budějovice, Czech Republic) for help with 3D reconstruction, O. Benada (Institute of Microbiology, Prague, Czech Republic) for assistance in electron microscopy of mtDNA, D.A. Maslov and S. Yasuhira (University of California) for kindly providing a *D. papillatum* genomic clone, subclones, and preliminary sequence data; D. Tremblay for his contribution to mtDNA localization at an early stage of this project; A. G. B. Simpson (Dalhousie University, Halifax, Nova Scotia, Canada) for helpful discussions on mitochondrial ultrastructure; and B. F. Lang, A. Hauth, and L. Koski (Université de Montréal) for helpful discussions on the manuscript.

This research was supported by grants 5022302 and Z60220518 from the Grant Agency of the Czech Academy of Sciences and grant 6007665801 from the Ministry of Education of the Czech Republic to J.L., and a Special Project grant (SP-34) from the Canadian Institutes of Health Research (CIHR), Genetics Program, to G.B. The Canadian Institute for Advanced Research (CIAR), Program for Evolutionary Biology, provided salary and interaction support to G.B. Equipment used in the context of this project was financed by CIHR (MOP-15331) and Genome-Canada/Genome-Quebec (PEP project).

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Bendich, A. J. 1993. Reaching for the ring: the study of mitochondrial genome structure. *Curr. Genet.* 24:279-290.
- Bloom, D., A. de Haan, M. van den Berg, P. Sloof, M. Jirák, J. Lukes, and R. Benne. 1998. RNA editing in the free-living bodonid *Bodo saltans*. *Nucleic Acids Res.* 26:1205-1213.
- Bloom, D., A. de Haan, J. van den Berg, M. van den Berg, P. Sloof, M. Jirák, J. Lukes, and R. Benne. 2000. Mitochondrial minicircles in the free-living bodonid *Bodo saltans* contain two gRNA gene cassettes and are not found in large networks. *RNA* 6:121-135.
- Boer, P. H., and M. W. Gray. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55:399-411.
- Bonan, L. 1993. *Trans* splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* 7:40-46.
- Borst, P., F. Fase-Fowler, P. J. Weljers, J. D. Barry, L. Tedley, and K. Vickerman. 1985. Kinetoplast DNA from *Trypanosoma nax* and *T. congolense*. *Mol. Biochem. Parasitol.* 15:129-142.
- Borst, P., and A. M. Kroon. 1969. Mitochondrial DNA: physicochemical properties, replication, and genetic function. *Int. Rev. Cytol.* 26:167-190.
- Brugerolle, G., J. Lom, E. Nohýnková, and L. Joyon. 1979. Comparaison et évolution des structures cellulaires chez plusieurs espèces de bodonidés et cryptobidés appartenant aux genres *Bodo*, *Cryptobia* et *Trypanoplasma* (Kinetoplastida, Mastigophora). *Protistologica* 15:197-221.
- Burger, G., L. Forget, Y. Zhu, M. W. Gray, and B. F. Lang. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl. Acad. Sci. USA* 100:892-897.
- Burger, G., M. W. Gray, and B. F. Lang. 2003. Mitochondrial genomes—anything goes. *Trends Genet.* 19:709-716.
- Burger, G., Y. Zhu, T. G. Littlejohn, S. J. Greenwood, M. N. Schnare, B. F. Lang, and M. W. Gray. 2000. Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J. Mol. Biol.* 297:365-380.
- Cavaller-Smith, T. 1981. Eukaryote kingdoms: seven or nine? *Biosystems* 14:461-481.
- Clayton, D. A., C. A. Smith, J. M. Jordan, M. Teplitz, and J. Vinograd. 1968. Occurrence of complex mitochondrial DNA in normal tissues. *Nature* 220:976-979.
- Cowan, R., C. M. Collis, and G. W. Grigg. 1987. Breakage of double-stranded DNA due to single-stranded nicking. *J. Theor. Biol.* 127:229-245.
- Doojles, D., I. Chaves, R. Kieft, A. Dirks-Mulder, W. Martin, and P. Borst. 2000. Base J originally found in Kinetoplastida is also a minor constituent of nuclear DNA of *Euglena gracilis*. *Nucleic Acids Res.* 28:3017-3021.
- Estévez, A. M., and L. Simpson. 1999. Uracine insertion/deletion RNA editing in trypanosome mitochondria—a review. *Gene* 240:247-260.
- Farmer, M. A., and R. E. Triemer. 1994. An ultrastructural study of *Leptomonas applanatum* (Preisig) n.g. (Euglenida). *J. Eukaryot. Microbiol.* 41:112-119.
- Feagin, J. E. 1994. The extrachromosomal DNAs of apicomplexan parasites. *Annu. Rev. Microbiol.* 48:81-104.

20. Forget, L., J. Ustinova, Z. Wang, V. A. Huss, and B. F. Lang. 2002. *Hydrophyllum curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Mol. Biol. Evol.* 19:310-319.
21. Funes, S., E. Davidson, A. Reyes-Prieto, S. Magallón, P. Herion, M. P. King, and D. González-Halphen. 2002. A green algal apicoplast ancestor. *Science* 296:2155.
22. Gill, J. I., and M. L. Gully. 1994. Immunoglobulin and T-cell receptor gene rearrangement. *Hematol. Oncol. Clin. N. Am.* 8:751-770.
23. Gray, M. W., B. F. Lang, and C. Burger. 2004. Mitochondria of protists. *Annu. Rev. Genet.* 36:477-524.
24. Griesman, K. 1914. Ueber marine Flagellaten. *Arch. Protistenkd.* 32:1-78.
25. Hajduk, S. L., A. M. Siqueira, and K. Vickerman. 1986. Kinoplast DNA of *Bodo caudatus*: a nonconcatenated structure. *Mol. Cell. Biol.* 6:4372-4378.
26. Heinonen, T. Y., M. N. Schnare, P. G. Young, and M. W. Gray. 1987. Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. Biol. Chem.* 262:2879-2887.
27. Jahn, C. L., and L. A. Klobutcher. 2002. Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.* 56:489-520.
28. Jandáček, J., and L. Kubínová. 1998. 3D reconstruction of surfaces captured by confocal microscopy. *Acta Stereol.* 17:259-264.
29. Kent, M. L., R. A. Elston, T. A. Norad, and T. K. Sawyer. 1987. An *Isomera*-like flagellate (*Protozoa: Mastigophora*) infection in larval geoduck clams, *Panope abruptus*. *J. Invertebr. Pathol.* 50:221-229.
30. Kivic, P. A., and P. L. Walne. 1984. An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. *Origins Life* 13:269-288.
31. Laatsch, T., S. Zauner, B. Stoebe-Maler, K. V. Kowalik, and U.-G. Maier. 2004. Plasmid derived single gene minicircles of the dinoflagellate *Ceratium hirundinella* are localized in the nucleus. *Mol. Biol. Evol.* 21:1318-1322.
32. Larsen, J., and D. J. Patterson. 1990. Some flagellates (*Protista*) from tropical marine sediments. *J. Nat. Hist.* 24:801-937.
33. Liang, X. H., A. Hartman, S. Ulliel, and S. Michaeli. 2003. *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. *Eukaryot. Cell* 2:830-840.
34. Lukeš, J., G. J. Arts, J. van den Burg, A. de Haan, F. Opperdoes, P. Sloof, and R. Benne. 1994. Novel pattern of editing regions in mitochondrial transcripts of the cryptobid *Typanoplasma borexii*. *EMBO J.* 13:5086-5098.
35. Lukeš, J., D. L. Guillebride, J. Votýpka, A. Zilková, R. Benne, and P. T. Englund. 2002. Kinoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell* 1:495-502.
36. Lukeš, J., M. Jirák, N. Avilyanov, and O. Benada. 1998. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helica*. *EMBO J.* 17:838-846.
37. Madison-Antenucci, S., J. Grams, and S. L. Hajduk. 2002. Editing machines: the complexities of trypanosome RNA editing. *Cell* 108:435-438.
38. Maslow, D. A., L. Hollar, P. Haghighat, and P. Newsham. 1998. Demonstration of mRNA editing and localization of guide RNA genes in kinetoplast-mitochondria of the plant trypanosomatid *Phytomonas serpens*. *Mol. Biochem. Parasitol.* 93:225-236.
39. Maslow, D. A., and L. Simpson. 1994. RNA editing and mitochondrial genomic organization in the cryptobid kinetoplastid protozoan *Typanoplasma borexii*. *Mol. Cell. Biol.* 14:8174-8182.
40. Maslow, D. A., S. Yamahira, and L. Simpson. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. *Protist* 150:33-42.
41. Michel, F., and B. Dujon. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial, chloroplast- and nuclear-encoded members. *EMBO J.* 2:33-38.
42. Montegut-Fellmer, A. E., and R. E. Triemer. 1994. Phylogeny of *Diplonema ambulator* (Larsen and Patterson). I. Homologies of the flagellar apparatus. *Eur. J. Protistol.* 30:227-237.
43. Moreira, D., P. López-García, and K. Vickerman. 2004. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int. J. Syst. Evol. Microbiol.* 54:1861-1875.
44. Myler, P. J., D. Glick, J. E. Feglia, T. H. Morales, and K. D. Stuart. 1993. Structural organization of the maxicircle variable region of *Typanosoma brucei*: identification of potential replication origins and topoisomerase II binding sites. *Nucleic Acids Res.* 21:687-694.
45. Nédélec, A. M., R. W. Lee, C. Lemieux, M. W. Gray, and G. Burger. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res.* 10:819-831.
46. Nosek, J., and L. Tomášová. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr. Genet.* 44:73-84.
47. Patterson, D. J. 1994. Protozoa: evolution and Systematics, p. 1-14. In K. Hausmann and N. Hillmann (ed.), *Progress in protozoology*. Gustav Fischer Verlag, Stuttgart, Germany.
48. Pearson, W. R. 2000. Flexible sequence similarity searching with the FASTA program package. *Methods Mol. Biol.* 132:185-219.
49. Pellegrini, M. 1980. Three-dimensional reconstruction of organelles in *Euglena gracilis* Z. I. Qualitative and quantitative changes of chloroplasts and mitochondrial reticulum in synchronous photoautotrophic culture. *J. Cell Sci.* 43:137-166.
50. Philippe, H., P. Loper, H. Brinkmann, K. Budin, A. Cernot, J. Laurent, D. Moreira, M. Muller, and H. Le Guyader. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. B* 267:1213-1221.
51. Pigram, W. J., W. Fuller, and M. E. Davies. 1973. Letter: unwinding the DNA helix by intercalation. *J. Mol. Biol.* 80:361-365.
52. Porter, D. 1973. *Isomera papillatum* sp. n., a new colorless marine flagellate: a light- and electron microscopic study. *J. Protozool.* 20:351-356.
53. Schnare, M. N., T. Y. Heinonen, P. G. Young, and M. W. Gray. 1986. A discontinuous small subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. Biol. Chem.* 261:5187-5193.
54. Schnepf, E. 1994. Light and electron microscopic observations in *Rhynchopus coenodictyonis* spec. nov., a colorless, phagotrophic euglenozoan with concealed flagella. *Arch. Protistenkd.* 144:63-74.
55. Schuster, F. L., S. Goldstein, and B. Hershenov. 1968. Ultrastructure of a flagellate, *Isomera nigrans* nov. gen. nov. sp., from a polluted marine habitat. *Protistologica* 4:141-149.
56. Shapiro, T. A., and P. T. Englund. 1995. The structure and replication of kinetoplast DNA. *Annu. Rev. Microbiol.* 49:117-143.
57. Sharp, P. A. 1987. Trans splicing: variation on a familiar theme? *Cell* 50:147-148.
58. Simpson, A. G. B. 1997. The identity and composition of the Euglenozoa. *Arch. Protistenkd.* 148:318-328.
59. Simpson, A. G. B., E. E. Gill, H. A. Callahan, R. W. Litaker, and A. J. Roger. 2004. Early evolution within kinetoplastids (Euglenozoa), and the late emergence of trypanosomatids. *Protist* 155:407-422.
60. Simpson, A. G. B., J. Lukeš, and A. J. Roger. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. *Mol. Biol. Evol.* 19:2071-2083.
61. Simpson, A. G. B., and A. J. Roger. 2004. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol. Phylogenet. Evol.* 30:201-212.
62. Simpson, L., and R. B. Emeson. 1996. RNA editing. *Annu. Rev. Neurosci.* 19:27-52.
63. Skuja, H. 1948. Taxonomie des Phytoplanktons einiger Seen in Uppland, Schweden. *Symb. Bot. Upsal.* 9:5-399.
64. Sogin, M. L., H. J. Elwood, and J. H. Gunderson. 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc. Natl. Acad. Sci. USA* 83:1383-1387.
65. Stodola, J. 1967. Encyclopedia of water plants. TFH Publications, Inc., Jersey City, NJ.
66. Stoths, P., M. Jirák, and J. Lukeš. 2001. Polykinetoplast DNA structure in *Dinastigella typaniformis* and *Dinastigella mimosa* (Kinetoplastida). *Mol. Biochem. Parasitol.* 113:323-326.
67. Sturm, N. R., and L. Simpson. 1990. Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from *Leishmania tarentolae* mitochondria: RNA editing intermediates. *Cell* 61:871-878.
68. Tessier, L. H., H. van der Spek, J. M. Gualberto, and J. M. Grienenberger. 1997. The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. *Curr. Genet.* 31:208-213.
69. Triemer, R. E., and M. A. Farmer. 1991. An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. *Protoplasma* 164:91-104.
70. Triemer, R. E., and M. A. Farmer. 1991. An ultrastructural organization of the heterotrophic euglenids and its evolutionary implications, p. 185-204. In D. J. Patterson and J. Larsen (ed.), *The biology of free-living heterotrophic flagellates*. Clarendon Press, Oxford, United Kingdom.
71. Turmel, M., C. Lemieux, G. Burger, B. F. Lang, C. Otis, I. Plante, and M. W. Gray. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell* 11:1717-1730.
72. Vickerman, K. 1977. DNA throughout the single mitochondrion of a kinetoplastid flagellate: observations on the ultrastructure of *Cryptobia vaginabae*. *J. Protozool.* 24:221-233.
73. von der Heyden, S., E. E. Chao, K. Vickerman, and T. Cavalier-Smith. 2004. Ribosomal RNA phylogeny of bodonid and diplomemid flagellates and the evolution of euglenozoa. *J. Eukaryot. Microbiol.* 51:402-416.
74. Xu, C. W., J. C. Hines, M. L. Engel, D. G. Russell, and D. S. Ray. 1996. Nucleus-encoded histone H1-like proteins are associated with kinetoplast DNA in the trypanosomatid *Cnithidia fasciculata*. *Mol. Cell. Biol.* 16:564-576.
75. Yasuhira, S., and L. Simpson. 1995. Minicircle-encoded guide RNAs from *Cnithidia fasciculata*. *RNA* 1:634-643, 1995.
76. Yasuhira, S., and L. Simpson. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and *hsp60*. *J. Mol. Evol.* 44:341-347.
77. Yurchenko, V., R. Hobza, O. Benada, and J. Lukeš. 1999. *Typanosoma avium*: large minicircles in the kinetoplast DNA. *Exp. Parasitol.* 92:215-218.
78. Zilková, A., M. Vancová, M. Jirák, and J. Lukeš. 2003. *Cruzella marina* (*Bodonina*, Kinetoplastida): nonconcatenated structure of polykinetoplast DNA. *Exp. Parasitol.* 104:159-161.

ERRATUM

Unique Mitochondrial Genome Structure in Diplonemids, the Sister Group of Kinetoplastids

William Marande, Julius Lukeš, and Gertraud Burger

Université de Montréal, Robert-Cedergren Centre for Bioinformatics and Genomics, Department of Biochemistry, 2900 Boulevard Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada; Institute of Parasitology, Czech Academy of Sciences and Faculty of Biology, University of South Bohemia, 37005 České Budějovice, Czech Republic; Program in Evolutionary Biology, Canadian Institute for Advanced Research, Toronto, Ontario, Canada

Volume 4, no. 6, p. 1137–1146, 2005. Page 1144, Table 1: the first two entries in the “Coding content(s)” column should read “Genes, gRNAs?”

2.2. Article 2 : Un seul exon par chromosome chez *Diplonema papillatum*

par William Marande et Gertraud Burger

en préparation pour le journal PNAS

Single-exon chromosomes in the mitochondrial genome of diplomonids

William Marande¹ and Gertraud Burger^{1,2}

¹Université de Montréal, Department of biochemistry, Robert-Cedergren Centre for Bioinformatics and Genomics, 2900 Edouard-Montpetit, Montreal, Quebec, H3T 1J4 Canada

²Corresponding author; email [REDACTED]

Keywords: mtDNA, gene fragmentation, U-insertion RNA editing, trans-splicing, protist, Euglenozoa

Abstract

Mitochondrial genomes of kinetoplastids are unusual in that they consist of a maxicircle and numerous minicircle chromosomes. The maxicircle bears conventional genes yet in encrypted fashion, while minicircles encode guide RNAs that direct uridine insertion/deletion editing of primary transcripts to generate translatable mRNAs. To understand the evolution of such a remarkable system, we are investigating diplomonids, the sistergroup of kinetoplastids.

As we have reported previously, the mtDNA *Diplonema papillatum* is composed of numerous, circular, 6- and 7-kbp chromosomes, termed A-class and B-class chromosomes. Here we show that each of these chromosomes carries a single gene fragment (exon) of 60-355 bp length. So far, we have identified 49 exons from a total of nine genes by sequencing ~250 kbp genomic and ~50 kbp cDNA. The *cox1* gene for example, which codes for subunit 1 of the cytochrome oxidase, is split up in nine separate exons, seven of which reside on chromosomes of class A and two of class B. We demonstrate that the contiguous *cox1* transcript is generated via a trans-splicing-like process, whose mechanism is yet undescribed. We also discovered RNA editing reminiscent of that in kinetoplastids, but featuring U insertions only and occurring but rarely. These findings suggest that multi-chromosome mtDNAs as well as U-based RNA editing originated in the common ancestor of diplomonids and kinetoplastids.

Taken together, the unique mitochondrial gene structure and expression of *Diplonema* makes this organism a prime candidate for the discovery of new strategies for genome rearrangements and novel mechanisms for gene expression.

Introduction

Mitochondrial genomes are generally thought to be small circular molecules, as initially described for human (Borst & Kroon, 1969). Yet, studies across eukaryotes, particularly protists, uncovered a broad range of mtDNA shapes: truly circular or covalently closed, linear monomeric, and linear concatenated designated *circular-mapping*, because such molecules appear circular in restriction analysis and DNA sequencing.

Typically, the mitochondrial genome consists of a single chromosome (in multiple copies), but numerous deviations are known today. Several (<10) chromosomes are found in the chytrid fungus *Spizellomyces punctatus* (Burger et al., 2003b), the nematode *Globodera* (Armstrong et al., 2000), the cnidarian *Hydra* (Pont-Kingdon et al., 2000), and in *Dicyema*, an animal of uncertain affiliation (Watanabe et al., 1999). Moreover, mitochondrial chromosomes in the hundreds occur in kinetoplastids (Shapiro & Englund, 1995; Lukes et al., 2002) and in a unicellular relative of animals, *Amoebidium parasiticum* (Burger et al., 2003a).

The number of genes contained in mtDNA varies from as few as five in the apicomplexans parasite *Plasmodium* up to 100 in jakobid flagellates (for reviews, see (Lang et al., 1999) and (Gray et al., 2004)). These genes are involved in at most five different biological processes: invariantly in respiration, oxidative phosphorylation and translation, and occasionally in transcription, RNA maturation and protein import. Mitochondrial introns, which in some organisms make up more than 70% of the genome, belong to either group I or group II.

Remarkably, the study of mitochondria has revealed several novel mechanisms of gene expression, which later were detected in the nucleus as well. One such innovative mechanism is trans-splicing of exons encoded in distant regions and often on different strands of the genome (Chapdelaine & Bonen, 1991; Malek et al., 1997; Malek & Knoop, 1998). Another intriguing phenomenon first discovered in mitochondria is RNA editing, a process that ‘corrects’ the sequence of a transcript. Several fundamentally different molecular mechanisms of RNA editing operate in mitochondria (reviewed in (Brennicke et al., 1999; Gott & Emeson, 2000)). C-to-U

substitution editing involves nucleotide modification and occurs in plant organelles (Steinhauser et al., 1999); tRNA editing replaces mispaired nucleotides with pairing ones and is found in diverse eukaryotes (Bullerwell & Gray, 2005); global substitution editing of unknown mechanism takes place massively in slime mold mitochondria; and finally uridine insertion/deletion editing (U-indel editing), which is templated by small guide RNAs (gRNAs), is restricted to kinetoplastids (Estevez & Simpson, 1999).

In eccentricity, mitochondria of kinetoplastids clearly stand out, as they display not only the most unusual genome architecture, but their gene expression mechanism is unconventional as well. It is not known how these particularities arose and whether they are shared by the other taxa of the Euglenozoa. This set the stage to investigate the sistergroup of kinetoplastids, the diplomemids.

Here we report that the free-living flagellate *Diplonema papillatum* has a multi-chromosome mitochondrial genome and that all genes are fragmented with a single exon per chromosome. Using the *cox1* gene as example, we demonstrate that its mRNA is generated by a trans-splicing-like process. In addition, we pin down a case of RNA editing reminiscent of U-indel editing known from kinetoplastids.

Results

In a previous study (Marande et al., 2005), we demonstrated by gel electrophoresis and Southern hybridization that *Diplonema* possesses numerous circular mitochondrial chromosomes of two sizes: 6 kbp, the preponderate class A, and 7 kbp, the minority class B. The *cox1* gene was shown to be fragmented in at least two pieces (exons), which are encoded on different chromosomes. Based on this information, we decided to analyze the genome and transcripts of *Diplonema* mitochondria, as described in the following.

1. Cloning of mitochondrial chromosomes

Two cloning approaches were employed. One exploits our previous finding that one portion of *cox1* (now exon #9; see Figure 3 A) is located on an A-class chromosome, while another (now exon #4) is encoded on a B-class chromosome. The circular chromosomes on which these exons reside were amplified by inverse PCR using primer pairs that bind to these exons and face 'outward'. The products obtained had the expected size, i.e., 6 kbp for the A-class chromosome and 7 kbp for the B-class chromosome. Another cloning approach involved digestion of *Diplonema* mtDNA with the restriction enzyme XhoI, which linearizes all chromosomes of B-class and ~50% of class A chromosomes (the other half of A class chromosomes is cut into two fragments, 2.9 and 3.1 kbp; data not shown). Both the PCR products and the XhoI-linearized DNA were ligated into a bluescript vector.

We succeeded to clone ~20 complete chromosomes of the A-class, but for class B, cloning of a complete chromosome failed. Among ~1,500 insert-containing clones obtained from class B DNA, none contained the expected 7-kbp insert, but instead only smaller ones (0.2 to 5.5 kbp). Apparently, the sequence is not tolerated and/or is excised once transformed into *E. coli*.

2. Sequence determination of A-class and B-class chromosomes

Six individual A-class chromosomes were sequenced completely, four of which are cloned PCR products and two are XhoI-linearized molecules. These chromosomes carry exon #9 of *cox1* (*cox1-e9*), *nad7-e5*, *nad7-e6*, *nad7-e8*, and a putative exon, designated A57. In the following, we will refer to these chromosomes as A3207:*cox1-e9*, and so on (Table 1). The A-class chromosomes sequenced have a size range of 5.754-5.852 kbp.

Since complete B-class chromosomes could not be cloned, we used the shotgun approach to clone and sequence the PCR product of *cox1-e4* carrying chromosomes. Note that the PCR products derive from a population of different molecules (see below). Hence, the assembled

sequence represents a ‘composite’ chromosome of 7.182 kbp that will be referred to as B:cox1-e4.

Based on the sequence conservation between chromosomes in each class, we designed primers for further PCR amplification of *Diplonema* mtDNA. These PCR products and 2.9 and 3.1-kbp XhoI restriction fragments (see above) were cloned and sequenced. To date, we have determined 249,071 bp, a data body sufficient to extract a global picture of the *Diplonema* mitochondrial genome. Sequences deposited in GenBank are listed in the Material and Methods section.

3. Makeup of A-class and B-class chromosomes

The average A+T content of *Diplonema* mtDNA is unusually low (46%), which is in agreement with the buoyant density estimation reported in previous studies (Maslov et al., 1999); no difference in nucleotide composition is seen between A-class and B-class chromosomes (for details on the nucleotide content, see Table 1). Both chromosome classes have a similar structure in that they are made up of two parts: (i) a large constant region (95 % of the total length) of nearly identical sequence between members of the same chromosomes class; and (ii) a short “cassette” (5% of the total length) that includes a particular exon and distinguishes individual chromosomes from one another (Figure 1).

The constant region of A-class chromosomes. The constant regions of A-class chromosomes are 5491-5544 bp long. Along the sequence, the nucleotide composition varies notably with two clearly distinct patterns (Figure 2 A). A ~1.5 kbp-long region alternates regularly between ~100 nt-long stretches predominated by either G+C or A+T (region I in Figure 2A). In the residual portion of the chromosome, the G+C content is essentially always above 50%. The GC skew peaks in a region corresponding to position 600 in chromosome A3207, suggesting the location of the replication origin.

Repeats (>12 nt) are moderate in size and recurrence, and oriented all in the same direction. Eight different motifs are present. Five (repa1 to repa5; length 45-132 nt) are arranged in two-copy tandems adjacent to cassettes. Three additional motifs (repa6 to repa8; 45-60 nt) occur in two or three copies dispersed over the chromosome (Table 2).

Multiple sequence alignments of six complete chromosomes showed ~125 occurrences of single nucleotide substitutions (Table 3). Any pair of A-class chromosomes differs by 34 to 65 positions, independent of whether the two chromosomes carry the same or a different cassette. The proportion of nucleotide transitions (53 %) is slightly exceeding that of transversions (47%). Generally, substitutions are distributed equally over the constant regions at 1.25 substitutions per 100 bp, with the exception of one region (corresponding to positions 4,800-4,900 in chromosome A3207), which accumulates 19 sites of substitution per 100 bp (region III in Figure 2A). Adjacent polymorphic sites are moderately linked with a maximum linkage distance of 30 nt.

Indels are rare and usually small (<4 nt), mostly involving a single nucleotide and occurring in tandem repeat regions. One indel (35 bp; position 4,393 in chromosome A3207) is noteworthy as it occurs in two chromosomes carrying different exons (A3207:cox1-e9 and A4005:nad7-e8), but only in one of the two chromosomes with identical exons (A3207:cox1-e9, but not A3208:cox1-e9). Most conspicuous is a region (position 2560 in chromosome A3207) displaying four different combinations of two repeat motifs, repa1 and repa8, resulting in a region that varies in size from 300 to 560 bp between different chromosomes.

The constant region of B-class chromosomes. This region is 6.872 kbp long and resembles that of A-class chromosomes in nucleotide composition, repeat type and frequency, and the position of the likely replication origin. The only notable difference is an additional 1.5-kbp stretch (approximate position 1,200-2,700 in chromosome B:cox1-e4) of predominant G+C content (54%). This stretch coincides with a tandem array of five copies of repeat motif repb1 (Table 4 and region IV in Figure 2C).

Sequence conservation of constant regions within a class and across classes. To determine the degree of variations between the constant regions of A-class chromosomes, we compared individually sequenced molecules. For B-chromosomes, we compared the readings that gave rise to the composite chromosome B:cox1-e4 together with 25 kbp partial sequences (see above). The constant regions of chromosomes from the same class is highly conserved (>97%) and the differences are more often substitutions (2 %) than indels (0.8 %). It should be noted that the degree of sequence variation reported is not due to PCR mis-incorporation or a polyclonal culture, because the same variability is observed when comparing directly cloned material with cloned PCR products, and mtDNA extracted from the original culture with that from a culture recently rendered monoclonal.

Finally, sequence comparison of A-class and B-class constant regions shows that ~2.6 kbp are essentially identical, representing 47% and 38% of these regions (see Figure 2).

The cassettes of chromosomes. Cassettes range from 192 to 357 bp for A-class chromosome and 174 to 458 bp for B-class chromosomes. Cassettes include stretches of 65 to 350 bp that share significant sequence similarity with known mtDNA-encoded genes. We will refer to these coding stretches as 'exons'. The sequence surrounding the exons, termed 'exon-flanking regions' is variable in length: 9 to 246 on the 5' side and 1 to 192 bp on the 3' side (Figure 1). Until now, we have sequenced 99 different cassettes, 80 on class A and 19 on class B chromosomes. In 49 cassettes, we identified exons of conserved genes (Table 6 and Table 7). The other cassettes are thought to contain exons as well, but sequence similarity to known genes may be too low to be readily recognizable.

4. Mitochondrial cDNA sequencing

Diplonema mitochondrial transcripts are poly-adenylated just like transcripts of nuclear genes. Therefore, we constructed a cDNA library of polyA-RNA purified from total RNA, using the SMART technique. Among the ~5,000 cDNA clones sequenced, 5 % (250) were of

identifiable mitochondrial origin. These mitochondrial ESTs assembled into a total of 40 contigs. We also generated cDNAs specifically from *cox1* and *nad7* transcripts by employing nested RT-PCR (two subsequent RT-PCR experiments with nested pairs of gene-specific primers, see Methods for details).

Based on the cDNA sequences, we inferred that *Diplonema* mitochondria use the conventional mitochondrial genetic code with TGA-Trp and GTG as additional start codon. Codon usage is unusual in that it is strongly biased against TTT and AAA (Table 5). At least two transcripts (*atp6* and *cox3*) end a few nucleotides upstream of the stop codon, similar as reported for plant mitochondria (Raczynska et al., 2006).

5. Gene structure

Comparison of cDNA and genomic DNA sequence allowed to pinpoint exactly the exon-exon boundaries and to annotate precisely the exons within the cassettes. Exons do not overlap and their concatenated sequence is identical with the cDNA sequence (with one exception, see later). The number of exons per gene ranges from 3 to >10 (Table 7). The contribution of A-class and B-class exons to each gene is similar across all genes, with a majority of A-class exons, and at most two B-class exons for a given gene. The *cox1* gene consists of nine exons, seven of which (all except exon #1 and #4) are encoded on A-class chromosomes (Figure 3A).

6. RNA expression and maturation of *cox1*

In the cDNA library, we encountered not only mature mRNAs, but also several transcript intermediates. For *cox1*, we found intermediates consisting of either exons #2, #3, #4, #5, #6, #7, #8, #9 (designated further on as exons #2..#9), exons #5..#9, or exon #9 alone (Figure 3 A). Further, certain intermediates start precisely at the 5' exon boundary while others carry a 5' untranslated region (UTR) of 220-700 nt length.

To better understand how the contiguous *cox1* mRNA is generated from multiple small exons, we conducted Northern hybridizations of polyA RNA. A *cox1*-e9 probe lit up nine distinct bands, the mature *cox1* mRNA (1.6 kb), an RNA species of ~250 nt (the size of exon #9) and seven bands of intermediate sizes (Figure 3 B, lane 1). This confirms the cDNA data, showing that exon #9 is transcribed individually, polyadenylated individually, and ligated with other exons to produce the mature mRNA.

A *cox1*-e5 probe revealed only five bands in Northern hybridization (Figure 3 B, lane 3), the mature *cox1* mRNA and four smaller transcripts between 1.45 kb and 0.8 kb in size. No transcript of the size of exon #5 could be detected, also after extended exposure. Apparently, transcripts of internal exons are not poly-adenylated. The size of the four smaller bands is consistent with transcripts including exons #2..#9, exons #3..#9, exons #4..#9 and exons #5..#9.

Incidentally, Northern hybridizations performed with different poly-A RNA preparations yielded unequal relative signal intensities of the 1.2-kb and 1.45-kb bands. This is likely due to different steady-state concentrations of the precursors, depending on the individual reaction kinetics of each step in the cascade.

Northern hybridization conducted with total RNA showed a much different pattern. A *cox1*-e5 probe lit up among others a band of 0.2 kb (Figure 3 B, lane 2), the expected size of an exon#5 transcript. This confirms that exon #5 is separately transcribed, but not polyadenylated. In addition, a *cox1*-e4 probe (Figure 3 B, lane 4 and 5) hybridized to four extra bands of intermediate size, not seen in experiments conducted with polyA-RNA. Therefore, precursor transcripts made up only from internal exons do exist. This implies that there is more than a single pathway for building a *cox1* mRNA.

To summarize, these experiments show that (i) exons are transcribed individually, (ii) the terminal exon transcript is the only one that undergoes polyadenylation, and (iii) the contiguous *cox1* mRNA is generated by exon joining.

7. RNA editing in *cox1* gene

Thorough comparison of the *cox1* genomic and cDNA sequences revealed an inconsistency in one location. At the junction of exons #4 and exon #5, six Ts are present in the cDNA but not in the corresponding genomic sequence (Figure 4 A). This insertion gives rise to three codons, CTT, TTT and TCG, specifying the residues Leu-Phe-Ser in the deduced *cox1* protein. Within the ~250 kbp of genomic sequence, there is not a single occurrence of six contiguous Ts (or As if on the complementary strand). Within the 34 kbp of mitochondrial cDNA sequence, the T-sextet in *cox1* is the only one observed.

Multiple alignment of *cox1* proteins from different species shows that the Leu-Phe-Ser tripeptide falls in a region that is relatively well conserved except for fast evolving species such as *Chlamydomonas* and *Trypanosoma* (Figure 4 B).

DISCUSSION

1. Large genome size and low coding density

The number of exons detected in *Diplonema* mtDNA and those inferred from cDNA sequences is ~100. Consequently, the mitochondrial genome of *Diplonema* consists of at least of 100 different chromosomes, summing up to a size above 620 kbp. This is significantly larger than mtDNA of *T. brucei*, whose size is ~123 kbp when considering the maxicircle of 23 kbp plus the ~100 distinct 1-kbp long minicircles (Hong M. et al 2003). Currently, the largest completely sequenced mtDNA is that of maize (569 kbp) with a portion of ~7 % coding regions (Clifton et al., 2004). The coding density of the *Diplonema* mtDNA is only 3%, and thus comparable with that of the human nuclear genome.

2. Unique genome architecture and gene structure

A mitochondrial genome composed of multiple chromosomes is not unique to *Diplonema* (see Introduction). Neither is the finding of a substantial core sequence conserved between the chromosomes of an organelle genome. For example, conserved cores were reported before for the minicircles of kinetoplastid mitochondria (Ray, 1989) and the minicircles of dinoflagellate plastids (Zhang et al., 1999). In addition, several organelle genomes are known with single-gene chromosomes, i.e., the plastid genomes of the dinoflagellates (Nisbet et al., 2004) and the mitochondrial genome of the animal *Dicyema* (Watanabe et al., 1999). The truly unique feature of the *Diplonema* mtDNA is that each of its hundred chromosomes carries a single small exon.

3. Expression of fragmented genes

Several mechanisms allow expression of fragmented genes. First DNA splicing joins fragments prior to transcription, as found in the nucleus of ciliates. Second, gene fragments are expressed as such, leading to fragmented gene products. For instance, the mitochondrial ribosomal RNA in *Plasmodium* (Vaidya et al., 1989; Feagin et al., 1997) and *Chlamydomonas* (Boer & Gray, 1988) consist of up to eight pieces that are believed to associate by hydrogen bonds only. Examples of mitochondrial proteins that are encoded and expressed in pieces include NADH dehydrogenase subunit 1 of *Tetrahymena pyriformis* (Edqvist et al., 2000) and cytochrome oxidase subunit 2 of *Scenedesmus* (Nedelcu et al., 2000). Third, trans-splicing mediated by group II introns joins exons post-transcriptionally, as observed in organelle genomes of land plants (Kohchi et al., 1988) and green alga (Goldschmidt-Clermont et al., 1991; Knoop et al., 1997; Turmel et al., 2002). Most recently, group I intron-mediated trans-splicing was discovered in mitochondria of the animal genus *Trichoplax* (Lang, Laforest and Burger, unpublished). In these cases, genes consist of a maximum of three transcription units.

In *Diplonema*, we ruled out DNA splicing, since a contiguous *cox1* gene could not be amplified by PCR spanning more than one exon (Marande et al 2005). Likewise, fragmented gene products are implausible, because the corresponding cDNAs are contiguous. Instead, we

demonstrated here that *D. papillatum* exons are transcribed independently and joined together subsequently.

Our cDNA data show that primary transcripts of individual exons include 5'UTRs and most likely 3'UTRs as well (which we did not detect probably due to the chosen experimental approach). These extensions are removed to allow base-precise joining of exons, since we found several cDNAs whose 5' end coincides with that of an exon. As the maturation of *Diplonema* mitochondrial transcripts involves excision of exon-adjacent regions followed by exon ligation, this process is formally identical with trans-splicing.

4. Communalities and differences of diplonemid and kinetoplastid mtDNAs

Class A and class B chromosomes in *Diplonema* mtDNA are somewhat reminiscent of minicircles of trypanosomatids, although in certain species, the degree of sequence conservation of the constant (core) regions is much lower (e.g., *C. fasciculata* with only 1% difference (Macina et al., 1986)), compared to *Diplonema*.

All genes currently identified in the mtDNA of *Diplonema* are also present in that of kinetoplastids. Among the ~50 hypothetical (unassigned) exons, about 15 should be missing exons of partial genes, and another 36 would be sufficient to account for unidentified but expected genes such as *rps12*, *nad1*, *nad8*, *nad9*, *rns* and *murfs* that are otherwise found in kinetoplastid mitochondria. Transfer RNA genes appear to be absent from mtDNA of *D. papillatum* as is the case for kinetoplastids where tRNAs are imported from the nucleus. These genes must have been lost from mtDNA already in the common ancestor of the two taxa. Interestingly, the mitochondrion-located *T. brucei* tRNA-Trp undergoes C to U modification in the first position of the anticodon, which changes the anticodon from CCA (only decoding TGG-Trp) to UCA (allowing recognition of both TGG-Trp and TGA-Trp) (Alfonzo et al., 1999). Thus, the dually-localized tRNA is secondarily adapted to the mitochondrial genetic code. It would be interesting to see if the same situation also exists in *Diplonema*.

Unlike in most eukaryotes, mitochondrial mRNAs of *Diplonema* are decorated with an A-tail, likely protecting RNA from degradation and relevant for translation initiation. A more complex situation occurs in kinetoplastids. In *T. brucei* mRNA stability depends not only of the poly(A) length but also of the editing degree (Gagliardi et al., 2004). Stable polyadenylated mitochondrial transcripts are also found in animals (recently reviewed in Slomovic et al., 2005) and dinoflagellates (Chaput et al., 2002). In contrast to the above cases and nuclear transcripts in general, polyA is a signal for RNA degradation in bacteria and plant mitochondria.

5. RNA editing in *cox1* mRNA: significance and evolution

We discovered six non-encoded uridines in the *cox1* transcript, at the junction of exons #4 and #5. These nucleotides code for relevant amino acids in the protein. In the *cox1* proteins of *Bos taurus* and *Paracoccus denitrificans*, this tripeptide (residues 264-266 in *B. taurus*) is located in a loop and engages in a 3D interaction with a second loop further toward the C-terminus (residues 328-330 in *B. taurus*). Interestingly, in the proteins of *B. taurus* and *P. denitrificans*, the first loop is positively charged (Lys and Arg) and the second loop contains small (Gly) and hydrophobic (Trp) residues. In contrast, the situation is inverse in the *Diplonema* protein: small (Ser) and hydrophobic (Leu, Phe) residues make up the first loop and positively charged amino acids (Arg) the second. Apparently, divergence of the primary sequence was constrained by the conservation of the tertiary structure, only allowing domain swapping.

Although not formally proven, these non-encoded Us originate most probably from insertion RNA editing, rather than from an exon. First, exons are all significantly longer than six nucleotides; and second, we did not find a cassette including six Us or As. U-insertion editing is an attractive hypothesis to explain these additional nucleotides, because of the massive U indel editing taking place in the closest neighbor of diplomemids, the kinetoplastids. Given the scarcity of occurrence and the exclusive insertions, we assume that *Diplonema* features a primitive form of U-indel editing seen in kinetoplastids.

6. Hypothesis: trans-splicing and RNA editing in one

We showed that in *Diplonema* mitochondria, separately transcribed exons are joined together to a mature mRNA by a process similar to trans-splicing. Trans-splicing implies that exons are bounded by intron sequences that either catalyze splicing themselves as do autocatalytic group I or II introns, or recruit specialized splicing factors. Yet, conserved signatures of group I, II, and spliceosomal introns appear to be absent. In addition, neighbor exons and their flanking regions do not display obvious conserved residues nor sequence complementarity.

Alternatively, trans-splicing could be directed by particular guide RNAs that recognize cognate neighbor exon transcripts and aligns them properly (Figure 5). The hypothesis of splice-guide RNAs (sgRNAs) could accommodate not only splicing but also U-insertion editing at exon-exon boundaries. The machinery to catalyze such a process must be endowed with endonuclease, ligase and polymerase activities and be able to specifically recognize mRNA-sgRNA duplexes. The ubiquitous DNA repair system could have given rise to the editosome in kinetoplastids (Worthey et al., 2003) and in parallel, to a splice-editosome in diplonemids. Crucial to this hypothesis is the existence of sgRNAs. Work in these lines is underway.

Conclusion

The mitochondrial genome of *Diplonema* is exceptional in every aspect: genome architecture, gene structure and gene expression. Moreover, other diplonemid species of the genera *Diplonema* (see annex 1 of this thesis) and *Rhynchopus* (Joannie Roy et al., 2007) display these features as well, providing valuable material for comparative studies. An intriguing question bears on the genetic makeup of mtDNA in the common ancestor kinetoplastids and diplonemids. The answer may be found in euglenids, a group that diverged prior to the kinetoplastid/diplonemid split. Indeed, a molecular-biology survey of the poorly studied euglenids, in particular of its basally diverging and primitive members, is overdue.

MATERIAL AND METHODS

Strain, culture and mtDNA extraction

Diplonema papillatum (ATCC 50162) was obtained from American Type Culture Collection. The organism was cultivated axenically at 22°C in artificial seawater enriched with 1% fetal bovine serum (Wisent) and 0.1% tryptone. Mitochondrial DNA was extracted from the organelle fraction isolated by differential centrifugation and a sucrose gradient as described in (Marande et al., 2005); see also (Lang & Burger, 2007).

PCR

We used the TaKaRa La-PCR kit as recommended, by adding 1% DMSO and 1M Betaine to the reaction mix. The primer pairs that allow the specific amplification of A and B class chromosomes are listed in Table 1. **Nested RT-PCR** was performed using poly(A) RNA as template. First strand (cDNA) synthesis was primed with an exon-specific oligonucleotide (dp76 for *cox1-e2* and dp94 for *nad7-e4*), using the reverse transcriptase of the SMART kit (TAKARA Bio Inc.). Initial PCR amplification was conducted for 35 cycles using the first strand primer plus the primer Smart IV that attaches to the most 5' end of the transcript. One fiftieth of the product was employed as template for a second PCR amplification for 35 cycles adding another exon-specific primer (dp87 for *cox1-e1* and dp100 for *nad7-e3*) that is located upstream the first primer, plus the primer Smart IV.

Cloning and DNA sequencing

For detailed protocols, see (Burger et al., 2007). In brief, PCR products and XhoI restriction fragments were separated on agarose gel. The PCR product of chromosome B (7 kbp) was

nebulized and then size-fractionated on agarose gel. DNA fractions were cloned in the vector pBFL6 (derived from pBluescript, Stratagene). cDNA libraries from poly(A) RNA were constructed using the components of the SMART kit, including the vector pDNRLib. The method has been described in detail elsewhere (Rodriguez-Ezpeleta et al., 2006). Plasmid DNA was extracted with the QIAprep 96 Turbo Miniprep Kit (Qiagen). For sequencing reactions, we used the dye terminator technology (kit ABI Prism Big Dye™ terminators version 3.0/3.1, Perkin-Elmer). Reactions were run on an MJ BaseStation automatic sequencer. We performed primer walking in the case of chromosome A (for primers, see Table 8), otherwise random sequencing. Readings from genomic clones were assembled using the GAP software suite (Staden, 1996), while cDNA sequences were assembled using phred, phrap, and consed (Ewing et al., 1998; Gordon et al., 1998). Sequences were deposited in GenBank. Accession numbers are listed in Table 9.

Sequence analysis

For multiple sequence alignments, we used ClustalW (Thompson et al., 1994) and Muscle (Edgar, 2004). Alignments were visualized and edited with the Genetic Data Environment (GDE) (Smith et al., 1994). Direct and dispersed repeated motifs were identified using the Complex TR software (Hauth & Joseph, 2002) implemented in AnaBench (Badidi et al., 2003) (<http://anabench.bcm.umontreal.ca/anabench>). We used tools of the EMBOSS suite, accessible via AnaBench, for the following analyses: COMPSEQ for di- and tri-nucleotide frequencies; CUSP for codon usage; FREAK for nucleotide distribution, using a window length of 50. GAP (Staden, 1996) was used for determining the frequency and distribution of nucleotide substitutions (transition/transversion) and indels.

Northern hybridization

RNAs were separated on a 1.5% agarose gel + formaldehyde 6%. PCR fragments used as probes were labeled by random priming in the presence of [α - 32 P]dATP (Perkin Elmer). Hybridizations were performed at 42⁰C in 4.5X SSPE (3.6M NaCl, 0.2M sodium phosphate, and 20mM EDTA [pH7.4]), plus 50% formamide, 0.1% SDS, 0.5% dry milk and 0.1mg/ml salmon sperm.

Analysis of protein 3D interaction

The structure of cytochrome c oxidase from *Bos taurus* (2OCC, (Yoshikawa et al., 1998) and *Paracoccus denitrificans* (1AR1, (Ostermeier et al., 1997), available in the Protein DataBase (<http://www.rcsb.org/pdb/home/home.do>), were inspected with the viewer Pymol (<http://pymol.sourceforge.net/>). We considered amino acids to be in contact if the distance between them was smaller than 6 (Miyazawa & Jernigan, 1999).

Table 1. Entirely sequenced mitochondrial chromosomes

Name	Gene & module number	Cloning & sequencing strategy	Length of element (bp)					% A+T content	
			chromo- some	cassette	upstream- flanking	gene exon	down- stream- flanking	constant region	cassette
A3207	<i>cox1</i> -m9	cloned PCR-amplified chromosome, primer- walking	5,852	311	9	250	52	46	45
A3208	<i>cox1</i> -m9	cloned PCR-amplified chromosome, primer walking	5,802	311	9	250	52	46	45
A3216	A57	cloned XhoI-linearized chromosome, primer walking	5,754	219	?	?	?	46	43
A3217	<i>nad7</i> -m5	cloned XhoI-linearized chromosome, primer walking	5,805	284	79	193	12	46	45
A4001	<i>nad7</i> -m6	cloned PCR-amplified chromosome, primer walking	5,795	272	37	65	170	46	44
A4005	<i>nad7</i> -m8	cloned PCR-amplified chromosome, primer walking	5,763	219	34	184	1	46	43
B3209	<i>cox1</i> -m4	PCR-amplified chromosome, random library	7,182	310	43	220	47	46	47

^a The chromosome name is built from chromosome class, contig id, gene, and exon number

Table 2: Sequence repeats in constant regions of chromosomes

Chromosome	Motif length	Copy number	Arrangement	Position	Similarity between copies
A3207					
repa1	132	2 ^a	tandem-direct	2567	100%
repa2	45	2	tandem-direct	3009	85%
repa3	40	2	tandem-direct	3734	85%
repa4	65	2	tandem-direct	3991	85%
repa5	64	2	tandem-direct	4181	85%
repa6	45	3	dispersed-direct	4798; 5332; 5649	100%
repa7;	60	2	dispersed-direct	4952; 5489	85%
repa8	170	1 ^b	tandem-direct	2831	100%
B3209					
repb1	345	5	tandem-direct	1375-3100	100%
repb2	90	3	tandem-direct	3400	90%
repb3	80	4	tandem-direct	5084	80%
repb4	45	3	dispersed-direct	6116; 6662; 6980	100%

^a repa1 is repeated 3 times in chromosome A4052:cox1-e3

^b repa8 is repeated 2 times in chromosome A3234:cox1-e8

Table 3. Sequence variation in constant regions of chromosomes

Chromosomes compared	Substitutions		Indels	
	transitions / transversions		total	per 1000 bp
	total	per 1000 bp		
A3207 vs A3208 (same gene module)	37 / 24	6.4 / 4.1	4	0.7
A3216 vs A3217 (different gene module)	15 / 19	2.6 / 3.3	14	2.4
A4001 vs A4005 (different gene module)	36 / 28	6.2 / 4.8	23	4.0
A4001 vs A3216 (different gene module)	25 / 22	4.3 / 3.8	19	3.3
A4001 vs A3217 (different gene module)	30 / 24	5.2 / 4.1	20	3.4
A3216 vs A4005 (different gene module)	36 / 29	6.2 / 5.0	20	3.4
A3217 vs A4005 (different gene module)	31 / 30	5.3 / 5.22	19	3.3
All four class A with different gene module	66 / 59	11.4 / 10.2	44	7.6
B-class with same gene module	13 / 41	1.8 / 5.7	2	0.3

In A-class chromosomes, the largest indel is 48-bp long, with insertion in chromosomes A3207 (at nucleotide position 1411) and A4005, and deletion in chromosomes A3208, A3216, A32017 and A4001. The second largest indel is 36 bp long, with insertion in chromosomes A3216 and A3217, and deletion in chromosomes A3207 (at nucleotide position 3094), A3208, A4001 and A4005. In B-class chromosomes, the largest indel is 50 bp long.

Table 4. Di- and tri-nucleotide frequencies of chromosome A3207

	Di-nucleotides (%)		Tri-nucleotides (%)			
		A	C	G	T	
AA	0.92	0.07	0.34	0.26	0.26	
AC	9.38	2.65	3.31	2.00	1.42	
AG	6.68	0.77	1.95	2.01	1.91	
AT	6.31	2.09	1.54	2.55	0.14	
CA	9.26	0.39	4.27	2.43	1.98	
CC	7.04	2.99	1.40	0.96	1.69	
CG	4.68	0.44	1.35	1.08	1.79	
CT	6.17	2.27	1.20	2.42	0.28	
GA	3.57	0.17	0.56	1.45	1.38	
GC	7.16	2.67	0.77	1.35	2.38	
GG	6.81	1.59	0.53	1.45	3.25	
GT	9.38	4.99	0.75	3.32	0.32	
TA	9.54	0.29	4.21	2.36	2.68	
TC	3.57	0.96	1.56	0.38	0.68	
TG	8.77	0.77	3.33	2.24	2.43	
TT	0.73	0.17	0.09	0.48	0.00	

Di-nucleotide frequencies <1% and trinucleotide frequencies <0.1% are shaded.

Table 5: Codon frequency of the *Diplonema* mitochondrial genome

F	TTT	3	(7 %)	S	TCT	17	(8 %)	Y	TAT	14	(14 %)	C	TGT	26	(36 %)
F	TTC	43	(93 %)	S	TCC	66	(29 %)	Y	TAC	86	(86 %)	C	TGC	46	(64 %)
L	TTA	9	(2 %)	S	TCA	21	(9 %)	*	TAA	3	(50 %)	W	TGA	26	(34 %)
L	TTG	19	(5 %)	S	TCG	7	(3 %)	*	TAG	3	(50 %)	W	TGG	51	(66 %)
L	CTT	11	(3 %)	P	CCT	18	(30 %)	H	CAT	49	(42 %)	R	CGT	17	(24 %)
L	CTC	102	(26 %)	P	CCC	11	(18 %)	H	CAC	69	(58 %)	R	CGC	7	(10 %)
L	CTA	121	(31 %)	P	CCA	24	(39 %)	Q	CAA	1	(2 %)	R	CGA	3	(4 %)
L	CTG	133	(34 %)	P	CCG	8	(13 %)	Q	CAG	52	(98 %)	R	CGG	13	(19 %)
I	ATT	2	(3 %)	T	ACT	26	(15 %)	N	AAT	4	(12 %)	S	AGT	29	(13 %)
I	ATC	43	(54 %)	T	ACC	55	(32 %)	N	AAC	28	(88 %)	S	AGC	85	(38 %)
I	ATA	34	(43 %)	T	ACA	55	(32 %)	K	AAA	0	(0 %)	R	AGA	7	(10 %)
M	ATG	113	(100 %)	T	ACG	38	(22 %)	K	AAG	21	(100 %)	R	AGG	23	(33 %)
V	GTT	16	(5 %)	A	GCT	82	(35 %)	D	GAT	24	(37 %)	G	GGT	74	(37 %)
V	GTC	83	(25 %)	A	GCC	44	(19 %)	D	GAC	41	(63 %)	G	GGC	44	(22 %)
V	GTA	96	(29 %)	A	GCA	80	(34 %)	E	GAA	7	(13 %)	G	GGA	53	(26 %)
V	GTG	138	(41 %)	A	GCG	31	(13 %)	E	GAG	47	(87 %)	G	GGG	29	(14 %)

Frequencies were inferred from available (partial and complete) coding cDNA sequences (7,593 bp). Percentages are calculated as the codon frequency per amino acid.

Table 6. Detailed listing of gene-containing cassettes and cDNAs

Gene	Gene exon number	Length (bp)			Chromosome	Strand of	
		Upstream-flanking	gene exon	downstream-flanking		gene exon	cDNA contig
<i>cox1</i>	1	52	194	?	B4615	-	dp4030
	2	35	125	63	A3223	-	
	3	26	253	34	A4052	-	
	4	43	220	47	B3209	-	
	5	63	179	24	A3238	-	
	6	35	169	80	A4020	-	
	7	116	89	36	A3369	-	
	8	102	110	23	A3234	-	
	9	9	250	52	A3207	+	
<i>cox2</i>	1	137	145	26	A3235	-	dp0321
	2	?	156	?	B4614	-	
	3	57	76	155	A3215	-	
	4	?	126	?	A4380	+	
<i>cox3</i>	1	28	162	?	A4016	-	
	2	?	?	9	A3202	-	
	3	45	272	6	A4039	-	
	4	17	232	55	A3319	+	
<i>cob</i>	1	5	217	54	A3335	-	dp4278
	2	95	155	36	A3359	+	
	3	99	138	53	A4219	+	
	4	54	198	42	A4226	+	
	5	22	280	15	A4266	+	
	6	39	123	44	B4155	+	
<i>atp6</i>	(1)	?	>182	37	A4045	+	dp4033
	(2)	?	206	20	A4014	-	
	(3)	62	139	50	A4246	+	

<i>nad4</i>	(1)	43	165	25	A3366	+	dp0719
	(2)	<u>246</u>	120	31	B3511	+	
	(3)	73	169	17	A4236	+	
<i>nad5</i>	(1)	<u>0</u>	218	29	A4213	+	dp1393
	(2)	151	<u>60</u>	29	A4273	+	
	(3)	45	183	42	A3331	+	
	(4)		?		?	?	
	(5)	64	122	80	A4527	+	
	(6)	115	201	11	A3229	+	
	(7)	116	90	3	A4240	-	
	(8)	11	201	102	A3213	+	
	(9)	66	115	92	A3248	+	
	(10)	135	70	17	A3225	-	
<i>nad7</i>	1	62	200	34	A4242	-	dp4285
	2	89	75	110	A4268	+	
	3	149	133	7	A3311	-	
	4	?	193	?	A4634	?	
	5	79	193	12	A3217	-	
	6	37	66	<u>192</u>	A4035	+	
	7	31	169	53	B4147	-	
	8	34	184	<u>1</u>	A4005	-	
	9	?	75	?	A4519	+	
<i>rnl</i>	>1	?	<u>351</u>	?	B4133	?	dp0340

Shortest and longest gene modules and flanking regions are underlined. Module numbers in parentheses indicate that one or more 5' modules remain undetermined. +, -, positive or negative strand with respect to the orientation of the module in chromosome A3207.

Table 7. Gene content and structure of *D. papillatum* mtDNA

Gene ^a	Length of coding regions (bp)		Number of identified gene exons in <i>Diplonema</i>	Number of inferred gene exons in <i>Diplonema</i> ^b	
	<i>T. brucei</i> ^c	<i>D. papillatum</i>	Class A + class B	Minimum	Expected
<i>cox1</i>	1,647	1605	7+2	9	9
<i>cox2</i>	630	504	3+1	4	4
<i>cox3</i>	864	813	4+0	4	4
<i>cob</i>	1,089	1,095	5+1	6	6
<i>nad1</i>	956	?	?	-	~5?
<i>nad4</i>	1,311	>453	2+1	~8	~8
<i>nad5</i>	1,770	>1,389	10+0	~12	~12
<i>nad7</i>	1,246	1,152	8+1	9	9
<i>nad8</i>	435	?	?	-	~3?
<i>nad9</i>	348	?	?	-	~2?
<i>atp6</i>	693	>582	3+0	~4	~4
<i>rps12</i>	255	?	?	-	~2?
<i>rns</i>	650	?	?	-	~4?
<i>rnl</i>	1,300	>350	0+1	~7	~7
Total	13,194	>7,943	42+7	63	~79

^a Gene products are: *atp6*, ATP synthase subunit 6; *cob*, cytochrome b apoprotein; *cox1-cox3*, cytochrome c oxidase subunits; *nad1-nad9*, NADH dehydrogenase subunits; *rps12*, ribosomal protein S12; *rnl*, LSU ribosomal RNA; *rns*, SSU ribosomal RNA; ?, not observed but expected.

^b Numbers have been estimated assuming an average module length of 170 bp. The minimum number is based on the currently identified gene set; the expected number assumes a gene set identical to that of trypanosomatids.

^c Sizes have been calculated from GenBank acc. nr. M94286.

Table 8. Primers

Name	Sequence	Used for
dp1	GCTGTTGATGCTCAGCAGAGCGGT	Primer walking, chromosome A and B
dp2	GCACTCGCTGTGGAGTGCGTGTAT	Primer walking, chromosome A and B
dp3	GATAGCATACCATCTACCGTCCTACA	Primer walking, chromosome A
dp4	GCTCTGCTGCTCCGCTGCACAGTG	Primer walking, chromosome A
dp5	CCTCTACACGTAGTAGATGCATATC	Primer walking, chromosome A
dp6	GTAGTGCTACGAGGAGCGGTAGAT	Primer walking, chromosome A
dp7	GCTGTAGAGCTACGTGCTGCTACAT	Primer walking, chromosome B
dp8	CGTGCATCCATGCACAGCCGGTC	Primer walking, chromosome B
dp11	GGTAGATGGACGGTATGCCTA	Primer walking, chromosome B
dp15	GCATCTACATCCTACCCTCTACGT	Primer walking, chromosome B
dp17	CCGCTCTGCTGAGCATCAACAGC	Primer walking, chromosome A and B
dp18	ATACACGCACTCCACAGCGAGTGC	Primer walking, chromosome A
dp19	GCTACATGCTATCCTCTCCCGTAG	Primer walking, chromosome

		B
dp20	GCACCACAGGATGCTAGTACATGTACAGC	Primer walking, chromosome B
dp21	GCGGTGTATTGGTACTAGCAGTACGGCGT	Primer walking, chromosome B
dp22	GGGATATGCACAGGGAATGCA	Primer walking, chromosome A and B
dp40	CGTAGCAGCAGCACGTAGTGTATGCA	Primer walking, chromosome A
dp44	GGGTACAGATCCACTTGGGA	Primer walking, chromosome A
dp45	GCGTAGTAAGGTAATGGTAT	Primer walking, chromosome A
dp63	GCTAGTACCACCTCGTAGCATG	Primer walking, chromosome B
dp64	GGCATCAACGCACCTGTGCAT	Primer walking, chromosome A
dp26	GGTCTTCGGGCACCCAGAG	PCR
dp27	TACCCTACGTACACGTGCAC	PCR
dp38	GACTACCAGTATAACCACAGG	PCR
dp39	AGCATCCATGCATCTGGAGG	PCR
dp90	GCTGATGTGCAGCTCTCCCTTTGG A	PCR
dp91	GAGTGCGTCCTGCAGACCTAGGA	PCR
dp93	CTGGTGCTAGCTGAGACCTTAGG	PCR
dp94	CGCTCCGGAGTAGGACACCAGATG	PCR and 1 st -strand cDNA of nested RT-PCR for <i>nad7</i>
dp100	CCAGTGTAGTGCTCCGGAT	2 nd PCR of nested RT-PCR

		for <i>nad7</i>
dp76	GCC ATC CAC TAG CAG CCA TGC TAC	1 st -strand cDNA of nested RT-PCR for <i>cox1</i>
dp87	AAG AGC ATG AGG ATG CCA TGC	2 nd PCR of nested RT-PCR for <i>cox1</i>

Table 9. Sequences deposited in GenBank

Accession Number	Contig_id	Class	Gene, exon	Generated by
Complete chromosomes				
	3209	B	<i>cox1</i> , exon 4	PCR
	3207	A	<i>cox1</i> , exon 9	PCR
	3217	A	<i>nad7</i> , exon 5	restriction
Cassettes				
	4687	A	<i>cox1</i> , exon1	PCR
	3223-4	A	<i>cox1</i> , exon2	restriction
	3249	A	<i>cox1</i> , exon 3	restriction
	3238	A	<i>cox1</i> , exon 5	restriction
	4020	A	<i>cox1</i> , exon 6	PCR
	3369	A	<i>cox1</i> , exon 7	restriction
	3234	A	<i>cox1</i> , exon 8	restriction
Full-length cDNAs				
	dp4278	-	<i>cob</i>	library
	dp4030	-	<i>cox1</i> , +5'UTR	library and nested PCR
	dp0321	-	<i>cox2</i> , + 5'UTR	library
	dp0301	-	<i>cox3</i> , +5'UTR	library
	dp4284	-	<i>nad7</i>	library and nested PCR

References

1. Borst, P. & Kroon, A. M. (1969) *Int Rev Cytol* **26**, 107-90.
2. Burger, G., Gray, M. W. & Lang, B. F. (2003) *Trends Genet* **19**, 709-716.
3. Armstrong, M. R., Blok, V. C. & Phillips, M. S. (2000) *Genetics* **154**, 181-92.
4. Pont-Kingdon, G., Vassort, C. G., Warrior, R., Okimoto, R., Beagley, C. T. & Wolstenholme, D. R. (2000) *J Mol Evol* **51**, 404-15.
5. Watanabe, K. I., Bessho, Y., Kawasaki, M. & Hori, H. (1999) *J Mol Biol* **286**, 645-50.
6. Shapiro, T. A. & Englund, P. T. (1995) *Annu Rev Microbiol* **49**, 117-43.
7. Lukes, J., Guilbride, D. L., Votypka, J., Zikova, A., Benne, R. & Englund, P. T. (2002) *Eukaryot Cell* **1**, 495-502.
8. Burger, G., Forget, L., Zhu, Y., Gray, M. W. & Lang, B. F. (2003) *Proc Natl Acad Sci U S A* **100**, 892-7.
9. Lang, B. F., Gray, M. W. & Burger, G. (1999) *Annu Rev Genet* **33**, 351-97.
10. Gray, M. W., Lang, B. F. & Burger, G. (2004) *Annu Rev Genet* **38**, 477-524.
11. Malek, O., Brennicke, A. & Knoop, V. (1997) *Proc Natl Acad Sci U S A* **94**, 553-8.
12. Malek, O. & Knoop, V. (1998) *RNA* **4**, 1599-609.
13. Chapdelaine, Y. & Bonen, L. (1991) *Cell* **65**, 465-72.
14. Gott, J. M. & Emeson, R. B. (2000) *Annu Rev Genet* **34**, 499-531.
15. Brennicke, A., Marchfelder, A. & Binder, S. (1999) *FEMS Microbiol Rev* **23**, 297-316.
16. Steinhäuser, S., Beckert, S., Capesius, I., Malek, O. & Knoop, V. (1999) *J Mol Evol* **48**, 303-12.
17. Bullerwell, C. E. & Gray, M. W. (2005) *J Biol Chem* **280**, 2463-70.
18. Estevez, A. M. & Simpson, L. (1999) *Gene* **240**, 247-60.
19. Marande, W., Lukes, J. & Burger, G. (2005) *Eukaryot Cell* **4**, 1137-46.
20. Maslov, D. A., Yasuhira, S. & Simpson, L. (1999) *Protist* **150**, 33-42.
21. Raczynska, K. D., Le Ret, M., Rurek, M., Bonnard, G., Augustyniak, H. & Gualberto, J. M. (2006) *FEBS Lett* **580**, 5641-6.
22. Clifton, S. W., Minx, P., Fauron, C. M., Gibson, M., Allen, J. O., Sun, H., Thompson, M., Barbazuk, W. B., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R. K. & Newton, K. J. (2004) *Plant Physiol* **136**, 3486-503.
23. Ray, D. S. (1989) *Mol Cell Biol* **9**, 1365-7.
24. Zhang, Z., Green, B. R. & Cavalier-Smith, T. (1999) *Nature* **400**, 155-9.
25. Nisbet, R. E., Koumandou, L. V., Barbrook, A. C. & Howe, C. J. (2004) *Gene* **331**, 141-7.
26. Vaidya, A. B., Akella, R. & Suplick, K. (1989) *Mol Biochem Parasitol* **35**, 97-107.
27. Feagin, J. E., Mericle, B. L., Werner, E. & Morris, M. (1997) *Nucleic Acids Res* **25**, 438-46.
28. Boer, P. H. & Gray, M. W. (1988) *Cell* **55**, 399-411.
29. Edqvist, J., Burger, G. & Gray, M. W. (2000) *J Mol Biol* **297**, 381-93.
30. Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W. & Burger, G. (2000) *Genome Res* **10**, 819-31.

31. Kohchi, T., Umesono, K., Ogura, Y., Komine, Y., Nakahigashi, K., Komano, T., Yamada, Y., Ozeki, H. & Ohyama, K. (1988) *Nucleic Acids Res* **16**, 10025-36.
32. Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. & Rochaix, J. D. (1991) *Cell* **65**, 135-43.
33. Knoop, V., Altwasser, M. & Brennicke, A. (1997) *Mol Gen Genet* **255**, 269-76.
34. Macina, R. A., Sanchez, D. O., Gluschankof, D. A., Burrone, O. R. & Frasch, A. C. (1986) *Mol Biochem Parasitol* **21**, 25-32.
35. Alfonzo, J. D., Blanc, V., Estevez, A. M., Rubio, M. A. & Simpson, L. (1999) *EMBO J* **18**, 7056-62.
36. Gagliardi, D., Stepien, P. P., Temperley, R. J., Lightowlers, R. N. & Chrzanowska-Lightowlers, Z. M. (2004) *Trends Genet* **20**, 260-7.
37. Slomovic, S., Laufer, D., Geiger, D. & Schuster, G. (2005) *Mol Cell Biol* **25**, 6427-35.
38. Chaput, H., Wang, Y. & Morse, D. (2002) *Protist* **153**, 111-22.
39. Worthey, E. A., Schnauffer, A., Mian, I. S., Stuart, K. & Salavati, R. (2003) *Nucleic Acids Res* **31**, 6392-408.
40. Joannie Roy, Drahomira Faktorova, Lukes, J. & Burger, a. G. (2007) *Protists* **in press**.
41. Lang, B. F. & Burger, G. (2007) *Nat Protoc* **2**, 652-60.
42. Burger, G., Lavrov, D. V., Forget, L. & Lang, B. F. (2007) *Nat Protoc* **2**, 603-14.
43. Rodriguez-Ezpeleta, N., Teijeiro, S., Forget, L., Burger, G. & Lang, B. F. (2006) in *Methods in Molecular Biology: Methods in ESTs*, ed. Parkinson, J. (Humana Press, Totowa, NJ), Vol. in press.
44. Staden, R. (1996) *Mol Biotechnol* **5**, 233-41.
45. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res* **8**, 175-85.
46. Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res* **8**, 195-202.
47. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res* **22**, 4673-80.
48. Edgar, R. C. (2004) *BMC Bioinformatics* **5**, 113.
49. Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. (1994) *Comput Appl Biosci* **10**, 671-5.
50. Hauth, A. M. & Joseph, D. A. (2002) *Bioinformatics* **18 Suppl 1**, S31-7.
51. Badidi, E., De Sousa, C., Lang, B. F. & Burger, G. (2003) *BMC Bioinformatics* **4**, 63.
52. Yoshikawa, S., Shinzawa-Itoh, K. & Tsukihara, T. (1998) *J Bioenerg Biomembr* **30**, 7-14.
53. Ostermeier, C., Harrenga, A., Ermler, U. & Michel, H. (1997) *Proc Natl Acad Sci U S A* **94**, 10547-53.
54. Miyazawa, S. & Jernigan, R. L. (1999) *Proteins* **34**, 49-68.

Figure Legends

Figure 1: Mitochondrial chromosome architecture of *D. papillatum*: a single gene fragment per circular chromosome.

The exon region is represented by the grey box, the exon-flanking region by the grey line and the constant region by the black line. Exon length vary from 64 to 351 bp, 5' exon flanking region from 9 to 246 bp and 3' exon flanking region from 1 to 192 bp.

Figure 2: Nucleotide distribution in chromosome A and B.

(A) **Chromosome A3207**. In pink is represented, the G+C distribution and in bleu is the A+T. The region from nucleotide position 590 to 1910 alternates between ~100-nt long stretches of 60% G+C and 40% G+C (region I under brackets). The residual portion of the chromosome is homogenously G+C-rich (60-77% G+C) with only a few short (< 50 nt) stretches where the percentage decreases to 50%. Four regions (180 to 680 nt long) have a high G+C content due principally to Cs. These are located at positions 4290-4915, 5100-5450, 5500-5750, and 4340-4520 of chromosome A3207:cox1-e9 GenBank Acc.#). The cassette position is indicated as II and III represents the region where there are 19 substitutions over 100 bp. (B) Shared sequence between chromosomes A and B are the grey lines and the cassette is the grey box. Triangles represent the position of the possible replication origin. Arrows in grey line correspond to repb1 motif (C) **Chromosome B3209**. Region IV represents the region of repb1 repeated 5 times

Figure 3: *Cox1* gene expression in *D. papillatum*.

(A) Description of *cox1* transcripts identified in the cDNA library obtained from polyadenylated RNA. Exons are represented by box with its corresponding number inside. Between exon #4 and #5, 6 T are added representing the editing site. Grey boxes represent exons encoded by chromosome B-class. The discontinuous lines represent the 5' UTR and the numbers, the size in base pair. (B) Northern hybridization with *cox1*'s exon #9 (lane 1), exon #5 (lane 2 and 3), with exon #4 (lane 4 and 5) and exon #1 (lane 6) as probes on poly-adenylated RNA (polyA) or total

RNA (total) probes. The black triangles highlight the bands. In lane 5, the fourth band is not visible in this short exposition but exists. Note that in experiments with total RNA, the 1.6-kb band has a much higher relative intensity than with poly(A)RNA. The most likely reason is that the ultimate transcript maturation step involves shortening of the A-tail, with the consequence that this product is less enriched during poly(A) RNA purification.

Figure 4: RNA editing in *cox1* gene of *D. papillatum*.

(A) Comparison of genomic and cDNA sequences at the editing site. The lowercase letters represent the nucleic acids that flank the exon #4 and #5. (B) Multiple alignment of *cox1* protein sequences with a focus on the editing site (grey box on the left). Arrows represent the interaction between these amino acids with the amino acids of the grey box on the right at the 3D level.

Figure 5 : Hypothesis of trans splicing mediated by Spliced guide RNA.

Exons (yellow and orange boxes) and the sgRNA (purple bar) are transcribed from different chromosomes. The sgRNA aligns adjacent exons to allow ligation

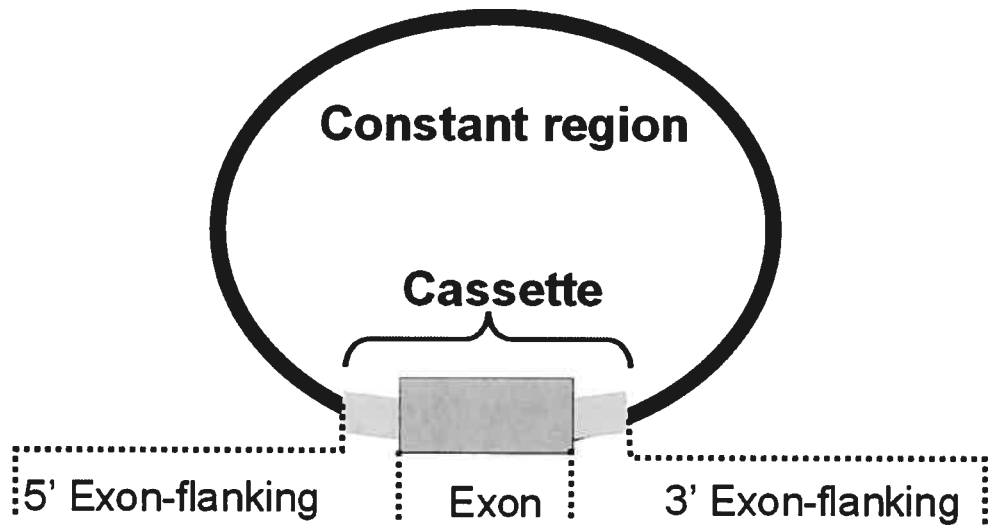


Figure 1

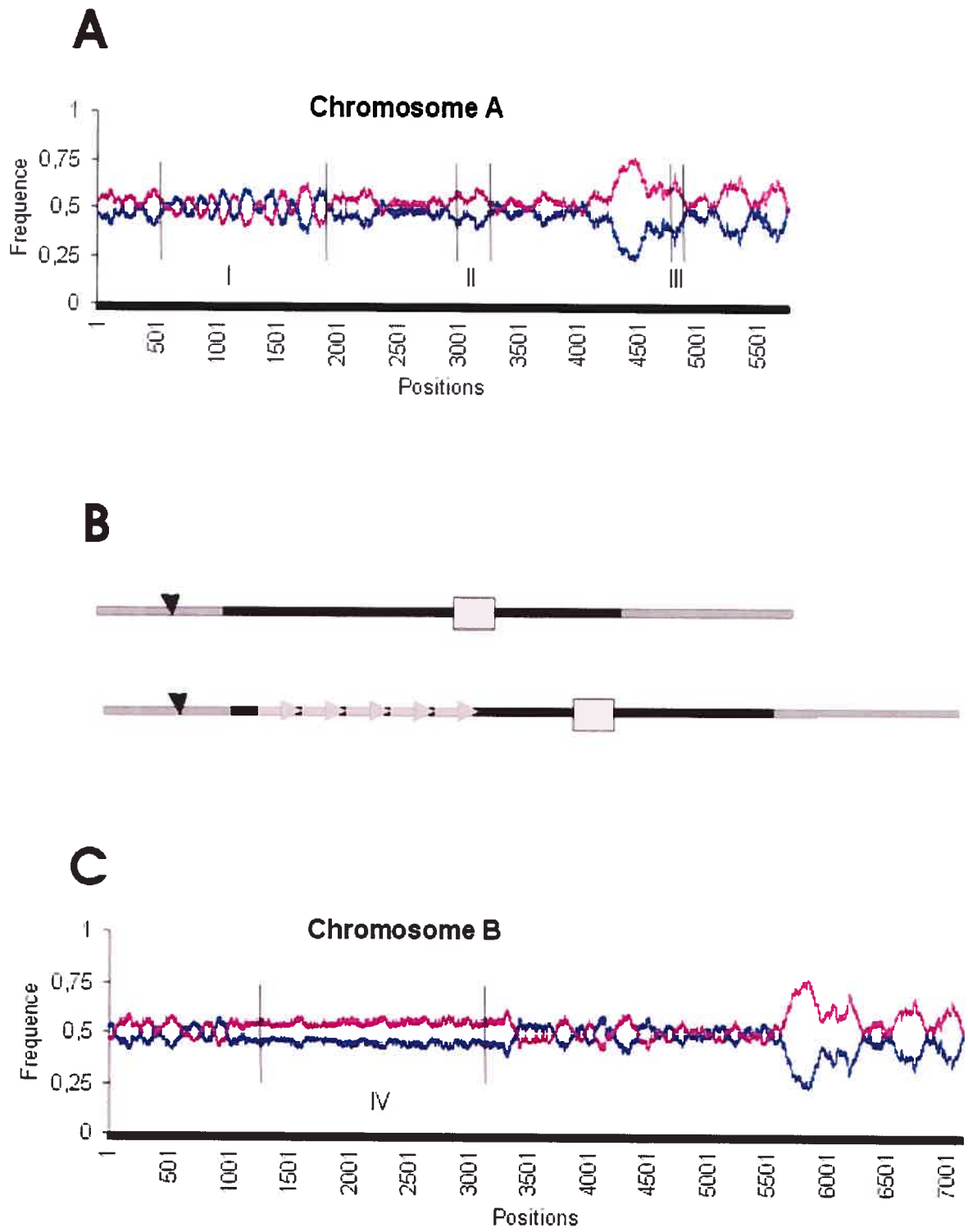


Figure 2

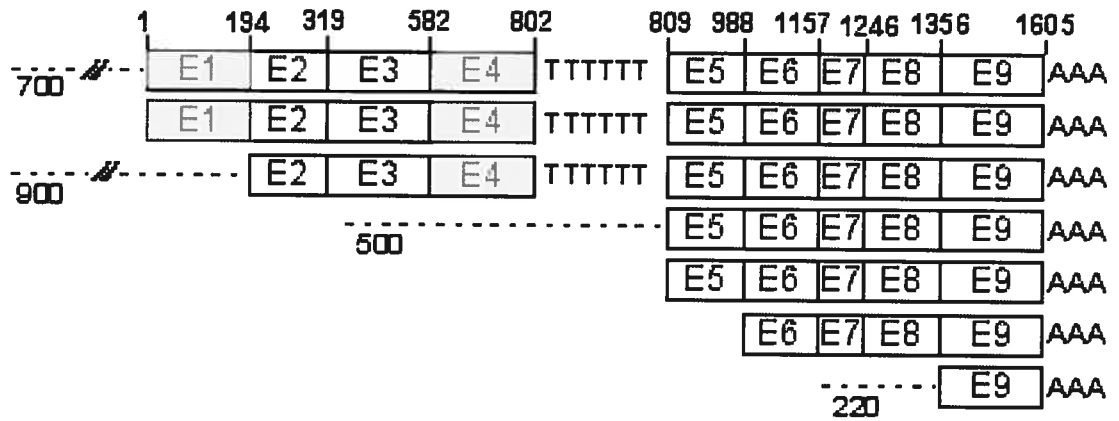
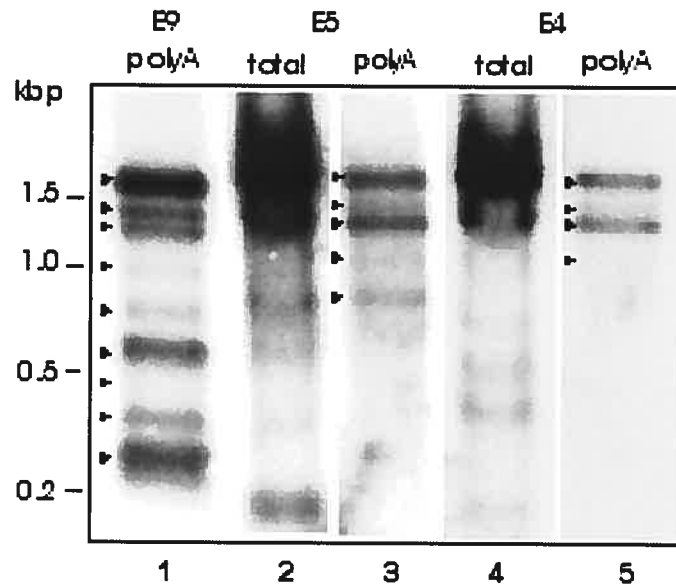
A**B**

Figure 3

A

exon #4
 5' ... GGAC cggac ... 3'
 cDNA ... GGAC TTTTTT CGCTC ...
 5' ... tccac CGCTC ... 3'
 exon #5

B

3D interaction

Para.deni-bc	KKE	IFGYLPMVLAMAAIGILGFVVAHHMYTAGMSLTQQAYFMLATMTIAVPTGIKVF	SW-IATM	WGG						
Recl.amer-mt	RRE	IFGYLGMVYAMSSIGVLGFIVWAHHMYTVGMDVDTRAYFTAATMVI	IAPTGIKIFSW-LATM	WGG						
Naeg.grub-mt	KKR	VFGHVPMIAAMLMIGLIGFIVWAHHMYTS	GIDTSTKAYFTAATMVI	IAPTGIKIFNWIATM	WGG					
Bos.taur-mt	KKE	PFGYMGVMVAMMSIGFLGFIVWAHHMFTV	GMDVDTRAYFTSATMI	IAPTGVKVF	SW-LATL	HGG				
Marc.poly-mt	RKE	VFGYLGMYAMISIGVLGFIVWAHHMFTV	GLDVDTRAYFTAATMI	IAPTGIKIF	SW-IATM	WGG				
Acan.cast-mt	NKS	IFGYIGMYAMLSIAVLGFIVWAHHMYTV	GLDVDTRAYFTAATMI	IAPTGIKIF	SW-IATL	WGG				
Cryp.coni-mt	NKG	IFGVKGMISAMSAIGFLGFLVWAHHMYTV	GLDVDTRAYFTAATMI	IAPTGIKIF	SW-LATL	WGG				
Chla.rein-mt	QKI	IFGNPSMIFAMSCISLLGSVVWGHMYT	IGLETDRSYFSGVTIL	ISLPTGTR	IFNWLSTYL	GNP				
Plas.viva-mt	QKE	VFGLTGMICAMGAI	SLLGFIVWAHHMFTVGLDLDTV	VAYFTSATMI	IAPTGMKIF	SW-MATI	YSG			
Dict.disc-mt	CRS	LFGNDSMILAMSCIA	ILGSVVAHHMYTTGLEVDTRAFF	TSSTIL	ISLPTGTR	IFNWICTY	MGS			
Eugl.grac-mt	HKD	IFGREGMMYCIIS	IGVVGYFVWAHHMFTVGLDIDRS	SYFSIATS	IISLPTS	VKMF	SY-INTW	ASG		
Leis.tare-mt	FRC	VFSTVAMIYSMLIA	ILGMFVWAHHMFVVGMDVDS	RAYFGVS	ILIGLPTC	VKLF	NW-IYSF	LYT		
Tryp.borr-mt	LRG	VFSFVTMIYSMLC	IVIGFFVWAHHMFVIGMDIDSR	AYFGLSLL	IGLPTC	IKIF	NW-FYSF	LEY		
Dipl.papi-mt	LES	LYNMLGMVYAMIA	I	AVVGYFVWAHHMFTVGLDVDTR	VYFSSATLL	I	ALPTS	IKVF	SW-MVGL	RRM

Figure 4

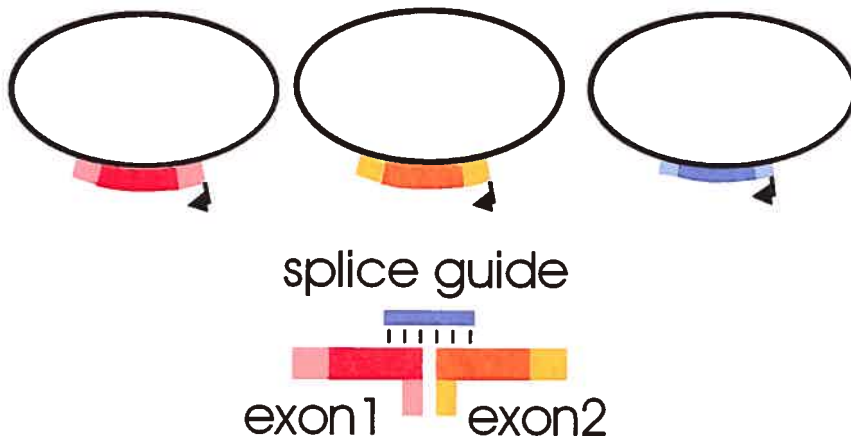


Figure 5

Chapitre 3. Conclusion

Le génome mitochondrial de *D. papillatum* est unique parmi les eucaryotes : les gènes sont tous fragmentés et chaque exon est situé sur un chromosome circulaire différent. Ce génome est aussi différent de celui du groupe frère de *Diplonema*, les kinétoplastides, car il ne forme pas de réseau et la dualité maxicercle / minicercle n'existe pas. L'édition de l'ARNm du gène *cox1* observée chez *D. papillatum* est modeste comparée à l'édition massive subie par les ARNm des kinétoplastides et correspond uniquement à de l'insertion de U.

Le mécanisme de ségrégation des chromosomes multiples qui n'ont pas de centromère est un problème encore mal compris. Chez les trypanosomatides, les chromosomes sont liés dans un réseau et sa duplication lors de la division cellulaire fournit un réseau complet à chaque cellule fille. Une autre solution est d'avoir suffisamment de redondance dans les copies pour que la répartition au hasard des chromosomes n'engendre pas la perte fatale d'un type de chromosome. Ce cas est illustré chez *C. helicis* qui possède >8000 minicercles. Si nous considérons à 200 le nombre d'ARNg nécessaire, il y aurait ~40 copies pour chaque minicercles, ce qui est suffisant pour permettre une répartition stochastique sans perte d'un type de chromosome. Chez *D. papillatum*, il y a plusieurs copies du même chromosome, donc nous pouvons supposer que la ségrégation au hasard s'applique aussi pour l'ADNmt de *Diplonema*.

Nous avons montré que les exons de *D. papillatum* sont transcrits séparément puis ligés ensembles créant un ARNm mature. Les étapes qui sont probablement impliquées dans la synthèse de l'ARNm continu incluent la transcription indépendante des exons avec leur région flanquante; le traitement des extrémités 5' et 3' des transcrits primaires; polyadénylation du transcrit de l'exon situé à l'extrémité 3' du gène; ligation des exons adjacents (figure 12).

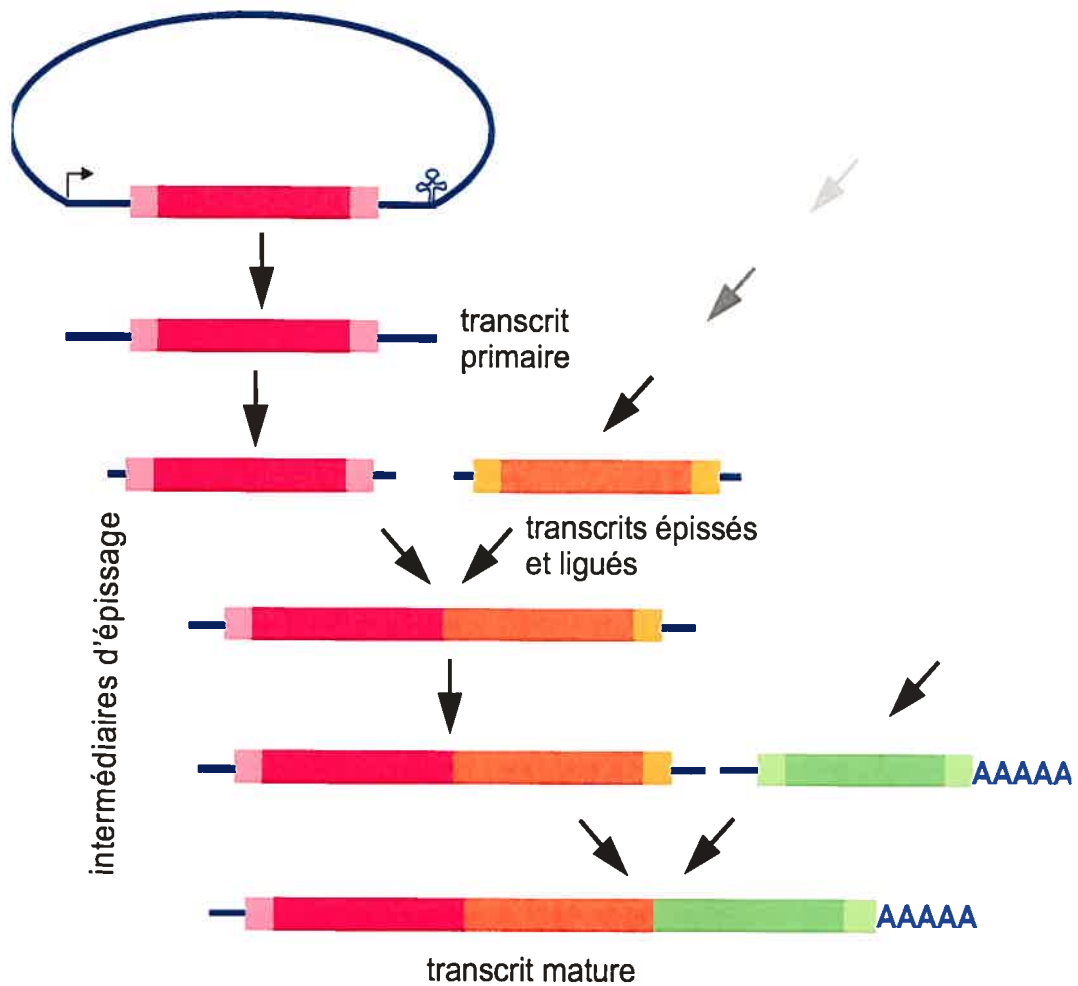


Figure 12 : Étapes d'expression des gènes de *D. papillatum*

La flèche courbée et le palindrome indiquent l'initiation et la terminaison de la transcription. Les séquences codantes de la cassette (les rectangles) sont en couleurs foncées et les régions adjacentes en pastel (courtoisie de Gertraud Burger).

Nous avons estimé à 100 le nombre d'exons et donc à 100 le nombre de transcrits. Cette complexité nécessite un système sélectif pour assembler les transcrits de manière spécifique.

3.1. Mécanismes potentiels d'épissage en trans

3.1.1. Via un intron discontinu

Ce mécanisme est très répandu au sein des génomes mitochondriaux des plantes. Une fois transcrits, les précurseurs s'associent par appariement afin de reconstruire une structure d'intron fonctionnel, permettant l'excision des séquences introniques et la ligation des exons. La spécificité de ce mécanisme réside dans l'appariement entre les régions exoniques ou flanquantes.

3.1.2. Via une molécule de type ARNg

Il est concevable qu'un ARN de type ARNg aligne deux exons consécutifs en formant un duplexe, permettant la jonction des exons par une réaction de ligation (figure 13). Chez *Diplonema*, l'hybridation de l'ARNg hypothétique avec les transcrits ciblerait spécifiquement les régions à éliminer (les régions non codantes) et les régions à liquer (les exons). Cette hypothèse accommoderait la ligation spécifique des transcrits des exons et l'insertion de U dans le transcrit du gène *cox1* (voir figure 14).

De plus, ce mécanisme hypothétique serait réalisable en prenant le même jeu de protéines que celui de l'éditosome soit une endonucléase, qui éliminerait les extensions non codantes 5' et 3', une TUTase et une ligase. Ce mécanisme nécessiterait aussi des protéines qui reconnaissent le duplexe ARNg/ARNm (voir figure 14).

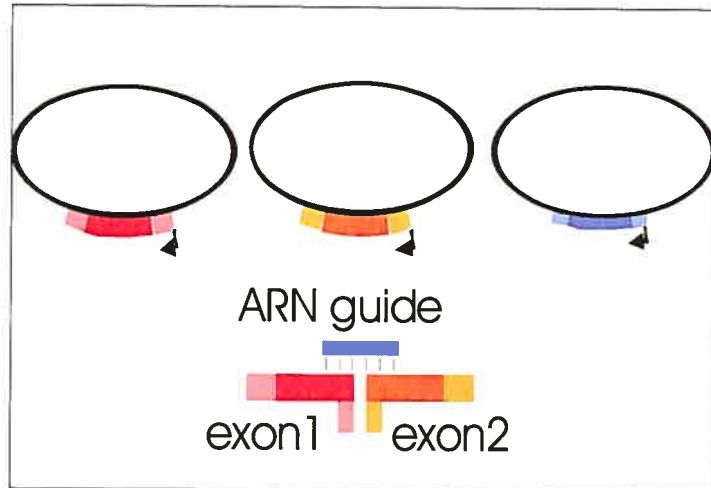


Figure 13 : Mécanismes d'épissage en trans médié par une molécule de type ARN guide
(Courtoisie de Gertraud Burger)

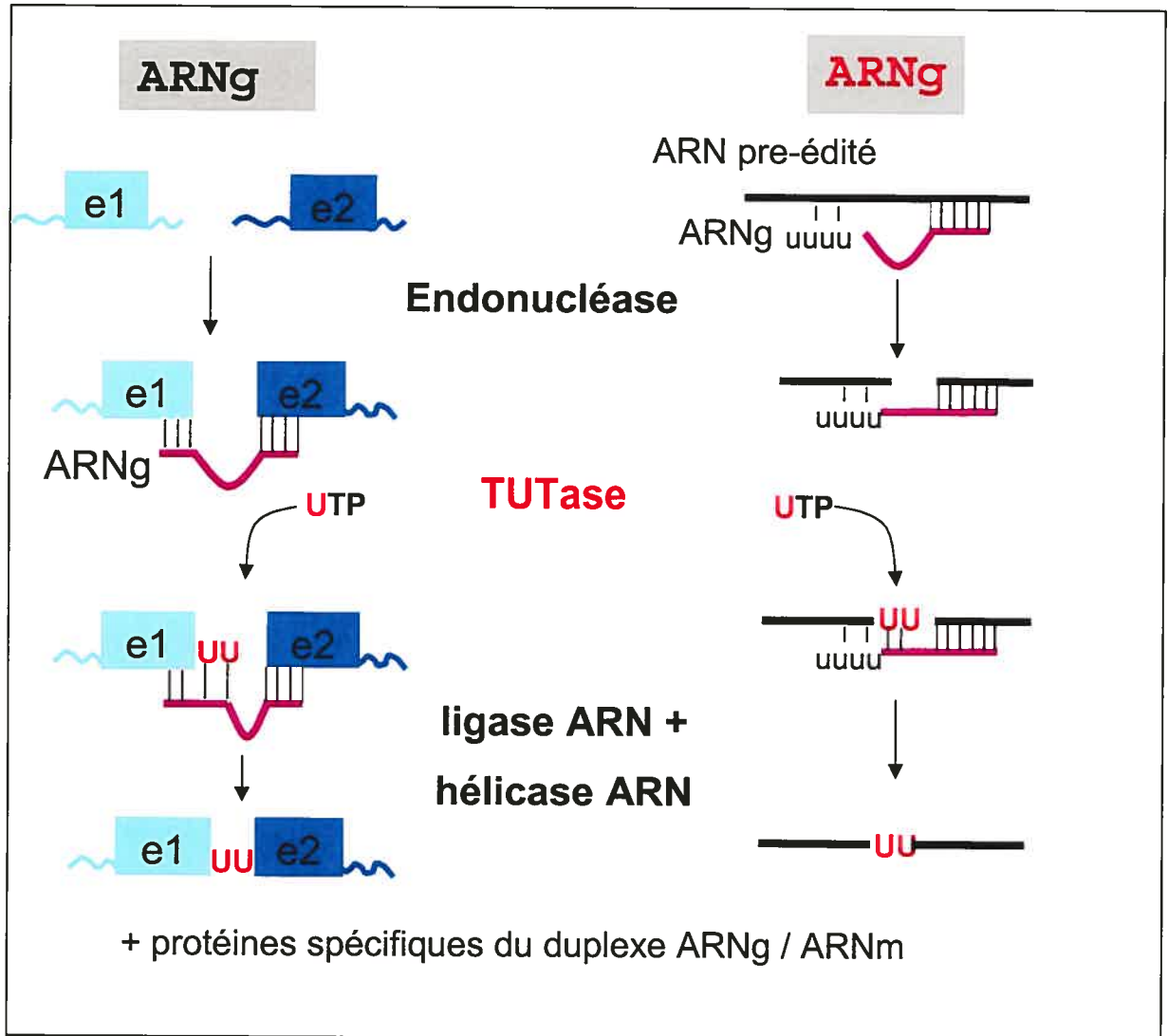


Figure 14 : Parallèle entre l'hypothèse ARNg chez *D. papillatum* et l'édition d'ARN chez les trypanosomatides.

Le mécanisme hypothétique de *Diplonema* est présenté à gauche et celui des trypanosomatides à droite.

3.2. ADNmt et expression des gènes au sein des diplonémides

3.2.1. ADN mitochondrial

L'ADNmt de *D. papillatum* étant inhabituel, la question se pose si ces caractéristiques représentent un caprice particulier de cette espèce ou si elles étaient déjà présentes dans l'ancêtre commun des diplonémides. Les résultats préliminaires sur l'ADNmt de *D. ambulator*, *D. sp2* et *Rhynchopus euleeides* indiquent que la structure multi-chromosomes est commune à tous les diplonémides (voir annexe 1). L'analyse de la séquence des chromosomes de chacune des espèces, avec notamment la caractérisation des régions conservées au sein du groupe, apportera des indices sur l'évolution de cet ADNmt. De plus, connaître le degré de variabilité de la fragmentation des gènes à travers les diplonémides promet de donner des indices sur le mécanisme d'épissage en trans.

3.2.2. Expression du gène *cox1*

Le gène *cox1* est composé de 9 exons chez *D. papillatum*. Nous avons montré que ce même gène était fragmenté chez *R. euleeides* et *D. ambulator* (voir annexe 1). Par extrapolation, nous assumons que la fragmentation des gènes est une caractéristique commune aux diplonémides.

Nous avons séquencé l'ADNc de *cox1* chez ces 2 espèces et nous avons observé que les 6 T ajoutés par édition d'ARN chez *D. papillatum* sont présents (voir annexe 2), même si la séquence d'ADN et protéique du gène *cox1* chez les diplonémides ne sont pas les mêmes. Ce résultat indique que cette région ne semble pas subir la même pression évolutive que le reste du gène. Ceci est un nouvel argument qui va en faveur de l'hypothèse de l'édition d'ARN chez *D. papillatum* et qu'il s'agit probablement d'une insertion de 6 U au niveau ARN. L'édition de l'ARN par insertion de U chez les diplonémides pourrait correspondre à un état primitif de ce qui est observé chez les kinétoplastides. Des

expériences de comparaison d'ADNc et d'ADNmt chez *R. euleeides* et *D. ambulator* permettront de confirmer notre hypothèse établissant que l'édition d'ARN est un mécanisme présent chez les diplonémides.

3.3. Les perspectives

Plusieurs questions concernant le génome mitochondrial de *D. papillatum* devront être résolues prochainement. Dans la partie suivante, je propose des approches afin d'y répondre.

3.3.1. Connaissance du contenu en gène

Jusqu'à maintenant, le séquençage de l'ADNmt de *D. papillatum* nous a fourni ~100 exons potentiels dont seulement 49 ont pu être identifiés clairement. La petite taille des exons ainsi que le peu de conservation dans la séquence protéique des gènes de *Diplonema* pourraient expliquer ce problème. Le séquençage d'une banque d'ADNc mitochondriaux donnera l'information requise. Dans le cas où la banque d'ADNc mitochondrial n'est pas efficace, il sera toujours possible d'obtenir des ADNc partiels par RT-PCR en utilisant un oligonucléotide spécifique d'une cassette et un autre qui se fixe sur la queue poly(A).

3.3.2. Cartographie des ARN

Les séquences issues de la banque d'ADNc n'ont pas permis de caractériser précisément les extrémités des ADNc. Ces données sont nécessaires pour identifier les signaux d'initiation et de terminaison de la transcription ainsi que les séquences qui pourraient être impliquées dans l'épissage en trans. L'approche pour identifier les extrémités 5' est le 5' RACE ; pour les extrémités 3', il sera nécessaire d'ajouter une queue synthétique poly(A), (C), (T), ou (G) à l'aide d'une poly(A) polymérase avant de faire une RT-PCR en utilisant l'amorce complémentaire de la queue. Les techniques telles que

l'extension d'amorce, « nucléase S1 » et circularisation d'ARN pourront être utilisées pour confirmer les résultats du 5' RACE.

3.3.3. Test des hypothèses d'épissage en trans

Épissage en trans médié par un intron : test d'épissage in vitro

Les analyses des séquences génomiques n'ont pas révélé de motifs typiques d'intron de groupe I, II ou spliceosomal. De plus, les séquences des exons ou des régions flanquantes des exons ne démontrent pas de complémentarité évidente permettant un appariement spécifique. Un essai d'épissage *in vitro* pourrait donner cette information. Deux exons adjacents seront synthétisés *in vitro* et leur capacité à éliminer les régions 5' et 3' UTR puis à se liguer sera expérimentée. Différentes conditions de MgCl₂ et de GTP (ou autre source d'énergie) ainsi que différents extraits acellulaires (mitochondrial ou total) de *D. papillatum* seront ajoutés. Une fois ce test mis au point, la réduction de la séquence non-codante des transcrits sera envisagée afin d'identifier quelles sont les séquences impliquées dans l'épissage.

Existence d'ARNg

Lors des constructions de banque d'ADNc, nous avons effectué une séparation en fonction de la taille, éliminant tous les petits ARN incluant les ARNg potentiels. Afin de démontrer leur existence, nous testerons par Northern blot des sondes qui reconnaissent les jonctions, définies selon des hypothèses différentes : (1) l'ARNg est complémentaire des extrémités de deux exons voisins; (2) l'ARNg est complémentaire de l'exon et de la région non-codante adjacente; (3) l'ARNg est complémentaire uniquement de la région non-codante adjacente. Une alternative serait d'utiliser un ADN simple brin correspondant à la séquence de l'ARN mature afin d'allumer tous les ARNg potentiels.

Une autre approche sera effectuée en synthétisant des oligonucléotides biotinylés complémentaires des ARNg hypothétiques. L'interaction de la biotine avec des billes magnétiques enduites de streptavidines permettra d'extraire les ARNg de l'ARN total (Haukanes & Kvam, 1993).

3.3.4. Vue d'ensemble de l'ADNmt des Euglenozoa

Les cultures de différentes espèces du genre diplonémide, notamment, *D. ambulator*, *D. sp2* et *Rhynchopus euleeides* sont disponibles au laboratoire. L'extraction de l'ADNmt de ces organismes a déjà été effectuée. L'amplification spécifique des chromosomes portant des exons de *cox1* est en cours. Le produit obtenu sera cloné et séquencé. Comme l'ADNc du gène *cox1* est connu pour chacune des espèces nommées ci-dessus, la fragmentation du gène *cox1* sera déduite par la comparaison des séquences génomiques et d'ADNc. De plus, la fragmentation sera confirmée par Northern.

Les tentatives de résolution du génome mitochondrial des euglénides, par l'intermédiaire d'*E. gracilis*, ont échoué. Il serait intéressant d'analyser l'ADNmt d'autres espèces de ce groupe. Un premier candidat pourrait être *Petalomonas cantuscigny* dont la séquence continue de 25 kbp issue de l'ADNmt n'a révélé qu'une partie du gène *atp6* suggérant des gènes cryptiques et de l'édition massive d'ARN (Roy et al., 2007). L'objectif est de savoir si les euglénides ont un génome proche de celui des kinétoplastides ou des diplonémides ou complètement différent.

Le génome mitochondrial des kinétoplastides est bien caractérisé, toutefois, peu de données sont disponibles chez les membres non-pathogènes des bodonides. L'analyse de *B. saltans* qui se trouve à la base des trypanosomatides pourrait apporter des réponses sur l'évolution de l'ADNk en réseau. L'investigation du génome d'*Hemistasia*, qui est un kinétoplastide avec une morphologie similaire aux diplonémides (Simpson et al., 2002) fournirait, peut-être, un lien évolutif entre les kinétoplastides et les diplonémides.

Le mot de la fin...

L'étude par biologie moléculaire d'une seule espèce des Euglenozoa a déniché une nouvelle structure de gène et un nouveau mécanisme de maturation des ARN. Fait sûr, nous ne sommes pas au bout de nos surprises avec les Euglenozoa.

Une question qui m'a souvent été posée lors des différentes présentations est :

Quel est l'avantage pour un organisme d'avoir un génome aussi complexe?

Maintenant ma réponse serait la suivante.

Je ne connais pas l'avantage que l'organisme a eu au moment où c'est apparu. Ce qui est sûr, c'est que ceci ne l'a pas empêché de survivre. Je pense que l'organisme s'est adapté comme il a pu face à un événement qui a perturbé complètement son génome. A priori, cette organisation de génome semble complexe et très coûteuse en énergie pour la cellule mais il y a encore beaucoup d'inconnues et peut-être que si on prenait le point de vue de *Diplonema*, avoir un génome unique tel que les mammifères serait pour lui insensé et sûrement très complexe.

Bibliographie

- Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 29:380-395.
- Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* 430:45-50.
- Alfonzo JD, Thiemann O, Simpson L. 1997. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. *Nucleic Acids Res* 25:3751-3759.
- Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133-140.
- Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* 154:181-192.
- Backert S, Nielsen BL, Borner T. 1997. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci* 2:477-483.
- Beagley CT, Okimoto R, Wolstenholme DR. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* 148:1091-1108.
- Bendich AJ. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J Mol Biol* 255:564-588.
- Bendich AJ. 2001. The form of chromosomal DNA molecules in bacterial cells. *Biochimie* 83:177-186.
- Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46:819-826.
- Boer PH, Gray MW. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55:399-411.
- Bohnert HJ, Herrmann RG. 1974. The genomic complexity of *Acanthamoeba castellanii* mitochondrial DNA. *Eur J Biochem* 50:83-90.
- Bonen L. 2006. Mitochondrial genes leave home. *The New phytologist* 172:379-381.
- Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends Genet* 17:322-331.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767-1780.
- Borst P. 1972. Mitochondrial nucleic acids. *Annual review of biochemistry* 41:333-376.
- Borst P, Kroon AM. 1969. Mitochondrial DNA: physicochemical properties, replication, and genetic function. *Int Rev Cytol* 26:107-190.
- Boxma B, de Graaf RM, van der Staay GW, van Alen TA, Ricard G, Gabaldon T, van Hoek AH, Moon-van der Staay SY, Koopman WJ, van Hellemond JJ, Tielens AG, Friedrich T, Veenhuis M, Huynen MA, Hackstein JH. 2005. An anaerobic mitochondrion that produces hydrogen. *Nature* 434:74-79.

- Breglia SA, Slamovits CH, Leander BS. 2007. Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from hsp90 gene sequences. *The Journal of eukaryotic microbiology* 54:86-92.
- Brennicke A, Marchfelder A, Binder S. 1999. RNA editing. *FEMS Microbiol Rev* 23:297-316.
- Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Curr Opin Microbiol* 7:528-534.
- Bullerwell CE, Leigh J, Forget L, Lang BF. 2003a. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res* 31:759-768.
- Bullerwell CE, Leigh J, Seif E, Longcore JE, Lang BF. 2003b. Evolution of the fungi and their mitochondrial genomes. In: Arora D, Khachatourians GG, eds. *Applied Mycology and Biotechnology*. Amsterdam: Elsevier Science. pp 133-160.
- Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003a. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci U S A* 100:892-897.
- Burger G, Gray MW, Lang BF. 2003b. Mitochondrial genomes - anything goes. *Trends Genet* 19:709-716.
- Busse I, Preisfeld A. 2002. Phylogenetic position of *Rhynchopus* sp. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. *Gene* 284:83-91.
- Busse I, Preisfeld A. 2003. Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of *Distigma* and *Astasia* (Euglenozoa). *International journal of systematic and evolutionary microbiology* 53:617-624.
- Carle GF, Olson MV. 1985. An electrophoretic karyotype for yeast. *Proc Natl Acad Sci U S A* 82:3756-3760.
- Cavalier-Smith T. 1981. Eukaryote kingdoms: seven or nine? *Biosystems* 14:461-481.
- Chapdelaine Y, Bonen L. 1991. The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell* 65:465-472.
- Chiu N, Chiu A, Suyama Y. 1975. Native and imported transfer RNA in mitochondria. *J Mol Biol* 99:37-50.
- Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol* 136:3486-3503.
- Cummings DJ, McNally KL, Domenico JM, Matsuura ET. 1990. The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr Genet* 17:375-402.
- Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A* 103:8751-8756.
- Denovan-Wright EM, Lee RW. 1995. Evidence that the fragmented ribosomal RNAs of *Chlamydomonas* mitochondria are associated with ribosomes. *FEBS Lett* 370:222-226.

- Feagin JE. 1992. The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol Biochem Parasitol* 52:145-148.
- Feagin JE. 2000. Mitochondrial genome diversity in parasites. *Int J Parasitol* 30:371-390.
- Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53:413-422.
- Feagin JE, Mericle BL, Werner E, Morris M. 1997. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res* 25:438-446.
- Fonty G, Crouse EJ, Stutz E, Bernardi G. 1975. The mitochondrial genome of *Euglena gracilis*. *Eur J Biochem* 54:367-372.
- Foury F, Roganti T, Lecrenier N, Purnelle B. 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* 440:325-331.
- Giege P, Brennicke A. 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A* 96:15324-15329.
- Gott JM, Emeson RB. 2000. Functions and mechanisms of RNA editing. *Annu Rev Genet* 34:499-531.
- Grams J, McManus MT, Hajduk SL. 2000. Processing of polycistronic guide RNAs is associated with RNA editing complexes in *Trypanosoma brucei*. *EMBO J* 19:5525-5532.
- Gray MW. 1998. Rickettsia, typhus and the mitochondrial connection. *Nature* 396:109-110.
- Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu Rev Genet* 38:477-524.
- Griessman K. 1914. Ueber marine Flagellaten. *Arch Protistenkd* 32:1-78.
- Guerineau M, Grandchamp C, Slonimski PP. 1975. Rearrangement of mitochondrial DNA molecules during the differentiation of mitochondria in yeast. I.-Electron microscopic studies of size and shape. *Biochimie* 57:917-929.
- Hajduk SL, Siqueira AM, Vickerman K. 1986. Kinetoplast DNA of *Bodo caudatus*: a noncatenated structure. *Mol Cell Biol* 6:4372-4378.
- Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* 21:3537-3544.
- 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 research* 31:5907-5916.
- Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. *Trends Genet* 21:111-119.
- Haukanes BI, Kvam C. 1993. Application of magnetic beads in bioassays. *Bio/technology (Nature Publishing Company)* 11:60-63.
- Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci U S A* 101:10639-10643.
- Hong M, Simpson L. 2003. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. *Protist* 154:265-279.

- Itoh T, Martin W, Nei M. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci U S A* 99:12944-12948.
- Kamikawa R, Inagaki Y, Sako Y. 2007. Fragmentation of Mitochondrial Large Subunit rRNA in the Dinoflagellate *Alexandrium catenella* and the Evolution of rRNA structure in Alveolate Mitochondria. *Protist* 158:239-245.
- Koslowsky DJ, Yahampath G. 1997. Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in *Trypanosoma brucei* are independent events. *Mol Biochem Parasitol* 90:81-94.
- Kovac L, Lazowska J, Slonimski PP. 1984. A yeast with linear molecules of mitochondrial DNA. *Mol Gen Genet* 197:420-424.
- Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147-157.
- Kuck U, Jekosch K, Holzamer P. 2000. DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. *Gene* 253:13-18.
- Kuhn K, Weihe A, Borner T. 2005. Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Res* 33:337-346.
- Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet* 38:1-35.
- Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387:493-497.
- Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33:351-397.
- Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet* 23:119-125.
- Laurent M, Steinert M. 1970. [Observation of kinetoplast DNA by electron microscopy]. *Archives internationales de physiologie et de biochimie* 78:1003-1004.
- Leander BS, Triemer RE, Farmer MA. 2001. Character evolution in heterotrophic euglenids. *European Journal of Protistology* 37:337-356.
- Lin S, Zhang H, Spencer DF, Norman JE, Gray MW. 2002. Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J Mol Biol* 320:727-739.
- Lukes J, Guilbride DL, Votypka J, Zikova A, Benne R, Englund PT. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot Cell* 1:495-502.
- Lukes J, Jirku M, Avliyakov N, Benada O. 1998. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia heliciis*. *Embo J* 17:838-846.
- Mahendran R, Spottswood MR, Miller DL. 1991. RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. *Nature* 349:434-438.
- Malek O, Brennicke A, Knoop V. 1997. Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proc Natl Acad Sci U S A* 94:553-558.
- 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* 4:1599-1609.

- Maslov DA, Simpson L. 1994. RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan *Trypanoplasma borreli*. *Mol Cell Biol* 14:8174-8182.
- Maslov DA, Yasuhira S, Simpson L. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. *Protist* 150:33-42.
- Michel F, Jacquier A, Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64:867-881.
- Moreira D, Lopez-Garcia P, Rodriguez-Valera F. 2001. New insights into the phylogenetic position of diplomids: G+C content bias, differences of evolutionary rate and a new environmental sequence. *International journal of systematic and evolutionary microbiology* 51:2211-2219.
- Murray HL, Mikheeva S, Coljee VW, Turczyk BM, Donahue WF, Bar-Shalom A, Jarrell KA. 2001. Excision of group II introns as circles. *Mol Cell* 8:201-211.
- Nass MM, Schori L, Ben-Shaul Y, Edelman M. 1974. Size and configuration of mitochondrial DNA in *Euglena gracilis*. *Biochim Biophys Acta* 374:283-291.
- Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* 10:819-831.
- Nosek J, Tomaska L. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr Genet* 44:73-84.
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. 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* 268:434-445.
- Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 29:380-395.
- Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* 430:45-50.
- Alfonzo JD, Blanc V, Estevez AM, Rubio MA, Simpson L. 1999. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J* 18:7056-7062.
- Alfonzo JD, Thiemann O, Simpson L. 1997. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. *Nucleic Acids Res* 25:3751-3759.
- Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133-140.
- Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. *Genetics* 154:181-192.
- Backert S, Nielsen BL, Borner T. 1997. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci* 2:477-483.

- Badidi E, De Sousa C, Lang BF, Burger G. 2003. AnaBench: a Web/CORBA-based workbench for biomolecular sequence analysis. *BMC Bioinformatics* 4:63.
- Beagley CT, Okimoto R, Wolstenholme DR. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* 148:1091-1108.
- Bendich AJ. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J Mol Biol* 255:564-588.
- Bendich AJ. 2001. The form of chromosomal DNA molecules in bacterial cells. *Biochimie* 83:177-186.
- Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46:819-826.
- Boer PH, Gray MW. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55:399-411.
- Bohnert HJ, Herrmann RG. 1974. The genomic complexity of *Acanthamoeba castellanii* mitochondrial DNA. *Eur J Biochem* 50:83-90.
- Bonen L. 2006. Mitochondrial genes leave home. *The New phytologist* 172:379-381.
- Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends Genet* 17:322-331.
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767-1780.
- Borst P. 1972. Mitochondrial nucleic acids. *Annual review of biochemistry* 41:333-376.
- Borst P, Kroon AM. 1969. Mitochondrial DNA: physicochemical properties, replication, and genetic function. *Int Rev Cytol* 26:107-190.
- Boxma B, de Graaf RM, van der Staay GW, van Alen TA, Ricard G, Gabaldon T, van Hoek AH, Moon-van der Staay SY, Koopman WJ, van Hellemond JJ, Tielens AG, Friedrich T, Veenhuis M, Huynen MA, Hackstein JH. 2005. An anaerobic mitochondrion that produces hydrogen. *Nature* 434:74-79.
- Breglia SA, Slamovits CH, Leander BS. 2007. Phylogeny of phagotrophic euglenids (Euglenozoa) as inferred from *hsp90* gene sequences. *The Journal of eukaryotic microbiology* 54:86-92.
- Brennicke A, Marchfelder A, Binder S. 1999. RNA editing. *FEMS Microbiol Rev* 23:297-316.
- Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Curr Opin Microbiol* 7:528-534.
- Bullerwell CE, Gray MW. 2005. In vitro characterization of a tRNA editing activity in the mitochondria of *Spizellomyces punctatus*, a chytridiomycete fungus. *The Journal of biological chemistry* 280:2463-2470.
- Bullerwell CE, Leigh J, Forget L, Lang BF. 2003a. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res* 31:759-768.
- Bullerwell CE, Leigh J, Seif E, Longcore JE, Lang BF. 2003b. Evolution of the fungi and their mitochondrial genomes. In: Arora D, Khachatourians GG, eds. *Applied Mycology and Biotechnology*. Amsterdam: Elsevier Science. pp 133-160.

- Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003a. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci U S A* 100:892-897.
- Burger G, Gray MW, Lang BF. 2003b. Mitochondrial genomes - anything goes. *Trends Genet* 19:709-716.
- Burger G, Lavrov DV, Forget L, Lang BF. 2007. Sequencing complete mitochondrial and plastid genomes. *Nat Protoc* 2:603-614.
- Busse I, Preisfeld A. 2002. Phylogenetic position of *Rhynchopus* sp. and *Diplonema ambulator* as indicated by analyses of euglenozoan small subunit ribosomal DNA. *Gene* 284:83-91.
- Busse I, Preisfeld A. 2003. Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of *Distigma* and *Astasia* (Euglenozoa). *International journal of systematic and evolutionary microbiology* 53:617-624.
- Carle GF, Olson MV. 1985. An electrophoretic karyotype for yeast. *Proc Natl Acad Sci U S A* 82:3756-3760.
- Cavalier-Smith T. 1981. Eukaryote kingdoms: seven or nine? *Biosystems* 14:461-481.
- Chapdelaine Y, Bonen L. 1991. The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell* 65:465-472.
- Chaput H, Wang Y, Morse D. 2002. Polyadenylated transcripts containing random gene fragments are expressed in dinoflagellate mitochondria. *Protist* 153:111-122.
- Chiu N, Chiu A, Suyama Y. 1975. Native and imported transfer RNA in mitochondria. *J Mol Biol* 99:37-50.
- Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, Meyer L, Wilson RK, Newton KJ. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol* 136:3486-3503.
- Cummings DJ, McNally KL, Domenico JM, Matsuura ET. 1990. The complete DNA sequence of the mitochondrial genome of *Podospira anserina*. *Curr Genet* 17:375-402.
- Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci U S A* 103:8751-8756.
- Denovan-Wright EM, Lee RW. 1995. Evidence that the fragmented ribosomal RNAs of *Chlamydomonas* mitochondria are associated with ribosomes. *FEBS Lett* 370:222-226.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Edqvist J, Burger G, Gray MW. 2000. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J Mol Biol* 297:381-393.
- Estevez AM, Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. *Gene* 240:247-260.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-185.

- Feagin JE. 1992. The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol Biochem Parasitol* 52:145-148.
- Feagin JE. 2000. Mitochondrial genome diversity in parasites. *Int J Parasitol* 30:371-390.
- Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53:413-422.
- Feagin JE, Mericle BL, Werner E, Morris M. 1997. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res* 25:438-446.
- Fonty G, Crouse EJ, Stutz E, Bernardi G. 1975. The mitochondrial genome of *Euglena gracilis*. *Eur J Biochem* 54:367-372.
- Foury F, Roganti T, Lecrenier N, Purnelle B. 1998. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* 440:325-331.
- Gagliardi D, Stepien PP, Temperley RJ, Lightowers RN, Chrzanowska-Lightowers ZM. 2004. Messenger RNA stability in mitochondria: different means to an end. *Trends Genet* 20:260-267.
- Giege P, Brennicke A. 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A* 96:15324-15329.
- Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. 1991. A small chloroplast RNA may be required for trans-splicing in *Chlamydomonas reinhardtii*. *Cell* 65:135-143.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195-202.
- Gott JM, Emeson RB. 2000. Functions and mechanisms of RNA editing. *Annu Rev Genet* 34:499-531.
- Grams J, McManus MT, Hajduk SL. 2000. Processing of polycistronic guide RNAs is associated with RNA editing complexes in *Trypanosoma brucei*. *EMBO J* 19:5525-5532.
- Gray MW. 1998. Rickettsia, typhus and the mitochondrial connection. *Nature* 396:109-110.
- Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu Rev Genet* 38:477-524.
- Griessman K. 1914. Ueber marine Flagellaten. *Arch Protistenkd* 32:1-78.
- Guerineau M, Grandchamp C, Slonimski PP. 1975. Rearrangement of mitochondrial DNA molecules during the differentiation of mitochondria in yeast. I.-Electron microscopic studies of size and shape. *Biochimie* 57:917-929.
- Hajduk SL, Siqueira AM, Vickerman K. 1986. Kinetoplast DNA of *Bodo caudatus*: a noncatenated structure. *Mol Cell Biol* 6:4372-4378.
- Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* 21:3537-3544.
- 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 research* 31:5907-5916.

- Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. *Trends Genet* 21:111-119.
- Haukanes BI, Kvam C. 1993. Application of magnetic beads in bioassays. *Bio/technology (Nature Publishing Company)* 11:60-63.
- Hauth AM, Joseph DA. 2002. Beyond tandem repeats: complex pattern structures and distant regions of similarity. *Bioinformatics* 18 Suppl 1:S31-37.
- Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci U S A* 101:10639-10643.
- Hong M, Simpson L. 2003. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. *Protist* 154:265-279.
- Itoh T, Martin W, Nei M. 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci U S A* 99:12944-12948.
- Joannie Roy, Drahomira Faktorova, Lukes J, Burger aG. 2007. Unusual Mitochondrial Genome Structures throughout Euglenozoa. *Protists in press*.
- Kamikawa R, Inagaki Y, Sako Y. 2007. Fragmentation of Mitochondrial Large Subunit rRNA in the Dinoflagellate *Alexandrium catenella* and the Evolution of rRNA structure in Alveolate Mitochondria. *Protist* 158:239-245.
- Knoop V, Altwasser M, Brennicke A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet* 255:269-276.
- Kohchi T, Umesono K, Ogura Y, Komine Y, Nakahigashi K, Komano T, Yamada Y, Ozeki H, Ohyama K. 1988. A nicked group II intron and trans-splicing in liverwort, *Marchantia polymorpha*, chloroplasts. *Nucleic Acids Res* 16:10025-10036.
- Koslowsky DJ, Yahampath G. 1997. Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in *Trypanosoma brucei* are independent events. *Mol Biochem Parasitol* 90:81-94.
- Kovac L, Lazowska J, Slonimski PP. 1984. A yeast with linear molecules of mitochondrial DNA. *Mol Gen Genet* 197:420-424.
- Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147-157.
- Kuck U, Jekosch K, Holzamer P. 2000. DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. *Gene* 253:13-18.
- Kuhn K, Weihe A, Borner T. 2005. Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Res* 33:337-346.
- Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet* 38:1-35.
- Lang BF, Burger G. 2007. Purification of mitochondrial and plastid DNA. *Nat Protoc* 2:652-660.
- Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387:493-497.
- Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33:351-397.

- Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet* 23:119-125.
- Laurent M, Steinert M. 1970. [Observation of kinetoplast DNA by electron microscopy]. *Archives internationales de physiologie et de biochimie* 78:1003-1004.
- Leander BS, Triemer RE, Farmer MA. 2001. Character evolution in heterotrophic euglenids. *European Journal of Protistology* 37:337-356.
- Li-Pook-Than J, Bonen L. 2006. Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res* 34:2782-2790.
- Lin S, Zhang H, Spencer DF, Norman JE, Gray MW. 2002. Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J Mol Biol* 320:727-739.
- Lukes J, Guilbride DL, Votypka J, Zikova A, Benne R, Englund PT. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot Cell* 1:495-502.
- Lukes J, Jirku M, Avliyakov N, Benada O. 1998. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia heliciis*. *Embo J* 17:838-846.
- Macina RA, Sanchez DO, Gluschankof DA, Burrone OR, Frasch AC. 1986. Sequence diversity in the kinetoplast DNA minicircles of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 21:25-32.
- Mahendran R, Spottswood MR, Miller DL. 1991. RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. *Nature* 349:434-438.
- Malek O, Brennicke A, Knoop V. 1997. Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proc Natl Acad Sci U S A* 94:553-558.
- 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* 4:1599-1609.
- Marande W, Lukes J, Burger G. 2005. Unique mitochondrial genome structure in diplomonads, the sister group of kinetoplastids. *Eukaryot Cell* 4:1137-1146.
- Maslov DA, Simpson L. 1994. RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan *Trypanoplasma borreli*. *Mol Cell Biol* 14:8174-8182.
- Maslov DA, Yasuhira S, Simpson L. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial COI protein sequences. *Protist* 150:33-42.
- Michel F, Jacquier A, Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64:867-881.
- Miyazawa S, Jernigan RL. 1999. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins* 34:49-68.
- Moreira D, Lopez-Garcia P, Rodriguez-Valera F. 2001. New insights into the phylogenetic position of diplomonads: G+C content bias, differences of evolutionary rate and a new environmental sequence. *International journal of systematic and evolutionary microbiology* 51:2211-2219.
- Murray HL, Mikheeva S, Coljee VW, Turczyk BM, Donahue WF, Bar-Shalom A, Jarrell KA. 2001. Excision of group II introns as circles. *Mol Cell* 8:201-211.

- Nass MM, Schori L, Ben-Shaul Y, Edelman M. 1974. Size and configuration of mitochondrial DNA in *Euglena gracilis*. *Biochim Biophys Acta* 374:283-291.
- Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* 10:819-831.
- Nisbet RE, Koumandou LV, Barbrook AC, Howe CJ. 2004. Novel plastid gene minicircles in the dinoflagellate *Amphidinium operculatum*. *Gene* 331:141-147.
- Nosek J, Tomaska L. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Curr Genet* 44:73-84.
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. 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* 268:434-445.
- Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T. 2006. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *The Journal of biological chemistry* 281:37661-37667.
- Ostermeier C, Harrenga A, Ermler U, Michel H. 1997. Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody FV fragment. *Proc Natl Acad Sci U S A* 94:10547-10553.
- Perez-Martinez X, Antaramian A, Vazquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, Gonzalez-Halphen D. 2001. Subunit II of cytochrome c oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *The Journal of biological chemistry* 276:11302-11309.
- Pont-Kingdon G, Vassort CG, Warrior R, Okimoto R, Beagley CT, Wolstenholme DR. 2000. Mitochondrial DNA of *Hydra attenuata* (Cnidaria): a sequence that includes an end of one linear molecule and the genes for 1-rRNA, tRNA(f-Met), tRNA(Trp), COII, and ATPase8. *J Mol Evol* 51:404-415.
- Porter D. 1973. *Isonema papillatum* sp. n., a new colorless marine flagellate: A light- and electron microscopic study. *J Protozool* 20:351-356.
- Preisfeld A, Busse I, Klingberg M, Talke S, Ruppel HG. 2001. Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *International journal of systematic and evolutionary microbiology* 51:751-758.
- Raczynska KD, Le Ret M, Rurek M, Bonnard G, Augustyniak H, Gualberto JM. 2006. Plant mitochondrial genes can be expressed from mRNAs lacking stop codons. *FEBS Lett* 580:5641-5646.
- Ray DS. 1989. Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes. *Mol Cell Biol* 9:1365-1367.
- Reichert AS, Neupert W. 2004. Mitochondriomics or what makes us breathe. *Trends Genet* 20:555-562.
- Renger HC, Wolstenholme DR. 1970. Kinetoplast deoxyribonucleic acid of the hemoflagellate *Trypanosoma lewisi*. *The Journal of cell biology* 47:689-702.

- Riou G, Delain E. 1969. Electron microscopy of the circular kinetoplasmic DNA from *Trypanosoma cruzi*: occurrence of catenated forms. *Proceedings of the National Academy of Sciences of the United States of America* 62:210-217.
- Rodriguez-Ezpeleta N, Teijeiro S, Forget L, Burger G, Lang BF. 2006. Generation of cDNA libraries: protists and fungi. In: Parkinson J, ed. *Methods in Molecular Biology: Methods in ESTs*. Totowa, NJ: Humana Press.
- Roy J, Faktorova D, Lukes J, Burger G. 2007. Unusual Mitochondrial Genome Structures throughout Euglenozoa. *Protist in press*.
- Ryan R, Grant D, Chiang KS, Swift H. 1978. Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences of the United States of America* 75:3268-3272.
- Schafer B. 2005. RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe*. *Gene* 354:80-85.
- Scheffler IE. 2001. A century of mitochondrial research: achievements and perspectives. *Mitochondrion* 1:3-31.
- Schuster FL, Goldstein S, Hershenov B. 1968. Ultrastructure of a flagellate, *Isonema nigricans* nov. gen. nov. sp., from a polluted marine habitat. *Protistologica* 4:141-149.
- Shapiro TA, Englund PT. 1995. The structure and replication of kinetoplast DNA. *Annu Rev Microbiol* 49:117-143.
- Shapiro TA, Klein VA, Englund PT. 1999. Isolation of kinetoplast DNA. *Methods in molecular biology (Clifton, NJ)* 94:61-67.
- Shikanai T. 2006. RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol Life Sci* 63:698-708.
- Simpson AG, Lukes J, Roger AJ. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. *Mol Biol Evol* 19:2071-2083.
- Simpson AG, Roger AJ. 2004. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol Phylogenet Evol* 30:201-212.
- Simpson AGB. 1997. The identity and composition of the Euglenozoa. *Arch Protistenkd* 148:318-328.
- Simpson L, Da Silva A. 1971. Isolation and characterization of kinetoplast DNA from *Leishmania tarentolae*. *J Mol Biol* 56:443-473.
- Slomovic S, Laufer D, Geiger D, Schuster G. 2005. Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. *Mol Cell Biol* 25:6427-6435.
- Smith SB, Bendich AJ. 1990. Electrophoretic charge density and persistence length of DNA as measured by fluorescence microscopy. *Biopolymers* 29:1167-1173.
- Smith SW, Overbeek R, Woese CR, Gilbert W, Gillevet PM. 1994. The genetic data environment an expandable GUI for multiple sequence analysis. *Comput Appl Biosci* 10:671-675.
- Staden R. 1996. The Staden sequence analysis package. *Mol Biotechnol* 5:233-241.
- Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing. *J Mol Evol* 48:303-312.

- Stoddard BL. 2005. Homing endonuclease structure and function. *Quarterly reviews of biophysics* 38:49-95.
- Stolba P, Jirku M, Lukes J. 2001. Polykinetoplast DNA structure in *Dimastigella trypaniformis* and *Dimastigella mimosa* (Kinetoplastida). *Mol Biochem Parasitol* 113:323-326.
- Stuart K, Panigrahi AK. 2002. RNA editing: complexity and complications. *Mol Microbiol* 45:591-596.
- Suyama Y. 1986. Two dimensional polyacrylamide gel electrophoresis analysis of Tetrahymena mitochondrial tRNA. *Curr Genet* 10:411-420.
- Suyama Y, Miura K. 1968. Size and Structural Variations of Mitochondrial DNA. *Proc Natl Acad Sci U S A* 60:235-242.
- Tessier LH, van der Speck H, Gualberto JM, Grienenberger JM. 1997. The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. *Curr Genet* 31:208-213.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
- Triemer RE, Ott DW. 1990. Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema*. *Europ J Protistol* 25:316-320.
- Turmel M, Otis C, Lemieux C. 2002. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol Biol Evol* 19:24-38.
- Vaidya AB, Akella R, Suplick K. 1989. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol* 35:97-107.
- Vickerman K. 2000. Diplonemids. In: Lee JJ, Leedale GF, Bradbury P, eds. *An illustrated guide to the protozoa*. Kansas: Allen Press, Lawrence. pp 1157-1159.
- Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). *Cell* 25:793-803.
- Waring RB, Davies RW, Scazzocchio C, Brown TA. 1982. Internal structure of a mitochondrial intron of *Aspergillus nidulans*. *Proc Natl Acad Sci U S A* 79:6332-6336.
- Watanabe KI, Bessho Y, Kawasaki M, Hori H. 1999. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. *J Mol Biol* 286:645-650.
- Wesolowski M, Fukuhara H. 1981. Linear mitochondrial deoxyribonucleic acid from the yeast *Hansenula mrakii*. *Mol Cell Biol* 1:387-393.
- Worthey EA, Schnauffer A, Mian IS, Stuart K, Salavati R. 2003. Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res* 31:6392-6408.
- Yasuhira S, Simpson L. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. *J Mol Evol* 44:341-347.
- Yoshikawa S, Shinzawa-Itoh K, Tsukihara T. 1998. Crystal structure of bovine heart cytochrome c oxidase at 2.8 Å resolution. *J Bioenerg Biomembr* 30:7-14.

- Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155-159.
- Zikova A, Vancova M, Jirku M, Lukes J. 2003. *Cruzella marina* (Bodonina, Kinetoplastida): non-catenated structure of poly-kinetoplast DNA. *Exp Parasitol* 104:159-161.
- Perez-Martinez X, Antaramian A, Vazquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, Gonzalez-Halphen D. 2001. Subunit II of cytochrome c oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem* 276:11302-11309.
- Porter D. 1973. *Isonema papillatum* sp. n., a new colorless marine flagellate: A light- and electron microscopic study. *J Protozool* 20:351-356.
- Preisfeld A, Busse I, Klingberg M, Talke S, Ruppel HG. 2001. Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *International journal of systematic and evolutionary microbiology* 51:751-758.
- Reichert AS, Neupert W. 2004. Mitochondriomics or what makes us breathe. *Trends Genet* 20:555-562.
- Renger HC, Wolstenholme DR. 1970. Kinetoplast deoxyribonucleic acid of the hemoflagellate *Trypanosoma lewisi*. *The Journal of cell biology* 47:689-702.
- Riou G, Delain E. 1969. Electron microscopy of the circular kinetoplastic DNA from *Trypanosoma cruzi*: occurrence of catenated forms. *Proceedings of the National Academy of Sciences of the United States of America* 62:210-217.
- Roy J, Faktorova D, Lukes J, Burger G. 2007. Unusual Mitochondrial Genome Structures throughout Euglenozoa. *Protist in press*.
- Ryan R, Grant D, Chiang KS, Swift H. 1978. Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences of the United States of America* 75:3268-3272.
- Schafer B. 2005. RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe*. *Gene* 354:80-85.
- Scheffler IE. 2001. A century of mitochondrial research: achievements and perspectives. *Mitochondrion* 1:3-31.
- Schuster FL, Goldstein S, Hershenov B. 1968. Ultrastructure of a flagellate, *Isonema nigricans* nov. gen. nov. sp., from a polluted marine habitat. *Protistologica* 4:141-149.
- Shapiro TA, Klein VA, Englund PT. 1999. Isolation of kinetoplast DNA. *Methods in molecular biology (Clifton, NJ)* 94:61-67.
- Simpson AG, Lukes J, Roger AJ. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. *Mol Biol Evol* 19:2071-2083.
- Simpson AG, Roger AJ. 2004. Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol Phylogenet Evol* 30:201-212.
- Simpson AGB. 1997. The identity and composition of the Euglenozoa. *Arch Protistenkd* 148:318-328.

- Simpson L, Da Silva A. 1971. Isolation and characterization of kinetoplast DNA from *Leishmania tarentolae*. *J Mol Biol* 56:443-473.
- Smith SB, Bendich AJ. 1990. Electrophoretic charge density and persistence length of DNA as measured by fluorescence microscopy. *Biopolymers* 29:1167-1173.
- Stoddard BL. 2005. Homing endonuclease structure and function. *Quarterly reviews of biophysics* 38:49-95.
- Stolba P, Jirku M, Lukes J. 2001. Polykinetoplast DNA structure in *Dimastigella trypaniformis* and *Dimastigella mimosa* (Kinetoplastida). *Mol Biochem Parasitol* 113:323-326.
- Stuart K, Panigrahi AK. 2002. RNA editing: complexity and complications. *Mol Microbiol* 45:591-596.
- Suyama Y. 1986. Two dimensional polyacrylamide gel electrophoresis analysis of *Tetrahymena* mitochondrial tRNA. *Curr Genet* 10:411-420.
- Suyama Y, Miura K. 1968. Size and Structural Variations of Mitochondrial DNA. *Proc Natl Acad Sci U S A* 60:235-242.
- Tessier LH, van der Speck H, Gualberto JM, Grienenberger JM. 1997. The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. *Curr Genet* 31:208-213.
- Triemer RE, Ott DW. 1990. Ultrastructure of *Diplonema ambulator* Larsen & Patterson (Euglenozoa) and its relationship to *Isonema*. *Europ J Protistol* 25:316-320.
- Vaidya AB, Akella R, Suplick K. 1989. Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol* 35:97-107.
- Vickerman K. 2000. Diplonemids. In: Lee JJ, Leedale GF, Bradbury P, eds. *An illustrated guide to the protozoa*. Kansas: Allen Press, Lawrence. pp 1157-1159.
- Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). *Cell* 25:793-803.
- Waring RB, Davies RW, Scazzocchio C, Brown TA. 1982. Internal structure of a mitochondrial intron of *Aspergillus nidulans*. *Proc Natl Acad Sci U S A* 79:6332-6336.
- Watanabe KI, Bessho Y, Kawasaki M, Hori H. 1999. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. *J Mol Biol* 286:645-650.
- Wesolowski M, Fukuhara H. 1981. Linear mitochondrial deoxyribonucleic acid from the yeast *Hansenula mrakii*. *Mol Cell Biol* 1:387-393.
- Yasuhira S, Simpson L. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. *J Mol Evol* 44:341-347.
- Zikova A, Vancova M, Jirku M, Lukes J. 2003. *Cruzella marina* (Bodonina, Kinetoplastida): non-catenated structure of poly-kinetoplast DNA. *Exp Parasitol* 104:159-161.

Annexe 1 : L'ADNmt des diplonémides

Tableau A1: ADNmt parmi les diplonémides

Espèces	Taille des chromosomes ^a	Fragmentation de <i>cox1</i> ^b
<i>D. papillatum</i>	6 et 7 kbp	oui
<i>D. ambulator</i>	5 et 7 kbp	oui
<i>D. sp.2</i>	5 kbp	?
<i>R. euleeides</i>	7 et 7.8 kbp	oui

^a L'ADNmt de *R. euleeides* a été extrait par gradient de CICs comme présenté dans notre premier article (partie 2.1). L'ADNmt de *D. sp2* et *D. ambulator* a été purifié à partir d'un extrait d'ADN total. La taille des chromosomes a été déterminée après migration sur gel d'agarose.

^b La fragmentation du gène *cox1* pour *R. euleeides* et *D. ambulator* a été déduite d'une hybridation d'ARN avec l'extrémité 3' de l'ADNc de *cox1* comme sonde (équivalent de la taille de l'exon 9 de *cox1* chez *D. papillatum*). Dans les deux cas, les hybridations révélèrent un patron similaire à celui connu chez *D. papillatum* avec 5 bandes au minimum.

Annexe 2 : l'expression du gène *cox1* au sein des diplonémides

<i>D. papillatum</i>	CAT	CGA	GGA	GGA	CTT	TTT	TCG	CTC	TAC	AAC
<i>D. ambulator</i>	CAC	AGG	GTA	GCT	GTT	TTT	TCA	CTA	TTC	AAT
<i>R. euleeides</i>	CAT	AGG	ACT	GCT	GTT	TTT	TCT	GTG	TAC	AAC
<i>D. papillatum</i>	H	R	G	G	L	F	S	L	Y	N
<i>D. ambulator</i>	H	R	V	A	V	F	S	L	F	N
<i>R. euleeides</i>	H	R	T	A	V	F	S	V	Y	N

Figure A1: Séquence de l'ADNc et de la protéine du gène *cox1* chez les diplonémides. La boîte rose souligne les 6T présents chez tous les ADNc de *cox1* des diplonémides.

