



HAL
open science

Evolution des génomes microsporidiens et mécanisme d'adaptation moléculaire chez les parasites intracellulaires obligatoires

Abdel Belkorchia

► **To cite this version:**

Abdel Belkorchia. Evolution des génomes microsporidiens et mécanisme d'adaptation moléculaire chez les parasites intracellulaires obligatoires. Biochimie, Biologie Moléculaire. Université Blaise Pascal - Clermont-Ferrand II; Université d'Auvergne - Clermont-Ferrand I, 2007. Français. NNT : 2007CLF21766 . tel-00718201

HAL Id: tel-00718201

<https://theses.hal.science/tel-00718201>

Submitted on 16 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE BLAISE PASCAL

UNIVERSITE D'AUVERGNE

N° D.U. 1766

Année 2007

**Ecole doctorale
des Sciences de la Vie et de la Santé
N° d'ordre 463**

THESE

pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : MICROBIOLOGIE

Présentée et soutenue publiquement par

Abdel BELKORCHIA

le 26 septembre 2007

**EVOLUTION DES GENOMES MICROSPORIDIENS ET MECANISMES
D'ADAPTATION MOLECULAIRE CHEZ LES PARASITES
INTRACELLULAIRES OBLIGATOIRES**

JURY

Présidente : Dr. Catherine BOURGOUIN, Institut Pasteur, Paris

Rapporteurs : Dr. Joël GRILLASCA, Université Sud Toulon-Var

Dr. Stanislas TOMAVO, Université de Lille

Examineur : Dr. Patrick WINCKER, Genoscope, Paris

Directeurs de thèse : Pr. Pierre PEYRET, Université Blaise Pascal, Clermont-Ferrand

Dr. Eric PEYRETAILLADE, Université Blaise Pascal, Clermont-Ferrand

Laboratoire de Biologie des Protistes UMR CNRS 6023, Equipe de Génomique Intégrée des Interactions Microbiennes, Université Blaise Pascal, Clermont-Ferrand.

UNIVERSITE BLAISE PASCAL

UNIVERSITE D'AUVERGNE

N° D.U. 1766

Année 2007

**Ecole doctorale
des Sciences de la Vie et de la Santé
N° d'ordre 463**

THESE

pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : MICROBIOLOGIE

Présentée et soutenue publiquement par

Abdel BELKORCHIA

le 26 septembre 2007

**EVOLUTION DES GENOMES MICROSPORIDIENS ET MECANISMES
D'ADAPTATION MOLECULAIRE CHEZ LES PARASITES
INTRACELLULAIRES OBLIGATOIRES**

JURY

Présidente : Dr. Catherine BOURGOUIN, Institut Pasteur, Paris

Rapporteurs : Dr. Joël GRILLASCA, Université Sud Toulon-Var

Dr. Stanislas TOMAVO, Université de Lille

Examineur : Dr. Patrick WINCKER, Genoscope, Paris

Directeurs de thèse : Pr. Pierre PEYRET, Université Blaise Pascal, Clermont-Ferrand

Dr. Eric PEYRETAILLADE, Université Blaise Pascal, Clermont-Ferrand

Laboratoire de Biologie des Protistes UMR CNRS 6023, Equipe de Génomique Intégrée des Interactions Microbiennes, Université Blaise Pascal, Clermont-Ferrand.

Remerciements

Ce travail a été réalisé au sein du Laboratoire de Biologie des Protistes, UMR CNRS 6023 de l'Université Blaise Pascal. Je tiens tout d'abord à remercier Monsieur Christian Amblard, Directeur du laboratoire, pour m'avoir accueilli dans son unité.

Je tiens à remercier le Conseil Général et la Communauté d'agglomération du Bassin d'Aurillac pour leur participation au financement de ce doctorat réalisé à l'IUT Génie Biologique d'Aurillac.

Je souhaite exprimer mes remerciements les plus sincères à Messieurs Joël GRILLASCA et Stanislas TOMAVO pour m'avoir fait l'honneur d'accepter d'être rapporteurs de cette thèse, ainsi qu'à Monsieur Patrick WINCKER et Madame Catherine BOURGOUIN pour avoir accepté de juger ce travail.

Ces travaux de thèse n'auraient jamais pu être réalisés sans le soutien de Pierre PEYRET dans la recherche d'un financement. Je tiens donc à lui rendre hommage non seulement pour son aide, mais également pour m'avoir encadré durant ces 4 années et soutenu en toutes circonstances... Merci Pierre pour toute la confiance que tu m'as témoignée, ton aide précieuse dans l'avancement de ces travaux, et tes corrections constructives indispensables à la rédaction de ce mémoire. Trouve ici toute l'expression de ma gratitude pour tout ce que j'ai pu apprendre à tes côtés à la fois sur le plan scientifique, mais également humain. Merci pour ces discussions très enrichissantes et ton éloquence admirable (!)...Enfin, pour m'avoir donné (avec Frédéric Delbac) l'envie de m'engager sur cette voie il y a quelques années...

Un grand merci également à Eric PEYRETAILLADE, pour son soutien et son encadrement, son aide dans l'avancée des travaux, la rédaction de ce mémoire et des publications, et pour m'avoir permis de partager avec lui une infime (!) partie de ses connaissances. Ce fut un véritable plaisir et une chance d'avoir pu travailler avec toi. Je te remercie Eric pour ces moments de détente indissociables du sérieux expérimental !

Que Corinne Biderre-Petit trouve ici toute l'expression de ma reconnaissance pour son aide dans l'avancée de ces travaux et la rédaction des publications, mais aussi pour avoir partagé avec moi ses connaissances et compétences scientifiques. Merci Coco pour ton soutien... Très heureux pour toi que la roue ait finie par tourner dans le bon sens... !

Merci à tous les membres de l'équipe qui ont participé au bon déroulement de cette thèse par leur bonne humeur : Brigitte, pour ta disponibilité et ton aide, et le plaisir que j'ai eu à travailler avec toi (notamment au Sénégal !), Anne, Cécile, Sébastien, Jérôme, Jean-Baptiste, Mohiédine, Delphine, aux stagiaires Eric, Bruno, Yann, Xavier, pour ces moments passés ensemble (j'en oublie sûrement...).

Merci Olivier pour ces moments privilégiés de détente et ces instants de poésie partagés...(!), Stéphanie, Philippe, Geneviève, Muriel et Valérie, ma famille d'adoption (!), pour ces 4 années passées avec vous et votre soutien.

Un grand merci à Frédéric Delbac pour ces discussions très enrichissantes. Merci Fred pour ton aide et tes nombreux conseils.

J'aimerais également remercier Michel Hébraud et Christophe Chambon pour leur accueil chaleureux au sein de la plate-forme protéomique de l'INRA de Theix.

Merci Nicolas pour ton amitié et ces festins improvisés au goût si relevé, Eric, pour tes coups de gueule légitimes (!) et Patricia, pour la rapidité d'obtention des publications !

Mes remerciements vont également à tous mes collègues d'enseignement de l'IUT Génie Biologique d'Aubière et d'Aurillac avec qui j'ai pris énormément de plaisir à travailler.

Un grand merci à Olivier Camarès pour son assistance technique, Maryse, Claude et Jean-Paul pour votre disponibilité et votre accueil toujours chaleureux.

Merci Robert, Odile et Matthieu, pour votre soutien et la confiance que vous me témoignez...

A Ahmed, Karim, Cécile et Alex, Fred et Aude, Damien et Annabelle pour votre amitié.

André, je voudrais te remercier pour tout ce que tu as fait pour moi durant ces dernières années. Ton amitié, ton soutien, ton écoute, ont été pour moi l'occasion de me réaliser à travers ce cursus universitaire, parfois hésitant, et pour lequel tu m'as toujours soutenu et encouragé. Merci pour ta confiance Dédé, et pour m'avoir donné cet amour de la microbiologie à travers ton expérience et ta passion pour les champignons.

Pour terminer, j'aimerais dédier cette thèse à ma famille : à mes parents, toujours présents, pour m'avoir soutenu et permis de faire des études dans les meilleures conditions, à mes frères, pour leur aide en toutes circonstances, et enfin à ma sœur Safya, pour avoir apporté à cette grande famille cette joie et ce plaisir de nous retrouver tous ensemble. Mon plus grand regret : celui de ne pas avoir été à tes côtés pour te soutenir davantage et t'aider dans tes études...

A mon neveu Ahmed, mes nièces Amel et Nour, pour tous ces moments passés avec vous, pour m'avoir redonné mes 4 ans et permis de savourer quelques rares moments dans l'insouciance la plus totale. Merci à vous.

« Lorsque tout me ravit, j'ignore si quelque chose me séduit.
Elle éblouit comme l'Aurore et console comme la Nuit ... »

A Valérie...

SOMMAIRE

SOMMAIRE

LISTE DES ABREVIATIONS	1
RESUME / ABSTRACT	4
INTRODUCTION	5
<hr/>	
ETUDE BIBLIOGRAPHIQUE	8
<hr/>	
I/ LES MICROSPORIDIÉS	8
I.1. Introduction	8
I.1.1. Un phylum très diversifié	8
I.1.2. Tableaux cliniques	9
I.1.3. Epidémiologie	10
I.1.4. Diagnostic	11
I.1.5. Thérapie	12
I.2. Phylogénie	12
I.3. Organisation structurale de la spore	15
I.3.1. L'élément infectieux : le sporoplasme	15
I.3.2. L'appareil invasif	15
I.3.3. L'enveloppe sporale	17
II/ GENOMIQUE STRUCTURALE DES MICROSPORIDIÉS	18
II.1. Structure des génomes microsporidiens	18
II.1.1. Caryotype moléculaire	18
II.1.2. Organisation génique	19
II.1.2.1. Les séquences codantes	19
II.1.2.2. Les séquences intergéniques	21
II.1.2.3. Les introns	23
II.1.3. Les séquences répétées	24
II.1.3.1. Les éléments transposables	24
II.1.3.2. Les séquences satellites	26
II.1.3.3. Les ARN ribosomiques microsporidiens	27
II. 2. Génomique comparative	29

III/ REGULATION GENIQUE CHEZ LES PROTOZOAIRE PARASITES	32
III.1. Régulation transcriptionnelle	32
III.1.1. La pré-initiation de la transcription	32
III.1.2. Le complexe basal d'initiation de la transcription	33
III.1.3. L'élongation de la transcription	34
III.1.4. La terminaison de la transcription	34
III.1.5. Les signaux de transcription chez les protozoaires	35
III.2. Mécanismes de régulation post-transcriptionnelle	40
III.2.1. L'épissage en trans (trans-splicing)	40
III.2.2. Epissage alternatif	41
III.2.3. Régulation par expression d'ARN antisens	42
III.2.4. Régulation par ARN interférence	43
III.3. Régulation traductionnelle	44
III.3.1. L'initiation de la traduction chez les eucaryotes	45
III.3.2. Balayage de l'extrémité 5' UTR (ou scanning)	46
III.3.3. Reconnaissance du codon AUG	47
III.3.4. 5' UTR réduit et absence de scanning	48
IV/ INTERACTIONS HOTES PARASITES	49
IV.1. Le modèle microspodien	49
IV.1.1. Les différents modes d'infestation	49
IV.1.2. Germination sporale et internalisation du parasite	50
IV.1.3. Cycle de développement	52
IV.2. <i>Plasmodium falciparum</i>	53
IV.2.1. Cycle de développement de <i>P. falciparum</i>	53
IV.2.2. Mécanisme d'adaptation parasitaire	55
IV.2.3. Etudes des interactions <i>Anopheles gambiae</i> / <i>P. falciparum</i>	57

V/ LE MODELE D'ETUDE <i>BRACHIOLA ALGERAE</i> /<i>ANOPHELES GAMBIAE</i>	61
V.1. La microsporidie <i>Brachiola algerae</i>	61
V.1.1. Cycle cellulaire	61
V.1.2. Capacité d'adaptation	61
V.1.3. Réduction de la compétence vectorielle des moustiques	63
V.2. <i>Anopheles gambiae</i>	63
V.2.1. Le genre <i>Anopheles</i>	63
V.2.2. L'immunité chez les insectes	64
V.2.2.1. L'immunité cellulaire	64
V.2.2.2. L'immunité humorale	64
V.3. Immunité et compétence vectorielle	65
V.4. Invalidation de gènes candidats	66
<hr/>	
MATERIEL ET METHODES	68
<hr/>	
I. ETUDE DES MICROSPORIDIIES	68
I.1. Les espèces microsporidiennes étudiées	68
I.2. Production et purification des parasites	68
I.2.1. <i>Encephalitozoon cuniculi</i>	68
I.2.2. <i>Brachiola algerae</i>	68
I.3. Mise en évidence d'une organisation en opéron chez <i>E. cuniculi</i>	69
I.3.1. Extraction des ARN totaux	69
I.3.2. RACE-PCR et clonages	69
I.3.3. Clonage des produits PCR	69
I.3.3.1. Ligatures	69
I.3.3.2. Préparation des bactéries chimio-compétentes et transformation	70
I.3.3.3. Extraction d'ADN plasmidique	70
I.3.4. Séquençage des vecteurs recombinants	70
I.3.5. Calcul des énergies libres d'appariement (SSU ARNr-ARNm)	70

I.4. Etude du génome de <i>B. algerae</i>	70
I.4.1. Réalisation de caryotypes moléculaires	70
I.4.1.1. Préparation des blocs d'agarose	70
I.4.1.2. Electrophorèse en champs pulsés	71
I.4.2. Stratégie de séquençage	71
I.4.2.1 Réalisation des banques d'ADN génomique	71
I.4.2.2 Analyses bioinformatiques	72
I.4.3. Détection des unités ADNr par Southern-blot	72
I.4.3.1. Transfert sur membrane de nylon	72
I.4.3.2. Préparation et marquage des sondes	72
I.4.3.2.1. Préparation des sondes	72
I.4.3.2.2. Marquage des sondes	73
I.4.3.3. Hybridation et autoradiographie	73
I.4.4. Construction de la structure secondaire de l'ARNr 23S	73
II. ETUDE DES RELATIONS HOTES-PATHOGENES	74
II.1. Description des lignées cellulaires d'<i>An. gambiae</i> Sua4.0 et mos 55	74
II.2. Culture des lignées cellulaires	74
II.3. Etude de l'expression des variants du gène <i>cap</i>	74
II.3.1. Matériel biologique étudié	74
II.3.2. Stimulations antigéniques	74
II.3.3. Extraction des ARNs et RT-PCR	75
II.3.4. Analyses bioinformatiques	75
II.4. Essais de transfection de la lignée cellulaire Sua4.0	75
II.4.1. Constructions plasmidiques réalisées	75
II.4.2. Les différentes techniques de transfection utilisées	76
II.4.2.1. La lipofection	76
II.4.2.2. La magnétofection	77
II.4.3. Quantification des transformants	77
II.4.4. Inactivation par ARN interférence	77
II.4.4.1. Préparation des ARNdb	77
II.4.4.2. Transfection de la lignée hémocytaire Sua4.0	77

II.5. Identification des protéines codées par le gène <i>cap</i>	78
II.5.1. Production d'anticorps polyclonaux	78
II.5.1.1. Expression d'une protéine 14Yde recombinante chez <i>E. coli</i>	78
II.5.1.2. Immunisation	78
II.5.2. Electrophorèse monodimensionnelle	78
II.5.3. Electrophorèse bidimensionnelle	79
II.5.4. Western blot	79
II.5.5. Identification protéique par spectrométrie de masse	79
II.6. Essais d'infestation des cellules d'insectes par <i>B. algerae</i>	80
II.6.1. Infestation des lignées cellulaires	80
II.6.2. Mise en évidence de l'infestation	80
II.6.2.1. Production d'un sérum total anti- <i>B. algerae</i>	80
II.6.2.2. IFA	81
<hr/>	
RESULTATS	82
<hr/>	
I. UN NOUVEAU MECANISME DE REGULATION TRANSCRIPTIONNELLE CHEZ LA MICROSPORIDIE <i>ENCEPHALITOOZON CUNICULI</i>	82
A. SYNOPSIS	82
B. PUBLICATION 1 : « Genome evolution of the eukaryotic parasite <i>Encephalitozoon cuniculi</i> is related to polycistronic gene clusters »	84
C. DISCUSSION	99
II. <i>BRACHIOLA ALGERAE</i> : MICROSPORIDIE MONTRANT DE REMARQUABLES CAPACITES D'ADAPTATION A LA VIE PARASITAIRE ET PRESENTANT UN GENOME COMPLEXE DE 23 MPB	103
A. SYNOPSIS	103
B. PUBLICATION 2 : « <i>In vitro</i> propagation of the microsporidian pathogen <i>Brachiola algerae</i> and studies of its chromosome and ribosomal DNA organization in the context of the complete genome sequencing project »	105

C. DISCUSSION	115
III. DEVELOPPEMENT DE NOUVEAUX OUTILS DE GENOMIQUE FONCTIONNELLE CHEZ <i>AN. GAMBIAE</i>	122
III.1. Transfection des lignées cellulaires d' <i>An. gambiae</i>	123
III.2. Evaluation de l'efficacité de l'ARN interférence en culture de cellules d' <i>An. gambiae</i>	125
III.3. Etude de gènes candidats surexprimés chez le moustique au cours d'une infestation parasitaire	127
III.3.1. <i>14Yde</i> : un nouveau variant d'épissage du gène <i>cap</i> ?	127
III.3.2. Caractérisation d'expression des variants du gène <i>14Yde</i> par RT-PCR	129
III.3.3. Recherche d'expression des protéines codées par le gène <i>cap</i>	129
III.4. CONCLUSION	132
CONCLUSION-PERSPECTIVES	135
REFERENCES BIBLIOGRAPHIQUES	138

LISTE DES ABREVIATIONS

A : adénine
aa : acides aminés
ACT : Artemis Comparison Tool
ADN : Acide DésoxyriboNucléique
ARN : Acide RiboNucléique
ARNi : ARN interference
ARNnc : ARN non codant
ATP : Adénosine TriPhosphate
BAC : Bacterial Artificial Chromosome
BCIP : 5-Bromo-4-Chloro-3-Indolyl Phosphate
BLAST : Basic Local Alignment Search Tool
BREs : TFIIB Recognition Elements
C : cytosine
CDK : Cyclin-Dependent Kinase
CDS : Coding DNA Sequence
CF : Cleavage Factor
CHEF : Coutour-clamped Homogeneous Electric Field
CPSF : Cleavage and Polyadenylation Specificity Factor
CS : Circumsporozoite
CstF : Cleavage Stimulatory Factor
CTD : Carboxy-Terminal-Domain
CTL4 : C-type Lectin 4
CTLMA2 : C-type Lectin Mannose binding 2
CTRP : Circumsporozoite and TRAP-Related Protein
Da : Dalton
DAPI : 4',6-DiAmidino-2-PhenylIndole
DCE : Downstream Core Element
dNTP : DésoxyNucléotide TriPhosphate
DO : Densité Optique
DPE : Downstream Promoter Element
DSE : Downstream Sequence Element
DSIF : DRB-Sensitivity Inducing Factor
DTT : DiThioThréitol
EDTA : Ethylene Diamine Tetra-acetic Acid
EGFP : Enhanced Green Fluorescent Protein
EST : Expressed Sequence Tag
ET : Elément transposable
G : guanine
GAG : GlycoAminoGlycanes
GST : Glutathion-S-Transférase
GT : GuanlylTransférase
HFF : Human Foreskin Fibroblast
His : histidine

HSP 70 : Heat Shock Protein 70
HVS : Herpès Virus Simplex
IFA : ImmunoFluorescent Assay
IMPDH: Inosine Monophosphate Deshydrogenase
Inr : Initiator Element
IPTG : Iso Propyl- β -D Thio-Galactopyranoside
ITS : Intergenic Transcribed Spacer
kb : kilobase
kDa : kiloDalton
LB: Luria Bertani
LMP : Low Melting Point
LRIM 1 : Leucin rich-repeat Immune gene 1
LSU : Large SubUnit
MAOP : Membrane Attack Ookinete Protein
Mb : mégabase
MDCK : Madin-Darby Canine Kidney
MEM : Milieu Minimum Essentiel
miARN : microARN
Mpb : Méga-paires de bases
MT : MethylTransférase
MTE: Motif Ten Element
NBT : Nitro Blue Tetrazolium
NELF : Negative Elongation Factor Complex
nt : nucléotide
ORF : Open Reading Frame
PABP : Poly(A) Binding Protein
pb : paire de bases
PBS : Phosphate Buffer Saline
PCR : Polymerase Chain Reaction
PRRs : Pattern Recognition Receptors
P-TEFb: Positive Transcription Elongation Factor b
PVDF : PolyVinylidène DiFluoride
RACE : Rapid Amplification of cDNA Ends
RT : Reverse Transcripase
SDS : Sodium Dodecyl Sulfate
SDS-PAGE : Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
shRNA : short hairpin RNA
siARN : small interfering ARN
SIDA : Syndrome d'Immuno Déficience Acquis
snRNP : Small Nuclear RiboNucleoProtein
SOAP : Secreted Ookinete Adhesive Protein
SSU : Small SubUnit
SVF : Sérum de Veau Fœtal
T : thymine
TAF : TBP-Associated Factor
TBE : Tris Borate EDTA

TBP : TATA Binding Protein
TBP: TriButylPhosphine
TBS : Tris Buffer Saline
TEP 1 : Thioester-containing Protein 1
TRAP : Thrombospondin-Related Adhesive Protein
TRITC : TetramethylRhodamine IsoThioCyanate
UTR : UnTranslated Region
UV : Ultra Violet
VIH : Virus de l'Immunodéficience Humaine
VSG : Variable Surface Glycoprotein
WARP : von Willebrand factor A domain Related Protein
X-Gal : 5-Bromo-4-chloro-3-indolyl- β -D-Galactoside

Résumé

Les microsporidies sont des parasites intracellulaires obligatoires caractérisés par un mode d'infestation original et, pour certaines espèces, par un très fort degré de compaction du génome nucléaire. Afin de contribuer à une meilleure compréhension des processus d'expression génique chez ces pathogènes opportunistes de l'Homme, des travaux portant sur les mécanismes de régulation transcriptionnelle ont été entrepris. Nos résultats montrent qu'une organisation polycistronique des gènes est retrouvée chez *Encephalitozoon cuniculi* (parasite de mammifères ayant une taille de génome de 2,9 Mb). Au cours du processus de transcription, une maturation aléatoire des messagers conduit à la formation d'une population d'ARNm comprenant des ARNm monocistroniques et polycistroniques. De plus, afin d'améliorer nos connaissances sur la structure des génomes microsporidiens mais également sur leurs capacités d'adaptation, le séquençage du génome de *Brachiola algerae* (microsporidie parasite de mammifères et d'insectes) a été initié en collaboration avec le Genoscope. Trente bandes d'ADN chromosomique de 160 à 2240 kb ont pu être visualisées, pour une taille haploïde de génome estimée à 23 Mb. L'hybridation moléculaire d'une sonde ciblant l'ARNr 16S nous a permis de localiser des unités d'ADNr sur au moins 9 des 30 bandes d'ADN chromosomique. Les premières analyses de ce génome AT riche (>75%) montrent la présence de plusieurs classes d'éléments transposables ainsi que des gènes codant des enzymes impliquées dans le processus d'ARN interférence. Enfin, le développement d'un modèle d'étude *in vitro* des interactions *An. gambiae* / pathogènes a été réalisé, grâce à la démonstration des capacités d'infestation des lignées cellulaires Sua4.0 (cellules hématocytaires) et Mos 55 (cellules embryonnaires) par *B. algerae*. L'invalidation génique par ARN interférence sur ces lignées cellulaires devrait concourir à une meilleure compréhension des mécanismes d'adaptation des parasites *P. falciparum* et *B. algerae* chez *An. gambiae*.

Mots-clé : Microsporidies, génome, opéron, régulation transcriptionnelle, adaptation parasitaire

Abstract

Microsporidia are obligate intracellular parasites characterized by a unique invasive mechanism, and, in some species, by a highly compact nuclear genome. To improve our knowledge of gene expression in these opportunistic pathogens, experiments on transcriptional regulation mechanisms were undertaken. In the microsporidia *Encephalitozoon cuniculi* (a mammalian parasite with a genome size of 2.9 Mb), we demonstrated a polycistronic gene organization. At the time of transcription, the transcripts are randomly processed leading to an mRNA population composed of monocistronic and polycistronic mRNA. To study the structure of microsporidian genomes and their adaptation capacities, the sequencing of *Brachiola algerae* genome (an insect and mammalian microsporidia) has been initiated in collaboration with Genoscope. Thirty chromosome bands from 160 to 2240 kb were identified, giving a haploid genome size estimated of 23 Mb. The study of the chromosomal distribution of the 16S rRNA unit showed the presence of 16S copy on at least 9 chromosome bands. The first analysis of this AT rich genome (more than 75%) revealed the presence of different kinds of transposable elements as well as genes encoding for enzymes involved in the RNA interference process. We further demonstrated that two *Anopheles gambiae* cellular lineages Sua4.0 (haemocytic cell) and Mos 55 (embryonic cells) can be infected by *B. algerae* providing a model to study *in vitro* the interaction between *An. gambiae* and its pathogens. The successful use of the RNA interference in these two types of cells will contribute to a better understanding of the adaptation mechanisms of pathogens like *P. falciparum* and *B. algerae*.

Key words: Microsporidia, genome, operon, transcriptional regulation, parasitic adaptation

INTRODUCTION

Introduction

Dans le cas d'un parasitisme intracellulaire obligatoire, la relation de dépendance du parasite envers l'hôte est poussée à l'extrême. Ainsi, plusieurs questions fondamentales de parasitologie consistent à comprendre quels sont les mécanismes mis en place par un organisme pour pénétrer, survivre et se multiplier à l'intérieur d'une cellule hôte. De plus, d'un point de vue adaptatif, quel peut être l'impact d'un tel mode de développement sur l'évolution des génomes des parasites et les mécanismes de régulations géniques qui en découlent (Meissner et Soldati, 2005) ? La maîtrise de cette connaissance apparaît comme un pré-requis indispensable à la mise en place de stratégies préventives et thérapeutiques des maladies infectieuses.

Dans le monde très diversifié des eucaryotes unicellulaires, le phylum des *Apicomplexa* est un exemple de groupe représenté par de très nombreuses espèces ayant une importance médicale et vétérinaire considérable, certaines espèces faisant l'objet de recherches intensives (*Plasmodium falciparum*, *Toxoplasma gondii*, etc.). Le phylum des *Microspora* est un autre exemple important regroupant de très nombreuses espèces dont l'importance médicale a été révélée avec la pandémie de SIDA. Ces organismes sont reconnus comme étant des pathogènes opportunistes majeurs responsables de nombreuses infections chez les immunodéprimés (Didier, 2005). Les microsporidies se distinguent principalement des autres parasites intracellulaires par leur mode d'invasion unique, l'extrême réduction de la taille du génome nucléaire de certaines espèces, mais également leur très grande capacité d'adaptation (Franzen et *al.*, 2006).

L'étude du génome de la microsporidie *Encephalitozoon cuniculi* a révélé l'existence d'un génome extrêmement compact de 2,9 Mpb, codant pour environ 2000 CDS (Katinka et *al.*, 2001). L'analyse préalable d'un contig de 4,8 kb avait permis de mettre en évidence une forte réduction de la taille des régions intergéniques, laissant supposer l'existence d'un

système de régulation transcriptionnelle original (Duffieux et *al.*, 1998). Dans le but d'améliorer nos connaissances sur les mécanismes de régulation transcriptionnelle chez les microsporidies, nous avons entrepris dans la première partie de cette thèse, une caractérisation précise des messagers microsporidiens. La mise en évidence de différentes populations d'ARNm, comprenant des ARN monocistroniques et polycistroniques, nous a permis d'envisager l'existence d'une organisation de type opéron. Cette organisation génique pourrait être la conséquence d'une adaptation extrême à la vie parasitaire dans une niche écologique stable.

Chez *E. cuniculi*, l'adaptation à un mode de vie intracellulaire obligatoire a conduit à une perte de gènes, voire à une élimination complète de certaines voies métaboliques. Aussi, dans un second temps, afin d'évaluer les capacités d'évolution des génomes microsporidiens en relation avec leurs capacités d'infestation, nous avons entrepris le séquençage d'une autre espèce microsporidienne, *Brachiola algerae*, et ce, en collaboration avec le Genoscope. Les données de résolution du caryotype moléculaire par électrophorèse en champs pulsés, la reconstruction de la structure secondaire de l'ARNr 23S, et les premiers résultats d'annotation seront présentés. L'annotation fonctionnelle de ce génome a révélé l'existence d'un répertoire génique conséquent probablement à l'origine du large spectre d'hôtes de cette espèce. Il a de plus été montré la présence d'un grand nombre d'éléments mobiles qui étaient absents du génome d'*E. cuniculi*. Ces éléments mobiles pourraient participer aux mécanismes d'adaptations parasitaires suite à des changements environnementaux brutaux (infestations d'hôtes très différents tels que les mammifères et les diptères).

Enfin, dans une troisième partie nous présenterons le développement d'un modèle d'étude *in vitro* des interactions *An. gambiae* / pathogènes (*P. falciparum* et *B. algerae*). Dans ce contexte, nous avons préalablement démontré les capacités d'infestation des lignées cellulaires Sua4.0 (cellules hémocytaires) et Mos55 (cellules embryonnaires) par la

microsporidie *B. algerae*. Nous nous sommes ensuite attachés à optimiser la transfection de ces lignées cellulaires et à développer les outils de génétique fonctionnelle permettant l'inactivation génique par ARN interférence. Les résultats préliminaires sont prometteurs et devraient permettre l'utilisation de ce modèle cellulaire d'*An. gambiae* pour l'optimisation de l'interférence ARN par ARNdb, siARN ou miARN. Ces données seront alors utilisées pour effectuer une inactivation efficace *in toto* des gènes candidats de l'anophèle potentiellement impliqués dans le développement de pathogènes comme *P. falciparum* ou comme la microsporidie *B. algerae*. L'utilisation de ces modèles cellulaires devraient également contribuer à une meilleure compréhension des mécanismes d'adaptation hôte / pathogènes, notamment pour les interactions entre le moustique *An. gambiae* et la microsporidie *B. algerae*. En effet, ce pathogène est capable d'effectuer l'intégralité de son cycle dans ces lignées cellulaires.

BIBLIOGRAPHIE

I / LES MICROSPORIDIÉS

I.1. Introduction

Les microsporidies sont des organismes eucaryotes, parasites intracellulaires obligatoires, capables d'infester quasiment tout le règne animal (revue par **Didier et Weiss, 2006**). L'intérêt pour ces parasites s'est accru ces dernières décennies en raison de leur implication croissante en pathologie humaine, notamment chez les personnes atteintes du SIDA (**Desportes, 1985 ; Didier, 2005**). Même si les modalités de transmission parasitaire restent à définir, la caractérisation d'espèces possédant le même génotype, chez l'Homme et l'animal, conforte l'hypothèse d'une transmission zoonotique (**Mathis et al., 2005**). Ces organismes, ubiquitaires, existent dans l'environnement sous une forme de résistance sporale, et se distinguent dans le monde du vivant par une structure d'invasion unique, le tube polaire. L'étude du génome de certaines espèces a révélé un fort degré de compaction génique et l'élimination de nombreuses voies métaboliques en accord avec leur statut de parasites intracellulaires obligatoires (**Katinka et al., 2001 ; Keeling et al., 2004**).

I.1.1. Un phylum très diversifié

Le phylum des microsporidies regroupe quelques 1200 espèces réparties en plus de 143 genres (**Didier, 2005**). Ces parasites eucaryotes unicellulaires forment des spores très résistantes dont la taille, en fonction des espèces, varie de 1-3 μm à 1,5-4 μm . Les premiers cas de microsporidiose remontent au XIX^{ème} siècle. L'agent étiologique de la pébrine, *Nosema bombycis*, qui dévasta les élevages de vers à soie en Europe fut identifié par Nägeli en 1857. La première microsporidie parasite de mammifères fut isolée en 1922 au niveau du cerveau d'un lapin présentant une encéphalite granulomateuse : elle fut nommée *Encephalitozoon cuniculi* (**Levaditi et al., 1923**). Cette espèce a depuis été décrite chez plusieurs espèces de rongeurs et de nombreux mammifères dont l'Homme (**Didier, 2005**). Le

Tableau 1. Microsporidioses humaines et manifestations cliniques (d'après Franzen et Müller, 2001). Les espèces (1) n'ont été mises en évidence que chez des patients atteints de SIDA et les espèces (2) n'ont été identifiées que chez des individus immunocompétents. Les autres espèces ont été répertoriées dans les deux cas de figure.

Espèces microsporidiennes	Manifestations cliniques
<i>Enterocytozoon bieneusi</i>	Diarrhées, cholangites, cholécystites, pneumonies, bronchites, sinusites, rhinites
<i>Encephalitozoon intestinalis</i>	Diarrhées, cholangites, cholécystites, néphrites, infections urinaires, pneumonies, sinusites, rhinites, kératoconjonctivites, infections disséminées
<i>Encephalitozoon hellem</i> ⁽¹⁾	Kératoconjonctivites, pneumonies, sinusites, rhinites, bronchites, néphrites, urétrites, prostatites, cystites, infections disséminées
<i>Encephalitozoon cuniculi</i>	Hépatites, péritonites, encéphalites, infections intestinales, infections urinaires, kératoconjonctivites, sinusites, rhinites, infections disséminées.
<i>Trachipleistophora hominis</i> ⁽¹⁾	Infections musculaires, kératoconjonctivites, sinusites, rhinites
<i>Trachipleistophora anthropophthora</i> ⁽¹⁾	Encéphalites, infections musculaires, infections disséminées
<i>Pleistophora</i> spp.	Infections musculaires
<i>Vittaforma corneae</i> ⁽²⁾	Kératites, infections urinaires
<i>Nosema ocularum</i> ⁽²⁾	Kératoconjonctivites
<i>Brachiola connori</i> ⁽²⁾	Infections disséminées
<i>Brachiola vesicularum</i> ⁽¹⁾	Infections musculaires
<i>Brachiola algerae</i>	Kératoconjonctivites, infections musculaires, atteintes cutanées
<i>Microsporidium africanum</i> ⁽²⁾	Infections oculaires
<i>Microsporidium ceylonensis</i> ⁽²⁾	Infections oculaires

premier cas de microsporidiose humaine a, quant à lui, été rapporté en 1959 chez un enfant japonais de 9 ans souffrant de désordres neurologiques (**Matsubayashi et al., 1959**). Les microsporidies furent cependant très peu étudiées avant 1985 et la pandémie de SIDA.

I.1.2. Tableaux cliniques

La majorité des cas de microsporidioses recensés concernent des sujets immunodéprimés. Huit genres de microsporidies sont pathogènes pour l'Homme et provoquent des pathologies diverses pouvant conduire à la mort (**Tableau 1**). Les principales espèces responsables de pathologies chez l'Homme sont *Enterocytozoon bieneusi* et celles appartenant au genre *Encephalitozoon*. Ainsi, 14 espèces sont aujourd'hui reconnues comme de véritables pathogènes opportunistes (**Didier et Weiss, 2006**). L'intensité et la durée des manifestations cliniques (diarrhées chroniques, douleurs abdominales, cachexie, inflammation des voies biliaires, conjonctivite, myosite, etc.) dépendent du niveau de réactivité du système immunitaire chez les individus immunodéficients (**Maggi et al., 2000 ; Tumwine et al., 2005**). Ainsi, il a été montré que la prise de traitements antiretroviraux, qui stimulent la réponse immunitaire en augmentant le taux de lymphocytes TCD4+, entraîne la disparition des diarrhées chroniques (**Maggi et al., 2000 ; Wiwanitkit, 2006**). Certaines espèces, telle qu'*E. bieneusi* ont un tropisme particulier pour les systèmes hépatobiliaire et pulmonaire (**Weber et al., 2000 ; Sodqi et al., 2004**). D'autres, comme celles appartenant au genre *Encephalitozoon*, se caractérisent par leur capacité de dissémination dans tout l'organisme (**Fournier et al., 2000 ; Tosoni et al., 2002**). Enfin, quelques espèces des genres *Brachiola* ou *Trachipleistophora*, isolées principalement chez des personnes immunodéficientes, peuvent être responsables d'atteintes localisées (cornée) ou disséminées (muscle, rein, cerveau, etc.) (**Coyle et al., 2004 ; Juarez et al., 2005 ; Curry et al., 2005**).

I.1.3. Epidémiologie

La prévalence des microsporidies dans la population générale est difficile à appréhender car les études réalisées donnent des résultats variables selon la localisation géographique et la technique employée. Des études basées sur la recherche d'anticorps anti-microsporidiens établissent ainsi une prévalence allant de 1,3 à 22 % parmi des donneurs de sang, des femmes enceintes, des employés d'abattoir ou encore des patients atteints de diarrhées chroniques (**van Gool et al., 1997 ; Kucerova-Pospisilova et Ditrich, 1998 ; Enriquez et al., 1998 ; Del Aguila et al., 2001 ; Kucerova-Pospisilova et al., 2001**). Ces données suggèrent que les microsporidioses sont communes dans la population humaine bien qu'elles soient la plupart du temps asymptomatiques. Un diagnostic basé sur des approches microscopiques et/ou moléculaires (PCR) donnent des valeurs voisines, comprises entre 3,3 et 17,4%, pour une population d'enfants, de personnes âgées et de voyageurs en zones tropicales chez qui les symptômes finissent par disparaître au bout de quelques jours (**Valperga et al., 1999 ; Muller et al., 2001 ; Lores et al., 2002 ; Wichro et al., 2005**). La prévalence des microsporidioses au sein de populations composées d'individus atteints du VIH varie, quant à elle, de 5 à 50% en fonction de la localisation géographique et des techniques de détection (**Weber et al., 2000 ; Deplazes et al., 2000**).

Aucune donnée n'existe à l'heure actuelle sur les possibilités d'une transmission à partir de porteurs sains, ni concernant la probable réactivation de formes latentes au cours d'une défaillance du système immunitaire. Cependant, des cas de microsporidioses sont fréquemment rapportés suite à des greffes d'organes sans que la source de contamination puisse être identifiée avec précision. L'hypothèse d'une transmission à partir du donneur, ou d'une réactivation du développement parasitaire chez le receveur au cours du traitement immunosuppresseur demande à être confirmée (**Barsoum, 2006**).

Etant donné le caractère ubiquitaire de ces parasites et la mise en évidence de génotypes microsporidiens communs chez l'Homme et l'animal, plusieurs sources et modes de transmission ont été proposés (**Dascomb et al., 2000 ; Didier et al., 2004 ; Haro et al., 2005**). Chez l'Homme, une transmission horizontale semble privilégiée, aucune transmission verticale n'ayant été caractérisée si ce n'est chez d'autres mammifères (**Snowden et Shadduck, 1999**). La contamination de l'environnement se fait par libération des spores *via* les fèces, les produits d'excrétions ou la mort de l'organisme hôte. L'ingestion d'eau ou d'aliments contaminés donne alors lieu à une contamination, la résistance des spores assurant une survie à très long terme dans le milieu extérieur (**Didier et al., 2004 ; Mathis et al., 2005**). Des études ont révélé la présence de spores d'*E. bienewisi* dans différents échantillons d'eau montrant qu'il existe des réservoirs environnementaux (**Dowd et al., 1998**). De même, une étude récente montre que des oiseaux aquatiques (cygnes, canards sauvages, etc.) constituent des réservoirs d'espèces microsporidiennes (*E. hellem*, *E. intestinalis*) capables d'infester l'Homme (**Slodkiewicz-Kowalska et al., 2006**). Une transmission directe ou indirecte à partir d'animaux contaminés, d'objets ou d'aliments souillés est également possible (**Mathis et al., 2005**). Enfin, plusieurs cas de microsporidioses chez l'Homme sont dus à l'espèce *B. algerae* (parasite du moustique), laissant supposer une possible transmission par des vecteurs (**Cali et al., 2004 ; Coyle et al., 2004 ; Didier et al., 2004**). La caractérisation d'un même génotype, à la fois chez l'Homme et le moustique, vient également conforter l'hypothèse d'une transmission possible par les moustiques (**Visvesvara et al., 2005**).

I.1.4. Diagnostic

Une fois le diagnostic posé et la suspicion d'une microsporidiose évoquée, différentes approches peuvent être employées afin de caractériser avec plus ou moins de spécificité (selon

la technique utilisée) l'espèce incriminée (**Didier et al., 2006**). Outre leur simplicité et leur rapidité d'application, les colorations de matériels biologiques donnent des résultats variables en fonction des colorants utilisés (Trichrome modifié, Calcofluor White M2R, Uvitex 2B) et ne permettent pas d'identifier précisément l'espèce microsporidienne (**Garcia, 2002**). Grâce à l'obtention d'anticorps spécifiques, l'immunofluorescence directe, beaucoup plus sensible, peut être appliquée à l'étude d'échantillons cliniques et permettre une identification précise de l'espèce (**Sheoran et al., 2005**). Une autre technique, tout aussi sensible et spécifique, est l'immunofluorescence indirecte qui permet de révéler la présence d'anticorps dirigés contre le parasite dans le sérum du patient (**van Gool et al., 2004**). Enfin, des approches moléculaires basées sur l'amplification du gène codant l'ARNr 16S ou la région séparant les ARNr 16S et 23S (ITS1) assurent une discrimination suffisante entre différents isolats appartenant à la même espèce, et permettent une détection de plusieurs espèces au sein d'un même échantillon (**Peuvel et al., 2000 ; Wang et al., 2005**).

I.1.5. Thérapie

La recherche de substances anti-microsporidiennes se heurte à de nombreuses difficultés en raison de la localisation intracellulaire des parasites, de la résistance de la paroi sporale épaisse, et de la faible connaissance de leur métabolisme. Actuellement, deux molécules sont principalement utilisées en traitement : l'albendazole, un inhibiteur de la polymérisation des microtubules et donc de la division cellulaire, et la fumagilline, un inhibiteur de l'aminopeptidase de type II, mais dont la cytotoxicité limite son utilisation systémique (**Tremoulet et al., 2004 ; Zhang et al., 2005**).

I. 2. Phylogénie

En analysant le degré de divergence de certains marqueurs génétiques, il est possible de dresser un arbre phylogénétique faisant apparaître l'émergence des espèces les unes par

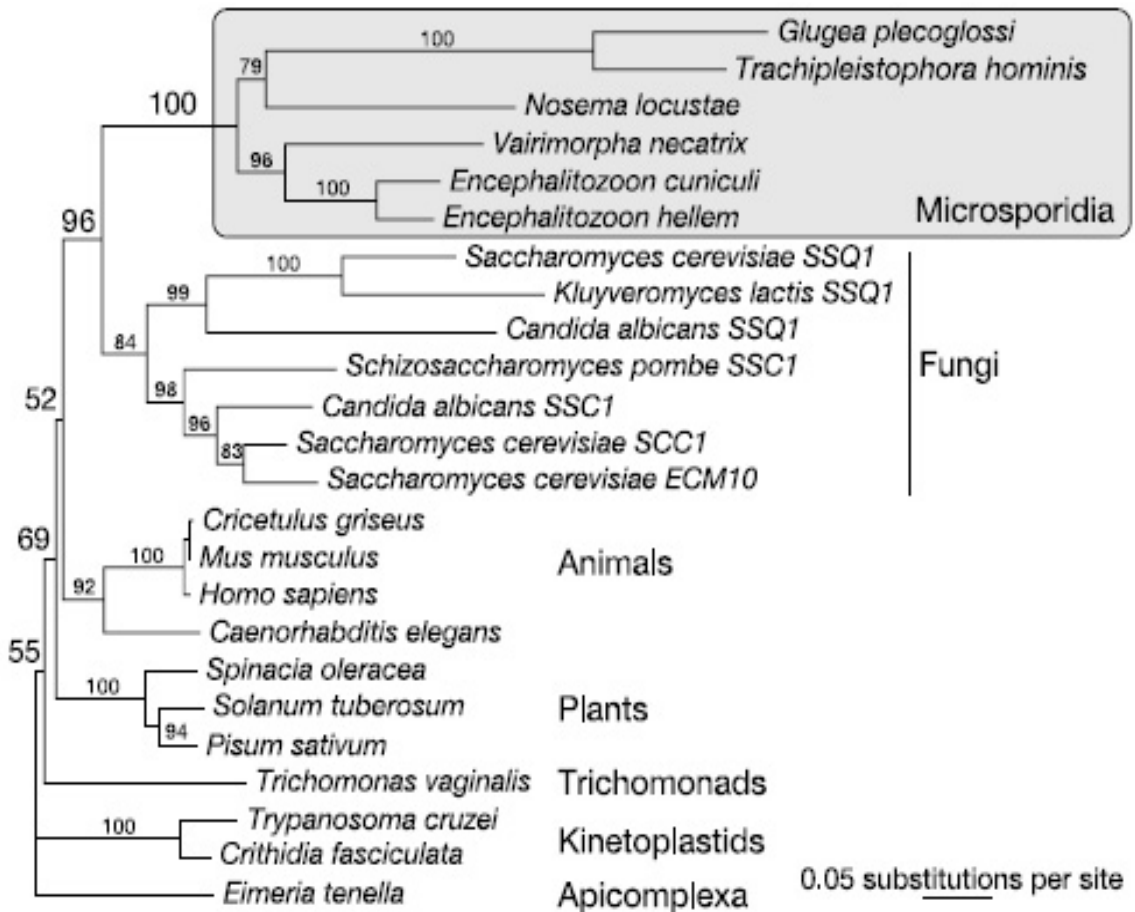


Figure 1. Phylogénie basée sur la séquence du gène codant pour une protéine HSP70 mitochondriale (d'après Williams *et al.*, 2002). Ces données montrent une émergence tardive des microsporidies après l'endosymbiose mitochondriale et suggère que ces micro-organismes sont apparentés aux champignons.

rapport aux autres au cours de l'évolution (**James et al., 2006**). La détermination de la position des espèces au sein de l'arbre est multifactorielle : elle est entre autre, fonction du (des) marqueur(s) génétique(s) choisi(s), de leur vitesse d'évolution, mais également du choix des algorithmes bioinformatiques utilisés. Si la position des microsporidies dans l'arbre du vivant a été l'objet de nombreuses controverses, c'est probablement en raison des particularités propres à ce Phylum.

Au même titre que les espèces appartenant aux genres *Entamoeba*, *Giardia*, ou *Trichomonas*, les microsporidies ont tout d'abord été considérées comme des organismes primitifs (*Archezoa*) ayant émergé avant l'endosymbiose (**Cavalier-Smith, 1987**). Corrélée à l'absence de structures eucaryotes classiques (constante de sédimentation ribosomale de 70S, ARNr 16 et 23S, appareil de Golgi atypique...), cette classification est apparue parfaitement légitime, d'autant plus qu'elle a été confortée par les résultats des premières analyses phylogénétiques portant sur l'ARNr 16S, l'aminoacyl-ARNt synthétase ou encore les facteurs d'élongation EF1 α et EF2 (**Curgi et al., 1980 ; Vossbrinck et al., 1987 ; Patterson, 1994 ; Brown et Doolittle, 1995 ; Kamaishi et al., 1996a et b**). Cependant, sur la base de nouvelles analyses phylogénétiques portant sur les tubulines α et β , des données contradictoires émergent et révèlent une proximité phylogénétique entre les microsporidies et les champignons (**Keeling et Doolittle, 1996 ; Edlind et al., 1996**). Ces derniers résultats ont été confirmés par des analyses phylogénétiques complémentaires (**Fig. 1**) (**Fast et al., 1999 ; Hirt et al., 1999 ; Williams et al., 2002 ; Thomarat et al., 2004 ; Gill et Fast, 2006**).

Les résultats de ces études sont en accord avec des caractères physiologiques retrouvés chez les champignons : présence de tréhalose et de chitine dans la spore, appariement des noyaux en diplocaryons chez certains genres microsporidiens, mitoses fermées (pleuromitose) avec présence de plaques polaires (**Vivarès et al., 1977 ; Dolgikh et Semenov, 2003**). La caractérisation de gènes nucléaires codant des protéines mitochondriales (ex : HSP70, sous

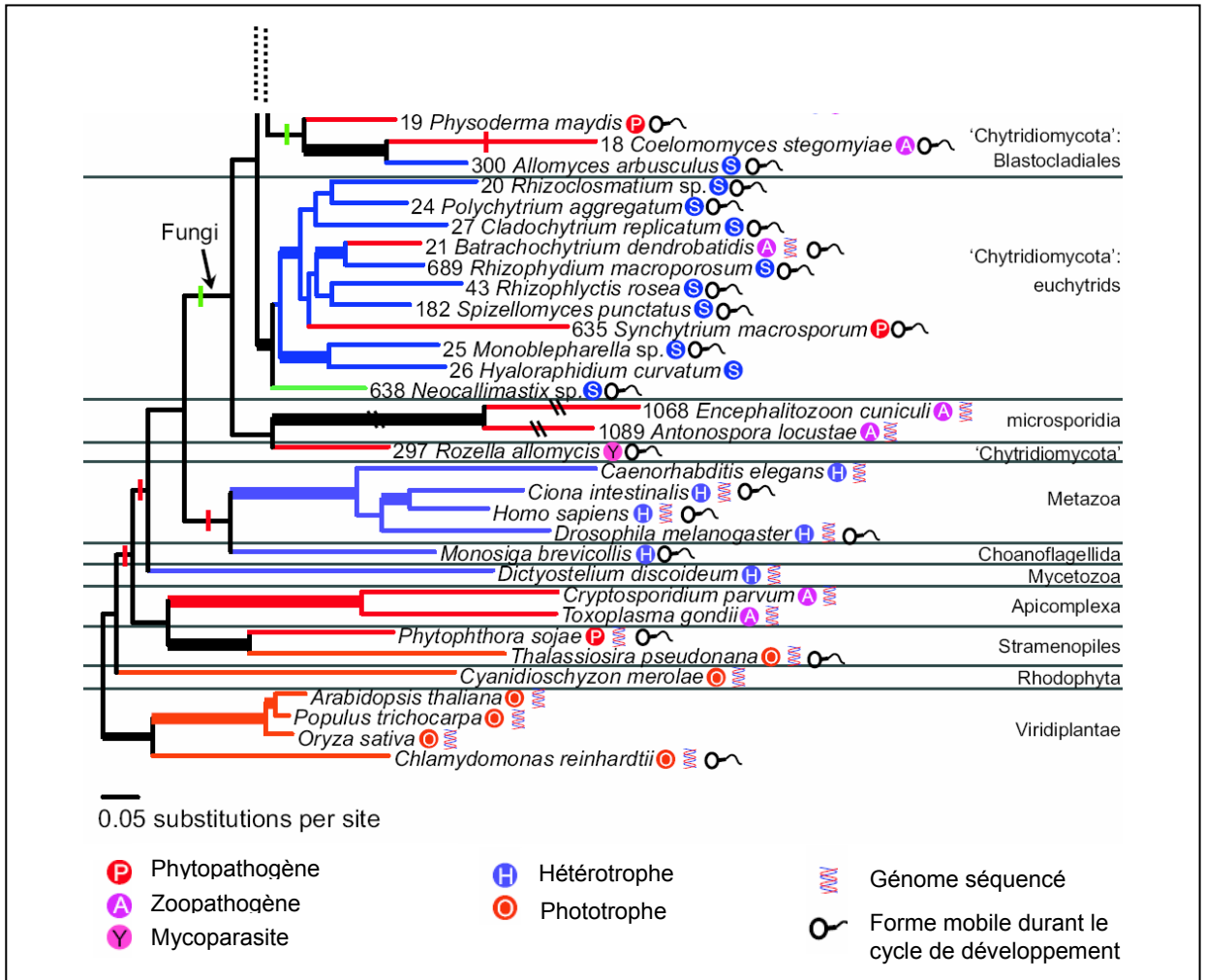


Figure 2. Arbre phylogénétique basé sur l'analyse de six gènes (d'après James *et al.*, 2006). L'analyse a été réalisée sur 199 groupes taxonomiques et porte sur les gènes codant l'ARNr 5,8 S, 18S et 28S, le facteur d'élongation *EF1 α* , et les deux sous-unités de l'ARN polymérase II (*RPB1* et *RPB2*).

unités α et β de la pyruvate deshydrogénase E1, ferredoxine) confirme une perte secondaire des mitochondries après l'endosymbiose (**Peyretailade et al., 1998a ; Fast et al., 2001 ; Katinka et al., 2001**).

Comment expliquer alors que les premières analyses phylogénétiques aient pu donner des résultats erronés? Il apparaît que les gènes codant les ARNr 16 et 23S microsporidiens ont en réalité évolué beaucoup plus rapidement que chez les autres espèces, ce qui a eu pour effet de placer ce phylum à proximité du groupe externe indépendamment des relations phylogénétiques réelles, c'est-à-dire à la base de l'arbre phylogénétique. Ce phénomène est communément appelé "phénomène d'attraction artéfactuelle des longues branches" (**van de Peer et al., 2000**).

Alors que l'ensemble des analyses phylogénétiques place le phylum des microsporidies à proximité des Ascomycètes, Basidiomycètes ou encore des Zygomycètes, voire même le considère comme un phylum distinct des Champignons (**Keeling et al., 2003 ; Gill and Fast, 2006 ; Liu et al., 2006**), une étude phylogénétique s'appuyant sur une nouvelle hypothèse remet en question les conclusions jusque là avancées (**Fig. 2**) (**James et al., 2006**). S'appuyant sur l'analyse de 6 gènes conservés au sein du phylum des Champignons, cette étude phylogénétique laisse apparaître une proximité très forte des microsporidies avec un chytride endopathogène, *Rozella allomycis*. L'ancêtre des champignons, probablement très proche des chytrides actuels (champignons dont l'un des stade de développement est unicellulaire flagellé), aurait au cours de l'évolution privilégié un mode de nutrition statique, entraînant la perte du flagelle et l'apparition séquentielle des différents groupes fongiques terrestres. Cette perte coïnciderait avec l'apparition de nouvelles formes de dissémination des spores par voie aérienne. Chez les microsporidies, le flagelle aurait évolué pour former une structure d'invasion unique, le tube polaire.

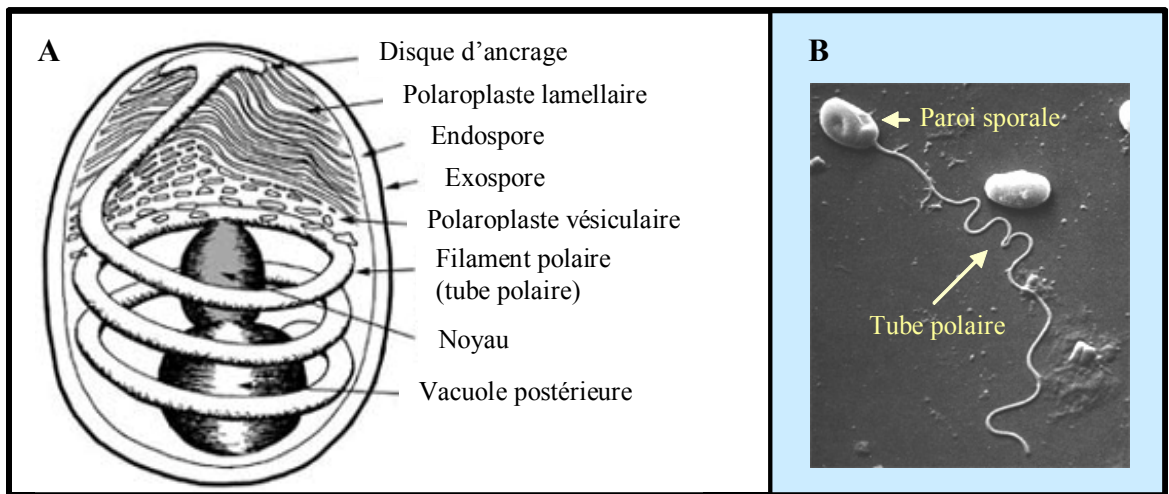


Figure 3. Spores des microsporidies. (A) Représentation schématique de l'ultrastructure d'une spore de microsporidie. (B) Spore d'*Encephalitozoon cuniculi* dévaginée et observée au microscope à épifluorescence. Après extrusion, le tube polaire peut atteindre 100 fois la longueur de la spore (50 à 100 μm).

I.3. Organisation structurale de la spore

D'une taille variant de 1 à 20 μm et délimitée par une paroi épaisse, la spore renferme le sporoplasme ainsi que l'appareil invasif (**Fig. 3**). Elle représente la forme de résistance et le stade de dissémination. Bien que les microsporidies présentent des spécificités d'hôtes très différentes, elles se caractérisent par un même mécanisme d'invasion. Ce processus, unique dans le monde du vivant, implique l'extrusion brutale du tube polaire qui reste attaché au pôle antérieur de la spore et peut aller transpercer la membrane de la cellule hôte pour délivrer le sporoplasme infectieux dans le cytoplasme de la cellule hôte (**Franzen, 2005**).

I.3.1. L'élément infectieux : le sporoplasme

Le sporoplasme désigne l'élément infectieux qui est transféré à l'intérieur de la cellule hôte *via* le tube polaire. Il est composé de l'appareil nucléaire et du cytoplasme renfermant de nombreux ribosomes ainsi que du réticulum endoplasmique (RE). Selon l'espèce, deux types de conformations nucléaires existent : monocaryotique ou diplocaryotique. Dans ce dernier cas, les deux noyaux accolés se divisent de façon synchrone pour former un diplocaryon (**Vavra, 1976**). La ploïdie de ces noyaux reste incertaine.

Le sporoplasme se caractérise par une très grande richesse en ribosomes. La mise en évidence d'activités enzymatiques classiquement associées aux compartiments *cis* et *trans*-golgien au niveau de structures vésiculaires et tubulaires (**Takvorian et Cali, 1994 et 1996 ; Sokolova et al., 2001**) confirme l'existence d'un appareil de Golgi atypique sans saccules (**Vavra, 1965**).

I.3.2. L'appareil invasif

Les microsporidies se caractérisent par un appareil d'infestation original constitué de quatre éléments structuraux : le disque d'ancrage, le tube polaire, le polaroplaste, et la vacuole postérieure (**Fig. 3**).

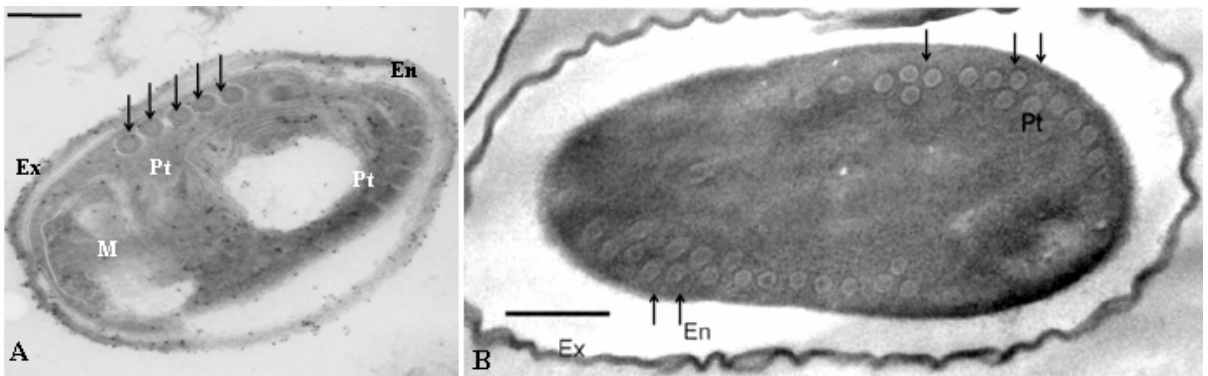


Figure 4. Le tube polaire : appareil invasif des microsporidies. (A) Chez *Encephalitozoon hellem*, le tube polaire est enroulé en 5 spires organisées en une seule couche (flèches). Echelle : 0,2 μm **(B)** Coupe d'une spore d'*Antonospora locustae* montrant les spires du tube polaire organisées en deux couches (flèches). Echelle : 0,5 μm . (d'après Slamovits *et al.*, 2004). Pt : tube polaire; M : manubrium; Ex : exospore; En : endospore.

Le **disque d'ancrage** (ou capuchon polaire) se situe au pôle antérieur de la spore, au niveau d'une région où l'on observe un amincissement de l'endospore correspondant à la zone de rupture pariétale au cours du processus d'extrusion. Il est relié au tube polaire.

Le **tube polaire (ou filament polaire)** se présente sous la forme d'une structure tubulaire enroulée à l'intérieur du cytoplasme. Le nombre de tours de spires, qui est utilisé comme un critère d'identification, varie de 4 à 30. Délimité par une membrane, le tube polaire est constitué de 6 couches concentriques. Seules 3 couches sont observées après extrusion (**Canning et al., 1992 ; Chioralia et al., 1998**). Chez *Encephalitozoon*, 5 à 7 tours de spires sont observés (**Fig. 4**). Chez *Antonospora locustae* et *Paranosema grylli*, le filament polaire, organisé en une ou deux rangées, présente respectivement 15 à 18, et 18 à 20 tours de spires (**Slamovits et al., 2004 ; Sokolova et al., 2003**). Cette structure, enroulée à l'intérieur de la spore, est délimitée par une membrane et semble être constituée de plusieurs couches concentriques denses ou claires aux électrons organisées autour d'un corps central (**Canning et al., 1992 ; Chioralia et al., 1998**).

Le **polaroplaste** est un système membranaire situé dans la partie antérieure de la spore. Très volumineux, il peut occuper jusqu'aux 2/3 du volume sporal. Il est constitué d'une région lamellaire formée par l'empilement régulier de saccules très aplatis, et d'une région vésiculaire. Lors de la germination (extrusion du tube polaire) et du passage du sporoplasme dans le tube polaire, le polaroplaste lamellaire serait transféré pour être à l'origine de la membrane délimitant le sporoplasme libéré (**Weidner et al., 1984**).

Enfin, la **vacuole postérieure** est une structure de taille variable, claire aux électrons, située au pôle basal de la spore. Son gonflement interviendrait dans les processus d'extrusion du tube polaire et de transfert du sporoplasme. Cette vacuole est absente chez le genre *Nosema*.

I.3.3. L'enveloppe sporale

Epaisse et rigide, la paroi protège le sporoplasme contre les stress environnementaux permettant notamment la survie des parasites après leur libération des cellules hôtes. La rigidité de la paroi permet également de maintenir une pression hydrostatique élevée et sa perméabilité aqueuse joue un rôle dans l'extrusion du tube polaire (**Frixione *et al.*, 1997**). Certains constituants (glycoprotéines notamment) localisés à la surface de la paroi sporale pourraient aussi être impliqués dans le processus d'initiation de l'invasion, en interagissant avec des récepteurs cellulaires (**Bohne *et al.*, 2000**).

De nature protéo-chitineuse, la paroi est constituée de deux couches recouvrant une membrane plasmique : la couche la plus externe est appelée exospore, la plus interne étant l'endospore. D'épaisseur variable (15 à 100 nm), l'exospore est une région dense aux électrons constituée principalement de glycoprotéines (**Vavra, 1976 ; Vivarès *et al.*, 1976**).

Plus épaisse, l'endospore est une couche claire aux électrons sans organisation apparente (**Bigliardi et Sacchi, 2001**). Elle est constituée de protéines et de polysaccharides, la chitine étant l'un des composants majeurs (**Vavra, 1976 ; van Gool *et al.*, 1993**). Solidaire de la membrane plasmique qui délimite le sporoplasme, elle recouvre la spore avec une épaisseur constante de 100 nm. Elle présente cependant une épaisseur réduite au pôle antérieur de la spore au niveau duquel aura lieu l'extrusion du tube polaire (**Chioralia *et al.*, 1998**).

Tableau 2. Estimation de la taille du génome haploïde et du nombre de chromosomes chez quelques microsporidies (évaluation d'après les caryotypes moléculaires obtenus par électrophorèse en champs pulsés). Nbre : nombre ; chrs : chromosome.

Espèces	Taille (Mpb)	Nbre de bandes	Nbre de chrs	Tailles des bandes (kb)	Auteurs
<i>Glugea atherinae</i>	19,5	16	-	420-2700	Bidierre <i>et al.</i> , 1994
<i>G. stephani</i>	16,8	15	-	340-2654	Amigo <i>et al.</i> , 2002
<i>Nosema bombycis</i>	15,3	18	-	380-1500	Kawakami <i>et al.</i> , 1994
<i>N. pyrausta</i>	10,5	13	-	440-1390	Munderloh <i>et al.</i> , 1990
<i>N. furnacalis</i>	10,2	13	-	440-1360	Munderloh <i>et al.</i> , 1990
<i>Vairimorpha sp</i>	10,2	8	-	720-1790	Malone et Mc Ivor, 1993
<i>Vavraia oncoperae</i>					
chenille	10,2	16	-	140-1830	Malone et McIvor, 1993
ver	8,0	14	-	130-1930	
<i>N. costelytrae</i>	7,4	8	-	290-1810	Malone et McIvor, 1993
<i>Spraguea lophii</i>					Amigo <i>et al.</i> , 2002
chez <i>L. piscatorius</i>	7,3	10	15	266-1076	
chez <i>L. budegassa</i>	7,3	11	15	271-1120	
<i>S. lophii</i>	6,7	13	15	202-878	Mansour <i>et al.</i> , 2004
<i>S. lophii</i>	6,2	12	-	230-980	Bidierre <i>et al.</i> , 1994
<i>N. locustae</i>	5,3	18	-	139-651	Streett, 1994
<i>Encephalitozoon cuniculi</i>	2,9	11	11	217-315	Bidierre <i>et al.</i> , 1995
<i>E. hellem</i>	2,6	7-13	-	205-272	Bidierre <i>et al.</i> , 1999
					Delarbre <i>et al.</i> , 2001
<i>E. intestinalis</i>	2,3	10	-		Bidierre <i>et al.</i> , 1999

II/ GENOMIQUE STRUCTURALE DES MICROSPORIDIÉS

II.1. Structure des génomes microsporidiens

II.1.1. Caryotype moléculaire

Les premières informations sur la structure des génomes de microsporidies datent des années 1990. L'étude d'espèces parasites d'insectes a révélé des génomes haploïdes particulièrement petits, compris entre 7 et 10 Mpb, laissant supposer une relative simplicité d'organisation génomique (**Munderloh et al., 1990 ; Malone et McIvor, 1993**). Ce phylum regroupe quelques 1200 espèces dont la taille du génome se caractérise par une très grande hétérogénéité. En effet, des variations de taille considérables sont observées, de 2,3 Mpb pour *E. intestinalis* à 19,5 Mpb pour *Glugea atherinae* (**Tableau 2**), (**Biderre et al., 1994 et 1999**). De la même manière, selon les espèces, le nombre de bandes d'ADN chromosomique peut varier de 8 à 18 bandes respectivement pour les espèces du genre *Vairimorpha sp.* et *G. atherinae*. A noter également la très grande variabilité de taille des chromosomes, de 130 kb chez *V. oncoperae* à 2700 kb chez *G. atherinae* (**Biderre et al., 1994**).

Le genre *Encephalitozoon* regroupe des espèces parasites de mammifères (*E. cuniculi*, *E. hellem* et *E. intestinalis*) dont les génomes sont à l'heure actuelle montrés comme étant les plus réduits pour des organismes eucaryotes. Chez ces trois espèces, le génome se compose de 10 à 13 bandes d'ADN chromosomique dont les tailles varient de 175 à 315 kb (**Peyretailade et al., 1998b ; Vivarès et Méténier, 2000**). Chez *E. cuniculi* ou encore *E. hellem*, des variations intraspécifiques (nombre de bandes d'ADN chromosomique, taille des chromosomes, taille globale du génome) ont également été rapportées (**Biderre et al., 1997 ; Sobottka et al., 1999 ; Delabre et al., 2001**). Ces données suggèrent des mécanismes de remaniements chromosomiques et une diploïdie probable des génomes du genre *Encephalitozoon*. Ces polymorphismes caryotypiques ne sont cependant pas reliés à une spécificité d'hôte.

Tableau 3. Estimation de la taille du génome et du nombre de gènes de plusieurs organismes procaryotes et eucaryotes.

Génome	Taxonomie	Taille (Mpb)	Nombre putatif de gènes
<i>Bigelowiella natans</i>	nucléomorphe	0,37	331
<i>Rickettsia prowazekii</i>	bactérie	1,1	834
<i>Encephalitozoon cuniculi</i>	microsporidie	2,9	1997
<i>Escherichia coli</i>	bactérie	4,6	4377
<i>Giardia lamblia</i>	diplomonadine	12	~ 5000
<i>Saccharomyces cerevisiæ</i>	levure	12	5885
<i>Plasmodium falciparum</i>	apicomplexa	23	6500
<i>Leishmania major</i>	kinétoplastidé	28	8600
<i>Caenorhabditis elegans</i>	animal	97	17085
<i>Arabidopsis thaliana</i>	plante	117	25498
<i>Drosophila melanogaster</i>	animal	180	14000
<i>Homo sapiens</i>	animal	3286	30000 à 40000

L'analyse des caryotypes moléculaires n'est pas suffisamment discriminante pour définir avec exactitude le nombre de chromosomes. En effet, des chromosomes hétérologues de même taille peuvent co-migrer dans le gel, alors que des chromosomes homologues de taille différente peuvent être différenciés. Pour pallier à ces problèmes, des approches de cartographie physique et d'électrophorèse en champs pulsés en deux dimensions ont été mises en oeuvre pour l'espèce *E. cuniculi*. La première a permis de déterminer des groupes de liaison spécifiques de chaque bande d'ADN chromosomique, mettant en évidence 11 chromosomes hétérologues chez cette espèce (**Biderre *et al.*, 1997**). La seconde a montré la présence de deux chromosomes homologues III présentant une différence de taille de 3 kb ce qui conforte l'idée d'une diploïdie (**Brugère *et al.*, 2000a**).

II.1.2. Organisation génique

Au cours de l'évolution, et probablement en raison de leur statut de parasites intracellulaires obligatoires, les endoparasites voient la taille de leur génome se réduire (**Tableau 3**), (**Khachane *et al.*, 2007**). Si les causes de cette réduction ou les contraintes qui en sont à l'origine ne peuvent être identifiées précisément, cette réduction touche toutes les séquences du génome, qu'il s'agisse de régions codantes (domaines protéiques) ou non codantes (introns, régions intergéniques), allant parfois jusqu'à l'élimination de gènes voire de voies métaboliques complètes (**Keeling et Slamovits, 2005**).

II.1.2.1. Les séquences codantes

Le programme de séquençage du génome complet d'*E. cuniculi* a permis d'apporter de nombreuses réponses sur la compaction du génome de cette espèce (**Katinka *et al.*, 2001**). Le premier point concerne le nombre de séquences codantes (CDS) : seulement 1997 ont été prédites, ce qui représente le plus petit nombre de gènes connu chez un eucaryote, exception faite du génome du nucléomorphe de la chlorarachniophyte *Bigelowiella natans* qui n'en

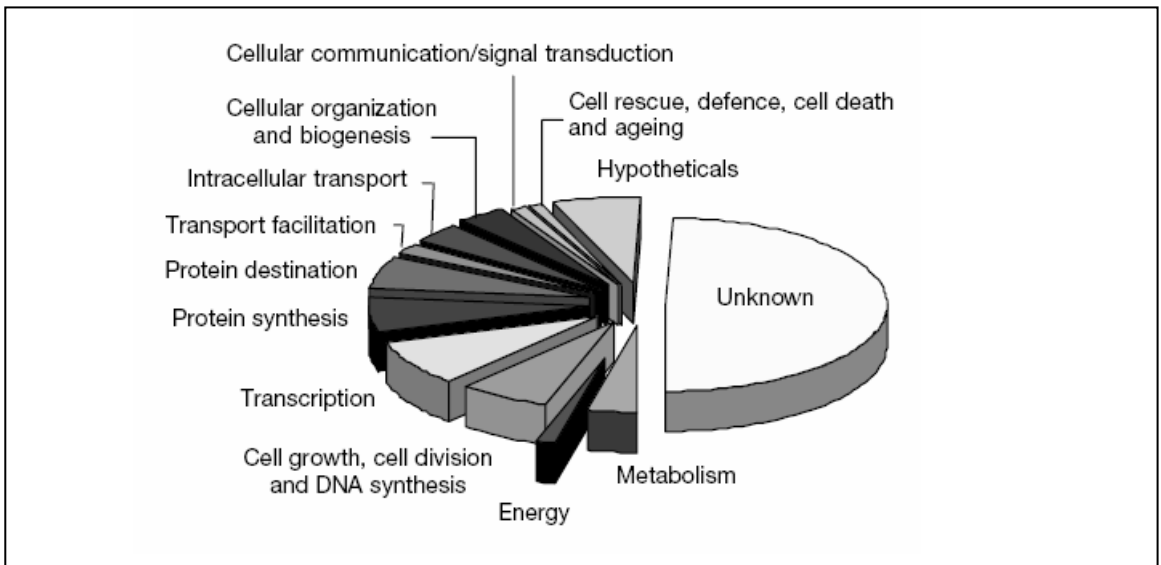


Figure 5. Distribution en classes fonctionnelles des protéines prédites chez *E. cuniculi* (d'après Katinka et al., 2001).

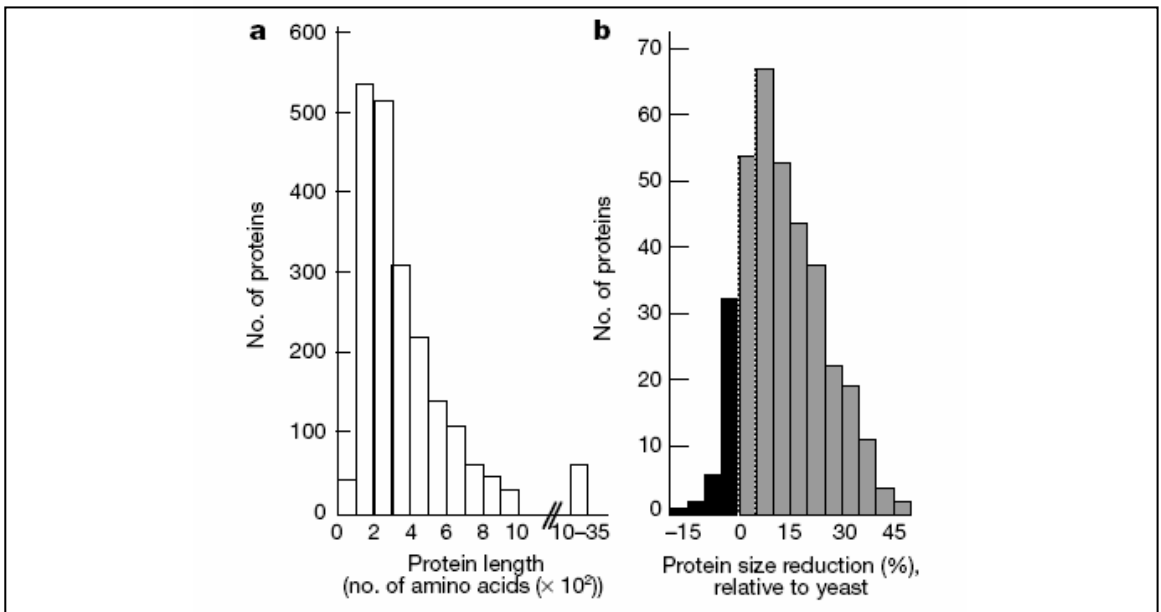


Figure 6. Comparaison de la taille des protéines chez *E. cuniculi* par rapport à leurs homologues chez *S. cerevisiae* (d'après Katinka et al., 2001). **a.** Distribution des protéines d'*E. cuniculi* en fonction de leur taille (seulement 6 protéines ont plus de 2000 acides aminés). **b.** Degré de réduction de la taille des protéines d'*E. cuniculi* par rapport à leur homologue chez *S. cerevisiae* (n=350). En gris, protéines pour lesquelles une réduction est observée. En noir, protéines dont la taille apparaît plus importante chez *E. cuniculi*.

compte que 331 (**Tableau 3**), (**Gilson et al., 2006**). Chez la bactérie *Escherichia coli* (souche K12), 4288 séquences codantes ont été dénombrées (**Blattner et al., 1997**), soit deux fois plus que chez la microsporidie *E. cuniculi*.

Chez *E. cuniculi*, 44% des gènes identifiés codent des protéines de fonctions connues, 6% pour des protéines hypothétiques conservées, et 50% des protéines sans fonction prédite (**Fig. 5**). La densité des gènes est très élevée avec presque un gène par kilobase. C'est la plus forte densité observée pour l'ensemble des génomes nucléaires eucaryotes séquencés en totalité, avec celle du nucléomorphe de *Guillardia theta* (1 gène / 977 pb). Une densité comparable, estimée à 0,97 gène / kb contre 0,94 chez *E. cuniculi* a également été décrite chez une autre espèce de microsporidie, *Antonospora locustae*, pour laquelle 685 kb ont été séquencés (**Keeling et Slamovits, 2004**).

Pour comparaison, la densité observée chez *Saccharomyces cerevisiae* est de 1 CDS / 2 kb, chez *Leishmania major* 1 CDS / 3,26 kb et chez *Trypanosoma cruzi* 1 CDS / 4,5 kb. La densité génique trouvée dans les génomes de certaines espèces de microsporidies et du nucléomorphe pourrait donc représenter la densité maximale possible pour un génome nucléaire.

Il faut également noter que chez *E. cuniculi*, plus de 85% des protéines prédites présentent une taille 15% inférieure à celle de leur orthologue chez la levure (**Fig. 6**). La perte de certaines protéines et de certaines voies métaboliques aurait entraîné une diminution des interactions fonctionnelles entre protéines d'où la disparition des domaines impliqués dans ces interactions pour les protéines maintenues (**Vivarès et al., 2002**).

L'examen des gènes assignés à des catégories fonctionnelles montre une simplification ou une disparition de certaines voies métaboliques : biosynthèse des acides aminés limitée, absence de la chaîne respiratoire et de la voie de biosynthèse des acides gras. **Katinka et al. (2001)** suggèrent que la forte dépendance des parasites intracellulaires pour les métabolites de

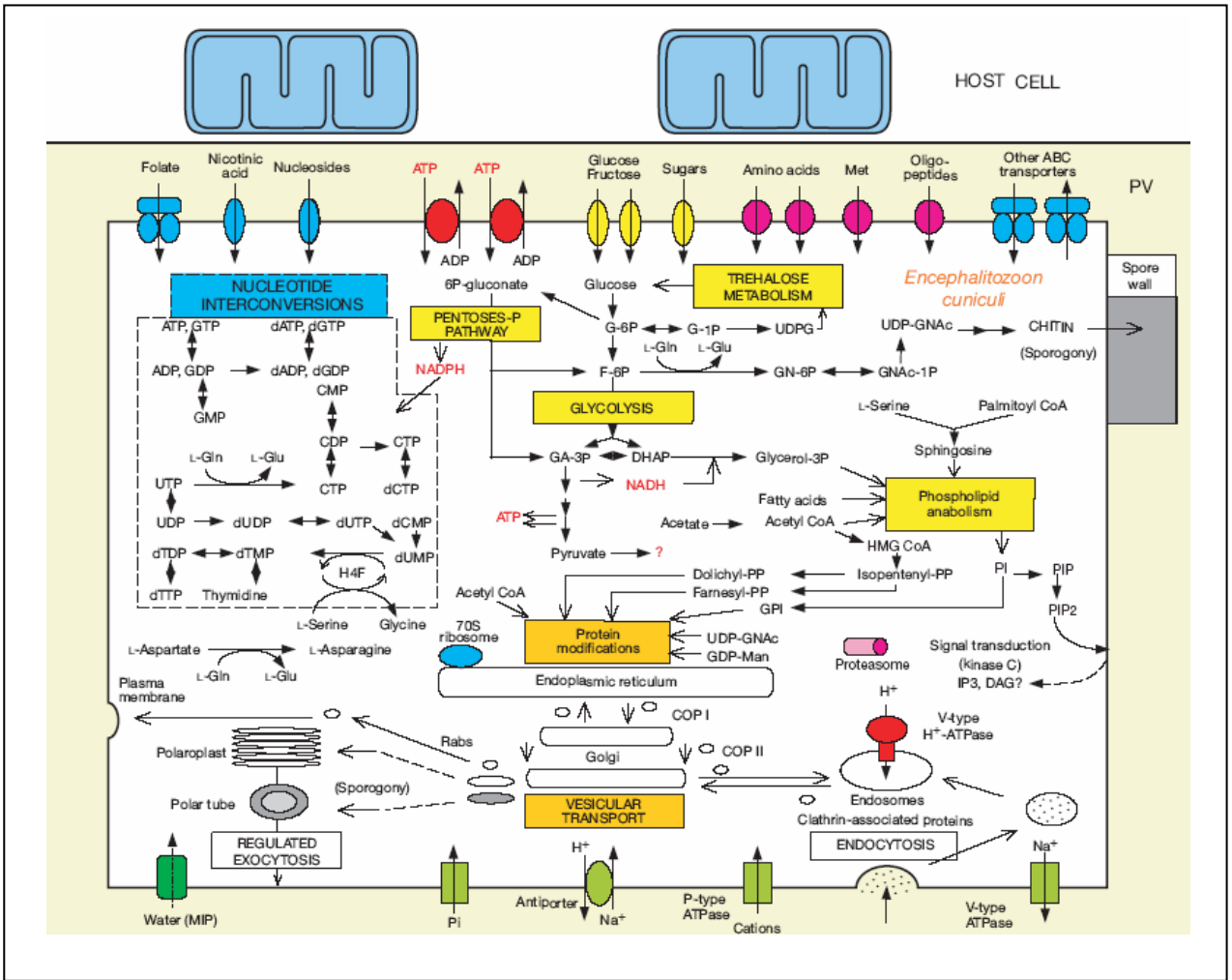


Figure 7. Reconstitution du métabolisme et des voies de transport d'*E. cuniculi* déduits de l'analyse du génome. Les voies de biosynthèse des nucléotides et de production d'énergie, sont indiquées. Le tréhalose semble être la source majeure de glucose. Le métabolisme de la chitine nécessaire à la différenciation de la paroi sporale au cours de la sporogénèse est représenté. Les transporteurs potentiels associés à la membrane plasmique sont indiqués ainsi que leurs substrats. Les voies d'endocytose et de transport vésiculaire impliquant l'appareil de Golgi polarisé *cis-trans* sont indiquées. Les points d'interrogation signalent les incertitudes majeures concernant le pyruvate et la nature des messagers secondaires des voies de transduction de signaux. Le parasite est représenté au sein d'une vacuole parasitophore (PV) dans le cytoplasme de la cellule hôte avec un recrutement de mitochondries près de la membrane de la vacuole parasitophore (d'après Katinka *et al.*, 2001).

l'hôte s'explique par l'absence de gènes impliqués dans de nombreuses voies métaboliques comme par exemple le cycle des acides tricarboxyliques (**Fig.7**), (ou cycle de Krebs). L'absence de la voie de dégradation des acides gras et du complexe ATP synthétase F0-F1 illustre fortement la dépendance énergétique de ce parasite à l'égard de son hôte. Au contraire, les voies concernant une variété d'autres processus essentiels (réplication de l'ADN, ribosomes, splicéosome ou encore la glycolyse) sont partiellement ou totalement conservées. L'élimination de gènes conduit à un génome codant un nombre réduit de protéines. Le protéome simplifié d'*E. cuniculi* pourrait donc constituer un bon outil pour la caractérisation des réseaux d'interactions minimales (**Brosson et al., 2006**).

II.1.2.2. Les séquences intergéniques

La majorité des gènes d'*E. cuniculi* sont de petite taille par rapport à leurs homologues chez la levure (**Katinka et al., 2001**). Cette réduction ne permet pas d'expliquer à elle seule la faible taille observée pour ce génome. En effet, si tous les gènes réduits étaient de taille identique à celle de leurs homologues (levure), la taille de ce génome n'augmenterait que de 0,25 Mpb (**Vivares et al., 2002**). Cette réduction est donc également liée à la forte compaction des gènes et donc à une réduction importante de la taille des régions intergéniques. En effet la taille moyenne de ces régions n'est que de 129 pb (**Katinka et al., 2001**) et la région intergénique la plus grande observée sur le chromosome I ne fait que de 639 pb (**Peyret et al., 2001**).

Une densité génique comparable a également été décrite sur certains contigs chez *A. locustae* (**Keeling et Slamovits, 2004**). Ces mêmes auteurs ont néanmoins caractérisé des régions intergéniques pouvant atteindre 3,8 kb. Cette particularité explique que malgré une taille de génome estimée à 5,4 Mpb (deux fois la taille du génome d'*E. cuniculi*) seulement 2600 CDS aient pu être identifiés jusqu'à présent.

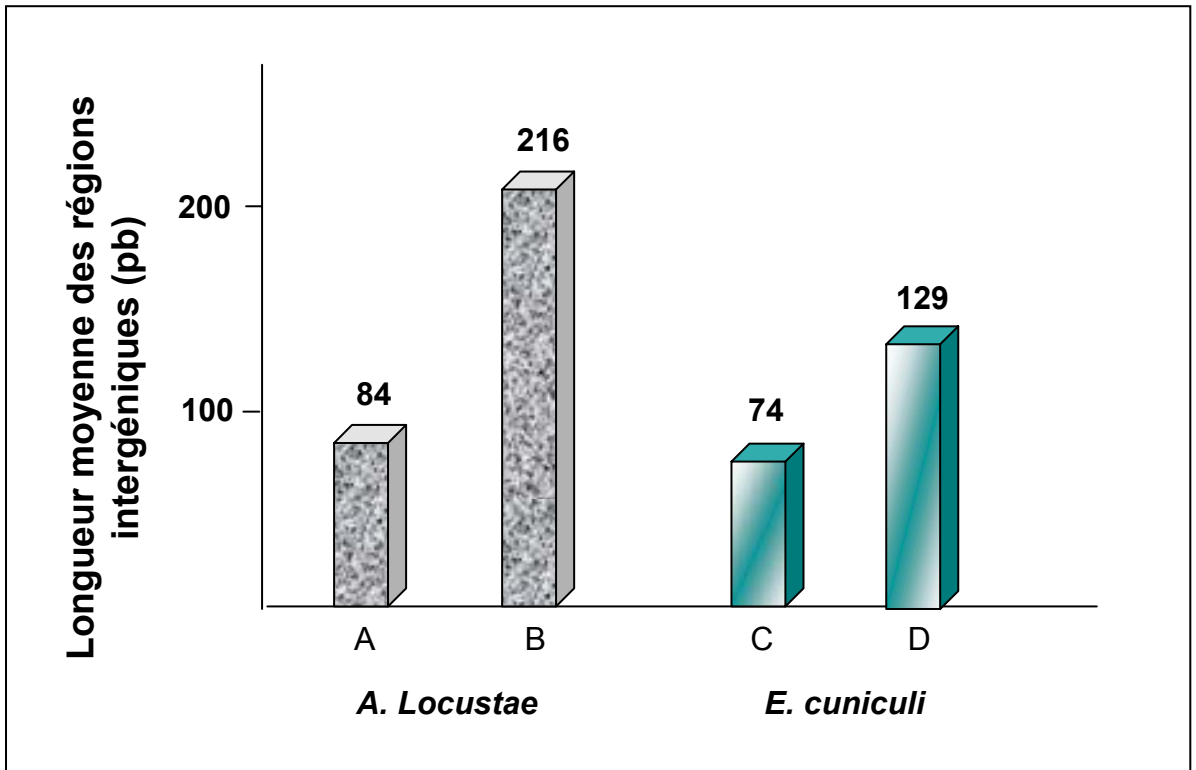


Figure 8. Représentation graphique des moyennes de taille des régions intergéniques chez *E. cuniculi* et *A. locustae* (d'après Slamovits et al., 2004). **A** et **C**: taille moyenne des espaces intergéniques des gènes conservés entre *A. locustae* et *E. cuniculi*. **B**: taille moyenne des espaces intergéniques des gènes d'*A. locustae* dont l'ordre n'est pas conservé chez *E. cuniculi*. **D**: taille moyenne des espaces intergéniques pour l'ensemble des gènes chez *E. cuniculi*.

L'étude de la conservation de la synténie entre les espèces *E. cuniculi* et *A. locustae* a révélé que 13% des gènes sont présents dans le même contexte et 30% sont séparés par un petit nombre de réarrangements (**Slamovits et al., 2004**). La taille des régions intergéniques séparant deux gènes dont l'ordre est conservé entre ces deux espèces est beaucoup plus faible que la taille moyenne des régions intergéniques (**Fig. 8**). Cette diminution importante de taille pourrait augmenter la stabilité du génome en diminuant le nombre de recombinaisons (**Slamovits et al., 2004**).

Le séquençage des génomes des nucléomorphes de *G. theta* et *B. natans* a permis de mettre en évidence des génomes dont le degré de compaction est similaire à celui de l'espèce *E. cuniculi*, la taille moyenne des régions intergéniques étant respectivement de 70 et 113 pb pour chacune des espèces (**Douglas et al., 2001 ; Gilson et McFadden, 2001**). L'analyse de banques d'ESTs de ces génomes fortement compactés (*G. theta*, *B. natans* et *A. locustae*) a révélé l'existence de nombreux transcrits dont l'extrémité 3' UTR se situe dans la région codante du gène situé en aval (**Williams et al., 2005**). D'après ces auteurs, une réduction de taille importante des régions intergéniques pourrait être à l'origine de la perte des signaux de fin de transcription. L'utilisation de sites cryptiques localisés en aval serait à l'origine de la formation dans certains cas de transcrits englobant plusieurs gènes. Chez les apicomplexes, notamment chez les espèces des genres *Cryptosporidium* sp. et *Theilaria* sp., la taille moyenne des régions intergéniques peut varier de 369 pb (*T. annulata*) à plus de 700 pb (*C. hominis*) (revue par **Keeling et Slamovits, 2005**). La réduction de taille des espaces intergéniques représente probablement l'un des éléments les plus prépondérants à l'origine de la compaction du génome chez les apicomplexes.

II.1.2.3. Les introns

Les introns de type splicéosomien ont été décrits dans les gènes de plantes, d'animaux et de plusieurs protistes unicellulaires, mais sont absents de ceux de bactéries et d'archées. Grâce au séquençage complet du génome d'*E. cuniculi*, 13 introns de ce type ont été mis en évidence parmi lesquels 10 sont présents dans des gènes codant pour des protéines ribosomiques. Une étude plus fine de l'un d'entre eux, localisé dans le gène codant la protéine L27a entrant dans la composition de la grande sous-unité ribosomique, a été réalisée par **Biderre et al. (1998)**. Cet intron de petite taille (28 pb) possède un pourcentage de résidus adénine et thymine de l'ordre de 61%, et est localisé juste après le codon d'initiation de traduction ATG. Il possède les bornes 5' GT---AG 3' caractéristiques des introns splicéosomiens, et la séquence consensus de 6 pb GTAAGT à son extrémité 5' généralement trouvée chez les Eucaryotes supérieurs. La présence de cet intron est en accord avec la découverte d'une partie des composants du complexe ribonucléoprotéique responsable de l'épissage, le splicéosome (**Katinka et al., 2001**).

Deux autres introns, présentant des tailles de 16 et 42 pb ont été mis en évidence dans 2 des 44 gènes codant pour les ARNt (ARNt isoleucine et ARNt tyrosine). Le mécanisme permettant leur épissage est différent de celui des introns splicéosomiens mais nécessite tout de même la présence d'enzymes spécialisées. Une recherche par homologie (blast) a permis d'identifier le gène codant l'endonucléase responsable de cet épissage. Une approche similaire visant à caractériser le gène codant la ligase kinase s'est révélée infructueuse probablement en raison d'une divergence importante de la séquence nucléique.

Le nombre d'introns prédit pour le génome d'*E. cuniculi* est le plus faible jamais observé pour un eucaryote. Notons néanmoins que chez l'algue rouge, *Cyanidoschyzon merolae*, des introns ne sont retrouvés que dans 0,5% des gènes contre 0,6% chez *E. cuniculi* (**Keeling et Slamovits, 2005**).

Alors que certains organismes (*E. cuniculi*, *S. cerevisiae*, etc.) ont préféré éliminer la plupart des introns pour arriver à un degré de compaction relativement important de leur génome, d'autres (*Bigelowiella Natans* ou encore *Paramecium tetraurelia*) ont opté de manière privilégiée pour une diminution de leur taille (Abrahamsen et al., 2004 ; Zagulski et al., 2004 ; Keeling et Slamovits, 2005 ; Gilson et al., 2006). Chez *P. falciparum*, malgré l'absence d'introns dans 46 % des gènes contre 30% environ chez *Theilaria* sp, son génome présente une densité génique plus faible en raison de la taille importante de ses introns (179 pb) et des régions intergéniques (1694 pb) (Gardner et al., 2002 ; Keeling et Slamovits, 2005).

II.1.3. Les séquences répétées

II.1.3.1. Les éléments transposables

Les éléments transposables (ETs) sont des séquences d'ADN répétées de quelques kb capables de se déplacer au sein du génome (Capy et al., 1998). Deux groupes d'éléments transposables peuvent être distingués en fonction de leur mode de transposition : les éléments de classe I transposent par un intermédiaire à ARN (rétrotransposon) et ceux de classe II utilisent un intermédiaire à ADN (transposon). En raison de leur capacité de déplacement au sein du génome, les éléments transposables peuvent être à l'origine de remaniements génétiques très importants (translocation, inversion, délétion). Ils sont donc considérés comme des éléments pouvant jouer un rôle capital dans l'évolution structurale des génomes (Kazazian, 2004 ; Le Rouzic et al., 2007).

La mise en évidence d'un ORF codant une reverse transcriptase chez *Spraguea lophii* suggérait l'implication possible d'éléments transposables dans le processus de compaction génomique chez les microsporidies. Le séquençage complet du génome d'*E. cuniculi* n'a cependant pas permis de révéler l'existence de tels éléments (Hinkle et al., 1997 ; katinka et

al., 2001). Néanmoins, la caractérisation d'une séquence génomique codant une reverse transcriptase chez l'espèce *Vittaforma corinae* a permis de conforter l'hypothèse d'un rôle important joué par ces éléments au cours de l'évolution des génomes microsporidiens (Mittleider *et al.*, 2002).

Très récemment, 8 familles d'éléments transposables (*Nbr*) de type *Ty3/gypsy* ont été identifiées au sein du génome de *Nosema bombycis* (Xu *et al.*, 2006). Ces éléments présentent de grande similarité avec d'autres éléments caractérisés chez les levures. La mise en évidence de mutations non sens au sein de ces différentes familles d'ETs (rétrotransposons) indique que bon nombre de ces éléments seraient actifs. De même, la caractérisation de remaniements génétiques ayant entraîné une perte de conservation de la synténie entre les espèces *E. cuniculi* et *N. bombycis* sur certains contigs où ont été localisés des éléments transposables, suggère de nouveau une participation active de ces éléments dans l'évolution des génomes microsporidiens (Xu *et al.*, 2006). Chez *E. cuniculi*, l'hypothèse d'une perte secondaire de ces éléments au cours de l'évolution a également été avancée (Bringaud *et al.*, 2006).

Si l'on considère l'ensemble des protozoaires, la plus grande diversité d'éléments transposables est principalement retrouvée chez les *Trypanosomatidae* parasites de l'Homme (*T. brucei* et *T. cruzi*), et d'insectes (*Crithidia fasciculata*). Ces éléments, potentiellement fonctionnels, apparaissent dispersés au sein du génome (Bhattacharya *et al.*, 2002). Chez *L. major*, phylogénétiquement très proche des trypanosomes *T. brucei* et *T. cruzi*, de nombreux éléments semblent avoir dégénéré et perdu leur autonomie (Bringaud *et al.*, 2006).

Trois familles d'éléments transposables ont été caractérisées chez le protozoaire flagellé, *Giardia lamblia*, dont deux en position subtélomérique (région connue pour être le siège de nombreuses recombinaisons) et une en position interne (Arkhipova et Morrison, 2001).

Enfin chez *Entamoeba histolytica*, plusieurs classes de rétrotransposons ont été identifiées (LINE, SINE). Ces éléments, présents en grand nombre de copies semblent dispersés au sein du génome (**Bakre et al., 2005**). La proportion relative d'éléments transposables retrouvée au niveau du génome de ces protozoaires varie de 1 à 40%, la plus faible et la plus forte proportion étant respectivement retrouvée chez *G. lamblia* et *T. cruzi*.

La présence de rétrotransposons dans les génomes de la plupart des organismes eucaryotes suggère une acquisition précoce de ces éléments au cours de l'évolution (**Bowen et Jordan, 2002**). En raison de leur capacité de déplacement au sein des génomes et des risques d'intégration au niveau de séquences codantes, des mécanismes de régulation cellulaires ont été mis en place afin de limiter leurs déplacements (ARN interférence).

II.1.3.2. Les séquences satellites

Le génome d'*E. cuniculi* contient très peu de séquences satellites (**Biderre et al., 1997**). La recherche de répétitions en tandem situées dans les régions subtélomériques du chromosome I a révélé la présence de 11 classes de répétitions de part et d'autre des unités d'ADNr (**Peyret et al., 2001**). Chez *A. locustae*, des séquences minisatellites correspondant à 5 répétitions d'un motif de 84 pb ont été identifiées au niveau d'une région intergénique (**Fast et al., 2003**).

Une très grande diversité de séquences satellites a cependant été mise en évidence chez de nombreux protozoaires parasites (revue par **Wickstead et al., 2003**). Différentes classes d'ADN satellite ont ainsi été caractérisées chez *P. falciparum*, *L. major*, *T. brucei*, *T. cruzi* et *G. lamblia*. Ces séquences répétées diffèrent dans la taille de leur unité, leur distribution au sein des génomes, et la plupart des répétitions apparaissent être entièrement spécifiques d'espèce. L'étude des génomes de *T. cruzi* et *T. brucei*, deux espèces phylogénétiquement proches, a révélé l'existence d'ADN satellite répété en tandem pouvant

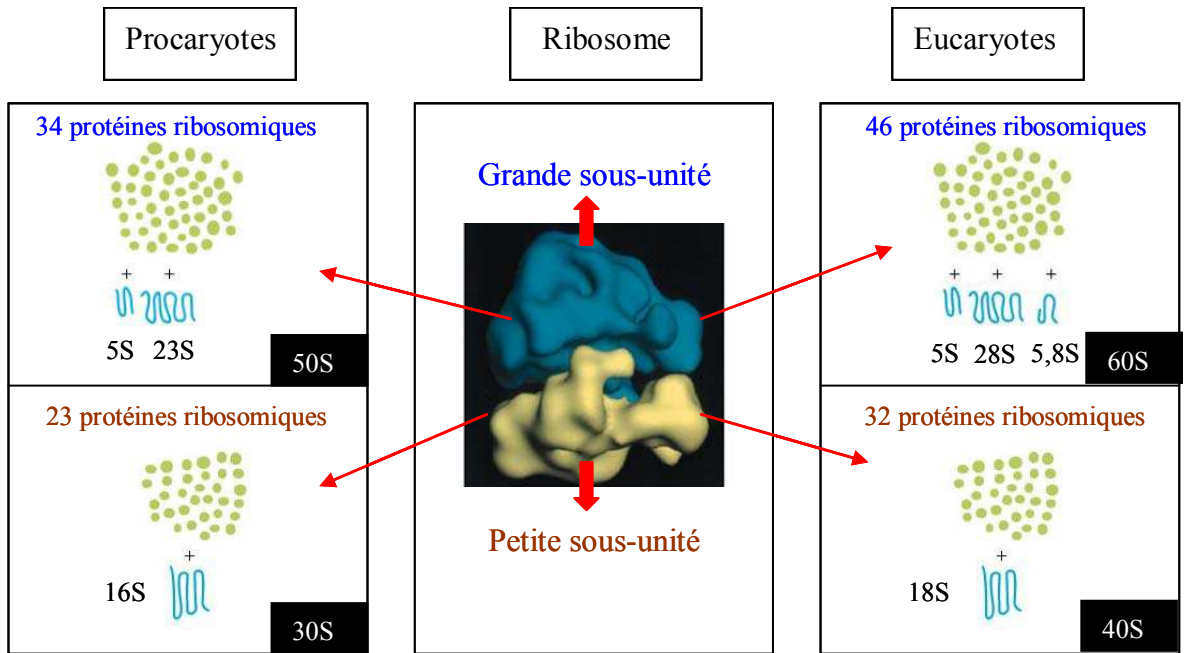


Figure 9. Représentation schématique de la structure et de la composition en ARNr et protéines d'un ribosome chez les Procaryotes et les Eucaryotes.

représenter jusqu'à 9% de leur génome. Chez *P. falciparum*, des infestations répétées réalisées chez la souris entraînent une perte drastique des séquences répétées. Corréliées à une diminution importante de la viabilité des gamétocytes, ces observations suggèrent un rôle essentiel de ces répétitions dans le cycle parasitaire (**Frontali et al., 1994**).

Une classe de répétition semble faire exception à la règle : celle localisée au niveau des régions subtélomériques. Ces répétitions apparaissent très bien conservées chez les différentes espèces citées précédemment, probablement en raison de leur importance dans le maintien de la structure des chromosomes (**Mylers et al., 1999**). Chez *T. brucei*, les gènes codant la protéine de surface VSG (Variable Surface Glycoprotein) sont organisés en clusters et à proximité des régions télomériques. Siège de processus de recombinaison importants, ces régions participent ainsi à la très grande variabilité antigénique parasitaire (**Cross et al., 1998**).

II.1.3.3. Les ARN ribosomiques microsporidiens

La compaction et la réduction importante des génomes microsporidiens semblent également avoir affecté de manière importante les gènes codant pour les ARN ribosomiques (**Refardt et Mouton, 2006**). Avec une constante de sédimentation de 70S, les ribosomes microsporidiens sont de type "procaryotique" (**Fig. 9**). En raison de leur extrême réduction, les ARN ribosomiques sont de types 16S et 23S, et figurent parmi les plus petits ARNr connus. Classiquement, chaque unité transcriptionnelle d'ADNr se compose d'un gène codant l'ARNr 16S (SSU), d'un espaceur interne (ITS1) et enfin d'un gène codant l'ARNr 23S (LSU). Ce dernier, en raison de l'absence d'ITS2 est fusionné à l'ARNr 5,8S (**Vossbrinck et Woese, 1986**).

La séquence complète d'une unité d'ADNr a été caractérisée pour la première fois chez l'espèce *E. cuniculi* (**Peyretaillade et al., 1998b**). Avec une taille de 2449 nucléotides,

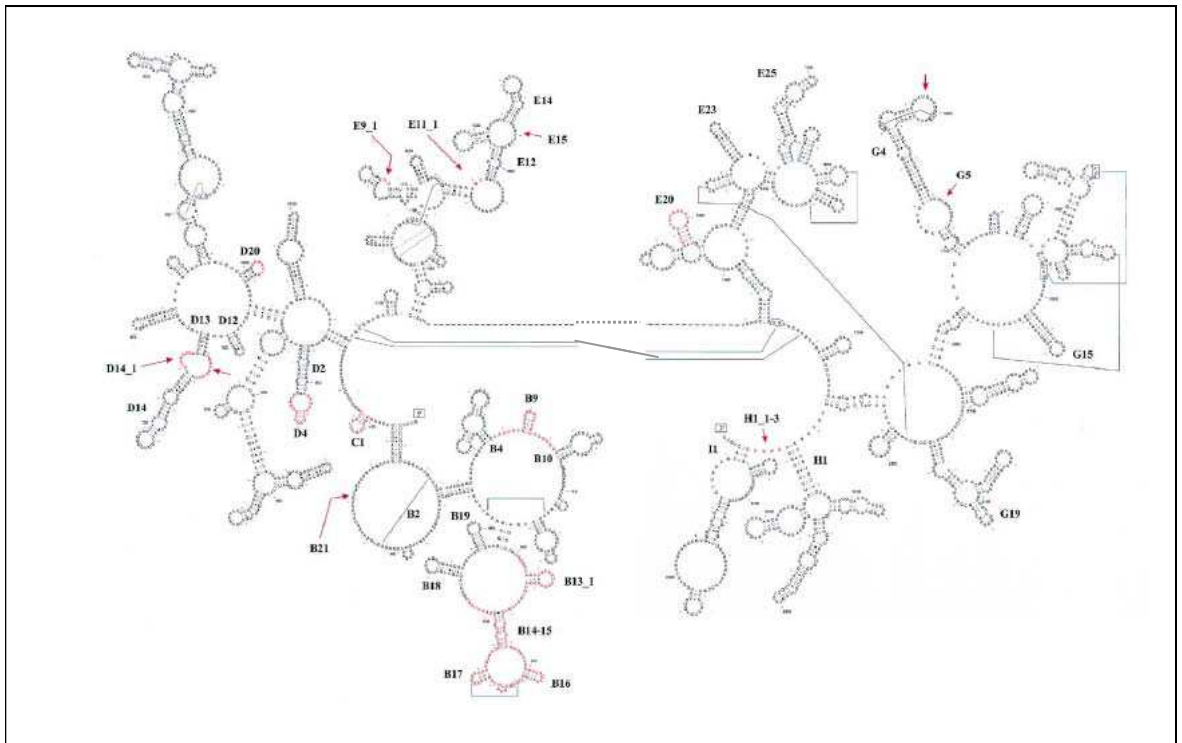


Figure 10. Structure secondaire de l'ARNr 23S d'*E. cuniculi*. (d'après Peyretailade et *al.*, 1998). Les régions variables en longueur et séquence chez *E. cuniculi* sont indiquées en rouge et / ou sont identifiées à l'aide d'une flèche rouge. Les interactions tertiaires sont indiquées par des lignes continues (Schnare et *al.*, 1996).

l'ADNr 23S d'*E. cuniculi* est le plus petit connu chez les Eucaryotes mais aussi chez les Procaryotes. La détermination de sa structure secondaire montre que cette dernière présente toutes les caractéristiques typiques des ARNr eucaryotes malgré la réduction ou l'absence de certains domaines (**Fig. 10**). Depuis, des unités complètes ont été caractérisées chez plusieurs espèces microsporidiennes dont *Nosema apis*, *Heterosporis anguillarum*, *N. bombycis*, *N. spodopterae*, *N. antheraeae*, *Glugoides intestinalis* (**Gatehouse and Malone, 1998; Tsai et al., 2005 ; Huang et al., 2004 ; Wang et al., 2006 ; Refardt et Mouton, 2006**). Chez ces quatre dernières espèces, des réarrangements de l'unité d'ADNr ont conduit à une inversion de l'ordre des gènes (**Refardt et Mouton, 2006**). Cette organisation originale des unités d'ADNr n'est pas retrouvée chez toutes les espèces du genre *Nosema* malgré la très forte proximité phylogénétique des espèces qui en font partie. L'existence d'un groupe unique d'espèce au sein de ce genre a ainsi été remise en cause (**Tsai et al., 2005**).

L'analyse du génome de *N. bombycis* a révélé l'existence de gènes complets et partiels codant l'ARNr 16S (**Iiyama et al., 2004**). La mise en évidence d'un élément transposable (transposon-like) en amont d'une des copies laisse supposer l'implication d'éléments mobiles dans ces remaniements. De plus, l'hypothèse de réarrangements localisés faisant suite à un processus de compaction de plusieurs unités en tandem, et comprenant *in fine* le dernier (LSU) et le premier gène (SSU) de deux unités consécutives a été émise (**Weider et al., 2005**). Chez *N. apis*, une organisation similaire des gènes au sein des unités d'ADNr a également été décrite (**Gatehouse et Malone, 1998**).

La localisation chromosomique des unités d'ADNr apparaît très hétérogène en fonction des espèces. Les résultats obtenus suite à l'hybridation des caryotypes moléculaires, avec une sonde correspondant à l'ARNr 16S des espèces *E. cuniculi* et *E. intestinalis*, montrent une dispersion inattendue des unités d'ADNr sur l'ensemble des bandes chromosomiques. Chez *E. cuniculi*, chaque chromosome est porteur de deux unités

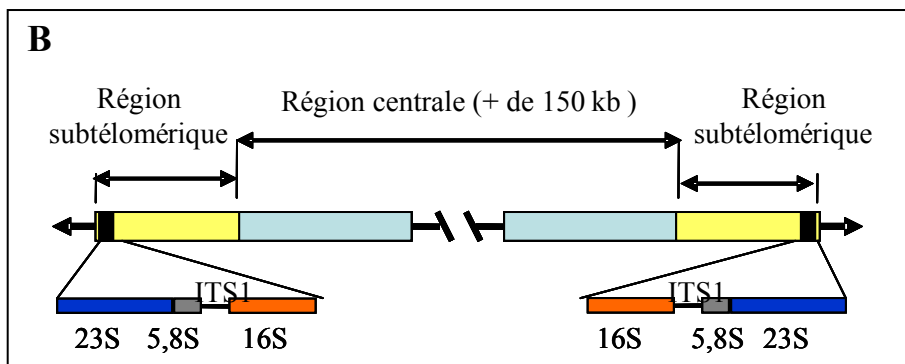
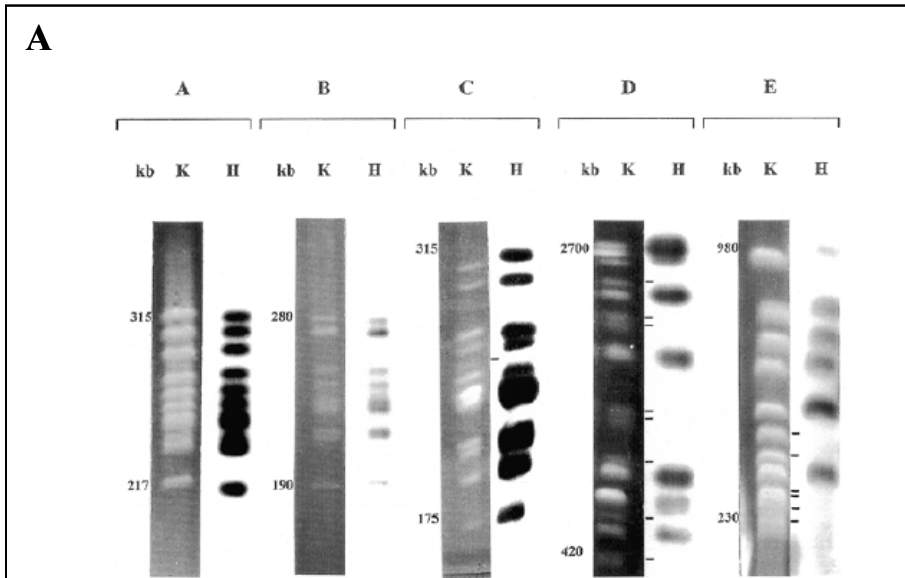


Figure 11. Localisation, chez différentes espèces microsporidiennes, des unités d'ADNr, et organisation structurale chez *E. cuniculi*.

(A) Résultats de l'hybridation réalisée avec une sonde correspondant à une partie de l'ADNr 16S permettant de visualiser la localisation des gènes de l'ADNr dans le génome de 3 espèces de microsporidies pathogènes de l'Homme (A : *E. cuniculi*, B : *E. intestinalis*, C : *E. hellem*) et de 2 parasites de poisson (D : *G. atherinae*, E : *S. lophii*). Les tailles du plus petit et du plus grand chromosome sont indiquées en kb, sur la gauche du caryotype moléculaire pour chaque espèce. K : caryotype moléculaire ; H : profil d'hybridation. Les chromosomes qui ne sont pas marqués par la sonde sont identifiés sur la droite du caryotype à l'aide d'un tiret. (B) Représentation schématique de l'organisation des chromosomes de l'espèce *E. cuniculi* déduite des données de séquençage du chromosome I (Peyret et al., 2001) et de cartographie physique (Brugère et al., 2000). Chaque chromosome est divisé en 2 principaux domaines : une partie centrale spécifique de chacun et composée essentiellement de séquences uniques ; 2 domaines terminaux symétriques (de 15 à 33 kb selon les chromosomes) contenant chacun 1 unité de transcription de l'ADNr et des séquences satellites présentant une organisation similaire. Les régions télomériques, estimées à environ 9 kb, sont représentées par les flèches noires et sont situées à l'extrémité de chaque unité d'ADNr.

ribosomiques, chacune en position subtélomérique et orientées de manière symétrique (**Fig. 11**), (**Peyret et al., 2001 ; Katinka et al., 2001**). Pour la troisième espèce du genre *Encephalitozoon*, *E. hellem*, seule 1 des 13 bandes d'ADN chromosomique semble être dépourvue d'unité ribosomique (**Peyretailade et al., 1998b**). A l'inverse, chez *N. bombycis*, les unités d'ADNr ne sont portées que par un seul chromosome (**Kawakami et al., 1994**). Pour les deux espèces parasites de poissons, *S. lophii* et *G. atherinae*, seule la moitié des bandes d'ADN chromosomique constituant leur caryotype moléculaire présente un signal positif (**Fig. 11**) (**Peyretailade et al., 1998b**).

Concernant le gène codant pour l'ARNr 5S, 3 copies, chacune sur un chromosome différent (V, VII et IX) ont été caractérisées chez *E. cuniculi* (**Katinka et al., 2001**). A l'inverse, chez les espèces *N. bombycis* et *N. spodopterae*, le gène codant pour l'ARNr 5S est situé respectivement à 279 pb et 277 pb en aval de l'unité d'ADNr (**Huang et al., 2004 ; Tsai et al., 2005**).

II. 2. Génomique comparative

La génomique comparative a été utilisée initialement pour la réalisation de cartes génétiques chez les Vertébrés (**Graves, 1998**). Basée initialement sur la recherche d'homologie entre séquences géniques afin de faciliter l'annotation syntaxique des génomes, cette discipline comprend aujourd'hui un ensemble d'axes de recherche qui portent sur : 1) la composition relative des génomes, 2) leur organisation structurale, 3) l'identification de gènes orthologues ou familles multigéniques, 4) la conservation de l'ordre des gènes ou synténie, 5) l'identification de gènes spécifiques d'espèces tels que ceux impliqués dans la spécificité d'hôtes et la pathogénicité, 6) l'identification de cibles thérapeutiques et enfin 7) l'évolution des génomes (**Carlton, 2003 ; Keeling et al., 2005 ; Luscher et al., 2007**).

Les premières études de génomique comparative réalisées chez les microsporidies ont permis de mettre en évidence une très grande hétérogénéité des profils caryotypiques (**Biderre et al., 1994**). De telles différences ne peuvent s'expliquer que par un degré de compaction moindre et/ou l'existence de répertoires géniques beaucoup plus conséquents au sein des génomes de taille importante. Afin d'apporter quelques éléments de réponse quant aux nombreuses questions qui entourent la structure génomique des microsporidies, l'étude comparative des génomes d'*E. cuniculi* (2.9 Mpb) et d'*A. locustae* (5.4 Mpb) a été initiée (**Slamovits et al., 2004**). *A. locustae* montre un degré de compaction génique comparable à celui d'*E. cuniculi* (0.94 gène/kb contre 0.97 chez *E. cuniculi*) ainsi que la présence d'espaces intergéniques très réduits. L'étude de certains fragments génomiques a permis de révéler l'existence de différences de compaction. Celles-ci laissent supposer l'existence de vitesses d'évolution distinctes en fonction du type d'interaction hôte-parasite considéré (**Slamovits et al., 2004**).

De manière surprenante, alors que ces deux espèces sont phylogénétiquement éloignées, cette même étude a montré qu'elles possèdent en commun près de 94% des gènes étudiés: ainsi, parmi les 183 gènes identifiés, 175 présentent des homologues avec *E. cuniculi*, 4 semblent être spécifiques d'*A. locustae* et 3 présentent des homologues avec des séquences identifiées chez d'autres organismes. A cette liste, on peut également ajouter deux autres gènes codant pour des protéines mitochondriales récemment découverts chez *A. locustae* et non détectés chez *E. cuniculi*: un transporteur de pyruvate et une protéase de la membrane interne mitochondriale (**Williams et Keeling, 2005**). A la lumière des résultats obtenus au cours de cette première analyse comparative, aucune donnée significative ne permet de justifier des différences de taille aussi importantes de leur génome, comme notamment la caractérisation d'une proportion importante de gènes spécifiques d'*A. locustae* (**Keeling et Slamovits, 2004**). Cependant, notons que ces résultats d'analyse ne portent que sur 10% du

génomique d'*A. locustae*. L'annotation du génome de cette microsporidie se poursuit actuellement, plus de 2600 CDS ont été annotés jusqu'à présent soit 30% de plus que chez *E. cuniculi*, ce qui est encore bien trop insuffisant pour expliquer des différences de taille de leur génome aussi importantes.

La comparaison des données génomiques obtenues pour *P. falciparum* et *P. yoelii* a mis en évidence 60% de gènes orthologues communs aux deux espèces et 40% de gènes spécifiques d'espèces ou appartenant à des familles multigéniques ayant évolué différemment (**Carlton et al., 2003**). La différence majeure entre ces deux génomes repose sur la taille des gènes chez *P. falciparum*. Celle-ci est en moyenne deux fois plus importante que chez *P. yoelii* (**Gardner et al., 2002**). L'analyse d'une région de 50 kb dont l'organisation et la structure génomique apparaissent très conservées entre ces deux espèces, a montré une remarquable conservation de la synténie pour 75% des gènes (**Carlton et al., 2003**).

Une approche comparative de l'architecture des génomes de *T. brucei*, *T. cruzi* et *L. major*, trois kinétoplastidés dont les distances phylogénétiques sont importantes, a non seulement révélé 6200 gènes orthologues chez ces trois espèces, mais également un très haut degré de conservation de la structure de leurs génomes (**El-Sayed et al., 2005**). A l'inverse, chez les espèces *E. invadens* et *E. histolytica*, ou encore les apicomplexes *C. parvum* et *P. falciparum*, une telle conservation n'est pas observée (**Wang et al., 2003 ; Bankier et al., 2003**).

III/ REGULATION GENIQUE CHEZ LES PROTOZOAIRES PARASITES

III.1. Régulation transcriptionnelle

Chez les eucaryotes, la régulation de l'expression des gènes nécessite la coordination spatio-temporelle de facteurs transcriptionnels, généraux et spécifiques, capables de reconnaître des motifs localisés dans les régions promotrices (revue par **Heintzman et Ren, 2007**). La fixation de ces facteurs transcriptionnels au niveau des séquences promotrices (séquences *cis*) nécessite que la chromatine soit dans un état conformationnel adéquate (**Mellor, 2005**).

III.1.1. La pré-initiation de la transcription

Plusieurs régions situées en amont et en aval du site d'initiation de la transcription sont de grande importance pour la transcription et constituent le "cœur" du promoteur (**Juven-Gershon et Kadonaga, 2006**). Parmi ces éléments figurent la boîte TATA, située 25 à 30 nt en amont du site +1 de transcription, l'élément Inr (Initiator Element) qui englobe le site +1 de transcription, les éléments BRE (TFIIB Recognition Elements) situés de part et d'autre de la boîte TATA, et enfin, plusieurs motifs localisés en aval du site +1 de transcription dont MTE (Motif Ten Element), DPE (Downstream Promoter Element) ou encore DCE (Downstream Core Element) (**Juven-Gershon et Kadonaga, 2006**). L'étape de pré-initiation représente une étape essentielle dans la régulation de l'expression génique. Celle-ci est responsable du recrutement de la machinerie basale de transcription ou complexe de pré-initiation, traditionnellement représenté par l'ARN polymérase II et les facteurs généraux de transcription (GTFs pour General Transcription Factors) que sont TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, TFIIH. La reconnaissance de chaque élément constituant la cœur du promoteur se fait de manière séquentielle (**Woychik et Hampsey, 2002**). Plus de 2000 facteurs de transcription spécifiques de certains gènes ou groupes de gènes ont été caractérisés au niveau

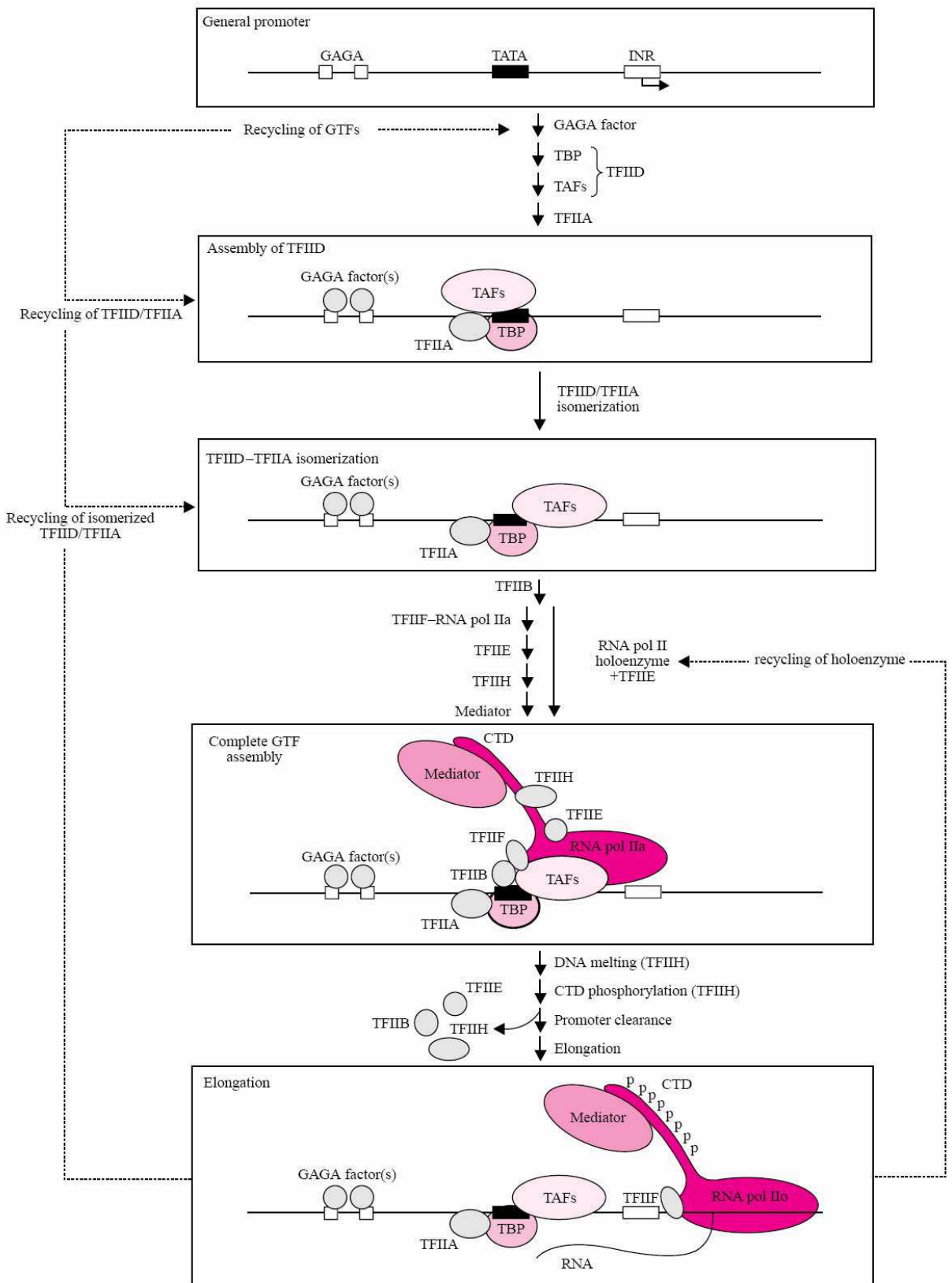


Figure 12. Représentation schématique du modèle d'assemblage séquentiel du complexe de pré-initiation sur le promoteur, et activation de l'étape d'élongation (D'après Ogbourne et Antalis, 1998).

du génome humain (**Brivanlou et Darnell, 2002**). Ces facteurs reconnaissent des séquences d'ADN spécifiques en amont du promoteur, et interagissent avec de nombreux co-régulateurs pour assurer la régulation de la transcription.

III.1.2. Le complexe basal d'initiation de la transcription

Le complexe TFIID, *via* la sous-unité TBP (TATA Binding Protein), se lie à la boîte TATA et sert de plate-forme à l'assemblage de la machinerie transcriptionnelle (**Fig. 12**) (**Lee et Young, 1998**). Les protéines TAFs (TBP-Associated Factors), qui entrent dans la composition du complexe TFIID, participent à la régulation de la transcription en augmentant l'affinité de TBP pour la boîte TATA, mais également en modulant l'activité d'autres facteurs transcriptionnels (**Albright et Tjian, 2000**).

Les facteurs TFIIA et TFIIB participent entre autre à la stabilisation du complexe TFIID en interagissant avec TBP et les séquences nucléiques bordant la boîte TATA. Le rôle de TFIIB est primordial puisque ce facteur assure également le recrutement de l'ARN polymérase II au niveau du site de démarrage de la transcription *via* son interaction avec TFIIF (revue par **Nikolov et Burley, 1997**). Enfin le facteur TFIIE, qui interagit directement avec l'ARN polymérase II, stimule l'activité du facteur TFIIH essentielle aux étapes d'initiation et d'élongation de la transcription (**Roeder, 1996**).

La grande sous-unité de l'ARN polymérase II (*cpb1*) possède un domaine carboxy-terminal (CTD) composé de motifs répétés (YSPTSPS). Deux résidus sérine (Ser2 et Ser5) par motif peuvent être phosphorylés. Chez les mammifères et la levure *S. cerevisiae*, le nombre d'unités répétées est respectivement de 52 et 26. Chez *S. cerevisiae*, des modifications du nombre d'unités répétées (de 26 à 11) entraînent un dysfonctionnement de la machinerie transcriptionnelle (**West et Corden, 1995**).

L'initiation de la transcription consiste en l'avancée de l'ARN pol II et la synthèse des 10 à 15 premiers nucléotides (revue par **Conaway et al., 2000**). Cette étape assure la libération du promoteur et permet le recrutement d'un nouveau complexe d'initiation de la transcription.

III.1.3. L'élongation de la transcription

L'hyperphosphorylation du domaine CTD s'accompagne d'un changement conformationnel de l'ARN polymérase qui peut alors entamer la phase d'élongation (**Dahmus, 1996**). Au cours de celle-ci, plusieurs facteurs d'élongation (activateurs ou inhibiteurs) peuvent intervenir afin de réguler le niveau d'activité de la machinerie transcriptionnelle (**Conaway et al., 2000**). Certains de ces facteurs, appartenant au complexe positif d'élongation comme P-TEFb (Positif Transcription Elongation Factor b) ou encore les Elongin A et A2, interviennent principalement en permettant la suppression des pauses de la polymérase ; d'autres, comme DSIF (DRB-Sensitivity Inducing Factor) ou NELF (Negative Elongation Factor Complex), inhibent l'étape d'élongation de manière synergique en interagissant directement avec l'ARN polymérase (**Yamagushi et al., 1999**).

III.1.4. La terminaison de la transcription

L'étape de terminaison de la transcription nécessite la reconnaissance d'un signal de polyadénylation (AAUAAA) dont la dégénérescence apparaît plus importante chez *S. cerevisiae* que chez les eucaryotes supérieurs (**Zhao et al., 1999**). Chez les mammifères, le clivage a lieu entre ce motif de polyadénylation et l'élément DSE (Downstream Sequence Element) riche en résidus uracile (ou GU), et ceci, de manière privilégiée au niveau du dinucléotide CA (pour revue **Proudfoot, 2004**). Les facteurs CPSF (Cleavage and Polyadenylation Specificity Factor) et CstF (Cleavage Stimulatory Factor) se fixent

respectivement sur les motifs AAUAAA et DSE, et participent de manière active à la reconnaissance des signaux de clivage. Deux facteurs essentiels, CF I et CF II (Cleavage factor I et II) assurent le clivage du transcrit. Cette étape de clivage nécessite une étroite interaction entre l'ensemble des facteurs précédemment cités mais également la poly(A) polymérase, dont la processivité est stimulée par la fixation des protéines PABP (PolyA Binding Protein) au niveau de la queue poly(A) en formation (pour revue **Wahle et Ruegsegger, 1999**).

Enfin, soulignons l'importance du domaine CTD qui joue un rôle crucial dans le recrutement séquentiel des différents complexes enzymatiques et facteurs de transcription au cours des étapes d'initiation, d'élongation et de terminaison, mais également dans la coordination des processus co-transcriptionnels (**Maniatis et al., 2002 ; Proudfoot, 2004**).

III.1.5. Les signaux de transcription chez les protozoaires

Chez les Apicomplexa comme *T. gondii* et *P. falciparum*, les différents éléments *cis* comme la boîte TATA ou la boîte CAAT ne sont pas conservés (revue par **Meissner et Soldati, 2005**). En revanche, un motif composé de 6 répétitions de 27 pb, avec comme séquence consensus central TGAGAGC, semble être impliqué dans l'initiation de la transcription du gène *sag1* (**Soldati et Boothroyd, 1995**). D'autres promoteurs présentant cette séquence ont pu être caractérisés, suggérant un rôle important de cette séquence dans le choix du site d'initiation de transcription. Chez *Plasmodium*, l'étude du génome et les expériences de transfection ont permis d'identifier un îlot G-box palindromique (**Militello et al., 2004**).

De nombreux composants de la machinerie de transcription eucaryote ont été mis en évidence chez *T. gondii* et *P. falciparum*. Cependant, certaines exceptions méritent d'être soulignées comme par exemple l'absence apparente de la plupart des facteurs TFIIA, TFIIE β ,

TFIIIF et de la plupart de facteurs associés à TBP. Ceci suggère l'existence d'un appareil de transcription primitif mais néanmoins suffisant pour assurer la régulation de la machinerie transcriptionnelle. Pour de nombreux transactivateurs comme VP16 de l'herpès virus simplex (HVS), plusieurs interactions avec des facteurs de transcription (TFIIA, TFIIB, TFIID ou TFIIF) et un facteur associé à TBP (TAF31) ont été décrites pour l'activation *in vitro* (**Flint et Shenk, 1997 ; Uesugi et al., 1997**). Aucune transactivation à l'aide de VP16 n'a pu être mise en évidence chez *T. gondii* et *P. falciparum*, ce qui semble suggérer l'existence d'un mécanisme de régulation transcriptionnel différent (revue par **Meissner et Soldati, 2005**). Cependant, chez *P. falciparum*, l'utilisation de la méthode HCA (Hydrophobic Cluster Analysis), basée sur la recherche de déterminants structuraux universels impliqués dans le repliement protéique, a révélé l'existence de facteurs de transcription non caractérisés par simple recherche d'homologie (**Callebaut et al., 2005**). L'identification, chez *P. falciparum*, de ces nouveaux facteurs de transcription remet en cause l'idée d'une simplification à l'extrême des mécanismes de régulation de la transcription chez les Apicomplexa.

Chez les *Trypanosomatidae* (*T. brucei*, *T. cruzi* et *L. major*), le mécanisme de transcription est particulier puisque des pré-ARNm sont synthétisés sous forme polycistronique. Des motifs, localisés au niveau des régions intergéniques, permettent d'assurer la polyadénylation et le capping des différents gènes du messenger polycistronique (revue par **Palenchar et Bellofatto, 2006**). Chez ces organismes, les mécanismes d'initiation de la transcription sont peu connus, et seul un nombre limité de promoteurs ont été caractérisés jusqu'à présent (**Ivens et al., 2005**). Ainsi, l'absence apparente d'un système de régulation de la transcription des gènes de classe II laissait donc suggérer une initiation aléatoire de la transcription (**McAndrew et al., 1998**). Cependant, le séquençage du chromosome I de *L. major* a révélé une organisation génomique particulière : la totalité des 79

gènes est regroupée au sein de deux unités polycistroniques divergentes (**Myler et al., 1999**). L'étude des mécanismes de transcription de ces gènes a permis d'identifier un site spécifique d'initiation de la transcription localisé entre les deux unités (**Martinez-Calvillo et al., 2003**). Un promoteur bidirectionnel assurerait la transcription de chacune des unités polycistroniques jusqu'au niveau des régions télomériques.

En ce qui concerne la machinerie de transcription, des facteurs généraux de transcription sont absents (TFIIB, TFIIE, TFIIIF, certaines sous-unités de TFIIF et co-facteurs transcriptionnels). Chez les eucaryotes supérieurs, la protéine TBP joue un rôle prépondérant dans la formation du complexe de pré-initiation. Celle-ci présente deux doublets de phénylalanine au niveau de son domaine carboxy-terminal conservé, ces acides aminés jouant un rôle essentiel dans la liaison à la boîte TATA. Chez *Trypanosoma*, la protéine TBP ne présente pas le premier doublet de phénylalanine, suggérant qu'elle se lie différemment à l'ADN. De plus, l'absence de répétitions heptapeptidiques au niveau du domaine CTD de l'ARN polymérase implique, chez *Trypanosoma*, une orchestration différente des mécanismes de régulation transcriptionnelle (**Ivens et al., 2005 ; Palenchar et Bellofatto, 2006**). Aucune hypothèse n'a cependant été émise quant à ces mécanismes. En ce qui concerne la maturation de chacun des messagers du polycistron, l'épissage en *trans* des gènes situés en aval semble davantage définir le site de polyadénylation du gène en amont que les motifs consensus usuels (**Matthews et al., 1994**).

Chez *Trichomonas vaginalis*, un motif conservé de 13 nucléotides, riche en pyrimidine, et comprenant le site +1 de transcription, a été caractérisé en amont du codon d'initiation de traduction (**Quon et al., 1994 ; pour revue Marmorstein, 2003**). Ce motif présente un domaine très conservé d'initiation de la transcription (Inr) CA₊₁YT/A et apparaît suffisant pour assurer, *in vitro*, une transcription efficace. Néanmoins, l'existence d'éléments

distants impliqués dans la régulation de l'expression de certains gènes ne peut être exclue (**Liston et al., 1998**). Ainsi, pour certains gènes, une boîte TATA située 30 nt en amont du site +1 de transcription est parfois retrouvée. De plus, certains gènes présentent une région 5'UTR anormalement riche en résidus adénine et thymine pouvant être impliquée dans la régulation de la transcription (**Hahn et al., 1989 ; Singer et al., 1990**). Concernant le recrutement de la machinerie transcriptionnelle chez *T. vaginalis*, la recherche de facteurs d'initiation a permis la caractérisation d'une protéine de 39 kDa, appelée IBP39, capable d'interagir avec l'élément Inr (**Liston et al., 2001**). Le domaine C-Terminal de cette protéine assure le recrutement de l'ARN polymérase II en interagissant avec le domaine CTD (**Lau et al., 2003 ; Schumacher et al., 2003 ; Lau et al., 2006**).

Chez *Entamoeba histolytica*, un motif consensus AAAAATTCA pouvant jouer le même rôle que l'élément Inr chez *T. vaginalis* a pu être caractérisé (**Buss et al., 1995 ; Singh et al., 1997**). De plus, deux éléments *cis*, l'un correspondant à une boîte TATA-like située 30 nt en amont du site + 1 de transcription et l'autre à une boîte GAAC située entre les deux, ont été caractérisés (**Purdy et al., 1996**). Chez *G. lamblia*, 21 des 28 sous-unités protéiques composant chacune des ARN polymérases et seulement 4 des 12 facteurs d'initiation de la transcription (TAFs) ont pu être identifiés. La divergence importante de certains facteurs d'initiation de la transcription tel TBP, ou encore l'absence du facteur TFIIB, révèlent une machinerie transcriptionnelle primitive en accord avec l'émergence précoce de ces parasites (**Best et al., 2004**).

Chez les microsporidies, peu de données concernant l'expression des gènes sont disponibles. Les gènes microsporidiens présentent des régions AT riches en amont de l'ATG qui ont permis la caractérisation de deux séquences consensus: les boîtes A (TATAWA) et B

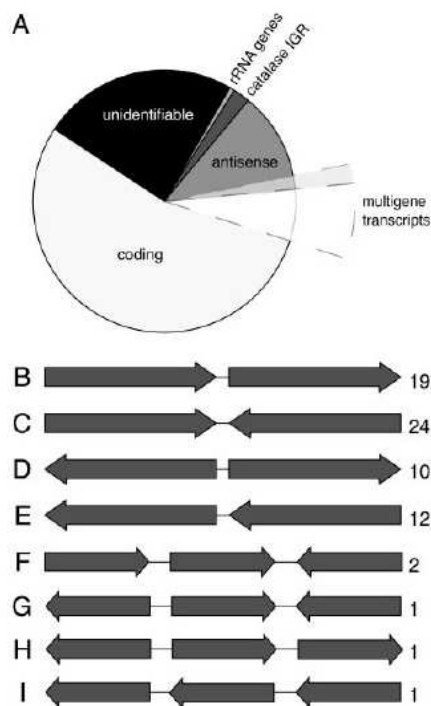


Figure 13. Distribution des différents types de messagers identifiés dans la banque d'ESTs d' *A. locustae*. (A) Classement des 1146 ARNm en fonction du type de transcrit. 24% des messagers ne présentent aucune homologie avec des séquences connues. 1% correspondent à des fragments des opérons ARNr et 2% à des éléments répétés présents dans la région intergénique de la catalase d'*A. locustae*. 13% représentent des transcrits antisens dont 15% sont des antisens de plus d'un gène. 61% représentent des messagers sens de gènes connus dont 11% sont des transcrits contenant deux gènes ou plus en orientation sens ou antisens. (B-I) Représentation schématique des transcrits codant pour plus d'un gène. Le chiffre à droite indique le nombre de transcrits identifiés pour chaque cas. (d'après Williams *et al.*, 2005).

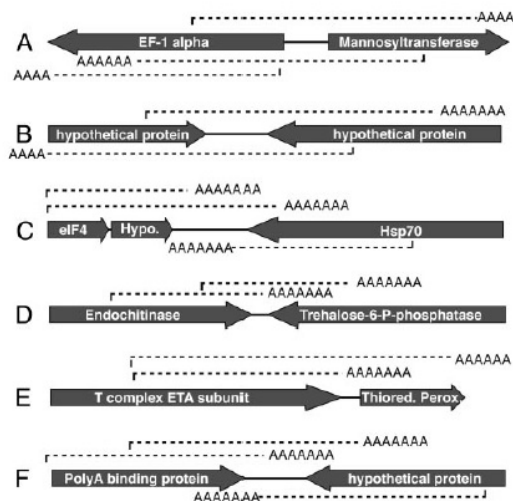


Figure 14. Exemple de messagers identifiés dans la banque d'ESTs d' *A. locustae*. Les flèches grises représentent la position et l'orientation des gènes dans leur contexte génomique (le nom des gènes est mentionné). Les pointillés, orientés par la queue poly(A), indiquent les transcrits détectés dans la banque d'ESTs (d'après Williams *et al.*, 2005).

(SAAAAT) situées respectivement 20 et 30-60 nucléotides en amont du codon d'initiation de traduction, assimilables aux boîtes TATA et CAAT classiquement retrouvées au niveau des promoteurs eucaryotes (**Duffieux et al., 1998**). Le séquençage du génome d'*E. cuniculi* a révélé la présence des composants de la machinerie de transcription typiquement eucaryotes : ARN polymérase II, facteurs de transcription TFIIA, TFIID, TFIIE, TFIIIF et la plupart des sous-unités de TFIIH. Les différents gènes codant les enzymes impliquées dans la formation de la coiffe (triphosphatase, guanylyltransférase et méthyltransférase) ou encore la polyadénylation des messagers sont également présents (**Katinka et al., 2001 ; Hausmann et al., 2002**). La séquence consensus AAUAAA correspondant au signal de polyadénylation des eucaryotes a pu être identifiée pour de nombreux transcrits. Il a également été montré par des approches de 3' RACE-PCR que les régions 3' UTR présentent des tailles variables allant de 25 à 175 nt (**Broussolle et al., 1999 ; Delbac et al., 2001a, Peuvel et al., 2002**).

Récemment, chez *A. locustae*, l'analyse d'une banque d'ESTs a permis de mettre en évidence l'existence de nombreux messagers codant plusieurs gènes (**Fig. 13**) (**Williams et al., 2005**). L'ensemble des données collectées suggère la perte de signaux de terminaison au niveau de la région 3' UTR et l'utilisation de signaux cryptiques. Dans certains cas, les messagers identifiés commencent par un antisens suggérant que l'initiation de la transcription a lieu dans le gène amont (**Fig. 14**). Notons également qu'une fraction importante de ces transcrits (13%) code pour des gènes antisens, laissant supposer l'existence d'un mécanisme de régulation post-transcriptionnelle. Ces auteurs proposent que la compaction du génome d'*A. locustae*, caractérisée par une importante réduction des régions intergéniques, a entraîné la perte des signaux d'initiation et de terminaison de la transcription indispensables à la régulation de l'expression des gènes. Dans la mesure où leur approche expérimentale ne permet pas de définir avec certitude l'extrémité 5' des messagers, une réserve peut être émise quant à l'hypothèse d'une perte des signaux d'initiation.

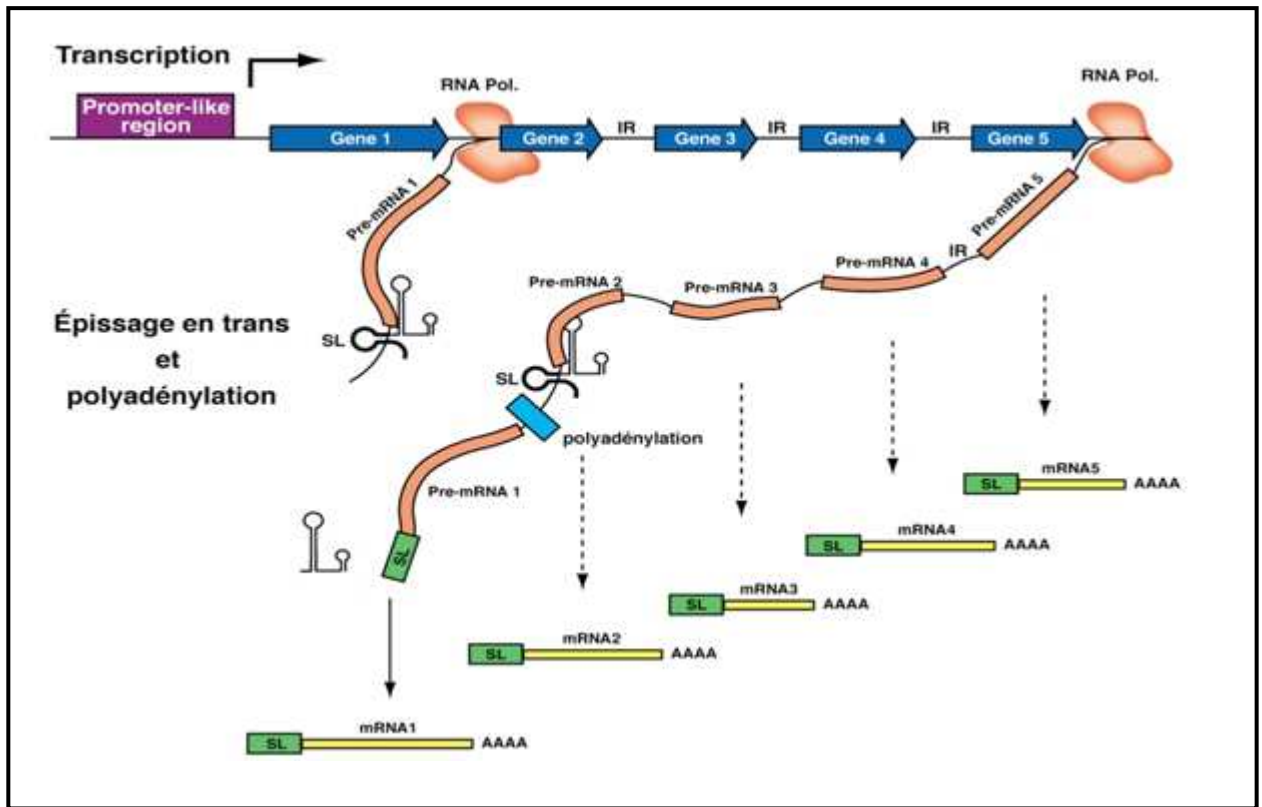


Figure 15. Transcription polycistronique et épissage en trans chez les *Kinetoplastidae* (d'après les travaux de thèse de Chou M., 2005).

Au cours de la transcription, l'ARN polymérase génère de longs pré-ARNm. La maturation de chaque cistron nécessite deux réactions de clivage. La première, une réaction d'épissage en *trans*, fait intervenir un ARN (SL) de 39 nt, dont le rôle est d'assurer la maturation en 5' de chaque ARNm monocistronique. La seconde, qui correspond à une réaction de clivage de l'extrémité 3', est suivie d'une polyadénylation.

III.2. Mécanismes de régulation post-transcriptionnelle

Chez les protozoaires parasites, la régulation de l'expression des gènes se fait de manière privilégiée à un niveau post-transcriptionnel et selon différentes modalités (**Meissner et Soldati, 2005**).

III.2.1. L'épissage en trans (trans-splicing)

Chez les *Kinetoplastidae*, la transcription des unités polycistroniques par l'ARN polymérase de type II génère des polycistrons pouvant atteindre plusieurs dizaines de kb (**McDonagh et al., 2000 ; Monnerat et al., 2004 ; Benz et al., 2005**). Par un processus d'épissage en *trans*, chaque monocistron subit une étape de maturation caractérisée par l'ajout d'un mini-exon en 5' (séquence leader) et d'une queue poly(A) en 3' (**Fig. 15**).

Chez *T. brucei*, tous les facteurs appartenant au *trans*-spliceosome ont été caractérisés (pour revue **Liang et al., 2003**). Un complexe snRNP (small nuclear RiboNucleoProtein) assure le transfert d'une séquence d'ARN leader de 39 à 41 nt au niveau de l'extrémité 5'UTR de chaque cistron (**Milhausen et al., 1984 ; Perry et al., 1987 ; Milhausen et al., 1984**). Cette séquence apparaît relativement bien conservée, les 6 premiers nucléotides étant invariablement retrouvés chez tous les *Kinetoplastidae*. Au niveau de la séquence d'ARN leader, les 4 nucléotides situés après la 7 méthyl-guanosine (m⁷G) présentent une signature structurale spécifique (**Bangs et al., 1992**). Une autre caractéristique semble être la présence, 100 à 400 nt en amont du site 3' d'épissage, d'un motif polypyrimidine (polyY) de taille variable. **Siegel et al. (2005)** ont montré que la taille optimale de ce motif est de 25 nt, mais aussi, que la distance séparant ce motif du site 3' d'épissage a un impact direct sur l'efficacité du processus.

III.2.2. Epissage alternatif

Au cours de la transcription, les ARN pré-messagers subissent un processus d'épissage en *cis* permettant d'éliminer les introns (*cis-splicing*). Ce processus génère un ARNm mature formé par l'association des exons. Au cours de l'épissage alternatif, un gène peut générer plusieurs populations d'ARNm (ou variants d'épissage). Chacun de ces messagers correspond à l'association de différentes combinaisons d'exons. La complexité combinatoire qui en résulte contribue à la diversité biologique et permet, à partir d'un gène unique, la synthèse de plusieurs protéines dont les propriétés biologiques diffèrent selon les tissus ou encore le stade de développement. D'un point de vue phylogénétique, ce mécanisme est apparu très tôt et semble très bien conservé dans le règne animal. Chez les protozoaires, très peu d'études relatent de manière exhaustive la mise en évidence d'un tel mécanisme, laissant supposer un rôle mineur de ce processus en matière de régulation post-transcriptionnelle.

Chez les *Trypanosomatidae*, où les gènes sont organisés en unités polycistroniques et apparemment dépourvus d'introns, il était admis que seul un mécanisme d'épissage en *trans* assurait la maturation de chaque transcrit (**Mottram et al., 1989 ; Ares et Weiser, 1995**). Les premiers travaux ayant permis de mettre en évidence l'existence d'introns chez *T. brucei* et l'existence d'un mécanisme d'épissage alternatif ont été réalisés par **Mair et al. (2002)**, avec la caractérisation d'un intron au niveau du gène codant la poly (A) polymérase. Chez *T. cruzi*, plusieurs variants d'épissage du gène *LYT1*, impliqué dans le cycle lytique, sont différenciellement exprimés au cours du cycle de développement parasitaire (**Manning-Cela et al., 2002**).

Enfin, chez *T. gondii*, les myosines B et C qui ne diffèrent que par leur extrémité C-terminale, sont obtenues par épissage alternatif (**Delbac et al., 2001b**). Chez ce même organisme, une approche visant à caractériser de nouvelles cibles thérapeutiques (voies de

synthèses des purines) a permis d'identifier deux variants du gène *IMPDH* (Inosine MonoPhosphate Deshydrogénase) (Sullivan *et al.*, 2005).

Ce mode de régulation post-transcriptionnel ne semble cependant pas être à l'origine d'une diversité protéique importante chez les protozoaires.

III.2.3. Régulation par expression d'ARN antisens

De telles molécules représentent plus de 20% des transcrits chez *G. lamblia* (Elmendorf *et al.*, 2001). Ces auteurs ont suggéré que ce taux anormalement élevé était dû à une perte du mécanisme de contrôle du processus de transcription entraînant l'utilisation de promoteurs cryptiques. Cependant, les travaux décrits par Von Allmen *et al.* (2005) montrent que la présence des ARN antisens semble régulée. En effet, leur synthèse ne se fait pas de manière constitutive mais est directement liée à la production de l'ARN sens correspondant. Chez l'espèce *P. falciparum*, les ARNs antisens sont également fortement représentés (Gunasekera *et al.*, 2004). Militello *et al.* (2005) proposent que leur synthèse soit dépendante de la structure de la chromatine (structure ouverte) rendant les deux brins compatibles pour la transcription par l'ARN polymérase II. Ces ARNs antisens, comme c'est le cas chez d'autres organismes, peuvent réguler la transcription des ARNs sens en entrant en compétition avec les facteurs de transcription. Ils peuvent également participer à la régulation de la traduction ou à la stabilité des ARNs suite à leur association avec les ARNs sens (Carmichael 2003). Des ARNs antisens ont également été décrits chez les kinétoplastidés (Martinez-Calvillo *et al.*, 2003 ; Monnerat *et al.*, 2004). Chez *Leishmania*, Dumas *et al.* (2006) ont mis en évidence l'expression au stade amastigote d'une nouvelle classe d'ARN non codant (ARNnc) de 300 à 600 nt, impliqués probablement dans la différenciation parasitaire. Des ARNs antisens, pouvant jouer un rôle dans la régulation post-transcriptionnelle de ces ARNs non codant, ont également été caractérisés.

Différents résultats laissent penser qu'une telle régulation existe chez les microsporidies. En effet, le séquençage d'une banque d'ESTs correspondant à l'espèce *A. locustae* a permis d'isoler 13% de transcrits codant des gènes antisens (**Williams et al., 2005**).

III.2.4. Régulation par ARN interférence

Initialement découvert chez le nématode *Caenorhabditis elegans*, l'ARN interférence (ARNi) désigne un mécanisme naturel d'inhibition post-transcriptionnelle faisant suite à l'introduction ou l'expression endogène d'un ARN double brin (**Fire et al., 1998**). Ce processus, relativement bien conservé, a été caractérisé chez de très nombreux eucaryotes (plantes, mammifères, protozoaires, nématodes, etc.) (**Agrawal et al., 2003**). Les fonctions de ce processus sont multiples : protection du génome contre les virus ou les transposons qui, au cours de leur cycle de réplication, passent par des intermédiaires ARNs double brin, mais également régulation de l'expression des gènes par les microARNs (**Zhang et al., 2007**).

Chez les protozoaires parasites, le mécanisme d'ARNi a initialement été décrit chez *T. brucei* (revue par **Motyka and Englund, 2004 ; Ullu et al., 2004**). Chez ce parasite, plusieurs enzymes impliquées dans le processus (TbAGO1, TbDcl1) ont pu être identifiées (**Shi et al., 2006a et b**). Contrairement à ce qui se passe chez *T. brucei* et *T. congolense*, tous les efforts développés pour induire une dégradation ciblée d'ARNm par ARNi chez *L. major* et *T. cruzi* sont restés vains (**Robinson et Beverley, 2003**). L'analyse du génome ces deux derniers parasites n'a pas révélé l'existence de gènes orthologues codant pour des protéines impliquées dans le processus d'interférence. Des approches visant à caractériser la fonctionnalité du mécanisme chez *T. cruzi* ont été réalisées sans succès (**DaRocha et al., 2004**). Cette approche s'est néanmoins révélée fonctionnelle chez *T. gondii* et *E. histolytica* (**Al-Anouti et Ananvoranich, 2002 ; Kaur et Lohia, 2004**). Le séquençage du génome d'*E. histolytica* a d'ailleurs permis, par la suite, de mettre en évidence plusieurs familles de

microARNs, confirmant l'existence d'un mécanisme de régulation post-transcriptionnelle par ARN interférence (**De et al., 2006**).

Chez *P. falciparum*, plusieurs études suggèrent l'existence du mécanisme d'ARNi malgré les difficultés liées à l'identification de gènes impliqués dans ce mécanisme (**McRobert et McConkey, 2002 ; Malhotra et al., 2002 ; Mohammed et al., 2003**).

Enfin, le séquençage complet du génome d'*E. cuniculi* n'a pas révélé la présence de gènes codant les protéines AGO1 ou encore Dicer, ce qui aurait pu ouvrir la voie à l'application d'une technique de génétique reverse efficace (**Katinka et al., 2001**).

III.3. Régulation traductionnelle

La synthèse protéique, qui suit et affine le contrôle exercé lors de la transcription, représente une étape importante de l'expression des gènes (**Mathews et al., 2000**). Un contrôle efficace de la traduction des ARNm aboutit à une modulation rapide des quantités de protéines dans un contexte spatio-temporel, à l'échelle de la cellule et des tissus. La traduction se déroule classiquement en trois phases : l'initiation, l'élongation et la terminaison. L'initiation de la synthèse protéique demeure une phase régulatrice primordiale de la synthèse protéique (**Gingras et al., 1999**).

Les ARNm eucaryotes possèdent à leur extrémité 5' UTR une structure essentielle : la coiffe (m^7 Guanosine). L'addition de cette coiffe (ou capping), qui correspond à la maturation de l'extrémité 5' du messenger, a lieu dès le début de la transcription (**Maniatis et al., 2001**). En effet, la phosphorylation du domaine CTD de l'ARN polymérase II permet le recrutement d'un complexe enzymatique composé d'une ARN triphosphatase, d'une guanylyltransférase (GT) ainsi que d'une méthyltransférase (MT), dont le rôle est d'assurer l'ajout de cette coiffe (**Proudfoot et al., 2002**). Celle-ci possède plusieurs rôles distincts (protection contre la dégradation par des exonucléases, exportation de l'ARNm vers le cytoplasme) mais ne sera

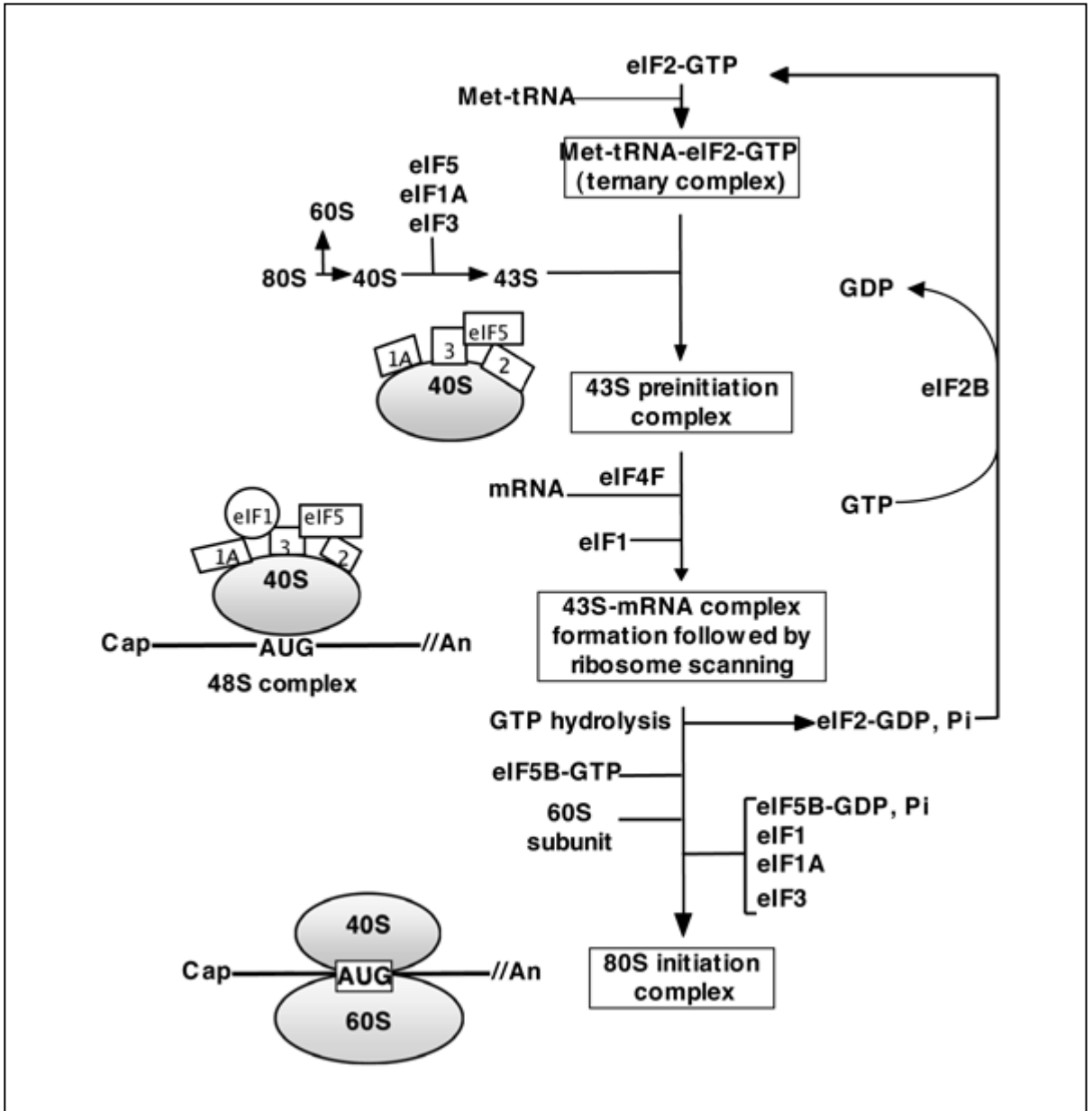


Figure 16. Diagramme illustrant la formation séquentielle du complexe d'initiation de la traduction chez les eucaryotes (d'après Lopez-Lastra et al., 2005).

La sous-unité 40S, en interaction avec de nombreux facteurs d'initiation (eIFs) s'associe au complexe ternaire : une fois assemblé, ce complexe de préinitiation 43S est recruté à l'extrémité 5' de l'ARNm par l'intermédiaire du facteur eIF4F (initiation "cap-dépendant"). Ce complexe glisse ensuite le long de l'ARNm jusqu'au codon d'initiation de la traduction (scanning), où a lieu la formation du complexe d'initiation 80S. Après relargage, les facteurs d'initiation sont recyclés.

abordé ici que son implication dans le recrutement des facteurs d'initiation de la traduction nécessaires à la fixation de la sous-unité 40S du ribosome.

III.3.1. L'initiation de la traduction chez les eucaryotes

L'initiation de la traduction est une étape complexe qui nécessite le recrutement séquentiel de nombreux facteurs eIFs (eukaryotic translation Initiation Factors), dont le rôle est d'assurer la formation d'un complexe ARNm-ribosome actif essentiel à la synthèse protéique (Merrick *et al.*, 2004 ; revue par Lopez-Lastra *et al.*, 2005). Un diagramme illustrant les différentes étapes de la formation du complexe d'initiation chez les eucaryotes est présenté **Figure 16**. Pour la majorité des ARNm cellulaires, la traduction est dépendante de la coiffe (cap-dépendant) (Von Der Haar *et al.*, 2004).

Le complexe eIF4F, complexe multimérique composé des facteurs d'initiation eIF4G, eIF4E, eIF4A et eIF4B, assure le recrutement efficace de la sous-unité ribosomique 40S au niveau de l'extrémité 5' des ARNm. Le facteur eIF4E joue un rôle central dans la régulation de la synthèse protéique en se fixant de manière spécifique à la coiffe, mais également en interagissant avec le facteur eIF4G dont la fonction essentielle est le recrutement des facteurs eIF4A et eIF4B (revue par Kapp et Lorsch, 2004). Ces derniers vont assurer l'élimination des structures secondaires de l'ARNm grâce à l'activité ARN hélicase du facteur eIF4B, eIF4A ayant pour fonction principale de stimuler l'activité du facteur eIF4B (Rogers *et al.*, 1999). L'interaction entre les facteurs eIF4E et eIF4G est nécessaire à la stabilisation du complexe eIF4E-coiffe (Von Der Haar *et al.*, 2000). Une fois le complexe eIF4F (eIF4G+eIF4E+eIF4A+eIF4B) formé, la sous-unité 40S du ribosome est recrutée grâce notamment aux facteurs eIF1 et eIF1A (Marcotrigiano *et al.*, 2001). Le facteur eIF1A catalyse également le transfert du complexe ternaire ARNt (Met)_i-eIF2•GTP au sein de la sous-unité 40S (Chaudhuri *et al.*, 1999). Une fois en place, cet énorme complexe, aussi

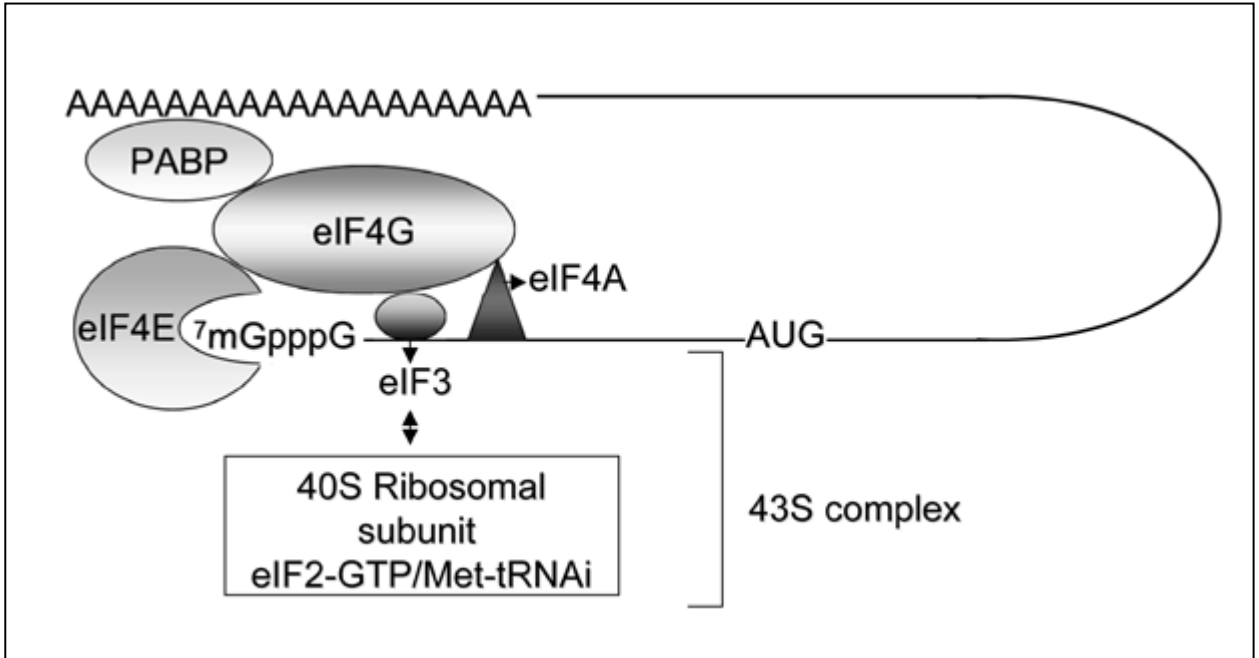


Figure 17. Recrutement de la sous-unité 40S à l'extrémité 5' de l'ARNm (d'après Lopez-Lastra et *al.*, 2005). Le complexe eIF4F interagit avec l'extrémité 5' de l'ARNm (coiffe) via le facteur eIF4E et la queue poly(A) par l'intermédiaire du facteur eIF4G (interaction indirecte via la protéine PABP). Le facteur eIF3 assure le recrutement du complexe de préinitiation 43S à l'extrémité 5' de l'ARNm.

appelé complexe 43S, va glisser jusqu'au codon initiateur AUG en réalisant un véritable balayage ou scanning (**Fig. 17**).

La protéine de liaison à la queue poly(A), PABP (poly(A) Binding Protein), joue également un rôle prépondérant dans la formation du complexe d'initiation eIF4F. En effet, en se liant au facteur eIF4G, elle assure une meilleure affinité du facteur eIF4E pour la coiffe, mais provoque également une circularisation de l'ARNm (**Wells et al., 1998 ; Von Der Haar et al., 2000**). En interagissant avec le facteur eIF4G, la protéine PABP augmente l'efficacité de traduction en stimulant la formation du complexe ribosomique 80S (**Kahvejian et al., 2005**).

III.3.2. Balayage de l'extrémité 5' UTR (ou scanning)

Une fois le complexe 43S formé à l'extrémité 5' de l'ARNm, celui-ci va balayer la région 5' non traduite de l'ARNm jusqu'à trouver le codon initiateur AUG (**Kozak, 2002**). Chez les mammifères et *S. cerevisiae*, la taille moyenne des régions 5' UTR est respectivement de 90 et 50 nt (**Kozak, 1987 ; Cigan et Donahue, 1987**). Le scanning de l'extrémité 5' UTR nécessite la coopération synergique des facteurs eIF1 et eIF1A et leur fixation au niveau du complexe 40S (**Kapp et Lorsch, 2004**). Ceux-ci induisent un changement conformationnel de la sous-unité 40S qui va alors pouvoir non seulement entamer efficacement la phase de scanning, mais également positionner le bon codon initiateur au niveau du site actif (**Passmore et al., 2007**). Lorsque le complexe rencontre le codon initiateur, l'appariement codon anti-codon entre le codon AUG et l'ARNt(Met)_i provoque une pause (**Cigan et al., 1988**). Cette pause permet l'action du facteur eIF5 qui, en interagissant avec eIF3, provoque l'hydrolyse du GTP porté par le facteur eIF2 (**Phan et al., 1998 ; Asano et al., 2000**). L'hydrolyse irréversible du GTP, qui doit avoir lieu précisément au niveau du bon codon initiateur de traduction représente une étape cruciale (**Majumdar et**

Maitra, 2005). Cette hydrolyse entraîne la dissociation de tous les facteurs eIFs ; la sous-unité 60S peut alors se fixer au complexe 40S ARNt(Met)_i.

III.3.3. Reconnaissance du codon AUG

Différents travaux indiquent que l'interaction ARNt (Met)_i –AUG est essentielle, mais pas suffisante pour promouvoir efficacement le démarrage de la traduction. Les facteurs eIF2, eIF5 et eIF1 sont impliqués dans la fidélité de sélection du site d'initiation (**Pestova et Kolupaeva, 2002**). La plupart des ARNm des eucaryotes présentent un contexte qui facilite la reconnaissance de l'AUG par le complexe 43S. Chez les Vertébrés, cette séquence consensus, appelée séquence Kozak, est la suivante : GCCRCCAUGG où R est une purine (**Kozak, 1986 et 1987**). Chez les sporozoaires parasites une autre séquence consensus a été caractérisée : U/ANAAAAUGA. Dans ce consensus, le résidu A en position -3 joue un rôle prépondérant (**Yamauchi et al., 1991**). Un rôle de certains résidus situés en aval de l'AUG, comme par exemple le résidu guanine en position +4 par rapport à l'AUG est également avancé (**Kozak, 1997**). Chez *T. gondii*, l'analyse de plusieurs régions 5' UTR a permis de définir une nouvelle séquence consensus qui diverge de celles définies pour les eucaryotes supérieurs : gNCAAaATGg (**Seeber et al., 1997**).

Récemment, l'hypothèse d'une traduction plus efficace indirectement liée à la présence d'une guanine en position +4, mais directement liée à la nature du second acide aminé de la chaîne polypeptidique a été confortée par une analyse portant sur 34169 séquences protéiques humaines (**Xia, 2007**). Cet auteur propose que le clivage de la méthionine, d'autant plus efficace que le second acide aminé a un faible encombrement stérique, privilégie la présence d'une alanine (GCN) ou d'une glycine (GGN) en seconde position (où N est une base quelconque), et favorise de ce fait la présence d'une guanine en position +4.

III.3.4. 5' UTR réduit et absence de scanning

Les approches de 5' RACE-PCR, ont permis de montrer que chez des protozoaires amitochondriaux appartenant aux genres *Entamoeba*, *Trichomonas* ou *Giardia*, les séquences 5' UTR étaient fortement réduites (**Bruchhaus et al., 1993** ; **Davis-Hayman et al., 2000** ; **Adam, 2000**). Une telle situation est également retrouvée chez les microsporidies (**Fuglsang, 2005**).

Chez *G. lamblia*, la taille des régions 5' UTR varie de 0 à 14 nt (**Adam, 2000**). La recherche par homologie de facteurs d'initiation de traduction a révélé l'absence de certains facteurs tels eIF4B et eIF4H, tous deux impliqués dans l'élimination des structures secondaires dans la région 5' UTR du messager, mais également du facteur eIF4G, qui joue un rôle central dans la formation du complexe de pré-initiation et l'étape de balayage de la région 5' UTR (**Dever, 1999** ; **Prévot et al., 2003**). Sur la base de ces constatations, un modèle basé sur le recrutement "cap-dépendant" des facteurs généraux d'initiation de traduction, et ceci à proximité du codon initiateur AUG, a été proposé (**Li et Wang, 2004**). Ces auteurs ont ainsi montré que le premier codon AUG en aval de la coiffe est le seul codon initiateur, et que l'efficacité de traduction est maximale lorsque qu'un seul nucléotide sépare le codon AUG de la coiffe. Ainsi, bien que dépendant de la coiffe, le mécanisme de traduction ne nécessite pas l'étape de scanning nécessaire à la reconnaissance du bon codon d'initiation (**Li et Wang, 2004**).

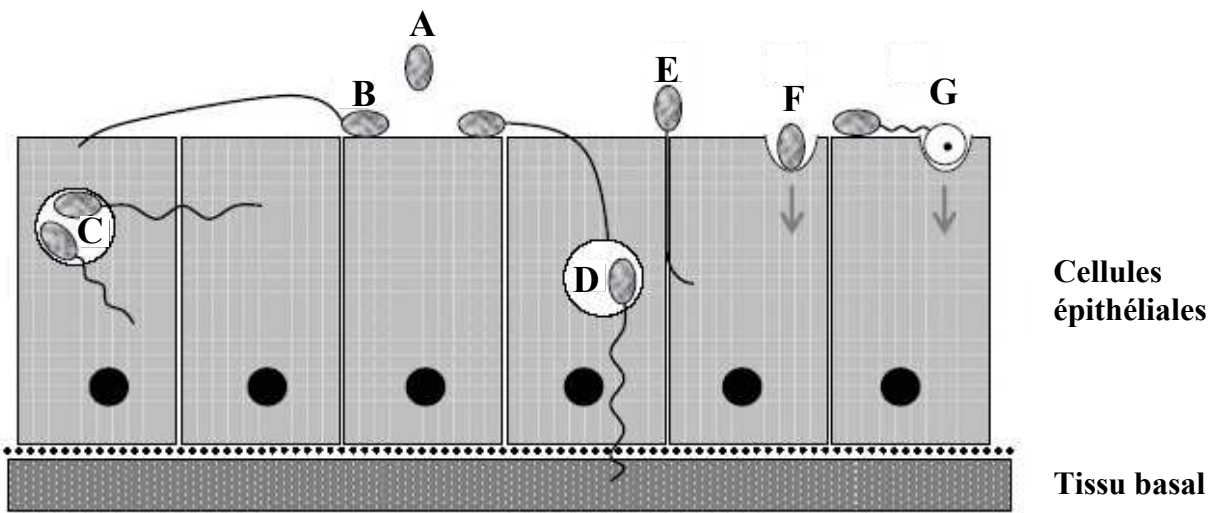


Figure 18. Les différentes voies hypothétiques d'entrée d'*E. cuniculi* dans les cellules épithéliales. (A) Spore mature libre. (B) Adhésion de la spore à la surface de la cellule hôte suivie de la dévagination du tube polaire qui traverse la membrane de la cellule hôte et libère le sporoplasme dans le cytoplasme de la cellule cible. (C) Infestation secondaire de la même cellule ou d'une cellule adjacente par dévagination du tube polaire de spores encore contenues dans une vacuole parasitophore. (D) Infestation secondaire d'une cellule sous jacente voire d'autre tissu. (E) Pénétration basolatérale par insinuation des tubes polaires dévaginés au niveau de jonctions intercellulaires. F et G. Internalisation par endocytose directe de la spore (F) ou du sporoplasme (G) (modifié d'après Vivarès et Méténier, 2001 et Franzen, 2005).

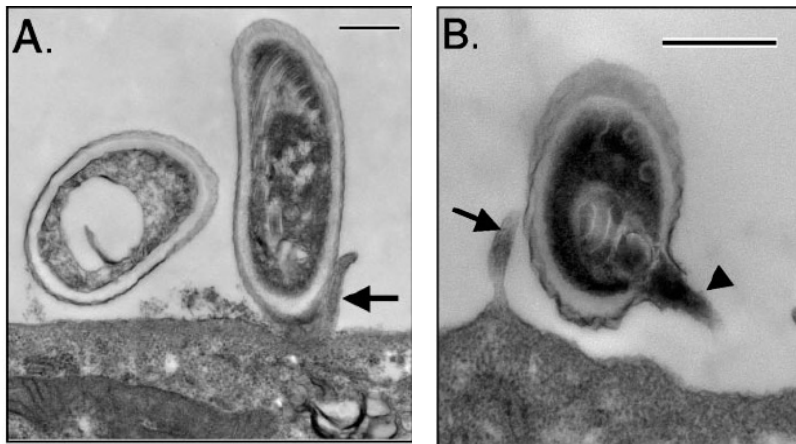


Figure 19. L'attachement des spores à la surface des cellules hôte pourrait être le précurseur de la dévagination. Images en microscopie électronique à transmission de spores d'*E. intestinalis* attachées à la surface de cellules Caco-2. Les spores attachées semblent être en contact direct avec la surface de la cellule hôte ou des microvillosités : ces spores présentent un tube polaire intact non dévaginé, encore enroulé à l'intérieur de la spore (A), d'autres présentent un tube polaire dévaginé (B). Les flèches montrent les microvillosités de la cellule hôte, l'extrémité de la flèche indique le tube polaire dévaginé d'*E.intestinalis*. Echelle : 500 nm (d'après Hayman *et al.*, 2005).

IV/ INTERACTIONS HOTES PARASITES

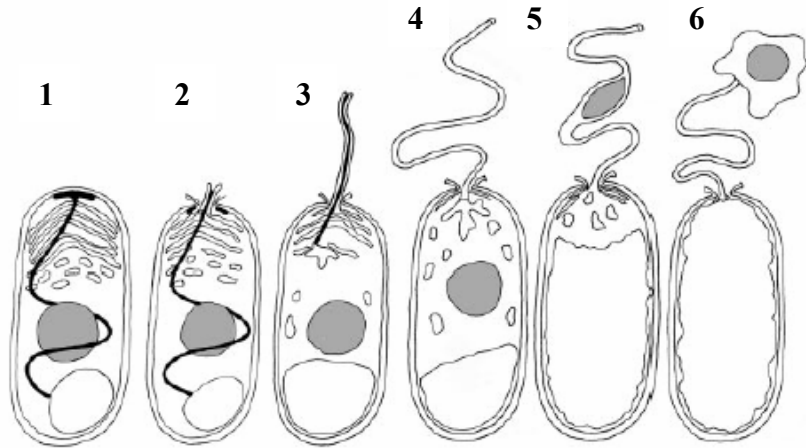
IV.1. Le modèle microsporidien

IV.1.1. Les différents modes d'infestation

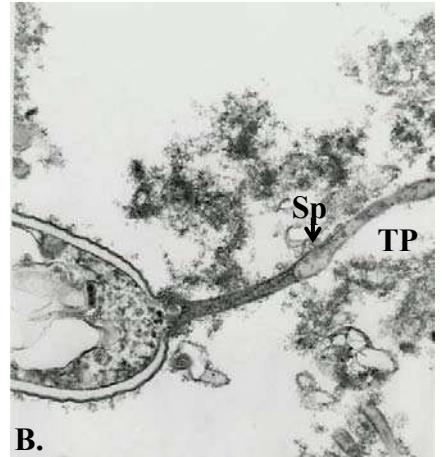
Deux principaux mécanismes sont proposés pour expliquer le mode d'envahissement des cellules hôtes par les microsporidies (**Fig. 18**). Le premier mécanisme est un processus actif au cours duquel le tube polaire semble être capable de percer la membrane de la cellule hôte, agissant ainsi comme l'aiguille d'une seringue. La dévagination du tube polaire permet alors de créer un pont à travers lequel le sporoplasme infectieux va être transféré directement dans le cytoplasme de la cellule hôte. L'adhésion des spores microsporidiennes à la surface des cellules cibles pourrait être l'événement initial nécessaire à la dévagination du tube polaire. Des observations en microscopie électronique montrent en effet que les spores microsporidiennes s'attachent à la membrane de la cellule hôte avant l'invasion (**Fig. 19**). De plus, il a été récemment démontré que des glycosaminoglycanes sulfatés (GAGs) présents à la surface de la cellule hôte jouent un rôle important dans l'attachement des spores d'*E. intestinalis* (**Hayman et al., 2005**). Chez *E. cuniculi* et *E. intestinalis*, il a ainsi été montré qu'une protéine constitutive de la paroi sporale, EnP1, est impliquée dans la reconnaissance spécifique de résidus GAGs à la surface des cellules (**Southern et al., 2007**).

Les molécules présentes à la surface de la spore et impliquées dans l'interaction avec les résidus GAGs cellulaires restent à déterminer.

Le second mécanisme décrit est un mécanisme d'invasion par phagocytose ou endocytose des spores microsporidiennes ou de leur sporoplasme (**Fig. 18F et G**). L'inhibition de l'invasion par la cytochalasine D indique que ce processus est dépendant de l'actine (**Magaud et al., 1997 ; Couzinet et al., 2000 ; Foucault et Drancourt, 2000 ; Takvorian et al., 2005**). Les microsporidies, comme d'autres parasites intracellulaires, sont



A.



B.

Figure 20. Représentation schématique du processus de dévagination d'une spore microsporidienne. Le sporoplasme ou matériel infectieux est expulsé hors de la spore par le tube polaire. Le polaroplasme lamellaire forme la future membrane du sporoplasme injecté dans le cytoplasme de la cellule hôte. **(1)** Spore mature où sont représentés le tube polaire (noir), le noyau (gris), le polaroplasme et la vacuole postérieure. **(2)** Suite à une augmentation de la pression osmotique se produit un gonflement du polaroplasme et de la vacuole postérieure entraînant la rupture du disque d'ancrage et le début de la dévagination du tube polaire à la manière d'un doigt de gant inversé. **(3)** La dévagination du tube polaire se poursuit. **(4)** Lorsque le tube polaire est complètement dévaginé le sporoplasme est poussé vers le tube polaire puis il est transféré à l'intérieur de celui-ci **(5)**. **(6)** Le sporoplasme sort du tube polaire et se retrouve entouré d'une membrane ayant comme origine le polaroplasme lamellaire (d'après Keeling et Fast, 2002). **(B)** Image du passage du sporoplasme infectieux à travers le tube polaire dévaginé observé en microscopie électronique à transmission. Sp : sporoplasme, TP : tube polaire (cliché F. Delbac).

donc capables d'entraîner des remaniements du cytosquelette de la cellule hôte pour faciliter leur entrée.

Des études récentes semblent montrer que l'infestation par phagocytose des spores est 10 fois plus fréquente que l'infestation par injection du sporoplasme *via* le tube polaire (**Franzen et al., 2005**). Cependant **Leitch et al. (2005)** ont montré que la phagocytose n'est pas le mode d'infestation privilégié dans les cellules épithéliales intestinales. De tels résultats laissent supposer que le mode d'infestation serait dépendant du type cellulaire infesté.

Par ce processus d'infestation, alors que la majorité des spores phagocytées sont dégradées, certaines « dévagent » leur tube polaire avant la fusion lysosomale et injectent leur sporoplasme au niveau du cytoplasme échappant ainsi aux phagosomes. Dans certains cas, la dévagination du tube polaire peut également permettre le passage du sporoplasme vers une autre cellule. Une fois dans le cytoplasme, elles peuvent alors initier leur cycle de multiplication intracellulaire. Des infestations secondaires peuvent ensuite avoir lieu lorsque des spores matures, produites dans une première cellule infestée, germent pour infester la même cellule hôte ou une cellule adjacente (**Fig. 18C et D**). Le tube polaire serait alors un moyen d'échapper à la lyse après la phagocytose (**Franzen et al., 2005**).

IV.1.2. Germination sporale et internalisation du parasite

Le processus d'extrusion du tube polaire et le transfert du matériel génétique sont réalisés en moins de 2 secondes (**Frixione et al., 1992**) ; ils peuvent être divisés en 4 étapes : (1) l'activation par divers stimuli appropriés, (2) l'augmentation de la pression osmotique intrasporale, (3) la dévagination du tube polaire par « inversion » et (4) la libération du sporoplasme au travers de ce tube creux (**Fig. 20**).

Encore hypothétiques, les stimuli déclencheurs de la germination ne sont pas universels et diffèrent selon les espèces microsporidiennes. Ceci reflète probablement une

adaptation des microsporidies à leur hôte et aux conditions environnementales, ces parasites étant capables d'infester un très grand nombre de Vertébrés, d'Invertébrés ainsi que différents tissus et types cellulaires. Ainsi, l'existence de stimuli spécifiques nécessaires à l'initiation de la germination permettrait d'éviter des libérations accidentelles du sporoplasme dans l'environnement et pourrait contribuer à la spécificité d'hôte (**Xu et Weiss, 2005**).

In vitro, selon l'espèce de microsporidie considérée, différentes conditions semblent être à l'origine d'une activation de la germination : les variations de pH, l'addition de cations monovalents (Na^+ , K^+) ou d'anions, la présence de peroxyde d'hydrogène ou de faibles doses de radiations UV (**Undeen et Avery, 1988 ; Undeen et Vander Meer, 1990 ; Leitch et al., 1993 ; Frixione et al., 1994**). Un rôle du calcium, séquestré au niveau du polaroplaste, est suggéré car l'addition du ionophore de calcium A23187 stimule la germination (**Weidner et Byrd, 1982**). De plus, **Leitch et al. (1993)** ont montré que les ions Ca^{2+} et leur interaction avec le cytosquelette influencent la dévagination : des traitements à la nifédipine, agent bloquant des canaux calciques, ou à la cytochalasine D qui détruit les filaments d'actine inhibent la germination d'*E. hellem*.

Quels que soient les stimuli nécessaires à l'activation de la germination, leur conséquence est une augmentation de la pression osmotique intrasporale (**Lom et Vavra, 1963 ; Weidner et Byrd, 1982**). Cette dernière est responsable du gonflement du polaroplaste et de la vacuole postérieure produisant ainsi la force nécessaire à la rupture de la paroi au niveau du capuchon polaire, région où son épaisseur est réduite, et permettant alors l'extrusion du tube polaire et la libération du sporoplasme (**Fig. 20**). Une des causes de l'augmentation de la pression osmotique pourrait être l'hydrolyse en deux molécules de glucose du tréhalose, disaccharide non réduit et forme majoritaire de stockage glucidique de la spore. En effet, simultanément à la germination, une forte diminution du taux de tréhalose est corrélée avec une augmentation stoechiométrique de la quantité de sucres réducteurs

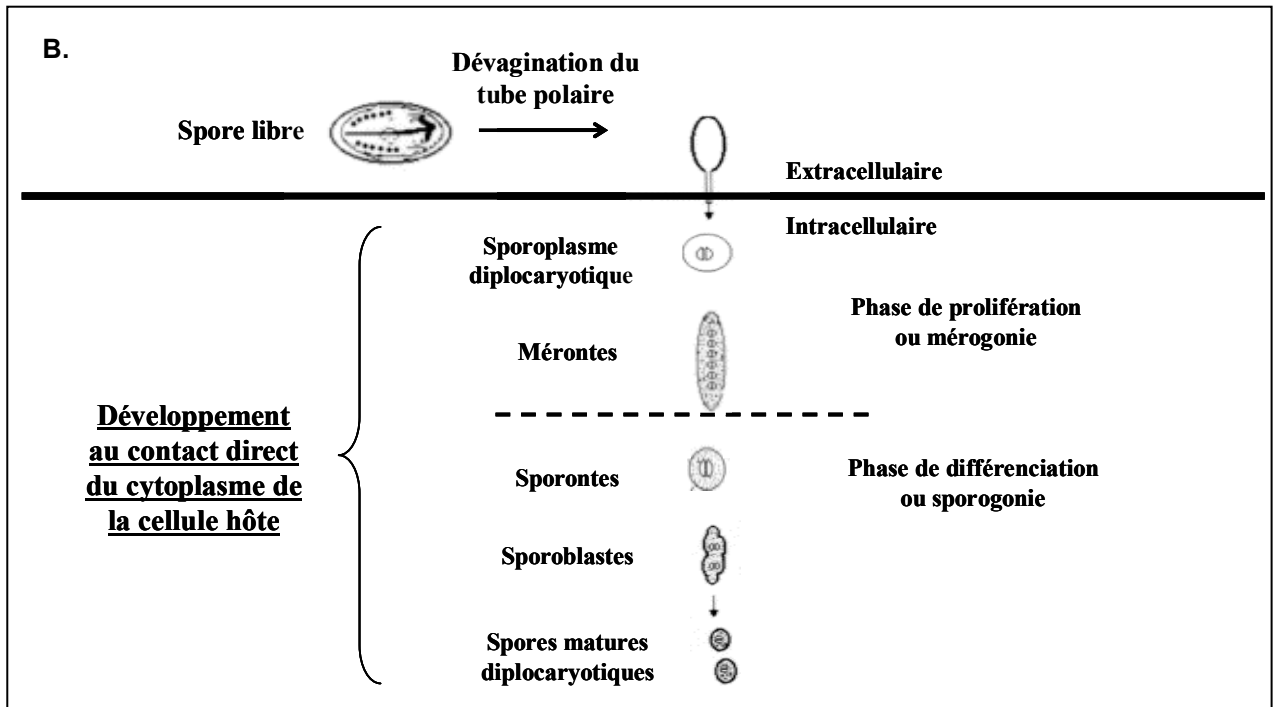
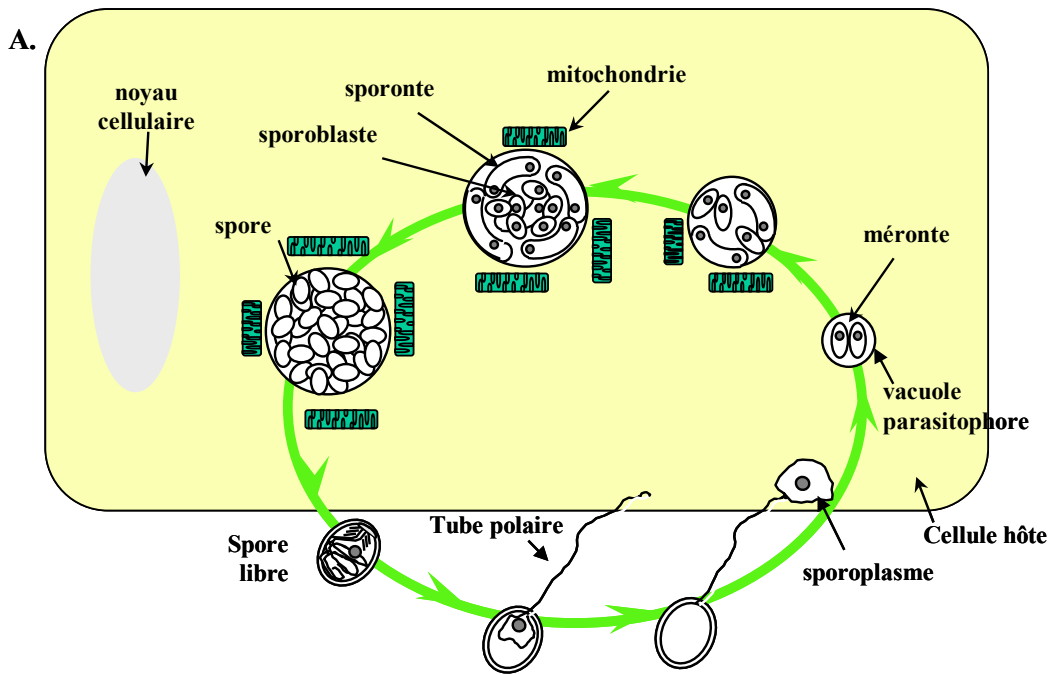


Figure 21. Cycle de développement des microsporidies. (A) Cycle de développement du genre *Encephalitozoon*. La spore extracellulaire libère son sporoplasme infectieux dans le cytoplasme de la cellule hôte *via* le tube polaire « dévaginé » pour initier un nouveau cycle. Ce cycle se déroule au sein d'une vacuole parasitophore à proximité des mitochondries de l'hôte. **(B)** Cycle de développement des microsporidies des genres *Brachiola* et *Antonospora*. Après extrusion du tube polaire, le sporoplasme se développe directement au contact du cytoplasme sans vacuole parasitophore. Les mérontes se divisent par fissions binaires. Les sporontes subissent une division binaire pour donner deux sporoblastes (modifié d'après Weiss, 2001).

(Undeen et Van der Meer, 1994). De plus, des canaux transmembranaires de type aquaporine pourraient jouer un rôle important en facilitant l'entrée rapide d'eau à l'intérieur de la spore (Ghosh *et al.*, 2005).

IV.1.3. Cycle de développement

Il existe une grande diversité de cycles de développement chez les microsporidies. Dans tous les cas, trois grandes phases caractéristiques peuvent être distinguées (**Fig. 21**) : une phase d'invasion, une phase de multiplication intensive ou mérogonie et une phase de différenciation comprenant la sporogonie et la sporogénèse.

✓ la phase **d'invasion** permet le transfert du sporoplasme vers le cytoplasme de la cellule hôte (Lom et Vavra, 1963).

✓ La seconde partie du cycle intracellulaire correspond à une phase de **multiplication** ou **mérogonie**. De forme allongée, les mérontes sont délimités par une membrane plasmique, et comportent, selon l'espèce considérée, un ou deux noyaux. Leur cytoplasme contient de nombreux ribosomes et des structures membranaires assimilées au réticulum endoplasmique et à l'appareil de Golgi. Ils se divisent soit par fission binaire (*Encephalitozoon*, *Nosema*) soit par scissions multiples au sein d'un plasmode végétatif plurinucléé (*Pleistophora*, *Enterocytozoon*) (Cali et Takvorian, 1999). Les microsporidies du genre *Encephalitozoon* réalisent leur cycle de développement au sein d'une vacuole parasitophore dont l'origine semble être la membrane plasmique de la cellule hôte (Magaud *et al.*, 1997 ; Foucault et Drancourt, 2000). Pour les autres genres microsporidiens, le cycle de développement a lieu au contact direct du cytoplasme de la cellule hôte (**Fig. 21**).

✓ La troisième phase est une phase de **différenciation** comprenant la **sporogonie** et la **sporogénèse**. La **sporogonie** commence lorsque les mérontes se transforment en sporontes. Les sporontes sont caractérisés par le dépôt de matériel dense aux électrons sur la membrane

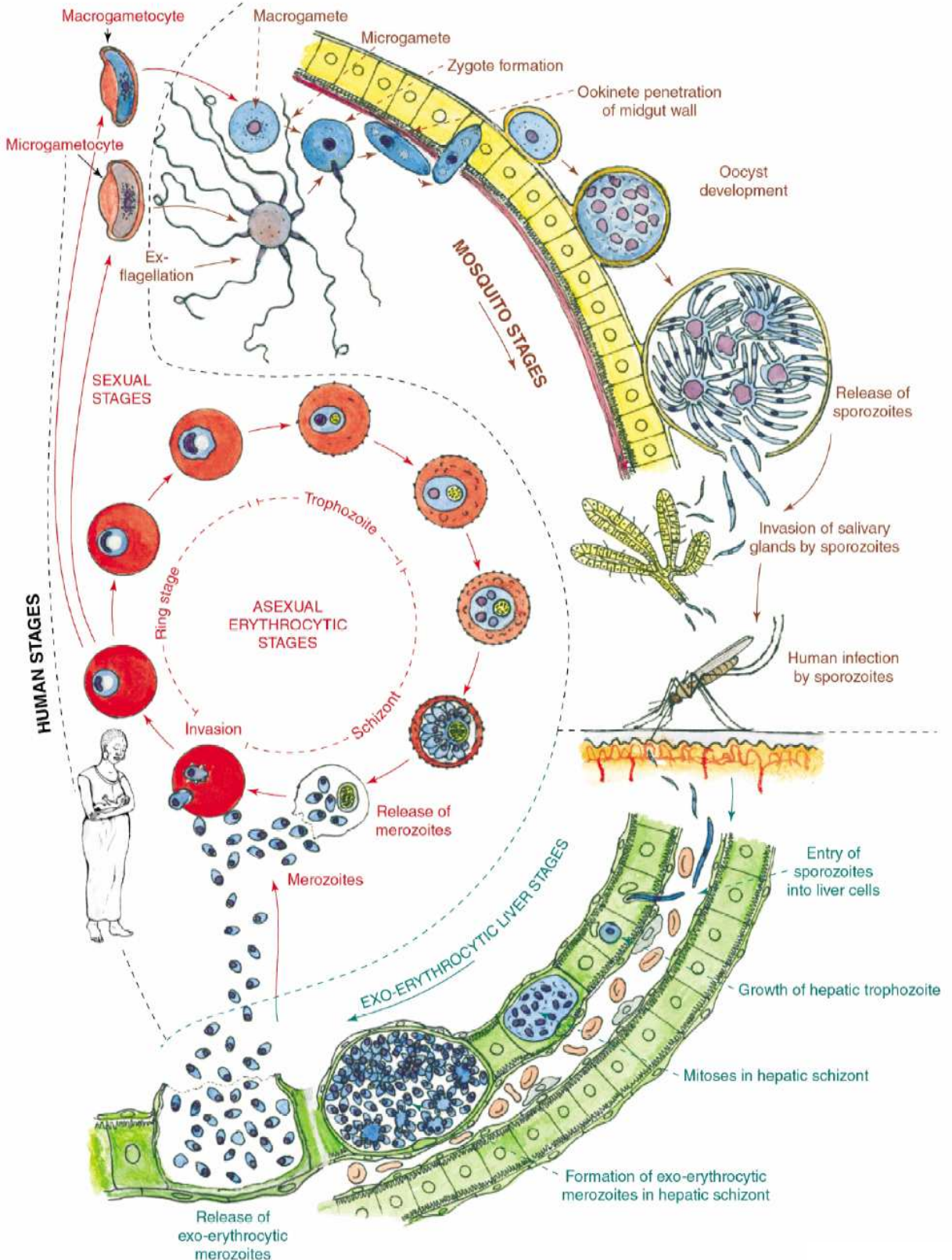


Figure 22. Le développement de *Plasmodium falciparum* dans ses deux hôtes (d'après L. Bannister & G. Mitchell, 2003). Sont ici représentés les principaux stades du développement du parasite dans le foie et les globules rouges chez l'Homme, ainsi que dans le tube digestif et les glandes salivaires chez le moustique.

plasmique évoquant la future exospore. Ils subissent alors soit une division binaire pour donner directement deux sporoblastes (*Encephalitozoon*, *Nosema*), soit deux divisions successives conduisant à quatre sporoblastes (*Septata*), soit une série de divisions multiples à l'origine de l'élargissement de la cellule appelée alors plasmode sporogonial (*Pleistophora*). A partir de cette cellule multinucléée, plus de 100 spores pourront être produites.

La **sporogénèse**, maturation des sporoblastes en spores, est associée à la formation de la paroi sporale et à la différenciation de l'appareil invasif (tube polaire, capuchon polaire, polaroplaste et vacuole postérieure). Les sporoblastes diminuent en taille. Leur cytoplasme devient plus dense aux électrons et plus complexe en raison d'une production importante de ribosomes, de l'élargissement de l'appareil de Golgi mais aussi du fait de la présence de nombreux corps denses aux électrons impliqués dans la construction du tube polaire. Après maturation des sporoblastes en spores, celles-ci sont libérées dans l'environnement par éclatement des cellules hôtes.

IV.2. *Plasmodium falciparum*

IV.2.1. Cycle de développement de *P. falciparum*

Les plasmodiums ont un cycle de développement relativement complexe. Il fait intervenir des stades parasitaires différents où les plasmodies changent d'aspect, de taille, et où l'on observe une variabilité antigénique très importante (**Ghosh et al., 2000 ; Bannister et Mitchell, 2003**).

P. falciparum nécessite deux hôtes successifs pour se développer : l'Homme, qui lui sert d'hôte intermédiaire, et un moustique, qui lui sert de vecteur (hôte définitif). A l'intérieur de ses deux hôtes consécutifs, le parasite suit un développement complexe qui comporte deux étapes essentielles : une phase de multiplication asexuée qui se déroule chez l'Homme, et une phase de multiplication sexuée chez le moustique (**Fig. 22**). Au cours de son repas sanguin, le

moustique infecté injecte dans la circulation sanguine de l'Homme le parasite (sous forme de sporozoïtes) où il reste de quelques minutes à une demi-heure avant de disparaître. Beaucoup d'entre eux sont phagocytés par des macrophages, mais quelques uns envahissent les hépatocytes où ils subiront la première phase de leur développement : la **schizogonie pré-érythrocytaire** ou hépatocytaire. Les sporozoïtes deviennent alors des trophozoïtes hépatiques qui se multiplient intensément dans une même vacuole parasitophore jusqu'à devenir un schizonte hépatique multinucléé. Une semaine après la piqûre de l'anophèle, les schizontes éclatent et libèrent dans la circulation sanguine des milliers de mérozoïtes. Cette phase hépatique est asymptomatique, on parle alors d'incubation.

Intervient ensuite la deuxième phase du développement parasitaire, quand les mérozoïtes pénètrent dans des globules rouges : la **schizogonie érythrocytaire**. Cette étape est, comme l'étape précédente, une étape de prolifération intense. Les mérozoïtes sont alors au stade anneau ; en se multipliant, ils deviennent des trophozoïtes puis des schizontes érythrocytaires. Après 2 ou 3 jours, le globule rouge éclate libérant ainsi 8 ou 16 nouveaux mérozoïtes qui pourront alors infecter de nouveaux globules rouges. Ce sont ces cycles de multiplication dans le sang, et plus particulièrement l'explosion des globules rouges, qui provoquent les fièvres périodiques caractéristiques du paludisme ; plus le nombre de parasites augmente, plus la personne infectée est malade.

Néanmoins, il existe une alternative pour certains mérozoïtes : l'entrée dans la **gamétocytogenèse**. Les parasites se transforment en gamétocytes mâles (microgamètes) ou femelles (macrogamètes) qui ne continueront leur différenciation que s'ils sont « absorbés » par un moustique lors de son repas sanguin. Après fécondation dans l'intestin du moustique, le zygote ainsi formé devient en quelques heures un ookinète mobile. Il entamera ensuite la dernière phase de son développement : la **sporogonie**. Dans la membrane stomacale, le parasite grandit rapidement, formant un oocyste sphérique. Au bout d'une semaine, selon la

température ambiante, l'oocyste commence un processus de divisions internes, avec formation de sporozoïtes en forme de petits vers. L'oocyste explose 14 jours après le début de ce processus de division. Les sporozoïtes ainsi libérés gagnent ensuite les glandes salivaires du moustique. A cours d'un prochain repas sanguin, ceux-ci seront injectés à une prochaine victime.

Pour résumer, *Plasmodium* subit plusieurs phases de prolifération intense aussi bien chez le moustique lors de la sporogonie que chez l'Homme, pendant la schizogonie exoérythrocytaire et la schizogonie érythrocytaire, cette dernière phase étant la phase responsable de la maladie. Il passe aussi par une phase de différenciation cellulaire, associée à un arrêt de la prolifération : la gamétocytogenèse, qui, elle, est responsable de la dissémination de la maladie.

IV.2.2. Mécanisme d'adaptation parasitaire

Au cours, de son développement chez *An. gambiae*, *P. falciparum* traverse deux membranes épithéliales (tubes digestifs et glandes salivaires) au niveau desquelles ont été caractérisées des modifications structurales et biochimiques (**Whitten et al., 2006 ; Michel et Kafatos, 2005**). Les mécanismes d'adaptation mis en place par le parasite chez le moustique se définissent ainsi par l'expression de gènes lui assurant : 1) "l'échappement" face aux mécanismes de défense, 2) les ressources nécessaires à l'accomplissement de ses phases de multiplication et de différenciation, 3) le contrôle d'expression de certains gènes du moustique afin d'assurer l'intégralité de son cycle et enfin, 4) son interaction avec les deux membranes épithéliales en question.

Des expériences de knock-out ont permis de caractériser des protéines parasitaires indispensables à l'invasion de l'épithélium intestinal. Un certain nombre d'entre elles sont synthétisées au niveau des structures sécrétoires que sont les micronèmes, d'autres,

correspondent à des protéines de surface dont le rôle apparaît tout aussi essentiel dans le mécanisme d'invasion (revue par **Barillas-Murry et Kumar, 2005**).

Dessens et al (1999) ont ainsi montré que l'absence d'expression au stade ookinète des protéines CS (Circumsporozoite), TRAP (Thrombospondin-Related Adhesive Protein) et CTRP (Circumsporozoite and TRAP-Related Protein) réduit fortement la mobilité de l'ookinète et inhibe complètement sa capacité à envahir le tissu intestinal. Les protéines solubles WARP (von Willebrand factor A domain-Related Protein) et SOAP (Secreted Ookinete Adhesive Protein) sont des protéines riches en résidus cystéine. Synthétisées au niveau des micronèmes, celles-ci sont capables de former des complexes macromoléculaires *via* la mise en place de ponts disulfures (**Yuda et al., 2001 ; Dessens et al., 2003**). Chez *P. berghei*, l'inactivation par transgénèse du gène codant la protéine SOAP se traduit par une incapacité des parasites à traverser l'épithélium intestinal (**Dessens et al., 2003**).

Kadota et al. (2004) ont étudié le rôle d'une protéine synthétisée au niveau des micronèmes, la protéine MAOP (Membrane-Attack Ookinète Protein). La mutation du gène codant cette protéine a permis de mettre en évidence son rôle dans la perforation de la membrane intestinale.

Le passage de l'ookinète à travers l'épithélium intestinal implique la reconnaissance de récepteurs ou protéines associés à cette membrane. A partir d'un extrait protéique intestinal d'*A. aegypti*, **Dinglasan et al.** (2003) ont réussi à produire des anticorps monoclonaux. L'un d'eux, l'anticorps MG96 reconnaît chez *Anopheles sp.* plusieurs protéines glycosylées exprimées spécifiquement à la surface des cellules intestinales. Associé à un repas sanguin infesté par *Plasmodium*, cet anticorps est capable d'inhiber complètement la formation d'oocystes chez le moustique, révélant ainsi le rôle joué par ces protéines dans l'interaction hôte-parasite.

Des travaux menés par **Ghosh et al.** (2001) avaient permis d'identifier un peptide de 12 acides aminés, SM1 (pour Salivary and Midgut peptide 1) ayant la capacité de se fixer de manière spécifique au niveau du lobe distal des glandes salivaires ainsi qu'au niveau de la membrane intestinal, lieux privilégiés d'interaction avec le parasite. L'expression de ce peptide au niveau intestinal chez un moustique transgénique (*An. stephensi*) inhibe également le développement parasitaire (**Ito et al., 2002**).

La phospholipase A2 (PLA2), une protéine contenue dans le venin de serpents ou d'abeilles, présente des propriétés intrinsèques très intéressantes : insertion stable dans la double couche phospholipidique, faible poids moléculaire et stabilité même sous une forme inactive. En raison de ces propriétés, **Zieler et al.** (2001) ont cherché à déterminer si cette enzyme pouvait, en se liant à l'épithélium intestinal du moustique, bloquer le cycle parasitaire. Leurs travaux, portant sur la capacité d'invasion de *P. gallinaceum* et *P. falciparum* ont permis de caractériser une diminution significative du développement des oocystes en présence de phospholipase A2. L'expression constitutive de cette enzyme chez un moustique transgénique (*An. stephensi*), permet, comme précédemment, de limiter fortement le développement des oocystes (**Moreira et al., 2002**). Enfin, chez *An. aegypti*, l'inoculation d'anticorps dirigés contre la protéine aaSGS1 (*Aedes aegypti* Salivary Gland Surface protein), une protéine spécifiquement exprimée au niveau des glandes salivaires, a révélé l'implication de cette protéine dans le processus d'invasion (**Korochkina et al., 2006**).

IV.2.3. Etudes des interactions *An. gambiae* / *P.falciparum*

Grâce aux données de séquençage de plusieurs espèces d'insectes, l'annotation du génome d'*An. gambiae* a été grandement facilitée, permettant de caractériser l'existence d'au moins 14 000 gènes (**Holt et al., 2002**). Couplée au séquençage du génome de *P. falciparum*, la connaissance du répertoire génique d'*An. gambiae* permet d'initier des études exhaustives,

rendant possibles la compréhension des interactions hôte-parasite et la découverte de nouvelles cibles thérapeutiques. En effet, parmi les relations entre les 3 acteurs de la maladie parasitaire, la plus étudiée est l'interaction entre *Plasmodium* et l'être humain infecté. Or, cette recherche ne s'est pas encore traduite par des avancées décisives dans la mise au point d'un vaccin, même si plusieurs protéines parasitaires capables de conférer une immunité à long terme semblent prometteuses (**Schofield et al., 2002 ; Perlaza et al., 2003**), tout comme la production chez l'Homme d'anticorps capables de bloquer le développement parasitaire chez le moustique (transmission blocking vaccines), (**Lavazec et al., 2007**). En revanche, les interactions entre le parasite et l'insecte sont restées peu explorées. La compréhension des mécanismes d'interaction entre le pathogène (*P. falciparum*) et son vecteur (*An. gambiae*) permettrait d'identifier des gènes cibles indispensables au cycle du parasite.

Ainsi, plusieurs équipes de par le monde ont concentré leurs recherches sur l'étude des interactions *An. gambiae* / *P. falciparum*, impliquant principalement des analyses de type transcriptomique (biopuce à ADNc, banques soustractives, differential display) axées sur l'identification de gènes du moustique dérégulés au cours de l'invasion, ou impliqués dans la réponse immunitaire après stimulation antigénique (**Dimopoulos et al., 2000 ; Oduol et al., 2000 ; Bonnet et al., 2001 ; Christophides et al., 2002 ; Osta et al., 2004 ; Chen et al., 2004 ; Abraham et al., 2004 ; Srinivasan et al., 2004 ; Xu et al., 2005 ; Vlachou et al., 2005 ; Dong et al., 2006**).

L'avènement du séquençage des génomes d'*An. gambiae* et *P. berghei* a permis également d'initier, par une approche biopuce à ADN, l'étude du profil global d'expression des gènes du moustique et du parasite au cours de l'invasion (**Xu et al., 2005**). D'autres études ont quant à elles permis l'identification de protéines du moustique dérégulées au cours de l'invasion grâce à la conjugaison d'analyses protéiques bidimensionnelles et des technologies de spectrométrie de masse. La mise au point de biopuces à ADN exhaustives

couvrant la totalité des gènes d'*Anopheles* par les firmes Affymetrix et Agilent, constitue une nouvelle avancée dans l'étude complète des interactions *Anopheles* / *Plasmodium*.

Par ces approches, un grand nombre de gènes intervenant dans le développement parasitaire ont été identifiés, et l'implication de certains de ces gènes dans le développement du parasite validée grâce à leur inactivation par ARNi. Ainsi, chez *An. gambiae*, l'inactivation des gènes codant les protéines TEP1 (Thioester-containing Protein 1) et LRIM1 (Leucin rich-repeat Immune gene 1) favorise le développement parasitaire (**Blandin et al., 2004 ; Osta et al., 2004**). A l'inverse, l'inactivation des gènes codant les protéines CTL4 (C-Type Lectin 4) et CTLMA2 (CTL Mannose binding 2) se traduit par une diminution des capacités d'échappement parasitaire, 54 à 97% des ookinètes étant détruits par mélanisation (**Osta et al., 2004**). Chez *An. gambiae*, la démonstration de l'efficacité du processus d'interférence par ARN au niveau des glandes salivaires ouvre des perspectives d'études des gènes impliqués dans la maturation et la transmission des parasites (**Boisson et al., 2006**). L'inactivation spécifique d'un gène nécessite néanmoins l'injection de 5 à 30 fois plus d'ARNdb dans l'hémolymphe que pour une inactivation au niveau du tube digestif, probablement en raison de la très faible perméabilité cellulaire de ce tissu (**Boisson et al., 2006**).

En fonction de l'espèce parasitaire considérée (*P. falciparum*/*P. berghei*), l'extinction de certains gènes d'*An. gambiae* impliqués dans la réponse immunitaire a un impact différent sur le déroulement du cycle parasitaire (**Dong et al., 2006**).

La mise au point d'un système de transgénèse stable capable d'assurer l'expression d'un ARNdb et donc, l'inactivation d'un gène impliqué dans le développement de *Plasmodium*, représente une stratégie de choix dans l'interruption du cycle parasitaire chez le moustique. *In fine*, cette approche s'appuie sur la création de lignées transgéniques stables et réfractaires à toute transmission parasitaire, mais surtout capable de supplanter les populations de moustiques autochtones. Très prometteuse, cette stratégie a montré l'avantage certain

(fitness) que confère un gène anti-parasitaire chez un moustique transgénique (**Marrelli et al., 2007**).

Tableau 4. Différences majeurs observées entre les deux microsporidies modèles étudiées : *Encephalitozoon cuniculi* et *Brachiola algerae*.

Microsporidie	<i>E.cuniculi</i>	<i>B. algerae</i>
Taille spore (µm)	2-2,5 x 1-1,5	2,3-3,9 x 3,6-5,4
Tours de spires du tube polaire	5 à 6	8 à 11
Noyau(x)	1 (monocaryon)	2 (diplocaryon)
Interface hôte-parasite	vacuole parasitophore	contact du cytoplasme
Taille du génome nucléaire	2,9 Mpb	23 Mpb
Hôtes majeurs	mammifères	diptères (Anophèles)

V/ LE MODELE D'ETUDE *BRACHIOLA ALGERAE* / *ANOPHELES GAMBIAE*

V.1. La microsporidie *Brachiola algerae*

V.1.1. Cycle cellulaire

Tout au long des stades de développement intracellulaire, le "noyau diplocaryotique" est maintenu. La multiplication végétative et la différenciation se déroulent directement au contact du cytoplasme de la cellule hôte. Les spores obtenues sont ovoïdes et mesurent approximativement $4,4 \times 2,8 \mu\text{m}$. Il est intéressant de noter les différences de développement par rapport à la première microsporidie séquencée, *E. cuniculi*. En effet, cette dernière possède un noyau monocaryotique, son cycle de développement se déroule à l'intérieur d'une vacuole parasitophore et les spores sont de plus petite taille ($2,5 \times 1,5 \mu\text{m}$). Leurs principales différences sont résumées dans le **Tableau 4**. Sur la base d'analyses ultrastructurales et moléculaires, *B. algerae* a été récemment transféré dans le genre *Anncaliia* (**Franzen et al., 2006**).

V.1.2. Capacité d'adaptation

La majorité des espèces microsporidiennes se caractérisent à la fois par un spectre d'hôtes relativement étroit mais également par une spécificité tissulaire, exception faite pour certaines espèces telles qu'*E. cuniculi*, *A. locustae* ou encore *B. algerae*. Parasite d'insectes, cette dernière espèce montre en effet des capacités d'infestation surprenantes. Les capacités de développement des espèces microsporidiennes sont fonction de la température de l'organisme hôte. A ce titre, deux types de pathogènes peuvent être distingués:

- ✓ Les pathogènes d'organismes poïkilothermes, individus dont la température corporelle varie (insectes, poissons).
- ✓ Les pathogènes d'organismes homéothermes, individus contrôlant leur température interne (mammifères).

Ainsi, la température de développement des microsporidies d'insectes a toujours été considérée comme un facteur limitant la transmission vers les mammifères. Cependant, l'injection de spores de *B. algerae* au niveau de la queue et des coussinets d'une souris athymique a permis de mettre en évidence un développement important de cette espèce au point d'inoculation, laissant entrevoir des capacités d'adaptation importante à une température proche de 37°C (**Alger et Undeen, 1980 ; Trammer et al., 1997**). La recherche des températures permissives au développement de *B. algerae* en culture cellulaire a montré que cette microsporidie était capable de se multiplier dans une large gamme de températures allant de 24 à 38°C, et dans des types cellulaires variés (**Trammer et al., 1999 ; Lowman et al., 2000 ; Franzen et al., 2006**).

Récemment, cette microsporidie a été identifiée au niveau de la cornée d'un patient immunocompétent (**Visvesvara et al., 1999**). Cet isolat humain a également été utilisé pour infecter des souris immunodéficientes (**Koudela et al., 2001**). Après l'application oculaire, le développement parasitaire entraîne une infection sévère du foie de la souris sans créer de lésion au niveau de l'oeil. La découverte, chez une personne immunocompétente, de lésions oculaires dues à *B. algerae* (**Visvesvara et al., 1999**), et l'implication de cette microsporidie dans le décès d'une personne immunodéprimée (**Coyle et al., 2004**), ont permis d'affirmer le caractère opportuniste de cette espèce. Celle-ci possède des potentialités d'adaptation importantes, que ce soit au niveau de son spectre d'hôte (insectes, mammifères), des types cellulaires infestés (muscle, rein, cornée, intestin) ou encore de sa capacité à se développer dans une large gamme de températures.

Enfin, des études ont montré qu'une infection par *B. algerae* du moustique parasité par *P. falciparum* entraînait une réduction du développement de l'agent du paludisme, mettant ainsi en lumière un contrôle biologique éventuel (**Margos et al., 1992**).

V.1.3. Réduction de la compétence vectorielle des moustiques

Fox et Weiser (1959) ont réalisé les premiers travaux montrant qu'une infection par *B. algerae* de moustiques *An. gambiae* entraînait une réduction du développement et de la transmission de *P. falciparum*. Cependant, ces auteurs ont imputé la diminution de la compétence vectorielle du moustique à une destruction du tissu intestinal par *B. algerae*.

Par la suite, d'autres travaux ont permis l'hypothèse d'une utilisation possible de cette microsporidie dans la lutte biologique, en raison de sa capacité à diminuer la transmission parasitaire. En effet, cette microsporidie est capable d'augmenter la mortalité des moustiques durant leur cycle de développement (larvicide), en diminuant la fertilité et la longévité des femelles, mais également en perturbant le cycle parasitaire (**Anthony et al., 1972 ; Savages et al., 1972 ; Gajanana et al., 1979 ; Geetha et al., 1979**). Plus récemment, **Margos et al. (1992)** ont montré qu'une infestation simultanée d'*An. stephensi* par *P. falciparum* et *B. algerae* était à l'origine d'une réduction moyenne de 69% du nombre d'oocystes. De la même manière, l'impact sur le développement de *P. yoleii* a mis en évidence une réduction oocystique de 85% ainsi qu'une diminution du pouvoir invasif des sporozoïtes (**Schenker et al., 1992**).

Dans un contexte de séquençage des génomes de *P. falciparum* et d'*An. gambiae*, il apparaît ainsi judicieux de porter également notre attention sur les capacités d'adaptation de *B. algerae*, notamment dans le cadre d'études de post-génomiques visant à améliorer nos connaissances sur les mécanismes d'interactions parasitaires au sein de ce modèle tripartite.

V.2. *Anopheles gambiae*

V.2.1. Le genre *Anopheles*

Toutes les espèces vectrices de *Plasmodium* appartiennent au genre *Anopheles*. La classification des espèces d'*Anopheles* a débuté en 1818, lorsque Meigen en a réalisé les

premières descriptions. Depuis, plus de 420 espèces d'*Anopheles* ont été identifiées, et 68 sont responsables de la transmission des parasites à l'Homme. Seule la femelle, hématophage, assure la transmission de la maladie.

V.2.2. L'immunité chez les insectes

L'étude des mécanismes de défense chez les insectes a révélé l'existence d'un système très efficace basée exclusivement sur une réponse immunitaire innée (revue par **Lemaître et Hoffman, 2007**). Elle repose sur la mise en place d'une réponse globale, présentant une spécificité relative vis-à-vis d'un certain nombre de pathogènes (bactéries Gram positif ou négatif, champignons, parasites), mais jamais adaptative (**Hoffmann et al., 1995 et 1996 ; 1996 ; Hultmark, 2003**). Phylogénétiquement la plus ancienne, car retrouvée vraisemblablement chez tous les métazoaires, l'immunité chez les insectes est fondée sur des réactions cellulaires et humorales. Celle-ci s'appuie sur l'existence de récepteurs cellulaires capables de reconnaître des éléments étrangers ainsi que la synthèse de peptides antimicrobiens (**Janeway, 1989 ; Hoffmann et Reichhart 2002**).

V.2.2.1. L'immunité cellulaire

Elle est assurée par les hémocytes, cellules présentes au niveau de l'hémolymphe, dont la fonction est d'assurer l'encapsulation ou la phagocytose des particules étrangères (**Meister et al., 2004**). Cette immunité cellulaire participe indirectement à l'activation de l'immunité humorale *via* la synthèse de zymogènes (pro-phénoloxydases), de protéines cytokines-like ou de molécules effectrices.

V.2.2.2. L'immunité humorale

La réponse humorale consiste d'une part, en l'activation de cascades protéolytiques (coagulation et mélanisation), et d'autre part, en la synthèse de peptides antimicrobiens dont

l'expression est sous le contrôle des voies de régulation Toll et Imd (revue par **Lemaître et Hoffmann, 2007**). L'injection de bactéries ou une simple blessure induit chez les insectes les plus évolués la synthèse, entre autre, d'une batterie de peptides antimicrobiens. Ces molécules antibiotiques sont produites principalement par le corps gras, un organe fonctionnellement analogue au foie des Vertébrés, et sont sécrétées dans l'hémolymphe (**Meister et al., 1997 ; Hoffmann et al., 1996**).

Cette réponse immunitaire assurerait une discrimination entre le soi et le non soi infectieux grâce à des récepteurs "PRRs" (Pattern Recognition Receptors), capables de reconnaître des molécules aux motifs structuraux invariants présents à la surface des microorganismes (LPS, mureïne, Beta 1,3 glucane...) mais absents des cellules eucaryotes (**Janeway et al., 2002 ; Gottart et al., 2006**). Lorsque ces récepteurs sont stimulés, ils déclenchent immédiatement des mécanismes effecteurs tels que l'opsonisation, la phagocytose, ou la cascade du complément.

V.3. Immunité et compétence vectorielle

La notion même de l'existence d'un système immunitaire chez le moustique s'opposant au développement du parasite fut avancée par **Huff** en 1927. Ces constatations ont été confirmées par la mise en évidence d'une perte parasitaire au cours des différentes phases de développement chez le moustique. En fonction du couple vecteur-parasite considéré, la capacité vectorielle du moustique est sous la dépendance de facteurs liés à des interactions spécifiques (**Alavi et al., 2003**). Il est néanmoins admis que chez le moustique, il existe une immunité de base définissant le niveau d'interaction possible (**Sinden et al., 2004**). A ce titre, chez les espèces réfractaires à toute transmission, **Schwartz et al. (2002)** expliquent leur faible représentativité au sein des populations totales par le coût physiologique dû à la destruction totale des parasites par le système immunitaire. Le niveau de susceptibilité des

souches et leur capacité à assurer avec plus ou moins d'efficacité le cycle du parasite apparaît être sous la dépendance de nombreux facteurs alléliques mais également environnementaux (Niare et al., 2002 ; Lambrechts et al., 2006).

V.4. Invalidation de gènes candidats

Le concept même d'une altération permanente des compétences vectorielles du moustique, et ceci bien évidemment, en vue de la création de lignées réfractaires à toute transmission parasitaire, fut proposé en 1968 par **Chris Curtis** de l'Ecole de Médecine et Hygiène Tropicale de Londres.

C'est grâce à l'avènement de la biologie moléculaire ainsi qu'aux récents progrès dans le domaine des manipulations génétiques que cette idée novatrice fut envisagée au travers notamment de l'utilisation de virus recombinants (Carlson et al., 1995 ; Olson et al., 1996 ; Higgs et al., 1996) puis d'éléments transposables (pour revue voir Moreira et al., 2002 ; Christophides, 2005).

Ainsi, l'élément P (Rubin et Spradling, 1982), un élément génétique mobile caractérisé chez *D. melanogaster*, a été utilisé pour l'intégration de séquences d'ADN chez deux moustiques vecteurs de maladies tropicales, *Aedes* et *Anopheles* (Miller et al., 1987 ; Mc grane et al., 1988 ; Morris et al., 1989). Le très faible taux d'intégration de cet élément transposable chez des espèces non drosophiliennes a conduit les scientifiques à rechercher des éléments transposables beaucoup plus efficaces afin d'envisager l'intégration stable d'un transgène chez quelques espèces vectrices appartenant aux genres *Aedes* et *Anopheles*. Ainsi, ces recherches se sont traduites par la découverte et l'utilisation de nouveaux éléments tels *Minos* (Franz et al., 1991 ; Catteruccia et al., 2000 ; Lombardo et al., 2005 ; Yoshida et Watanabe, 2006), *Hermes* (O' Brachta et al., 1996 ; Jasinskiene et al., 1998), *Mariner* (Garza et al., 1991 ; Coates et al., 1998), ou encore *PiggyBac* (Fraser et al., 1996 ;

Grossman et al., 2000 ; Ito et al., 2002 ; Nolan et al., 2002 ; Rodriguez et al., 2006) dont l'intégration apparaît "séquence-spécifique" (**Grossman et al., 2001**).

Chez *Anopheles* sp., l'élément *PiggyBac* se présente comme un outil idéal de transformation en raison de sa capacité d'intégration stable au sein du génome (**Grossman et al., 2001 ; Rodriguez et al., 2006**). Il devient alors possible d'envisager, après l'identification de gènes potentiellement impliqués dans le cycle de *Plasmodium*, de mettre en place différentes stratégies visant à interrompre leur expression, ou au contraire, à stimuler l'expression de gènes dont les produits s'opposent au développement du pathogène (**Ito et al., 2002 ; Marrelli et al., 2006**). Au cours du cycle parasitaire, compte tenu de l'étroite interaction des parasites avec certains tissus (épithélium intestinal, salivaires) et afin d'éviter toutes inactivations pléiotropes, des promoteurs tissus-spécifiques et inductibles ont été recherchés en vue d'une inactivation spatio-temporelle ciblée (**Ito et al., 2002**).

MATERIEL ET METHODES

I. ETUDE DES MICROSPORIDIÉS

I.1. Les espèces microsporidiennes étudiées

Deux espèces microsporidiennes ont été étudiées : *Encephalitozoon cuniculi* et *Brachiola algerae*.

I.2. Production et purification des parasites

I.2.1. *Encephalitozoon cuniculi*

E. cuniculi, isolat de souris (GB-M1), fourni par le Pr. E.U. Canning (Imperial College of London) et isolat de référence choisi pour le séquençage du génome, est cultivé *in vitro* sur des cellules épithéliales MDCK (Madin-Darby Canine Kidney) ou des cellules fibroblastiques HFF (Human Foreskin Fibroblast). Les cultures s'effectuent dans une atmosphère enrichie à 5% en CO₂ à 37°C en milieu MEM (Minimum Essential Medium, Life technologies) supplémenté en sérum de veau foetal 5% (Hyclone, Perbio), L-glutamine 2 mM (Invitrogen), fungizone 2,5 µg.ml⁻¹ (Invitrogen) et divers antibiotiques (ampicilline 100 µg.ml⁻¹, streptomycine 100 µg.ml⁻¹, pénicilline 100 U.ml⁻¹). Lorsque les cellules sont à confluence, l'infestation est réalisée avec 10⁸ spores d'*E. cuniculi* par flacon de 25 cm². Les spores libérées par lyse des cellules hôtes, sont collectées tous les deux jours dans le surnageant de culture. Après centrifugation (2 min, 6000 g) et lavages en PBS, le culot de spores est repris dans de l'eau distillée stérile et conservé à 4°C.

I.2.2. *Brachiola algerae*

B. algerae, isolat fourni par le Dr. A. Undeen, est cultivé *in vitro* sur des cellules fibroblastiques HFF (Human Foreskin Fibroblast). Les cultures s'effectuent en milieu L-15 (Leibovitz, Life technologies) supplémenté en sérum de veau foetal 10% et divers antibiotiques (voir précédemment) à 28°C. Lorsque les cellules sont à confluence, l'infestation est réalisée avec 2 x 10⁷ spores de *B. algerae* par flacon de 80 cm². Après 4 à 6 semaines de culture, les spores sont libérées des cellules hôtes par lyse mécanique. Pour cela, l'échantillon biologique est passé au travers d'une seringue munie d'une aiguille 25G. Après centrifugation (2 min, 6000 g) et lavages en PBS, les spores purifiées sont conservées à 4°C pour de nouvelles infestations ou la réalisation de caryotypes moléculaires.

Tableau 5. Séquences des amorces utilisées pour la réalisation des RACE-PCR.
(les amorces précédées du chiffre 3 ou 5 ont été respectivement utilisées pour les 3' et 5' RACE-PCR).

Amorces	Gène	Séquences
3'ubiq	ECU02_0740	5' - GGATACCACCAGACCAGCAGAGGCTG -3'
5'ubiq		5' - CTCGATACTGTCCGAAGGCTCGACT -3'
3'pap	ECU02_0730	5' - GCGTCAAAGTTTCGAGGCGGTTGACG -3'
5'pap		5' - CCCTTACTATGCACCCCCAGTCTG -3'
3'h2a	ECU02_0720	5' - GGACCAGGATGAGGATATCGAAGG -3'
5'h2a		5' - TGATCTCGCTGATTAGGTACATTAC -3'
3'httr	ECU02_0710	5' - GATACCGACCTGGCGCTTGAACACG -3'
5'httr		5' - CGTGCTCAGACGCAGATGCAACAAC -3'
3'tpp	ECU01_0870	5' - AGCATCTATAGGCCTTCATCGACTC -3'
5'tpp		5' - AAGGATCTTGCATGGAGAATAGGCC -3'
3'01_0860	ECU01_0860	5' - AGAGTCTAGAGATGGCGATTTACAT -3'
5'01_0860		5' - GAAGCAGCAGCAGAATCCAGGCGCC -3'
3'10_1410	ECU10_1410	5' - CATTCTATTTCTTCTCGGGCACTCTGGG -3'
5'10_1410		5' - GGATATCAGCTTGTTCTGGATAACGGC -3'
3'reg6	ECU10_1420	5' - GGAAGAGGATCCTGCAAATCCATGC -3'
5'reg6		5' - CGCTGTGCTCGTCAAGGACTTCAAC -3'
3'gcn	ECU10_1430	5' - GGCGTCGAGTGCATTGTAGAGGAGG -3'
5'gcn		5' - GTCATAGTCGACAGCCAGAAACACGATC -3'
3'ysx3	ECU10_1440	5' - GGCCAGACTTTCTGAGCAGATGCAATCC -3'
5'ysx3		5' - GGATTGCATCTGCTCAGAAAGTCTGGCC -3'
3'pred	ECU10_1450	5' - TGGGTATGATGACAAGCCAAGGTCTTC -3'
5'pred		5' - CGAGTCCTCGAGCTGTGAGTACATGC -3'
3'garnt	ECU10_1460	5' - GAAGGGAGACAAGGACTTCTGAGGCTG -3'
5'garnt		5' - CATTCTTCAGCATGCTCTCATGC -3'

I.3. Mise en évidence d'une organisation en opéron chez *E. cuniculi*

I.3.1. Extraction des ARN totaux

Des cellules HFF à confluence sont infestées par 10^8 spores d'*E. cuniculi*. Les ARN totaux de cellules HFF infestées sont extraits 3 jours après infestation avec le kit RNeasy Mini kit (QIAGEN) selon les recommandations du fournisseur. Une analyse qualitative et quantitative des ARN totaux est réalisée sur Bioanalyser Agilent.

I.3.2. RACE-PCR et clonages

Des 5' et 3' RACE-PCR ont été réalisées afin d'amplifier chaque CDS d'un opéron potentiel à l'aide du kit SMARTTM RACE Amplification (Clontech), selon les recommandations du fournisseur. L'inverse transcription a été effectuée à partir de 5 µg d'ARN totaux de cellules HFF infestées par *E. cuniculi*, grâce à des amorces spécifiques définies à l'aide du logiciel GoArrays (**Rimour et al., 2005**). Cet algorithme permet de générer des amorces spécifiques des gènes d'*E. cuniculi* ciblés, tout en ne présentant aucune hybridation croisée avec les ARN des cellules hôtes. Ainsi, pour chaque CDS d'un opéron putatif, des amorces sens et antisens ont été définies respectivement pour les 3' et 5' RACE (**Tableau 5**).

Les réactions de PCR ont été réalisées selon les recommandations du fournisseur. Pour cela 3 étapes d'amplification ont été réalisées. Ces 3 étapes sont constituées respectivement de 5, 5 et 30 cycles. Ces cycles comprennent une dénaturation à 95°C (15 sec), une hybridation (15 sec) et enfin une élongation à 72°C (3 min). Cette hybridation est réalisée à 70°C pour la première étape, à 68°C pour la seconde et à 66°C pour la dernière. Ces 3 étapes sont précédées d'une dénaturation à 95°C (5 min) et suivies d'une élongation terminale à 72°C (7 min).

I.3.3. Clonage des produits PCR

I.3.3.1. Ligatures

Après migration sur gel d'agarose, les différentes bandes d'ADN amplifiées sont excisées du gel sous UV (254 nm). L'ADN est ensuite purifié en utilisant le kit Gel Extraction Kit (QIAGEN) et repris dans 30 µl d'eau distillée stérile. Les fragments PCR sont clonés grâce au kit pGEM-T Easy Vector (Promega) en utilisant 7 µl d'ADN purifié pour un volume réactionnel final de 10 µl.

I.3.3.2. Préparation des bactéries chimio-compétentes et transformation

Les cellules compétentes d'*E. coli*, souche XL1-Blue, sont préparées selon **Sambrook et al. (1989)**. Après transformation des bactéries avec le produit de ligature, celles-ci sont étalées sur milieu LB (tryptone 1%, NaCl 1%, extrait de levure 0,5% et agar 1,5%) supplémenté avec 100 µg.ml⁻¹ d'ampicilline, 0,1M d'IPTG (Sigma) et 40 µg.ml⁻¹ de X-Gal (Molecular biology). Les boîtes de Petri sont incubées une nuit à 37°C.

I.3.3.3. Extraction d'ADN plasmidique

L'ADN plasmidique est extrait selon la technique de la lyse alcaline (**Sambrook et al., 1989**) mais sans étape de purification au phénol. 2 µl des 20 µl de la préparation d'ADN sont digérés par l'enzyme de restriction *EcoRI* pour assurer la libération du fragment cloné dont la présence sera vérifiée par électrophorèse sur gel d'agarose 1%.

I.3.4. Séquençage des vecteurs recombinants

Les clones recombinants sont séquencés par la méthode de **Sanger et al. (1977)** avec le kit Big Dye Terminator v1.1 Cycle Sequencing (Applied Biosystems). Les réactions sont réalisées dans un volume final de 20 µl contenant 500 ng d'ADN plasmidique, 3,3 pmol d'amorce (SP6 ou T7), 3 µl de tampon 5X (Applied Biosystems) et 2 µl de mix (Applied Biosystems) selon les conditions préconisées par le fournisseur. Le séquençage est réalisé à l'aide du séquenceur ABI PRISM 310 Genetic Analyzer (Applied Biosystems).

I.3.5. Calcul des énergies libres d'appariement (SSU ARNr-ARNm)

Cette analyse a été réalisée selon la méthode décrite par **Osada et al. (1999)**. L'énergie libre d'appariement entre l'ARNr 16S et l'ARNm (50 nt de part et d'autre du codon d'initiation de la traduction) a été estimée. Cette analyse a été réalisée en parallèle chez la levure *S. cerevisiae* ainsi que chez 3 espèces bactériennes : *Haemophilus influenzae*, *Bacillus subtilis* et *Helicobacter pylori*.

I.4. Etude du génome de *B. algerae*

I.4.1. Réalisation de caryotypes moléculaires

I.4.1.1. Préparation des blocs d'agarose

Les spores de *B. algerae* purifiées et conservées en PBS à 4°C comme décrit précédemment, sont soumises à un traitement à la désoxyribonucléase I dans un volume final

de 1 ml (solution à 250 U.ml⁻¹, NaCl 0,1 M), 30 min à 37°C, afin d'éliminer l'ADN résiduel des cellules hôtes. L'action de la DNase est ensuite stoppée par un traitement de 20 min en EDTA 0,1 M suivi de plusieurs lavages en PBS. La concentration en spores est évaluée à l'aide d'une cellule de Malassez. Les spores sont incluses dans de l'agarose à bas point de fusion (LMP, Biorad) 1% en PBS 1X à raison de 5.10⁸ spores/bloc. Les blocs sont ensuite traités par une solution de lyse (EDTA 0,5 M, N-Lauryl sarcosine 1%, et protéinase K 2mg.ml⁻¹) pendant 48 h à 42°C, permettant la libération de l'ADN des spores dans le bloc d'agarose.

1.4.1.2. Electrophorèse en champs pulsés

Pour la résolution des différents caryotypes moléculaires, le système CHEF (contour-clamped homogeneous electric field) a été employé (CHEF-DR[®] System, Biorad). Les blocs d'agarose équilibrés 5 x 30 min en tampon TBE 0,5X, ainsi qu'une échelle de poids moléculaire (*S. cerevisiae*, souche YNN295, Biorad) sont déposés dans les puits d'un gel d'agarose GIGAPHOR (Eurobio) à 1,5% en TBE 0,5X. La résolution du caryotype moléculaire de *B. algerae* a nécessité l'utilisation de plusieurs programmes de séparation des bandes d'ADN chromosomique. La visualisation de l'ensemble des bandes a été obtenue suite à une migration de 48 h sous un courant électrique de 5,5 V/cm, à une température de 14°C en tampon TBE 0,5X, avec une progression linéaire des temps de pulsation de 60 à 120 sec, sous un angle de 120°. L'optimisation de résolution des plus petites bandes caryotypiques (160 à 370 kb) a nécessité une migration à 5,5 V/cm durant 40 h (temps de pulsation de 15 à 25 sec et de 20 à 29 sec durant respectivement 16 et 24 h). La séparation des bandes comprises entre 365 et 680 kb a été réalisée à 5 V/cm pendant 42 h (temps de pulsation de 66 à 71 sec et de 56 à 63 sec pendant 21 h). Enfin, une migration à 5 V/cm durant 49 h (temps de pulsation de 66 à 71 sec et de 56 à 63 sec pendant 28 et 21 h respectivement) a permis de séparer les molécules dont les tailles étaient comprises entre 500 et 1000 kb. Pour terminer la séparation dans la zone supérieure du caryotype moléculaire (bandes de 970 à 2240 kb), une migration durant 48 h à 5 V/cm a été réalisée (temps de pulsation de 120 à 145 sec). Les gels sont colorés 1 h à l'aide d'une solution de bromure d'éthidium (0,4 µg.ml⁻¹).

I.4.2. Stratégie de séquençage de *B. algerae*

1.4.2.1. Réalisation des banques d'ADN génomique

La purification des parasites et l'extraction de l'ADN ont été réalisées en utilisant les mêmes techniques que pour le séquençage du génome d'*E. cuniculi* (Katinka *et al.*, 2001). La

seconde phase, réalisée au Genoscope, a porté sur l'obtention de banques d'ADN génomique à partir du génome complet. Cette étape a permis d'obtenir trois banques plasmidiques (vecteurs pCNS et pCDNA), contenant des inserts de 2-3 kb ou de 7 kb et trois banques en vecteur BAC (vecteur pBELOBAC11) contenant des inserts de 25 à 50 kb. La troisième phase, toujours réalisée au Genoscope et correspondant au séquençage proprement dit, a permis d'obtenir 105019 séquences à partir des clones des 6 banques génomiques.

1.4.2.2. Analyses bioinformatiques

Une première analyse des séquences obtenues a été réalisée dans le but d'estimer la qualité des banques génomiques (biais de clonage, contamination) mais également la qualité des séquences. Pour cela, les séquences ont tout d'abord été traitées avec le logiciel pregap4 du package Staden (**Staden et al., 2000**) pour éliminer les régions correspondant au vecteur de clonage et aux régions de mauvaise qualité. Par la suite, une analyse comparative de chacune des séquences restantes a été effectuée à la fois contre les séquences du génome humain, du génome d'*E. coli* et des vecteurs de clonage, en utilisant le logiciel blastN. Les séquences spécifiques de *B. algerae* ont été assemblées en utilisant le logiciel gap4.

1.4.3. Détection des unités ADN_r par Southern-blot

1.4.3.1. Transfert sur membrane de nylon

Après migration et séparation des bandes d'ADN chromosomique par électrophorèse en champs pulsés, les gels sont traités 15 min avec une solution de HCl 0,25 M afin de dépuriner partiellement l'ADN. Celui-ci est ensuite dénaturé 30 min dans une solution alcaline composée de NaOH 0,5 M et NaCl 1 M. Le gel est ensuite neutralisé dans une solution de Tris-HCl 0,5 M et NaCl 0,5 M pendant 30 min. Le transfert est effectué par capillarité sur une membrane de nylon chargée positivement (BrightStarTM-Plus, Ambion) en SSPE 20 X (NaCl 3,6 M, NaH₂PO₄ 0,2 M, EDTA 20 mM).

1.4.3.2. Préparation et marquage des sondes

1.4.3.2.1. Préparation des sondes

Afin d'extraire l'ADN génomique destiné aux amplifications PCR, les spores de *B. algerae* purifiées ($5 \cdot 10^8$) sont reprises dans 200 µl d'eau stérile puis mises à ébullition 10 min (**Ombrouck et al., 1997**). Les réactions d'amplification sont réalisées à l'aide d'un thermocycleur (Thermal Cycler, Biorad) dans un volume de 50 µl avec 10 ng d'ADN, 1 µM de chaque amorce (directe : 5' AGAGCGTATTCAGTGGTC 3' et inverse : 5'

GCATACAGCACATCACAC 3'), 0,25 μM de chaque dNTP, 2,5 mM de MgCl_2 et 2,5 U de Taq DNA Polymerase (Clontech). Après une première étape de dénaturation de 3 min à 94°C, 35 cycles sont réalisés. Chacun de ces cycles comprend une dénaturation de 30 sec à 94°C, une hybridation de 30 sec à 42-65°C (selon le T_m des amorces) et une élongation de 2 min à 72°C. Une élongation finale est effectuée 10 min à 72°C.

Le produit d'amplification (327 pb) est déposé sur gel d'agarose 1%, visualisés sous UV après traitement au bromure d'éthidium (0,4 $\mu\text{g.ml}^{-1}$), découpés et purifiés sur colonne (gel extraction Kit, QIAGEN). Ils sont ensuite clonés en vecteur TOPO II.1 (TOPO-TA Cloning Kit, Invitrogen) pour être séquencés comme vu précédemment (I §3.4).

I.4.3.2.2. Marquage des sondes

Après digestion enzymatique des plasmides séquencés (4.3.2.1) et migration sur gel d'agarose, les inserts sont découpés puis élués dans 50 μl d'eau distillée stérile en utilisant le kit Gel Extraction Kit (QIAGEN). Les fragments d'ADN sont ensuite marqués avec le désoxyribonucléide [$\alpha^{32}\text{P}$]dCTP (Amersham) par la technique de marquage aléatoire (« random priming ») selon les recommandations du fournisseur (HexalabelTM DNA labeling Kit, Fermentas).

I.4.1.3. Hybridation et autoradiographie

Les membranes de nylon sont préhybridées dans un four à hybridation (Hybridazer HBI-TECHNE) 1h à 65°C, dans une solution SSPE 5X, SDS 1% et lait écrémé à 4 $\mu\text{g.ml}^{-1}$. Les sondes radiomarquées sont dénaturées 5 min à 100°C, puis ajoutées à la solution de préhybridation. L'hybridation est effectuée à 65°C pendant 16 h puis les membranes sont lavées deux fois 30 min à 65°C en SSPE 0.1 X et SDS 0,1 %. Elles sont ensuite mises au contact de films autoradiographiques BIOMAX MR[®] (Kodak) 24 h à -70°C.

I.4.4. Construction de la structure secondaire de l'ARNr 23S

La structure secondaire de l'ARNr 23S de *B. algerae* a été reconstruite manuellement. L'annotation des différentes hélices et domaines de la structure secondaire a été réalisée selon la nomenclature définie (De Rijk et al., 1998a). En ce qui concerne les régions hypervariables (notées V1 à V12), leur numérotation a été réalisée en accord avec l'annotation définie pour les ARN 23S eucaryotes (De Rijk et al., 1998b, Wuyts et al., 2001).

II. ETUDE DES RELATIONS HOTES-PATHOGENES

II.1. Description des lignées cellulaires d'*An. gambiae* : Sua4.0 et Mos 55

Nous disposons au sein du laboratoire de deux lignées cellulaires d'*An. gambiae*, une lignée hémocytaire Sua4.0 (fournie par F. Catteruccia, Imperial College of Science, London) impliquée dans les défenses immunitaires du moustique, ainsi qu'une lignée embryonnaire Mos 55 (fournie par E. Saraiva, Universidade Federal do Rio de Janeiro, Brésil).

II.2. Culture des lignées cellulaires

La culture s'effectue en milieu Schneider's *Drosophila* medium (Life technologies) supplémenté en sérum de veau fœtal 10 %, et divers antibiotiques (streptomycine 100 $\mu\text{g.ml}^{-1}$, pénicilline 100 U. ml^{-1} et fungizone 2,5 $\mu\text{g.ml}^{-1}$) à 25°C. Après 1 semaine de culture, le tapis cellulaire est décollé par agitation. Les cellules sont remises en suspension dans un volume final de 20 ml et le repiquage est réalisé au facteur de dilution souhaité (1/3 à 1/5).

II.3. Etude de l'expression des variants du gène *cap*

II.3.1. Matériel biologique étudié

La mise en évidence de l'expression de l'ensemble des variants du gène *cap* (c-Cbl associated protein), constitutive ou induite par stimulation antigénique a été réalisée chez les deux lignées cellulaires (Sua4.0 et Mos 55). La caractérisation de l'expression du transcrite *14Yde* (un des variants du gène *cap*) à différents stades de développement (stades œuf, larves et nymphes) ainsi que chez le moustique adultes (mâles ou femelles), en l'absence ou non de repas sanguins (femelles gorgées et non gorgées) a également été effectuée.

II.3.2. Stimulations antigéniques

Ces stimulations ont été réalisées uniquement en culture cellulaire sur les deux lignées à disposition (Sua4.0 et Mos 55) par ajout dans le milieu de culture de LPS (10 $\mu\text{g.ml}^{-1}$), Zymosan (100 $\mu\text{g.ml}^{-1}$) ou encore H₂O₂ (20 mM). Après 4 h d'induction, les milieux sont éliminés et les cellules lavées en PBS 1X à deux reprises, puis incubées à 25°C dans du milieu Schneider's supplémenté avec 10% de SVF en présence d'antibiotiques.

Tableau 6. Séquences des amorces utilisées pour l'amplification des variants du gène *cap* ou le gène codant l'actine chez *An. gambiae*.

Amorces	Séquences
14V1D	5' -CCAAAGAGCAATTAGAGCATC-3'
14V1R	5' -TACTCGAACCGGATCGGC-3'
14V2D	5' -TATGGGTCAGGAGAGCGG-3'
14V2R	5' -CATCCCACCACTCTTTGG-3'
14V3D	5' -TG TAGAGATACTGACCGACATC -3'
14V3R	5' -ATCCTGTGCGTGCGGAC-3'
14V4D	5' -GCCAACAGTTCGAACAGCG-3'
14V4R	5' -GGATGACGATCGTTCGGC-3'
14V5D	5' -CACGTTACGATCAGAAATTGATG-3'
ActineD	5' -CATGGAAGCGTGCGGCAT-3'
ActineR	5' -TTAGAAGCACTTTCGGTGCA-3'

II.3.3. Extraction des ARNs et RT-PCR

Aux temps T+ 1, 6, 12 et 24 h, après stimulation antigénique, les ARN totaux des cellules sont extraits avec le kit RNeasy Mini kit (QIAGEN) selon les recommandations du fournisseur. L'extraction des ARNs totaux des échantillons de moustiques à différents stades a été réalisée à l'aide du kit TriReagent RNA Isolation Reagent (Sigma). Une analyse qualitative et quantitative des ARNs totaux est réalisée sur Bioanalyser Agilent.

5 µg d'ARN totaux sont mis en présence de l'amorce inverse spécifique du ou des variants considérés à une concentration finale de 1 µM, puis dénaturés 10 min à 70°C La réaction de transcription inverse est effectuée à 42°C pendant 1 h après addition de 2,5U de reverse transcriptase MMLV (Promega), 10U de RNAsine (Promega), 4 µl de tampon MMLV 5X et 1 µl de dNTPs (10 mM) pour un volume final de 20 µl. Les réactions PCR sont réalisées en utilisant 5 µl d'ADNc obtenu précédemment après dilution au 1/10^{ème}. Les séquences des amorces utilisées sont répertoriées dans le **Tableau 6**. L'amplification est assurée par une ADN polymérase ThermoStart (Invitrogen) et comprend une dénaturation initiale à 94°C pendant 20 min, suivie de 30 cycles comportant trois étapes : une dénaturation à 94°C pendant 30 sec, une hybridation à 60°C pendant 20 sec et une élongation à 72°C (1 à 2 min 30 sec). Après une élongation finale de 10 min, les produits d'amplification sont analysés sur gel d'agarose et visualisés par coloration au bromure d'éthidium (0,4 µg.ml⁻¹). Les bandes d'intérêt sont découpées et purifiées (gel extraction Kit, QIAGEN), clonées en vecteur TOPO II.1 (TOPO-TA Cloning Kit, Invitrogen) puis séquencées.

II.3.4. Analyses bioinformatiques

Les électrophorégrammes obtenus après séquençage sont alignés grâce au Staden Package version 2000. Les programmes BLAST (Basic Local Alignment Search Tool, **Altschul et al., 1990**) sont utilisés pour les recherches d'homologies avec les séquences nucléiques et protéiques disponibles dans les banques de données. L'alignement des différents ADNc et ESTs avec la région génomique a été réalisée avec le logiciel SIM4 (PBIL, <http://pbil.univ-lyon1.fr>).

II.4. Essais de transfection de la lignée cellulaire Sua4.0

II.4.1. Constructions plasmidiques réalisées

Les différents vecteurs ont été construits à partir du plasmide pAc5.1a, (**Fig.23**), (Invitrogen). Ce plasmide permet d'initier l'expression d'un gène d'intérêt sous contrôle d'un promoteur actine de drosophile. Le variant d4 du gène codant la protéine EGFP (Enhancer

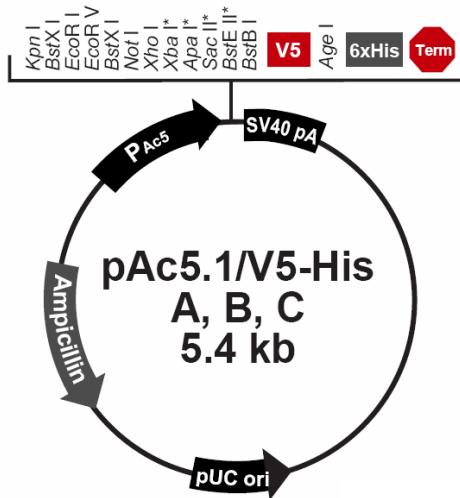


Figure 23. Carte du vecteur d'expression pAc5.1/V5-His (d'après les informations d'Invitrogen, <http://invitrogen.com>). Le gène d4EGFP a été cloné entre les sites de restriction *EcoRI* et *XbaI* du vecteur pAc5.1(A).

Tableau 7. Séquences des amorces utilisées pour l'amplification et le clonage du gène rapporteur codant d4EGFP dans le vecteur d'expression homologue pAc5.1 (A) (Invitrogen). Les sites de restriction utilisés (*EcoRI* et *XbaI*) sont indiqués en italique. La séquence soulignée correspond à la séquence Kozak de référence déduite de l'analyse des données expérimentales (ESTs). En rouge, sont indiquées les bases ayant été modifiées par rapport à la séquence Kozak de référence.

Amorces	Séquences
d4EGFP-AK(<i>EcoRI</i>)	5' -G <i>GAATTC</i> ACAAACATGGTGAGCAAGGGCGAG-3'
d4EGFP-NK1	5' -G <i>GAATTC</i> ACAAAATGGTGAGCAAGGGCGAG-3'
d4EGFP-NK2	5' -G <i>GAATTC</i> AAAAAATGGTGAGCAAGGGCGAG-3'
d4EGFP-NK3	5' -G <i>GAATTC</i> ACAAAATGGAGAGCAAGGGCGAG-3'
d4EGFP-NK4	5' -G <i>GAATTC</i> ACAAAATGAAAGAGCAAGGGCGAG-3'
d4EGFP-NK5	5' -G <i>GAATTC</i> AAAAAATGAAAGAGCAAGGGCGAG-3'
d4EGFP-NK6	5' -G <i>GAATTC</i> AAAAAATGAAAGCAAGGGCGAG-3'
D4EGFP-R (<i>XbaI</i>)	5' -GC <i>TCTAGA</i> CACATTGATCCTAGCAGAAGC-3'

Green Fluorescent Protein) d'*Aequoria victoria* a été utilisé comme gène rapporteur et cloné entre les sites *EcoRI* et *XbaI* du vecteur pAc5.1a (pAc5.1d4EGFP). Ce gène de 843 pb a été amplifié en utilisant comme matrice le plasmide pd4EGFP-N1 (Clontech). Les différentes constructions visant à étudier l'impact d'une modification nucléotidique dans l'environnement du codon initiateur de traduction (séquence Kozak) ont été réalisées par mutagenèse dirigée. Ces modifications ont été introduites grâce à l'utilisation d'oligonucléotides mutés. Les réactions PCR sont réalisées en utilisant 200 ng de matrice ADN (vecteur pAc5.1d4EGFP) et 1U d'ADN polymérase haute fidélité (Platinum Taq High fidelity, Invitrogen). Pour cela 2 étapes d'amplification ont été réalisées. Ces deux étapes sont constituées respectivement de 5 et 30 cycles. Ces cycles comprennent une dénaturation à 95°C (30 sec), une hybridation (20 sec) et enfin une élongation à 68°C (90 sec). Cette hybridation est réalisée à 60°C pour la première étape, et à 65°C pour la seconde. Ces deux étapes sont précédées d'une dénaturation à 95°C (5 min) et suivies d'une élongation terminale à 72°C (7 min). Les produits d'amplification sont analysés sur gel d'agarose et visualisés par coloration au bromure d'éthidium (0,4 µg.ml⁻¹). Les bandes d'intérêt sont découpées et purifiées (gel extraction Kit, QIAGEN), puis digérées de manière séquentielle à l'aide des enzymes *EcoRI* et *XbaI*. Après clonage, les différentes constructions sont vérifiées par séquençage. Les amorces utilisées pour les clonages sont regroupées dans le **Tableau 7**.

II.4.2. Les différentes techniques de transfection utilisées

II.4.2.1. La lipofection

24 h avant transfection, des cellules Sua4.0 sont repiquées sur lamelles de verre en boîte 6 puits à raison de 4 x 10⁶ cellules/puits. L'efficacité de différents agents de transfection tels que la Lipofectin (Invitrogen), la Flyfectin (OZ Biosciences), le Dreamfect (OZ Biosciences), ou encore la Cellfectin (Invitrogen), a été testée sur des tapis cellulaires à 80 % de confluence, selon les recommandations des fournisseurs. En raison de son efficacité, l'utilisation de la Cellfectin a été privilégiée pour la mise au point des conditions de transfection. Plusieurs paramètres ont été étudiés : concentrations en ADN plasmidique (de 2 à 50 µg), en Cellfectin (2 à 25 µl), temps d'incubation du mélange Cellfectin/ADN avec les cellules, présence ou absence de sérum de veau fœtal. Les cultures sont ensuite réalisées en milieu Schneider's supplémenté avec 10% de sérum de veau fœtal.

II.4.2.2 La magnétofection

24 h avant transfection, des cellules Sua4.0 sont repiquées en boîte 6 puits à raison de 4×10^6 cellules/puits. La transfection des cellules par le vecteur d'expression pAc5.1d4EGFP a été réalisée en association avec les billes magnétiques (OZ Biosciences) et la Cellfectin. L'ADN plasmidique (10 à 30 μg d'ADN/ μl de billes) est incubé 15 min avec les billes magnétiques en milieu MEM sans sérum. 50 μl à 200 μl du mélange ADN/billes sont mis en contact avec les cellules. L'ensemble est placé 30 min sur une plaque magnétique. Les cultures sont ensuite réalisées en milieu Schneider's supplémenté avec 10% de sérum de veau foetal.

II.4.3. Quantification des transformants

Les cellules sont fixées dans du méthanol froid à -80°C , 24, 48, ou 72 h après transfection. Après marquage du noyau au DAPI (10 $\mu\text{g}.\text{ml}^{-1}$), les cellules sont montées entre lame et lamelle en présence de moviol. Les préparations sont observées à l'aide d'un microscope à épifluorescence LEICA DMR.

II.4.4. Inactivation par ARN interférence

II.4.3.1. Préparation des ARNdb

La production d'ARNdb (600 pb) ciblant le messager codant l'EGFP a été réalisée par transcription *in vitro* à l'aide du kit MEGAscript® (Ambion) selon les recommandations du fournisseur.

II.4.3.2. Transfection de la lignée hématocytaire Sua4.0

Brièvement, des cellules Sua4.0 ont été repiquées en boîte 6 puits à raison de $1,5 \times 10^6$ cellules/puits 24 h avant l'étape de transfection. La transfection a été réalisée avec le vecteur d'expression pAc5.1d4EGFP à raison de 20 μg d'ADN/puits. 15 μl de Cellfectin ont été utilisés. 48 h après transfection, 2 ml de milieu Schneider's (sans sérum ni antibiotiques) contenant 20 μg d'ARNdb ont été mis en contact avec les cellules pendant 12 h. 48 h, 72 h et 120 h après ajout des ARNdb, les cellules sont fixées et montées en lame et lamelles comme précédemment (II.4.3) pour dénombrement.

II.5. Identification des protéines codées par le gène *cap*

II.5.1. Production d'anticorps polyclonaux

II.5.1.1. Expression d'une protéine 14YDE recombinante chez *Escherichia coli*

Le gène codant 14Yde a été cloné dans le vecteur d'expression pGEX-4T1 (Pharmacia) entre les sites *Bam*H1 et *Eco*RI (directe : 5' CGGGATCCACAAATA-TGTTTGCAGAG 3' et inverse : 5' CGGAATTCCATTTTGTGACGTAGTT 3'). Le vecteur d'expression de type procaryote utilisé possède en amont du site multiple de clonage une séquence codant la Glutathion-S-Transférase (GST) permettant une purification de la protéine recombinante par chromatographie d'affinité sur colonne de glutathion. L'expression des protéines recombinantes est réalisée chez *E. coli* BL21⁺ en milieu LB contenant 100 µg.ml⁻¹ d'ampicilline. Dès que la DO_{600nm} atteint 0,5, l'expression des protéines recombinantes est induite par ajout d'IPTG 2 mM. Après 4 h, les bactéries sont centrifugées à 5000 rpm pendant 3 minutes. Les protéines bactériennes totales sont solubilisées dans du tampon de Laemmli contenant 2,5% de SDS et 100 mM de DTT puis sont analysées en SDS-PAGE (Laemmli, 1970).

II.5.1.2. Immunisation

La production des anticorps est réalisée chez des souris CD1 à partir des bandes protéiques séparées sur gel de polyacrylamide. Après coloration des protéines au bleu de Coomassie, la bande protéique correspondant à la protéine recombinante d'intérêt est découpée et broyée dans 200 µl d'eau stérile à l'aide d'un Potter. Les échantillons ainsi obtenus sont homogénéisés avec de l'adjuvant de Freund complet (Sigma) et injectés dans la cavité intra-péritonéale des souris. De l'adjuvant de Freund incomplet est ensuite utilisé pour les injections réalisées 14, 21 et 30 jours après la première injection. Le sang de l'animal est prélevé une semaine après la dernière immunisation pour évaluer la réponse immunitaire.

II.5.2. Electrophorèse monodimensionnelle

L'extraction des protéines totales de moustiques mâles et femelles est réalisée par broyage mécanique à l'aide d'un Potter en tampon de Laemmli (SDS 2,5% et DTT 100 mM). Les échantillons sont ensuite placés 15 min à ébullition et centrifugés 5 min à 14000 g afin d'éliminer les débris cellulaires, et analysés sur gel de polyacrylamide 10-12%. La migration s'effectue à 25 mA par gel, puis les gels sont colorés au bleu de Coomassie (Biorad).

II.5.3. Electrophorèse bidimensionnelle

Les protéines totales d'échantillons de moustiques mâles sont extraites en tampon de solubilisation à l'aide d'un Potter (Urée 8M, Thiourée 4M, CHAPS 4%, Tris HCl 20 mM, EDTA 5 mM, MgCl₂ 5 mM) en présence d'inhibiteur de protéases (P8215, P8340, Sigma). Les protéines de chaque échantillon biologique à traiter (20 moustiques) sont extraites dans 100 µl de tampon de solubilisation. 5 cycles de broyage (30 sec)/glace (30 sec) sont appliqués à l'échantillon. 100 µl de tampon de solubilisation ainsi que 10 µl de TBP 0.2 M sont ensuite ajoutés à l'échantillon. Après centrifugation (10000 g pendant 20 min), les protéines sont dosées par colorimétrie à l'aide du Kit « Bio-Rad Protein Assay » (BioRad), selon la technique de Bradford. Les protéines sont ensuite précipitées à l'acétone glaciale, reprises en tampon de solubilisation (Urée 7M, Thiourée 2M, CHAPS 4%) pour obtenir une concentration finale de 5 µg.µl⁻¹ et conservées à -20°C. Les quantités de protéines totales déposées ont varié de 40 à 60 µg pour la réalisation des cartes de référence, et de 200 à 800 µg pour les cartes destinées aux prélèvements des spots protéiques. A la solution protéique, ont été ajoutés 4 µl de TBP (0,2 M) et 15 µl d'ampholytes (pH 3/10). La migration des échantillons protéiques en 1^{ère} dimension est réalisée avec le système Multiphor II (Amersham) à une température constante de 19°C pendant 22 h (36000 Vh total). La seconde migration est réalisée en condition dénaturante, à 15 mA/gel sur gel de polyacrylamide 10 %.

II.5.4. Western blot

Les protéines séparées par électrophorèse SDS-PAGE sont transférées sur membrane de PVDF (PolyVinylidene DiFluoride, PolyLabo) en utilisant un système de transfert semi-sec (Amersham). Une étape de saturation est effectuée en présence de PBS-lait écrémé 5% pendant 1 h. La membrane est ensuite incubée pendant 3 h avec les anticorps anti-V5 couplé AP (alcaline phosphatase) ou anti-14Yde dilués respectivement au 1/500^e ou 1/2000^e en PBS-Triton X-100 0,1%. Après 3 lavages en PBS-Triton X-100 0,1%, la membrane est traitée 1 h avec l'anticorps secondaire anti-IgG de souris couplés à la phosphatase alcaline (Promega) dilué au 1/10000. Trois lavages en PBS-Triton X-100 0,1% suivis d'un lavage en PBS sont effectués. La révélation est réalisée en présence de NBT et BCIP selon les recommandations du fournisseur (Promega).

II.5.5. Identification protéique par spectrométrie de masse

Les bandes protéiques et spots d'intérêt ont été extraits manuellement d'un gel SDS-PAGE monodimensionnel ou bidimensionnel coloré au bleu de Coomassie. Après plusieurs

lavages en solutions de décoloration (25 mM de NH_4HCO_3 pendant 15 min suivi d'un bain de 15 min avec une solution d'acétonitrile 50% (v/v) contenant 25 mM de NH_4HCO_3), le gel est déshydraté avec une solution d'acétonitrile 100%. Après séchage, le gel est incubé avec une solution de réduction (NH_4HCO_3 , 10 mM DTT) pendant 1 h à 37°C puis avec une solution d'alkylation (NH_4HCO_3 , 55 mM d'iodoacétamide) pendant 30 min à 37°C. Plusieurs lavages avec les solutions de décoloration sont réalisés avant un lavage final en eau pure pendant 15 min avant un traitement à l'acétonitrile 100%. Selon la quantité de protéines, 2 à 3 μl de trypsine modifiée diluée (0,1 $\mu\text{g}\cdot\text{ml}$ dans 25 mM de NH_4HCO_3 , Promega) sont ajoutés. Après 30 min d'incubation, 7 μl de 25 mM de NH_4HCO_3 sont ajoutés sur le gel pour une incubation 5 h à 37°C. Pour l'analyse en MALDI-TOF, une fraction de 0,5 μl est mélangée avec 0,5 μl de solution de matrice (acide α -cyano-4-hydroxycinnamique saturé à 50% avec une solution d'acétonitrile 60%-acide trifluoroacétique 0,1%). La solution résultante est déposée sur la plaque et rincée avec 2 μl de TFA 0,1% pendant 30 sec. Le mélange de peptides est analysé par un spectromètre de masse de type MALDI-TOF avec une moyenne de masse allant de 0 à 4200 Da. Dans la plupart des cas, les protéines ayant un score supérieur à 70 et une moyenne de couverture de 20% par rapport aux séquences des banques de référence sont automatiquement validées. Pour les autres, une vérification manuelle est réalisée.

II.6. Essais d'infestation des cellules d'insectes par *B. algerae*

II.6.1. Infestation des lignées cellulaires

Des cellules HFF, Sua4.0 et Mos 55 cultivées sur lamelles de verre sont infestées avec des spores purifiées de *B. algerae*. En fonction des lignées cellulaires, une à deux semaines après infestation, les cellules sont fixées 20 min en présence de méthanol froid (-80°C). Dans le cas de marquage de spores extracellulaires, des frottis de spores de *B. algerae* sont réalisés sur des lames de verre recouvertes de polylysine. Les préparations sont fixées de la même manière en présence de méthanol.

II.6.2. Mise en évidence de l'infestation par *B. algerae*

*II.6.2.1. Production d'un sérum total anti *B. algerae**

Des spores de *B. algerae*, purifiées et conservées à 4°C en PBS, sont fixées en paraformaldehyde 4%-glutaraldehyde 0.1% pendant 20 min à température ambiante. L'échantillon biologique est ensuite placé 15 min à ébullition puis 10 cycles de congélation-décongélation en azote liquide sont réalisés. L'immunisation de souris CD1 est réalisée comme vu précédemment par injection intra-péritonéale de $2 \cdot 10^6$ spores traitées.

II. 6.2.2. IFA

Après une étape de perméabilisation de 30 min avec une solution de PBS-Triton X-100 0,5%, une étape de saturation en PBS-lait écrémé 5% est réalisée pendant une heure. Le matériel biologique est alors incubé 1 h avec les anticorps primaires (sérum total anti-*B. algerae*) dilués en PBS/Triton X-100 0,1% (1/100^{ème}). Plusieurs lavages en PBS-Triton X-100 0,1% sont réalisés avant d'incuber les cellules 1 h avec les anticorps secondaires (anti-IgG de souris ou de lapin) conjugués au fluorochrome Alexa 488 (dilué au 1/1000^{ème}, Molecular Probes). Afin de mieux visualiser les foyers infectieux ou les noyaux des spores, une coloration au DAPI est réalisée. Les préparations sont observées à l'aide d'un microscope à épifluorescence LEICA DMR. Le traitement des images est réalisé à l'aide du logiciel Adobe Photoshop 7.0.

RESULTATS

I. Un nouveau mécanisme de régulation transcriptionnelle chez la microsporidie *Encephalitozoon cuniculi*

A. SYNOPSIS

L'organisation des gènes en opéron est peu répandue chez les eucaryotes, contrairement aux procaryotes où elle permet de regrouper la plupart du temps des gènes impliqués au niveau d'une même fonction biologique ou d'une même voie métabolique, assurant ainsi une régulation transcriptionnelle et traductionnelle commune (**Salgado et al., 2000**). Chez les eucaryotes, quelques cas isolés d'organisation en opérons ont été décrits chez la drosophile, les mammifères et les plantes. Cependant, ce type d'organisation est beaucoup plus fréquent chez les nématodes, notamment chez l'espèce *Caenorhabditis elegans* et les Trypanosomes (**Blumenthal, 2004**). Chez ces deux groupes d'eucaryotes, les opérons sont transcrits sous forme d'un polycistron à partir duquel sont libérés les ARNm monocistroniques, suite à un clivage et à une réaction d'épissage en *trans* faisant intervenir un ARN apportant la coiffe de chaque ARNm monocistronique (**Blumenthal, 2004**).

Chez *E. cuniculi*, l'analyse d'un contig de 4,8 kb réalisée par **Duffieux et al. (1998)** a permis de montrer que les espaces entre deux CDS successifs pouvaient être réduits à 34 pb. La longueur moyenne entre deux CDS ayant la même orientation, calculée pour l'ensemble des gènes, est quant à elle de seulement 129 pb (**Keeling et Slamovits, 2004**).

Une telle réduction de taille laisse difficilement envisager la présence au sein de séquences aussi courtes, de motifs nucléiques impliqués à la fois dans l'initiation de la transcription du gène le plus en aval, et dans la fin de transcription du gène le plus en amont.

Dans le but de caractériser une organisation polycistronique chez *E. cuniculi*, des expériences de RACE-PCR ont été réalisées sur plusieurs opérons putatifs.

L'étude des mécanismes transcriptionnels d'un opéron composé de 4 gènes de même orientation et séparés par des espaces intergéniques de 136 pb, 77pb et 97 pb, a permis de caractériser plusieurs populations d'ARNm et a révélé un mécanisme de régulation

transcriptionnelle original. Les résultats expérimentaux montrent que l'ARNm polycistronique subit une maturation ne faisant pas intervenir un processus d'épissage en *trans*. Ce processus est à l'origine de la formation de messagers monocistroniques et polycistroniques.

Ces approches expérimentales et une analyse « *in silico* » du protéome d'*E. cuniculi* permettent de proposer un nouveau modèle de régulation transcriptionnelle basé sur la reconnaissance aléatoire du signal de clivage/polyadénylation assurant l'étape de maturation de chaque ARNm monocistronique au cours de la transcription.

Des expériences de RACE-PCR ciblant des gènes de même orientation et séparés par des espaces intergéniques de 150 à 596 pb suggèrent qu'en moyenne, la distance séparant deux unités transcriptionnelles se situe entre 300 et 400 pb.

Une telle organisation génomique, définie par une transcription polycistronique et une maturation aléatoire reflète probablement des mécanismes d'adaptation mis en place par *E. cuniculi* au cours de l'évolution en raison de son statut de parasite intracellulaire obligatoire.

B. PUBLICATION 1: « Genome evolution of the eukaryotic parasite *Encephalitozoon cuniculi* is related to polycistronic gene clusters »

Genome evolution of the eukaryotic parasite *Encephalitozoon cuniculi* is related to polycistronic gene clusters

Keywords: Microsporidia; genome evolution; transcriptional regulation; RACE-PCR

Abdel Belkorchia, Cécile Militon, Eric Peyretailade, Sébastien Rimour, Philippe Veisseire, Olivier Gonçalves and Pierre Peyret*

LBP UMR CNRS 6023
Génomique Intégrée des Interactions Microbiennes
Université Blaise Pascal– Institut Universitaire de Technologie
Clermont-Ferrand Universities BP 86
63177 Aubiere Cedex
France

Abbreviations: RACE, rapid amplification of cDNA end; UTR, untranslated region; CDS, coding DNA sequence; DB, downstream box.

* Corresponding author: LBP UMR CNRS 6023, 63177 Aubiere Cedex, France.
Tel: +33 (0)4.73 40 76 04. Fax : +33 (0)4 73 40 76 70.
E-mail address: pierre.peyret@univ-bpclermont.fr

Abstract

Microsporidia are obligate intracellular eukaryotic parasites with genomes as small as 2.3 Mbp. The 2.9 Mbp genome of the microsporidian *Encephalitozoon cuniculi* has been fully sequenced and genome annotation has revealed a gene density of 0.97 genes per kilobase. The aim of this study is to characterize transcriptional regulations that could explain genes expression in a highly compact genome. To explore such regulations of genes expression, we performed 5' and 3' RACE PCR experiments on different mRNAs encoded by genes in close proximity and present on the same DNA strand. Analysis of the RACE-PCR fragments revealed that polycistronic transcription occurs in the *E. cuniculi* species. In a context of genome reduction, polycistronic gene clusters should be an efficient solution for genes expression. The studies have also showed a highly reduction of the 5'-UTR regions (<7 nts). The polycistronic organization and the translational machinery simplification in *E. cuniculi* could be the consequence of the simplification of biological process during the adaptation to the obligate parasitic lifestyle.

1. Introduction

It has been recognized that operons represent a common form of gene organization in bacteria and archaea. In *Escherichia coli*, for example, the 4000 genes are included into at least 700 operon structures (Salgado et al., 2000). By contrast, the genes of eukaryotes are generally considered to be monocistronic, each with its own promoter at the 5' end and a transcription terminator at the 3' end. Nevertheless such polycistronic organizations have been characterized in isolated cases in flies, mammals and plants and are frequently found in trypanosome and nematodes. In these two latter groups, operons are transcribed as extended polycistronic pre-mRNA and individual mRNAs are cleaved from this precursor by a trans-splicing reaction, adding a capped RNA, the spliced leader (SL) to the 5' end and by polyadenylation at the 3' end (Blumenthal and Gleason, 2003). In the nematode *Caenorhabditis elegans*, at least 15% of the genes are contained in operons (Blumenthal et al., 2002). This organism is characterized by a highly compact genome and the polycistronic transcription units comprise multiple genes that are generally separated by ~100 bp.

Microsporidia are amitochondriate eukaryotes. However, all the most recent phylogenetic studies have argued against an early origin of microsporidia and supported a placement of these organisms among fungi (James et al., 2006). These organisms are also characterized by small genomes ranging from 19.6 Mb to only 2.3 Mb. *Encephalitozoon cuniculi* harbours the smallest sequenced eukaryotic genome which is divided in 11 chromosomes (Katinka et al., 2001; Peyret et al., 2001). Genome compaction is reflected by the shortness of most putative proteins relative to their eukaryote orthologues but also by the clustering in close proximity with the same orientation of many genes. Analysis of a DNA contig of 4.3 kbp (Duffieux et al., 1998) showed that these intergenic spacers may be reduced to 34 bp. Statistical analysis performed on the 1997 CDSs isolated from the *E. cuniculi*

genome have revealed that the average space between two consecutive CDS with the same orientation is only to 156 bp (Keeling and Slamovits, 2004). Such shortness spacers are unlikely to present both promoter sequences of downstream genes as well as terminator signals of upstream ones. Therefore this genomic structure seems to be in agreement with an operon organization. Such organization has been so validated on different gene clusters of the *E. cuniculi* genome in the present study.

2. Materials and methods

2.1. Cell culture

Cell culture and infestation by *Encephalitozoon cuniculi* were made as previously described (Peyret et al., 2001).

2.2. RNA extraction

Total RNA was extracted from *E. cuniculi* infected HFF cells using RNeasy Midi Kit (Qiagen). RNA quality and quantity were estimated using the Agilent bioanalyser with the RNA 6000 Nano LabChip® Kit.

2.3. Isolation of full length cDNAs and characterization of the operon organization

5' and 3' RACE analyses were carried out to amplify cDNA ends of putative operons using the SMART™ RACE Amplification kit according to the recommendations of the manufacturer (CLONTECH). Reverse Transcriptions were performed on 5 µg *E. cuniculi* total RNA extract from infected cells. Primer design for RACE-PCR was performed using a modified program based on the algorithm developed by Rimour et al. (2005). This algorithm permitted us to identify specific primers which are unable to cross hybridize with host cells

RNA. Two primers were then defined for each of the four genes. An antisense one, for 5' RACE-PCR experiments and a sense one, for 3' RACE-PCR experiments. After electrophoresis migration in an agarose gel, PCR products were purified with the QIAquick Gel Extraction kit (QIAGEN), cloned into the pGEM[®]-T Easy Vector (PROMEGA) and sequenced with the Big-Dye[®] Terminator v1.1 kit (APPLIED BIOSYSTEMS) on the ABI PRISM 310 Genetic Analyzer (APPLIED BIOSYSTEMS) according to the recommendations of these different manufacturers.

2.4. Free energy estimation between SSU rRNA and translated gene

According to methods described by Osada et al. (1999), the free energy values corresponding to base pairing between *E. cuniculi* 16S rRNA sequence and sequences including 50 nucleotides upstream and downstream each annotated start codon have been calculated. Such analyses were also performed for *Saccharomyces cerevisiae* and three bacterial species, *Haemophilus influenzae*, *Bacillus subtilis* and *Helicobacter pylori*.

3. Results

3.1. Polycistronic transcription in *E. cuniculi*

In order to validate a polycistronic organization in *E. cuniculi* genome, 5' and 3' RACE experiments were performed using specific primers designed from the ubiquitin (*ubiq*), the polyA polymerase (*pap*), the histone H2A (*h2a*) and the histone transcription regulator (*htr*) Coding DNA Sequences (CDSs). These CDSs, located on the chromosome II of *E. cuniculi* (ECU02_0710: CAD25101, ECU02_0720: CAD25102, ECU02_0730: CAD25103 and ECU02_0740i: CAD25104), have the same orientation and are separated by intergenic spacers of 136 bp, 77 bp and 97 bp respectively.

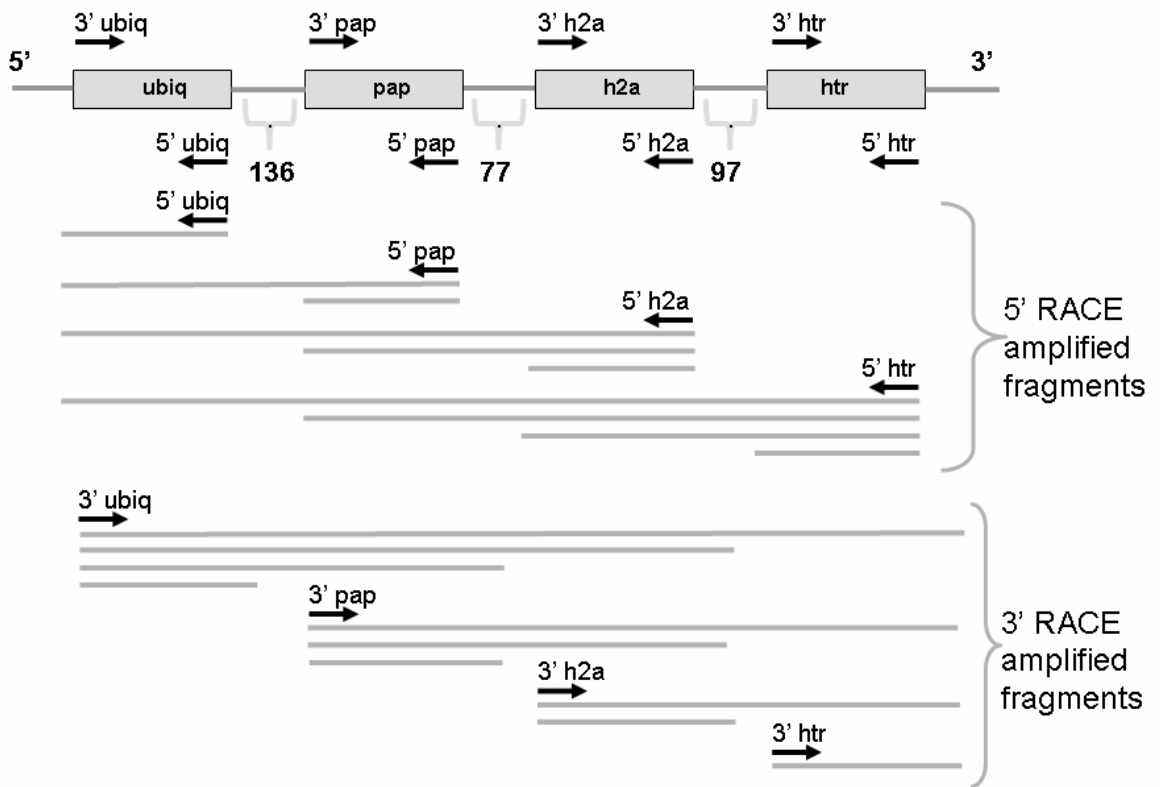


Fig. 1. Schematic representation of a polycistronic gene organization in *E. cuniculi* genome. 5' and 3' RACE products were generated using specific primers designed from the *ubiquitin* (*ubiq*), the *polyA polymerase* (*pap*), the *histone H2A* (*h2a*) and the *histone transcription regulator* (*htr*) Coding DNA Sequences drawn in grey. Multiple RACE products obtained with a specific primer indicate a new eukaryotic transcriptional regulation coupled to processing. Intercistronic spacers are indicated in bp and localized with grey brace brackets. Arrows represent the orientation of the specific primers.

Contrary to expectations, these RACE experiments enabled the amplification of one to four DNA fragments depending on the primer used (Fig. 1). All the amplified fragments were cloned and sequenced.

5' RACE PCR performed with the 5'-ubiq primer revealed only one DNA fragment. Sequence analysis demonstrated that this experiment allowed us to amplify the *ubiq* gene 5' end and showed that the 5' UTR region is reduced to 6 nts. Using the 5'-pap primer, two PCR products were obtained. Sequence of the largest DNA fragment allowed us to confirm the *ubiq* gene 5' end and is thus in agreement with an operon organization. But, unexpectedly, analysis of the smallest sequence revealed the 5' end of the *pap* gene and located it at 7 or 3 nts upstream of the putative AUG, depending on the clone product. With the 5'-h2a and the 5'-htr specific primers, 3 and 4 PCR products were obtained respectively. In each case, analysis of the smallest fragment allowed the characterization of 5' end of the *h2a* and *htr* genes positioned at 1 and 4 nts upstream of their respective putative AUG. Sequences studies of the higher fragments reinforced the operon organization and confirmed the 5' end of the *ubiq* and *pap* genes for the two experiments and the 5' end of *h2a* gene for the experiment performed with the htr primer.

For 3' RACE PCR, a single fragment was amplified using the 3'-htr primer. Sequence analysis of this fragment placed the polyadenylation site 82 nts downstream of the stop codon. This polyadenylation site was confirmed by sequencing the highest PCR products obtained for each of the 3'-h2a, the 3'-pap and the 3'-ubiq primers. These primers also allowed the polyadenylation sites of the *h2a* (primers 3'-h2a, 3'-pap and 3'-ubiq) and the *pap* gene (primers 3'-pap and 3'-ubiq) to be determined. They were respectively located 36 and 44 nts downstream of their stop codon. Using the 3'-ubiq primer, the polyadenylation site of the *ubiq* gene could also be defined (16 nts downstream of the stop codon) in addition to the polyadenylation sites of *pap*, *h2a* and *htr* genes.

3' RACE experiments revealed that the histone *h2a* transcript is polyadenylated. Primary and secondary structure analyses performed on 3' UTR of this latter mRNA and on other putative members of this histone family genes located in *E. cuniculi* genome did not reveal a conserved sequence or a stem loop structure implicated in mRNA stability as in metazoan organisms. Furthermore, genes coding for stem loop binding protein (Dominski et al., 1999) and U7 snRNA genes (Bond et al., 1991), required for the 3' end processing and the turnover of these mRNA in metazoan species, were absent in the *E. cuniculi* genome. Therefore, final 3' end maturation of the histone family mRNA results only in a polyadenylation step to ensure their stability.

This four genes cluster is framed by CDSs in opposite direction. Ecu02_0700 is located at 416 bp from Ecu02_0710 and Ecu02_0750i also in opposite direction is localized at 241 bp from Ecu_0740i. We did not observe transcription overrun. This observation reinforces the operon organization hypothesis for *Encephalitozoon cuniculi*. Such an organization has also been validated for two other complete operons comprised of 2 and 6 genes, located respectively on chromosome I (ECU01_0860: CAD24956 and ECU01_0870: CAD24957 coding respectively for an hypothetical protein and the trehalose-6-phosphate phosphatase) and chromosome X (ECU10_1460: CAD25865, ECU10_1450: CAD25864, ECU10_1440: CAD25863, ECU10_1430: CAD25862, ECU10_1420: CAD25861, and ECU10_1410: CAD25860 coding respectively for the Glutaminyl tRNA synthetase, the proteasome B-type subunit delta chain PCRD, the integral membrane protein YSX3, the transcriptional activator GCN, the proteasome regulatory subunit 6A and an hypothetical protein). RACE-PCR fragments size analysis coupled to sequencing always confirmed operon organization (data not shown). It should be noted that the space between the two CDSs of the former operon is 150 bp and the 5 spaces of the latter varies from 48 to 125 bp. As for the first

genes cluster we did not observe transcriptional overrun in the genes in opposite direction located upstream and downstream of the genes cluster.

To estimate the minimum sequence length between two independent operons, 5' RACE-PCR experiments were performed on consecutive CDSs separated by a spacer greater than 150 bp. Genes and potential intergenic sizes are indicated in Table 1.

Multiple RACE-PCR products were systematically obtained with intercistronic space size below 270 bp. Sequencing of these PCR fragments confirm polycistronic organization. For spacing greater than 379 bp, only one RACE PCR corresponding to the 5' or the 3' maturation sites from the CDS targeted was obtained (Table 1).

For spacers ranging from 309 to 379, both situations have been observed. These analyses suggest that the size delineating two independent transcription units lies between 300 and 400 bp.

All these results are in agreement with a polycistronic transcription in *E. cuniculi*. Another hypothesis to explain these results could consists in considering that each monocistronic gene is under the control of a promoter and transcription read through the downstream region. Nevertheless, no multiple 3' RACE-PCR fragments were obtained for the last gene of each polycistronic structure tested. Furthermore, for genes, harbouring a same orientation and not in an operon structure, no multiple 5' and 3' RACE PCR products were obtained respectively for the downstream and the upstream gene. All these results also argue against the transcription read through the downstream region for the tested genes and reinforce the polycistronic organization.

3.2. Reduced 5' UTR is sufficient for translation

5' RACE PCR experiments show that the 5' UTR regions are highly reduced and that their size varies between 0 to 7 nucleotides for tested genes.

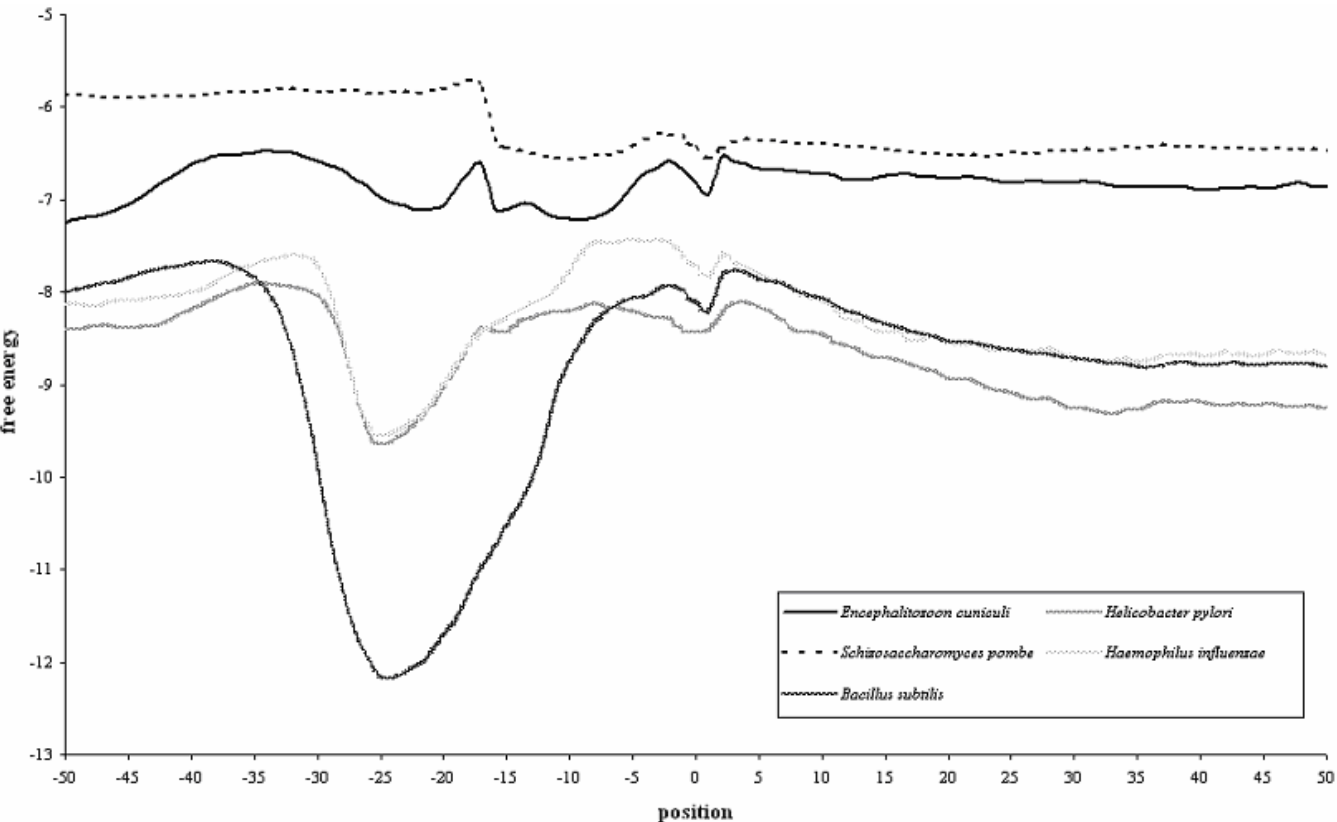


Fig. 2. Absence of downstream box (DB) to ensure translational machinery fixation in *E. cuniculi*.

The free energy value corresponding to base pairing between *E. cuniculi* 16S rRNA sequence and sequences including 50 nucleotides upstream and downstream each annotated start codon have been calculated. Such analyses were also performed for *Saccharomyces cerevisiae* and three bacterial species, *Haemophilus influenzae*, *Bacillus subtilis* and *Helicobacter pylori*. For the considered sequence portion, only a free energy value reduction was observed in the sequence upstream the start codon for the three bacterial species. Such free energy reduction is in agreement with the interaction between 3' SSU-RNA end and Shine Dalgarno sequence to ensure ribosomal fixation to perform translation. Like in *Saccharomyces cerevisiae* sequences, no sharp fall in free energy value was determined downstream the start codon sequence in *E. cuniculi* suggesting the absence of a putative downstream box (BD) to ensure translational machinery fixation.

We have also calculated the free energy values corresponding to base pairing between *E. cuniculi* 16S rRNA sequence and a 50 large nucleotide window on all annotated CDSs (Fig. 2). For the considered window, no sharp fall in free energy value was observed for *E. cuniculi* downstream the start codon suggesting the lack of a potential base pairing with 16S rRNA and consequently the absence of a putative downstream box (DB) to ensure translational machinery fixation. This result was also confirmed by Fuglsang non-randomness analysis where the data calculated did not support the hypothesis of a potential DB (Fuglsang, 2005). Furthermore, 5' RACE PCR experiments show that the 5' UTR regions are highly reduced and that their size varies between 0 to 7 nucleotides for tested genes. So no interaction could then be possible between 16S rRNA and such 5' UTR. All these results are in agreement with a translation mechanism based on mRNA capping to ensure ribosomal complex fixation.

4. Discussion

Our 5' and 3' RACE PCR experiments have revealed that polycistronic transcription process occurs in *E. cuniculi* species. Nevertheless, these results permitted us to characterize a mixed mature mRNA population corresponding to all the combinatorial maturation possibilities (Fig. 1). It appears that there are polycistronic transcription units in which RNAs can begin at the 5' end of any gene in the unit and end at the 3' end of any gene in the unit. We could postulate that the pre-mRNA is stochastically processed during the transcription step leading to these mature mRNAs. Nevertheless, we could not exclude a post-transcriptional maturation of the pre-mRNA leading to the formation of monocistronic gene like described in *Trypanosoma cruzi* (Jäger et al., 2007). However, the 5' RACE-PCR analysis did not allow us to detect the presence of a short splice leader (SL) added at the 5' end of the different process transcripts. It may thus be concluded that these processing mechanism do not imply a trans-

splicing maturation of the polycistronic pre-mRNA as described in trypanosomes (Vanhamme and Pays, 1995) but also in euglena (Frantz et al., 2000), nematodes (Spieth et al., 1993), flatworms (Davis, 1997), cnidarians (Stover and Steele, 2001) and, more recently, in the chordate *Oikopleura dioica* (Ganot et al., 2004). Indeed in these organisms, polycistronic precursors are processed into monocistronic mRNAs by conventional cleavage and polyadenylation at the 3' ends and trans-splicing at the 5' ends using a short splice-leader.

Recently published work has demonstrated the presence of several genes on a same mRNA for another microsporidian species (*Antonospora locustae*, genome size 5.4 Mb), for nucleomorphs of the cryptomonad *Guillardia theta* (genome size of 551 kbp) as well as for the chlorarachniophyte *Bigelowiella natans* (genome size of 380 kbp) (Williams et al., 2005). Multigene transcripts are more common in nucleomorphs (nearly 100% of mRNAs described) and virtually every cDNAs begin with a sense-strand gene. For the microsporidian *Antonospora locustae*, mRNA molecules that encode sequence from more than one gene are also highly common. In the present study we only characterized few operons in *E. cuniculi*. It should be interesting to evaluate the proportion of these structures in the complete genome.

Characterization, of polycistronic organization in *E. cuniculi* let suggest us that, in a process of simplification like genome compaction, such organization should be a more efficient solution compared with embedment of the DNA signals (promoter, termination) in adjacent genes. The Microsporidia and nucleomorphs are essential models for understanding the compaction of the eukaryotic nuclear genome with gene density as high as 0.97, 1.02 and 0.88 genes per kilobase for *E. cuniculi*, the nucleomorph of *Guillardia theta* and the nucleomorph of *Bigelowiella natans* respectively.

This polycistronic organization may be also the consequence of the simplification of biological process during the adaptation to the obligate parasitic lifestyle. The genome has therefore evolved to eliminate all the genes that may be considered unnecessary in this new

environment. Numerous metabolic pathways were eliminated since the parasite can use the resources of the host cell directly with efficient transporters (Katinka et al., 2001). Inside the host cell, the environment is extremely stable and therefore the pathogen does not need to adapt rapidly to environmental perturbations contrary to other organisms that are in direct contact with the outside environment. In fact, obligate parasites need to develop faster in the host cell using nutritional resources and must eliminate laborious gene regulations adapted to higher organisms and/or organisms that are not in a protected intracellular environment. A first example of such simplification in *E. cuniculi* was described for the processing of the rRNA unit (Peyretailade et al., 2001). Indeed primer extension and S1 nuclease protection experiments indicate a simple processing to ensure 16S and 23S rRNA release. Furthermore the presumed ITS1 sequence, characterized by strain-dependent polymorphism, remains linked to 16S rRNA. In the present study, we have demonstrated a simplification of the transcription process due to a polycistronic organization but also a probable simplification of the translation due to the reduction of the 5' UTR region of each mRNA molecule. Indeed, similarly short 5'UTR have also been observed from deeply branched anaerobic protozoa (*Entamoeba histolytica*, *Trichomonas vaginalis* and *Giardia lamblia*). In *Giardia lamblia*, experimental evidence has suggested that such 5' UTR size reduction associated with incomplete protein synthetic machinery should be sufficient for recruiting ribosomes and initiating translation without the requirement of scanning (Li and Wang, 2004). Our results suggest a cap dependent translation but without scanning in *E. cuniculi* conducted by simplified eukaryotic translational machinery in relation to the highly reduced 5' UTR.

It should be finally noted that tested operons do not include genes required for the accomplishment of an entire cellular function. Indeed only two genes encoding product implicated in the proteasome metabolism are located on a same transcription unit. In *C. elegans*, such an association is also not the rule (Blumenthal and Gleason, 2003). Although

these authors described some operons in which genes were associated to ensure a part of a metabolic pathway or a part of a protein complex, they also demonstrated that many operons contain genes that have no related function. Concerning expression level in this organism, it has been demonstrated that although clustered genes are expressed from a same promoter, they are often differentially regulated. This could be explained in worms by different efficiencies of RNA processing and/or by different mRNA stabilities (Blumenthal and Gleason, 2003). To validate such regulation in *E. cuniculi*, the level of each monocistronic gene of a same operon should be evaluated by real-time PCR.

Acknowledgments

Cécile Militon and Abdel Belkorchia were supported by a grant from "Ministère de l'éducation, de la recherche et de la technologie". We thank Kate James for expert reading of the manuscript.

References

- Blumenthal, T., et al., 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature* 417, 851-4.
- Blumenthal, T., Gleason, K.S., 2003. *Caenorhabditis elegans* operons: form and function. *Nat. Rev. Genet.* 4, 112-20.
- Bond, U.M., Yario, T.A., Steitz, J.A., 1991. Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both *in vivo* and *in vitro*. *Genes Dev.* 5, 1709-22.
- Davis, R.E., 1997. Surprising diversity and distribution of spliced leader RNAs in flatworms. *Mol. Biochem. Parasitol.* 87, 29-48.
- Dominski, Z., Zheng, L.X., Sanchez, R., Marzluff, W.F., 1999. Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. *Mol. Cell. Biol.* 19, 3561-70.
- Duffieux, F., Peyret, P., Roe, B.A., Vivares, C.P., 1998. First report on the systematic sequencing of the small genome of *Encephalitozoon cuniculi* (Protozoa, Microspora): gene organization of a 4.3 kbp region on chromosome I. *Microb. Comp. Genomics* 3, 1-11.
- Frantz, C., Ebel, C., Paulus, F. and Imbault, P., 2000. Characterization of trans-splicing in Euglenoids. *Curr. Genet* 37, 349-55.

- Fuglsang, A., 2005. A cytosine-rich region upstream of start codons serving as a signal for initiation of translation in *Encephalitozoon cuniculi*? Mol. Biochem. Parasitol. 140, 69-73.
- Ganot, P., Kallesoe, T., Reinhardt, R., Chourrout, D., Thompson, E.M., 2004. Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. Mol. Cell. Biol. 24, 7795-805.
- Jäger V., De Gaudenzi J.G., Cassalo A., D'Orso I., Frasch A., 2007. mRNA maturation by two-step trans-splicing/polyadenylation processing in Trypanosomes. Proc. Natl. Acad. Sci. U. S. A. 104, 2035-2042.
- Katinka, M.D., et al., 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414, 450-3.
- Keeling, P.J., Slamovits, C.H., 2004. Simplicity and complexity of microsporidian genomes. Eukaryot. Cell 3, 1363-9.
- Li, L., Wang, C.C., 2004. Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. J. Biol. Chem. 279, 14656-64.
- Osada, Y., Saito, R. and Tomita, M., 1999. Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes. Bioinformatics 15, 578-81.
- Peyret, P., et al., 2001. Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora). Genome Res 11, 198-207.
- Peyretaillade, E., et al., 1998. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res. 26, 3513-20.
- Peyretaillade, E., Peyret, P., Metenier, G., Vivares, C.P., Prensier, G., 2001. The identification of rRNA maturation sites in the microsporidian *Encephalitozoon cuniculi* argues against the full excision of presumed ITS1 sequence. J Eukaryot. Microbiol. Suppl. 60S-62S.
- Rimour, S., Hill, D., Militon, C., Peyret, P., 2005. GoArrays: highly dynamic and efficient microarray probe design. Bioinformatics 21, 1094-103.
- Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., Collado-Vides, J., 2000 Operons in *Escherichia coli*: genomic analyses and predictions. Proc. Natl. Acad. Sci. U. S. A. 97, 6652-7.
- Spieth, J., Brooke, G., Kuersten, S., Lea, K., Blumenthal, T., 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell 73, 521-32.
- Stover, N.A., Steele, R.E., 2001. Trans-spliced leader addition to mRNAs in a cnidarian. Proc. Natl. Acad. Sci. U. S. A. 98, 5693-8.
- Vanhamme, L., Pays, E., 1995. Control of gene expression in trypanosomes. Microbiol. Rev. 59, 223-40.
- Williams, B.A., Slamovits, C.H., Patron, N.J., Fast, N.M., Keeling, P.J., 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc. Natl. Acad. Sci. U. S. A. 102, 10936-41.

Table 1. Determination of the minimum sequence length that separates two operons.

5' RACE-PCRs were conducted on the most distal CDS of putative operons. Multiple PCR fragments amplification indicates a polycistronic organization mentioned by YES. Generation of only one PCR product indicates a separation of the CDS in a distinct operon (NO). Distance between the two most distal CDS in a potential operon is indicated in bp. The CDSs name and the chromosomal location are also mentioned.

Upstream CDS	Downstream CDS	Space between consecutive CDS bp	Operon structure
HYPOTHETICAL PROTEIN (ECU01_0860: CAD24956)	TREHALOSE-6-PHOSPHATE PHOSPHATASE (ECU01_0870: CAD24957)	150	YES
GLYCEROL 3-PHOSPHATE DEHYDROGENASE (ECU05_0270: CAD26543)	PHOSPHOMANNOMUTASE (ECU05_0260: CAD26542)	185	YES
SPORE WALL PROTEIN 1 (ECU10_1660: CAD25887)	ATP-DEPENDENT RNA HELICASE (ECU10_1680: CAD25888)	200	YES
20S PROTEASOME ALPHA SUBUNIT (C2) (ECU11_1670: CAD26077)	RING ZINC FINGER TRANSCRIPTION NEGATIVE REGULATOR FACTOR (ECU11_1660: CAD26076)	249	YES
DNA-DIRECTED RNA POLYMERASE II 19kDa SUBUNIT (ECU07_1750: CAD215706)	MADS domain (ECU07_1740: CAD25705)	270	YES
HYPOTHETICAL PROTEIN (ECU02_1390)	similarity to HYPOTHETICAL WD-REPEAT PROTEIN YN57_yeast (ECU02_1400)	309	NO
GLUTAMYL tRNA SYNTHETASE (ECU02_1210: CAD25150)	similarity TO HOMEBOX DOMAIN-CONTAINING PROTEIN (ECU02_1200: CAD25149)	312	YES
CTP SYNTHASE (ECU11_0880: CAD25997)	U5 ASSOCIATED snRNP (ECU11_0870: CAD25998)	318	NO
VACUOLAR ATP SYNTHASE SUBUNIT B (ECU10_1040: CAD25823)	ATP-binding domain (ECU10_1050: CAD25824)	322	YES
TRANSCRIPTION INITIATION FACTOR TFIID 20/15kDa SUBUNIT (ECU02_1500: CAD25179)	ALANYL tRNA SYNTHETASE (ECU02_1490: CAD25178)	342	NO
STRUCTURE-SPECIFIC RECOGNITION PROTEIN (ECU07_1010: CAD25634)	60S RIBOSOMAL PROTEIN L37A (ECU07_1005: CAD25633)	345	YES
DOLICHOL-PHOSPHATE MANNOSYLTRANSFERASE (ECU04_1060: CAD25294)	DNA TOPOISOMERASE III (ECU04_1070: CAD25295)	364	YES
similarity to HYPOTHETICAL PROTEIN YM63_CAEEL (ECU01_0360: CAD24905)	similarity to CHROMODOMAIN HELICASE DNA-BINDING PROTEIN 2 (ECU01_0350: CAD24904)	368	NO
60S RIBOSOMAL PROTEIN L30 (ECU05_1490: CAD26669)	similarity to HYPOTHETICAL WD-REPEAT PROTEIN YER2_yeast (ECU05_1500: CAD26670)	379	YES
HYPOTHETICAL PROTEIN (ECU10_1030: CAD)	VACUOLAR ATP SYNTHASE SUBUNIT B (ECU10_1040: CAD)	394	NO

HYPOTHETICAL PROTEIN (ECU04_0410: CAD25228)	26S PROTEASOME REGULATORY SUBUNIT 12(ECU04_0420: CAD25229)	399	NO
BELONGS TO THE ABC TRANSPORTER SUPERFAMILY, GCN20 HOMOLOG (ECU05_1190: CAD26639)	similarity to HYPOTHETICAL TRANSMEMBRANE PROTEIN YSV4_CAEEL (ECU05_1180: CAD26638)	421	NO
SER/THR PROTEIN PHOSPHATASE PP2C HOMOLOG 1 (ECU07_1520 ;CAD25683)	CATION-TRANSPORTING ATPase (ECU07_1500 ;CAD25682)	429	NO
similarity to HYPOTHETICAL TRANSMEMBRANE PROTEIN YADA (ECU11_1700: CAD26080)	similarity to Myb-related transcription factor (ECU11_1690: CAD26079)	502	NO
ZINC PROTEASE (INSULINASE FAMILY) (ECU06_0750: CAD25435)	MEI2-RELATED PROTEIN (ECU06_0740 : CAD25434)	552	NO
similarity to HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN YQ55_CAEEL (ECU03_1540: CAD26297)	similarity to HELICASE MOT1 (ECU03_1530: CAD26296)	596	NO

I. C. DISCUSSION

La mise en évidence de gènes orientés dans le même sens et séparés par des espaces intergéniques de courte taille laissait supposer l'existence, chez *E. cuniculi*, d'un mécanisme transcriptionnel original. Nos résultats montrent qu'une organisation polycistronique des gènes est retrouvée chez l'espèce *E. cuniculi*. Au cours du processus de transcription, chaque unité transcriptionnelle subit une maturation par clivage aléatoire aboutissant à la formation d'une population d'ARNm comprenant des ARNm monocistroniques et polycistroniques. Chacun des cistrons possède une coiffe et une queue poly(A). Ce processus ne fait pas intervenir un processus d'épissage en *trans* comme chez les trypanosomes et *C. elegans* (Blumenthal et Gleason, 2003).

L'étude des régions intergéniques permettrait de mieux cibler les portions d'ADN génomique où sont localisées les régions régulatrices (promoteurs et terminateurs), mais également, concourrait à l'identification des motifs nucléiques consensus nécessaires à ces régulations. Ce travail nécessite tout d'abord d'extraire ces régions potentiellement régulatrices des unités transcriptionnelles. Corrélé à l'étude du niveau d'expression de chaque opéron par une approche biopuce à ADN, un classement des régions promotrices pourra être envisagé. Ce classement sera réalisé en fonction du niveau d'expression de chacune de ces régions mais également en fonction du moment où l'expression est déclenchée pendant le cycle cellulaire du parasite. En effet, l'expression contrôlée de certains gènes en fonction du stade de développement parasitaire ou encore en réponse aux variations environnementales n'est pas à exclure, comme cela a déjà été démontré chez des protozoaires parasites tels que *L. major* (Rochette et al., 2005 ; Dumas et al., 2006), *P. falciparum* (Kyes et al., 2003 ; Sharp et al., 2006), *T. cruzi* (Gomes et al., 2004 ; Perez-Diaz et al., 2007), *Giardia* sp. (Que et al., 1996 ; Hehl et al., 2000) et *Entamoeba* sp. (Sanchez et al., 1994 ; Gilchrist et al., 2006 ; Ehrenkaufer et al., 2007).

Dès lors, une synchronisation du cycle cellulaire du parasite ou tout du moins une extraction de l'ADN correspondant aux différents stades de développement du parasite sera donc nécessaire.

Une analyse à l'aide du logiciel MEME pourrait être envisagée afin de caractériser une séquence promotrice minimum (**Bailey et Elkan 1995**). Ce logiciel permet en effet, à partir d'un jeu de séquences, de définir un ou plusieurs motifs communs à ces séquences. Il sera alors possible d'identifier une séquence promotrice correspondant à un niveau d'expression privilégiée, et à un stade parasitaire donné.

La caractérisation de telles régions régulatrices est nécessaire à la création de constructions en transgénèse pour des approches de génomique fonctionnelle. Il est probable, de ce fait, que les résultats infructueux en matière de transfection chez les microsporidies (*E. cuniculi*) soient la conséquence d'un mauvais choix des régions régulatrices. Le développement d'un système de transfection efficace chez *E. cuniculi* permettra, par la suite, d'analyser plus finement les signaux et les mécanismes de régulation de la transcription.

La faible taille supposée des régions promotrices laisse également envisager la possibilité d'utiliser des promoteurs bidirectionnels comme cela est le cas chez *P. falciparum* (**Fernandez-Becarra et al., 2003**). Ainsi, aussi bien au niveau de régions intergéniques de courte taille séparant deux gènes divergents, qu'en amont de certains gènes nécessitant un fort niveau d'expression, nous avons déjà pu caractériser une séquence promotrice potentielle correspondant au palindrome AAAATTTT. Ce promoteur putatif est actuellement testé dans différentes constructions afin de valider sa fonctionnalité.

Chez *E. cuniculi*, les mécanismes de régulation de la transcription pourront être également appréhendés grâce à la conception d'une biopuce à ADN dédiée à la caractérisation d'ARNs antisens. Différents résultats laissent penser qu'une telle régulation par ARNs antisens existe chez les microsporidies. En effet, le séquençage d'une banque d'ESTs

correspondant à l'espèce *Antonospora locustae* a permis d'isoler différents ADNc correspondant à des ARN antisens (**Williams et al., 2005**). Chez *E. cuniculi*, des expérimentations de 3' RACE PCR ont également permis de montrer que les extrémités 3' de deux gènes convergents se chevauchent. L'hybridation d'une telle biopuce avec des ARNs extraits à partir d'une cinétique d'infestation du parasite permettra donc de mettre en évidence ces ARNs antisens et de suivre leurs variations d'expression au cours du cycle en fonction du niveau d'expression des ARNs sens correspondant.

De nombreux exemples de transcrits antisens naturels ont été décrits à la fois chez des eucaryotes et des procaryotes, mais aussi chez les virus à ARN et à ADN. Ainsi, de telles molécules représentent plus de 20% des transcrits chez *G. lamblia* (**Elmendorf et al., 2001**). Ces auteurs ont suggéré que ce taux anormalement élevé était dû à une perte du mécanisme de contrôle du processus de transcription entraînant l'utilisation de promoteurs cryptiques. Cependant, les travaux décrits par **Von Allmen et al. (2005)** montrent que la présence des ARNs antisens semble régulée. En effet, leur synthèse ne se fait pas de manière constitutive mais est directement liée à la production de l'ARN sens correspondant. Chez l'espèce *P. falciparum*, les ARNs antisens sont également fortement représentés (**Gunasekera et al., 2004**). **Militello et al. (2005)** proposent que leur synthèse soit dépendante de la structure de la chromatine. Ces ARNs antisens, comme c'est le cas chez d'autres organismes, peuvent réguler la transcription des ARNs sens en entrant en compétition avec les facteurs de transcription. Ils peuvent également participer à la régulation de la traduction ou à la stabilité des ARNs suite à leur association avec les ARNs sens (**Carmichael, 2003**). Des ARNs antisens ont également été décrits chez les kinétoplastidés (**Martinez-Calvillo et al., 2003 ; Monnerat et al., 2004**).

Enfin, il serait important de vérifier si ce nouveau système de régulation transcriptionnelle résulte d'une forte compaction du génome. En effet, chez d'autres

organismes caractérisés par une forte réduction de la taille de leur génome, la transcription polycistronique semble être présente. C'est notamment le cas pour les nucléomorphes des espèces *Guillardia theta* et *Bigelowiella natans* caractérisées respectivement par des génomes de 551 et 380 kb (Williams *et al.*, 2005). Cette transcription polycistronique existe également chez *A. locustae*, une autre espèce microsporidienne dont le génome présente une taille de 5,4 Mpb (Williams *et al.*, 2005). Chez cette espèce, les auteurs ont également décrit des transcrits pouvant englober plusieurs gènes n'ayant pas la même orientation en raison d'une perte de signaux de fin de terminaison.

Des expérimentations de 5' et 3' RACE-PCR devront être réalisées sur des espèces microsporidiennes caractérisées par des génomes de plus grande taille. L'espèce *B. algerae* semble appropriée pour ce genre d'étude, car cet organisme est constitué du génome le plus grand connu à ce jour pour une microsporidie (génome haploïde estimé à 23 Mpb) (Belkorchia *et al.*, 2007, soumis). Le séquençage de ce génome, en cours, permettra de cibler les gènes regroupés potentiellement en opéron.

Une telle organisation génomique, définie par une transcription polycistronique et une maturation aléatoire, pourrait refléter des mécanismes d'adaptation mis en place par *E. cuniculi* au cours de l'évolution en raison de son statut de parasite intracellulaire obligatoire.

II. *Brachiola algerae* : microsporidie montrant de remarquables capacités d'adaptation à la vie parasitaire et présentant un génome complexe de 23 Mpb

A. SYNOPSIS

Le phylum des microsporidies regroupe plus de 1200 espèces réparties en 143 genres. Ces organismes intracellulaires obligatoires sont capables d'infester tout le règne animal, des Invertébrés aux Vertébrés, avec une préférence non négligeable pour les arthropodes. La température de développement des microsporidies d'insectes a longtemps été considérée comme un des facteurs limitant la transmission parasitaire vers les mammifères (**Trammer, 1997**). Le genre *Brachiola* regroupe plusieurs espèces parasites d'insectes. Celles-ci se distinguent des espèces appartenant au genre *Nosema* par leur capacité de développement à des températures plus élevées. L'une d'elle, *B. algerae*, démontre des capacités d'infestation et d'adaptation métabolique surprenantes. Compte tenu de son implication croissante en pathologie humaine, *B. algerae* se présente aujourd'hui comme une véritable espèce opportuniste en émergence (**Visvesvara et al., 1999 ; Coyle et al., 2004 ; Cali et al., 2005**).

Dans le cadre d'études comparatives des mécanismes d'adaptations de cette microsporidie sur plusieurs modèles cellulaires d'insectes et de mammifères, nous avons développé au laboratoire deux modèles cellulaires d'étude des interactions *B. algerae*-*An. gambiae*, et mis en évidence des capacités de développement au sein d'une lignée hémocytaire normalement impliquée dans les défenses immunitaires. Compte tenu de son spectre d'hôtes relativement large et de ses capacités d'adaptation lui assurant un développement dans plusieurs types cellulaires, *B. algerae* se présente comme un modèle microsporidien parasitaire d'excellence. De même, afin d'approfondir la connaissance de ce pathogène opportuniste, le séquençage de son génome a été initié par notre équipe de recherche en collaboration avec le Genoscope. La résolution de son caryotype moléculaire a révélé un génome composé d'au moins 30 bandes d'ADN chromosomique. La taille de son

génomique haploïde, estimée à 23 Mpb, est la plus importante connue à l'heure actuelle pour une microsporidie.

Les premières données d'annotation ont déjà révélé l'existence de plus de 3500 CDS. L'existence d'un répertoire génique aussi conséquent pourrait expliquer les facultés d'adaptation très importantes de *B. algerae* notamment son large spectre d'hôtes.

Les premiers assemblages de séquences ont permis de reconstruire un contig de 12269 pb sur lequel ont été caractérisées deux unités transcriptionnelles ribosomales (ADNr) de même orientation. L'analyse des régions intergéniques en aval de chaque unité a révélé l'existence de motifs répétés. Ces régions, très conservées, laissent supposer l'existence de plus de deux unités ADNr en tandem. L'hybridation du caryotype moléculaire avec une sonde ciblant l'ARNr 16S révèle la présence de ces unités sur au moins 9 des 30 bandes d'ADN chromosomique. La reconstruction de la structure secondaire de l'ARNr 23S a été réalisée et montre une conservation globale du core propre à ce groupe phylogénétique. L'analyse des premiers jeux de données a également permis la mise en évidence de nombreuses classes d'éléments transposables.

B. PUBLICATION 2: « *In vitro* propagation of the microsporidian pathogen *Brachiola algerae* and studies of its chromosome and ribosomal DNA organization in the context of the complete genome sequencing project »

In vitro propagation of the microsporidian pathogen *Brachiola algerae* and studies of its chromosome and ribosomal DNA organization in the context of the complete genome sequencing project[☆]

Abdel Belkorchia^{a,b}, Corinne Biderre^{a,b}, Cécile Militon^{a,b}, Valérie Polonais^{a,c}, Patrick Wincker^d,
Claire Jubin^c, Frédéric Delbac^{a,c}, Eric Peyretailade^{a,b}, Pierre Peyret^{a,b,*}

^a Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal, 24 avenue des Landais, Campus des Cézeaux, 63177 Aubière Cedex, France

^b Génomique Intégrée des Interactions Microbiennes, France

^c Parasitologie Moléculaire et Cellulaire, France

^d Genoscope and CNRS UMR 8030, 2 rue Gaston Crémieux, 91057 Evry, France

Received 3 April 2007; received in revised form 2 July 2007; accepted 6 September 2007

Available online 20 September 2007

Abstract

Brachiola algerae has a broad host spectrum from human to mosquitoes. The successful infection of two mosquito cell lines (Mos55: embryonic cells and Sua 4.0: hemocyte-like cells) and a human cell line (HFF) highlights the efficient adaptive capacity of this microsporidian pathogen. The molecular karyotype of this microsporidian species was determined in the context of the *B. algerae* genome sequencing project, showing that its haploid genome consists of 30 chromosomal-sized DNAs ranging from 160 to 2240 kbp giving an estimated genome size of 23 Mbp. A contig of 12,269 bp including the DNA sequence of the *B. algerae* ribosomal transcription unit has been built from initial genomic sequences and the secondary structure of the large subunit rRNA constructed. The data obtained indicate that *B. algerae* should be an excellent parasitic model to understand genome evolution in relation to infectious capacity.

© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: *Brachiola algerae*; Microsporidia; Molecular karyotype; rDNA transcription unit; LSU rRNA gene secondary structure; Insect cell lines

1. Introduction

Microsporidia are obligate intracellular parasites which can infect a wide variety of organisms including invertebrates (particularly insects) and vertebrates. Recent phylogenetic analyses have suggested that these pathogens have a fungal origin, a placement consistent with their possession of the shared traits of closed mitosis and spores that contain chitin and trehalose [1].

Amongst the 150 or so Microsporidian genera reported today, 7 are able to cause opportunistic infections in acquired immunodeficiency syndrome (AIDS) patients and can lead to several syndromes in immunocompetent hosts [2,3]. In the

genus *Brachiola*, all three species characterized so far cause opportunistic infections: *B. connori* [4], *B. algerae* [5] and the type species *B. vesicularum* which was identified from an AIDS patient with myositis [6]. A developmental feature of the genus *Brachiola* is the "thickened" plasmalemma which is the result of electron dense secretions on the parasite surface produced from the sporoplasm through to the spore stage. Throughout the life cycle, members of this microsporidian genus possess a diplo-karyotic apparatus which consists of two morphologically identical adjacent nuclei that divide synchronously and equally. It is supposed that the paired nuclei never recombine and therefore evolved independently. Another important feature is that all developmental parasitic stages divide and differentiate in direct contact with the host cell cytoplasm without any parasitophorous vacuole (reviewed in [7]).

The species *Brachiola algerae* (formerly named *Nosema algerae* [8] and recently transferred to the genus *Anncaliia* [9] was originally isolated from larvae of the *Anopheles stephensis*

[☆] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

* Corresponding author. Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal, 24 avenue des Landais, Campus des Cézeaux, 63177 Aubière Cedex, France. Tel.: +33 473407604; fax: +33 4 73407670.

E-mail address: pierre.peyret@univ-bpclermont.fr (P. Peyret).

mosquito and was later found in four other mosquito species. Since then, it has been shown that this microsporidian species had one of the broadest *in vivo* and *in vitro* host ranges known, infecting many different species of insects and capable of development in vertebrate hosts. Indeed, *in vitro* studies demonstrated that *B. algerae* had the ability to infect and successfully grow in a wide variety of cells from both invertebrate and vertebrate hosts at various temperatures from 24 to 38 °C [10]. In humans several reports indicate that *B. algerae* is the causative agent of multiple superficial and deep tissue infections both in immunocompetent [11,12] and immunocompromised patients [13,14]. Furthermore, some studies have shown that *B. algerae* infection of mosquitoes parasitized by *Plasmodium falciparum* leads to the reduction in development of the causative agent of malaria, suggesting a possible biological control [15].

Although *B. algerae* is a microsporidian model of interest because of its ability to develop *in vitro* in several cell lines and to produce spores with a size ($4.4 \times 2.8 \mu\text{m}$) larger than those of *Encephalitozoon* sp., as yet very little molecular data is available and this data largely comprises of SSU rRNA sequences [16,17]. The sequencing of the whole genome of *B. algerae*, in which our lab is involved, is in progress (http://www.cns.fr/externe/English/Projets/Projet_KI/KI.html) and should be helpful not only for a better understanding of its genome organization, but also to understand specific host–parasite relationships. At the time of writing, the *B. algerae* genomic database consists of 130,000 shotgun reads, totalling a threefold coverage of the haploid genome. The present study deals with the karyotyping of this species by contour-clamped homogeneous field (CHEF) gel electrophoresis, the determination of the complete rRNA transcription unit sequence and the construction of the secondary structure of the LSU rRNA gene. In spite of many papers concerning the *in vitro* propagation of this parasite, very few studies have been conducted to establish *B. algerae* in mosquito cell lines. In this paper, we demonstrate the possible development of *B. algerae* in two *Anopheles gambiae* lines, Mos55 and Sua 4.0.

2. Materials and methods

2.1. Cell cultures

Human foreskin fibroblasts (HFF) were maintained in a culture medium of L-15 (Leibovitz, LifeTechnologies) containing 10% foetal calf serum (FCS) and 2 mM L-glutamine, and grown at 37 °C in a 5% CO₂ atmosphere. Two mosquito cell lines from *A. gambiae*, Sua 4.0 (kindly supplied by F. Catteruccia, Imperial College of Science, London) and Mos55 (kindly supplied by E. Saraiva, University of Federal do Rio de Janeiro, Brazil) were grown at 25 °C in Schneider's medium (Life Technologies) supplemented with 10% FCS.

2.2. Spore propagation

The *B. algerae* isolate used in this study was kindly provided by Pr. W.A. Maier (University of Sigmund-Freud, Bonn) and

corresponds to the microsporidium isolated by Dr. A. Undeen from its original mosquito host (*Anopheles stephensi*). HFF cells were grown to confluence, and 200 μl of *B. algerae* spores (10^8 ml^{-1}) were added to each of the cell cultures. Spores were collected from infected cell cultures weekly. Collected culture medium was centrifuged at 6000 *g* for 2 min, the supernatant was removed, and the pelleted spores were resuspended in sterile distilled water and stored at 4 °C.

2.3. Antibody production

Antisera to *B. algerae* spore proteins were produced in SWISS mice. *B. algerae*-containing HFF cells were scraped off six weeks after inoculation. Spores were treated with 1% sodium dodecyl sulphate (SDS) for 10 min at 50 °C and washed with phosphate-buffered saline (PBS). Purified spores were fixed with 4% paraformaldehyde–0.1% glutaraldehyde for 20 min at room temperature (RT). Spores were then disrupted by boiling for 15 min followed by 10 freeze–thaw cycles in liquid nitrogen. Samples ($\sim 2 \cdot 10^6$ spores for each injection) were homogenized with Freund's adjuvant for the first injection and Freund's incomplete adjuvant for the next injections (days 14, 21 and 28). Mice were injected intraperitoneally with the above samples and sera were collected 1 week after the last injection and stored at –20 °C.

2.4. Indirect immunofluorescence assay

B. algerae thin smears were prepared on polylysine-treated cover glasses. Slides were air dried and *B. algerae* spores were fixed with 100% methanol for 30 min at –80 °C. Slides were saturated with PBS-5% skimmed milk for 1 h and incubated with mouse antisera at 1:200 in PBS-0.1% Triton X-100 for 1 h at RT. After washing three times with PBS-0.1% Triton X-100, slides were incubated with a dilution of 1:1000 Alexa 488-conjugated goat anti-mouse IgG (Sigma) for 1 h at RT.

Mosquito and human cell lines were grown to 80% confluence on glass coverslips in 6-well plates and $3 \cdot 10^6$ freshly purified spores were added to each well. Fifteen days after this inoculation, cells were fixed with 100% cold methanol at –80 °C and then permeabilized with PBS-0.5% Triton X-100 for 30 min at RT. Preparations were saturated with PBS-5% skimmed milk for 1 h and incubated with mouse antisera at 1:200 in PBS-0.1% Triton X-100 for 1 h at RT. After several washes in the same buffer, the cells were then incubated with a dilution of 1:1000 Alexa 488-conjugated goat anti-mouse IgG (Sigma) for 1 h at RT. 4,6-Diamidino-2-phenylindole (DAPI) staining was used to visualise genomic DNA in the diplokaryon nuclei of extra- and intracellular parasites as well as host cell nucleus. Preparations were then examined with a Leica epifluorescence microscope. Images were analyzed using Adobe Photoshop 7.0 software.

2.5. Spore purification and plug preparation

Purification of *B. algerae* spores and preparation of pulsed field gel electrophoresis (PFGE) plugs was performed according

to Biderre et al. [18]. Briefly, *B. algerae* infected-HFF cells were scraped off 6 weeks post-inoculation and passed 3 times through a 27-gauge needle to liberate intracellular spores. Free spores were purified from the disrupted host cell fractions by 3 successive centrifugations at 1000 *g*. To remove host DNA contamination, the spore pellet was incubated with 20 U of Dnase I (Boehringer, Mannheim) at 30 °C for 30 min. Enzymatic activity of DNase I was inhibited by addition of 1/10 volume of 0.2 M EDTA. Spores were washed 3 times in PBS. The spore pellet was mixed with 1% boiled Gigaphor (Eurobio) agarose (6.10^8 spores / plug) and cooled at 37 °C in a plug mould. The plugs were chilled on ice and incubated in a lysis buffer consisting of 500 mM EDTA, 1% *N*-lauryl sarcosine, 2 mg ml⁻¹ proteinase K, for 48 h at 42 °C. The plugs were stored at 4 °C until used.

2.6. PFGE

Contour-clamped homogeneous electric field PFGE was performed using the CHEF-DR III (Biorad) system. Separation of chromosomes was achieved using a 1.5% Gigaphor (Eurobio) agarose gel in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8.0) at 14 °C with a field angle of 120°. Chromosomal DNA bands of *Saccharomyces cerevisiae* strain YNN295 (Biorad) were used as molecular markers. Five different running conditions were necessary to obtain a fine resolution for each chromosomal size class. Separation of a full set of *B. algerae* chromosomal DNA bands was achieved using a field strength of 5.5 V/cm, a pulse ramp from 60 to 120 s and a running time of 48 h. For separation of chromosomes smaller than 400 kbp, the running time was 40 h using a pulse ramp from 15 to 26 s during 16 h and from 20 to 29 s during 24 h, and a 5.5 V/cm field strength. For chromosomes between 300 and 700 kbp, the running time was 42 h using a pulse ramp from 66 to 71 s during 21 h and from 56 to 63 s during 21 h, and a 5 V/cm field strength. For chromosomes between 500 and 1000 kbp, the running time was 49 h with a pulse ramp from 70 to 79 s during 28 h and from 56 to 63 s during 21 h, and a 5 V/cm strength field. For chromosomes larger than 900 kbp, the following running conditions were used: 48 h using a 5 V/cm field strength, and a pulse ramp from 120 to 145 s. The gels were stained for 1 h in ethidium bromide solution at a final concentration of 0.4 µg ml⁻¹. Southern blotting of *B. algerae* molecular karyotypes were performed on Nylon membrane (Ambion) and hybridizations were performed according to Biderre et al. [19] with an SSU rRNA probe amplified by polymerase chain reaction (PCR). The forward primer 5'-AGAGCGTATTCAGTGGTC-3' and the reverse primer 5'-GCATAGAGCACATCACAC-3' were used, yielding a 373 bp DNA fragment.

2.7. rDNA subunit identification

Ribosomal RNA subunit sequences were identified by analysis of data generated by the *B. algerae* whole genome sequencing project which was initiated in collaboration between Genoscope (Evry, France) and the lab "Biologie des Protistes"

(UMR CNRS 6023, France). One hundred and thirty thousand initial random shotgun reads were obtained by sequencing clones from two plasmid libraries (inserts of 3 and 10 kbp) and a miniBAC library (inserts of about 25 kbp) constructed with *B. algerae* genomic DNAs. Using the Staden package [20], a genomic fragment of 12,269 bp encompassing the complete sequence of the rDNA operon, a second partial LSU rRNA gene and flanking sequences of the transcription unit, was determined. The sequence was deposited in the EMBL database under the accession number AM422905.

2.8. LSU rRNA gene secondary structure construction

The secondary structure of the *B. algerae* LSU rRNA gene was constructed by a manual method. The helices in the LSU rRNA secondary structure elements were located and labelled based on a database of LSU rRNA secondary structure [21] while the hypervariable areas (V1–V12) were numbered in accordance with all known eukaryotic LSU rRNAs [22,23].

3. Results and discussion

3.1. *Brachiola algerae* can propagate in human and hemocyte-like cell lines

B. algerae is a favourable model for host–parasite interaction studies mainly because this microsporidium possesses the capacity to adapt to a wide range of hosts, being able to infect insects and mammals, notably humans [9]. A wide variety of cells have been used to produce *B. algerae* species *in vitro*. Mammalian cell lines are widely preferred mainly because this microsporidium has been identified as an opportunistic pathogen in immunosuppressed and immunocompetent humans and because *B. algerae* spores isolated from mosquitoes were able to grow efficiently in these cell types and produce viable spores [24]. No mosquito cell line for *B. algerae* propagation has been clearly reported in the literature. However, a *B. algerae*-mosquito cell line model system would be of great interest, in particular for host–parasite interaction studies as the complete human and *A. gambiae* genome sequences are now available. In this context we tested the ability of this microsporidium to develop *in vitro* within one human and two *A. gambiae* cell lines.

A mouse anti-*B. algerae* spore antiserum was used in immunofluorescence assays either with human (HFF) or insect (Mos55 and Sua 4.0) cell lines infected by *B. algerae* or with purified extracellular spores (Fig. 1). A strong fluorescent signal was observed with the cell surface of extracellular parasites (spores) fixed on glass slides (Fig. 1; 3A). Surprisingly, the antiserum was not reactive with extruded polar tubes. When applying antibodies to infected human cell cultures, spores were clearly identified within the host cell cytoplasm or released into the medium after host cell lysis. The extracellular spores gave a stronger fluorescent signal than the ones inside the host cell cytoplasm (Fig. 1; 3B, arrow). No fluorescence was seen for the host cells and the youngest developing stages of the parasite (meronts) (Fig. 1C and 3C, unfilled red arrows).

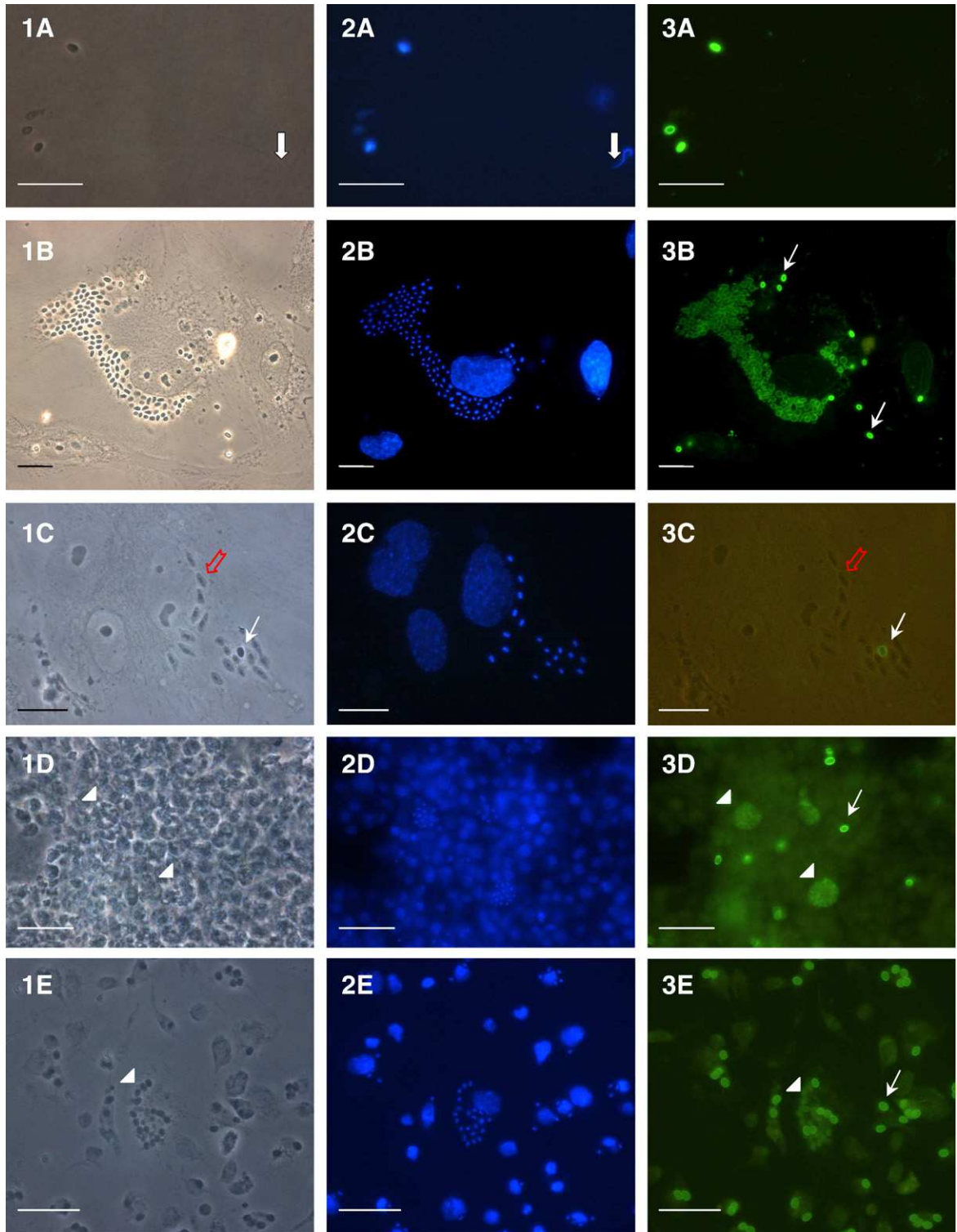


Fig. 1. Immunofluorescence assays with anti-*Brachiola algerae* spore antiserum. Row 1: micrographs taken by the phase-contrast microscope; row 2: samples stained by DAPI; row 3: samples labelled by indirect immunofluorescence assay staining with anti-*B. algerae* spore antiserum and then with goat anti-mouse IgG conjugated with Alexa 488. Line A corresponds to purified spores; lines B and C are *B. algerae* infected-HFF cells; lines D and E are *B. algerae* infected-Mos55 and -Sua.4 cells (*Anopheles gambiae* cell lines) respectively. The antiserum shows a strong positive reaction to *B. algerae* spores (white arrows; 3B, 3C, 3D and 3E) and some spores show a weak positive reaction (arrowheads; 3D and 3E), especially the intracellular spores. The developing stages in *B. algerae* infected-HFF cells show a negative reaction (unfilled red arrows; 3C). The plain arrows show a diplokaryotic nucleus in the extruded polar tube (1A and 2A). Scale: 20 μ m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We then tested two *A. gambiae* cell lines (Mos55 and Sua 4.0) for *B. algerae* development. Microscopy and immunofluorescence assay experiments show the presence of spores inside the insect cell cytoplasm (Fig. 1; 3D and E, arrowheads). These data suggest that *B. algerae* is able to develop in these two different cell types. However, a longer time is necessary to detect infected mosquito cells (10–15 days for Mos55 and Sua 4.0) compared to HFF cells (4 days). Furthermore, few spores seem to be produced with insect cell lines and released in the culture medium. Nevertheless, *B. algerae* seems to be able to infest and develop in hemocyte-like cells normally involved in the innate immune response. Understanding host–pathogen relationships should be of great interest to discover the mechanisms used by the pathogen to escape host defences.

3.2. Molecular karyotyping of *B. algerae* predicts a haploid genome size greater than 23 Mbp

The determination of the *B. algerae* molecular karyotype has been undertaken by PFGE within the framework of the *B. algerae* genome sequencing project. As shown in Fig. 2A this molecular karyotype exhibits a complex pattern with chromosomal-size bands ranging in size from 160 to 2240 kbp.

Due to co-migration of many fairly similar sized molecules, it was difficult to resolve all the chromosomal-sized DNA bands in a single separation step. Several variants of PFGE conditions were used to improve the resolution of four particular zones of the molecular karyotype as described in Fig. 2 from panels B to

E. Analysis of the entire set of electrophoregrams allowed us to recognize 30 chromosomal-sized DNA bands as schematized in Fig. 2F.

Amongst the 30 chromosomal DNA bands resolved by PFGE, 8 present a stronger intensity and 7 are less intense. Non-stoichiometric staining intensities of the chromosomal bands could be explained by several factors. The stronger staining intensity could result from the co-migration of heterologous DNA molecules of the same size and the faint one, from the presence of only a single copy of homologous chromosomes in the band, one of the homologs having a different size. Differences in staining intensities might also come from variations in the staining rate of ethidium bromide. A final explanation might be due to possible differential ploidy states in the individual nuclei of the diplokaryon, which have not been determined. Therefore, it is difficult to determine the exact chromosome number in a karyotype only on the basis of PFGE data particularly when the pattern of this karyotype is complex and where no molecular markers are available. Nevertheless, if we consider all the bands resolved, we can estimate the genome size of *B. algerae* at about 23 Mbp. This size range is in reasonable agreement with a recent study indicating that the haploid genome size of *B. algerae* should be between 15 and 20 Mbp [25]. Consequently, *B. algerae* is the microsporidian species with the largest genome, greater than the monokaryotic species *Glugea atherinae* whose genome was estimated at 19.5 Mbp [26].

Molecular karyotypes have been reported for 15 microsporidian species including 8 species parasitizing insects [27,28], and 3 specific to mammals belonging to the *Encephalitozoon*

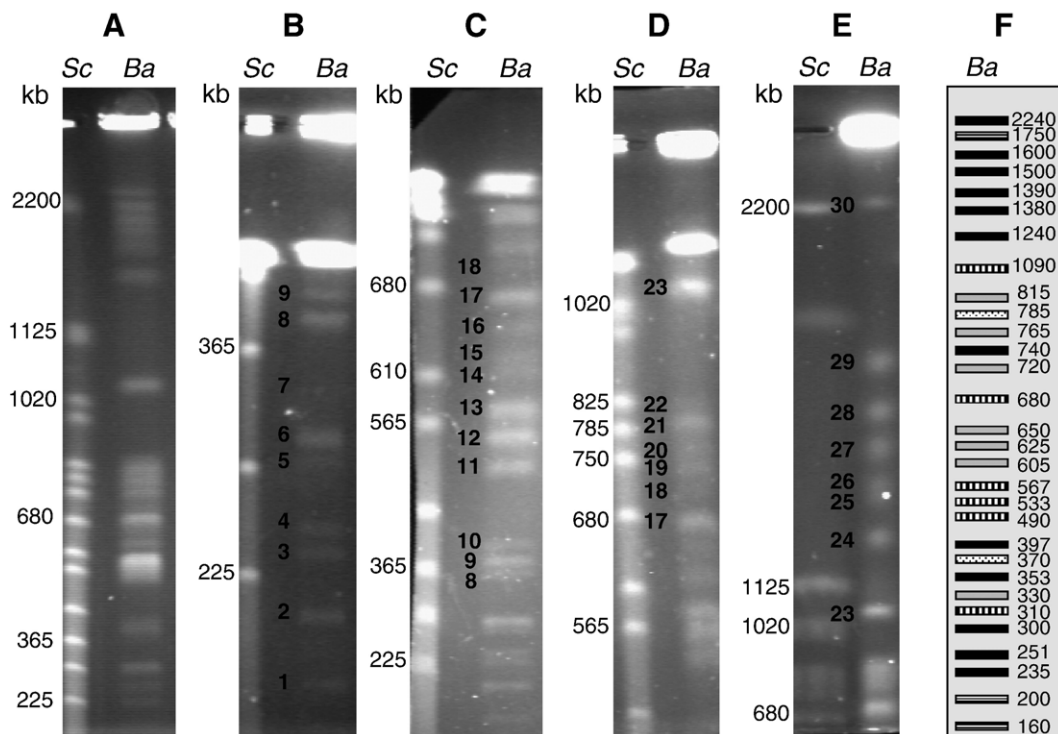


Fig. 2. Molecular karyotype of *Brachiola algerae*. The chromosome sizes of *Saccharomyces cerevisiae* YNN295 (Sc) are molecular markers. A: the chromosomal bands of *B. algerae* (Ba) showing the sizes from 160 and 2240 kb. B–E: the chromosomal bands in each zone. F: a diagram of the total number of bands found (30 bands), gray bands are less intense, dotted bands are of stronger intensity, horizontally and vertically striped bands are weakly and strongly hybridized with the SSU rDNA probe, respectively, black bands are of moderate intensity.

genus [18,29] (Table 1). Insect-infecting microsporidian molecular karyotypes differ widely in their number of chromosome-sized DNAs (from 8 chromosomal bands for *Vairimorpha* sp. and *N. costelytrae* to 31 for *Paranosema grylli*), in contrast, the karyotype of the mammal-infecting *Encephalitozoon* species microsporidia comprises of 10 to 12 chromosome-sized DNAs. Whereas *B. algerae* shows, with the species *P. grylli* a parasite of crickets [27], the greatest number of chromosomal bands (30 and 31 bands respectively), its chromosome size range is closer to that observed for the two fish microsporidium unikaryotic species belonging to the genus *Glugea* [26,30].

The genome sizes estimated so far for 15 microsporidia species are very diverse, the smallest one being estimated at 2.3 Mbp (*E. intestinalis*) [29] and the biggest one at 19.5 Mbp (*G. atherinae*) [26], and potentially more for *B. algerae* (Table 1). However, only the extremely reduced genome of *E. cuniculi* has been sequenced. This genome of 2.9 Mbp is distributed in 11 chromosomes and comprises of ~2000 protein-coding genes [31]. Analysis of the content of the *E. cuniculi* genome shows that its high compaction mainly results in the loss of many biosynthetic pathways. Partial genome sequencing of a second microsporidian species, *Antonosporea locustae*, which is estimated at 5.3 Mbp and thus of size twofold greater than that of *E. cuniculi*, also displays the same characteristic of genome compaction and stability [32]. The observed diversity in microsporidian genome size and content is highly surprising. Genome sizes can vary 8 fold between *E. cuniculi* and *B. algerae*. A possible explanation has recently been brought by Xu et al. [33] with the discovery of long-terminal repeat retrotransposon (*NBr*) elements in the genome of *N. bombycis*

(estimated at 15.3 Mbp). The different families of transposable elements are known to be at the origin of size variation and genome evolution of many eukaryotic organisms including protozoan parasites [34,35]. Using the dataset of *B. algerae* genomic shotgun reads, we have identified a high proportion of such elements which could correspond to several retrotransposon families but we have yet to characterize the groups to which they belong (data not shown). All these results strongly suggest that transposable elements might play a major role in size variation observed between the different microsporidian species.

Moreover, the fact that *B. algerae* infects a broad host spectrum from insects to mammals with the possibility of multiplication and differentiation over a large temperature range, supposes the capacity to adapt to various metabolisms. Such adaptations require the presence of all the genes implicated in these multiple metabolic pathways to ensure host infection and parasite multiplication. Indeed, the first results of the structural annotation of the *B. algerae* genome has allowed the identification of more than 3500 coding DNA sequences, which represent almost twice the number described for the *E. cuniculi* genome [31].

3.3. Characterization of a complete rDNA unit and evidence for the presence of multiple copies

The systematic sequencing project of the *B. algerae* genomic DNA has allowed us to determine the complete sequence of a contig of 12,269 bp which encompasses a whole rRNA transcription unit (SSU and LSU rRNA genes) with its flanking regions as well as a partial rDNA unit.

Table 1
Comparison of the size (kb) and the number of *Brachiola algerae* chromosomal DNA bands with those of the known microsporidian species

Species	Nucleus state	Host	Band number	Chrs number	Band size range	Genome size estimate (Mbp)
Mammal species						
<i>Encephalitozoon cuniculi</i>	Monokaryotic	Human, dog, rodents	11	11	217–315	2.9
<i>E. intestinalis</i>	Monokaryotic	Human	10		182–279	2.3
<i>E. hellem</i>	Monokaryotic	Human	7–12		175–315	2.4–2.9
Fish species						
<i>Glugea atherinae</i>	Monokaryotic	Atherine	16		420–2700	19.5
<i>G. stephani</i>	Monokaryotic	Flounder	15		340–2654	16.7
<i>Spraguea lophii</i>	Diplokaryotic	Angler	Min 10	15	230–1120	6.2–7.3
Crustacea species						
<i>Nadelspora canceri</i>	Monokaryotic	Crabs	10		280–1879	7.3–7.4
Insect species						
<i>Vairimorpha</i> sp.	Mono- and diplokaryotic		Min 8		720–1790	9.2
<i>Vavraia oncooperae</i>	Monokaryotic	Moth	14–16		130–1930	8–10.2
<i>Nosema costelytrae</i>	Diplokaryotic	Grass grub	Min 8		290–1810	7.4
<i>N. bombycis</i>	Diplokaryotic	Silkworm	18		380–1500	15.3
<i>N. furnacalis</i>	Diplokaryotic	Corn borer	13		440–1360	10.2
<i>N. pyrausta</i>	Diplokaryotic	Corn borer	13		440–1390	10.5
<i>Antonosporea locustae</i>	Diplokaryotic	Grasshopper	18		139–651	5.4
<i>Paranosema grylli</i>	Diplokaryotic	Cricket	31	17	135–485	4.5–4.7
Multiple-host species						
<i>Brachiola algerae</i>	Diplokaryotic	Human, rodent mosquitoes	30		160–2240	23

Chrs: chromosome; Mbp: Megabase pair; Min: minimum.

Table 2
Analysis of the repetitive sequences located in the 5' and 3' IGS (intergenic sequences) regions of the rRNA transcription unit

A							
	Sequence						
Rep1	TTATTAGAGTGGAGTAGGT						
Rep2	AATATTTT						
Rep3	TTGAATTTATAATGTTAAAATGAAAAGTGTATTTTGATTGGCTGAAAATTCATTGTTGAAACTCTAAAAAGTATTTTGATTGG						
Rep4	TTTATATATGTTGAAACTATAAAGAAGATTGATTGGTTGAT						
Rep5	TTATATAGGTTAAGTCTAGCTATTTCTCCCCCAAAGAAGCTTGAGGGGGGCGGGAAAATAGCTATCAATCAGAAC						
Rep6	CTAAACCTTGCTTATTTCCAAGATAGAATT						
Rep7	TTTATTGTAATAACAAACATGCAAAACCTAAACCACATTCAGAATTTAAGAAAGAATTGTTGAAAACATGAAA TGGAACCAAAAACCTAACCAATCAGAATTTGAAATACTAGATCTAATTTTGACCAATAAACATAAAA						

B							
	Rep1	Rep2	Rep3	Rep4	Rep5	Rep6	Rep7
Repeat size (nt)	19	8	83	42	76	30	140
Repeat number in 5'	2	2	11	0	2	2	3
Repeat number in 3'	4	2	5	2	2	1	6
Repeat organization	Tandem	Disjoined	Tandem	Tandem	Disjoined	Disjoined	Tandem
Repeat positions in 5'	588–618	631–638 655–662	681–1291		1477–1552 1573–1646	2172–2201 2413–2442	2609–2990
Repeat positions in 3'	9444–9512	9525–9532 9549–9556	9577–9687 9822–9853	9700–9816	9987–10,072 10,093–10,166	11,040–11,069	11,236–12,025

A: the sequences of seven repeats, Rep-1–Rep-7 identified in the 5' and 3' IGS regions. B: the characteristics of each repeat with their size and location in the 5' and 3' IGS regions. The number of repeats corresponds to the copy number of each repeat.

From the complete sequence of the rRNA unit, the SSU rDNA gene was estimated at 1391 nucleotides (positions 5462–6852), with a GC percent of 46.7%. It presents 99% homology to two complete SSU rRNA sequences of *B. algerae* present in GenBank (accession nos. AF069063 and AY230191). The internal transcript spacer (ITS) size was estimated at 22 nt (positions 6853–6874) from the secondary structure of the LSU rRNA gene, and its GC content at only 13.6%. The ITS size is thus lower than the 41 nt size reported previously (Genbank accession no. AY230191) [16].

Two LSU rRNA genes were identified in the contig sequence: a partial one of 384 nt (positions 1–384) and a complete one estimated at 2557 nt (positions 6875–9430). The GC content of the complete LSU rRNA gene is 50.4%, which is significantly higher than the GC content of species that belong to the *Nosema* genus in which *B. algerae* had been firstly classified (31.9% for *N. bombycis* (Genbank accession no. AY259631), 31.8% for *N. spodopterae* (Genbank accession no. AY211390), 32.9% for *N. ceranae* (Genbank accession no. DQ078785), 35.3% for *N. apis* (Genbank accession no. U90885)). In contrast, this percentage is closer to those determined for the species *E. cuniculi* (53%) (Genbank accession no. AJ005581) and *H. anguillarum* (58.9%) (Genbank accession no. AF387331). The LSU rRNA sequence

identities of *B. algerae* with *Microsporidium* 57864, *N. ceranae*, *H. anguillarum*, *N. bombycis* and *E. cuniculi* are 56.9, 58.8, 60.5, 61 and 63.7% respectively, using Needleman-Wunsch global alignment algorithm (<http://bioweb.pasteur.fr/seqanal/EMBOSS/>). The putative start and terminal regions of the LSU rRNA gene were determined by comparison with the microsporidian LSU rRNA sequences from databases and by manual secondary structure prediction. With 2557 nt, the *B. algerae* LSU rRNA gene represents the second longest sequence described so far for microsporidia, after that of *H. anguillarum* with 2664 nt [36].

The flanking regions of the whole rDNA transcription unit consist of a 5' intergenic spacer (IGS) region estimated at 5077 nt upstream of the SSU rRNA gene (positions 385–5461) and a 3' one of 2838 nt downstream of the LSU rRNA gene (positions 9431–12269). They are respectively 28.5% and 27% GC-rich, which contrasts with the GC percentage of the rRNA genes (see above). No open reading frame and no sequence homologous to either tRNA or 5S rRNA genes were detected outside the coding regions of SSU and LSU rRNA genes. Both IGS regions contain sequences with a very high homology with various identical, tandemly or interspersed repeated elements (Table 2, Fig. 3). Seven different elements from 8 to 140 nt have been characterized of which six are common between the two

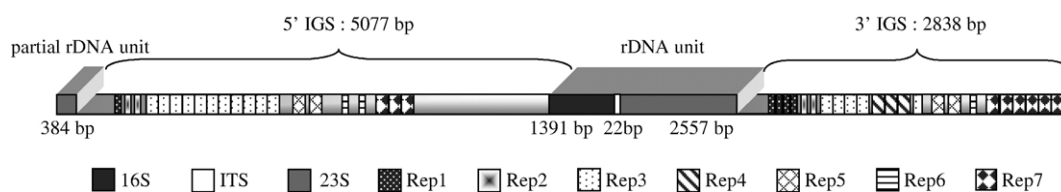


Fig. 3. Schematic representation of the distribution of the repetitive sequences in both the 5' and 3' intergenic regions of the rRNA transcription unit of *B. algerae*.

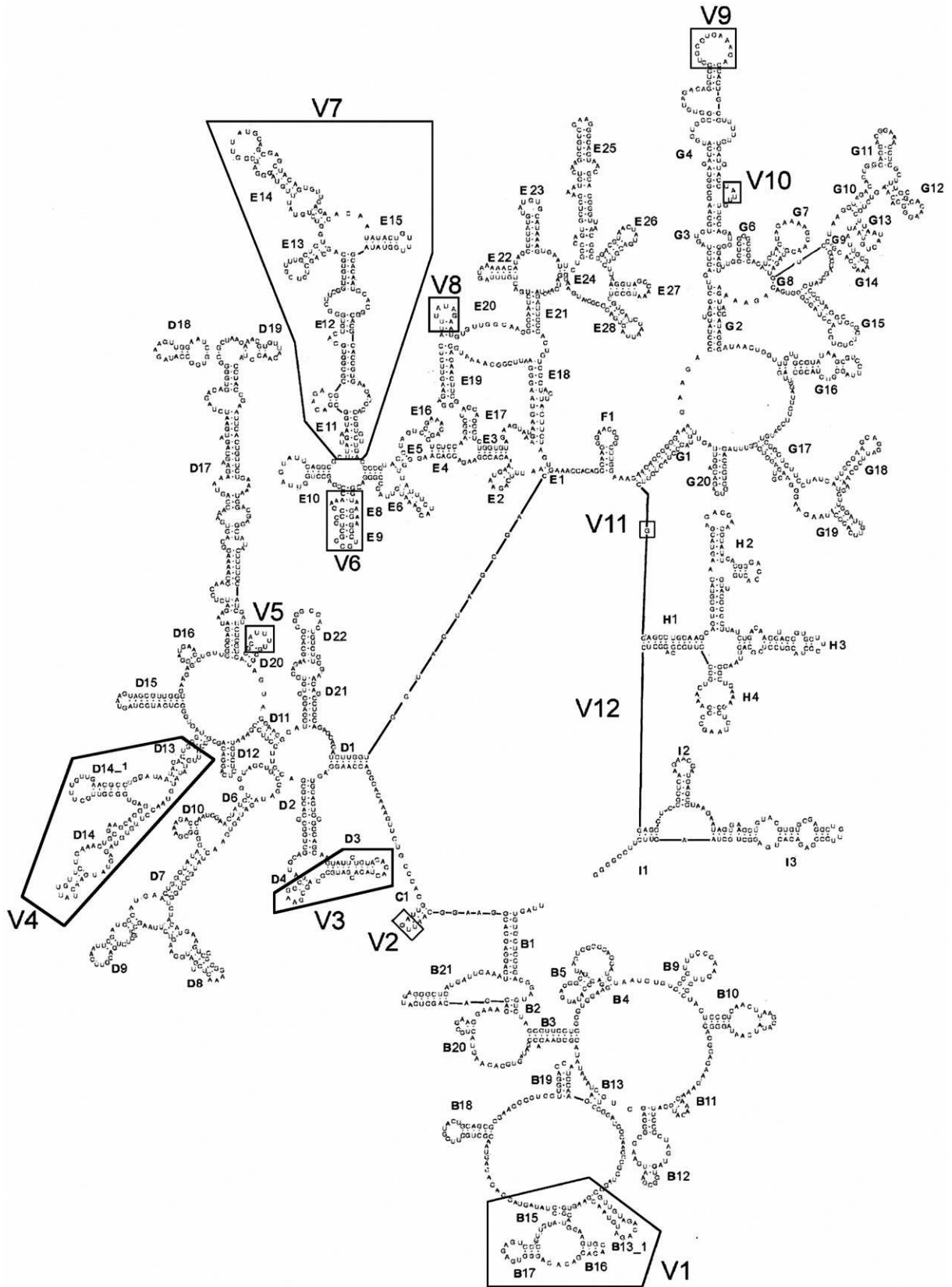


Fig. 4. Model of the secondary structure of *Brachiola algerae* large subunit (LSU) rRNA gene. Helix numbering is according to De Rijk et al. [21] and hypervariable area numbering (V1 to V12) is according to De Rijk et al. and Wuyts et al. [22,23].

regions. The elements named Rep1 (19 nt), Rep3 (83 nt), Rep4 (42 nt) and Rep7 (140 nt) are tandemly arranged with an alternation of perfect and imperfect copies. The number of these copies varies from 1 to 11 and is different according to the IGS region. The Rep4 element is only found in the 3' IGS region. Concerning the disjointed elements named Rep2 (8 nt), Rep5 (76 nt) and Rep6 (30 nt), two perfect copies were identified for each except for Rep6 which has a single copy in the 3' IGS region (Fig. 3).

The identification of a part of a second LSU rRNA gene in the DNA contig seems to indicate that the rRNA genes could be arranged in a head-to-tail tandemly repeated manner in *B. algerae*. Given the size of the 5' IGS region and the one of the complete rRNA repeat, we can suppose that the rRNA gene repeat unit of *B. algerae* is close to 9 kbp. This size is twice smaller than the one demonstrated for *N. apis* which presents the same head-to-tail arrangement for its rRNA gene repeats [37].

The *B. algerae* chromosomal bands separated by CHEF electrophoresis using different conditions were hybridized with an SSU rDNA probe (positions 6115–6488). The rDNA unit location is restricted to 9 chromosomal bands (numbers 1–2–6–11–12–13–17–23 and 29) (Fig. 1F of which 6 present a stronger intensity) (data not shown). Consequently, the number of chromosomal bands carrying ribosomal transcription units might be underestimated. The *B. algerae* rDNA unit distribution contrasts with scattered distribution over all chromosomes for *E. cuniculi* species [29] and location to a single large chromosome for *N. bombycis* [38]. This distribution on only some bands of the molecular karyotype is closer to that observed for the fish parasites *S. lophii* and *G. atherinae* [29].

The analysis of the 130,000 initial random shotgun reads has allowed us to isolate only one contig containing the rRNA transcription unit. However, given the haploid size estimate of *B. algerae* genome (23 Mbp), the 130,000 shotgun sequences represent a threefold genome coverage. This coverage for the rRNA contig is about 71×. Assuming that no bias in cloning was introduced, we estimate the number of rRNA unit is close to 25. Furthermore, individual analysis of the shotgun reads composing this contig does not reveal sequence variability. This result argues for a high conservation of this rRNA operon sequence and its flanking regions for all rRNA transcription units present on the different DNA chromosomal bands. Such conservation of the rRNA sequence and its flanking regions has also been observed for the 22 rRNA units of the *E. cuniculi* genome which are located in sub-telomeric position [39]. Such a location might be conceivable in *B. algerae* but with a head-to-tail arrangement.

3.4. Characterization of the LSU rRNA gene secondary structure

The secondary structure of the LSU rRNA gene of *B. algerae* is showed in Fig. 4 and is basically similar to that of the other microsporidian species. Eight groups (B to I) can be distinguished clockwise from a core area. Twelve hypervariable areas (V1–V12) are also shown in Fig. 4. Six helices (B6, B7, B8, B14, D5 and G5) are missing. Based on the secondary

structure of the eukaryotic LSU rRNA gene [40], 5 areas of the hypervariable regions are almost entirely missing (V2, V8, V10, V11 and V12) and 4 areas are extremely reduced (V1, V3, V5 and V6). Only 6 microsporidian LSU rRNA sequences have been reported so far and it is difficult to compare them. The secondary structure of *B. algerae* LSU rRNA gene is closer to that of *H. anguillarum* with the common presence of helices B21, E9, E15 which are absent in all other species studied so far [41]. However, in contrary to this last species, the *B. algerae* LSU rRNA gene does not have the G5 helix and has a less reduced D20 helix. The similarity of the conformation observed between these two species with conservation of a greater number of helices might be explained by the fact that they have the longest sequences. The *B. algerae* LSU rRNA gene diverges from all the other microsporidian species by the presence of the helix D14-1 in the V4 area.

In conclusion, our studies allow us to estimate the molecular size of the *B. algerae* haploid genome (23 Mbp) and to describe the complexity of this genome. This is an essential step to define the complete genome sequencing strategy and to determine the number of sequences necessary for the whole coverage of this genome. In this paper we report the first sequence organization derived from shotgun genome sequencing the description of more than 12 kbp which encompasses the rRNA unit. The completion of the genome sequence of *B. algerae* and its comparison to other microsporidian sequences will be also important for a better understanding of metabolic capacities and host–parasite relationships of this multi-host parasite. We have also shown that *B. algerae* is able to invade hemocyte-like cells from mosquitoes supposing that it could possess several strategies to escape host cell defences. In addition it is also important to note that the sequences of the genomes of *P. falciparum* (causative agent of malaria), the *Anopheles* mosquito (vector of *P. falciparum*) and human are now available and should provide interesting data for further post-genomic approaches.

Acknowledgments

We are grateful to Dr. Wyatt Paul for reviewing the English version of the manuscript. This work was supported by the CNRS. The sequencing at Genoscope was funded by CNRG. A. B. was supported by a grant from “Communauté d’Agglomération d’Aurillac” and C.B by a grant from “Auvergne Council”.

References

- [1] James JY, Kauff F, Schoch CL, et al. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 2006;443:818–22.
- [2] Desportes-Livage I, Le Charpentier A, Galian A, et al. Occurrence of a new microsporidian: *Enterocytozoon bienewisi* n.g., n.sp. in the enterocytes of a human patient with AIDS. *J Protozool* 1985;32:250–4.
- [3] Snowden KF. Zoonotic microsporidia from animals and arthropods with a discussion of human infections. In: Lindsay DS, Weiss LM, editors. *Opportunistic infections: toxoplasma, sarcocystis, and microsporidia*. Boston, Mass: Kluwer Academic Publishers; 2004. p. 123–34.
- [4] Sprague V. *Nosema connori* n. sp., a microsporidian parasite of man. *Trans Am Microsc Soc* 1974;93:400–3.
- [5] Lowman PM, Takvorian PM, Cali A. The effects of elevated temperatures and various time–temperature combinations on the development of

- Brachiola (Nosema) algerae* N. Comb. in mammalian cell culture. J Eukaryot Microbiol 2000;47:221–34.
- [6] Cali A, Takvorian PM, Lewin S, et al. *Brachiola vesicularum*, n. g., n. sp., a new microsporidium associated with AIDS and myositis. J Eukaryot Microbiol 1998;45:240–51.
- [7] Takvorian PM, Weiss LM, Cali A. The early events of *Brachiola algerae* (Microsporidia) infection: spore germination, sporoplasm structure, and development within host cells. Folia Parasitol 2005;52:118–29 (Prague).
- [8] Vavra J, Undeen AH. *Nosema algerae* n. sp. (Cnidospora, Microsporida) a pathogen in a laboratory colony of *Anopheles stephensi* Liston (Diptera, Culicidae). J Protozool 1970;17:240–9.
- [9] Franzen C, Nasonova ES, Scholmerich J, Issi IV. Transfer of the members of the genus *Brachiola* (microsporidia) to the genus *Anncaliia* based on ultrastructural and molecular data. J Eukaryot Microbiol 2006;53:26–35.
- [10] Kucerova Z, Moura H, Visvesvara GS, Leitch GJ. Differences between *Brachiola (Nosema) algerae* isolates of human and insect origin when tested using an *in vitro* spore germination assay and a cultured cell infection assay. J Eukaryot Microbiol 2004;51:339–43.
- [11] Visvesvara GS, Belloso M, Moura H, et al. Isolation of *Nosema algerae* from the cornea of an immunocompetent patient. J Eukaryot Microbiol 1999;46:10S.
- [12] Visvesvara GS, Moura H, Leitch GJ, Schwartz DA, Xiao LX. Public health importance of *Brachiola algerae* — an emerging pathogen of humans. Folia Parasitol 2005;52:83–94.
- [13] Coyle CM, Weiss LM, Rhodes III LV, et al. Fatal myositis due to the microsporidian *Brachiola algerae*, a mosquito pathogen. N Engl J Med 2004;351:42–7.
- [14] Cali A, Weiss LM, Takvorian PM. A review of the development of two types of human skeletal muscle infections from microsporidia associated with pathology in invertebrates and cold-blooded vertebrates. Folia Parasitol 2005;52:51–61 (Prague).
- [15] Margos G, Maier WA, Seitz HM. The effect of nosematosis on the development of *Plasmodium falciparum* in *Anopheles stephensi*. Parasitol Res 1992;78:168–71.
- [16] Muller A, Trammer T, Chioralia G, Seitz HM, Diehl V, Franzen C. Ribosomal RNA of *Nosema algerae* and phylogenetic relationship to other microsporidia. Parasitol Res 2000;86:18–23.
- [17] Vossbrinck CR, Andreadis TG, Vavra J, Becnel JJ. Molecular phylogeny and evolution of mosquito parasitic Microsporidia (Microsporidia: Amblyosporidae). J Eukaryot Microbiol 2004;51:88–95.
- [18] Biderre C, Pages M, Metenier G, Canning EU, Vivares CP. Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. Mol Biochem Parasitol 1995;74:229–31.
- [19] Biderre C, Duffieux F, Peyretailade E, et al. Mapping of repetitive and non-repetitive DNA probes to chromosomes of the microsporidian *Encephalitozoon cuniculi*. Gene 1997;191:39–45.
- [20] Staden R, Beal KF, Bonfield JK. The Staden package, 1998. Methods Mol Biol 2000;132:115–30.
- [21] De Rijk P, Caers A, Van de Peer Y, De Wachter R. Database on the structure of large ribosomal subunit RNA. Nucleic Acids Res 1998;26:183–6.
- [22] De Rijk P, Gatehouse HS, De Wachter R. The secondary structure of *Nosema apis* large subunit ribosomal RNA. Biochim Biophys Acta 1998;1442:326–8.
- [23] Wuyts J, De Rijk P, Van de Peer Y, Winkelmans T, De Wachter R. The European Large Subunit Ribosomal RNA Database. Nucleic Acids Res 2001;29:175–7.
- [24] Moura H, da Silva AJ, Moura IN, et al. Characterization of *Nosema algerae* isolates after continuous cultivation in mammalian cells at 37 degrees C. J Eukaryot Microbiol 1999;46:14S–6S.
- [25] Dia, N., Lavie, L., Metenier, G., Toguebaye, B.S., Vivares, C.P., Cornillot, E., in press. InterB multigenic family, a gene repertoire associated with subterminal chromosome regions of *Encephalitozoon cuniculi* and conserved in several human-infecting microsporidian species. Curr Genet, in press.
- [26] Biderre C, Pages M, Metenier G, et al. On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates. C R Acad Sci III 1994;317:399–404.
- [27] Nasonova E, Cornillot E, Metenier G, et al. Chromosomal composition of the genome in the monomorphic diplokaryotic microsporidium *Paranozema grylli*: analysis by two-dimensional pulsed-field gel electrophoresis. Folia Parasitol 2005;52 (Prague), 14S–W.
- [28] Amogan HP, Martinez JP, Ciuffetti LM, Field KG, Reno PW. Karyotype and genome size of *Nadelspora canceri* determined by pulsed field gel electrophoresis. Acta Protozool 2006;45:249–54.
- [29] Peyretailade E, Biderre C, Peyret P, et al. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res 1998;26:3513–20.
- [30] Amigó JM, Gracia MP, Salvadó H, Vivarès CP. Pulsed field gel electrophoresis of three microsporidian parasites of fish. Acta Protozool 2002;41:11–6.
- [31] Katinka MD, Duprat S, Cornillot E, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 2001;414:450–3.
- [32] Slamovits CH, Fast NM, Law JS, Keeling PJ. Genome compaction and stability in microsporidian intracellular parasites. Curr Biol 2004;14:891–6.
- [33] Xu J, Pan G, Fang L, et al. The varying microsporidian genome: existence of long-terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis*. Int J Parasitol 2006;36:1049–56.
- [34] Wickstead B, Ersfeld K, Gull K. Repetitive elements in genomes of parasitic protozoa. Microbio Mol Biol Rev 2003;67:360–75.
- [35] Barke AA, Rawal K, Ramaswamy R, Bhattacharya A, Bhattacharya S. The LINEs and SINEs of *Entamoeba histolytica*: comparative analysis and genomic distribution. Exp Parasitol 2005;110:207–13.
- [36] Tsai SJ, Kou GH, Lo CF, Wang CH. Complete sequence and structure of ribosomal RNA gene of *Heterosporis anguillarum*. Dis Aquat Organ 2002;49:199–206.
- [37] Gatehouse HS, Malone LA. The ribosomal RNA gene region of *Nosema apis* (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. J Invertebr Pathol 1998;71:97–105.
- [38] Kawakami Y, Inoue T, Ito K, et al. Identification of a chromosome harboring the small subunit ribosomal RNA gene of *Nosema bombycis*. J Invertebr Pathol 1994;64:147–8.
- [39] Brugere JF, Cornillot E, Metenier G, Bensimon A, Vivares CP. *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. Nucleic Acids Res 2000;28:2026–33.
- [40] De Rijk P, De Wachter R. RnaViz, a program for the visualisation of RNA secondary structure. Nucleic Acids Res 1997;25:4679–84.
- [41] Tsai SJ, Huang WF, Wang CH. Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene. J Eukaryot Microbiol 2005;52:52–4.

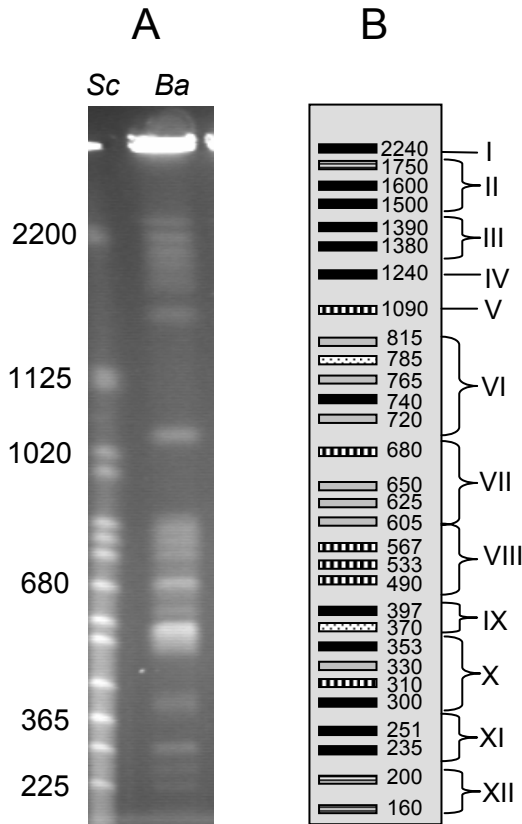


Figure 24. Définition des 12 groupes chromosomiques constituant le caryotype moléculaire de *B. algerae*. (A) . Résolution du caryotype moléculaire de *B. algerae* après migration électrophorétique en champs pulsés. (B). Sur cette représentation, les 12 groupes chromosomiques définis pour l'approche shotgun génome partiel sont indiqués. Sc (échelle *Saccharomyces cerevisiae* YNN295), Ba (*Brachiola algerae*). Les tailles de certains fragments de l'échelle sont indiquées à droite du gel.

C. DISCUSSION

Afin d'approfondir nos connaissances sur les pathogènes microsporidiens (relation hôtes-parasites, structure des génomes, capacités métaboliques) le séquençage du génome de *B. algerae* a été entrepris en collaboration avec le Genoscope. Une stratégie de séquençage par shotgun complet a tout d'abord été privilégiée. Une première tentative d'assemblage a permis de générer 7671 contigs dont 8 seulement présentent des tailles de 20 à 25 kb. La difficulté d'obtenir des contigs de grande taille résulte d'une richesse importante en séquences répétées et en éléments transposables. Suite à la résolution du caryotype moléculaire, une seconde approche de séquençage par chromosomes ou groupes de chromosomes a été envisagée. Pour cela, 12 groupes chromosomiques ont été définis (**Fig. 24**). Les premières données de séquences issues des deux premiers groupes séquencés ont permis d'améliorer la qualité de l'assemblage sans pour autant permettre l'obtention de contigs de plus de 30 kb. Le séquençage se poursuit pour améliorer le niveau d'assemblage.

Une première analyse des séquences a révélé que ce génome était particulièrement riche en adénine et thymine (76%), et, tout comme le génome des espèces microsporidiennes pour lesquelles des données de séquençage sont disponibles, quasiment dépourvu d'introns. En raison de ces deux caractéristiques, l'identification structurale des différents gènes a été réalisée en recherchant les cadres de lecture ouverts les plus longs. Cette annotation, réalisée sur les contigs obtenus par l'approche shotgun complet, a permis de caractériser 3431 CDS de plus de 300 nt, soit près de 1,5 fois plus que chez l'espèce *E. cuniculi* (1997 CDS). Ce nombre est également supérieur à celui retrouvé chez *A. locustae* (2600 CDS). Cependant, pour cette dernière espèce, le génome n'est pas complet et l'annotation des gènes partiels situés à l'extrémité des contigs n'a pas été réalisée.

L'annotation fonctionnelle des 3431 CDS de *B. algerae* a été réalisée sur la base d'une analyse comparative. Pour cela, la séquence de chaque CDS a été comparée à l'ensemble des

données de protéomes disponibles pour les espèces microsporidiennes (*E. cuniculi* : génome complet 1997 CDS ; *A. locustae* : génome incomplet 2605 CDS) puis à l'ensemble de séquences protéiques disponibles dans les banques Swissprot et TrEMBL. Les premiers résultats ont permis d'attribuer une fonction à 235 gènes. Cette annotation se poursuit actuellement.

L'existence d'un répertoire génique aussi conséquent est sans aucun doute à l'origine des capacités d'adaptation très importantes de *B. algerae*. Pour preuve, *B. algerae* est capable d'initier un cycle de développement au sein de cellules hématocytaires. Ces cellules de l'immunité, retrouvées au niveau de l'hémolymphe chez le moustique ont pour rôle d'assurer la destruction des éléments pathogènes. Cette capacité de prolifération à l'intérieur même des cellules de l'immunité, et plus précisément des macrophages, caractérise plusieurs protozoaires (*T. gondii*, *T. cruzi* et *Leishmania* sp.) responsables d'infections chroniques (**Denkers et Butcher, 2005**). Chacun de ces parasites met en place une stratégie d'échappement parasitaire spécifique, en modifiant notamment l'activation de voies de transduction indispensables à une réponse immunitaire efficace. L'étude des mécanismes d'invasion de *B. algerae*, réalisées sur plusieurs types cellulaires (lignée hématocytaire, macrophages, HFF) devra être envisagée.

De plus, l'identification complète de ce répertoire représente également, dans le cadre d'approches post-génomiques, un atout supplémentaire pour orienter les recherches expérimentales visant à mieux comprendre le cycle de développement des microsporidies, leurs modes de transmission ou encore les mécanismes d'échappement parasitaire. Les produits de ces gènes pourraient représenter des cibles parasitaires potentielles et conduire au développement de nouvelles thérapeutiques efficaces (**Didier et al., 2005**). De telles analyses ont ainsi été menées pour des espèces du genre *Plasmodium*. Elles ont par exemple permis de caractériser des protéines kinases (CDKs) qui, du point de vue structural, sont différentes de

celles retrouvées chez l'hôte (**Kappes et al., 1999 ; Doerig et al., 2002 ; Ward et al., 2004**). Ces approches ont ainsi conduit à la recherche et à l'identification de nouvelles molécules ciblant de manière spécifique ces enzymes, et donc, capables d'exercer un effet antiparasitaire (**Cassandra et al., 2003, Woodard et al., 2003 ; revue par Keenan et al., 2005**).

Parmi les gènes identifiés chez *B. algerae*, sans homologues chez d'autres espèces microsporidiennes, il faut noter la présence de deux gènes (*Ago1*, *Dicer-like4*) codant pour des protéines intervenant dans les mécanismes d'ARNi. Ce résultat est d'autant plus intéressant qu'il permet d'émettre des hypothèses quant à l'évolution des génomes microsporidiens (régulation de l'expression des éléments transposables) mais également fournit un outil de génomique fonctionnelle puissant. Par conséquent, cet organisme représente probablement aujourd'hui le meilleur modèle microsporidien pour le développement d'outils de génomique fonctionnelle utilisant les approches de transfection et d'ARNi. Le génome de cet organisme étant riche en régions non codantes, il existe une plus forte probabilité d'intégrer dans ces régions des constructions d'intérêt et de générer des lignées stables sans modification d'expression des gènes endogènes.

Contrairement à *E. cuniculi* qui se développe à l'intérieur d'une vacuole parasitophore, *B. algerae* se multiplie au contact direct du cytoplasme de la cellule hôte (pour revue voir **Fasshauer et al., 2005**). Cette spécificité pourrait grandement faciliter la mise au point de conditions de transfection, aucun système n'ayant été développé à l'heure actuelle.

L'annotation du génome d'*E. cuniculi* a permis de mettre en évidence un nombre réduit de gènes dupliqués (**Katinka et al., 2001**). La taille importante du génome de *B. algerae* pourrait être la conséquence de phénomènes de duplication de gènes entraînant l'apparition de familles multigéniques ou de phénomènes de duplication de régions génomiques comme l'a montré la comparaison de plusieurs génomes de levures (**Dujon et al., 2004**). La vérification de cette hypothèse nécessite cependant d'être couplée à une analyse de

conservation de la synténie entre les deux espèces. L'outil ACT (Artemis Comparison Tool), développé par l'institut du Sanger (**Carver et al., 2005**), semble ici le mieux adapté.

Entre les espèces *E. cuniculi* et *A. locustae*, un degré inattendu de conservation de synténie est observé ; 13% des gènes sont présents dans le même contexte et 30% sont séparés par un petit nombre de réarrangements (**Slamovits et al., 2004**).

Chez les protozoaires parasites, le degré de conservation de la synténie est variable. Concernant les *Trypanosomatidae* par exemple, dont les espèces ont divergé il y a plus de 400 millions d'années, de grandes régions de synténie ont été décrites alors qu'une forte divergence des séquences est observée (**Ghedini et al., 2004**). Ainsi, chez *T. brucei* et *L. major*, 68 et 75% respectivement des gènes sont retrouvés dans un même contexte génomique, soit 110 régions synténiques dont les tailles varient de 19,9 à 30,7 Mb. En raison de l'organisation polycistronique qui caractérise ces génomes, l'hypothèse d'une forte pression de sélection indispensable au maintien de l'ordre des gènes est évoquée (**El-Sayed et al., 2005**). Le degré de conservation de synténie est également élevé entre les espèces du genre *Plasmodium* (**Carlton et al., 2002**). Ainsi, entre les espèces *P. falciparum* et *P. yoelii*, une étude comparative portant sur une région synténique de 50 kb a révélé une conservation de l'ordre des gènes pour 75% d'entre eux. Une diminution de degré de conservation de la synténie est néanmoins constatée lorsque les distances phylogénétiques augmentent (**Carlton et al., 2002**).

Une telle conservation n'est pas observée entre les espèces *Entamoeba invadens* et *E. histolytica* bien que ces deux espèces aient divergé beaucoup plus tardivement (**Wang et al., 2003**). L'analyse structurale de leur génome a révélé l'existence d'une fraction importante de séquences répétées (**Loftus et al., 2005**). De tels réarrangements ou processus de recombinaison pourraient être la conséquence, au sein de leurs génomes, de cette fraction importante de séquences répétées.

Cette analyse de conservation de la synténie donnera également des informations sur l'organisation des chromosomes de *B. algerae*, tout comme l'analyse portant sur les contigs où sont localisées les unités d'ADNr dont la position chromosomique devra être définie. En effet, le nombre élevé de séquences entrant dans la composition de certains contigs portant ces unités laisse supposer l'existence d'un nombre relativement important d'unités d'ADNr au sein du génome. Une telle situation est notamment retrouvée chez *E. cuniculi*, où 22 copies ont pu être caractérisées en position subtélomériques (1 unité pour chaque extrémité chromosomique). Une telle organisation pourrait être appréhendée par des approches de cartographie physique utilisant la KARD-2D (**Brugère et al., 2000**).

L'analyse de certains contigs révèle la présence de régions intergéniques de taille supérieure à celles identifiées chez les espèces *E. cuniculi* et *A. locustae*, probablement en raison d'un système de régulation transcriptionnelle différent. En effet, il a été montré que la faible taille des espaces intergéniques chez *E. cuniculi* est liée à une organisation structurale de type polycistronique. La structure du génome de *B. algerae* ne semble donc pas répondre à une telle organisation. Deux hypothèses peuvent néanmoins être émises concernant les modalités de régulation de la transcription chez cette espèce: hypothèse 1) les régions intergéniques contiennent les signaux de fin de transcription et séquences promotrices des gènes situés respectivement en amont et en aval, ce qui signifie que le génome de *B. algerae* est constitué principalement d'unités transcriptionnelles monocistroniques ; hypothèse 2) bien qu'ayant une taille supérieure, les régions intergéniques ne contiennent pas les signaux d'initiation et de fin de transcription et correspondent réellement à des régions intercistroniques (organisation polycistronique). Le développement d'une étude comparative portant sur l'analyse des génomes de *B. algerae* et d'*E. cuniculi* pourrait nous aider à mieux appréhender les mécanismes de régulations transcriptionnelles propres à chacune des ces microsporidies. L'analyse comparative de leur génome pourrait également mettre en lumière

les mécanismes pouvant être à l'origine du processus de compaction des génomes microsporidiens.

L'analyse des premiers jeux de données a révélé au sein du génome de *B. algerae* l'existence de nombreuses classes d'ETs (*Ty3-Gypsy*, *Copia*, LINE). Cette grande hétérogénéité de classes pourrait être corrélée au large spectre d'hôtes de cette espèce microsporidienne. En effet, l'hypothèse d'un transfert horizontal d'ETs de la cellule hôte vers le génome microsporidien ne peut être exclue, les parasites intracellulaires étant, au même titre que les virus, considérés comme des vecteurs potentiels de séquences génétiques, et donc d'ETs (revue par **Silva et al., 2004**).

De part leur capacité à se déplacer au sein des chromosomes, les ETs sont considérés comme importants dans l'évolution de la structure des génomes. Il serait intéressant de savoir si ces éléments sont actifs et donc capables d'initier des cycles de réplication. Pour cela, des expérimentations de RT-PCR devront être menées en ciblant différentes parties des cadres de lecture ouverts. L'analyse des différentes classes d'éléments mobiles retrouvés chez *B. algerae* devrait nous apporter quelques éléments de réponse quant à leur rôle probable dans la plasticité des génomes microsporidiens, et leur influence sur l'organisation chromosomique. Notons également que ces ETs pourraient jouer un rôle primordial dans les stratégies qui seront déployées afin d'assurer des transformations stables de cette espèce.

Le séquençage du génome d'*E. cuniculi* n'a pas révélé la présence d'ETs connus, ni de transcriptase inverse, enzyme indispensable à leur mobilité. Cette enzyme a néanmoins été caractérisée au sein du génome des deux espèces microsporidiennes, *V. corinae* (**Mittleider et al, 2002**) et *S. lophii* (**Hinkle et al, 1997**). Pour *S. lophii*, l'enzyme détectée présente de fortes homologies avec les éléments mobiles de type « gypsy » appartenant à la catégorie des rétrotransposons. Très récemment, 8 familles d'éléments transposables (*Nbr*) de type *Ty3/gypsy* ont été identifiées au sein du génome de *N. bombycis* (**Xu et al 2006**). La mise en

évidence de mutations non sens au sein de ces différentes familles d'ETs (rétrotransposons) indique que ces éléments seraient actifs.

Chez les protozoaires parasites, plusieurs types d'ETs ont été décrits et la plus grande diversité est enregistrée chez les trypanosomes (**Bhattacharya et al, 2002**). L'analyse de leur séquence a montré, qu'à l'exception d'une classe d'éléments trouvée chez *T. cruzi*, ce sont tous des rétrotransposons. *G. lamblia* contient trois familles d'ETs dont deux sont en position subtélomériques (**Arkhipova et Morrison, 2001**) alors que chez *E. histolytica*, de nombreux rétrotransposons sont dispersés dans son génome. La présence de ces ETs dans les génomes de la plupart de ces microorganismes suggère une acquisition précoce de ces éléments au cours de l'évolution. Certaines espèces comme *E. cuniculi* auraient ainsi subi une perte secondaire de ces éléments.

III. Développement de nouveaux outils de génomique fonctionnelle chez *Anopheles gambiae*

De nombreux laboratoires travaillent depuis quelques années sur l'étude des interactions *Anopheles-Plasmodium* (revue par **Barillas-Murry et Kumar, 2005**). Ces thématiques scientifiques sont regroupées autour de deux grands domaines d'études que sont : 1) les mécanismes de défense déclenchés par l'insecte au cours de l'invasion (mécanismes moléculaires et cellulaires de la réponse immunitaire), et 2) les stratégies d'échappement parasitaire lui permettant d'assurer son cycle de développement (**Michel et Kafatos, 2005 ; Whitten et al., 2006 ; Dong et al., 2006**). Ces travaux ont permis l'identification d'un grand nombre de gènes du moustique dont l'expression semble régulée par le parasite (**Vlachou et al., 2005 ; Dong et al., 2006**). Ces gènes, potentiellement impliqués dans le cycle de *Plasmodium*, représentent de véritables gènes cibles dont l'inactivation permettrait de bloquer le cycle de maturation parasitaire chez le moustique, et donc, la transmission du paludisme. Cette approche nécessite le développement d'un système efficace de génétique inverse permettant d'inactiver l'expression d'un ou de plusieurs gènes. Beaucoup plus facile à mettre en œuvre que la recombinaison homologe, l'ARNi se présente comme un outil de choix dans l'inhibition post-transcriptionnelle d'un gène d'intérêt (**Dong et al., 2006**).

Ainsi, l'objectif que nous nous sommes fixés dans le cadre du projet "Génomique fonctionnelle chez *Anopheles gambiae* : développement de nouveaux outils" (2002-2004) coordonné par le Dr. Catherine Bourgoïn suite à l'appel d'offre « Post-séquençage *Anopheles* », consiste à développer, chez *An. gambiae*, des outils efficaces d'interférence ARN. Quatre équipes de recherche sont impliquées dans ce projet. Notre équipe maîtrisant les modèles cellulaires d'*An. gambiae*, ces systèmes *in vitro* permettront de valider les efficacités des invalidations par ARNi. Les systèmes d'invalidation génique (ARNdb, siARN, miARN) et constructions les plus efficaces seront alors testés *in toto* par l'équipe de C. Bourgoïn (Biologie et Génétique du Paludisme, Institut Pasteur, Paris). Cette équipe de recherche, qui a identifié plusieurs gènes cibles potentiellement impliqués dans le cycle de maturation du parasite, est chargée d'assurer le développement de lignées transgéniques chez *An. gambiae*. L'équipe de Christophe Antonieswki du laboratoire de Biologie du Développement (Génétique et Epigénétique de la Drosophile, Institut Pasteur, Paris), possède de solides compétences en matière d'ARNi chez la drosophile. Sur ce même modèle, l'équipe de Chantal Vaury du laboratoire Interactions génétiques et cellulaires au cours de la différenciation (INSERM 384, Faculté de Médecine, Clermont-Ferrand), a réalisé des

avancées importantes dans l'élaboration des vecteurs de transgénèse nécessaires au développement du projet.

Notre contribution dans ce projet nous permettra également d'acquérir une maîtrise des outils de génomique fonctionnelle pour l'étude de notre modèle biologique (interactions *B. algerae* / *Anopheles gambiae*).

III.1. Transfection des lignées cellulaires d'*An. gambiae*

La détermination de la conformation d'ARNdb la plus efficace à induire l'inactivation d'un gène a été tout d'abord appréhendée sur deux lignées cellulaires d'*An. gambiae* : une lignée hémocytaire (Sua4.0) impliquée dans les défenses immunitaires ainsi qu'une lignée embryonnaire (Mos55).

Dans un premier temps, l'utilisation du gène rapporteur codant l'EGFP (Enhanced Green Fluorescent Protein) permettant une analyse directe a été privilégiée pour évaluer le système de transfection le plus efficace sur les deux lignées cellulaires. Ce gène a été placé sous contrôle du promoteur actine de drosophile (Hoa *et al.*, 2003 ; Brown *et al.*, 2003). La construction plasmidique a été réalisée par l'équipe de C. Antoniewski.

Les deux lignées utilisées ne présentant pas d'inhibition de contact, elles forment des amas cellulaires rendant difficile la quantification précise des taux de transfection. Ainsi, les essais de quantification en cytométrie de flux (Bioanalyser, FACS) présentés ci-après se sont avérés inappropriés en raison de la formation de ces agrégats cellulaires indissociables. L'estimation visuelle du nombre de cellules transfectées a donc été privilégiée dans un premier temps. Pour cela, le nombre moyen de cellules transfectées pour chaque condition a été évalué sur 50 champs (Leica, grossissement x 200).

Trois techniques de transformation ont été utilisées au cours de notre étude, la magnétofection, l'électroporation et la lipofection. Les premiers essais par électroporation se sont avérés inefficaces. Cette technique s'est montrée particulièrement inadaptée à la transformation cellulaire de nos lignées d'insectes en raison d'un taux de lyse cellulaire important et de l'absence de cellules exprimant l'EGFP. Nous avons donc privilégié l'utilisation de la lipofection. Différents agents de transfection ont été testés (Polyfect, Lipofectin, Dreamfect, Genejuice) dont certains réputés pour leur efficacité sur cellules d'insectes (Cellfectin, Flyfectin). Les taux de transfection les plus importants ont été obtenus avec la Cellfectin. La mise au point des conditions de transfection a été uniquement réalisée sur la lignée Sua4.0 qui a démontré dès les premiers essais une plus grande sensibilité aux agents de transfection. Ces essais ont permis de mettre en évidence un taux maximum de

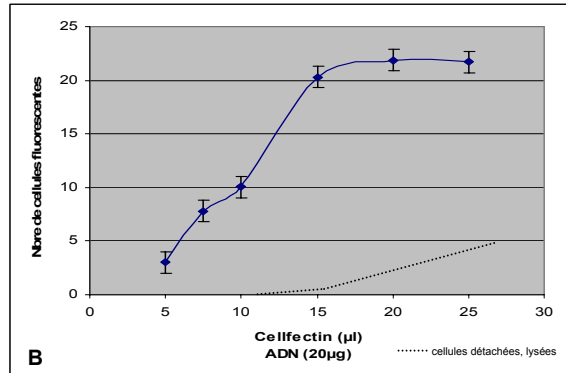
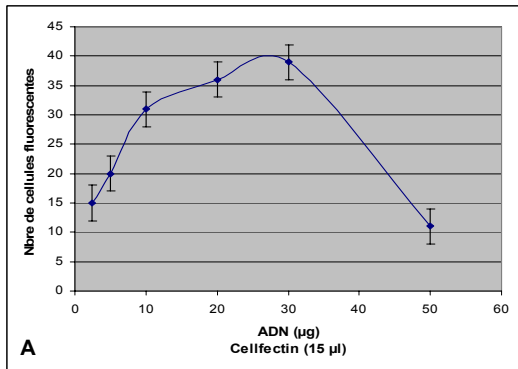


Figure 25. Optimisation de l'efficacité de transfection de cellules hémocytaires d'*Anopheles gambiae* (lignée Sua4.0). $1,5 \times 10^6$ cellules ont été transfectées dans chacune des expériences. **(A)** Variation des quantités d'ADN (quantité de Cellfectin fixée à 15 µl/puits). **(B)** Variation des quantités de Cellfectin (quantité d'ADN fixée à de 20 µg/puits). Au delà de 15 µl de Cellfectin par puits, un début de lyse cellulaire est observée (effets cytotoxiques). Les cellules ont été transfectées à l'aide du vecteur pAc5.1d4EGFP. 72 h après transfection, les cellules ont été montées entre lame et lamelle pour dénombrement (Nbre de cellules fluorescentes/champ).

cellules transfectées pour une quantité de Cellfectin de 15 µl /puits (confluence de 80%, quantité d'ADN fixée à 20 µg) (**Fig. 25**). Au-delà, le taux n'augmente pas de manière significative, et apparaît pour des quantités supérieures à 15 µl de Cellfectin un début de lyse cellulaire (effets cytotoxiques). La détermination des quantités optimales d'ADN plasmidique montre qu'entre 20 et 30 µg d'ADN, les taux de transfection sont les plus élevés (**Fig. 25**). Au-delà de 30 µg, on observe une diminution importante de l'efficacité de transfection.

Les premiers résultats concernant ces essais indiquent clairement que tous les autres agents de transfection utilisés ne sont pas adaptés à nos lignées cellulaires. De nombreux facteurs ont été modifiés (concentration en vecteur et en agent de transfection, temps d'incubation) sans variation significative de l'efficacité de transfection. Enfin, des essais faisant appel à un champ magnétique (magnétofection) combiné ou non à un agent de transfection ont été réalisés. Nous avons pu mettre en évidence une augmentation importante de l'efficacité de transfection (+ 40%) lorsque la magnétofection était associée à la Cellfectin.

Dans un second temps, de manière à confirmer les estimations visuelles des taux de transfection, des essais de quantification ont été réalisés à l'aide du Bioanalyser Agilent, uniquement à partir des essais de transfection utilisant la Cellfectin. Aucune estimation des taux de transfection à l'aide du FACS n'a été réalisée. Les essais ont permis de mettre en évidence une plus faible efficacité des taux de transfection par rapport aux quantifications en microscopie à épifluorescence. L'une des explications pourrait reposer sur un manque de sensibilité des chaînes de détection des cytomètres. En effet, la fluorescence faible de certaines cellules pouvant être assimilée à l'autofluorescence importante de la lignée cellulaire (Sua4.0). Cette faible fluorescence pourrait être la conséquence d'un manque d'efficacité de la traduction du messager codant l'EGFP en raison d'une séquence Kozak moins appropriée pour le modèle biologique *Anopheles*. Nous ne pouvons cependant pas exclure une moindre efficacité du promoteur de drosophile dans le modèle *Anopheles* ou une stabilité des messagers diminuée. Nous avons entrepris néanmoins des modifications de la séquence Kozak basées sur l'exploitation des données d'ESTs de gènes fortement exprimés chez *An. gambiae* (actine, tubuline, cytochromes, HSP). Ces données nous ont permis de caractériser une séquence Kozak putative propre à *An. gambiae* (AAACAUGN). Cette séquence est semblable à la séquence Kozak de drosophile utilisée jusqu'à présent (C/AAAA/CAUGN).

Tableau 8. Environnement du codon d'initiation de traduction de l'EGFP testé dans les expériences de transfection des lignées cellulaires d'*Anopheles gambiae*.

AK (séquence Kozak de référence); NK1 à NK6 (séquences Kozak testées).

Construction	Séquences (5'→3')
AK	ACA AAC ATG GTG AGC
NK1	ACA AA A ATG GTG AGC
NK2	AAA AA A ATG GTG AGC
NK3	ACA AA A ATG G A G AGC
NK4	ACA AA A ATG A A G AGC
NK5	AAA AA A ATG A A G AGC
NK6	AAA AA A ATG A A A AGC

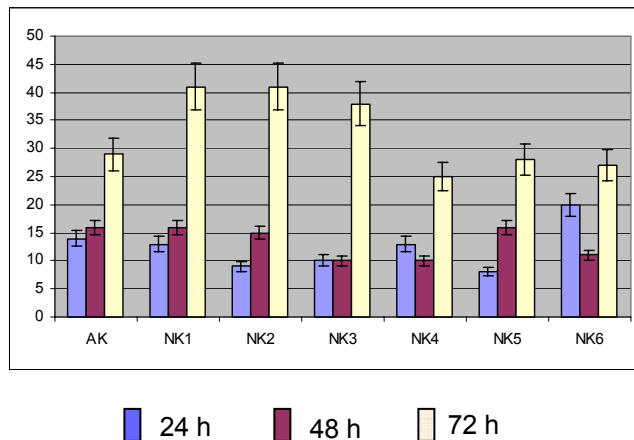


Figure 26. Estimation du nombre de cellules fluorescentes en fonction de l'environnement du codon d'initiation de traduction du gène rapporteur EGFP. Les quantités d'ADN plasmidique (pAc5.1d4EGFP) et d'agent de transfection (Cellfectin) ont été fixées à 20 µg et 15 µl respectivement. Les cellules ont été dénombrées 24, 48 et 72 h après transfection. AK (séquence Kozak de référence) ; NK1 à NK6 (séquences Kozak testées).

Cependant, plusieurs modifications de cette séquence ont été réalisées pour déterminer la séquence la plus efficace (**Tableau 8**). L'expression du gène rapporteur a été quantifiée pour chacune de ces constructions (**Fig. 26**).

Quelle que soit la séquence, aucune différence significative des taux de transfection n'est observée visuellement après 24 h ou 48 h de transfection. Cependant, après 72 h, le nombre de cellules pouvant être détecté est quasiment multiplié par un facteur 4 pour 3 des 7 constructions (NK1, NK2 et NK3). Un faible taux d'expression du gène codant l'EGFP durant les premières 48 h, associé également à une instabilité importante du messager dont le temps de demi-vie est de 4 h (variant d4 de la GFP) pourrait expliquer ces différences de résultats. L'augmentation de l'intensité de fluorescence au cours du temps (après 72 h) peut ainsi s'expliquer par une accumulation de l'EGFP rendant possible la détection d'un nombre plus important de cellules transfectées.

La guanine en +4 et l'adénine en position -1 semblent jouer un rôle important dans l'augmentation de la fluorescence. Cette modification se traduit par une augmentation d'environ 38% du nombre de cellules émettant une fluorescence caractérisable. L'observation au microscope à épifluorescence révèle également pour ces constructions une augmentation importante de la fluorescence (très supérieure au bruit de fond). La présence d'une adénine en -5 (NK1/NK2) ou en +5 ne semble pas déterminante. Enfin, la présence d'une suite contiguë de résidus adénine en amont de l'ATG (NK2) n'apparaît pas non plus primordiale.

III.2. Evaluation de l'efficacité de l'ARN interférence en culture de cellules d'*An. gambiae*

Compte tenu des essais de transfection très prometteurs sur la lignée Sua4.0, nous avons décidé de réaliser l'ensemble des tests permettant la détermination des meilleures conformations d'ARNdb sur cette lignée. Différentes possibilités ont pu être envisagées concernant les modes d'obtention de ces ARNdb. Pour la mise en place d'une interférence initiée par ARNdb de grande taille (600 pb), notre choix s'est porté sur une synthèse *in vitro* (Megascript-Ambion) d'un ARNdb permettant de cibler l'ARNm de l'EGFP. Des cellules Sua4.0 ont été repiquées en boîte 6 puits à raison de $1,5 \times 10^6$ cellules/puits 24 h avant l'étape de transfection. Celle-ci a été réalisée avec le vecteur d'expression pAc5.1-d4EGFP (20µg/puits) en utilisant la Cellfectin (15µl) comme agent de transfection. Après deux jours, les cellules ont été mises en contact avec les ARNdb anti-EGFP (incubation 12 h). Une diminution visible du nombre de cellules fluorescentes a pu être observée 48, 96 et 120 h après l'ajout des ARNdb. En raison de la formation d'agrégats cellulaires et des difficultés

liées à une estimation visuelle fiable de la décroissance de la fluorescence, aucun taux d'extinction n'a pu être défini.

Les miARNs, retrouvés chez un grand nombre d'organismes (Vertébrés et Invertébrés), et impliqués dans des processus de régulation d'expression génique ont été décrits comme des médiateurs beaucoup plus efficaces de l'ARNi. Ces ARNs non codants, de 21 à 25 nt, sont capables de réguler l'expression génique en s'appariant aux ARNm cibles dont ils sont partiellement complémentaires. Cette hybridation a pour effet de bloquer la traduction de l'ARNm ou d'entraîner sa dégradation par clivage spécifique à la manière des siARNs (Jaubert *et al.*, 2007 ; Zhang *et al.*, 2007).

Ces miARNs présentent l'avantage, du moins pour certains, d'être organisées en tandem (cluster), ce qui permet de contrôler d'une part, la synthèse de plusieurs ARNdb à partir d'une même construction, mais également de placer cette synthèse sous le contrôle d'un promoteur inductible. Il s'agit donc ici de reprogrammer des miARNs naturels afin de cibler l'extinction de gènes d'intérêt (GFP).

L'équipe de C. Antoniewski a développé plusieurs constructions nécessaires à la validation de systèmes d'expression de miARNs. Ces constructions sont destinées d'une part, à contrôler la fonctionnalité du système d'expression de miARNs, et d'autre part, à tester sa capacité à inactiver efficacement un gène cible (GFP).

Ces constructions ont pu être testées sur la lignée Sua4.0. Des essais de transfection à l'aide d'un vecteur témoin (GFP sous contrôle du promoteur α -tubuline de la drosophile) ont tout d'abord été réalisés. 24, 48 et 72 h après transfection, les cellules ont été fixées et montées entre lame et lamelle pour observation. Aucune fluorescence n'a été détectée, indiquant clairement une inefficacité du promoteur α -tubuline de la drosophile chez *An. gambiae*.

Les constructions utilisant le promoteur α -tubuline avaient été initialement développées pour la drosophile. D'autres constructions utilisant le promoteur actine 5C (ubiquitaire) de la drosophile devront ainsi être réalisées afin de valider l'approche d'inactivation par miARNs chez *An. gambiae*. De nombreux travaux ont en effet déjà permis de valider la fonctionnalité de ce promoteur chez cette espèce (Zhao et Eggleston, 1999 ; Hoa *et al.*, 2003 ; Brown *et al.*, 2003).

La démonstration d'une inactivation spécifique par ARNi au niveau du tube digestif (Ito *et al.*, 2002), mais également au niveau des glandes salivaires (Boisson *et al.*, 2006) a été

faite. En raison des avantages liés à l'utilisation de miARNs reprogrammés pour l'inactivation de gènes cibles, le développement de ces outils devront se poursuivre.

Au travers de nos résultats préliminaires, nous avons démontré l'efficacité de transfection de la lignée cellulaire Sua4.0 et des capacités d'inactivation des gènes par ARNi. Malgré les difficultés de quantification et les améliorations à apporter pour l'optimisation de l'inactivation génique par ARNi ce modèle cellulaire demeure un système prometteur (crible de constructions pour l'ARNi et modèle d'étude des interactions hôte-pathogènes). Nous avons alors étudié la possibilité d'invalider un gène candidat d'*An. gambiae* potentiellement impliqué dans le développement de *P. falciparum*.

III.3. Etude de gènes candidats surexprimés chez le moustique au cours d'une infestation parasitaire

III.3.1. 14Yde: un variant d'épissage du gène *cap* ?

Dans le cadre d'une collaboration avec l'équipe de C. Bourgoïn, nous nous sommes intéressés à un gène identifié par cette même équipe et dont l'expression est régulée chez le moustique au cours de l'invasion par *P. falciparum* (Bonnet et al., 2001). Ce gène apparaît surexprimé 24 h après un repas sanguin infesté, c'est-à-dire lors de la traversée de l'épithélium intestinal par l'ookinète. Dans le cas d'une surexpression placée sous contrôle parasitaire, ce gène constituerait un gène cible potentiel à inactiver. L'expression de ce gène pourrait également résulter d'un mécanisme de défense du moustique suite à l'invasion intestinal (réponse immunitaire). Ainsi, la surexpression ciblée de ce gène pourrait de la même manière limiter les capacités d'échappement parasitaire.

Après amplification complète du messager (5' et 3' RACE-PCR), le gène *14Yde* a été cloné et séquencé dans sa totalité par l'équipe de C. Bourgoïn. *14Yde* code pour une protéine de 357 acides aminés (39 kDa) qui présente trois domaines de type SH3 (src Homology-3). Ces motifs sont impliqués dans des interactions de type protéine-protéine à l'origine de la transduction de signaux, l'adhésion cellulaire ou encore le remaniement du cytosquelette (Ribon et al., 1998 ; Baumann et al., 2000; Yamazaki et al., 2002). Grâce aux données d'annotation du génome d'*An. gambiae* (<http://www.ensembl.org/index.html>), nous avons pu identifier le gène codant ce messager (ENSANGG0000006935). Ce dernier dénommé *cap* (pour c-Cbl associated protein) pourrait finalement générer 3 transcrits (notés T1, T2 et T3) résultant d'un épissage alternatif dont aucun ne correspond au variant *14Yde* (variant d'épissage). Nous avons analysé ce gène pour identifier d'autres variants possibles en utilisant

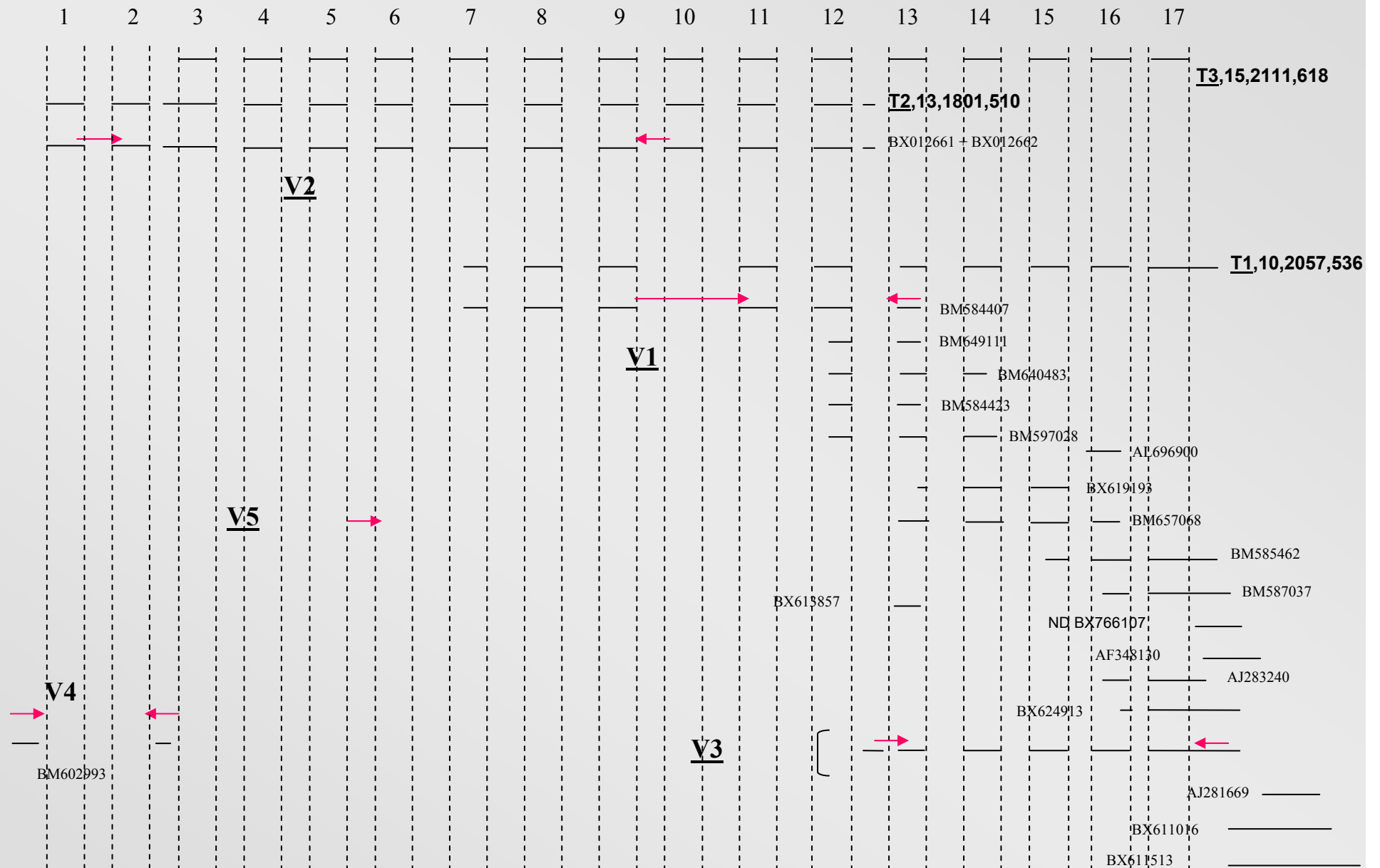


Figure 27. Représentation schématique de l'alignement des transcrits prédits automatiquement dans AnoBase (T1, T2 et T3) et des ESTs (BX, BM et AJ) avec la séquence génomique du gène *cap*. L'alignement du gène et des ESTs a été réalisé avec le logiciel SIM4 permettant d'identifier les exons et les introns. De 1 à 17 : exons du gène *cap* décrits dans AnoBase (ENSANGG00000006935). 3 transcrits d'épissage (notés T1, T2 et T3) sont prédits par l'annotation automatique. Les chiffres situés après ces variants indiquent respectivement le nombre d'exons, la taille en pb ainsi que le nombre d'acides aminés (aa) de la protéine codée par le messager en question. Les transcrits putatifs (T1 et T2) issus de l'annotation automatique sont validés par les données expérimentales (ESTs). Les variants identifiés par alignement de données EST sont notés de V1 à V5. V3 correspond au gène *14Yde* (*cap*). Les flèches en rouge indiquent la position des amorces utilisées pour l'amplification de chacun des variants (V1 à V5).

les données d'ESTs qui ont été générées par différents laboratoires. En effet, chez *An. gambiae*, une analyse basée sur l'exploitation des données de génomiques confrontées aux données expérimentales (ESTs) nous a permis de caractériser l'existence de 5 variants (**Fig. 27**).

Il apparaît que les transcrits putatifs (T1 et T2) issus de l'annotation automatique sont confirmés par nos analyses d'ESTs (V1 et V2), mais aucun des ESTs ne permet de conforter l'existence du transcrit 3 (2111 pb, 15 exons, 618 aa). De plus, le messager étudié (*14Yde*) semble donc correspondre à un nouveau variant non prédit par l'annotation automatique de ce gène. Ce nouveau variant est caractérisé par une région 3' UTR beaucoup plus longue s'étendant au-delà du site de clivage/polyadénylation définie par l'annotation des trois autres variants. Cependant, de nombreux ESTs (ex : BX611016, BX611513...) révèlent l'existence d'un ou plusieurs variants dont l'extrémité 3' UTR est également plus longue. Le variant V4, confirmé expérimentalement par l'EST BM602993 correspond à un autre variant non détecté par l'annotation automatique. Le variant V5, quant à lui, correspond à un variant putatif où seule l'amorce directe utilisée pour le caractériser est indiquée (amorce chevauchante exons 3-4). Les amorces inverses des variants V1, V2 et V3 ont ainsi été utilisées en combinaison avec l'amorce directe du variant 5 afin de caractériser l'existence d'un messager non détecté par l'annotation automatique.

Une approche par homologie a permis d'identifier l'homologue du gène *14Yde* d'*An. gambiae* chez *D. melanogaster*, le gène *dcap* (pour *Drosophila* c-Cbl-associated protein). Ce gène, par épissage alternatif, peut potentiellement générer 10 protéines putatives. Actuellement seuls 5 variants d'épissage ont été caractérisés expérimentalement (**Yamazaki et al., 2002**). Les 5 protéines codées par ces variants présentent une région C-terminale conservée possédant une identité de 62% avec la protéine produite par le variant *14Yde*. A l'heure actuelle, chez *An. gambiae*, aucune fonction biologique n'a été décrite pour cette protéine. Chez *D. melanogaster*, la protéine homologue (DCAPS), principalement exprimée au niveau du tube digestif et du corps gras à un stade embryonnaire tardif, serait impliquée dans le transport du glucose (**Yamazaki et al., 2002**). Nous avons alors entrepris l'étude de l'expression des différents variants chez *An. gambiae*.

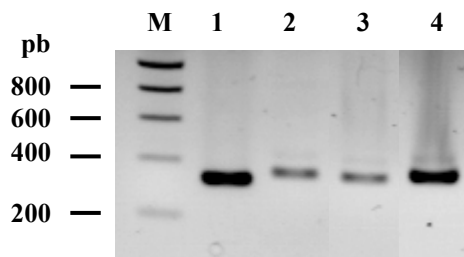


Figure 28. Analyse par RT-PCR de l'expression du gène codant 14YDE à partir d'ARN totaux extraits de moustiques adultes. Analyse sur gel d'agarose 1% du produit d'amplification (taille attendue 356 pb). Piste 1 : témoin actine-échantillon femelles gorgées (300 pb), piste 2 : femelles gorgées, piste 3 : femelles non gorgées, piste 4 : mâles. M : échelle de poids moléculaire (Eurobio).

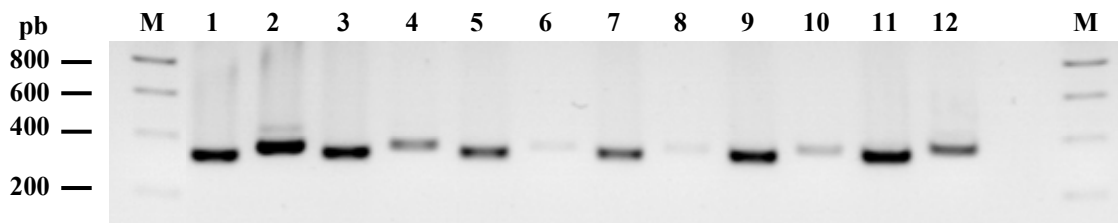


Figure 29. Analyse par RT-PCR de l'expression du gène codant 14YDE à partir d'ARN totaux extraits à différents stades de développement d'*Anopheles gambiae*. Analyse sur gel d'agarose 1% du produit d'amplification (taille attendue 356 pb). Piste 1, 3, 5, 7, 9, 11 : témoin actine pour chaque stade de développement (300 pb), piste 2 : stade oeuf, piste 4 : stade larvaire L2, piste 6 : stade larvaire L3, piste 8 : stade larvaire L4, piste 10 : stade nymphe jeune, piste 12 : stade nymphe âgée. M : échelle de poids moléculaire (Eurobio).

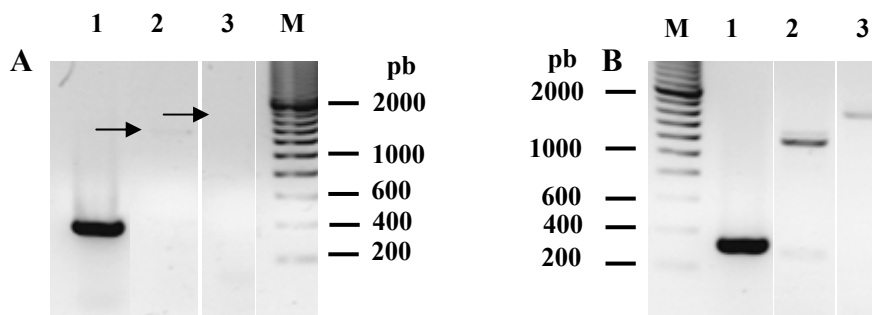


Figure 30. Analyse par RT-PCR de l'expression des variants V1 et V5 à partir d'ARN totaux extraits de moustiques adultes. Analyse sur gel d'agarose 1% des produits d'amplification (A) Femelles gorgées. Piste 1 : témoin actine (300 pb), piste 2 : variant V1 (1319 pb), piste 3 : variant V5 (1775 pb). (B) moustiques mâles. Piste 1 : témoin actine (300 pb), piste 2 : variant V1, piste 3 : variant V5. M : échelle de poids moléculaire (Eurobio).

III.3.2. Caractérisation d'expression des variants du gène *14Yde* par RT-PCR

Parallèlement aux essais mis en oeuvre pour caractériser les protéines codées par le gène *cap*, l'étude de l'expression des différents variants a été menée dans nos lignées cellulaires. En effet, ce gène est un bon candidat pour mener à bien le développement des approches d'inactivation par ARNi. Ce gène est en effet dérégulé par une infection parasitaire. Nous avons tout d'abord cherché à déterminer si l'expression de ces variants était constitutive, ou correspondait à une expression induite au cours d'un phénomène infectieux, c'est-à-dire à une réponse immunitaire non spécifique à *Plasmodium*. Pour cela, la caractérisation de manière systématique, de l'expression de l'ensemble des variants du gène *cap* avec ou sans stimulation antigénique (LPS, Zymosan, H₂O₂), a été entreprise sur les deux lignées cellulaires Sua4.0 et Mos 55.

Des amorces spécifiques ont été définies pour chaque variant. De manière à éviter toute amplification liée à une contamination ADN, chaque couple d'amorces destiné à l'amplification spécifique d'un variant d'intérêt est composé d'une amorce complémentaire d'une séquence localisée à la jonction de deux exons (amorces chevauchantes).

Quel que soit le variant considéré, aucune expression n'a pu être caractérisée en culture cellulaire avec ou sans stimulation antigénique.

Nous avons par la suite cherché à vérifier l'expression de ces variants à différents stades de développement (stade œuf, larve, nymphe, adultes). Le variant *14Yde* (V3) s'exprime ainsi chez les moustiques mâles et femelles (**Fig. 28**), mais également à tous les stades de développement (**Fig. 29**). Au stade adulte, l'expression des variants V1 et V5 a pu être caractérisée chez les moustiques mâles ainsi que chez les femelles gorgées (**Fig. 30**). Chez les femelles gorgées, bien que difficilement observable sur la **figure 30A**, une bande à la taille attendue est révélée sur banc UV pour chacun des variants. Aucune amplification n'a cependant été obtenue pour les variants V2 et V4. Nous avons alors tenté de caractériser la protéine codé par le gène *14Yde*.

III.3.3. Recherche d'expression des protéines codées par le gène *cap* (*14Yde*)

La recherche d'une expression en système homologue de la protéine 14YDE a été entreprise en utilisant différents extraits protéiques de moustiques. Après transfert sur membrane et hybridation avec un anticorps anti-peptide ciblant la partie N-terminale de la protéine, une bande de 45 kDa correspondant à la taille attendue de la protéine a pu être mise en évidence mais avec un très faible niveau de détection (**Fig. 31A**). Cependant, cette

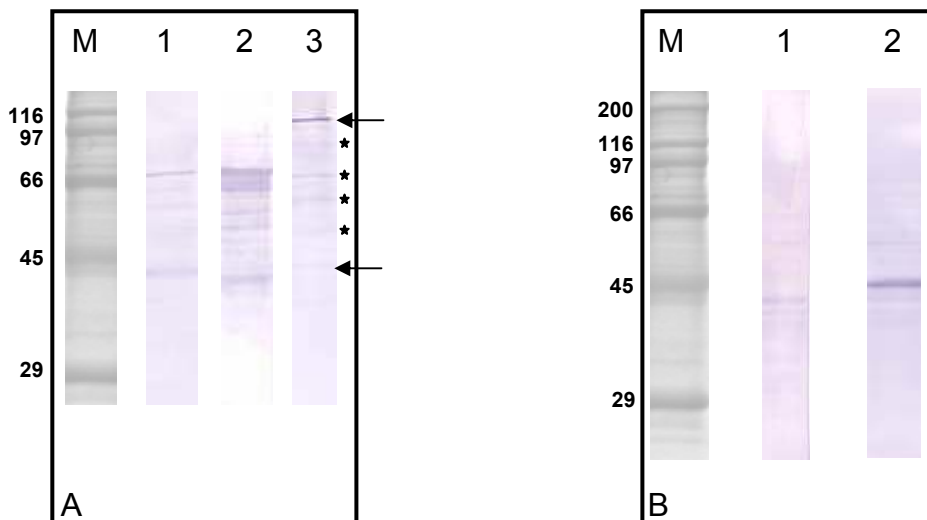


Figure 31. Détection de la protéine 14Yde en système homologue et expression de la protéine recombinante 14Yde-tagV5-His en système cellulaire d'*Anopheles gambiae*. (A) Western blot de protéines après induction en système hétérologue. Piste 1 : témoin non induit, piste 2 : après induction de la protéine recombinante GST-14Yde, piste 3 : échantillons protéiques de moustiques mâles. Les anticorps anti-peptide14Yde sont utilisés au 1/10 000^{ème} (pistes 1 et 2) et au 1/1000^{ème} (piste 3). Les anticorps secondaire anti-IgG de lapin couplés à la phosphatase alcaline sont utilisés au 1/10 000^{ème}. (B) Western blot réalisé à partir d'extraits de cellules transfectées. Piste 1 : cellules non transfectées (témoin), piste 2 : cellules exprimant la protéine 14Yde fusionnée au tag V5. Les anticorps anti-V5 couplés à la phosphatase alcaline sont utilisés au 1/500^{ème}. M : marqueur de poids moléculaire en kDa (Sigma).

Les flèches indiquent la position de la bande à la taille attendue (45 kDa) et celle donnant un marquage intense à 100 kDa. Les étoiles indiquent les bandes révélées ayant une masse moléculaire intermédiaire.

immunodétection a également permis la caractérisation d'une bande supplémentaire à 95-100 kDa. Cette bande apparaît beaucoup plus intense sur certains échantillons protéiques, notamment sur les extraits protéiques provenant de moustiques mâles. D'autres bandes, de poids moléculaires intermédiaires sont également détectées. L'anticorps anti-peptide a été au préalable testé sur un échantillon protéique issu d'une expression en système hétérologue (**Fig. 31A**). Une bande à 66 kDa correspondant à la taille de la protéine taguée de l'épitope GST (26 kDa) est révélée.

Afin d'écarter toute possibilité de formation d'un complexe dimérique de la protéine 14YDE (bande protéique à 100 kDa), nous avons réalisé le clonage du gène *14Yde* en vecteur d'expression pAc5.1 (tagué V5-His), puis effectué une expression en système homologue dans la lignée Sua4.0. L'immunodétection à l'aide d'un anticorps anti-V5 a permis la révélation d'une bande unique à 45 kDa (**Fig. 31B**). Ces premiers résultats semblent exclure l'existence d'un complexe dimérique formé par la protéine 14YDE, et confortent la détection d'une protéine d'environ 100 kDa ayant des déterminants antigéniques communs avec la protéine 14YDE. L'utilisation de cet anticorps anti-peptide sur un extrait protéique de cellules transfectées exprimant 14YDE ne permet pas de révéler de marquage spécifique. L'anticorps anti-peptide n'apparaît donc pas suffisamment spécifique pour détecter la protéine 14YDE.

Nous ne pouvons cependant pas exclure que le tag V5-His gêne la formation d'un possible dimère. De plus, la protéine de haut poids moléculaire pourrait correspondre à un nouveau variant d'épissage même si aucun des variants caractérisés ne code pour une protéine de 95-100 kDa. Enfin, il semble peu probable à la vue des tailles attendues que cette protéine de 100 kDa puisse correspondre à l'un des variants V1-V5 même en tenant compte de modifications post-traductionnelles possibles.

Afin de valider l'une ou l'autre de ces hypothèses, l'identification de cette protéine de 100 kDa a été réalisée par spectrométrie de masse. Pour cela, les extraits protéiques de moustiques mâles, chez lesquels l'expression de cette protéine semble la plus importante, ont été soumis à une migration électrophorétique bidimensionnelle. Suite à la mise au point des conditions de séparation (**Fig. 32**), les protéines ont été transférées sur membrane. Après hybridation avec l'anticorps anti-peptide, les westerns blots nous ont révélé certains spots protéiques d'intérêt dont la masse moléculaire coïncide avec la taille des bandes protéiques caractérisées en gel monodimensionnel (**Fig. 33**). La réalisation de gels semi-préparatifs en vue des prélèvements et des identifications protéiques en spectrométrie de masse a été effectuée (**Fig. 34**). Deux spots protéiques dont la masse moléculaire a été estimée à 45-50 kDa et présentant un pH acide pour l'un, et basique pour l'autre, ont pu être identifiés.

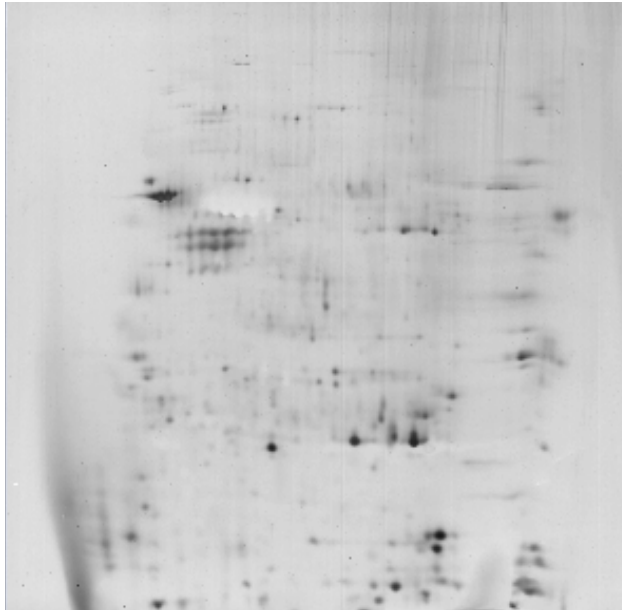


Figure 32. Carte d'électrophorèse bidimensionnelle d'un extrait de protéines de moustiques mâles (têtes). Gel analytique, gradient de pH 3-10, dépôt de 100 µg de protéines, coloration au nitrate d'argent.

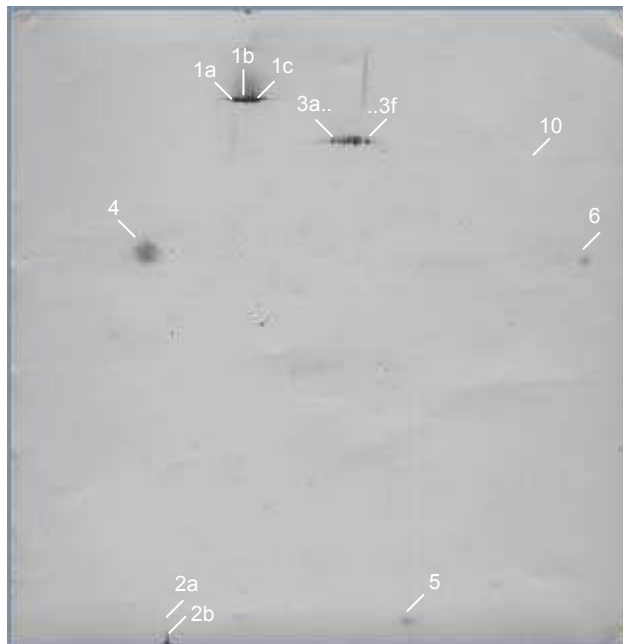


Figure 33. Western blot réalisé sur un extrait protéique de moustiques mâles (têtes). Plusieurs spots protéiques dont les masses moléculaires coïncident avec la taille des bandes protéiques en monodimensionnel sont révélés à 90 -100 kDa ainsi qu'à 70 kDa (isoformes). Deux spots dont la masse moléculaire (45 kDa) correspond à 14YDE sont révélés (pH acide et basique). Tous les spots protéiques ont été prélevés en vue de leur identification. Les anticorps anti-peptides dirigés contre la protéine 14YDE ont été utilisés au 1/2000^{ème}. Les anticorps anti-lapin couplés à la phosphatase alcaline ont été utilisés au 1/10 000^{ème}.

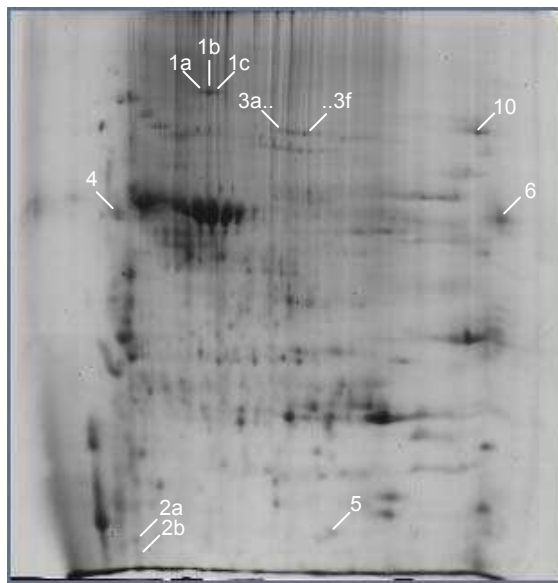


Figure 34. Carte d'électrophorèse bidimensionnelle réalisée à partir d'un extrait protéique de moustiques mâles (têtes). Dépôt de 800 μ g de protéines. Le gel a été Coloré au bleu de Coomassie colloïdal. Les numéros indiquent les spots prélevés et soumis à un identification en spectrométrie de masse MALDI-TOF.

Tableau 9. Liste des protéines prélevées sur gels d'électrophorèse bidimensionnelle et soumises à une identification en spectrométrie de masse MALDI-TOF.

MM : masse moléculaire (kDa), pI : point isoélectrique, Nbre : nombre, AA : acides aminés

Echantillons	MM exp. pI exp.	MM théo. pI théor.	SCORE	COUVERTURE	IDENTIFIANT	Nbre AA	PROTEINE
1b	100 5,6	102 5,45	276	34%	ENSANGP00000013163	877	Myosine
1c	100 5,7	102 5,45	288	34%	ENSANGP00000013163	877	Myosine
3a	80 6,5	78 6,13	96	17%	ENSANGP00000020648	688	Prophénoloxydase
3b	80 6,6	78 6,13	117	23%	ENSANGP00000020648	688	Prophénoloxydase
3c	80 6,7	78 6,13	180	29%	ENSANGP00000020648	688	Prophénoloxydase
3d	80 7	78 6,13	224	32%	ENSANGP00000020648	688	Prophénoloxydase
3e	80 7,1	78 6,13	121	27%	ENSANGP00000020648	688	Prophénoloxydase
4	50 4,3	51 4,53	88	29%	ENSANGP00000013960	483	Protéine synaptique
6	47 9	50 9,16	77	27%	ENSANGP00000018372	463	Facteur d'élongation TEF1- α

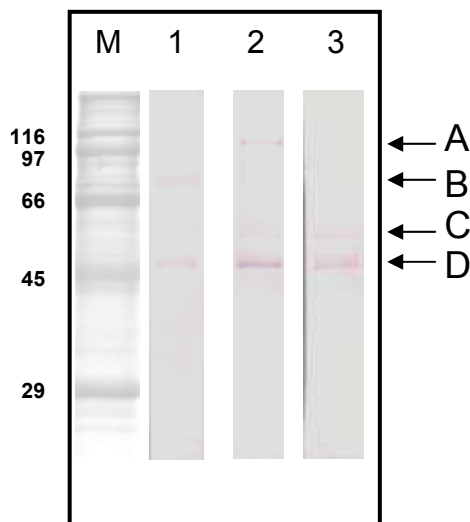


Figure 35. Les anticorps polyclonaux dirigés contre la protéine 14YDE reconnaissent plusieurs bandes protéiques sur différents extraits protéiques de moustiques. Marquage différentiel sur échantillons protéiques de moustiques. Piste 1: femelles gorgées, piste 2 : femelles non gorgées, piste 3 : mâles. Les anticorps anti-14YDE sont utilisés au 1/5000^{ème}. Les anticorps secondaire anti-IgG de souris couplés à la phosphatase alcaline sont utilisés au 1/10 000^{ème}. M : marqueur de poids moléculaire en kDa (Sigma). Les lettres A-D indiquent les bandes prélevées et soumises à une identification par spectrométrie de masse.

Tableau 10. Liste des protéines prélevées sur gels d'électrophorèse monodimensionnelle et soumises à une identification en spectrométrie de masse MALDI-TOF. Les lettres ainsi que les chiffres qui suivent indiquent respectivement la position des différentes bandes prélevées et les pistes correspondantes (Fig. 34).

MM : masse moléculaire (kDa), Nbre : nombre, AA : acides aminées

Echantillons	MM exp.	MM théo.	SCORE	COUVERTURE	IDENTIFIANT	Nbre AA	PROTEINE
A2	100	104	54	20%	ENSANGP00000011796	897	Apha-actinine
B1	70	35,9	30	18%	ENSANGP00000014053	320	AMPC dependant kinase
C1	50	51	108	17%	ENSANGP00000024137	463	ATP synthase (beta su)
C2	50	51	145	23%	ENSANGP00000024137	643	ATP synthase (beta su)
C3	50	51	106	29%	ENSANGP00000024137	463	ATP synthase (beta su)
D1	45	42	38	15%	ENSANGP00000022175	380	Actine
D2	45	42	82	31%	ENSANGP00000022175	380	Actine
D3	45	42	127	49%	ENSANGP00000022175	380	Actine

L'analyse en spectrométrie de masse de ces deux protéines a montré qu'aucune d'entre elles ne correspondait à une protéine codée par le gène *cap* (**Tableau 9**). De la même manière, des spots protéiques à 70 et 100 kDa ont pu être prélevés et identifiés (plusieurs isoformes). Aucun des spots ne correspond à une protéine codée par un des variants du gène d'intérêt. Les spots protéiques 1a, 2a, 2b, 5 et 10 n'apparaissent pas dans le tableau en raison des très faibles scores qui les caractérisent (identification impossible). Le peptide de synthèse utilisé pour la production d'anticorps monoclonaux partage au maximum 6 aa consécutifs (sur les 15 qui le composent) avec les protéines pour lesquelles une réaction croisée a été observée.

Ces résultats confirment le manque de spécificité des anticorps anti-peptides. De nouveaux anticorps dirigés contre la protéine 14YDE complète ont été produits (anticorps polyclonaux). Ceux-ci ont été validés sur des extraits protéiques cellulaires de cellules transfectées exprimant le gène *14Yde*. Une seule bande à 45 kDa est révélée (non montré). Testés sur de nouveaux extraits protéiques de moustiques mâles, ceux-ci ne révèlent plus de bande à 100 kDa. Néanmoins, les bandes en position intermédiaire ainsi que la bande à 45 kDa sont révélées. Ces nouveaux anticorps ont été utilisés afin de détecter l'expression de la protéine 14YDE et/ou des protéines codées par des variants d'épissage du même gène sur des extraits protéiques totaux de moustiques (femelles gorgées, femelles non gorgées, mâles). Après séparation des protéines par électrophorèse monodimensionnelle et transfert sur membrane, des westerns blot ont été réalisés (**Fig. 35**). Une bande protéique de 45 kDa correspondant à la taille attendue de la protéine 14YDE est observée chez les femelles gorgées, non gorgées et chez les mâles. De plus d'autres bandes protéiques sont mises en évidence dans les différents extraits protéiques. Des bandes à 70, 100 et 50 kDa sont détectées respectivement dans les extraits protéiques des femelles gorgées, non gorgées et chez les mâles. Toutes les bandes protéiques ont été prélevées et préparées en vue d'une identification par spectrométrie de masse. Aucune protéine codée par l'un des variants n'a pu être identifiée (**Tableau 10**). La recherche d'un peptide appartenant à l'un des variants possibles au niveau du spectre peptidique massique global de chaque bande protéique soumise à l'analyse s'est avérée négative.

III.4 Conclusion

La mise au point d'un système *in vitro* permettant de valider les efficacités d'inactivation par ARNi a été réalisée. Cette approche représente un préalable important aux essais qui seront effectués *in toto*, en proposant les systèmes d'inactivation génique et les constructions les plus efficaces.

Différents agents de transfection ont pu être testés sur deux lignées cellulaires d'*An. gambiae*, une lignée hémocytaire (Sua4.0) ainsi qu'une lignée embryonnaire (Mos 55). Les meilleurs taux de transfection ont été obtenus avec la Cellfectin. Cet agent de transfection s'est avéré relativement efficace sur la lignée Sua4.0 chez laquelle l'ensemble des essais ont été réalisés. La quantification précise des taux de transfection en cytométrie de flux (FACS, Bioanalyser) a été rendue difficile en raison de la formation d'agrégats cellulaires indissociables. Ainsi, une quantification visuelle permettant d'estimer le niveau d'expression du gène codant l'EGFP a été privilégiée. Celle-ci a permis de mettre en évidence un faible niveau de fluorescence associée à une autofluorescence importante des cellules. L'hypothèse d'une faible efficacité de traduction sous la dépendance d'un environnement inadéquat a été vérifiée. Les différentes constructions réalisées ont permis d'optimiser la "séquence Kozak". Il apparaît judicieux, compte tenu des difficultés rencontrées dans l'estimation des taux de transfection propres à chaque construction, d'envisager l'utilisation d'autres gènes rapporteurs basée sur la quantification d'activités enzymatiques (CAT, β -galactosidase, β -glucuronidase, luciférase). Enfin, des constructions utilisant un promoteur fort et constitutif spécifique d'*An. gambiae* permettraient également d'améliorer nettement les niveaux d'expression.

Concernant l'évaluation de l'efficacité d'une déplétion en ARNm suite à l'introduction d'ARNdb, deux approches ont été utilisées. La première, qui repose sur l'utilisation d'ARNdb de grande taille (600pb) s'est avérée particulièrement efficace dans l'inactivation post-transcriptionnelle du messenger codant l'EGFP. La seconde, basée sur l'utilisation d'un vecteur assurant l'expression d'un miARN a montré une inefficacité du promoteur α -tubuline de drosophile chez *An. gambiae*. La quantification précise de l'inactivation d'un gène rapporteur devra être réalisée par une approche mixte couplant RT-PCR quantitative, microscopie à fluorescence et/ou mesure d'une activité luciférase (Yu et al., 2002 ; Konet et al., 2007).

Chez *D. melanogaster*, 78 miARNs ont été identifiés jusqu'à présent (revue par Jaubert et al., 2007). Ces derniers, au carrefour de nombreuses voies de régulation cellulaire (différenciation musculaire, métabolisme énergétique, croissance cellulaire, homéostasie,

etc.), jouent un rôle essentiel dans les processus de régulation post-transcriptionnelle (**Jaubert et al., 2007**).

Compte tenu d'une part, des progrès récents en matière de reprogrammation de clusters contrôlant la synthèse de plusieurs miARNs chez la drosophile, et d'autre part, de la spécificité des ARNdb de petite taille, l'invalidation génique par miARNs semble prometteuse pour une inactivation ciblée et contrôlée chez le moustique. L'équipe dirigée par C. Antoniewski développe actuellement ces approches chez *D. melanogaster*.

Chez *An. gambiae*, une autre stratégie a été récemment développée (**Konet et al., 2007**). Celle-ci repose sur l'expression de siARNs en "épingle à cheveux" ou shRNA (short hairpin RNA) sous contrôle d'un promoteur U6. L'ensemble de ces travaux ouvrent ainsi de nouvelles perspectives quant à l'utilisation en transgénèse d'un système efficace d'expression d'ARNdb.

Dans le cadre d'une collaboration avec l'équipe de C. Bourguin, l'étude des interactions *Anopheles-Plasmodium* nous a conduit à étudier l'expression des variants du gène *cap*, et plus précisément du variant *14Yde* dont l'expression se voit stimuler 24 h après un repas sanguin infesté par *P. falciparum*. Ce variant pourrait représenter un gène candidat potentiel à inactiver. La confrontation des données théoriques d'annotation automatique avec les données expérimentales nous a permis de caractériser l'existence d'au moins 5 variants pour ce gène (dont *14Yde*). La caractérisation de l'expression de l'ensemble des variants du gène *cap* a été entreprise sur les deux lignées cellulaires Sua4.0 et Mos 55. Quelle que soit la lignée cellulaire, aucun variant du gène *cap* ne s'exprime. La mise en évidence de l'expression des variants à différents stades de développement a montré que le variant V3 (*14Yde*) s'exprime à tous les stades de développement. Ainsi, la mise en évidence d'une expression différentielle de certains de ces variants au cours de l'invasion intestinale devra être réalisée afin d'identifier précisément le ou les variants dont l'expression est dérégulée par une invasion parasitaire (RT-PCR quantitative, immunofluorescence). La fonction de protéine codée par ce gène est encore inconnue chez *An. gambiae*. La détermination de cette fonction permettrait de mieux comprendre les interactions mises en jeu.

Parallèlement, la recherche de l'expression du gène *14Yde* en système homologue a été réalisée. En raison d'un manque de spécificité des anticorps utilisés, l'expression du variant *14Yde* ou de protéines codées par d'autres variants du même gène n'a pu être caractérisée sur les échantillons protéiques de moustiques adultes (mâles, femelles gorgées et non gorgées). Cette difficulté devra être contournée par la production de nouveaux anticorps.

Enfin, le modèle cellulaire a été validé et devrait nous permettre également des études d'interactions de la microsporidie *B. algerae* avec le moustique *An. gambiae*. Nous avons effectivement démontré que cette microsporide pouvait envahir et se multiplier dans les cellules hémocytaires.

CONCLUSION-PERSPECTIVES

Conclusion-Perspectives

La compaction importante des génomes microsporidiens laissait présager l'existence d'un système de régulation transcriptionnelle particulier en raison de la très faible taille des régions intergéniques. Dans le but d'apporter de plus amples informations sur l'organisation des génomes microsporidiens, une partie de mon travail a consisté à étudier, chez *E. cuniculi*, l'existence d'une organisation de type polycistronique. Une étude réalisée sur 4 gènes potentiellement organisés en opéron a permis de caractériser plusieurs populations d'ARNm monocistroniques et polycistroniques. Ces approches expérimentales permettent de proposer un nouveau modèle de régulation transcriptionnelle basé sur la reconnaissance aléatoire du signal de clivage/polyadénylation.

Une meilleure connaissance de la structure des génomes microsporidiens, mais également de leur importantes capacités d'adaptation nous a conduit à séquencer le génome de *B. algerae*, une espèce microsporidienne qui a démontré *in vitro* une capacité de développement au sein de cellules hémocytaires d'insectes. La résolution de son caryotype moléculaire a révélé l'existence d'au moins 30 bandes d'ADN chromosomique, pour une taille estimée à 23 Mpb. L'analyse des premiers jeux de données nous a permis de caractériser, sur l'un des contigs, deux unités d'ADNr en tandem. En aval de chacune de ces unités, de nombreuses répétitions ont été identifiées, indiquant l'existence de plus de deux unités d'ADNr en tandem. L'hybridation moléculaire d'une sonde ciblant l'ARNr 16S nous a également permis de localiser ces unités sur au moins 9 des 30 bandes d'ADN chromosomique. L'annotation syntaxique laisse apparaître un répertoire génique tout aussi conséquent, 3500 CDS ayant été identifiés. Plusieurs classes d'ETs ont pu être caractérisées, ainsi que des gènes codant des enzymes impliquées dans le processus d'ARNi, faisant de *B. algerae* un modèle d'étude inespéré pour la mise au point d'outils de génétique reverse chez

les microsporidies. Au moins 3 arguments en faveur du développement de tels outils chez cette espèce microsporidienne peuvent être avancés. D'une part, l'existence de régions non codantes de taille suffisante favoriserait l'intégration stable de transgènes, et, d'autre part, limiterait l'impact d'une telle intégration sur l'expression de gènes endogènes. Enfin, la mise au point des conditions de transfection serait grandement facilitée par le développement, au contact direct du cytoplasme de la cellule hôte de cette microsporidie. Le séquençage ainsi que l'annotation fonctionnelle du génome de *B. algerae* sont en cours de réalisation. L'identification de l'ensemble des gènes composant le répertoire génique de *B. algerae* est indispensable à la caractérisation de gènes spécifiques des espèces microsporidiennes. Une première analyse comparative a permis de montrer que 17% des gènes de *B. algerae* étaient retrouvés chez au moins une des deux autres espèces étudiées, *E. cuniculi* et *A. locustae*. Ces gènes, communs à plusieurs espèces microsporidiennes, pourraient être impliqués dans l'organisation structurale de ces parasites, le mécanisme d'invasion ou encore la différenciation cellulaire (stades de développement et de résistance). La mise en place des conditions de transfection, couplée au développement d'outils d'inactivation génique chez *B. algerae*, est indispensable à une meilleure compréhension des mécanismes d'adaptation parasitaire chez les microsporidies.

Dans le cadre d'une collaboration visant à développer de nouveaux outils de génomiques fonctionnelles chez *An. gambiae*, un des objectifs de mon travail a été la mise au point des conditions de transfection sur deux lignées cellulaires d'*An. gambiae*, et ceci, afin de déterminer les meilleures conformations d'ARNdb capables d'entraîner une déplétion importante en ARNm. Ces travaux mériteraient d'être poursuivis chez *An. gambiae*, notamment en ce qui concerne l'utilisation de miARNs reprogrammés pour l'extinction ciblée de plusieurs gènes d'intérêt. Les travaux récents de **Konet et al. (2007)** ouvrent également de nouvelles perspectives quant à la mise en place d'une stratégie d'inactivation par ARN

interférence chez *An. gambiae*. Ces outils de génétique reverse, validés sur des modèles cellulaires, permettront le développement de nouveaux vecteurs de transgénèse indispensables à la démonstration, *in toto*, de l'implication de gènes candidats dans le cycle de maturation du parasite.

L'étude des interactions *An. gambiae*/*P. falciparum* nous a conduit à nous intéresser à un gène du moustique qui apparaît surexprimé 24 h après un repas sanguin. Ce gène, qui pourrait être impliqué dans le cycle de maturation du parasite, mais également dans d'autres voies de transduction cellulaire devra faire l'objet d'études complémentaires, au même titre que l'ensemble des variants du gène cap afin de déterminer précisément leur(s) rôle(s) au cours du processus d'invasion.

Enfin, le développement d'un modèle cellulaire d'étude des interactions *An. gambiae*/*B. algerae*, couplé au séquençage récent du génome de cette espèce microsporidienne, devrait également nous permettre de contribuer à l'identification des mécanismes d'adaptation des microsporidies à une vie parasitaire. En effet, nous devrions pouvoir disposer d'un modèle hôte-pathogène où les deux partenaires pourront subir des invalidations géniques par ARNi facilitant ainsi la compréhension de ces mécanismes complexes. Jusqu'à présent, aucun système de transfection ne s'est avéré efficace chez les microsporidies. Nos résultats ouvrent donc des perspectives excitantes dans un contexte de compréhension :

- du cycle de développement des microsporidies,
- des interactions hôte-pathogène avec plus particulièrement la transmission des pathogènes par les vecteurs que sont les moustiques,
- de l'évolution des génomes parasitaires.

REFERENCES BIBLIOGRAPHIQUES

- Abraham, E.G., Islam, S., Srinivasan, P., Ghosh, A.K., Valenzuela, J.G., Ribeiro, J.M., Kafatos, F.C., Dimopoulos, G., and Jacobs-Lorena, M.** (2004). Analysis of the *Plasmodium* and *Anopheles* transcriptional repertoire during ookinete development and midgut invasion. *J Biol Chem.*, **279**, 5573-5580.
- Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear, P.H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., and Kapur, V.** (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*, **304**, 441-445.
- Agabian, N.** (1990). Trans splicing of nuclear pre-mRNAs. *Cell* **61**, 1157-1160.
- Agrawal, N., Dasaradhi, P.V., Mohammed, A., Malhotra, P., Bhatnagar, R.K., and Mukherjee, S.K.** (2003). RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev.*, **67**, 657-685.
- Al-Anouti, F., and Ananvoranich, S.** (2002). Comparative analysis of antisense RNA, double-stranded RNA, and delta ribozyme-mediated gene regulation in *Toxoplasma gondii*. *Antisense Nucleic Acid Drug Dev.*, **12**, 275-281.
- Alavi, Y., Arai, M., Mendoza, J., Tufet-Bayona, M., Sinha, R., Fowler, K., Billker, O., Franke-Fayard, B., Janse, C.J., Waters, A., and Sinden, R.E.** (2003). The dynamics of interactions between *Plasmodium* and the mosquito: a study of the infectivity of *Plasmodium berghei* and *Plasmodium gallinaceum*, and their transmission by *Anopheles stephensi*, *Anopheles gambiae* and *Aedes aegypti*. *Int J Parasitol.*, **33**, 933-943.
- Albright, S.R., and Tjian, R.** (2000). TAFs revisited: more data reveal new twists and confirm old ideas. *Gene*, **242**, 1-13.
- Altmann, M., Edery, I., Sonenberg, N., and Trachsel, H.** (1985). Purification and characterization of protein synthesis initiation factor eIF-4E from the yeast *Saccharomyces cerevisiae*. *Biochemistry*, **24**, 6085-6089.
- Anthony, D., Darrel W., Savage, K.E., and Weidhaas, D.** (1972). Nosematosis: its effect on *Anopheles albimanus* Wiedemann and a population model of its relation to malaria transmission. *Proc. Helminthol. Soc. Wash.*, **39**, 428-433.
- Ares, M., Jr., and Weiser, B.** (1995). Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog Nucleic Acid Res Mol Biol.*, **50**, 131-159.
- Arkhipova, I.R., and Morrison, H.G.** (2001). Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. *Proc Natl Acad Sci U S A*, **98**, 14497-14502.
- Asano, K., Clayton, J., Shalev, A., and Hinnebusch, A.G.** (2000). A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNA(Met) is an important translation initiation intermediate *in vivo*. *Genes Dev.*, **14**, 2534-2546.
- Bakre, A.A., Rawal, K., Ramaswamy, R., Bhattacharya, A., and Bhattacharya, S.** (2005). The LINES and SINES of *Entamoeba histolytica*: comparative analysis and genomic distribution. *Exp Parasitol.*, **110**, 207-213.
- Bangs, J.D., Crain, P.F., Hashizume, T., McCloskey, J.A., and Boothroyd, J.C.** (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem.*, **267**, 9805-9815.
- Bankier, A.T., Spriggs, H.F., Fartmann, B., Konfortov, B.A., Madera, M., Vogel, C., Teichmann, S.A., Ivens, A., and Dear, P.H.** (2003). Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. *Genome Res.*, **13**, 1787-1799.
- Bannister, L., and Mitchell, G.** (2003). The ins, outs and roundabouts of malaria. *Trends Parasitol.*, **19**, 209-213.
- Barillas-Mury, C., and Kumar, S.** (2005). Plasmodium-mosquito interactions: a tale of dangerous liaisons. *Cell Microbiol.*, **7**, 1539-1545.
- Barsoum, R.S.** (2006). Parasitic infections in transplant recipients. *Nat Clin Pract Nephrol.*, **2**, 490-503.
- Bartel, D.P.** (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.
- Benz, C., Nilsson, D., Andersson, B., Clayton, C., and Guilbride, D.L.** (2005). Messenger RNA processing sites in *Trypanosoma brucei*. *Mol Biochem Parasitol.*, **143**, 125-134.

- Best, A.A., Morrison, H.G., McArthur, A.G., Sogin, M.L., and Olsen, G.J.** (2004). Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. *Genome Res.*, **14**, 1537-1547.
- Bhattacharya, S., Bakre, A., and Bhattacharya, A.** (2002). Mobile genetic elements in protozoan parasites. *J Genet.*, **81**, 73-86.
- Biderre, C., Metenier, G., and Vivares, C.P.** (1998). A small spliceosomal-type intron occurs in a ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi*. *Mol Biochem Parasitol.*, **94**, 283-286.
- Biderre, C., Pages, M., Metenier, G., David, D., Bata, J., Prensier, G., and Vivares, C.P.** (1994). On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates. *C R Acad Sci III*, **317**, 399-404.
- Biderre, C., Duffieux, F., Peyretailade, E., Glaser, P., Peyret, P., Danchin, A., Pages, M., Metenier, G., and Vivares, C.P.** (1997). Mapping of repetitive and non-repetitive DNA probes to chromosomes of the microsporidian *Encephalitozoon cuniculi*. *Gene*, **191**, 39-45.
- Biderre C., Canning E. U., Metenier G., and Vivares, C.P.** (1999). Comparison of two isolates of *Encephalitozoon hellem* and *intestinalis* (Microsporida) by pulsed-field gel electrophoresis. *European Journal of Protistology*, **35**, 194-196.
- Bigliardi, E., and Sacchi, L.** (2001). Cell biology and invasion of the microsporidia. *Microbes Infect.*, **3**, 373-379.
- Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., and Levashina, E.A.** (2004). Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell*, **116**, 661-670.
- Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y.** (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453-1474.
- Bohne, W., Ferguson, D.J., Kohler, K., and Gross, U.** (2000). Developmental expression of a tandemly repeated, glycine- and serine-rich spore wall protein in the microsporidian pathogen *Encephalitozoon cuniculi*. *Infect Immun.*, **68**, 2268-2275.
- Boisson, B., Jacques, J.C., Choumet, V., Martin, E., Xu, J., Vernick, K., and Bourgouin, C.** (2006). Gene silencing in mosquito salivary glands by RNAi. *FEBS Lett.*, **580**, 1988-1992.
- Bonnet, S., Prevot, G., Jacques, J.C., Boudin, C., and Bourgouin, C.** (2001). Transcripts of the malaria vector *Anopheles gambiae* that are differentially regulated in the midgut upon exposure to invasive stages of *Plasmodium falciparum*. *Cell Microbiol.*, **3**, 449-458.
- Bowen, N.J., and Jordan, I.K.** (2002). Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol.*, **4**, 65-76.
- Bringaud, F., Ghedin, E., Blandin, G., Bartholomeu, D.C., Caler, E., Levin, M.J., Baltz, T., and El-Sayed, N.M.** (2006). Evolution of non-LTR retrotransposons in the trypanosomatid genomes: *Leishmania major* has lost the active elements. *Mol Biochem Parasitol.*, **145**, 158-170.
- Brivanlou, A.H., and Darnell, J.E., Jr.** (2002). Signal transduction and the control of gene expression. *Science*, **295**, 813-818.
- Brosson, D., Kuhn, L., Delbac, F., Garin, J., C, P.V., and Texier, C.** (2006). Proteomic analysis of the eukaryotic parasite *Encephalitozoon cuniculi* (microsporidia): a reference map for proteins expressed in late sporogonial stages. *Proteomics*, **6**, 3625-3635.
- Broussolle, V., Fumel, S., Peyret, P., and Vivares, C.P.** (1999). Application of differential display RT-PCR to the analysis of gene expression in a host cell-microsporidian *E. cuniculi* interaction. *J Eukaryot Microbiol.*, **46**, 25S-26S.
- Brown, A.E., Bugeon, L., Crisanti, A., and Catteruccia, F.** (2003). Stable and heritable gene silencing in the malaria vector *Anopheles stephensi*. *Nucleic Acids Res.*, **31**, e85.
- Brown, J.R., and Doolittle, W.F.** (1995). Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci U S A*, **92**, 2441-2445.
- Bruchhaus, I., Leippe, M., Lioutas, C., and Tannich, E.** (1993). Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. *DNA Cell Biol.*, **12**, 925-933.

- Brugere, J.F., Cornillot, E., Metenier, G., Bensimon, A., and Vivares, C.P.** (2000). *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. *Nucleic Acids Res.*, **28**, 2026-2033.
- Bryan, R.T., Weber, R., and Schwartz, D.A.** (1997). Microsporidiosis in patients who are not infected with human immunodeficiency virus. *Clin Infect Dis.*, **24**, 534-535.
- Buss, H., Lioutas, C., Dobinsky, S., Nickel, R., and Tannich, E.** (1995). Analysis of the 170-kDa lectin gene promoter of *Entamoeba histolytica*. *Mol Biochem Parasitol.*, **72**, 1-10.
- Cali, A., Weiss, L.M., and Takvorian, P.M.** (2004). An analysis of the microsporidian genus *Brachiola*, with comparisons of human and insect isolates of *Brachiola algerae*. *J Eukaryot Microbiol.*, **51**, 678-685.
- Cali A., and P.M., T.** (1999). Developmental morphology and life cycles of the microsporidia. In M. Wittner, and L. M. Weiss (Eds). *The Microsporidia and Microsporidiosis*. American Society for Microbiology, Washington, D. C., 85-128.
- Callebaut, I., Prat, K., Meurice, E., Mornon, J.P., and Tomavo, S.** (2005). Prediction of the general transcription factors associated with RNA polymerase II in *Plasmodium falciparum*: conserved features and differences relative to other eukaryotes. *BMC Genomics*, **6**, 100.
- Canning E. U., Curry A., Lacey C. J. N., and D., F.J.** (1992). Ultrastructure of *Encephalitozoon* sp. infecting the conjunctival, corneal and nasal epithelia of a patient with AIDS. *Europ. J. Protistol.*, **28**, 226-237.
- Canning, E.U., and Lom, J.** (1986). *The microsporidia of vertebrates*. Academic Press, Inc. New York, N.Y.
- Capy, P., Bazin, C., Higuete, D., and Langin, T.** (1998). *Dynamics and Evolution of Transposable Elements*. Chapman & Hall, New York.
- Carlson, J., Olson, K., Higgs, S., and Beaty, B.** (1995). Molecular genetic manipulation of mosquito vectors. *Annu Rev Entomol.*, **40**, 359-388.
- Carlton, J.M.** (2003). Genome sequencing and comparative genomics of tropical disease pathogens. *Cell Microbiol.*, **5**, 861-873.
- Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Perteau, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., Peterson, J.D., Pop, M., Kosack, D.S., Shumway, M.F., Bidwell, S.L., Shallom, S.J., van Aken, S.E., Riedmuller, S.B., Feldblyum, T.V., Cho, J.K., Quackenbush, J., Sedegah, M., Shoabibi, A., Cummings, L.M., Florens, L., Yates, J.R., Raine, J.D., Sinden, R.E., Harris, M.A., Cunningham, D.A., Preiser, P.R., Bergman, L.W., Vaidya, A.B., van Lin, L.H., Janse, C.J., Waters, A.P., Smith, H.O., White, O.R., Salzberg, S.L., Venter, J.C., Fraser, C.M., Hoffman, S.L., Gardner, M.J., and Carucci, D.J.** (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*, **419**, 512-519.
- Carmichael, G.G.** (2003). Antisense starts making more sense. *Nat Biotechnol.*, **21**, 371-372.
- Catteruccia, F., Nolan, T., Blass, C., Muller, H.M., Crisanti, A., Kafatos, F.C., and Loukeris, T.G.** (2000). Toward *Anopheles* transformation: Minos element activity in anopheline cells and embryos. *Proc Natl Acad Sci U S A*, **97**, 2157-2162.
- Cavalier-Smith, T.** (1987). Eukaryotes with no mitochondria. *Nature*, **326**, 332-333.
- Chaudhuri, J., Chowdhury, D., and Maitra, U.** (1999). Distinct functions of eukaryotic translation initiation factors eIF1A and eIF3 in the formation of the 40 S ribosomal preinitiation complex. *J Biol Chem.*, **274**, 17975-17980.
- Chen, H., Wang, J., Liang, P., Karsay-Klein, M., James, A.A., Brazeau, D., and Yan, G.** (2004). Microarray analysis for identification of *Plasmodium*-refractoriness candidate genes in mosquitoes. *Genome*, **47**, 1061-1070.
- Chioralia, G., Trammer, T., Maier, W.A., and Seitz, H.M.** (1998). Morphologic changes in *Nosema algerae* (Microspora) during extrusion. *Parasitol Res.*, **84**, 123-131.
- Christophides, G.K.** (2005). Transgenic mosquitoes and malaria transmission. *Cell Microbiol.*, **7**, 325-333.

- Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N.T., Hoffmann, J.A., Kanzok, S.M., Letunic, I., Levashina, E.A., Loukeris, T.G., Lycett, G., Meister, S., Michel, K., Moita, L.F., Muller, H.M., Osta, M.A., Paskewitz, S.M., Reichhart, J.M., Rzhetsky, A., Troxler, L., Vernick, K.D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., and Kafatos, F.C.** (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science*, **298**, 159-165.
- Cigan, A.M., and Donahue, T.F.** (1987). Sequence and structural features associated with translational initiator regions in yeast--a review. *Gene*, **59**, 1-18.
- Cigan, A.M., Feng, L., and Donahue, T.F.** (1988). tRNAⁱ(met) functions in directing the scanning ribosome to the start site of translation. *Science*, **242**, 93-97.
- Coates, C.J., Jasinskiene, N., Miyashiro, L., and James, A.A.** (1998). Mariner transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc Natl Acad Sci U S A*, **95**, 3748-3751.
- Conaway, J.W., Shilatifard, A., Dvir, A., and Conaway, R.C.** (2000). Control of elongation by RNA polymerase II. *Trends Biochem Sci.*, **25**, 375-380.
- Couzinet, S., Cejas, E., Schittny, J., Deplazes, P., Weber, R., and Zimmerli, S.** (2000). Phagocytic uptake of *Encephalitozoon cuniculi* by nonprofessional phagocytes. *Infect Immun.*, **68**, 6939-6945.
- Coyle, C.M., Weiss, L.M., Rhodes, L.V., 3rd, Cali, A., Takvorian, P.M., Brown, D.F., Visvesvara, G.S., Xiao, L., Naktin, J., Young, E., Gareca, M., Colasante, G., and Wittner, M.** (2004). Fatal myositis due to the microsporidian *Brachiola algerae*, a mosquito pathogen. *N Engl J Med.*, **351**, 42-47.
- Cross, G.A., Wirtz, L.E., and Navarro, M.** (1998). Regulation of vsg expression site transcription and switching in *Trypanosoma brucei*. *Mol Biochem Parasitol.*, **91**, 77-91.
- Curgy J, Vavra J, and C.P, V.** (1980). Presence of ribosomal ARN with prokaryotic properties in Microsporidia, eukaryotic organisms. *Biol. Cell.*, **38**, 49-52.
- Curry, A., Beeching, N.J., Gilbert, J.D., Scott, G., Rowland, P.L., and Currie, B.J.** (2005). *Trachipleistophora hominis* infection in the myocardium and skeletal muscle of a patient with AIDS. *J Infect.*, **51**, e139-144.
- Curtis, C.F.** (1968). Possible use of translocations to fix desirable genes in insect pest populations. *Nature*, **218**, 368-369.
- Dahmus, M.E.** (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J Biol Chem.*, **271**, 19009-19012.
- DaRocha, W.D., Otsu, K., Teixeira, S.M., and Donelson, J.E.** (2004). Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol.*, **133**, 175-186.
- Dascomb, K., Frazer, T., Clark, R.A., Kissinger, P., and Didier, E.** (2000). Microsporidiosis and HIV. *J Acquir Immune Defic Syndr.*, **24**, 290-292.
- Davis-Hayman, S.R., Shah, P.H., Finley, R.W., Lushbaugh, W.B., and Meade, J.C.** (2000). *Trichomonas vaginalis*: analysis of a heat-inducible member of the cytosolic heat-shock-protein 70 multigene family. *Parasitol Res.*, **86**, 608-612.
- De Rijk, P., Gatehouse, H.S., and De Wachter, R.** (1998a). The secondary structure of *Nosema apis* large subunit ribosomal RNA. *Biochim Biophys Acta*, **1442**, 326-328.
- De Rijk, P., Caers, A., Van de Peer, Y., and De Wachter, R.** (1998b). Database on the structure of large ribosomal subunit RNA. *Nucleic Acids Res.*, **26**, 183-186.
- De, S., Pal, D., and Ghosh, S.K.** (2006). *Entamoeba histolytica* : computational identification of putative microRNA candidates. *Exp Parasitol.*, **113**, 239-243.
- del Aguila, C., Rueda, C., De la Camara, C., and Fenoy, S.** (2001). Seroprevalence of anti-*Encephalitozoon* antibodies in Spanish immunocompetent subjects. *J Eukaryot Microbiol., Suppl*, 75S-78S.
- Delarbre, S., Gatti, S., Scaglia, M., and Drancour, M.** (2001). Genetic diversity in the microsporidian *Encephalitozoon hellem* demonstrated by pulsed-field gel electrophoresis. *J Eukaryot Microbiol.*, **48**, 471-474.

- Delbac, F., Peuvel, I., Metenier, G., Peyretailade, E., and Vivares, C.P.** (2001a). Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of *ptp1* and *ptp2* genes in three *Encephalitozoon* species. *Infect Immun.*, **69**, 1016-1024.
- Delbac, F., Sanger, A., Neuhaus, E.M., Stratmann, R., Ajioka, J.W., Toursel, C., Herm-Gotz, A., Tomavo, S., Soldati, T., and Soldati, D.** (2001b). *Toxoplasma gondii* myosins B/C: one gene, two tails, two localizations, and a role in parasite division. *J Cell Biol.*, **155**, 613-623.
- Deplazes, P., Mathis, A., and Weber, R.** (2000). Epidemiology and zoonotic aspects of microsporidia of mammals and birds. *Contrib Microbiol.*, **6**, 236-260.
- Desportes-Livage, I.** (1996). Human microsporidiosis and AIDS: recent advances. *Parasite*, **3**, 107-113.
- Desportes, I., Le Charpentier, Y., Galian, A., Bernard, F., Cochand-Priollet, B., Lavergne, A., Ravisse, P., and Modigliani, R.** (1985). Occurrence of a new microsporidian: *Enterocytozoon bienewisi* n.g., n. sp., in the enterocytes of a human patient with AIDS. *J Protozool.*, **32**, 250-254.
- Dessens, J.T., Beetsma, A.L., Dimopoulos, G., Wengelnik, K., Crisanti, A., Kafatos, F.C., and Sinden, R.E.** (1999). CTRP is essential for mosquito infection by malaria ookinetes. *Embo J.*, **18**, 6221-6227.
- Dessens, J.T., Siden-Kiamos, I., Mendoza, J., Mahairaki, V., Khater, E., Vlachou, D., Xu, X.J., Kafatos, F.C., Louis, C., Dimopoulos, G., and Sinden, R.E.** (2003). SOAP, a novel malaria ookinete protein involved in mosquito midgut invasion and oocyst development. *Mol Microbiol.*, **49**, 319-329.
- Didier, E.S.** (1998). Microsporidiosis. *Clin Infect Dis.*, **27**, 1-7; quiz 8.
- Didier, E.S.** (2005). Microsporidiosis: an emerging and opportunistic infection in humans and animals. *Acta Trop.*, **94**, 61-76.
- Didier, E.S., and Weiss, L.M.** (2006). Microsporidiosis: current status. *Curr Opin Infect Dis.*, **19**, 485-492.
- Didier, E.S., Stovall, M.E., Green, L.C., Brindley, P.J., Sestak, K., and Didier, P.J.** (2004). Epidemiology of microsporidiosis: sources and modes of transmission. *Vet Parasitol.*, **126**, 145-166.
- Didier, E.S., Bowers, L., Stovall, M.E., Kuebler, D., Mittleider, D., Brindley, P.J., and Didier, P.J.** (2005). Antimicrosporidial activity of (fluoro)quinolones *in vitro* and *in vivo*. *Folia Parasitol (Praha)*, **52**, 173-181.
- Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultz, J., Benes, V., Bork, P., Ansoorge, W., Soares, M.B., and Kafatos, F.C.** (2000). *Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. *Proc Natl Acad Sci U S A*, **97**, 6619-6624.
- Dinglasan, R.R., Fields, I., Shahabuddin, M., Azad, A.F., and Sacci, J.B., Jr.** (2003). Monoclonal antibody MG96 completely blocks *Plasmodium yoelii* development in *Anopheles stephensi*. *Infect Immun.*, **71**, 6995-7001.
- Dolgikh, V.V., and Semenov, P.B.** (2003). Trehalose catabolism in microsporidia *Nosema grylli* spores. *Parazitologiya*, **37**, 333-342.
- Dominguez, D., Altmann, M., Benz, J., Baumann, U., and Trachsel, H.** (1999). Interaction of translation initiation factor eIF4G with eIF4A in the yeast *Saccharomyces cerevisiae*. *J Biol Chem.*, **274**, 26720-26726.
- Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E., and Dimopoulos, G.** (2006). *Anopheles gambiae* immune responses to human and rodent *Plasmodium* parasite species. *PLoS Pathog.*, **2**, e52.
- Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.T., Wu, X., Reith, M., Cavalier-Smith, T., and Maier, U.G.** (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, **410**, 1091-1096.
- Dowd, S.E., Gerba, C.P., and Pepper, I.L.** (1998). Confirmation of the human-pathogenic microsporidia *Enterocytozoon bienewisi*, *Encephalitozoon intestinalis*, and *Vittaforma corneae* in water. *Appl Environ Microbiol.*, **64**, 3332-3335.

- Duffieux, F., Peyret, P., Roe, B.A., and Vivares, C.P. (1998). First report on the systematic sequencing of the small genome of *Encephalitozoon cuniculi* (Protozoa, Microspora): gene organization of a 4.3 kbp region on chromosome I. *Microb Comp Genomics*, **3**, 1-11.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaiia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., and Souciet, J.L. (2004). Genome evolution in yeasts. *Nature*, **430**, 35-44.
- Dumas, C., Chow, C., Muller, M., and Papadopoulou, B. (2006). A novel class of developmentally regulated noncoding RNAs in *Leishmania*. *Eukaryot Cell*, **5**, 2033-2046.
- Edlind, T.D., Li, J., Visvesvara, G.S., Vodkin, M.H., McLaughlin, G.L., and Katiyar, S.K. (1996). Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Mol Phylogenet Evol.*, **5**, 359-367.
- El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renaud, H., Worthey, E.A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D.C., Haas, B.J., Tran, A.N., Wortman, J.R., Alsmark, U.C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J.M., Cerqueira, G.C., Creasy, T., Delcher, A.L., Djikeng, A., Embley, T.M., Hauser, C., Ivens, A.C., Kummerfeld, S.K., Pereira-Leal, J.B., Nilsson, D., Peterson, J., Salzberg, S.L., Shallom, J., Silva, J.C., Sundaram, J., Westenberger, S., White, O., Melville, S.E., Donelson, J.E., Andersson, B., Stuart, K.D., and Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science*, **309**, 404-409.
- Elmendorf, H.G., Singer, S.M., and Nash, T.E. (2001). The abundance of sterile transcripts in *Giardia lamblia*. *Nucleic Acids Res.*, **29**, 4674-4683.
- Enriquez, F.J., Taren, D., Cruz-Lopez, A., Muramoto, M., Palting, J.D., and Cruz, P. (1998). Prevalence of intestinal encephalitozoonosis in Mexico. *Clin Infect Dis.*, **26**, 1227-1229.
- Fast, N.M., and Keeling, P.J. (2001). Alpha and beta subunits of pyruvate dehydrogenase E1 from the microsporidian *Nosema locustae*: mitochondrion-derived carbon metabolism in microsporidia. *Mol Biochem Parasitol.*, **117**, 201-209.
- Fast, N.M., Logsdon, J.M., Jr., and Doolittle, W.F. (1999). Phylogenetic analysis of the TATA box binding protein (TBP) gene from *Nosema locustae*: evidence for a microsporidia-fungi relationship and spliceosomal intron loss. *Mol Biol Evol.*, **16**, 1415-1419.
- Fast, N.M., Law, J.S., Williams, B.A., and Keeling, P.J. (2003). Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution. *Eukaryot Cell*, **2**, 1069-1075.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806-811.
- Flint, J., and Shenk, T. (1997). Viral transactivating proteins. *Annu Rev Genet.*, **31**, 177-212.
- Foucault, C., and Drancourt, M. (2000). Actin mediates *Encephalitozoon intestinalis* entry into the human enterocyte-like cell line, Caco-2. *Microb Pathog.*, **28**, 51-58.
- Fournier, S., Liguory, O., Sarfati, C., David-Ouaknine, F., Derouin, F., Decazes, J.M., and Molina, J.M. (2000). Disseminated infection due to *Encephalitozoon cuniculi* in a patient with AIDS: case report and review. *HIV Med.*, **1**, 155-161.
- Fox, R.M., and Weiser, J. (1959). A microsporidian parasite of *Anopheles gambiae* in Liberia. *J Parasitol.*, **45**, 21-30.
- Franz, G., and Savakis, C. (1991). Minos, a new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. *Nucleic Acids Res.*, **19**, 6646.

- Franzen, C., Muller, A., Hartmann, P., and Salzberger, B.** (2005). Cell invasion and intracellular fate of *Encephalitozoon cuniculi* (Microsporidia). *Parasitology*, **130**, 285-292.
- Fraser, M.J., Ciszczon, T., Elick, T., and Bauser, C.** (1996). Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of *Lepidoptera*. *Insect Mol Biol.*, **5**, 141-151.
- Frixione, E., Ruiz, L., Cerbon, J., and Undeen, A.H.** (1997). Germination of *Nosema algerae* (Microspora) spores: conditional inhibition by D₂O, ethanol and Hg²⁺ suggests dependence of water influx upon membrane hydration and specific transmembrane pathways. *J Eukaryot Microbiol.*, **44**, 109-116.
- Frixione E., Ruiz L., and A.H., U.** (1994). Monovalent cations induce microsporidian spore germination *in vitro*. *J. Euk. Microbiol.*, **41**, 464-468.
- Frixione E., Ruiz L., Santillan M., de Vargas L.V., Tejero J.M., and A.H., U.** (1992). Dynamics of polar tube filament discharge and sporoplasm expulsion by microsporidian spores. *Cell. Motil. Cytoskel.*, **22**, 38-50.
- Frontali, C.** (1994). Genome plasticity in *Plasmodium*. *Genetica*, **94**, 91-100.
- Gajanana, A., Tewari, S.C., Reuben, R., and Rajagopalan, P.K.** (1979). Partial suppression of malaria parasites in *Aedes aegypti* and *Anopheles stephensi* doubly infected with *Nosema algerae* and *Plasmodium*. *Indian J Med Res.*, **70**, 417-423.
- Garcia, L.S.** (2002). Laboratory identification of the microsporidia. *J Clin Microbiol.*, **40**, 1892-1901.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., and Barrell, B.** (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498-511.
- Garza, D., Medhora, M., Koga, A., and Hartl, D.L.** (1991). Introduction of the transposable element mariner into the germline of *Drosophila melanogaster*. *Genetics*, **128**, 303-310.
- Gatehouse, H.S., and Malone, L.A.** (1998). The ribosomal RNA gene region of *Nosema apis* (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. *J Invertebr Pathol.*, **71**, 97-105.
- Geetha Bai, M., Das, P.K., Gajanana, A., and Rajagopalan, P.K.** (1979). Host-parasite relationship of *Nosema algerae*, a parasite of mosquitoes. *Indian J Med Res.*, **70**, 620-624.
- Germot, A., Philippe, H., and Le Guyader, H.** (1997). Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol.*, **87**, 159-168.
- Ghosh, A., Edwards, M.J., and Jacobs-Lorena, M.** (2000). The journey of the malaria parasite in the mosquito: hopes for the new century. *Parasitol Today*, **16**, 196-201.
- Ghosh, A.K., Ribolla, P.E., and Jacobs-Lorena, M.** (2001). Targeting *Plasmodium* ligands on mosquito salivary glands and midgut with a phage display peptide library. *Proc Natl Acad Sci U S A*, **98**, 13278-13281.
- Ghosh, K., Cappiello, C.D., McBride, S.M., Occi, J.L., Cali, A., Takvorian, P.M., McDonald, T.V., and Weiss, L.M.** (2006). Functional characterization of a putative aquaporin from *Encephalitozoon cuniculi*, a microsporidia pathogenic to humans. *Int J Parasitol.*, **36**, 57-62.
- Gill, E.E., and Fast, N.M.** (2006). Assessing the microsporidia-fungi relationship: Combined phylogenetic analysis of eight genes. *Gene*, **375**, 103-109.
- Gilson, P.R., and McFadden, G.I.** (2001). Genome sequencing. A grin without a cat. *Nature*, **410**, 1040-1041.
- Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J., and McFadden, G.I.** (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci U S A*, **103**, 9566-9571.
- Gingras, A.C., Raught, B., and Sonenberg, N.** (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem.*, **68**, 913-963.

- Gottar, M., Gobert, V., Matskevich, A.A., Reichhart, J.M., Wang, C., Butt, T.M., Belvin, M., Hoffmann, J.A., and Ferrandon, D.** (2006). Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell*, **127**, 1425-1437.
- Graves, J.A.** (1998). Background and Overview of Comparative Genomics. *Ilar J.*, **39**, 48-65.
- Grossman, G.L., Rafferty, C.S., Fraser, M.J., and Benedict, M.Q.** (2000). The piggyBac element is capable of precise excision and transposition in cells and embryos of the mosquito, *Anopheles gambiae*. *Insect Biochem Mol Biol.*, **30**, 909-914.
- Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., and Benedict, M.Q.** (2001). Germline transformation of the malaria vector, *Anopheles gambiae*, with the piggyBac transposable element. *Insect Mol Biol.*, **10**, 597-604.
- Gunasekera, A.M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D., and Wirth, D.F.** (2004). Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. *Mol Biochem Parasitol.*, **136**, 35-42.
- Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L.** (1989). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proc Natl Acad Sci U S A*, **86**, 5718-5722.
- Haro, M., Izquierdo, F., Henriques-Gil, N., Andres, I., Alonso, F., Fenoy, S., and del Aguila, C.** (2005). First detection and genotyping of human-associated microsporidia in pigeons from urban parks. *Appl Environ Microbiol.*, **71**, 3153-3157.
- Hausmann, S., Vivares, C.P., and Shuman, S.** (2002). Characterization of the mRNA capping apparatus of the microsporidian parasite *Encephalitozoon cuniculi*. *J Biol Chem.*, **277**, 96-103.
- Hayman, J.R., Southern, T.R., and Nash, T.E.** (2005). Role of sulfated glycans in adherence of the microsporidian *Encephalitozoon intestinalis* to host cells in vitro. *Infect Immun.*, **73**, 841-848.
- Hayman, J.R., Hayes, S.F., Amon, J., and Nash, T.E.** (2001). Developmental expression of two spore wall proteins during maturation of the microsporidian *Encephalitozoon intestinalis*. *Infect Immun.*, **69**, 7057-7066.
- Heintzman, N.D., and Ren, B.** (2007). The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell Mol Life Sci.*, **64**, 386-400.
- Higgs, S., Traul, D., Davis, B.S., Kamrud, K.I., Wilcox, C.L., and Beaty, B.J.** (1996). Green fluorescent protein expressed in living mosquitoes--without the requirement of transformation. *Biotechniques*, **21**, 660-664.
- Hinkle, G., Morrison, H.G., and Sogin, M.L.** (1997). Genes coding for reverse transcriptase, DNA-directed RNA polymerase, and chitin synthase from the microsporidian *Spraguea lophii*. *Biol Bull.*, **193**, 250-251.
- Hirt, R.P., Healy, B., Vossbrinck, C.R., Canning, E.U., and Embley, T.M.** (1997). A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr Biol.*, **7**, 995-998.
- Hirt, R.P., Logsdon, J.M., Jr., Healy, B., Dorey, M.W., Doolittle, W.F., and Embley, T.M.** (1999). Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci U S A*, **96**, 580-585.
- Ho, N.T., Keene, K.M., Olson, K.E., and Zheng, L.** (2003). Characterization of RNA interference in an *Anopheles gambiae* cell line. *Insect Biochem Mol Biol.*, **33**, 949-957.
- Hoffmann, J.A.** (1995). Innate immunity of insects. *Curr Opin Immunol.*, **7**, 4-10.
- Hoffmann, J.A., and Reichhart, J.M.** (2002). *Drosophila* innate immunity: an evolutionary perspective. *Nat Immunol.*, **3**, 121-126.
- Hoffmann, J.A., Reichhart, J.M., and Hetru, C.** (1996). Innate immunity in higher insects. *Curr Opin Immunol.*, **8**, 8-13.

- Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C., Kalush, F., Mural, R.J., Myers, E.W., Adams, M.D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P.T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., and Hoffman, S.L. (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, **298**, 129-149.
- Huang, W.F., Tsai, S.J., Lo, C.F., Soichi, Y., and Wang, C.H. (2004). The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. *Fungal Genet Biol.*, **41**, 473-481.
- Huff, C. (1927). Studies of the infectivity of plasmodia of birds for mosquitoes, with special reference to the problem of immunity in the mosquitoes. *Amer. Jour. Hyg.*, **7**, 706-734.
- Hultmark, D. (2003). *Drosophila* immunity: paths and patterns. *Curr Opin Immunol.*, **15**, 12-19.
- Iiyama, K., Chieda, Y., Yasunaga-Aoki, C., Hayasaka, S., and Shimizu, S. (2004). Analyses of the ribosomal DNA region in *Nosema bombycis* NIS 001. *J Eukaryot Microbiol.*, **51**, 598-604.
- Ishihara R, Y, H., and 1968. Some properties of ribosomes from the sporoplasm of *N. bombycis*. *J. Invertebr. Pathol.*, **377-385**. (1968). Some properties of ribosomes from the sporoplasm of *N. bombycis*. *J. Invertebr. Pathol.*, **11**, 377-385.
- Ito, J., Ghosh, A., Moreira, L.A., Wimmer, E.A., and Jacobs-Lorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature*, **417**, 452-455.
- Ivens, A.C., Peacock, C.S., Worthey, E.A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M.A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S.M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C.V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R.M., Cronin, A., Cruz, A.K., Davies, R.M., De Gaudenzi, J., Dobson, D.E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A.C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J.C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T.M., Price, C., Purnelle, B., Quail, M.A., Rabinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J.C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D.C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D.F., Blackwell, J.M., Stuart, K.D., Barrell, B., and Myler, P.J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science*, **309**, 436-442.

- James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H.T., Rauhut, A., Reeb, V., Arnold, A.E., Amtoft, A., Stajich, J.E., Hosaka, K., Sung, G.H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J.M., Slot, J.C., Wang, Z., Wilson, A.W., Schussler, A., Longcore, J.E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P.M., Powell, M.J., Taylor, J.W., White, M.M., Griffith, G.W., Davies, D.R., Humber, R.A., Morton, J.B., Sugiyama, J., Rossman, A.Y., Rogers, J.D., Pfister, D.H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R.A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R.A., Serdani, M., Crous, P.W., Hughes, K.W., Matsuura, K., Langer, E., Langer, G., Untereiner, W.A., Lucking, R., Budel, B., Geiser, D.M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D.S., Lutzoni, F., McLaughlin, D.J., Spatafora, J.W., and Vilgalys, R.** (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, **443**, 818-822.
- Janeway, C.A., Jr.** (1989). Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol.*, **54 Pt 1**, 1-13.
- Janeway, C.A., Jr., and Medzhitov, R.** (2002). Innate immune recognition. *Annu Rev Immunol.*, **20**, 197-216.
- Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Rafferty, C.S., James, A.A., and Collins, F.H.** (1998). Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the Hermes element from the housefly. *Proc Natl Acad Sci U S A*, **95**, 3743-3747.
- Jaubert, S., Mereau, A., Antoniewski, C., and Tagu, D.** (2007). MicroRNAs in *Drosophila*: The magic wand to enter the Chamber of Secrets? *Biochimie*. in press.
- Juarez, S.I., Putaporntip, C., Jongwutiwes, S., Ichinose, A., Yanagi, T., and Kanbara, H.** (2005). *In vitro* cultivation and electron microscopy characterization of *Trachipleistophora anthropophthera* isolated from the cornea of an AIDS patient. *J Eukaryot Microbiol.*, **52**, 179-190.
- Juven-Gershon, T., Hsu, J.Y., and Kadonaga, J.T.** (2006). Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans.*, **34**, 1047-1050.
- Kadota, K., Ishino, T., Matsuyama, T., Chinzei, Y., and Yuda, M.** (2004). Essential role of membrane-attack protein in malarial transmission to mosquito host. *Proc Natl Acad Sci U S A*, **101**, 16310-16315.
- Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N.** (2005). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev.*, **19**, 104-113.
- Kamaishi, T., Hashimoto, T., Nakamura, Y., Masuda, Y., Nakamura, F., Okamoto, K., Shimizu, M., and Hasegawa, M.** (1996a). Complete nucleotide sequences of the genes encoding translation elongation factors 1 alpha and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. *J Biochem (Tokyo)*, **120**, 1095-1103.
- Kamaishi, T., Hashimoto, T., Nakamura, Y., Nakamura, F., Murata, S., Okada, N., Okamoto, K., Shimizu, M., and Hasegawa, M.** (1996b). Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. *J Mol Evol.*, **42**, 257-263.
- Kapp, L.D., and Lorsch, J.R.** (2004). The molecular mechanics of eukaryotic translation. *Annu Rev Biochem.*, **73**, 657-704.
- Katinka, M.D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivares, C.P.** (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*, **414**, 450-453.
- Kaur, G., and Lohia, A.** (2004). Inhibition of gene expression with double strand RNA interference in *Entamoeba histolytica*. *Biochem Biophys Res Commun.*, **320**, 1118-1122.
- Kawakami Y, Inoue T, Ito K, Kitamizu K, Hanawa C, Ando T, Iwano H, and R., I.** (1994). Identification of a chromosome harboring the small subunit ribosomal RNA gene of *Nosema bombycis*. *J Invertebr Pathol.*, **64**, 147-148.
- Kazazian, H.H., Jr.** (2004). Mobile elements: drivers of genome evolution. *Science*, **303**, 1626-1632.

- Keeling, P.J.** (2003). Congruent evidence from *alpha-tubulin* and *beta-tubulin* gene phylogenies for a zygomycete origin of microsporidia. *Fungal Genet Biol.*, **38**, 298-309.
- Keeling, P.J., and Doolittle, W.F.** (1996). Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol Biol Evol.*, **13**, 1297-1305.
- Keeling, P.J., and Fast, N.M.** (2002). Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu Rev Microbiol.*, **56**, 93-116.
- Keeling, P.J., and Slamovits, C.H.** (2004). Simplicity and complexity of microsporidian genomes. *Eukaryot Cell.*, **3**, 1363-1369.
- Keeling, P.J., and Slamovits, C.H.** (2005). Causes and effects of nuclear genome reduction. *Curr Opin Genet Dev.*, **15**, 601-608.
- Keeling, P.J., Fast, N.M., Law, J.S., Williams, B.A., and Slamovits, C.H.** (2005). Comparative genomics of microsporidia. *Folia Parasitol (Praha)*, **52**, 8-14.
- Khachane, A.N., Timmis, K.N., and Martins dos Santos, V.A.** (2007). Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Mol Biol Evol.*, **24**, 449-456.
- Konet, D.S., Anderson, J., Piper, J., Akkina, R., Suchman, E., and Carlson, J.** (2007). Short-hairpin RNA expressed from polymerase III promoters mediates RNA interference in mosquito cells. *Insect Mol Biol.*, **16**, 199-206.
- Korochkina, S., Barreau, C., Pradel, G., Jeffery, E., Li, J., Natarajan, R., Shabanowitz, J., Hunt, D., Frevert, U., and Vernick, K.D.** (2006). A mosquito-specific protein family includes candidate receptors for malaria sporozoite invasion of salivary glands. *Cell Microbiol.*, **8**, 163-175.
- Koudela, B., Visvesvara, G.S., Moura, H., and Vavra, J.** (2001). The human isolate of *Brachiola algerae* (Phylum Microspora): development in SCID mice and description of its fine structure features. *Parasitology*, **123**, 153-162.
- Kozak, M.** (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283-292.
- Kozak, M.** (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125-8148.
- Kozak, M.** (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.*, **16**, 2482-2492.
- Kozak, M.** (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1-34.
- Kucerova-Pospisilova, Z., and Ditrich, O.** (1998). The serological surveillance of several groups of patients using antigens of *Encephalitozoon hellem* and *E. cuniculi* antibodies to microsporidia in patients. *Folia Parasitol (Praha)*, **45**, 108-112.
- Kucerova-Pospisilova, Z., Secor, W.E., Moura, H., Desportes-Livage, I., Detry, A., Bern, C., Leitch, G., and Visvesvara, G.S.** (2001). An ELISA test to detect human serum antibodies reactive with *Encephalitozoon intestinalis*. *J Eukaryot Microbiol., Suppl*, 73S-74S.
- Lambrechts, L., Chavatte, J.M., Snounou, G., and Koella, J.C.** (2006). Environmental influence on the genetic basis of mosquito resistance to malaria parasites. *Proc Biol Sci.*, **273**, 1501-1506.
- Lau, A.O., Liston, D.R., Vanacova, S., and Johnson, P.J.** (2003). *Trichomonas vaginalis* initiator binding protein, IBP39, contains a novel DNA binding motif. *Mol Biochem Parasitol.*, **130**, 167-171.
- Lau, A.O., Smith, A.J., Brown, M.T., and Johnson, P.J.** (2006). *Trichomonas vaginalis* initiator binding protein (IBP39) and RNA polymerase II large subunit carboxy terminal domain interaction. *Mol Biochem Parasitol.*, **150**, 56-62.
- Lavazec, C., Boudin, C., Lacroix, R., Bonnet, S., Diop, A., Thiberge, S., Boisson, B., Tahar, R., and Bourgoign, C.** (2007). Carboxypeptidases B of *Anopheles gambiae* as targets for a *Plasmodium falciparum* transmission-blocking vaccine. *Infect Immun.*, **75**, 1635-1642.
- Le Rouzic, A., Dupas, S., and Capy, P.** (2007). Genome ecosystem and transposable elements species. *Gene*, **390**, 214-220.
- Lee, T.I., and Young, R.A.** (1998). Regulation of gene expression by TBP-associated proteins. *Genes Dev.*, **12**, 1398-1408.

- Leitch, G.J., Visvesvara, G.S., and He, Q.** (1993). Inhibition of microsporidian spore germination. *Parasitol Today*, **9**, 422-424.
- Leitch, G.J., Ward, T.L., Shaw, A.P., and Newman, G.** (2005). Apical spore phagocytosis is not a significant route of infection of differentiated enterocytes by *Encephalitozoon intestinalis*. *Infect Immun.*, **73**, 7697-7704.
- Lemaitre, B., and Hoffmann, J.** (2007). The Host Defense of *Drosophila melanogaster*. *Annu Rev Immunol.*, **25**, 697-743.
- Levaditi C, Nicolau S, and R., S.** (1923). L'étiologie de l'encéphalite. *C.R. Hebd. Séances Acad. Sci. Paris*, **177**, 985-988.
- Li, L., and Wang, C.C.** (2004). Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. *J Biol Chem.*, **279**, 14656-14664.
- Liang, X.H., Haritan, A., Uliel, S., and Michaeli, S.** (2003). trans and cis splicing in *trypanosomatids*: mechanism, factors, and regulation. *Eukaryot Cell*, **2**, 830-840.
- Liston, D.R., and Johnson, P.J.** (1998). Gene Transcription in *Trichomonas vaginalis*. *Parasitol Today*, **14**, 261-265.
- Liston, D.R., Lau, A.O., Ortiz, D., Smale, S.T., and Johnson, P.J.** (2001). Initiator recognition in a primitive eukaryote: IBP39, an initiator-binding protein from *Trichomonas vaginalis*. *Mol Cell Biol.*, **21**, 7872-7882.
- Liu, Y.J., Hodson, M.C., and Hall, B.D.** (2006). Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes. *BMC Evol Biol.*, **6**, 74.
- Lom, J., and Nilsen, F.** (2003). Fish microsporidia: fine structural diversity and phylogeny. *Int J Parasitol.*, **33**, 107-127.
- Lom J., and J., V.** (1963). The mode of sporoplasm extrusion in microsporidian spores. *Acta Protozool.*, **1**, 81-92.
- Lombardo, F., Nolan, T., Lycett, G., Lanfrancotti, A., Stich, N., Catteruccia, F., Louis, C., Coluzzi, M., and Arca, B.** (2005). An *Anopheles gambiae* salivary gland promoter analysis in *Drosophila melanogaster* and *Anopheles stephensi*. *Insect Mol Biol.*, **14**, 207-216.
- Lopez-Lastra, M., Rivas, A., and Barria, M.I.** (2005). Protein synthesis in eukaryotes: the growing biological relevance of cap-independent translation initiation. *Biol Res.*, **38**, 121-146.
- Lores, B., Lopez-Miragaya, I., Arias, C., Fenoy, S., Torres, J., and del Aguila, C.** (2002). Intestinal microsporidiosis due to *Enterocytozoon bienewisi* in elderly human immunodeficiency virus--negative patients from Vigo, Spain. *Clin Infect Dis.*, **34**, 918-921.
- Lowman, P.M., Takvorian, P.M., and Cali, A.** (2000). The effects of elevated temperatures and various time-temperature combinations on the development of *Brachiola (Nosema) algerae* N. Comb. in mammalian cell culture. *J Eukaryot Microbiol.*, **47**, 221-234.
- Luscher, A., de Koning, H.P., and Maser, P.** (2007). Chemotherapeutic strategies against *Trypanosoma brucei*: drug targets vs. drug targeting. *Curr Pharm Des.*, **13**, 555-567.
- Magaud, A., Achbarou, A., and Desportes-Livage, I.** (1997). Cell invasion by the microsporidium *Encephalitozoon intestinalis*. *J Eukaryot Microbiol.*, **44**, 81S.
- Maggi, P., Larocca, A.M., Quarto, M., Serio, G., Brandonisio, O., Angarano, G., and Pastore, G.** (2000). Effect of antiretroviral therapy on cryptosporidiosis and microsporidiosis in patients infected with human immunodeficiency virus type 1. *Eur J Clin Microbiol Infect Dis.*, **19**, 213-217.
- Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H.O., Bishop, J.R., Falcone, F.H., Gavrilescu, C., Montgomery, J.L., Santori, M.I., Stern, L.S., Wang, Z., Ullu, E., and Tschudi, C.** (2000). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA*, **6**, 163-169.
- Majumdar, R., and Maitra, U.** (2005). Regulation of GTP hydrolysis prior to ribosomal AUG selection during eukaryotic translation initiation. *Embo J.*, **24**, 3737-3746.
- Malhotra, P., Dasaradhi, P.V., Kumar, A., Mohmmmed, A., Agrawal, N., Bhatnagar, R.K., and Chauhan, V.S.** (2002). Double-stranded RNA-mediated gene silencing of cysteine proteases (falcipain-1 and -2) of *Plasmodium falciparum*. *Mol Microbiol.*, **45**, 1245-1254.
- Maniatis, T., and Reed, R.** (2002). An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.

- Manning-Cela, R., Gonzalez, A., and Swindle, J.** (2002). Alternative splicing of LYT1 transcripts in *Trypanosoma cruzi*. *Infect Immun.*, **70**, 4726-4728.
- Marcotrigiano, J., Lomakin, I.B., Sonenberg, N., Pestova, T.V., Hellen, C.U., and Burley, S.K.** (2001). A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery. *Mol Cell.*, **7**, 193-203.
- Margos, G., Maier, W.A., and Seitz, H.M.** (1992). The effect of nosematosis on the development of *Plasmodium falciparum* in *Anopheles stephensi*. *Parasitol Res.*, **78**, 168-171.
- Marmorstein, R.** (2003). Transcription initiation at its most basic level. *Cell*, **115**, 370-372.
- Marrelli, M.T., Li, C., Rasgon, J.L., and Jacobs-Lorena, M.** (2007). Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on *Plasmodium*-infected blood. *Proc Natl Acad Sci U S A*, **104**, 5580-5583.
- Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J.** (2003). Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell.*, **11**, 1291-1299.
- Mathews MB, Sonenberg N, and JWB., H.** (2000). Origins and principles of translational control. In :Hershey JWB, Mathews MB, Sonenberg N, eds. *Translational control of gene expression*. ColdSpring Harbor, New York : Cold Spring Harbor Laboratory Press, 1-31.
- Mathis, A., Weber, R., and Deplazes, P.** (2005). Zoonotic potential of the microsporidia. *Clin Microbiol Rev.*, **18**, 423-445.
- Matsubayashi H., Koike T., Mikata T., and Hagiwara S.** (1959). A case of *Encephalitozoon* like body infection in man. *AMA Arch. Pathol.*, **67**, 181-187.
- Matsuo, H., Li, H., McGuire, A.M., Fletcher, C.M., Gingras, A.C., Sonenberg, N., and Wagner, G.** (1997). Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat Struct Biol.*, **4**, 717-724.
- Matthews, K.R., Tschudi, C., and Ullu, E.** (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. *Genes Dev.*, **8**, 491-501.
- McAndrew, M., Graham, S., Hartmann, C., and Clayton, C.** (1998). Testing promoter activity in the trypanosome genome: isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp Parasitol.*, **90**, 65-76.
- McArthur, A.G., Morrison, H.G., Nixon, J.E., Passamaneck, N.Q., Kim, U., Hinkle, G., Crocker, M.K., Holder, M.E., Farr, R., Reich, C.I., Olsen, G.E., Aley, S.B., Adam, R.D., Gillin, F.D., and Sogin, M.L.** (2000). The *Giardia* genome project database. *FEMS Microbiol Lett.*, **189**, 271-273.
- McDonagh, P.D., Myler, P.J., and Stuart, K.** (2000). The unusual gene organization of *Leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res.*, **28**, 2800-2803.
- McGrane, V., Carlson, J.O., Miller, B.R., and Beaty, B.J.** (1988). Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. *Am J Trop Med Hyg.*, **39**, 502-510.
- McRobert, L., and McConkey, G.A.** (2002). RNA interference (RNAi) inhibits growth of *Plasmodium falciparum*. *Mol Biochem Parasitol.*, **119**, 273-278.
- Meissner, M., and Soldati, D.** (2005). The transcription machinery and the molecular toolbox to control gene expression in *Toxoplasma gondii* and other protozoan parasites. *Microbes Infect.*, **7**, 1376-1384.
- Meister, M., Lemaître, B., and Hoffmann, J.A.** (1997). Antimicrobial peptide defense in *Drosophila*. *Bioessays*, **19**, 1019-1026.
- Meister, S., Koutsos, A.C., and Christophides, G.K.** (2004). The *Plasmodium* parasite--a 'new' challenge for insect innate immunity. *Int J Parasitol.*, **34**, 1473-1482.
- Mellor, J.** (2005). The dynamics of chromatin remodeling at promoters. *Mol Cell.*, **19**, 147-157.
- Merrick, W.C.** (2004). Cap-dependent and cap-independent translation in eukaryotic systems. *Gene*, **332**, 1-11.
- Michel, K., and Kafatos, F.C.** (2005). Mosquito immunity against *Plasmodium*. *Insect Biochem Mol Biol.*, **35**, 677-689.

- Milhausen, M., Nelson, R.G., Sather, S., Selkirk, M., and Agabian, N.** (1984). Identification of a small RNA containing the trypanosome spliced leader: a donor of shared 5' sequences of trypanosomatid mRNAs? *Cell*, **38**, 721-729.
- Militello, K.T., Dodge, M., Bethke, L., and Wirth, D.F.** (2004). Identification of regulatory elements in the *Plasmodium falciparum* genome. *Mol Biochem Parasitol.*, **134**, 75-88.
- Militello, K.T., Patel, V., Chessler, A.D., Fisher, J.K., Kasper, J.M., Gunasekera, A., and Wirth, D.F.** (2005). RNA polymerase II synthesizes antisense RNA in *Plasmodium falciparum*. *RNA*, **11**, 365-370.
- Miller, L.H., Sakai, R.K., Romans, P., Gwadz, R.W., Kantoff, P., and Coon, H.G.** (1987). Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science*, **237**, 779-781.
- Mittleider, D., Green, L.C., Mann, V.H., Michael, S.F., Didier, E.S., and Brindley, P.J.** (2002). Sequence survey of the genome of the opportunistic microsporidian pathogen, *Vittaforma corneae*. *J Eukaryot Microbiol.*, **49**, 393-401.
- Mohammed, A., Dasaradhi, P.V., Bhatnagar, R.K., Chauhan, V.S., and Malhotra, P.** (2003). In vivo gene silencing in *Plasmodium berghei*-a mouse malaria model. *Biochem Biophys Res Commun.*, **309**, 506-511.
- Monnerat, S., Martinez-Calvillo, S., Worthey, E., Myler, P.J., Stuart, K.D., and Fasel, N.** (2004). Genomic organization and gene expression in a chromosomal region of *Leishmania major*. *Mol Biochem Parasitol.*, **134**, 233-243.
- Moreira, L.A., Ghosh, A.K., Abraham, E.G., and Jacobs-Lorena, M.** (2002a). Genetic transformation of mosquitoes: a quest for malaria control. *Int J Parasitol.*, **32**, 1599-1605.
- Moreira, L.A., Ito, J., Ghosh, A., Devenport, M., Zieler, H., Abraham, E.G., Crisanti, A., Nolan, T., Catteruccia, F., and Jacobs-Lorena, M.** (2002b). Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. *J Biol Chem.*, **277**, 40839-40843.
- Morris, A.C., Eggleston, P., and Crampton, J.M.** (1989). Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Med Vet Entomol.*, **3**, 1-7.
- Mottram, J., Perry, K.L., Lizardi, P.M., Luhrmann, R., Agabian, N., and Nelson, R.G.** (1989). Isolation and sequence of four small nuclear U RNA genes of *Trypanosoma brucei subsp. brucei*: identification of the U2, U4, and U6 RNA analogs. *Mol Cell Biol.*, **9**, 1212-1223.
- Motyka, S.A., and Englund, P.T.** (2004). RNA interference for analysis of gene function in trypanosomatids. *Curr Opin Microbiol.*, **7**, 362-368.
- Muller, A., Bialek, R., Kamper, A., Fatkenheuer, G., Salzberger, B., and Franzen, C.** (2001). Detection of microsporidia in travelers with diarrhea. *J Clin Microbiol.*, **39**, 1630-1632.
- Munderloh, U.G., Kurtti, T.J., and Ross, S.E.** (1990). Electrophoretic characterization of chromosomal DNA from two microsporidia. *J Invertebr Pathol.*, **56**, 243-248.
- Myler, P.J., Audleman, L., deVos, T., Hixson, G., Kiser, P., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Fu, G., Ivens, A., and Stuart, K.** (1999). *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A*, **96**, 2902-2906.
- Niare, O., Markianos, K., Volz, J., Oduol, F., Toure, A., Bagayoko, M., Sangare, D., Traore, S.F., Wang, R., Blass, C., Dolo, G., Bouare, M., Kafatos, F.C., Kruglyak, L., Toure, Y.T., and Vernick, K.D.** (2002). Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. *Science*, **298**, 213-216.
- Nikolov, D.B., and Burley, S.K.** (1997). RNA polymerase II transcription initiation: a structural view. *Proc Natl Acad Sci U S A*, **94**, 15-22.
- Nolan, T., Bower, T.M., Brown, A.E., Crisanti, A., and Catteruccia, F.** (2002). piggyBac-mediated germline transformation of the malaria mosquito *Anopheles stephensi* using the red fluorescent protein dsRED as a selectable marker. *J Biol Chem.*, **277**, 8759-8762.
- O'Brochta, D.A., Warren, W.D., Saville, K.J., and Atkinson, P.W.** (1996). Hermes, a functional non-Drosophilid insect gene vector from *Musca domestica*. *Genetics*, **142**, 907-914.
- Oduol, F., Xu, J., Niare, O., Natarajan, R., and Vernick, K.D.** (2000). Genes identified by an expression screen of the vector mosquito *Anopheles gambiae* display differential molecular immune response to malaria parasites and bacteria. *Proc Natl Acad Sci U S A*, **97**, 11397-11402.

- Olson, K.E., Higgs, S., Gaines, P.J., Powers, A.M., Davis, B.S., Kamrud, K.I., Carlson, J.O., Blair, C.D., and Beaty, B.J.** (1996). Genetically engineered resistance to dengue-2 virus transmission in mosquitoes. *Science*, **272**, 884-886.
- Orenstein, J.M., Russo, P., Didier, E.S., Bowers, C., Bunin, N., and Teachey, D.T.** (2005). Fatal pulmonary microsporidiosis due to *Encephalitozoon cuniculi* following allogeneic bone marrow transplantation for acute myelogenous leukemia. *Ultrastruct Pathol.*, **29**, 269-276.
- Osta, M.A., Christophides, G.K., and Kafatos, F.C.** (2004). Effects of mosquito genes on *Plasmodium* development. *Science*, **303**, 2030-2032.
- Palenchar, J.B., and Bellofatto, V.** (2006). Gene transcription in Trypanosomes. *Mol Biochem Parasitol.*, **146**, 135-141.
- Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R., and Ramakrishnan, V.** (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell.*, **26**, 41-50.
- Patterson, D.J.** (1994). Protozoa: Evolution and Systematics. in *Progress in Protozoology. Proceedings of the IX International Congress of Protozoology, Berlin 1993.* (K. Hausmann and N. Hülsmann, eds.) Gustav Fischer Verlag, Stuttgart, Jena, New York., 1-14.
- Perlaza, B.L., Zapata, C., Valencia, A.Z., Hurtado, S., Quintero, G., Sauzet, J.P., Brahimi, K., Blanc, C., Arevalo-Herrera, M., Druilhe, P., and Herrera, S.** (2003). Immunogenicity and protective efficacy of *Plasmodium falciparum* liver-stage Ag-3 in Aotus lemurinus griseimembra monkeys. *Eur J Immunol.*, **33**, 1321-1327.
- Perry, K.L., Watkins, K.P., and Agabian, N.** (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *Proc Natl Acad Sci U S A*, **84**, 8190-8194.
- Pestova, T.V., and Kolupaeva, V.G.** (2002). The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.*, **16**, 2906-2922.
- Peuvel-Fanget, I., Polonais, V., Brosson, D., Texier, C., Kuhn, L., Peyret, P., Vivares, C., and Delbac, F.** (2006). *EnP1* and *EnP2*, two proteins associated with the *Encephalitozoon cuniculi* endospore, the chitin-rich inner layer of the microsporidian spore wall. *Int J Parasitol.*, **36**, 309-318.
- Peuvel, I., Delbac, F., Metenier, G., Peyret, P., and Vivares, C.P.** (2000). Polymorphism of the gene encoding a major polar tube protein PTP1 in two microsporidia of the genus *Encephalitozoon*. *Parasitology*, **121 Pt 6**, 581-587.
- Peuvel, I., Peyret, P., Metenier, G., Vivares, C.P., and Delbac, F.** (2002). The microsporidian polar tube: evidence for a third polar tube protein (PTP3) in *Encephalitozoon cuniculi*. *Mol Biochem Parasitol.*, **122**, 69-80.
- Peyret, P., Katinka, M.D., Duprat, S., Duffieux, F., Barbe, V., Barbazanges, M., Weissenbach, J., Saurin, W., and Vivares, C.P.** (2001). Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora). *Genome Res.*, **11**, 198-207.
- Peyretailade, E., Broussolle, V., Peyret, P., Metenier, G., Gouy, M., and Vivares, C.P.** (1998a). Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Mol Biol Evol.*, **15**, 683-689.
- Peyretailade, E., Biderre, C., Peyret, P., Duffieux, F., Metenier, G., Gouy, M., Michot, B., and Vivares, C.P.** (1998b). Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucleic Acids Res.*, **26**, 3513-3520.
- Phan, L., Zhang, X., Asano, K., Anderson, J., Vornlocher, H.P., Greenberg, J.R., Qin, J., and Hinnebusch, A.G.** (1998). Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. *Mol Cell Biol.*, **18**, 4935-4946.
- Proudfoot, N.** (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol.*, **16**, 272-278.
- Proudfoot, N.J., Furger, A., and Dye, M.J.** (2002). Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.

- Purdy, J.E., Pho, L.T., Mann, B.J., and Petri, W.A., Jr.** (1996). Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol Biochem Parasitol.*, **78**, 91-103.
- Quioco, F.A., Hu, G., and Gershon, P.D.** (2000). Structural basis of mRNA cap recognition by proteins. *Curr Opin Struct Biol.*, **10**, 78-86.
- Quon, D.V., Delgadillo, M.G., Khachi, A., Smale, S.T., and Johnson, P.J.** (1994). Similarity between a ubiquitous promoter element in an ancient eukaryote and mammalian initiator elements. *Proc Natl Acad Sci U S A*, **91**, 4579-4583.
- Raynaud, L., Delbac, F., Broussolle, V., Rabodonirina, M., Girault, V., Wallon, M., Cozon, G., Vivares, C.P., and Peyron, F.** (1998). Identification of *Encephalitozoon intestinalis* in travelers with chronic diarrhea by specific PCR amplification. *J Clin Microbiol.*, **36**, 37-40.
- Refardt, D., and Mouton, L.** (2007). Reverse arrangement of rRNA subunits in the microsporidium *Glugoides intestinalis*. *J Eukaryot Microbiol.*, **54**, 83-85.
- Robinson, K.A., and Beverley, S.M.** (2003). Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol.*, **128**, 217-228.
- Rodrigues, F., Oliveira, S., Rocha, B., and LA., M.** (2006). Germline transformation of *Aedes fluviatilis* (Diptera: Culicidae) with the piggyBac transposable element. *Mem Inst Oswaldo Cruz.*, **101**, 755-757.
- Roeder, R.G.** (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci.*, **21**, 327-335.
- Rogers, G.W., Jr., Richter, N.J., and Merrick, W.C.** (1999). Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *J Biol Chem.*, **274**, 12236-12244.
- Rubin, G.M., and Spradling, A.C.** (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science*, **218**, 348-353.
- Savage KE, Lowe RE, Hazard EI, and CS., L.** (1971). Studies of the transmission of *Plasmodium gallinaceum* by *Anopheles quadrimaculatus* infected with a *Nosema* sp. *Bull World Health Organ.*, **45**, 845-847.
- Schenker, W., Maier, W.A., and Seitz, H.M.** (1992). The effects of *Nosema algerae* on the development of *Plasmodium yoelii nigeriensis* in *Anopheles stephensi*. *Parasitol Res.*, **78**, 56-59.
- Schofield, L., Hewitt, M.C., Evans, K., Siomos, M.A., and Seeberger, P.H.** (2002). Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature*, **418**, 785-789.
- Schumacher, M.A., Lau, A.O., and Johnson, P.J.** (2003). Structural basis of core promoter recognition in a primitive eukaryote. *Cell*, **115**, 413-424.
- Schwartz, A., and Koella, J.C.** (2002). Melanization of *Plasmodium falciparum* and C-25 sephadex beads by field-caught *Anopheles gambiae* (Diptera: Culicidae) from southern Tanzania. *J Med Entomol.*, **39**, 84-88.
- Seeber, F.** (1997). Consensus sequence of translational initiation sites from *Toxoplasma gondii* genes. *Parasitol Res.*, **83**, 309-311.
- Shatkin, A.J., and Manley, J.L.** (2000). The ends of the affair: capping and polyadenylation. *Nat Struct Biol.*, **7**, 838-842.
- Sheoran, A.S., Feng, X., Singh, I., Chapman-Bonofiglio, S., Kitaka, S., Hanawalt, J., Nunnari, J., Mansfield, K., Tumwine, J.K., and Tzipori, S.** (2005). Monoclonal antibodies against *Enterocytozoon bieneusi* of human origin. *Clin Diagn Lab Immunol.*, **12**, 1109-1113.
- Shi, H., Tschudi, C., and Ullu, E.** (2006a). Functional replacement of *Trypanosoma brucei* Argonaute by the human slicer Argonaute2. *RNA*, **12**, 943-947.
- Shi, H., Tschudi, C., and Ullu, E.** (2006b). An unusual Dicer-like protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA*, **12**, 2063-2072.
- Siegel, T.N., Tan, K.S., and Cross, G.A.** (2005). Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Mol Cell Biol.*, **25**, 9586-9594.
- Sinden, R.E., Alavi, Y., and Raine, J.D.** (2004). Mosquito-malaria interactions: a reappraisal of the concepts of susceptibility and refractoriness. *Insect Biochem Mol Biol.*, **34**, 625-629.

- Singer, V.L., Wobbe, C.R., and Struhl, K.** (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev.*, **4**, 636-645.
- Singh, U., Rogers, J.B., Mann, B.J., and Petri, W.A., Jr.** (1997). Transcription initiation is controlled by three core promoter elements in the *hgl5* gene of the protozoan parasite *Entamoeba histolytica*. *Proc Natl Acad Sci U S A*, **94**, 8812-8817.
- Slamovits, C.H., Fast, N.M., Law, J.S., and Keeling, P.J.** (2004). Genome compaction and stability in microsporidian intracellular parasites. *Curr Biol.*, **14**, 891-896.
- Slodkiewicz-Kowalska, A., Graczyk, T.K., Tamang, L., Jedrzejewski, S., Nowosad, A., Zduniak, P., Solarczyk, P., Girouard, A.S., and Majewska, A.C.** (2006). Microsporidian species known to infect humans are present in aquatic birds: implications for transmission via water? *Appl Environ Microbiol.*, **72**, 4540-4544.
- Snowden KF, and JA, S.** (1999). Microsporidia of higher vertebrates. In: Wittner, M., Weiss, L. (Eds), *The microsporidia and microsporidiosis*. American Society of microbiology, Washington, DC., 393-419.
- Sobottka, I., Albrecht, H., Visvesvara, G.S., Pieniazek, N.J., Deplazes, P., Schwartz, D.A., Laufs, R., and Elsner, H.A.** (1999). Inter- and intra-species karyotype variations among microsporidia of the genus *Encephalitozoon* as determined by pulsed-field gel electrophoresis. *Scand J Infect Dis.*, **31**, 555-558.
- Sodqi M, Brazille P, Gonzalez-Canali G, Cornet M, Piketty C, and L., W.** (2004). Unusual pulmonary *Enterocytozoon bienewsi* microsporidiosis in an AIDS patient: case report and review. *Scand J Infect Dis.*, **36**, 230-231.
- Sokolova, Y., Snigirevskaya, E., Morzhina, E., Skarlato, S., Mironov, A., and Komissarchik, Y.** (2001). Visualization of early golgi compartments at proliferate and sporogenic stages of a microsporidian *Nosema grylli*. *J Eukaryot Microbiol.*, **Suppl**, 86S-87S.
- Sokolova, Y.Y., Dolgikh, V.V., Morzhina, E.V., Nasonova, E.S., Issi, I.V., Terry, R.S., Ironside, J.E., Smith, J.E., and Vossbrinck, C.R.** (2003). Establishment of the new genus *Paranosema* based on the ultrastructure and molecular phylogeny of the type species *Paranosema grylli* Gen. Nov., Comb. Nov. (Sokolova, Seleznirov, Dolgikh, Issi 1994), from the cricket *Gryllus bimaculatus* Deg. *J Invertebr Pathol.*, **84**, 159-172.
- Soldati, D., and Boothroyd, J.C.** (1995). A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Mol Cell Biol.*, **15**, 87-93.
- Srinivasan, P., Abraham, E.G., Ghosh, A.K., Valenzuela, J., Ribeiro, J.M., Dimopoulos, G., Kafatos, F.C., Adams, J.H., Fujioka, H., and Jacobs-Lorena, M.** (2004). Analysis of the *Plasmodium* and *Anopheles* transcriptomes during oocyst differentiation. *J Biol Chem.*, **279**, 5581-5587.
- Street, D.A.** (1994). Analysis of *Nosema locustae* (Microsporidia: Nosematidae) chromosomal DNA with pulsed-field gel electrophoresis. *J. Invertebr. Pathol.*, **63**, 301-303.
- Sullivan, W.J., Jr., Dixon, S.E., Li, C., Striepen, B., and Queener, S.F.** (2005). IMP dehydrogenase from the protozoan parasite *Toxoplasma gondii*. *Antimicrob Agents Chemother.*, **49**, 2172-2179.
- Takvorian, P.M., and Cali, A.** (1994). Enzyme histochemical identification of the Golgi apparatus in the microsporidian, *Glugea stephani*. *J Eukaryot Microbiol.*, **41**, 63S-64S.
- Takvorian, P.M., and Cali, A.** (1996). Polar tube formation and nucleoside diphosphatase activity in the microsporidian, *Glugea stephani*. *J Eukaryot Microbiol.*, **43**, 102S-103S.
- Takvorian, P.M., Weiss, L.M., and Cali, A.** (2005). The early events of *Brachiola algerae* (Microsporidia) infection: spore germination, sporoplasm structure, and development within host cells. *Folia Parasitol (Praha)*, **52**, 118-129.
- Thomarat, F., Vivares, C.P., and Gouy, M.** (2004). Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *J Mol Evol.*, **59**, 780-791.
- Tosoni, A., Nebuloni, M., Ferri, A., Bonetto, S., Antinori, S., Scaglia, M., Xiao, L., Moura, H., Visvesvara, G.S., Vago, L., and Costanzi, G.** (2002). Disseminated microsporidiosis caused by *Encephalitozoon cuniculi* III (dog type) in an Italian AIDS patient: a retrospective study. *Mod Pathol.*, **15**, 577-583.

- Trammer, T., Chioralia, G., Maier, W.A., and Seitz, H.M.** (1999). *In vitro* replication of *Nosema algerae* (Microsporidia), a parasite of anopheline mosquitoes, in human cells above 36 degrees C. *J Eukaryot Microbiol.*, **46**, 464-468.
- Tremoulet, A.H., Avila-Aguero, M.L., Paris, M.M., Canas-Coto, A., Ulloa-Gutierrez, R., and Faingezicht, I.** (2004). Albendazole therapy for *Microsporidium* diarrhea in immunocompetent Costa Rican children. *Pediatr Infect Dis J.*, **23**, 915-918.
- Tsai, S.J., Huang, W.F., and Wang, C.H.** (2005). Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene. *J Eukaryot Microbiol.*, **52**, 52-54.
- Tsai, S.J., Kou, G.H., Lo, C.F., and Wang, C.H.** (2002). Complete sequence and structure of ribosomal RNA gene of *Heterosporis anguillarum*. *Dis Aquat Organ.*, **49**, 199-206.
- Tumwine, J.K., Kekitiinwa, A., Bakeera-Kitaka, S., Ndeezi, G., Downing, R., Feng, X., Akiyoshi, D.E., and Tzipori, S.** (2005). Cryptosporidiosis and microsporidiosis in ugandan children with persistent diarrhea with and without concurrent infection with the human immunodeficiency virus. *Am J Trop Med Hyg.*, **73**, 921-925.
- Uesugi, M., Nyanguile, O., Lu, H., Levine, A.J., and Verdine, G.L.** (1997). Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science*, **277**, 1310-1313.
- Ullu, E., Matthews, K.R., and Tschudi, C.** (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. *Mol Cell Biol.*, **13**, 720-725.
- Ullu, E., Tschudi, C., and Chakraborty, T.** (2004). RNA interference in protozoan parasites. *Cell Microbiol.*, **6**, 509-519.
- Undeen A.H, and R.K., V.d.M.** (1994). Conversion of intrasporal trehalose into reducing sugars during germination of *Nosema algerae* (Protista : Microspora) spores : a quantitative study. *J. Eukaryot. Microbiol.*, **41**, 129-132.
- Undeen A.H., and S.W., A.** (1988). Effect of anions on the germination of *Nosema algerae* (Microspora : Nosematidae). *J. Invertebr. Pathol.*, **52**, 84-89.
- Undeen, A.H., and Vander Meer, R.K.** (1990). The effect of ultraviolet radiation on the germination of *Nosema algerae* Vavra and Undeen (Microsporida: Nosematidae) spores. *J Protozool.*, **37**, 194-199.
- Valperga, S.M., de Jogna Prat, S.A., de Valperga, G.J., Lazarte, S.G., de Trejo, A.V., Diaz, N., and Huttman, H.M.** (1999). Microsporidian spores in the stool specimens of toddlers, with or without diarrhea, from Tucuman, Argentina. *Rev Argent Microbiol.*, **31**, 157-164.
- Van de Peer, Y., Ben Ali, A., and Meyer, A.** (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene*, **246**, 1-8.
- van Gool, T., Canning, E.U., and Dankert, J.** (1994). An improved practical and sensitive technique for the detection of microsporidian spores in stool samples. *Trans R Soc Trop Med Hyg.*, **88**, 189-190.
- van Gool, T., Vetter, J.C., Weinmayr, B., Van Dam, A., Derouin, F., and Dankert, J.** (1997). High seroprevalence of *Encephalitozoon* species in immunocompetent subjects. *J Infect Dis.*, **175**, 1020-1024.
- van Gool, T., Biderre, C., Delbac, F., Wentink-Bonnema, E., Peek, R., and Vivares, C.P.** (2004). Serodiagnostic studies in an immunocompetent individual infected with *Encephalitozoon cuniculi*. *J Infect Dis.*, **189**, 2243-2249.
- van Gool, T., Snijders, F., Reiss, P., Eeftinck Schattenkerk, J.K., van den Bergh Weerman, M.A., Bartelsman, J.F., Bruins, J.J., Canning, E.U., and Dankert, J.** (1993). Diagnosis of intestinal and disseminated microsporidial infections in patients with HIV by a new rapid fluorescence technique. *J Clin Pathol.*, **46**, 694-699.
- Vavra, J.** (1965). Study by electron microscope of the morphology and development of some Microsporidia. *C. R. Acad. Sci. Hebd. Seance Acad. Sci. D.*, **261**, 3467-3470.
- Vavra, J.** (1976). Structure of the Microsporidia. In : "Comparative Pathobiology" (L. A. Bulla and T. C. Cheng, eds). Plenum Press, New york, **vol.1**, 1-85.
- Vavra J., and J.I.R., L.** (1999). Structure of the microsporidia. M. Wittner, and L. M. Weiss (Eds), 7-84.

- Visvesvara, G.S., Moura, H., Leitch, G.J., Schwartz, D.A., and Xiao, L.X.** (2005). Public health importance of *Brachiola algerae* (Microsporidia)-an emerging pathogen of humans. *Folia Parasitol (Praha)*, **52**, 83-94.
- Visvesvara, G.S., Belloso, M., Moura, H., Da Silva, A.J., Moura, I.N., Leitch, G.J., Schwartz, D.A., Chevez-Barrios, P., Wallace, S., Pieniazek, N.J., and Goosey, J.D.** (1999). Isolation of *Nosema algerae* from the cornea of an immunocompetent patient. *J Eukaryot Microbiol.*, **46**, 10S.
- Vivarès C.P., Bouix G, and J.F., M.** (1977). *Ormieresia carcini* gen. n., sp. n., Microsporidie du Crabe Méditerranéen, *Carcinus mediterraneus* Czerniavsky, 1884 : cycle évolutif et étude ultrastructurale. *J. Protozool.*, 83-94.
- Vivares, C.P., and Metenier, G.** (2000). Towards the minimal eukaryotic parasitic genome. *Curr Opin Microbiol.*, **3**, 463-467.
- Vivares, C.P., Loubes, C., and Bouix, G.** (1976). Further cytochemical researches on Microsporidia parasites of the Mediterranean shore crab, *Carcinus mediterraneus* Czerniavsky, 1884 (author's transl). *Ann Parasitol Hum Comp.*, **51**, 1-14.
- Vivares, C.P., Gouy, M., Thomarat, F., and Metenier, G.** (2002). Functional and evolutionary analysis of a eukaryotic parasitic genome. *Curr Opin Microbiol.*, **5**, 499-505.
- Vlachou, D., Schlegelmilch, T., Christophides, G.K., and Kafatos, F.C.** (2005). Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion. *Curr Biol.*, **15**, 1185-1195.
- von Allmen, N., Bienz, M., Hemphill, A., and Muller, N.** (2005). Quantitative assessment of sense and antisense transcripts from genes involved in antigenic variation (*vsp* genes) and encystation (*cwp 1* gene) of *Giardia lamblia* clone GS/M-83-H7. *Parasitology*, **130**, 389-396.
- von Der Haar, T., Ball, P.D., and McCarthy, J.E.** (2000). Stabilization of eukaryotic initiation factor 4E binding to the mRNA 5'-Cap by domains of eIF4G. *J Biol Chem.*, **275**, 30551-30555.
- von Der Haar, T., Gross, J.D., Wagner, G., and McCarthy, J.E.** (2004). The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nat Struct Mol Biol.*, **11**, 503-511.
- Vossbrinck, C.R., and Woese, C.R.** (1986). Eukaryotic ribosomes that lack a 5.8S RNA. *Nature*, **320**, 287-288.
- Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A., and Woese, C.R.** (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature*, **326**, 411-414.
- Wahle, E., and Ruegsegger, U.** (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev.*, **23**, 277-295.
- Wang, L.L., Chen, K.P., Zhang, Z., Yao, Q., Gao, G.T., and Zhao, Y.** (2006). Phylogenetic analysis of *Nosema antheraeae* (Microsporidia) isolated from Chinese oak silkworm, *Antheraea pernyi*. *J Eukaryot Microbiol.*, **53**, 310-313.
- Wang, Z., Orlandi, P.A., and Stenger, D.A.** (2005). Simultaneous detection of four human pathogenic microsporidian species from clinical samples by oligonucleotide microarray. *J Clin Microbiol.*, **43**, 4121-4128.
- Wang, Z., Samuelson, J., Clark, C.G., Eichinger, D., Paul, J., Van Dellen, K., Hall, N., Anderson, I., and Loftus, B.** (2003). Gene discovery in the *Entamoeba invadens* genome. *Mol Biochem Parasitol.*, **129**, 23-31.
- Weber R, Deplazes P, and D, S.** (2000). Diagnosis and clinical aspects of human microsporidiosis. *Contrib Microbiol.*, **6**, 166-192.
- Weider, L.J., Elser J.J., Crease T.J., Mateos M., Cotner J.B., and T.A., M.** (2005). The functional significance of ribosomal (r)DNA variation: Impacts on the evolutionary ecology of organisms, **36**, 219-242.
- Weidner, E., and Byrd, W.** (1982). The microsporidian spore invasion tube. II. Role of calcium in the activation of invasion tube discharge. *J Cell Biol.*, **93**, 970-975.
- Weidner E., Byrd W., Scarborough A., Pleshinger J., and D., S.** (1984). Microsporidian spore discharge and the transfert of polaroplast organelle membrane into plasma membrane. *J. Protozool.*, **31**, 195-198.

- Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B.** (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell.*, **2**, 135-140.
- West, M.L., and Corden, J.L.** (1995). Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics*, **140**, 1223-1233.
- Whitten, M.M., Shiao, S.H., and Levashina, E.A.** (2006). Mosquito midguts and malaria: cell biology, compartmentalization and immunology. *Parasite Immunol.*, **28**, 121-130.
- Wichro, E., Hoelzl, D., Krause, R., Bertha, G., Reinthaler, F., and Wenisch, C.** (2005). Microsporidiosis in travel-associated chronic diarrhea in immune-competent patients. *Am J Trop Med Hyg.*, **73**, 285-287.
- Wickstead, B., Ersfeld, K., and Gull, K.** (2003). Repetitive elements in genomes of parasitic protozoa. *Microbiol Mol Biol Rev.*, **67**, 360-375.
- Williams, B.A., and Keeling, P.J.** (2005). Microsporidian mitochondrial proteins: expression in *Antonospora locustae* spores and identification of genes coding for two further proteins. *J Eukaryot Microbiol.*, **52**, 271-276.
- Williams, B.A., Hirt, R.P., Lucocq, J.M., and Embley, T.M.** (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature*, **418**, 865-869.
- Williams, B.A., Slamovits, C.H., Patron, N.J., Fast, N.M., and Keeling, P.J.** (2005). A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci U S A*, **102**, 10936-10941.
- Wiwanitkit, V.** (2006). Intestinal parasite infestation in HIV infected patients. *Curr HIV Res.*, **4**, 87-96.
- Woychik, N.A., and Hampsey, M.** (2002). The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453-463.
- Xia, X.** (2007). The +4G Site in Kozak Consensus is not related to the efficiency of translation initiation. *PLoS ONE*, **2**, e188.
- Xu, J., Pan, G., Fang, L., Li, J., Tian, X., Li, T., Zhou, Z., and Xiang, Z.** (2006). The varying microsporidian genome: existence of long-terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis*. *Int J Parasitol.*, **36**, 1049-1056.
- Xu, X., Dong, Y., Abraham, E.G., Kocan, A., Srinivasan, P., Ghosh, A.K., Sinden, R.E., Ribeiro, J.M., Jacobs-Lorena, M., Kafatos, F.C., and Dimopoulos, G.** (2005). Transcriptome analysis of *Anopheles stephensi-Plasmodium berghei* interactions. *Mol Biochem Parasitol.*, **142**, 76-87.
- Xu, Y., and Weiss, L.M.** (2005). The microsporidian polar tube: a highly specialised invasion organelle. *Int J Parasitol.*, **35**, 941-953.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H.** (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, **97**, 41-51.
- Yamauchi, K.** (1991). The sequence flanking translational initiation site in protozoa. *Nucleic Acids Res.*, **19**, 2715-2720.
- Yoshida, S., and Watanabe, H.** (2006). Robust salivary gland-specific transgene expression in *Anopheles stephensi* mosquito. *Insect Mol Biol.*, **15**, 403-410.
- Yu, J.Y., DeRuiter, S.L., and Turner, D.L.** (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc Natl Acad Sci U S A*, **99**, 6047-6052.
- Yuda, M., Yano, K., Tsuboi, T., Torii, M., and Chinzei, Y.** (2001). von Willebrand Factor A domain-related protein, a novel microneme protein of the malaria ookinete highly conserved throughout *Plasmodium* parasites. *Mol Biochem Parasitol.*, **116**, 65-72.
- Zagulski, M., Nowak, J.K., Le Mouel, A., Nowacki, M., Migdalski, A., Gromadka, R., Noel, B., Blanc, I., Dessen, P., Wincker, P., Keller, A.M., Cohen, J., Meyer, E., and Sperling, L.** (2004). High coding density on the largest *Paramecium tetraurelia* somatic chromosome. *Curr Biol.*, **14**, 1397-1404.
- Zhang, B., Wang, Q., and Pan, X.** (2007). MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol.*, **210**, 279-289.

- Zhang, H., Huang, H., Cali, A., Takvorian, P.M., Feng, X., Zhou, G., and Weiss, L.M.** (2005). Investigations into microsporidian methionine aminopeptidase type 2: a therapeutic target for microsporidiosis. *Folia Parasitol (Praha)*, **52**, 182-192.
- Zhao, J., Hyman, L., and Moore, C.** (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev.*, **63**, 405-445.
- Zhao, Y.G., and Eggleston, P.** (1999). Comparative analysis of promoters for transient gene expression in cultured mosquito cells. *Insect Mol Biol.*, **8**, 31-38.
- Zieler, H., Keister, D.B., Dvorak, J.A., and Ribeiro, J.M.** (2001). A snake venom phospholipase A(2) blocks malaria parasite development in the mosquito midgut by inhibiting ookinete association with the midgut surface. *J Exp Biol.*, **204**, 4157-4167.

Résumé

Les microsporidies sont des parasites intracellulaires obligatoires caractérisés par un mode d'infestation original et, pour certaines espèces, par un très fort degré de compaction du génome nucléaire. Afin de contribuer à une meilleure compréhension des processus d'expression génique chez ces pathogènes opportunistes de l'Homme, des travaux portant sur les mécanismes de régulation transcriptionnelle ont été entrepris. Nos résultats montrent qu'une organisation polycistronique des gènes est retrouvée chez *Encephalitozoon cuniculi* (parasite de mammifères ayant une taille de génome de 2,9 Mb). Au cours du processus de transcription, une maturation aléatoire des messagers conduit à la formation d'une population d'ARNm comprenant des ARNm monocistroniques et polycistroniques. De plus, afin d'améliorer nos connaissances sur la structure des génomes microsporidiens mais également sur leurs capacités d'adaptation, le séquençage du génome de *Brachiola algerae* (microsporidie parasite de mammifères et d'insectes) a été initié en collaboration avec le Genoscope. Trente bandes d'ADN chromosomique de 160 à 2240 kb ont pu être visualisées, pour une taille haploïde de génome estimée à 23 Mb. L'hybridation moléculaire d'une sonde ciblant l'ARNr 16S nous a permis de localiser des unités d'ADNr sur au moins 9 des 30 bandes d'ADN chromosomique. Les premières analyses de ce génome AT riche (>75%) montrent la présence de plusieurs classes d'éléments transposables ainsi que des gènes codant des enzymes impliquées dans le processus d'ARN interférence. Enfin, le développement d'un modèle d'étude *in vitro* des interactions *An. gambiae* / pathogènes a été réalisé, grâce à la démonstration des capacités d'infestation des lignées cellulaires Sua4.0 (cellules hématocytaires) et Mos 55 (cellules embryonnaires) par *B. algerae*. L'invalidation génique par ARN interférence sur ces lignées cellulaires devrait concourir à une meilleure compréhension des mécanismes d'adaptation des parasites *P. falciparum* et *B. algerae* chez *An. gambiae*.

Mots-clé : Microsporidies, génome, opéron, régulation transcriptionnelle, adaptation parasitaire

Abstract

Microsporidia are obligate intracellular parasites characterized by a unique invasive mechanism, and, in some species, by a highly compact nuclear genome. To improve our knowledge of gene expression in these opportunistic pathogens, experiments on transcriptional regulation mechanisms were undertaken. In the microsporidia *Encephalitozoon cuniculi* (a mammalian parasite with a genome size of 2.9 Mb), we demonstrated a polycistronic gene organization. At the time of transcription, the transcripts are randomly processed leading to an mRNA population composed of monocistronic and polycistronic mRNA. To study the structure of microsporidian genomes and their adaptation capacities, the sequencing of *Brachiola algerae* genome (an insect and mammalian microsporidia) has been initiated in collaboration with Genoscope. Thirty chromosome bands from 160 to 2240 kb were identified, giving a haploid genome size estimated of 23 Mb. The study of the chromosomal distribution of the 16S rRNA unit showed the presence of 16S copy on at least 9 chromosome bands. The first analysis of this AT rich genome (more than 75%) revealed the presence of different kinds of transposable elements as well as genes encoding for enzymes involved in the RNA interference process. We further demonstrated that two *Anopheles gambiae* cellular lineages Sua4.0 (haemocytic cell) and Mos 55 (embryonic cells) can be infected by *B. algerae* providing a model to study *in vitro* the interaction between *An. gambiae* and its pathogens. The successful use of the RNA interference in these two types of cells will contribute to a better understanding of the adaptation mechanisms of pathogens like *P. falciparum* and *B. algerae*.

Key words: Microsporidia, genome, operon, transcriptional regulation, parasitic adaptation