



HAL
open science

**Rôle des ARN non-codants dans l'interaction
hôte-parasite : cas de l'interaction
Homme-Microsporidies**

Réginald Florian Kouassi Akossi

► **To cite this version:**

Réginald Florian Kouassi Akossi. Rôle des ARN non-codants dans l'interaction hôte-parasite : cas de l'interaction Homme-Microsporidies. Biologie moléculaire. Université Clermont Auvergne, 2024. Français. NNT : 2024UCFA0003 . tel-04611784

HAL Id: tel-04611784

<https://theses.hal.science/tel-04611784>

Submitted on 14 Jun 2024

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 CLERMONT AUVERGNE

Ecole Doctorale Sciences de le Vie, Santé, Agronomie et Environnement

N° d'ordre :

THESE

Présentée à l'Université Clermont Auvergne par

AKOSSI Réginald Florian Kouassi

En vue de l'obtention du grade de Docteur d'Université

Spécialité : Biologie moléculaire – Parasitologie

Soutenue publiquement le 9 Janvier 2024

***Rôle des ARN non-codants dans l'interaction hôte-parasite :
cas de l'interaction Homme-Microsporidie***

Directeur de thèse :

Dr. Eric PEYRETAILLADE, Maître de conférences HDR UCA, Laboratoire Microorganismes : Génome Environnement (LMGE)
UMR6023 Université Clermont Auvergne, CNRS

Co-Directeur :

Dr. Ivan WAWRZYNIAK, Ingénieur d'étude CNRS, Laboratoire Microorganismes : Génome Environnement (LMGE) UMR6023
Université Clermont Auvergne, CNRS

Présidente du jury :

Pr. Emilie BRASSET, Professeur des universités, Unité Génétique reproduction et développement (iGrED) CNRS, Université
Clermont Auvergne, Inserm, CHU Clermont Ferrand

Rapporteurs :

Dr. Mohamed-Ali HAKIMI, Directeur de recherche, Institut pour l'Avancée des Biosciences (IAB) Grenoble Équipe
Toxoplasmose & Hôte-parasite coévolution, Université Grenoble Alpes, Inserm, CNRS

Dr. Hervé SEITZ, Chargé de recherche Institut de génétique humaine UMR9002 CNRS, Université de Montpellier

Examinatrice :

Dr. Jessica BRYANT, Chargée de recherche Laboratory Biologie of host-parasite interactions, group Plasmodium Chromatin
and Transcription, Institut Pasteur Paris



REMERCIEMENTS

Tout d'abord je souhaiterais remercier Dr. Mohamed-Ali HAKIMI et Dr. Hervé SEITZ rapporteurs de cette thèse, ainsi que Pr. Emilie BRASSET et Dr. Jessica BRYANT, d'avoir accepté d'évaluer ce travail de thèse. Merci également d'avoir accepté de vous déplacer en personne pour l'éventuelle soutenance malgré vos agendas respectifs.

Ensuite je voudrais remercier Éric PEYRETAILLADE et Ivan WAWRZYNIAK, mes deux directeurs de thèse. Merci pour la confiance et le soutien tout au long ces années de thèse. Merci pour la liberté de réflexion et d'action que vous m'avez donnés et l'œil attentif que vous avez su porter tout au long du projet. Merci également d'avoir toujours été disponibles pour des réunions, des présentations ou juste des séances de brainstorming.

Merci également pour les moments plus conviviaux. Ivan, merci pour les longues journées alternant concentration intense et discussions diverses dans le bureau. Eric merci pour les entrées inopinées dès que tu avais deux secondes de libre à l'IUT, pour travailler et passer du temps avec nous. Merci également pour les séances de TP et TD et tous les enseignements que j'ai pu donner au cours de ma thèse. J'ai pu apprendre énormément et rire pas mal aussi.

Merci à toute l'équipe IHP, dont les différents membres faisaient eux aussi des entrées inopinées de temps en temps dans le bureau. Merci à tous les permanents de l'équipe qui bien que ne travaillant pas sur le projet proposaient toujours des conseils avisés couplés à des regards bienveillants et des encouragements. Bees chaleureuses à vous tous.

Merci maintenant à tous les non-permanents,

Aux mentors Maurines et Nastasia qui finissaient leur thèse quand je commençais la mienne mais qui m'ont permis de tout de suite être à l'aise au laboratoire.

Merci à Thania SBAGHDI ma co-thésarde de promo IHP, pour les bons baux, les bonnes blagues et les filons toujours de qualité. Merci aussi pour les séances de running du jeudi, où activement, et régulièrement, je vous regardais aller Ivan et toi depuis ma chaise.

Merci aux « Totally Spies » Eugénie CARRIERE, Juliette GILBERT, et Clémentine MATHIEN; vous êtes géniales les filles. Vraiment très content d'avoir pu passer cette deuxième moitié de thèse avec vous, le woohp sera toujours au top. Merci pour la sublime décoration des couloirs de notre équipe d'ailleurs, et dédicace à Clem ma budy-for-pauses-thé-mais-en-vrai-on-est-productifs-au-bout-du-compte.

Merci maintenant à tous les thésards du LMGE. Tout d'abord à ceux de ma promotion (la meilleure génération comme on nous appelle secrètement). En particulier Vincent FERNANDES (la team des « i » au Mõlkky, invaincue depuis la création du sport),



Lucie LAFFET (la team des « H », meilleur signe astrologique celtique, nous le savons bien) et Roxanne D'HOMMEE (team des « E » comme « E vous êtes sûr qu'on était là ? »)

Merci à Damien et Yoan, membres actifs de la confrérie de la cape et du cercle occulte du terminal respectivement

Merci à toute la génération actuelle de thésards également, cette dernière année a vraiment été un plaisir notamment grâce à vous.

Merci à tous mes amis qui suivaient ma thèse de plus ou moins loin. Qu'ils aient été en thèse eux aussi ou pas du tout, vous n'hésitez pas à envoyer de bonnes ondes et de la force.

Un grand merci à Victorien et Nathalie pour l'administration du laboratoire, et pour m'avoir accompagné plusieurs fois dans mes quêtes aux colis perdus.

Une mention à la dame du CROUS, qui n'a jamais trop posé de questions quand je commandais deux sandwiches un panini une pomme une banane et une bouteille d'eau (ce fut un aspect important de mon expérience de thèse).

Je voudrais maintenant dire un grand merci à Chloé COSTA, pour le SUR-soutien, en particulier pendant ces derniers mois de thèse. Les séances de zumba improvisées, les petites phrases de motivation à la Rocky et les bons petits plats aussi.

Merci également à toute la famille COSTA pour le soutien, et merci à Mr COSTA d'avoir pris le temps de jeter un coup d'œil rapide pour qu'il n'y ait pas trop fautes (un deuxième merci à Chloé, Ivan, Eric, Microsoft word, Google, Reverso, Larousse).

Merci à toute ma famille, à mes frères et sœurs, Xavier, Jeanine, Maxime et Aurélie AKOSSI ; qui ont joué et jouent encore un grand rôle dans la personne que je suis aujourd'hui. Merci aussi à mon neveu Evann et à sa maman.

Merci à Papa et Maman, qui ont toujours été là pour nous, et qui nous ont élevés avec amour et nous ont inculqué des valeurs que je suis content d'avoir aujourd'hui. Merci de me soutenir et de m'encourager inlassablement. Le fait que vous soyez fiers me rend encore plus heureux de faire ce que je fais et d'être qui je suis.

Enfin, je voudrais remercier Dieu, pour m'avoir permis de rencontrer toutes ces personnes et de les avoir dans ma vie, Je suis vraiment heureux de vous avoir rencontrés et je vous souhaite à tous, le meilleur,

A tous,

Akiba, nansin, asanté, thank you



INTRODUCTION

L'IMPORTANCE DES INTERACTIONS DANS LE VIVANT

Le vivant serait apparu sur terre il y a 3.8 à 4.3 milliards d'années. L'étude de son évolution et de sa diversification jusqu'à nos jours, décrit une longue histoire d'interactions, tant avec son environnement qu'entre espèces et individus (1). Le vivant, malgré sa diversité, admet comme élément commun de base la cellule (2). Ces cellules diffèrent grandement entre ordres et règnes mais partagent toutes la capacité d'intégrer des signaux issus de leurs environnements notamment ceux venant d'autres cellules. Cette intégration est centrale au fonctionnement, à la survie, et à l'évolution du vivant (3–5).

Les interactions cellulaires s'effectuent par le biais de molécules et de systèmes divers, plus ou moins conservés au cours de l'évolution (3). Elles peuvent avoir lieu entre cellules d'organismes différents ou au sein d'un même organisme (3,6,7). Parmi ces molécules, les glucides, les lipides et les protéines ainsi que certains de leurs dérivés métaboliques possèdent d'importants rôles de messagers inter-cellulaires ; une importante partie de la littérature du domaine leur est consacrée (3,7,8). Plus récemment, le rôle des acides nucléiques dans les interactions entre cellules a été mis en avant et semble représenter un versant crucial (9–11). Les ARNs non-codant notamment, sont un groupe de molécules hétérogènes dont le rôle dans les interactions entre cellules est de plus en plus décrit.

Une des aspirations de la biologie est de proposer par une étude scientifique rigoureuse, une compréhension plus globale du monde vivant et de son évolution. Dans cet optique, une meilleure appréhension des mécanismes d'interactions pourrait mener à des modèles plus complets et intégrés en biologie.



LE PARASITISME COMME MODE DE VIE

La diversité du vivant s'accompagne également d'une diversification des modes de vie. Certaines espèces sont plus vulnérables à leur environnement et plus dépendantes vis à vis d'autres organismes. Les parasites constituent un apex de cette notion de dépendance : un parasite est défini comme tout être se développant au détriment d'un autre, son hôte (12). Les parasites sont retrouvés dans quasiment tous les règnes (animaux, champignons, plantes, bactéries...) et leurs hôtes potentiels sont tout aussi variés (4,12,13).

L'interaction entre parasites et hôtes peut se faire de manière obligatoire ou facultative, permanente ou ponctuelle, exclusive vis-à-vis d'une espèce ou de manière plus éclectique. Dans le cas d'hôtes multiples, le parasitisme peut se faire de manière indépendante chez chaque hôte ou nécessiter une infection séquentielle des hôtes (4,12,13). Cette diversité de modes de parasitisme décrit un gradient de dépendance et suggère une grande diversité de mécanismes d'interactions et de processus cellulaires associés.

L'étude des parasites constitue également un enjeu sanitaire et économique important (14–17) : plusieurs maladies humaines graves telles que le paludisme ou encore la maladie du sommeil leurs sont associées (6,18). D'autres parasites quant à eux n'affectent pas directement l'Homme mais sont à l'origine de zoonoses avec de fortes répercussions en agriculture, dans l'élevage ou en entomoculture, s'accompagnant de réelles pertes économiques : on peut penser à la schistosomiase du bétail, aux cryptosporidioses bovines ou encore au phénomène d'effondrement des colonies d'abeilles impliquant les microsporidies (15,17,19).

Ainsi, l'étude des parasites et de leur biologie présente le double intérêt de permettre une meilleure compréhension des mécanismes d'interaction du vivant, mais également d'ouvrir la voie vers des solutions de lutte face aux parasitoses d'intérêts sanitaires et économiques.



CONTEXTE SCIENTIFIQUE GENERAL DE CE PROJET DE THESE

L'équipe IHP (Interactions hôtes parasites) du laboratoire LMGE (Laboratoire Microorganismes : Génome et Environnement) est une équipe spécialisée dans l'étude et la compréhension de certaines de ces relations de parasitisme. Elle se focalise principalement sur celles impliquant les microsporidies, un groupe de parasites intracellulaires obligatoires apparenté aux champignons et retrouvé comme infectant quasiment, tout le règne animal (20,21). C'est dans le cadre de leurs projets de recherche que s'inscrit ma thèse qui vise à fournir une meilleure compréhension des interactions hôtes-parasites par l'étude du rôle des ARN non-codants dans ces interactions. Pour ce faire, des méthodes de biologie cellulaire, de biologie moléculaire et de bio-informatique ont été appliquées à un modèle *in vitro* d'interactions entre des cellules humaines et la microsporidie *Anncaliia algerae*. Les résultats obtenus contribuent à la littérature et aux connaissances sur les mécanismes d'évolution et d'interaction des microsporidies ; mais également à la compréhension des parasites de manière plus générale, et du vivant.

Dans la suite de cette thèse, nous utiliserons le sens plus restrictif du mot « parasite » (sens selon Merck & Co. MSD (22)), faisant ainsi référence aux parasites eucaryotes unicellulaires et aux helminthes. Les helminthes ne seront cependant pas développés.



LISTE DES PUBLICATIONS ET DES COMMUNICATIONS ORALES

Les publications et communications suivantes ont pu découler de manière directe ou indirecte de ce projet de thèse :

Publications à lien direct:

-The intracellular parasite *Anncaliia algerae* induces a massive miRNA down-regulation in human cells

Reginald Florian Akossi, Frédéric Delbac, Hicham El Alaoui, Ivan Wawrzyniak, Eric Peyretailade,

Non-coding RNA Research, Volume 8, Issue 3, 2023, Pages 363-375, ISSN 2468-0540, <https://doi.org/10.1016/j.ncrna.2023.05.003>.

-MicroAnnot: A Dedicated Workflow for Accurate Microsporidian Genome Annotation

Jérémy Tournayre, Valérie Polonais, Ivan Wawrzyniak, **Reginald Florian Akossi**, Nicolas Parisot, Emmanuelle Lerat, Frédéric Delbac, Pierre Souvignet, Matthieu Reichstadt, Eric Peyretailade,

International Journal of Molecular Sciences, Preprints 2023, 2023101822. <https://doi.org/10.20944/preprints202310.1822.v1>.

Publication à lien indirect:

-Identification and localization of polar tube proteins in the extruded polar tube of the microsporidian *Anncaliia algerae*

Maurine Fayet, Nastasia Prybylski, Marie-Laure Collin, Eric Peyretailade, Ivan Wawrzyniak, Abdel Belkorchia, **Reginald Florian Akossi**, Marie Diogon, Hicham El Alaoui, Valérie Polonais, Frédéric Delbac,

Scientific Report 13, 8773 (2023). <https://doi.org/10.1038/s41598-023-35511-y>



Communications orales et posters :

-Communication orale, SFM 2023:

“Identification, and role of small interfering RNA in the minimalist parasite *Anncaliia algerae* during an infectious process”

-Communication orale, MicrosporidiaFest 2023:

“The intracellular parasite *Anncaliia algerae* induces a massive miRNA downregulation in human host cells”

-Communication orale, Journées de l’Ecole doctorale 2023 – prix session

“Identification and role of small interfering RNA in the minimalist parasite *Anncaliia algerae* during an infectious process”

-Communication orale Journée de l’AFEM 2022 -2eme ex-aequo prix jeune

“*Anncaliia algerae* derived iRNA as potential players in host-parasite interactions”

-MT180 2022 Demi-finaliste national

-Poster EMBO Parafrap 2022:

“Parasite *Anncaliia algerae* derived iRNA may target host cell genes during infection”



TABLE DES MATIERES

Remerciements.....	3
Introduction.....	7
L'importance des interactions dans le vivant.....	7
Le parasitisme comme mode de vie.....	9
Contexte scientifique général de ce projet de thèse	11
Liste des publications et des communications orales	13
Liste des figures et tables	21
Liste des abréviations	23
Section A : Etude bibliographique	25
Section A : Etude bibliographique	27
Partie I : Génomique parasitaire	27
Chapitre 1 : Les apicomplexés.....	27
I. Cycles parasitaires, pathogénie et génomique.....	29
I.a. Les <i>Plasmodium</i>	29
I.b. Les <i>Toxoplasma</i>	35
I.c. Les <i>Cryptosporidium</i>	39
II. Génomique comparative des apicomplexés.....	41
II.a Adaptations au parasitisme.....	41
II.b. Adaptations à la diversité d'hôtes et aux modes de vie.....	43
Chapitre 2 : Les kinétoplastidés	47
I. Cycles parasitaires et pathogénie	47
I.a. Les <i>Trypanosoma</i>	47
I.b. Les <i>Leishmania</i>	49
II. Génomique comparative des Kinétoplastidés.....	51
II.a. Génomique des Trypanosomatidés.....	51
II.b Génomique comparative des Trypanosomatidés	53
Chapitre 3 : Les Microsporidies.....	55
I. Les microsporidies généralités et biologie	55
I.a Généralités	55
I.b Biologie infectieuse des microsporidies.....	55
I.c. Les microsporidies écologie et distribution.....	57
I.d. Les microsporidies en santé et santé publique	59
II. Les microsporidies : génomique, post-génomique et évolution	59

II.a. Caractéristiques des génomes et gènes microsporidiens	59
II.b. Post-génomique microsporidienne.....	63
II.c La microsporidie <i>Anncaliia algerae</i>	65
Chapitre 4 : Les éléments transposables	67
I. Classification et structures.....	67
I.a La classe I : Les rétrotransposons	67
I.b. La classe II : Les transposons à ADN	71
II. Répartition et rôles.....	73
II.a. Répartition des éléments transposables dans le vivant.....	73
II.b. Rôles des éléments transposables	75
Partie II : Les ARN non codants	77
Chapitre 1 : La régulation par l'ARN.....	79
I. Mise en évidence historique de l'interférence par l'ARN.....	79
II. Les siRNA : biogénèse et rôles	85
II.a. Biogénèse et mode d'action.....	85
II.b. rôles des siRNA dans les interactions hôte-parasite	87
III. Les microRNA.....	91
III.a. Biogénèse et mode d'action.....	91
III.b. Rôles des miRNA dans les interactions hôte-parasite.....	95
IV. Les piRNA.....	99
IV.a. biogénèse et mode d'action.....	99
IV.b. Les piRNA et la régulation des éléments transposables	103
V. Les long ARNs non-codants	105
V.a. Types de LncRNA.....	105
V.b. Les LncRNA dans les interactions hôte-parasite	107
VI. Les circRNA.....	109
VI.a. Biogénèse des circRNA.....	109
VI.b. Les circRNA dans les interactions hôte-parasite.....	109
Chapitre 2 : Diversité évolutive des ARNnc et des processus d'ARN interférence	111
I. Perte de l'ARNi : l'exception killer chez <i>Saccharomyces cerevisiae</i> ; les centromères de <i>Ustilago maydis</i> ; <i>C.gatti</i> ; <i>T.cruzi</i> ; et les <i>Encephalitozoon</i>	111
II. Les ARNnc des parasites : Apicomplexés et Trypanosomatidés.....	117
III. Les ARNnc chez les microsporidies.....	121
Partie III : Objectifs de la thèse	125
Objectif 1 Validation d'un pipeline d'annotation automatique spécifique des génomes microsporidiens	125
Objectif 2 Caractérisation génomique et transcriptomique d' <i>Anncaliia algerae</i>	125

Objectif 3 Etude du rôle de petits ARNs non codants d' <i>Anncaliia algerae</i> dans l'interaction avec sa cellule hôte	127
Objectif 4 Etude de la réponse transcriptomique de la cellule hôte face à une infection par <i>Anncaliia algerae</i>	127
Section B : Etudes expérimentales	129
Section B Etudes expérimentales	131
Modèle d'étude et méthodes générales de la thèse	131
Partie I : Pipeline d'annotation spécifique des génomes microsporidiens : l'outil MicroAnnot	135
Partie II : Etude génomique et transcriptomique générale d' <i>Anncaliia algerae</i>	141
Life under strong transposable element diversity: Genomic and transcriptomic landscape of the microsporidian parasite <i>Anncaliia algerae</i>	143
1. Introduction	145
2. Results	157
3. Discussion	183
4. Conclusion	195
5. Material and methods	195
Partie III : Rôle et origine de petits ARNs interférents d' <i>Anncaliia algerae</i> dans l'interaction avec sa cellule hôte	227
Partie IV : Etude de la réponse transcriptomique de la cellule hôte face à une infection par <i>Anncaliia algerae</i>	279
SECTION C : DISCUSSION GENERALE, CONCLUSIONS ET PERSPECTIVES	285
Section C Discussion générale, conclusions et perspectives	287
Partie I Discussion générale et conclusions	287
Partie II Perspectives	299
SECTION D : RÉFÉRENCES	305
Section D Références	307
SECTION E : ANNEXES	339
Section E Annexes	341
Annexes section B Partie II	341
Annexes section B Partie III	353



LISTE DES FIGURES ET TABLES

Figure 1 : Arbre phylogénétique des apicomplexés d'intérêt humain	28
Figure 2 : Cycle parasitaire des <i>Plasmodium</i> chez l'homme.....	30
Figure 3 : Cycle parasitaire de <i>Toxoplasma gondii</i>	34
Figure 4 : Cycle parasitaire des <i>Cryptosporidium</i>	38
Figure 5 : Représentation schématique des relation phylogénétiques entre eucaryotes avec emphase sur les apicomplexés et les chrompodellids	40
Figure 6 : Exemple du cycle parasitaire de <i>Trypanosoma cruzi</i>	46
Figure 7 : Cycle parasitaire des <i>Leishmania</i>	48
Figure 8 : Structure générale d'une spore microsporidienne.....	54
Figure 9 : Cycle parasitaire des microsporidies	56
Figure 10 : Représentation des tailles de génomes et du nombre de gènes codant des protéines dans le vivant	58
Figure 11 : Phylogénie, hôtes et caractéristiques des microsporidies.....	60
Figure 12 : Schéma des voies d'infection par <i>Anncaliia algerae</i>	64
Tableau 1 : Classification des éléments transposables eucaryotes selon Wicker <i>et al</i> (144)	66
Figure 13 : Principaux mécanismes de transposition	68
Figure 14 : Structures des éléments transposables eucaryotes	70
Figure 15 : Distribution des éléments transposables dans la phylogénie eucaryote	72
Figure 16 : Différentes classes d'ARN	76
Figure 17 : Petits ARN interférents, partenaires Argonautes, et cibles principales.....	80
Tableau 2 : Types et fonctions des petits ARN interférents	82
Figure 18 : Voies de biogénèse des siRNA	84
Figure 19 : Régulation post-transcriptionnelle par les siRNA	86
Figure 20 : Voies canoniques et non canoniques de biogénèse des miRNA chez l'Homme.....	90
Figure 21 : Exemple d'interactions miRNA-cible	92
Figure 22 : Mécanismes de régulation post-transcriptionnelle des miRNA	94
Figure 23 : Biogénèse des piRNA et cycle ping-pong.....	98
Figure 24 : Fonctions des Longs ARN non-codants.....	104
Figure 25 : Biogénèse et fonctions des circRNA	108
Figure 26 : Arbre phylogénétique de 19 champignons et de la présence d'ARNi chez eux	112
Figure 27 : Conservation et perte de l'ARNi chez les trypanosomatidés.....	114
Figure 28 : Arbre phylogénétique de 21 microsporidies et présence ou absence du couple Dicer/Argonaute. 116	
Figure 29 : Régulation de gènes <i>var</i> par des LncRNA chez <i>P.falciparum</i>	118



LISTE DES ABREVIATIONS

A adénosine ; **ADN** acide désoxyribonucléique ; **ARN** acide ribonucléique ; **ARNnc** ARN non-codant ; **ARNr** ARN ribosomique ; **ARNt** ARN de transfert ; **b** base ; **C** cytosine ; **CDS** coding sequence ; **cf** confère ; **ET** élément transposable ; **G** guanosine ; **Kb** kilobase ; **Mb** mégabase ; **nt** nucléotide ; **ORF** open reading frame ; **snARN** small nuclear RNA ; **snoARN** small nucleolar RNA ; **T** thymidine ; **U** uridine ;



Section A : Etude bibliographique



SECTION A : ETUDE BIBLIOGRAPHIQUE

PARTIE I : GENOMIQUE PARASITAIRE

Les parasites ont une importance particulière en biologie en raison des enjeux sanitaires et économiques qu'ils représentent. Ce sont également des modèles intéressants pour l'étude de questions fondamentales en biologie, par leurs modes de vie et leurs caractéristiques uniques. Etudier les génomes et les comparer entre espèces éclaire grandement sur les processus d'adaptation et d'interaction mis en place par le vivant au cours de son évolution.

CHAPITRE 1 : LES APICOMPLEXES

Les apicomplexés sont des organismes unicellulaires parasites obligatoires d'animaux (23). Une de leur caractéristique commune réside dans la morphologie particulière de leurs cellules qui présentent un complexe apical permettant l'invasion des cellules hôtes (24). Les apicomplexés parasitant l'Homme sont regroupés dans la sous-classe des Eucoccidies, elle-même scindée en deux ordres : les Eucoccidés, parasites de divers tissus à l'exception du sang et infectant l'homme par voie féco-orale ; et les Haemosporididés, parasites de divers tissus dont le sang et transmis à l'homme par piqûre d'un arthropode hématophage (25) (**cf. figure 1** (26)). Ces parasites sont responsables de pathologies majeures chez l'Homme mais également chez les animaux (27).

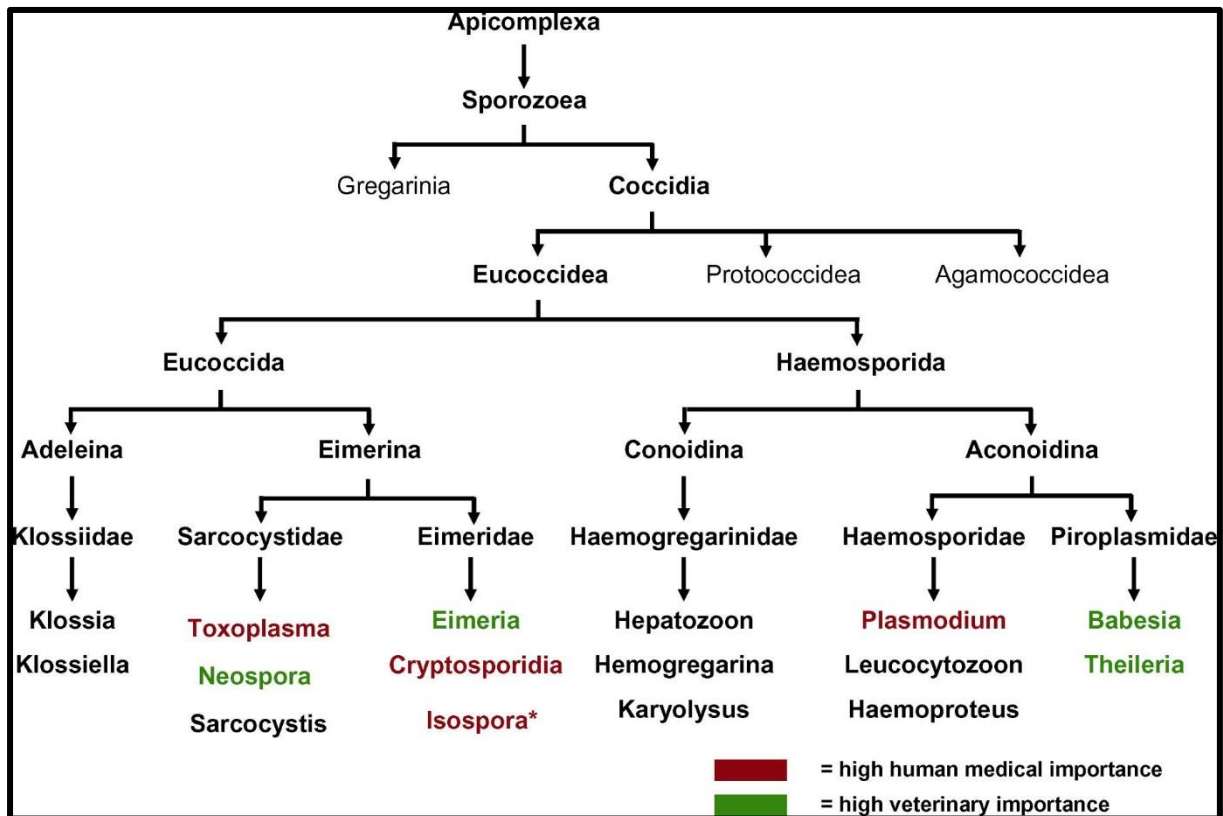


FIGURE 1 : ARBRE PHYLOGENETIQUE DES APICOMPLEXES D'INTERET HUMAIN

Arbre phylogénétique des apicomplexés parasitant l'Homme avec indiqués en rouge les parasites d'importance médicale et en vert ceux d'intérêt vétérinaire (Hans-Peter *et. al.*, 2009 (26)).

Les *Plasmodium* sont une famille des haemosporidés. Ces parasites sont hétéroxènes, c'est à dire effectuant leur cycle de vie dans au moins deux hôtes avec un hôte dit intermédiaire où le parasite effectue une réplique asexuée, et un hôte définitif, chez qui a lieu la reproduction sexuée du parasite (25) (cf. **figure 2**). L'hôte définitif des *Plasmodium* est un moustique du genre *Anopheles*. Leurs hôtes intermédiaires, eux, dépendront d'avantage de l'espèce de *Plasmodium* et peuvent inclure des oiseaux, des mammifères et même des reptiles (28,29).

Cinq espèces de *Plasmodium* sont connues comme infectant l'Homme et sont responsables du paludisme ; il s'agit de : *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* et *Plasmodium knowlesi* (28).

Dans les 5 cas le parasite est transmis à l'homme par la piqûre d'un moustique du genre *Anopheles* préalablement infecté au cours d'un repas de sang (28). Seuls les *Anopheles* femelles sont hématophages et la transmission du parasite ne se fait donc que par elles (28).

Chez l'anophèle, les gamètes mâles et femelles du parasite vont s'accoupler dans la lumière intestinale et former des œufs. Ces œufs évolueront en ookinètes, capables de traverser la barrière intestinale et de s'enkyster au niveau de sa lame basale formant des oocystes. La rupture de l'oocyste libèrera dans la circulation générale du moustique des sporozoïtes mobiles qui rejoindront les glandes salivaires d'où ils pourront infecter l'Homme lors d'un repas sanguin de l'anophèle femelle (25,28,30).

Chez l'homme, une fois dans la circulation sanguine, le parasite infecte en premier lieu le foie. Les sporozoïtes envahissent les hépatocytes et se multiplient activement (on parle de schizogonie hépatique). Ils sont ensuite libérés sous forme de mérozoïtes par lyse des cellules et vont retourner dans la circulation sanguine où ils infecteront les hématies. Dans les hématies les mérozoïtes vont passer par une phase dites en « anneau », puis devenir des trophozoïtes jeunes (phase d'alimentation active), des trophozoïtes âgés, et enfin se répliquer activement sous forme de schizontes jusqu'à la lyse de l'hématie, libérant de nouveaux

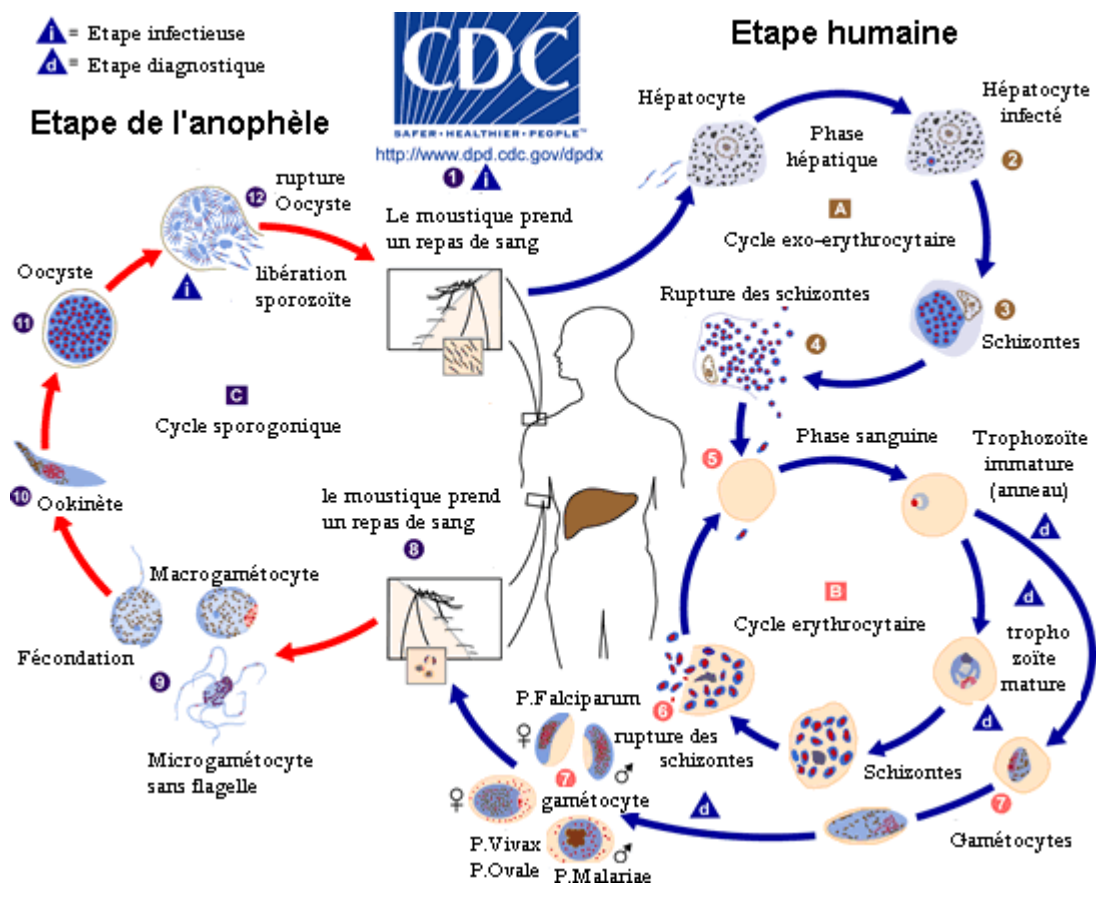


FIGURE 2 : CYCLE PARASITAIRE DES PLASMODIUM CHEZ L'HOMME

Représentation schématique du cycle parasitaire des plasmodiums chez l'Homme. Les flèches rouges représentent les étapes ayant lieu chez l'anophèle et les bleues celles se déroulant chez l'Homme. Les nombres indiquent l'enchaînement des différentes étapes et les lettres correspondent aux différents cycles. (D'après le site officiel du programme national de lutte contre le paludisme de Côte d'Ivoire (PNLPCI) (31), lui-même adapté du schéma du Centers for disease Control and Prevention américain (CDC) (32)).

mérozoïtes capables d'infecter d'autres hématies. Dans certains cas, les mérozoïtes qui infectent les hématies se différencient en gamétocytes mâles ou femelles. Ce sont ces mêmes gamétocytes qui pourront être récupérés par un anophèle lors d'un repas de sang et ainsi compléter le cycle biologique du parasite. Les différents hôtes intermédiaires constituent les réservoirs naturels des *Plasmodium* (25,28,30).

Les symptômes du paludisme varient en fonction des *Plasmodium* mais apparaissent principalement au cours de la phase sanguine du développement parasitaire chez l'Homme. Des dommages hépatiques sont également décrits, surtout lorsque l'infection est chronique ou récidivante (33).

En 2021 L'organisation mondiale pour la santé (OMS) estimait à 247 millions le nombre de cas annuels de paludisme. Le nombre de décès annuels associés était estimé à 619 000 et concernait principalement les populations des régions tropicales, en majorité localisées en Afrique avec 4 pays comptant pour plus de la moitié des décès mondiaux (Le Nigéria (31,3 %), la République démocratique du Congo (12,6 %), la Tanzanie (4,1 %) et le Niger (3,9 %)). *Plasmodium falciparum* est l'agent paludéen le plus virulent suivi de *Plasmodium vivax* (33).

Il existe des traitements et des prophylaxies antipaludéennes, et des vaccins sont également en développement (34). Plus la prise en charge est précoce plus les traitements sont efficaces (33). Des cas de résistance à certains traitements ont cependant été observés, notamment de la part de certaines souches de *P. falciparum* exprimant une protéine nommée PfCRT (*P.falciparum* Chloroquine Resistance Transporter) (35).

En termes de génomique, les *Plasmodium* sont haploïdes tout au long de leur cycle de vie. Leurs génomes sont de taille relativement réduite allant de 20 à 25 mégabases (Mb) pour 14 chromosomes. Un plastide circulaire d'environ 35 kilobases (Kb) et plusieurs copies d'ADN mitochondriaux d'environ 6 Kb sont également présents. Le pourcentage GC moyen des génomes est variable, allant de 19.3% à 42.4% selon les espèces. Il est souvent



inversement associé à l'abondance de séquences répétées simples (microsatellites) retrouvées dans les génomes (36). La structure chromosomique chez les *Plasmodium* est conservée au niveau du cœur des chromosomes, mais présente un fort polymorphisme de longueur et de séquence dans les régions sub-téломériques et téломériques. De nombreux sites de SNP (Single Nucléotide Polymorphisme, polymorphisme d'un seul nucléotide) et un nombre variable de copies de gènes (CNV, Copie Number Variation) sont également observés entre isolats d'une même espèce (36).

Les *Plasmodium* présentent entre 5009 (*P.vinckei*) et 8582 (*P.ovale*) gènes codants et non-codants (36,37). Les gènes codants des protéines orthologues entre *Plasmodium* sont généralement retrouvés en blocs de synténie. Ces blocs peuvent cependant être ordonnés différemment entre génomes (36). Des familles multigéniques existent chez les *Plasmodium* et sont particulièrement étudiées pour leurs implications dans la virulence du parasite et dans ses interactions avec ses hôtes intermédiaires (36,38–40). Il s'agit de gènes codant des protéines aux déterminants antigéniques variables, mais conservant leurs fonctions (variant surface antigens, (VSA)) (36,38–40). Ces protéines sont principalement exprimées à la surface des cellules infectées et des parasites (38,39). Elles sont impliquées dans l'adhésion du parasite aux cellules, dans la formation de structures infectieuses comme les rosettes d'érythrocytes (rosetting), et surtout, dans l'échappement immunitaire grâce à la variabilité de leurs déterminants antigéniques (36,38–40). Ces familles multigéniques sont retrouvées de manière conservée dans les régions sub-téломériques des chromosomes chez les *Plasmodium* (36). Le nombre de copies des gènes de ces familles peut varier entre individus et une grande diversité existe de par les phénomènes de mutations ponctuelles et de recombinaison. Ainsi, même une population clonale de *Plasmodium* pourra présenter une diversité pour ces gènes.

Chez *P.falciparum* par exemple, chaque individu posséderait une combinaison propre de 50 à 150 gènes *var*, une des familles multigéniques les plus étudiées (36)). Néanmoins, à un instant, un seul gène *var* est exprimé, permettant ainsi un échappement immunitaire efficace et chronique (on parle d'expression mutuellement exclusive des gènes *var*) (41). Les mécanismes sous-jacents de la régulation et du choix du gène *var* exprimé sont encore en investigation et impliqueraient notamment des régulations épigénétiques (41,42).

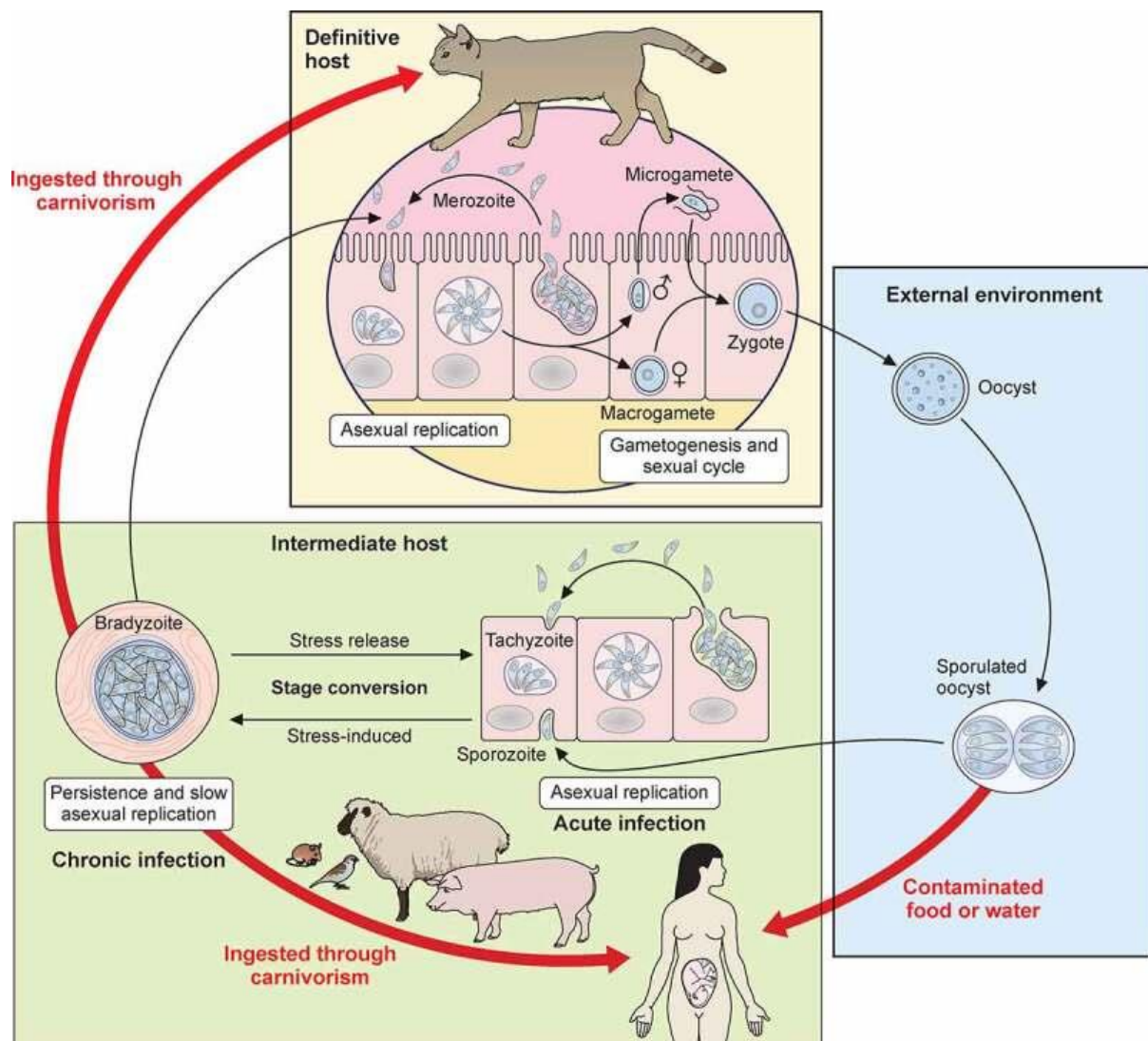


FIGURE 3 : CYCLE PARASITAIRE DE *TOXOPLASMA GONDII*

Représentation schématique du cycle parasitaire et des modes de transmission de *Toxoplasma gondii*. Les animaux représentés indiquent les hôtes. Le fœtus fait référence aux infections transplacentaires. (D'après Syrian G. Sanchez et Sébastien Besteiro, 2021 (43)).

I.B. LES *TOXOPLASMA*

Le genre *Toxoplasma* appartient à l'ordre des Eucoccidés et à la famille des Eimeriidés. La seule espèce connue de ce genre est *Toxoplasma gondii* qui infecte son hôte par voie féco-orale et possède un cycle monoxène chez son hôte définitif le chat (cycle complet avec reproduction asexuée et sexuée). Il existe tout-de-même des hôtes intermédiaires, servant de réservoirs environnementaux chez qui le parasite n'effectue qu'une reproduction asexuée. Ses hôtes intermédiaires peuvent appartenir à de nombreuses espèces d'oiseaux et de mammifères, l'Homme compris (**cf. figure 3**) (25). *T.gondii* est retrouvé à l'échelle mondiale sur tous les continents et infecterait près du tiers de la population mondiale (44).

L'hôte définitif, le chat, excrète des oocystes dans ses fèces. Ces oocystes encore non infectants le deviennent après sporulation et seront une source de contamination pour les autres hôtes par ingestion. Une transmission d'espèce en espèce le long de la chaîne alimentaire peut ainsi s'effectuer.

Chez l'hôte intermédiaire, les oocystes sont lysés libérant des sporozoïtes (**cf. figure 3**) (25,45,46). Les sporozoïtes de *T.gondii* envahiront la muqueuse intestinale et pourront infecter les macrophages s'y trouvant. Les parasites se transformeront en tachyzoïtes dans les macrophages (forme de répllication asexuée rapide), et entraîneront la dégénérescence progressive des cellules hôtes. Les tachyzoïtes seront ensuite libérés des cellules et se dissémineront dans l'hôte par les voies lymphatiques et sanguines.

T.gondii aura tendance à s'enkyster au niveau des muscles striés squelettiques et du cerveau sous forme de bradyzoïtes (forme à répllication lente). Enkysté, le parasite sera à l'abri du système immunitaire de l'hôte. Les félins s'infesteront alors par carnivorisme permettant un nouveau cycle sexué du parasite.

L'Homme se contamine principalement par consommation de légumes souillés ou par carnivorisme (poules, cochons, lapins, moutons... infestés) (**cf. figure 3**) (25,45,46). Chez l'homme les kystes pourront être réactivés suite à divers stimuli ou lors d'une diminution



des défenses immunitaires (SIDA, greffes d'organes...) (25,45). *T.gondii* représente un risque particulier chez la femme enceinte pour le fœtus en raison d'infections transplacentaires entraînant des toxoplasmoses congénitales ((45–47) qui peuvent induire des cécités et des retards mentaux chez l'enfant à naître. Chez les sujets immunodéprimés la toxoplasmose peut également s'accompagner d'adénopathies (état pathologique des nœuds lymphatiques) et d'atteintes du système nerveux central (48). La prise en charge des infections congénitales et des patients immunodéprimés implique souvent un traitement par pyriméthamine, un anti-protozoaire généraliste qui agit en bloquant la dihydrofolate réductase, inhibant ainsi la biosynthèse des bases puriques et pyrimidiques, et empêchant le parasite de synthétiser correctement son ADN (47,49). La toxoplasmose est principalement asymptomatique chez l'individu immunocompétent et ne nécessite généralement pas de traitement.

Le génome de *T. gondii* est constitué d'environ 63 à 69 Mb d'ADN nucléaire repartis en 14 chromosomes avec un pourcentage GC de l'ordre de 52%. On retrouve également un ADN plastidique de 35 Kb et un ADN mitochondrial de 5.9 Kb (50–52). Le génome nucléaire coderait pour 7860 à 8460 protéines en fonction des souches, avec en moyenne 5.9 exons par gène et environ 598 ARN non-codants (50,53). En Europe et en Amérique du Nord la population des *T.gondii* est constituée de trois souches principales numérotées de I à III et définies en fonction de leurs génotypes (54). Les souches I, II et III de *T. gondii* présentent des degrés de virulence différents et des prévalences variables (54). Les populations sont principalement clonales du fait de la domestication des chats et des efforts sanitaires de lutte contre la toxoplasmose qui ont conduit à une baisse de la transmission du parasite depuis le chat (54). La reproduction sexuée du parasite et le brassage génétique associé a maintenant principalement lieu dans les souches retrouvées chez des animaux non domestiqués (54–56).

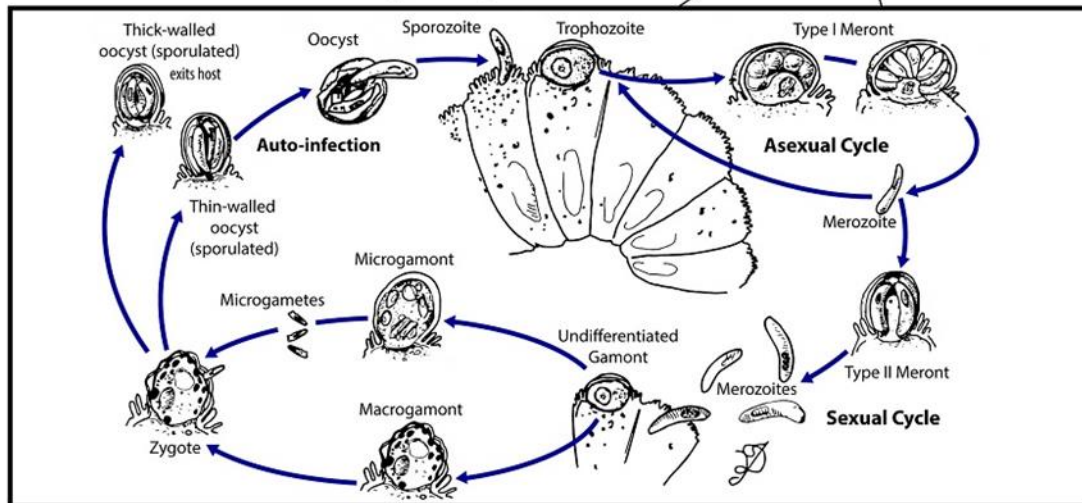
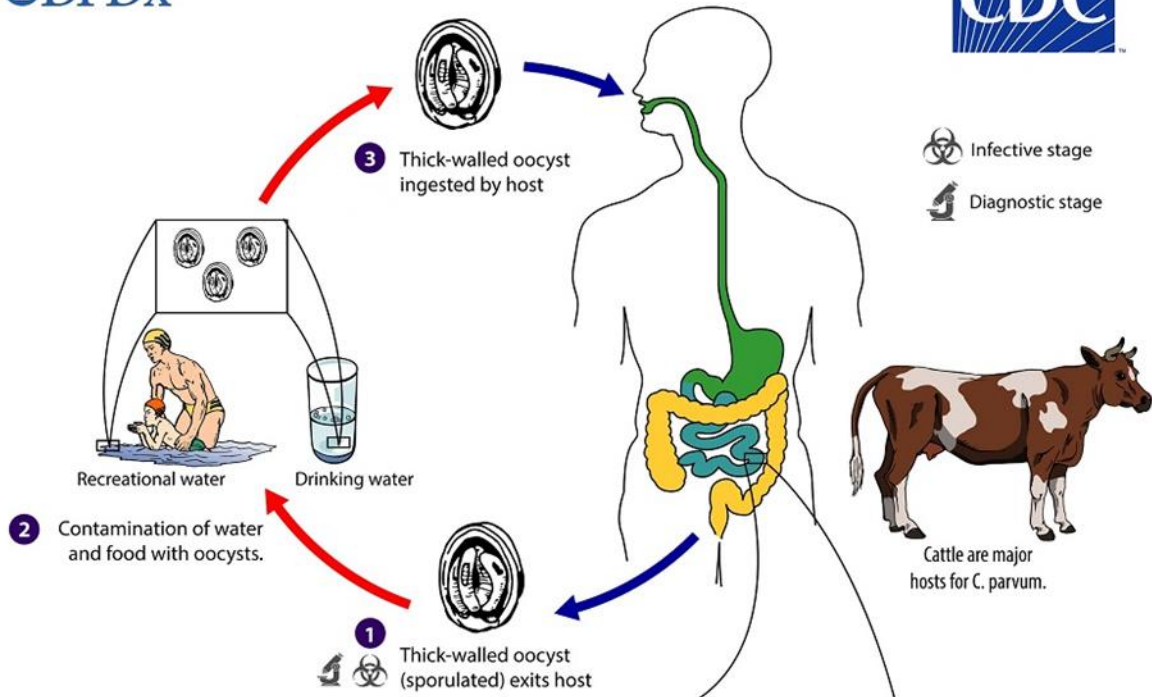


FIGURE 4 : CYCLE PARASITAIRE DES *CRYPTOSPORIDIUM*

Cycle parasitaire de *Cryptosporidium* spp. Les flèches rouges représentent les étapes ayant lieu dans l'environnement et les bleues celles se déroulant chez l'Homme. Les numéros indiquent l'enchaînement des étapes. L'encadré détaille le développement du parasite dans l'intestin moyen chez l'homme. (Adapté de Centers for disease Control and Prevention américain (CDC) (57)).

I.C. LES *CRYPTOSPORIDIUM*

Les *Cryptosporidium* sont de l'ordre des Eucoccididés, comme *T. gondii*, mais appartiennent à la famille des Cryptosporidiidés. Ce sont des parasites monoxènes, retrouvés principalement chez l'Homme, les veaux, les agneaux et les chevreaux (25). Les porcs et les chats peuvent être porteurs de certaines espèces, ainsi que de nombreux autres mammifères, des oiseaux, des reptiles et des poissons (58,59).

La contamination se fait par voie féco-orale et le cycle a lieu principalement au niveau de la muqueuse duodéno-jéjunale. Le parasite est intracellulaire mais extra-cytoplasmique (**cf. figure 4**) (25,60). Les *Cryptosporidium* sont ingérés sous forme d'oocystes matures contenant 4 sporozoïtes. Ces sporozoïtes sont libérés et adhèrent à la membrane des entérocytes (un type cellulaire formant l'épithélium intestinal et caractérisé par des microvillosités sur leur pôle apicale (61)). Il se forme ensuite une héli-vacuole parasitophore à la surface de l'entérocyte. La membrane du parasite et celle de sa vacuole fusionneront, laissant le parasite en contact direct avec le cytoplasme de l'entérocyte. S'ensuit une phase de prolifération (schizogonie) suivie de la libération de 8 mérozoïtes qui pourront à leur tour parasiter d'autres entérocytes (25,60).

Lorsque les mérozoïtes libérés parasitent de nouveaux entérocytes, une phase de réplication sexuée peut avoir lieu. Dans ce cas, il y a gamétogénèse et 8 à 16 gamètes mâles sans flagelle sont formées ainsi qu'un gamète femelle. Après fécondation a lieu une sporogénèse conduisant à deux types d'oocystes contenant chacun 4 sporozoïtes infestants : il y aura des oocystes à paroi épaisse (80%) (libérés dans les selles), et des oocystes à paroi fine (20%) (libérant leur contenu directement dans l'intestin et provoquant un cycle d'endo-infestation (25,60)).

Chez l'homme immunocompétent, les cryptosporidioses se manifestent principalement par des diarrhées qui se résolvent spontanément, l'atteinte perdure cependant chez les immunodéprimés. Au niveau vétérinaire, Les cryptosporidioses entraînent des pertes économiques lorsqu'elles touchent les veaux (25,60). Les antiparasitaires classiques montrent en générale peu d'efficacité (62,63). Seul le nitazoxanide a prouvé une efficacité clinique aux

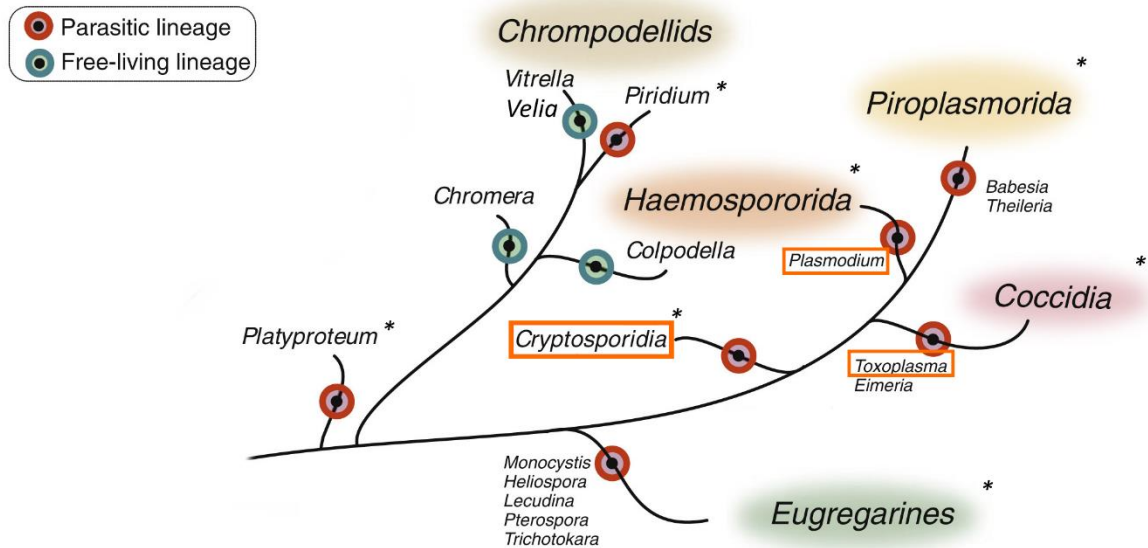


FIGURE 5 : REPRESENTATION SCHEMATIQUE DES RELATION PHYLOGENETIQUES ENTRE EUCARYOTES AVEC EMPHASE SUR LES APICOMPLEXES ET LES CHROMPODELLIDS

Représentation schématique de la phylogénie des apicomplexés et des chrompodellides. Les grandes familles sont entourées d'un halo coloré. *Plasmodium*, *Cryptosporidia* et *Toxoplasma* sont encadrés pour plus de visibilité. Les lignées parasitaires sont indiquées en rouge et les non-parasites en vert. Un astérisque « * » est ajouté à côté des familles ou groupes considérés comme des Apicomplexés. (Adaptée de Sonja Rueckert, Shweta V. Pipaliya et Joel B. Dacks 2019 (64)).

cryptosporidioses en induisant des lésions des membranes cellulaires ainsi que des perturbations du métabolisme énergétique mitochondrial (62,63).

Au moins 44 espèces de *Cryptosporidium* et plus de 120 génotypes ont été identifiés. *C. hominis* et *C. parvum* sont les deux espèces les plus retrouvées chez l'Homme (65). Les génomes des cryptosporidies sont parmi les plus petits génomes d'apicomplexés séquencés, avec environ 9.1 Mb répartis sur 8 chromosomes, un pourcentage GC d'environ 30%, et entre 5 et 20 % de gènes codants possédant des introns. L'annotation des génomes a permis de mettre en évidence entre 3900 et 3994 gènes dont 3805 codant des protéines (données basées principalement sur les espèces *C. hominis* et *C. parvum*) (66,67).

II. GENOMIQUE COMPARATIVE DES APICOMPLEXES

II.A ADAPTATIONS AU PARASITISME

L'étude de la parenté des apicomplexés révèle au moins trois apparitions distinctes du parasitisme dans ce groupe, deux aboutissant aux genres *Platyproteum* et *Piridium*, et une liant l'ensemble des autres apicomplexés dont les *Plasmodium*, *Toxoplasma*, et *Cryptosporidium* (cf. **figure 5**) (64,68,69). Les chromérides sont le groupe connu comme le plus proche phylogénétiquement des apicomplexés sur la base des données génomiques et moléculaires mais sont une famille d'algues photosynthétiques non parasitaires (64,68). Woo et collaborateurs (70) ont effectué une comparaison systématique du contenu en gène des apicomplexés et chromérides afin de caractériser les gènes perdus entre les deux lignées et mieux comprendre le passage d'un état libre à un état parasite au cours de l'évolution du vivant (70,71). Chez les apicomplexés, cette adaptation se serait faite principalement par une perte de gènes (68–70,72). L'étude par Woo et collaborateurs a rapporté une perte de gènes associée à la photosynthèse et à la synthèse des stérols. Les gènes associés au trafic endomembranaire étaient majoritairement tous conservés. La plupart des gènes associés au flagelle et à sa fonction étaient également retrouvés malgré l'absence de flagelle dans l'ensemble du groupe des apicomplexés et une perte progressive des gènes associés à mesure des spéciations chez ces parasites (70). La question demeure concernant les genres *Platyproteum* et *Piridium* qui n'étaient pas inclus dans l'étude.



Woo et collaborateurs ont également pu montrer en croisant des données de transcriptomique de *P. falciparum* et de la chroméride *Chromera velia*, cultivés tous deux *in vitro* sous diverses conditions de stress, qu'il y'avait une conservation de certains modules d'expression génique : ainsi, 80 gènes annotés comme impliqués dans l'invasion chez *P. falciparum* avaient vu leurs orthologues chez *C. velia* exprimés dans les conditions de culture. Ceci suggérerait une réorientation fonctionnelle des gènes entre ces deux espèces phylogénétiquement proches, *C. velia* n'étant pas parasite et n'entreprenant pas de processus d'invasion comme le ferait *P. falciparum* (68,70).

Une étude de la conservation de synténie entre le génome des plastides de *Piridium* (apicomplexés) et de *Vitrella* (chroméride) révèle une conservation quasi complète. Cela suggère que la perte de la photosynthèse a eu lieu tôt dans l'évolution du genre *Piridium* vers le parasitisme (64).

Toutes ces études dessinent un processus d'adaptation au parasitisme complexe chez les apicomplexés, médié par des pertes de gènes et des réorientations fonctionnelles de certains autres, tout ceci de manière parfois parallèle entre les lignées (64).

II.B. ADAPTATIONS A LA DIVERSITE D'HOTES ET AUX MODES DE VIE

Au sein du groupe des apicomplexés, la diversification semble s'être principalement effectuée par une acquisition de nouveaux gènes plutôt que par une perte, à l'exception notable des *Cryptosporidies* où la tendance opposée est observée (73–75). Chez les *Cryptosporidies* il semblerait que la réduction en taille du génome soit associée à la capacité du parasite à récupérer une grande diversité de nutriments dans la niche biologique constituée par ses hôtes (73).

La taille du génome des apicomplexés varie de 8.1 Mb chez le piroplasma *Babesia bovis* à 130 Mb chez la coccidie *Sarcocystis neurona*. Le nombre total de gènes peut également passer du simple au double entre les genres, allant de 3781 chez *B. bovis* à 8920 gènes identifiés pour *T. gondii* (73). Au sein de cette diversité génique, environ 1000 gènes constitueraient le corps commun des apicomplexés et ils ne représenteraient que 12 à 34% des gènes codants selon les espèces (74).



J. Wasmuth et collaborateurs ont effectué une étude portant sur les gènes des apicomplexés. Cette étude rapporte que de nombreux gènes spécifiques de genres sont associés à l'invasion de cellules hôtes ou à la survie parasitaire, tandis que ceux plus largement partagés entre genres sont principalement associés à des fonctions de ménage tels que la traduction et la transcription (75). Ceci témoignerait d'une évolution fortement orientée vers l'adaptation aux hôtes et aux différentes interactions avec ceux-ci.

Le groupe des « Grégarines » au sein des apicomplexés est composé principalement de parasites monoxènes extracellulaires d'invertébrés et d'urochordés (64,76). On compte parmi leur sein des parasites comme *Monocystis*, parasites de vers de terre (77). Ils sont considérés comme les apicomplexés les plus ancestraux (78) et constituent une opportunité intéressante d'étudier des génomes d'apicomplexés extracellulaire. Certaines grégarines ont été classées comme commensales ou mutualistes, remettant une fois de plus en cause la définition globale d'un apicomplexé (64). Les genres *Cryptosporidium*, *Platyproteum*, et *Piridium* sont particulièrement proches des grégarines, jusqu'à parfois avoir été classés comme tels (64).

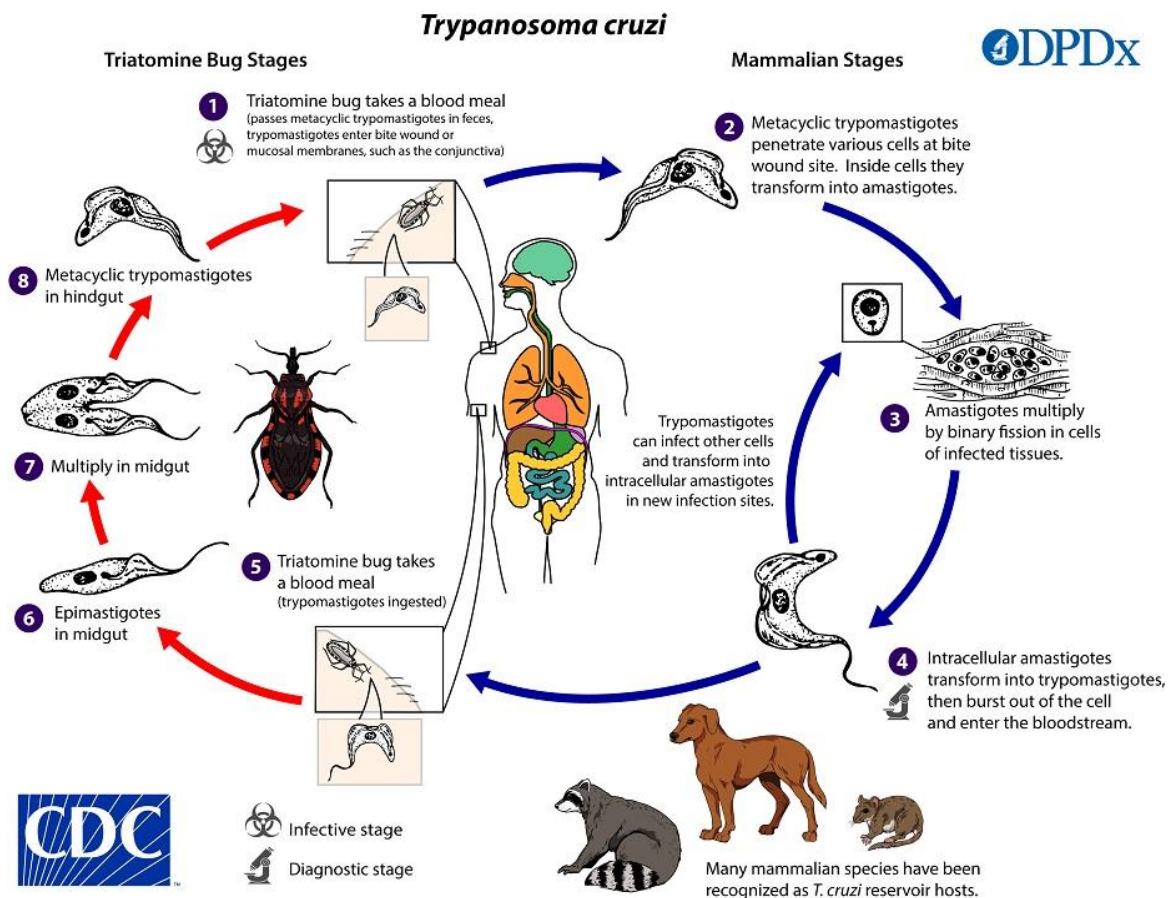


FIGURE 6 : EXEMPLE DU CYCLE PARASITAIRE DE *TRYPANOSOMA CRUZI*

Représentation schématique du cycle parasitaire de *Trypanosoma cruzi* responsable de la maladie de Chagas. Les flèches rouges représentent les étapes ayant lieu chez la punaise réduve et les bleues celles de déroulant chez l'Homme. Les numéros indiquent les différentes étapes et stades du parasite. (D'après *Centers for disease Control and Prevention* américain (CDC) (79))

CHAPITRE 2 : LES KINETOPLASTIDES

Les kinétoplastidés sont une classe d'eucaryotes unicellulaires flagellés. Ils sont caractérisés par la présence d'un kinétoplaste, une structure d'ADN mitochondrial concaténé, localisée au sein de leur large et unique mitochondrie, et retrouvée à la base du flagelle. Cette structure semble liée au cytosquelette et jouerait un rôle dans l'adaptation de ces organismes à leurs milieux (80). L'ADN du kinétoplaste représenterait près de 30% de l'ensemble de l'ADN des cellules (81). Les parasites de cette classe sont responsables de maladies souvent graves chez l'Homme (80).

I. CYCLES PARASITAIRES ET PATHOGENIE

I.A. LES *TRYPANOSOMA*

Les *Trypanosoma* sont des kinétoplastidés monoflagellés de la famille des trypanosomatidés. Ils sont parasites de mammifères dont l'Homme et transmis par des insectes piqueurs et hématophages, comme la mouche tsé-tsé ou la punaise réduve. Sept stades biologiques sont décrits, chacun caractérisé par une morphologie différente, adaptée à l'environnement du parasite. Chez l'hôte vertébré les *Trypanosoma* gardent principalement les formes dites trypomastigotes (forme d'infection extracellulaire) et amastigotes (intracellulaire) (25,80,82).

L'infection de l'Homme par *Trypanosoma cruzi* commence par l'entrée de formes dites trypomastigote dans le corps de l'hôte. Le parasite libéré par les fèces d'une punaise réduve va pénétrer le corps par l'entaille laissée suite à la morsure de la punaise (79,83). Le trypomastigote sera phagocyté par les macrophages de l'hôte et les infectera ainsi. Dans la cellule le parasite se différenciera en une forme dite amastigote qui sera initialement dans une vacuole parasitophore mais s'en échappera rapidement pour se multiplier au contact du cytoplasme de la cellule hôte (79,83). Ces formes amastigotes se retransformeront en trypomastigotes et sortiront de la cellule par lyse. Les trypomastigotes libérés pourront soit réinfecter d'autres cellules soit infecter une nouvelle punaise réduve à l'occasion d'une morsure (cf. figure 6) (79,83).

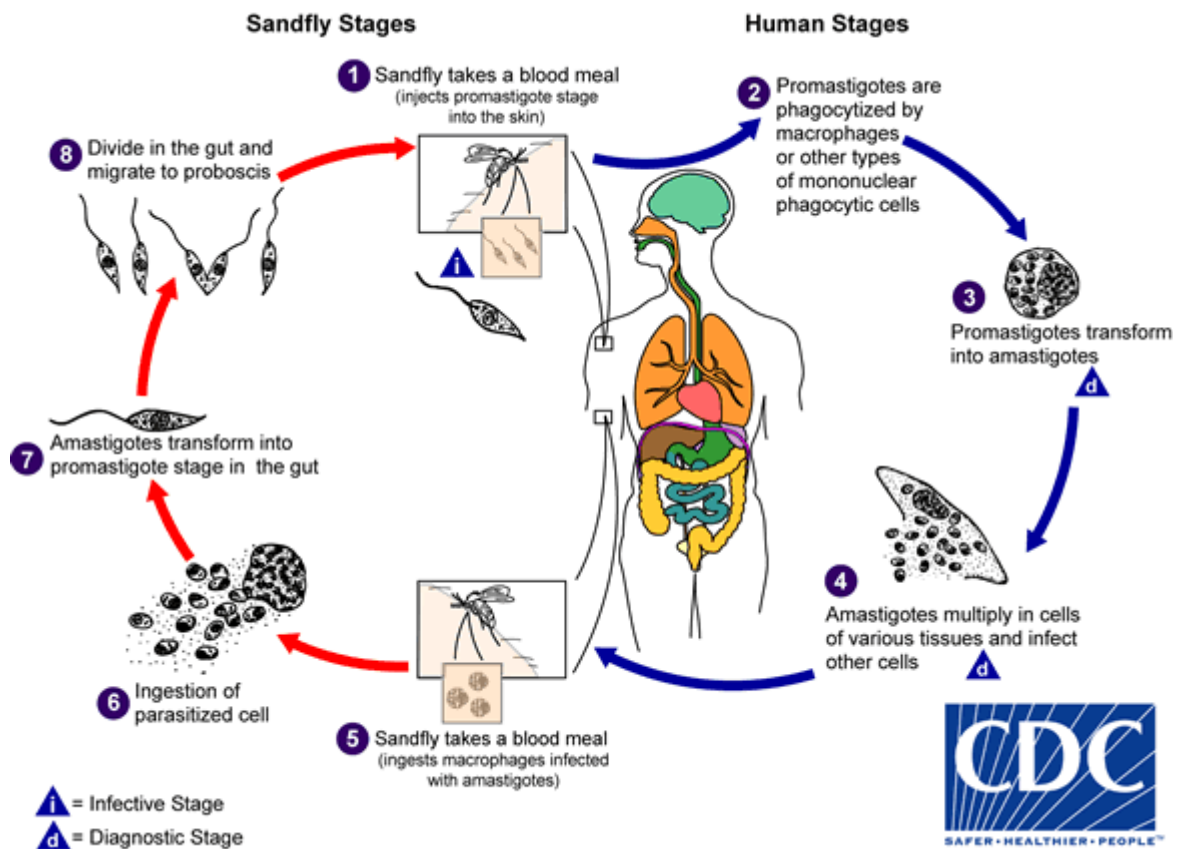


FIGURE 7 : CYCLE PARASITAIRE DES LEISHMANIA

Représentation schématique du cycle parasitaire des Leishmanies. Les flèches rouges représentent les étapes ayant lieu chez le phlébotome ou le *lutzomyia* et les bleues celles de déroulant chez l'Homme. Les numéros indiquent les différentes étapes et stades de l'infection par le parasite. (D'après Centers for disease Control and Prevention américain (CDC) (84))

Les *Trypanosoma* sont responsables de maladies graves chez l'Homme dont la maladie de Chagas (causée par *T. cruzi*, principalement présent en Amérique latine) et la maladie du sommeil (causée par *Trypanosoma brucei*, endémique de certaines régions d'Afrique subsaharienne) (25,80).

Les traitements des trypanosomiasés dépendent grandement du stade de la maladie, de l'agent parasitaire et de la précocité du diagnostic (25,80).

I.B. LES LEISHMANIA

Les *Leishmania* appartiennent aussi à la famille des trypanosomatidés. Ils sont hétéroxènes et se multiplient chez l'insecte sous leur forme mobile (promastigote), dans la lumière intestinale. La transmission aux mammifères dont l'Homme s'effectue lors d'une piqûre par une femelle de la sous-famille des *phlébotomes*, lorsque l'insecte régurgite une partie de son bol alimentaire, ou se fait écraser au contact de la peau. Dans le derme, les *Leishmania* sont alors phagocytés par des cellules phagocytaires mononucléés (monocytes, macrophages, cellules de Langerhans, cellules dendritiques...), et évoluent en forme amastigote (**cf. figure 7**) (25,80).

La Leishmaniose se présente sous différentes formes cliniques. Une forme cutanée, caractérisée par des lésions ulcéreuses et pouvant évoluer vers une forme cutanée diffuse ou cutanéomuqueuse, et une forme viscérale, plus grave, pouvant être accompagnée par une fièvre, une perte de poids, une anémie et une hépatosplénomégalie (25,80).

La prise en charge des Leishmanioses dépend de l'espèce de *Leishmania* et du type d'infection.



II. GENOMIQUE COMPARATIVE DES KINETOPLASTIDES

II.A. GENOMIQUE DES TRYPANOSOMATIDES

T. brucei, *L. major* et *T. cruzi* sont les espèces de Kinétoplastides les plus étudiées (85–87). *T. brucei* est diploïde. La taille de son génome haploïde est de 32 Mb réparties en 3 classes de chromosomes linéaires dont 11 chromosomes d'au moins 1 Mb, 1 à 5 chromosomes de taille dite intermédiaire (entre 200 et 700 Kb), et plus de 100 minichromosomes (de 50 à 150 Kb). *T. brucei* présente également un nombre inconnu d'ADN circulaires extra-chromosomiques contenant des séquences répétées (87–89). L'ensemble de ce génome contiendrait 9068 gènes dont 904 pseudogènes.

Le génome de *L. major* serait lui aneuploïde et présenterait un génome haploïde de 33 Mb cette fois réparti entre 36 chromosomes codant 8311 gènes dont 34 pseudogènes (86,90).

Le génome de *T. cruzi* est très longtemps resté fragmenté et incomplet. La présence plus élevée de séquences répétées en tandem dans le génome rendant son assemblage plus difficile que pour *T. brucei* et *L. major* (20-25% contre 50% pour *T. cruzi*) (85,86,91). Aujourd'hui, le génome quasi-complet de *T. cruzi* est disponible (91,92), et présente une taille de 55 Mb pour 28 chromosomes et environ 12 000 gènes (86). *T. cruzi* serait diploïde comme *T. brucei*, mais présenterait également certains chromosomes tri- ou tétraploïdes (92), avec des variations de ploïdie entre souches (91,93).

Une caractéristique importante des trypanosomatidés est leur capacité à faire varier les déterminants antigéniques de leurs protéines de surface, ce qui leur permet d'échapper au système immunitaire de leurs hôtes (comme ce qui est observé chez les *Plasmodium*). Cette variabilité antigénique est le fait d'un large répertoire de gènes dits VSG (variable surface antigènes, antigènes variables de surface), avec plus de 2500 copies chez *T. brucei* par exemple. Ces gènes sont régulés par divers mécanismes de compartimentalisation nucléaire et de modification épigénétiques aboutissant à l'expression d'une seule copie de ces gènes à la fois (88). Les gènes VSG seraient parfois transcrits sous forme de polycistrons, ensuite maturés en



plusieurs ARNm, dont un seul sera traduit (88). Les gènes VSG sont principalement retrouvés dans des régions sub-téломériques. Chez *T. brucei* ils sont principalement situés sur les 11 chromosomes d'au moins 1 Mb mais seraient également présents sur certains minichromosomes et chromosomes de taille intermédiaires (88).

Les génomes des *Trypanosomatidés* sont riches en séquences répétées (50% du génome de *T. cruzi*, 20-25% pour *T. brucei* et *L. major* (91,94)). Ces répétitions correspondent aux larges familles de VSG, aux répétitions sub-téломériques et aux éléments transposables (ET), uniquement représentés par des rétrotransposons dans ces génomes. Les ET constitueraient 12.6% du génome de *T. cruzi*, 5.7% de celui de *T. brucei* et seulement 1.6% de celui de *L. major* (95).

II.B GENOMIQUE COMPARATIVE DES TRYPANOSOMATIDES

L'étude comparative des génomes de *T. cruzi*, *T. brucei* et *L. major* révèle un corps commun d'environ 6200 gènes. La synténie globale est forte malgré des distances évolutives relativement importantes entre ces espèces. En effet, 68% des orthologues chez *T. brucei* et 75% chez *L. major* se trouveraient dans le même environnement génomique. Ces gènes feraient partie à 94% du corps commun à *T. cruzi*, *T. brucei* et *L. major* (85,86). Environ 32%, 26% et 12% des gènes seraient spécifiques aux espèces *T. brucei*, *T. cruzi* et *L. major* respectivement (85). La plupart de ces gènes semblent coder pour des protéines antigéniques de surface (86), qui seraient donc directement liées aux interactions avec leurs hôtes.

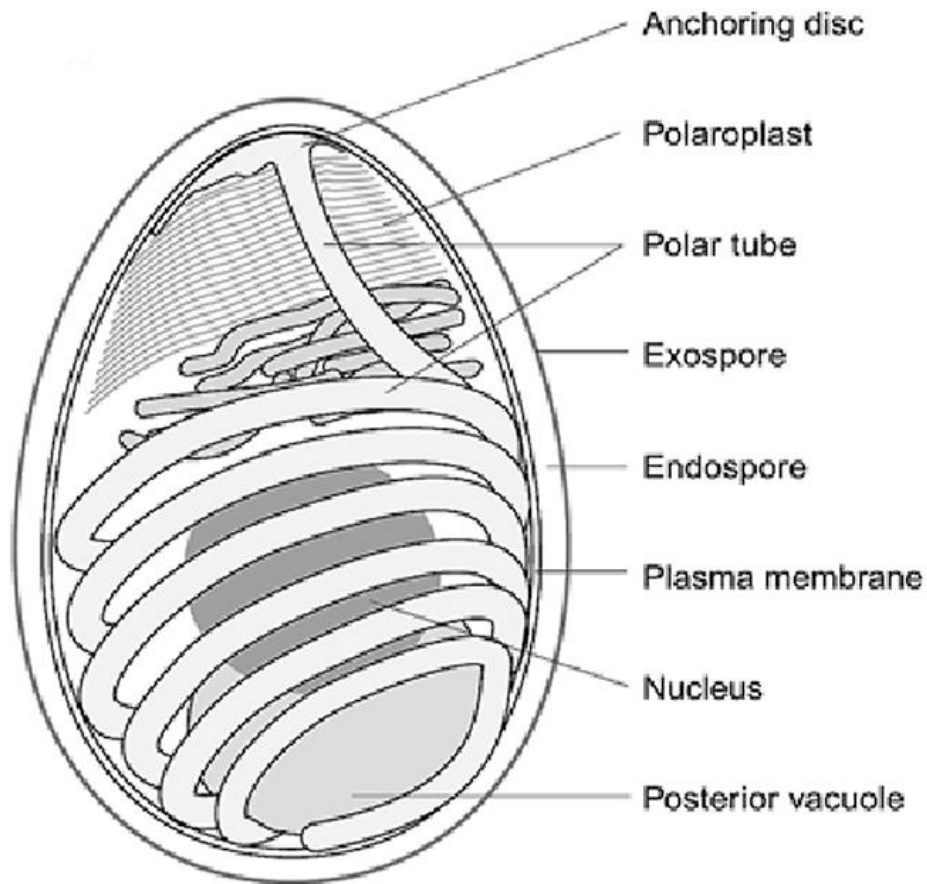


FIGURE 8 : STRUCTURE GENERALE D'UNE SPORE MICROSPORIDIENNE

Représentation schématique d'une spore de microsporidie dont la paroi est composée de trois couches : l'exospore qui est dense aux électrons ; l'endospore qui est épaisse mais translucide aux électrons ; et la membrane plasmique. Le tube polaire est l'appareil infectieux et est typique des microsporidies. Il est enroulé à l'intérieur de la spore avec un nombre de tours qui est spécifique à chaque espèce. Le tube polaire finit dans un disque d'ancrage en position apicale dans la spore. On retrouve également un noyau, une vacuole postérieure et un polaroplaste dans la spore. (D'après Caspar Franzen 2004 (96)).

CHAPITRE 3 : LES MICROSPORIDIES

I. LES MICROSPORIDIES GENERALITES ET BIOLOGIE

I.A GENERALITES

Les microsporidies sont un groupe de parasites intracellulaires obligatoires apparentés aux champignons (21). Ils présentent une large distribution géographique et un très large spectre d'hôtes allant des invertébrés aux vertébrés, dont l'Homme (97). On compte actuellement 45 familles, 218 genres, pour plus de 1600 espèces microsporidiennes (97). Les caractéristiques communes du groupe comprennent l'existence d'un stade sporal (stade de résistance et de dissémination) et un appareil infectieux unique appelé tube polaire (**cf. figure 8**) (21,98). D'autres caractéristiques incluent : un appareil de Golgi non-empilé, des ribosomes de petite taille (99), et une absence de peroxyosomes (21,98). Un reliquat de mitochondrie appelé mitosome existe chez les microsporidies. Les mitosomes sont dépourvus d'ADN et de la majeure partie des protéines constituant les mitochondries (100,101). Cependant, de nombreux gènes d'origine mitochondriale ont pu être identifiés dans le génome nucléaires des microsporidies (100–102).

I.B BIOLOGIE INFECTIEUSE DES MICROSPORIDIES

Les microsporidies sont retrouvées dans l'environnement sous forme de spores d'une taille comprise entre 1 et 12 μm selon les espèces (103). Ces spores infectent leurs hôtes par voie féco-orale ou par contact direct avec une surface lésée ou permissive. Le processus d'invasion des microsporidies fait intervenir le tube polaire et s'effectue rapidement, en moins de deux secondes *in vitro* (104). Une fois dans l'organisme hôte, la spore va dévagner son tube polaire. Le contenu cellulaire ou sporoplasme qui représente l'agent infectieux transitera à travers ce tube creux et se retrouvera libéré dans le cytoplasme d'une cellule hôte. Les mécanismes d'entrées exacts demeurent discutés dans la littérature (103). Cependant, des molécules de surface de la spore, du tube polaire ainsi que des récepteurs de la cellule hôte semblent impliqués dans le processus (103,105).

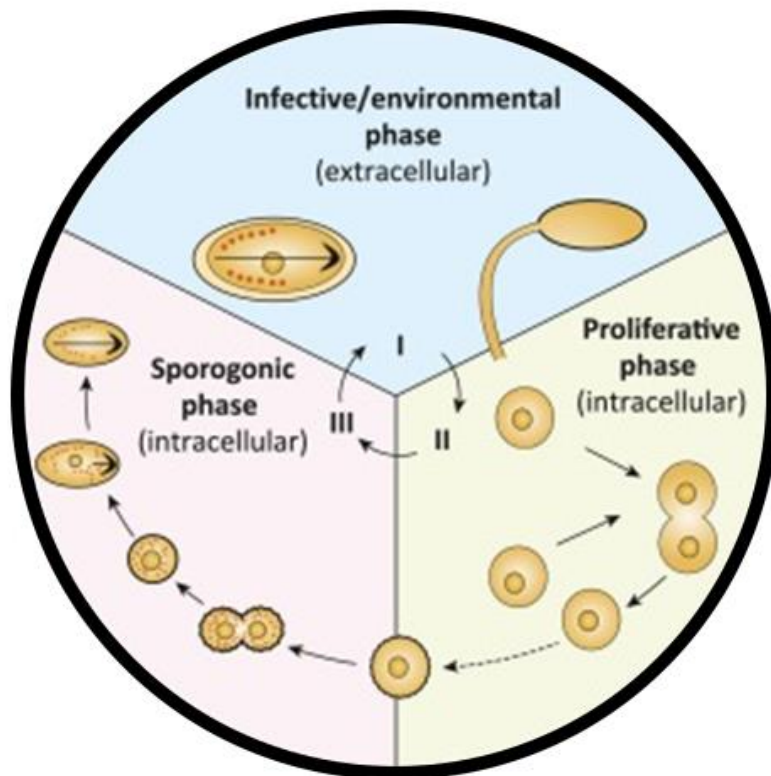


FIGURE 9 : CYCLE PARASITAIRE DES MICROSPORIDIES

Représentation schématique du cycle de développement intracellulaire des microsporidies. I) phase extracellulaire/infectieuse ; II phase proliférative ou mérogonie ; et III phase de différenciation sporogonique la sporogonie. (D'après Weiss et Takvorian 2021 (106)).

Une fois dans la cellule hôte, le parasite entrera dans une phase de prolifération active (mérogonie), suivi d'une phase de différenciation en spore (sporogonie) qui aboutira à la formation de nouvelles spores matures à l'intérieure de la cellule hôte (cf. **figure 9**) (21,105). Les spores seront ensuite libérées par lyse cellulaire ou excrétées par la cellule. La durée du cycle, la localisation subcellulaire ainsi que la morphologie des différents stades dépendent de l'espèce microsporidienne considérée (21,105).

La plupart des infections microsporidiennes ont lieu dans le tractus digestif, d'où elles peuvent ensuite disséminer dans l'organisme, sous certaines conditions, après lyse des premières cellules intestinales infectées. Différentes espèces microsporidiennes présentent un tropisme particulier pour différents tissus (*e.g* cornée, muscles striés squelettiques, tissu adipeux, encéphale). Les cellules phagocytaires seraient également impliquées dans le processus de dissémination en tant que transporteurs infectés ou « chevaux de Troie » (97,107,108).

I.C. LES MICROSPORIDIES ECOLOGIE ET DISTRIBUTION

Les microsporidies sont largement répandues dans les milieux aquatiques et les réseaux alimentaires (97,109,110). Etant capables d'infecter une grande diversité d'hôtes elles arrivent à se développer et à survivre dans de nombreux environnements. Leurs rôles et leurs impacts écologiques sont encore mal caractérisés malgré le nombre croissant d'études s'intéressant à leur présence et leur diversité dans l'environnement (97,109,110). Les microsporidioses représentent un enjeu économique en aquaculture, crevetticulture, ostréiculture ainsi qu'en apiculture (109).

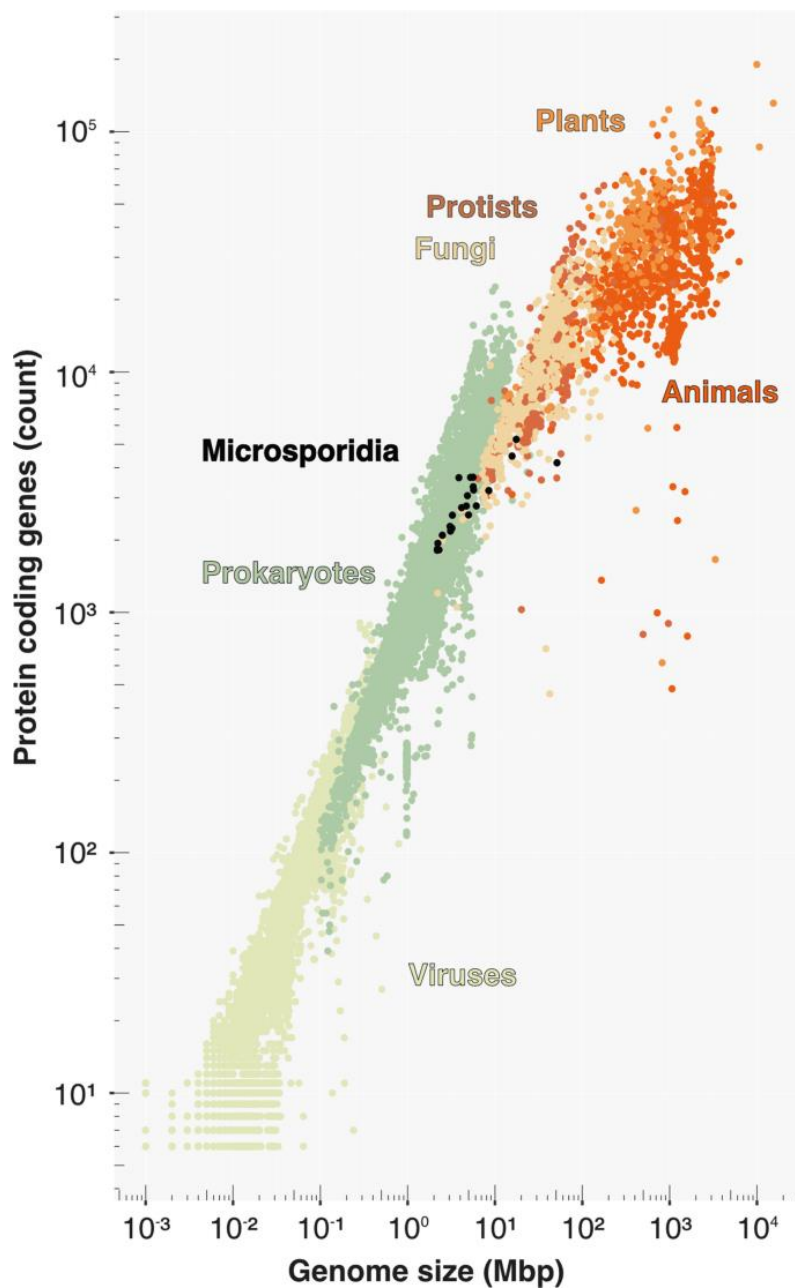


FIGURE 10 : REPRESENTATION DES TAILLES DE GENOMES ET DU NOMBRE DE GENES CODANT DES PROTEINES DANS LE VIVANT

Les microsporidies sont parmi les espèces eucaryotes avec les plus petits génomes. Toutes les entrées présentes sur les bases de données génomiques de NCBI au moment de la construction du graphique ont été incluses, tout en s'affranchissant des séquences partielles ainsi que des séquences aux annotations peu fiables. (D'après Jaspersen *et al.* 2022 (111)).

I.D. LES MICROSPORIDIES EN SANTE ET SANTE PUBLIQUE

Six genres de microsporidies sont connus comme infectant l'homme : *Enterocytozoon*, *Encephalitozoon*, *Anncaliia*, *Vittaforma*, *Trachipleistophora*, et *Pleistophora* (112).

Chez les individus immunocompétents la plupart des infections se limitent à des diarrhées, si elles ne sont pas complètement asymptomatiques (112). Les infections se résolvent en général spontanément après quelques jours, bien que quelques études fassent état d'infections persistantes et/ou latentes (113,114). Ces infections représenteraient un facteur de risques chez les patients atteints de cancers du côlon (114).

Chez les sujets immunodéprimés, les microsporidioses peuvent engager le pronostic vital des patients (112). Ces individus peuvent souffrir d'entéropathies, de myosites, de pneumopathies, d'urétrites, ou même de septicémies disséminées suite à l'infection (112). La pandémie de SIDA des années 80 a fait des microsporidioses un enjeu de santé publique (112). On estime le taux de prévalence mondiale des microsporidies chez l'Homme à environ 10.2% (115). Les principaux traitements contre les microsporidioses sont l'albendazole et la fumagilline, respectivement un inhibiteur de la polymérisation de la tubuline des nématodes, cestodes et de certains protozoaires, et un inhibiteur de la synthèse d'ARN des microsporidies (116,117).

II. LES MICROSPORIDIES : GENOMIQUE, POST-GENOMIQUE ET EVOLUTION

II.A. CARACTERISTIQUES DES GENOMES ET GENES MICROSPORIDIENS

Les microsporidies présentent parmi les plus petits génomes eucaryotes, avec l'espèce *Encephalitozoon intestinalis* qui possède un génome de seulement 2.3 Mb (cf. **figure 10**). Cette réduction génomique s'accompagne d'un nombre de gènes codants réduits (1833 protéines prédites chez *Encephalitozoon intestinalis*), de gènes d'une taille plus courte par rapport à leurs orthologues chez d'autres eucaryotes (ARNr inclus), et de régions intergéniques extrêmement réduites (parfois quelques nucléotides) (118,119). L'adaptation des microsporidies au parasitisme et à leurs hôtes serait passée par une réduction générale

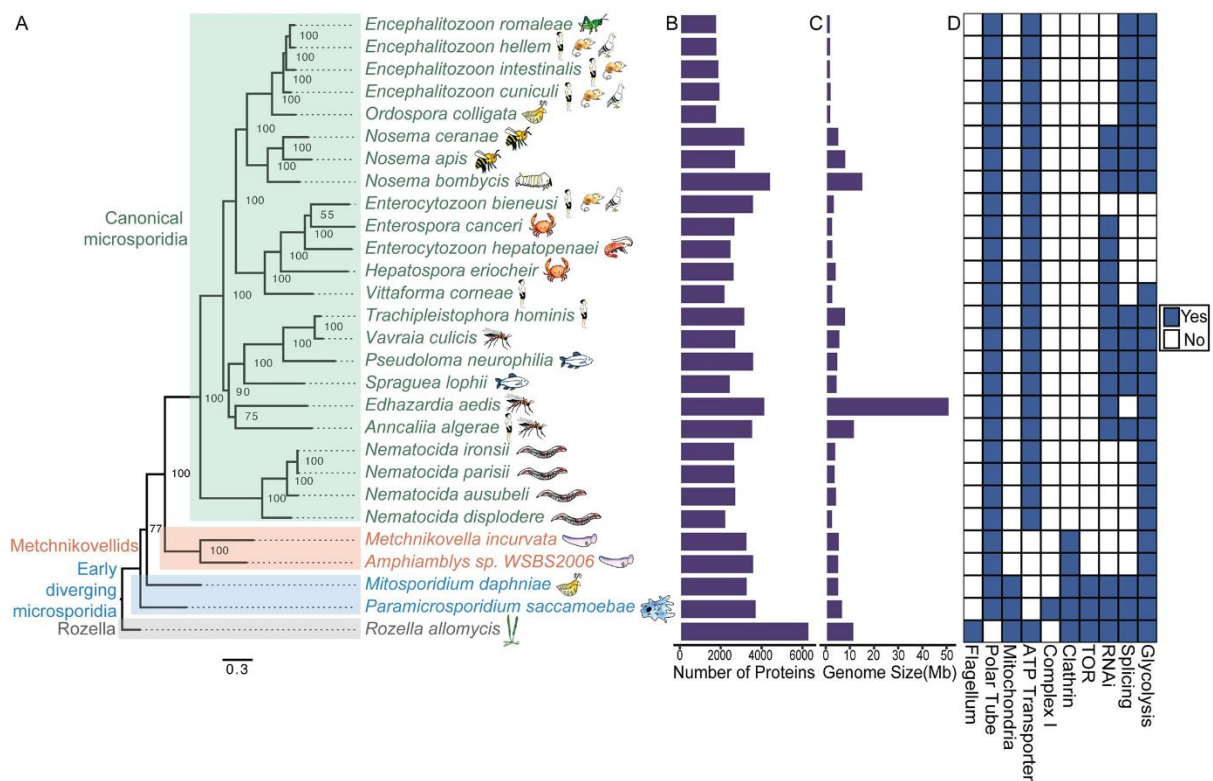


FIGURE 11 : PHYLOGENIE, HOTES ET CARACTERISTIQUES DES MICROSPORIDIES

(A) Arbre phylogénétique des microsporidies dont le génome a été séquencé. L'arbre est basé sur les protéines orthologues dont les gènes sont présents et en copie unique chez chacune des espèces (120,121). *Rozella allomyces* est utilisée comme racine en tant que groupe le plus proche phylogénétiquement. (B) Comparaison du nombre de gènes codant des protéines entre les différentes espèces. (C) Comparaison des tailles de génome. (D) Caractéristiques morphologiques et moléculaires. (D'après Wadi et Reinke 2020 (119)).

de leur répertoire de gènes codants, combinée à l'expansion de familles spécifiques de gènes comme les hexokinases, les métalloprotéase Ste24, les protéines chaperonnes Hsp90 et de nombreuses familles de transporteurs (122,123). Trente-deux familles de protéines ont été identifiées comme spécifiques aux microsporidies et seraient associées à la formation de la spore, la formation du tube polaire, mais également à d'autres fonctions encore non caractérisées (119,122,124).

La ploïdie est variable chez les microsporidies. Alors que certaines espèces semblent diploïdes tout le long de leur cycle de vie (e.g. *E. intestinalis*) d'autres espèces comme *Vairimorpha ceranae* (précédemment *Nosema ceranae*) sembleraient alterner entre des phases diploïdes et tétraploïdes (125). Des différences en termes de taille de génome existent également dans le phylum, et tranchent avec le paradigme de la réduction des génomes microsporidiens. L'espèce *Edhazardia aedis* par exemple possède le plus grand génome microsporidien enregistré avec 51.3 Mb (soit plus de 20 fois la taille de celui d'*Encephalitozoon intestinalis* (126,127)). De manière assez intéressante, son génome ne coderait qu'environ 4190 protéines, soit à peine 2.25 fois plus de protéines que *E. intestinalis* (cf. **figure 11**) (126). Cette différence s'expliquerait par un enrichissement en séquences non-codantes chez *E. aedis*, et une plus faible densité de gènes codants pour des protéines à l'échelle de son génome (119,127). Un autre facteur expliquant la disparité de taille entre les génomes microsporidiens serait l'expansion de familles d'ET chez certaines espèces (119,128,129). La présence de ces séquences répétées se manifeste notamment par des pourcentages GC variables dans le phylum, allant de 20% pour *Spraguea lophii* à 48% pour *E. cuniculi* (127,130). Une majorité de ces ET dériverait d'une longue histoire de transferts horizontaux avec des métazoaires (128). Tandis que certaines espèces comme *E. cuniculi* semblent avoir éliminé l'ensemble des ET de leurs génomes au cours de leur évolution, d'autres espèces en présentent une forte abondance et diversité comme *Anncaliia algerae* (118,128).



II.B. POST-GENOMIQUE MICROSPORIDIENNE

Bien que les microsporidies présentent les ARN polymérases classiquement retrouvées chez les eucaryotes, les processus de régulation de la transcription semblent simplifiés (131,132). Ainsi, des signaux de transcription de type « CCC » et « GGG » ont pu être identifiés en amont des gènes chez *E. cuniculi*. Ces signaux semblent être largement répandus chez les microsporidies et ne pas se limiter aux gènes codants des protéines mais aussi à ceux transcrits par l'ARN polymérase III. Un signal complémentaire de la forme « AAATTT » a été identifié encore plus en amont pour les ARNr ainsi que des gènes fortement exprimés. Ce signal pourrait correspondre à la boîte « TATA » classiquement retrouvée chez les eucaryotes (131). Au niveau des transcrits, les régions 5' et 3' UTR sont très réduites chez les microsporidies (131). Cependant, un phénomène de transcription non interrompue (overrun ou read-through transcriptionnel) causé par une reconnaissance stochastique des signaux de polyadénylation peut donner naissance à des transcrits plus longs qu'une séquence codante (CDS) chez certaines espèces comme *E. cuniculi* (131). Dans ces cas, l'ARN polymérase II, ne reconnaissant pas le signal présent en 3' d'un gène, continue sa transcription et donne naissance à des transcrits pouvant contenir plusieurs CDS consécutifs. Ce phénomène semble présent chez de nombreuses espèces microsporidiennes (131,133).

Les introns sont présents en nombre limité chez les microsporidies. Chez *E. cuniculi* par exemple, 16 introns de 23 à 52 nt ont été identifiés, principalement retrouvés en 5' de gènes codant des protéines ribosomales (134). Deux introns sont notables pour leur présence conservée entre orthologues à travers le phylum, il s'agit d'un intron dans un gène codant une polyA binding protein et la protéine ribosomique 40S/S23 (134,135). Ces deux introns présentent également une forte conservation de leurs séquences ce qui est surprenant étant donnée la forte vitesse d'évolution des séquences d'ADN non-codant et des séquences microsporidiennes en général (119,134,135). Ceci suggère une fonction importante de ces introns chez les microsporidies (134,135), potentiellement dans des processus de régulation de l'expression des gènes. Des modifications épigénétiques ont également été identifiées chez les *Encephalitozoon* et serviraient à réguler l'expression des ARNr au niveau des télomères par une hyper méthylation de l'ADN dans ces régions, notamment lors de l'entrée en dormance pendant la maturation des spores (136).

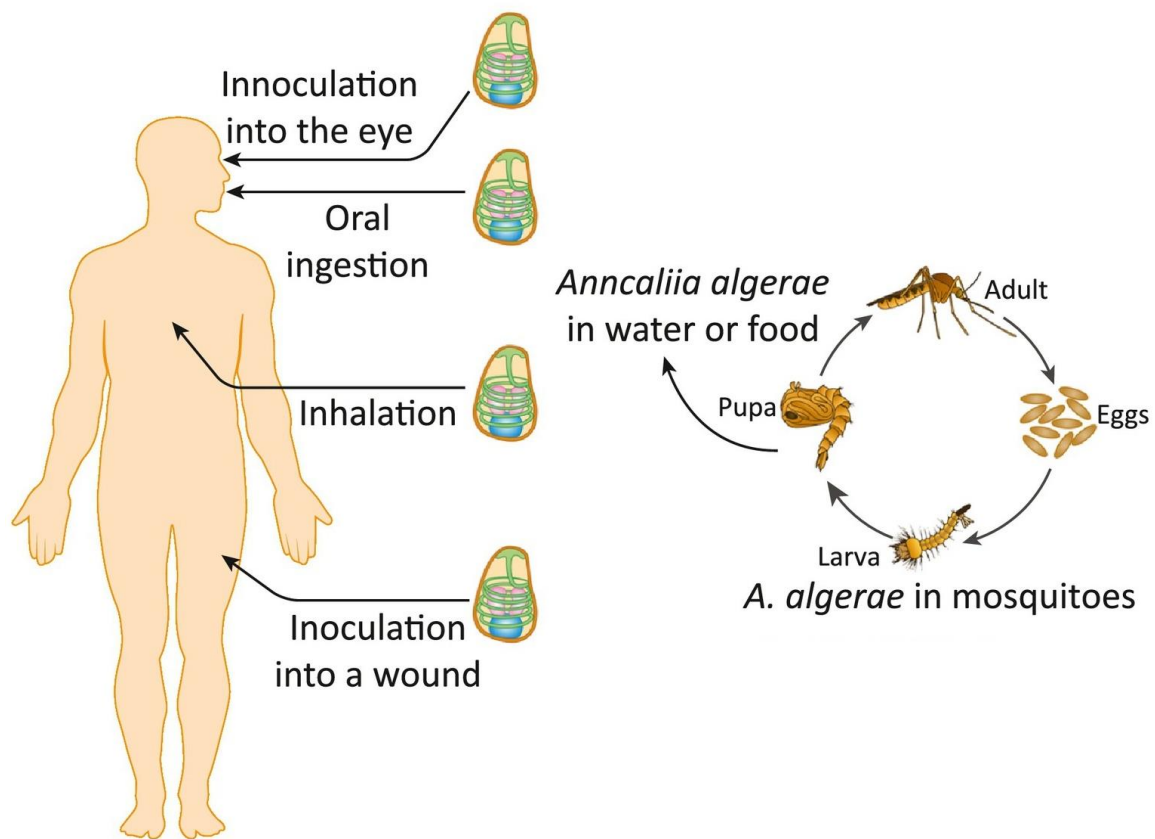


FIGURE 12 : SCHEMA DES VOIES D'INFECTION PAR ANNCALIIA ALGERAE

Représentation schématique des voies d'infection par *A. algerae* chez l'Homme et l'anophèle. *A. algerae* est capable d'infecter et de se développer dans de nombreux tissus ainsi que dans deux espèces très différentes physiologiquement et phylogénétiquement : l'Homme et l'anophèle. (Adapté de Weiss et Takvorian 2021 (106)).

II.C LA MICROSPORIDIE *ANNICALIA ALGERAE*

L'espèce *A. algerae* a un génome d'environ 12 Mb (137). Sa ploïdie était inconnue au début de cette thèse et son nombre de gènes codants était estimé à 2075 (138). Le génome d'*A.algerae* a une faible densité en gènes codants et présente un nombre conséquent de séquences d'ADN non-codant, comme c'est le cas pour le génome de *E.aedis* (127). *A.algerae* est également marquée par l'expansion de nombreux ET dans son génome (plus de 200 familles) (128).

Alors que la plupart des espèces microsporidiennes ont un spectre d'hôte relativement réduit, se spécialisant ainsi dans le parasitisme de quelques espèces ou groupes phylogénétiquement proches (119), *A.algerae* présente la particularité d'être capable d'infecter l'Homme et l'anophèle (cf. **figure 12**), deux espèces phylogénétiquement très éloignées et présentant des caractéristiques physiologiques très différentes (température, tissus, etc). Les espèces proches d'*A.algerae* sont des parasites d'insectes et l'on considère donc classiquement qu'*A.algerae* est un parasite de l'anophèle, qui a connu une expansion de son spectre d'hôte vers l'Homme (119,139). L'infection des deux hôtes se fait de manière indépendante et aucun cas de transmission directe entre anophèles et hommes n'a été décrit (106,139). *In vitro*, *A.algerae* présente également une forte adaptabilité et est capable de se développer dans de nombreux types cellulaires issus de nombreuses espèces (140). L'organisation chromosomique de ce parasite a initialement été étudiée par électrophorèse en champs pulsés (137,141), avant que le génome soit successivement séquencé par des approches Sanger, puis Illumina. Ainsi, les génomes de trois souches d'*A.algerae* sont accessibles dans des bases de données publiques (137). Le premier génome correspond à la souche Undeen, isolée d'une biopsie de muscle d'un patient et séquencé en Sanger. Les deux autres génomes ont été générés par séquençage Illumina et proviennent d'un insecte de la famille des anophèles (souche PRA 339) et d'une biopsie humaine (souche PRA 109) (137). Les trois souches semblent présenter des quantités d'ET différentes avec 1.64% chez PRA109, 1.34% chez PRA339 et 15.75% chez Undeen (129). C'est cette dernière souche qui a été étudiée dans le cadre de ce projet de thèse.

TABLEAU 1 : CLASSIFICATION DES ELEMENTS TRANSPOSABLES EUCARYOTES SELON WICKER ET AL (142)

(D'après Makalowski *et al* 2019 (143)).

Class	Order	Superfamily	Phylogenetic distribution
Class I (retrotransposons)	LTR	Copia	Plants, metazoans, fungi
		Gypsy	Plants, metazoans, fungi
		Bel-Pao	Metazoans
		Retrovirus	Metazoans
		ERV	Metazoans
	DIRS	DIRS	Plants, metazoans, fungi
		Ngaro	Metazoans, fungi
		VIPER	Trypanosomes
	PLE	Penelope	Plants, metazoans, fungi
	LINE	R2	Metazoans
		RTE	Metazoans
		Jockey	Metazoans
		L1	Plants, metazoans, fungi
	SINE	tRNA	Plants, metazoans, fungi
		7SL	Plants, metazoans, fungi
5S		Metazoans	
SVA _a		Primates	
Retrogenes _a		Plants, metazoans, fungi	
Class II (DNA transposons) Subclass 1	TIR	Tc1-Mariner	Plants, metazoans, fungi
		hAT	Plants, metazoans, fungi
		Mutator	Plants, metazoans, fungi
		Merlin	Metazoans
		Transib	Metazoans, fungi
		P	Plants, metazoans
		PiggyBac	Metazoans
		PIF-harbinger	Plants, metazoans, fungi
		CACTA	Plants, metazoans, fungi
	Crypton	Crypton	Fungi
Class II (DNA transposons) Subclass 2	Helitron	Helitron	Plants, metazoans, fungi
	Maverick	Maverick	Metazoans, fungi

CHAPITRE 4 : LES ELEMENTS TRANSPOSABLES

Les éléments transposables (ET) sont des séquences génétiques mobiles capables de se déplacer, se copier et se répandre de manière indépendante dans le génome de leurs hôtes. On les qualifie parfois de « séquences égoïstes » ou de « parasites génomiques ». Ces séquences sont présentes dans la plupart des génomes eucaryotes à quelques exceptions notables près et sont principalement actives dans les cellules de la lignée germinale des hôtes (144). La diversité, la distribution et les rôles que jouent les éléments transposables dans la biologie et l'évolution du vivant sont vastes et nombreux.

I. CLASSIFICATION ET STRUCTURES

Les ET sont regroupés en deux grandes classes en fonction de leur mécanisme de transposition. Ainsi, l'on distingue la classe I des rétrotransposons, qui passent sous forme d'ARN pour leur mobilisation ; et la classe II, les transposons à ADN, effectuant leur transposition uniquement sous forme d'ADN (144). Il existe au sein de ces classes des subdivisions en ordres (ou sous-classes), superfamilles et familles (**cf. tableau 1**) (143).

I.A LA CLASSE I : LES RETROTRANSPOSONS

Les rétrotransposons effectuent leurs transpositions par la production d'un ARNm intermédiaire, exporté dans le cytoplasme, rétro-transcrit en ADNc, et ensuite réacheminé dans le noyau où il pourra s'intégrer dans le génome. Le processus est qualifié de « copier-coller » car le loci initial n'est pas modifié au terme d'une nouvelle insertion dans le génome (**cf. figure 13**) (144).

Trois grandes sous-classes de rétrotransposons sont distinguées : il s'agit des éléments sans LTR (Long Terminal Repeats), des éléments avec LTR, et des éléments mobilisés par une tyrosine recombinase (YR) (144).

Les éléments non-LTR présentent la structure la plus simple. Ils sont généralement constitués de deux ORF (Open Reading Frame, cadres de lecture ouverts) ORF1 et ORF2. ORF2 assure les activités endonucléase et reverse transcriptase. Pour ces éléments, une fois le transcrit

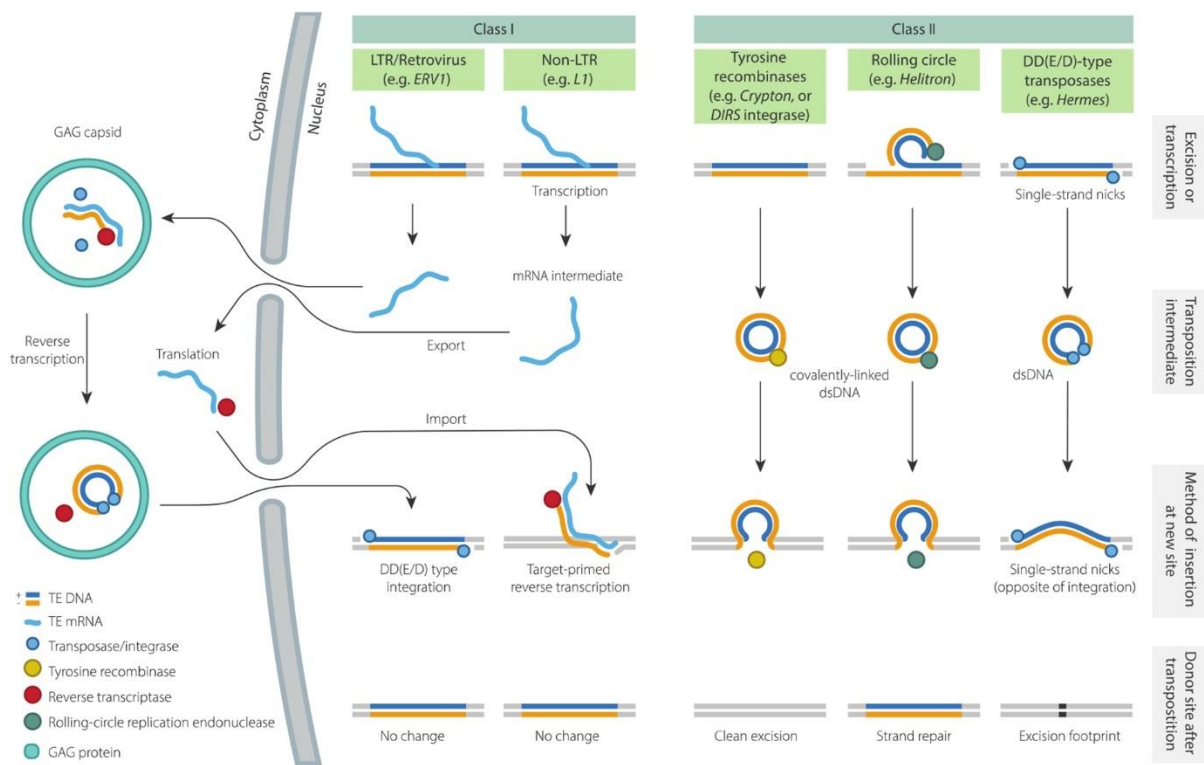


FIGURE 13 : PRINCIPAUX MECANISMES DE TRANSPOSITION

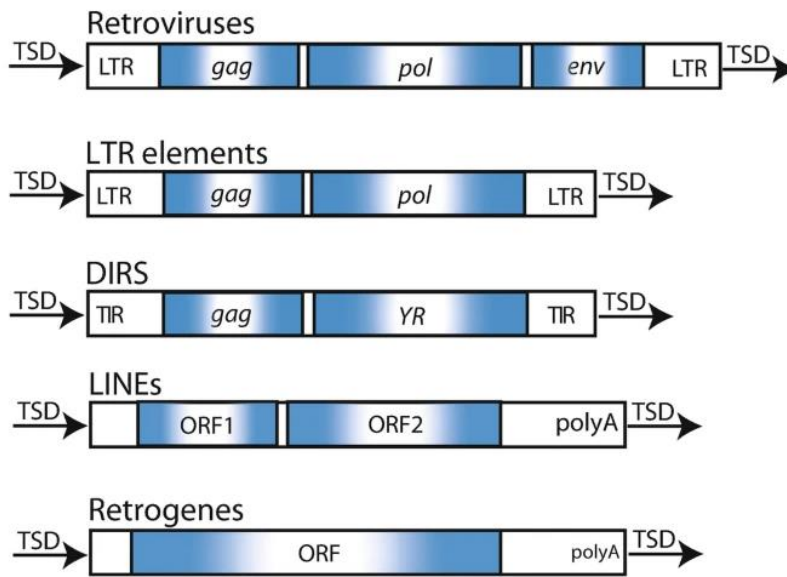
Cinq sous classes d'éléments transposables sont représentées. Dans la classe I les LTR et les non-LTR ; dans la classe II les tyrosines recombinases (YR), les Rolling circles (Helitrons) et les DD(E/D)-type transposases (ou TIR). Les YR-rétrotransposons et les Maverick/Polintons ne sont pas représentés. Les YR-rétrotransposons transposeraient via le même intermédiaire que les Cryptons. Les Maverick transposeraient par synthèse directe d'ADN (145). (D'après Jonathan N. Wells et Cédric Feschotte, 2020 (144)).

primaire exporté du noyau et traduit dans le cytoplasme, l'ARNm est réacheminé dans le noyau encore couplé à la reverse transcriptase. Une fois dans le noyau, la reverse transcriptase reconnaîtra un motif d'ADN (*e.g.* TT/AAAA pour l'élément non-LTR L1), et elle initiera une coupure simple brin. Il y aura ensuite hybridation de l'ARNm avec l'ADN du site reconnu, et la reverse transcription sera effectuée, suivi de l'intégration du cDNA synthétisé dans le génome (**cf. figure 13**) (144). Concernant l'ORF1, sa fonction est encore mal caractérisée. Il est dispensable dans certains éléments non-LTR et est même absent de certains d'entre eux (144,146). Dans d'autres cas cependant il est indispensable, et coderait une protéine impliquée dans la reconnaissance et le transport de l'ARNm vers le noyau (protéine chaperonne) (144,147).

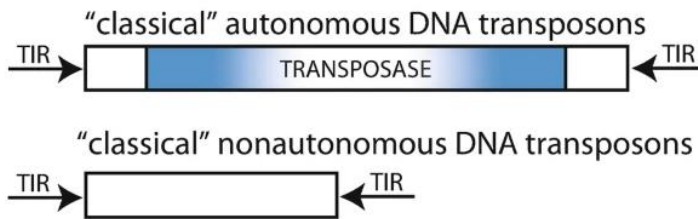
Les LTR dériveraient de rétrovirus endogènes (148). Ils sont constitués de deux gènes distincts : *gag* et *pol*. Ces gènes sont sous contrôle d'un promoteur localisé dans la région LTR amont et vont se retrouver associés au sein de la même unité transcriptionnelle pour former un ARNm polycistronique, sous l'action de l'ARN polymérase II. Les gènes *gag* et *pol* codent des poly-protéines qui seront maturées par l'action de la protéase codée par le gène de *pol*. Dans ce gène on retrouve également des portions codant pour les activités reverse transcriptases (RT), RNaseH, et intégrase (IN). *Pol* assure ainsi la maturation des polypeptides, la reverse transcription de l'ARNm initial en ADNc, l'acheminement de cet ADNc vers le noyau, et enfin, son intégration dans le génome lui-même. Le gène *gag* quant à lui, code pour des protéines qui s'assembleront en particule dans le cytoplasme, et c'est au sein de cette particule qu'aura lieu la reverse transcription (**cf. figure 13**) (144). Les régions LTR comme leur nom l'indique sont constituées de longues répétitions directes de séquences courtes et sont situées de part et d'autre du rétrotransposon (**cf. figure 14**) (149). Ces répétitions peuvent faire de plusieurs centaines à plusieurs milliers de paires de bases (143). Ces séquences présentent également des rôles de promoteurs chez les virus (150).

La troisième grande sous classe est composée des rétrotransposons YR. Cette classe est similaire structurellement aux LTR mais encode une tyrosine kinase (YR) à la place d'une intégrase. De plus, les rétrotransposons YR présentent bien des répétitions terminales mais celles-ci diffèrent entre les superfamilles majeures et se distinguent des LTR canoniques par leurs structures (94,144,151). Les rétrotransposons YR sont notamment représentés par les

Class I



Class II - subclass 1



Class II - subclass 2

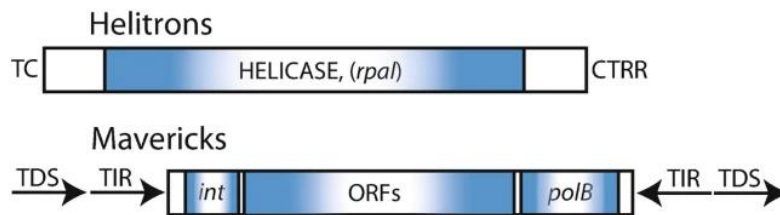


FIGURE 14 : STRUCTURES DES ELEMENTS TRANSPOSABLES EUCARYOTES

Structures des principaux éléments transposables eucaryotes. TSD = Target Site Duplication, duplication de quelques nucléotides autour du site d'insertion ; TIR = Terminal Inverted Repeats. (Adapté de Makalowski *et al* 2019 (143)).

éléments DIRS (**cf. figure 14**), ou dans les génomes de kinétoplastidés par les éléments VIPER et TATE (94). D'autres sous-classes de rétrotransposons existent mais sont moins représentées dans le vivant (**cf. tableau 1**) (143,144).

I.B. LA CLASSE II : LES TRANSPOSONS A ADN

4 ordres principaux de transposons à ADN sont connus : Les éléments se transposant par « couper-coller » et mobilisés par l'action des transposases DDE (nommés ainsi en liens avec les résidus aspartate et glutamate présents dans leur site catalytique) également appelés éléments TIR (Terminal Inverted Repeats) ; les tyrosine recombinases (YR) appelés *Cryptons* ; Les Helitrons qui se transposent par un système de cycle roulant ; et enfin les *Mavericks* (ou *Polintons*) qui s'auto-synthétiseraient (143,144).

Les transposons DDE et les *Cryptons* sont généralement constitués d'un seul ORF codant une recombinase entourée de petites séquences répétées inversées (short terminal inverted repeats, TIR). Contrairement aux *Cryptons* qui sont assez rares (144,152), les transposons DDE sont les ET les plus diversifiés et les plus répandus chez les eucaryotes avec plus de 17 superfamilles, différenciées par des transposases phylogénétiquement distinctes (144,153). La recombinase DDE serait même un des gènes les plus anciens et les plus répandus sur terre (154). La transposition par les transposases DDE est médiée par une attaque nucléophile d'une molécule d'eau à proximité des extrémités des TIR, ce qui conduirait à une excision directe de l'élément puis à une relocalisation. La multiplication et la diversification des familles de TIR se ferait par deux stratégies. La première consisterait en une transposition lors de la réplication de l'ADN de l'hôte d'un site déjà répliqué à un site en cours de réplication (144,155,156). La seconde impliquerait une recombinaison homologe après la cassure double brin de leur site d'excision (144,155,156). Ce second processus peut entraîner la perte de séquences internes donnant naissance à des ET ne possédant plus la transposase mais conservant les séquences TIR. Ces ET non-autonomes pourront s'appuyer sur les transposases d'autres ET, et parfois même se répliquer plus efficacement qu'eux. Ils formeront ainsi de larges familles comme les MITE (Miniature Inverted-repeat Transposable Elements) (144,155,156) (**cf. figure 14**).

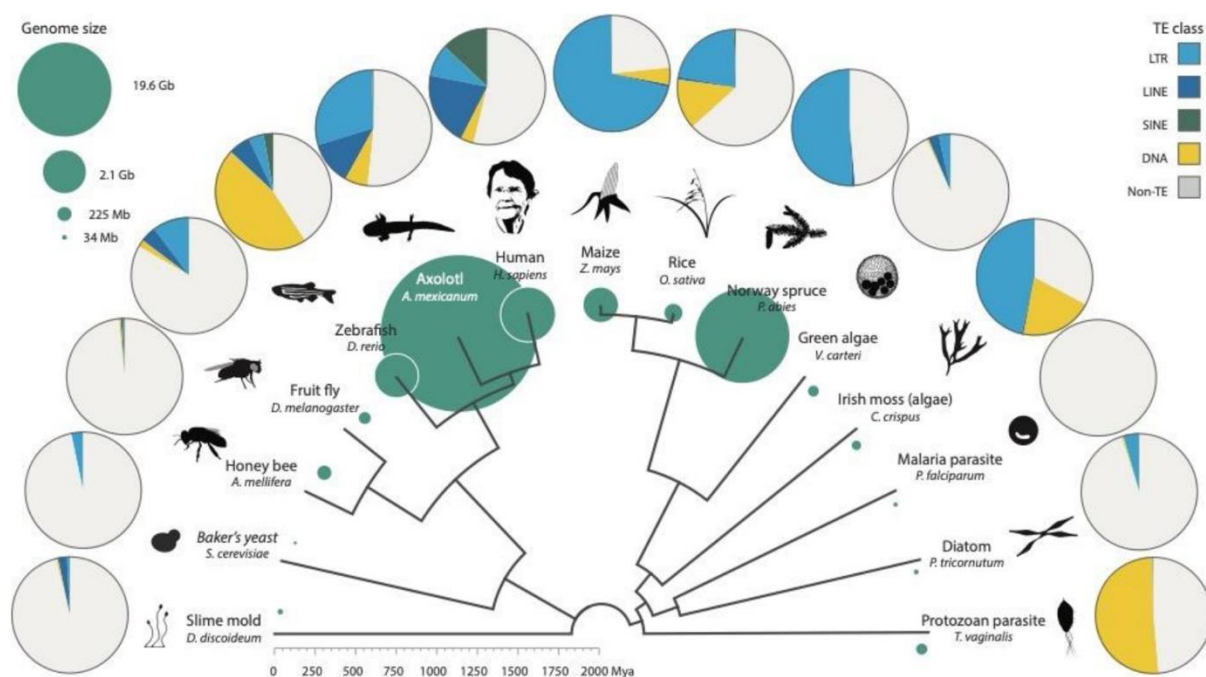


FIGURE 15 : DISTRIBUTION DES ELEMENTS TRANSPOSABLES DANS LA PHYLOGENIE EUKARYOTE

Représentation schématique de l'arbre phylogénétique des eucaryotes avec la taille des génomes représentée par des cercles verts de tailles différentes, et la diversité des ET représentée par des diagrammes en secteur. Les éléments transposables de classe II ont été regroupés sous « DNA ». Les données sont issues de l'analyse des génomes avec RepeatMasker (157). (Adapté de Jonathan N. Wells et Cédric Feschotte, 2020 (144), (158).

Les *Helitrons* sont abondants dans de nombreuses lignées d'eucaryotes. Cet ordre est principalement constitué d'ET non-autonomes et ne présentant pas de TIRs. Leur transposition s'effectue par la production d'un ADN double brin circulaire intermédiaire (cf. **figure 13**).

Enfin, les *Mavericks* (ou *Polintons*) constituent un groupe encore mal caractérisé. Ils sont remarquables par leurs tailles de 15 à 20 Kb et leur complexité avec jusqu'à 20 gènes codants, le tout entouré de long TIR de 400 à 700 paires de bases (144,145,159). Cette sous-classe d'ET est répandue chez les eucaryotes mais est généralement présente en faible nombre de copies à l'exception de quelques protistes comme *Trichomonas vaginalis* où ils représenteraient un tiers du génome (144,159). Les *Mavericks* présentent des similitudes avec différents groupes de virus à ADN double brin comme les adénovirus, notamment la présence d'une ADN polymérase B (polB), suggérant que ces éléments soient capables de se répliquer par synthèse directe d'ADN. Les *Mavericks* encodent également des DDE nucléases apparentées aux intégrases rétrovirales (144,145,159).

II. REPARTITION ET ROLES

Bien que très divers, les ET ne sont pas distribués de manière homogène entre les espèces eucaryotes (144). Les ET ont eu et ont encore un impact important sur l'évolution et la diversité des génomes eucaryotes (144).

II.A. REPARTITION DES ELEMENTS TRANSPOSABLES DANS LE VIVANT

Très peu d'espèces ne présentent pas d'ET dans leurs génomes. Quelques exceptions existent comme *Plasmodium falciparum*, *Toxoplasma gondii* ou encore les espèces du genre *Encephalitozoon* qui semblent avoir éliminé les ET de leurs génomes au cours de leur évolution (144). Cette perte d'ET ne concerne pas nécessairement les organismes qui leurs sont proches phylogénétiquement. En effet, certaines espèces proches peuvent présenter une forte abondance et une forte diversité d'ET (128,144,160). Le génome d'*A. algerae*



par exemple, a été envahi par une multitude de familles d'ET (112,128). Une étude de E.J Pritham *et al* décrit une diversité différente d'ET entre différentes espèces d'*Entamoeba*, avec des ratios Class I / Class II qui varient ; ainsi que des degrés divers d'amplification des ETs (160). L'abondance et la diversité des ET dans les génomes ne semble pas uniquement liées au mode de vie, à la taille des génomes, ou au genre considéré (**cf. figure 15**) (144). Le devenir des ET dans un génome dépend fortement de leurs fréquences de transposition, leurs préférences d'insertion mais également de leurs sensibilités aux processus de régulation mis en place par l'hôte (144).

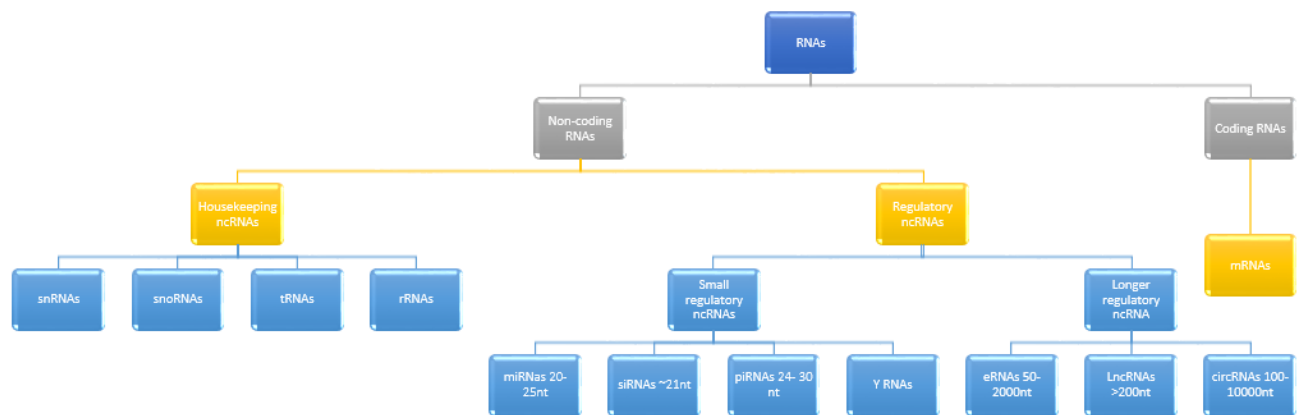
II.B. ROLES DES ELEMENTS TRANSPOSABLES

Les éléments transposables semblent avoir joué un rôle majeur dans l'évolution des génomes eucaryotes (161–163). Les ET semblent influencer les génomes par trois mécanismes principaux : tout d'abord par l'altération de la fonction et de la structure de certains gènes par insertion ; ensuite par l'induction de réarrangements chromosomiques conduisant à des changements notamment en termes de régulation et d'expression ; et enfin, en tant que source de nouvelles séquences codantes et non codantes permettant l'émergence d'innovations fonctionnelles et régulatrices (161,163).

Chez les microsporidies, les éléments transposables semblent représenter un facteur important d'évolution, d'adaptation et d'interaction avec leurs hôtes (124,128,164).

Au même titre que pour les ET, un intérêt croissant est né ces dernières années dans la littérature scientifique pour les rôles des ARN non-codants, notamment dans les phénomènes de communication intercellulaire (165–167), de régulation génique, et dans l'évolution des génomes de manière plus générale. Ces molécules pourraient constituer des acteurs importants dans le cadre des interactions hôte-parasite, notamment chez les microsporidies.

Les ARNs



Adapted from Ghildiyal et Zamore, 2009 Nat Rev Genet. and P.Zheng et al. Sep 2019 J Integr Bioinform.

FIGURE 16 : DIFFERENTES CLASSES D'ARN

Diagramme présentant les différentes classes d'ARN incluant les ARN codants, et les ARN non codants. Parmi les ARN non codants on retrouve les ARNnc de ménage avec les snRNA, les snoRNA, les tRNA et les rRNA. Pour les ARNnc régulateurs, on distingue les petits ARNnc (<200nt), principalement les miRNA (21-23nt), les siRNA (20-25nt), les piRNA (26-32nt), et les long ARN non-codants, dont les LncRNA, les eRNA et les circRNA (Nt = Nucléotides), (Les tailles de ces ARNnc varient en fonction des espèces mais également légèrement entre auteurs (168,169)). (Adapté de Ghildiyal et Zamore, 2009 et P.Zhang et al. 2019 (168,169)).

PARTIE II : LES ARN NON CODANTS

Le terme ARN non-codant (ARNnc) définit un groupe de molécules hétérogènes dans leur structure, leur biogénèse, et dans leur rôle biologique (168,170). Leur point commun réside principalement dans le fait qu'ils ne soient pas traduits en protéines contrairement aux ARN messagers (ARNm). Ils agissent en formant des complexes ribo-protéiques et influent sur l'expression des génomes (168,170). Les ARNnc participent à des phénomènes de régulation post-transcriptionnelle mais aussi épigénétique (168,170,171).

Deux catégories d'ARNnc peuvent être distinguées : les ARNnc dits de « ménage » qui interviennent dans des processus généraux de la cellule comme la transcription et la traduction ; et les ARNnc dits régulateurs (**cf. figure 16**) (168). Les ARNnc de ménage comprennent les snRNA (small nuclear RNA) impliqués dans la formation et la structure des complexes d'épissage des introns d'ARNm, les snoRNA (small nucleolar RNA) qui participent à la maturation des ARN ribosomiques dans le nucléole des noyaux cellulaires, les rRNA (ribosomal RNA) qui entrent dans la constitution des ribosomes (complexe central du processus de synthèse protéique), et les tRNA (transfer RNA), eux aussi impliqués dans la synthèse protéique en permettant la correspondance entre la séquence d'acide nucléique traduite et la séquence polypeptidique formée (170).

Les ARNnc régulateurs forment eux aussi un groupe hétérogène, avec des rôles très divers dont certains demeurent inconnus. Ils sont subdivisés en fonction de la taille de leurs séquences nucléotidiques et de leurs partenaires protéiques (168,170). On retrouve les petits ARNnc (sncRNA) (<200nt), impliqués dans des phénomènes de régulation épigénétique, transcriptionnelle et post-transcriptionnelle séquence spécifiques (microRNA (miRNA), small interfering RNA et ARN interference (siRNA), PIWI-interacting RNA (piRNA)) ; et les long ARNnc (lncRNA) (> 200nt), jouant eux aussi des rôles de régulation (post-)transcriptionnels et épigénétiques mais également souvent impliqués dans la structure et formation de complexes ribonucléoprotéiques (168,170,172). Les ARNnc sont d'importantes molécules régulatrices et leur action passe le plus souvent par une complémentarité en séquences nucléotidiques entre l'ARNnc et une autre séquence nucléotidique ciblée (ADN, ARNm) (168,170,173).



CHAPITRE 1 : LA REGULATION PAR L'ARN

I. MISE EN EVIDENCE HISTORIQUE DE L'INTERFERENCE PAR L'ARN

Le phénomène d'interférence par l'ARN (ARNi) a initialement été mis en évidence chez le pétunia indépendamment par Van der Krol *et al.* (174) et Napoli *et al.* (175).

Les deux groupes décrivaient un phénomène de répression de l'expression de certains gènes suite à l'introduction exogène soit de copies surnuméraires de ces gènes (van der Krol *et al.* (174)), soit d'un gène chimérique avec un promoteur non natif permettant sa surexpression (Napoli *et al.* (175)). Dans les deux cas cette répression n'affectait que les gènes homologues (174,175).

Ce phénomène a ensuite été observé chez d'autres plantes comme le tabac, la tomate ou encore *Arabidopsis thaliana* et fut baptisé « homology-dependent silencing » ou répression homologie-dépendante (176,177).

Une percée dans le domaine fut l'apport des travaux de Baulcombe *et al.* (178), qui ont suggéré l'implication des acides nucléiques dans le phénomène. L'ARN est présenté comme l'initiateur et la cible du phénomène, qui sera nommé répression post-transcriptionnelle de gènes (post-transcriptional gene silencing) (178,179).

D'autres études ont également confirmé l'implication de l'ARN dans ce processus, notamment chez le ver *C.elegans*, où une forte répression induite par l'introduction exogène d'ARN double brin (dsRNA) est décrite par A Fire *et al.* (180). Hamilton et Baulcombe (179) montrent également que l'interférence par l'ARN est médiée chez les plantes (tomate, tabac, *A.thaliana* et pétunia) par de petits ARN bicaténaire d'une taille d'environ 25 nt (siRNA, small interfering RNA) dont l'un des brins est antisens d'ARNm ((179).

Un autre évènement majeur dans la caractérisation de l'ARN interférence fut la mise en évidence de la machinerie protéique qui est associée à la maturation des ARN interférents et à l'activité d'ARN interférence par Bernstein *et al.* (181). Ils ont cherché à déterminer si l'étape de clivage des dsRNA en siRNA (qu'ils ont appelée phase initiatrice) et l'étape de clivage de

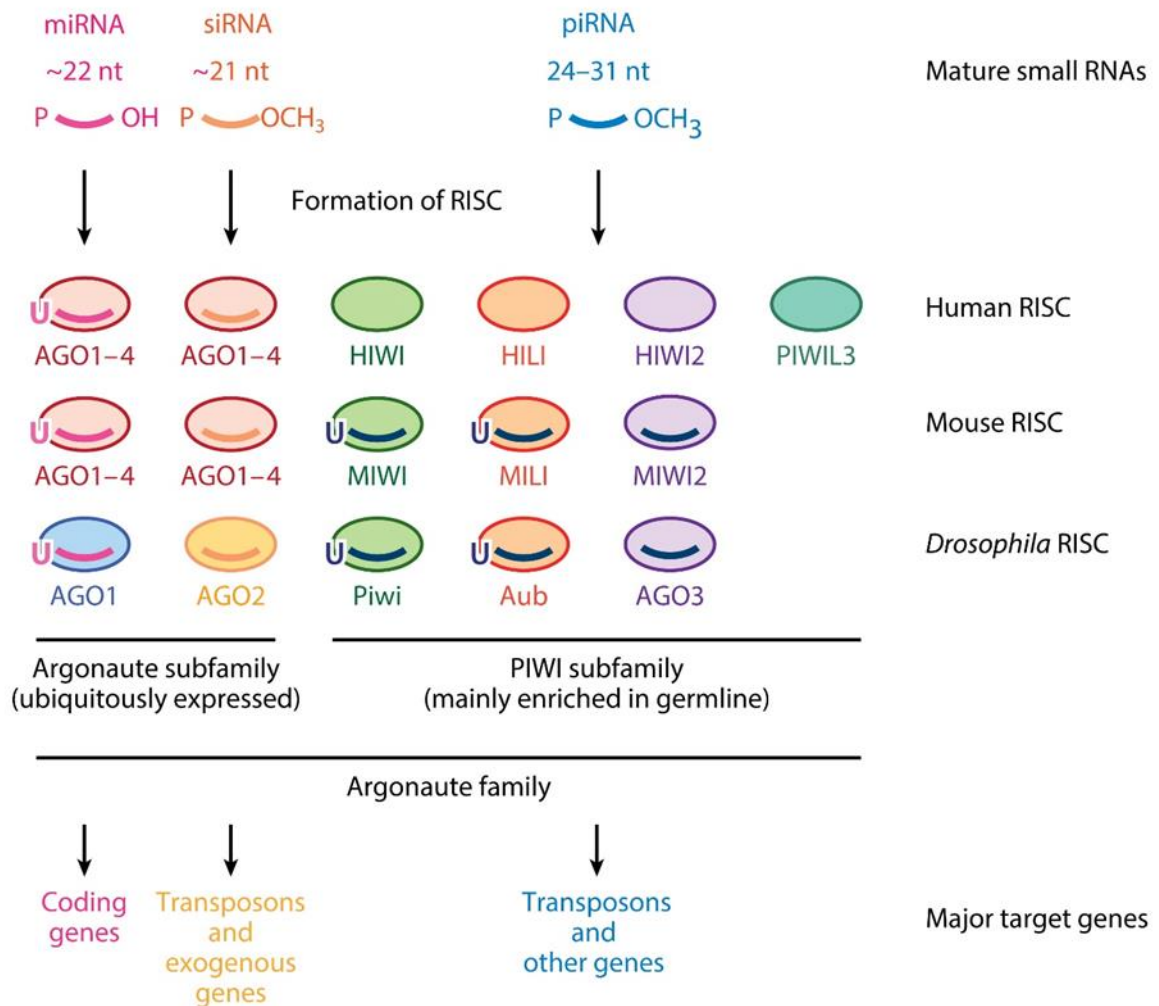


FIGURE 17 : PETITS ARN INTERFERENTS, PARTENAIRES ARGONAUTES, ET CIBLES PRINCIPALES

Exemple chez l'Homme, la souris et la drosophile. (Adapté de Iwasaki et al. 2015 (182)).

l'ARNm cible (nommée phase effectrice ou phase d'activité RNA-induced silencing complex (RISC)) impliquaient les mêmes enzymes. En travaillant sur des lysats de cellules de drosophile, ils ont pu montrer que les deux activités enzymatiques étaient portées par deux fractions protéiques différentes (181). Tabara *et al.* travaillant également sur la drosophile, identifient que seule une enzyme, la RNase III nommée Dicer, était capable de cliver les dsRNA en siRNA d'une taille caractéristique de 21 à 23 nucléotides. Dicer contient deux motifs RNase III, un domaine de liaison aux dsRNA, un domaine hélicase en position N-terminale, et un domaine PAZ. Le domaine PAZ est également présent sur les protéines RDE1, QDE2 et Argonaute, qui ont toutes les trois été associées à l'ARNi (181,183). Des orthologues de Dicer ont par la suite été observés chez tous les organismes étudiés avec une activité RNAi décrite (184).

En ce qui concerne la phase effectrice de l'ARNi, Martinez *et al.* (185), ont co-immunoprécipité le complexe protéique associé aux siRNA dans des cellules HeLa (lignée cellulaire humaine), et ont ainsi pu identifier les protéines Argonaute 1 (Ago 1) et Argonaute 2 (Ago 2) ainsi que de nombreuses autres protéines constituant le « RNA Induced Silencing Complexe » (ou RISC). La famille Argonaute regroupe des protéines ayant un domaine PIWI et un domaine PAZ (PIWI-Argonaute-zwille). Le domaine PIWI peut posséder une activité de clivage nucléotidique (activité slicer) qui n'est cependant pas retrouvée chez toutes les protéines Argonaute (172,186). Le domaine PAZ quant à lui, permet de lier de manière non spécifique les nucléotides en 3' des petits ARNs à la protéine (172,186). Les différentes protéines Argonaute peuvent être à l'origine de complexes protéiques distincts n'ayant pas la même affinité pour les différents types de petits ARN. Ainsi, chez *Drosophila melanogaster*, alors que la protéines AGO 1 se lie principalement aux miRNA, la protéine AGO 2 elle, est spécifique des siRNA. Ces deux protéines font pourtant partie de la même sous famille de protéines Argonautes, les AGO. Ce partitionnement entre différentes protéines Argonaute des petits ARNnc ne semble pas universel cependant, chez les vertébrés par exemple ceci n'est pas systématiquement observé. Cependant, Il existe une deuxième sous famille de protéines Argonautes nommée sous-famille PIWI, dont les protéines Piwi, aubergine et AGO3 de la drosophile font partie, et qui se lie spécifiquement aux piRNA, chez toute espèce en possédant (168,182,186) (**cf. figure 17**). La composition nucléotidique des ARN ainsi que certaines modifications enzymatiques

TABLEAU 2 : TYPES ET FONCTIONS DES PETITS ARN INTERFERENTS

Ago2, Argonaute2; casiRNA, cis-acting siRNA; DCL, Dicer-like; endo-siRNA, endogenous small interfering RNA; exo-siRNA, exogenous small interfering RNA; miRNA, microRNA; natsiRNA, natural antisense transcript-derived siRNA; piRNA, Piwi-interacting RNA; Pol II, RNA polymerase II; pri-miRNA, primary microRNA; RdRP, RNA-dependant RNA polymerase; tasiRNA, trans-acting siRNA ; la colonne « Proteins » correspond aux protéines impliquées dans la maturation de l'ARNnc. (D'après M.Ghildiyal and P.D.Zamore, 2009 (168)).

Name	Organism	Length (nt)	Proteins	Source of trigger	Function
miRNA	Plants, algae, animals, viruses, protists	20–25	Drosha (animals only) and Dicer	Pol II transcription (pri-miRNAs)	Regulation of mRNA stability, translation
casiRNA	Plants	24	DCL3	Transposons, repeats	Chromatin modification
tasiRNA	Plants	21	DCL4	miRNA-cleaved RNAs from the TAS loci	Post-transcriptional regulation
natsiRNA	Plants	22	DCL1	Bidirectional transcripts induced by stress	Regulation of stress-response genes
		24	DCL2		
		21	DCL1 and DCL2		
Exo-siRNA	Animals, fungi, protists	~21	Dicer	Transgenic, viral or other exogenous dsRNA	Post-transcriptional regulation, antiviral defense
	Plants	21 and 24			
Endo-siRNA	Plants, algae, animals, fungi, protists	~21	Dicer (except secondary siRNAs in <i>C. elegans</i> , which are products of RdRP transcription, and are therefore not technically siRNAs)	Structured loci, convergent and bidirectional transcription, mRNAs paired to antisense pseudogene transcripts	Post-transcriptional regulation of transcripts and transposons; transcriptional gene silencing
piRNA	Metazoans excluding <i>Trichoplax adhaerens</i>	24–30	Dicer-independent	Long, primary transcripts?	Transposon regulation, unknown functions
piRNA-like (soma)	<i>Drosophila melanogaster</i>	24–30	Dicer-independent	In <i>ago2</i> mutants in <i>Drosophila</i>	Unknown
21U-RNA piRNAs	<i>Caenorhabditis elegans</i>	21	Dicer-independent	Individual transcription of each piRNA?	Transposon regulation, unknown functions
26G RNA	<i>Caenorhabditis elegans</i>	26	RdRP?	Enriched in sperm	Unknown

qu'ils subissent influencent également l'affinité des protéines Argonaute pour ces petits ARN (168,187,188).

Aujourd'hui, les techniques de séquençage haut débit ainsi que le développement de nouvelles méthodes d'étude des ARNs permettent une meilleure identification de ces molécules et une connaissance de plus en plus approfondie des phénomènes biologiques qu'ils régissent (189,190). Les phénomènes d'ARN interférence ainsi que d'autres phénomènes de régulation par l'ARN ont ainsi été décrits chez de nombreux organismes (191) et de nombreux types de petits ARNnc régulateurs ont été identifiés avec des caractéristiques et des rôles propres (168) (**cf. tableau 2**). Les trois types de petits ARNnc régulateurs les plus étudiés sont les siRNA (21nt environ) ; les miRNA (20-25nt environ) ; et les piRNA (24-30nt environ) (168). Les fourchettes de tailles exactent varient en fonction des espèces et parfois des auteurs (168,169). Ces trois classes participent à des processus de régulation de l'expression de gènes lors de la transcription (Transcriptional Gene Silencing, TGS), ou après cette dernière (Post-Transcriptional Gene Silencing PTGS) (192).

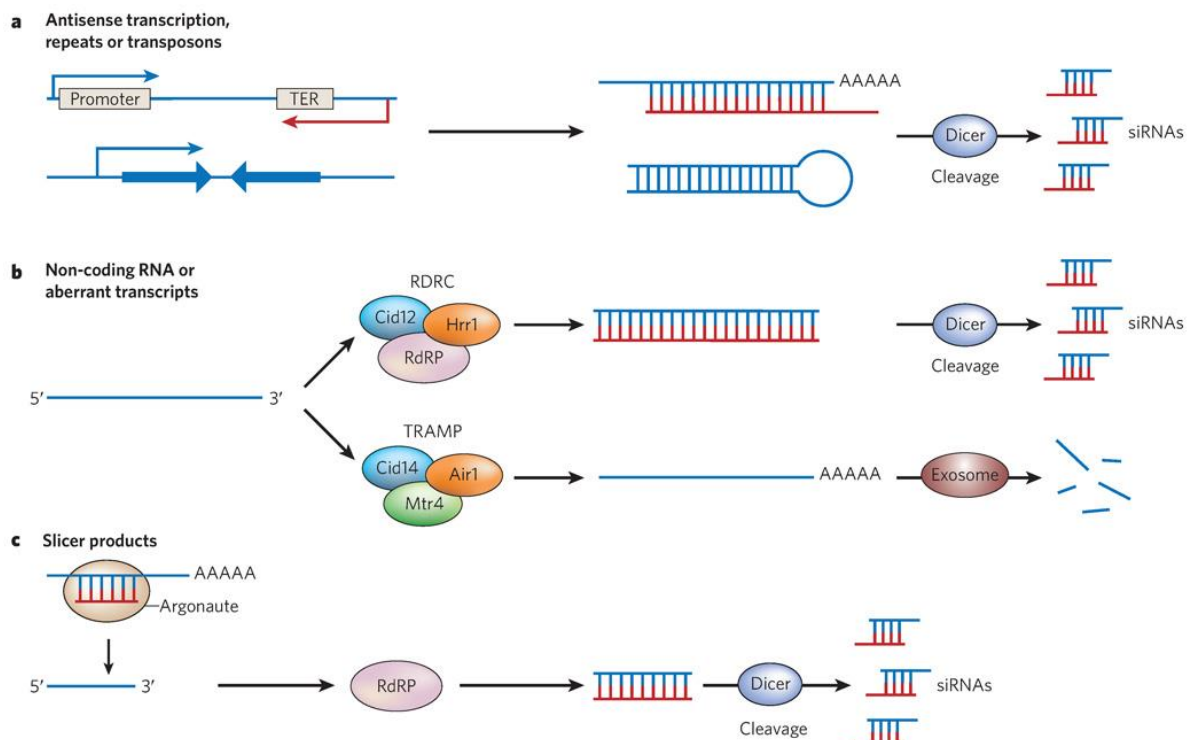


FIGURE 18 : VOIES DE BIOGENESE DES SIRNA

Voies de biogénèse des siRNA chez la levure *S.pombe*. (a) Génération de siRNA à partir de dsRNA issus d'une transcription convergente (ARN sens et antisens complémentaires) ou d'une transcription à travers des séquences répétées inversées (formation d'un ARN tige-boucle). TER = Transcription End Region, Region de terminaison de transcription. (b) siRNA issus d'ARN non-codants ou d'ARN aberrants pris en charge par les voies RDRC/Dicer et TRAMP/exosomes respectivement. (c) siRNA issus de l'action de RdRP (RNA dependant RNA polymerase) sur des produits de dégradation de l'activité slicer d'un complexe RISC donnant un transcrit d'ARN double brin complémentaire (dsRNA). Ce transcrit sera maturé en siRNA par Dicer. Les Drosophiles ainsi que les mammifères ne présentent pas de RdRPs. (D'après Moazed 2009 (193)).

II. LES siRNA : BIOGENESE ET ROLES

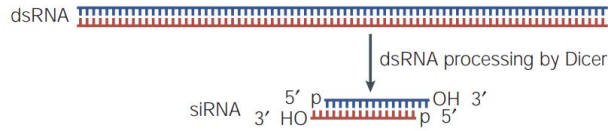
II.A. BIOGENESE ET MODE D'ACTION

Les siRNA ou petits ARN interférents sont retrouvés chez la plupart des Eucaryotes (*i.e.* plantes, animaux, champignons et protozoaires) (193). Ils découlent classiquement d'un long ARN double brin maturé par Dicer, libérant de petits ARN double brins avec des extrémités sortantes non appariées en 3' d'environ deux nucléotides. Les siRNA présentent une taille de 21 à 25 nt environ, mais qui peut varier entre organismes (194). Chez la drosophile les siRNA sont 2' O méthylés sur l'extrémité 3' OH par l'enzyme HEN1 (168). Cette méthylation influencerait les protéines et les complexes RISC avec lesquels ils interagiraient (168,187).

Les longs ARN double brin à l'origine des siRNA peuvent être introduits de manière exogène lors d'infections virales, de thérapies, ou d'expérimentations (siRNA exogènes), ou être issus de l'organisme lui-même (siRNA endogène). Différents processus peuvent conduire à la formation de siRNA endogènes : ils peuvent provenir de transcrits antisens complémentaires (pseudogènes, ARN non codants, promoteurs opposés sur deux brins, gènes dupliqués inversés (168,195)), d'une transcription de séquences répétées, de séquences non-codantes, de transposons présentant des structures en tige-boucle sur leurs transcrits, ou enfin de transcrits maturés ou partiellement dégradés et pris en charge par une RNA dépendante RNA polymérase (RdRP) (193)(**cf. figure 18**). Les mammifères ainsi que *Drosophila melanogaster* ne possèdent pas de RdRP et ne peuvent donc pas produire de siRNA par ce dernier mécanisme (193).

Une fois générés, les siRNA seront pris en charge par un complexe RISC et permettront la reconnaissance et le clivage spécifique de séquences d'ARN complémentaire au siRNA (**cf. figure 19**). Le complexe siRNA-RISC (siRISC) régule son transcrit cible par clivage endonucléolytique en vis-à-vis des nucléotides 10 et 11 du siRNA (**cf. figure 19**). Un certain degré de complémentarité est nécessaire au clivage (*i.e.* pas nécessairement parfaite en particulier en ce qui concerne l'extrémité 3' du siRNA (196)). Le clivage entraîne la répression post-transcriptionnelle (PTGS) de la cible par sa dégradation complète. Pour cela, la portion 3' de la cible sera dégradée par l'exoribonucléase 5'->3' XRN1, et la portion 5' par l'exosome (un complexe multimérique d'exonucléases 3'->5').

Initiation phase



Effector phase

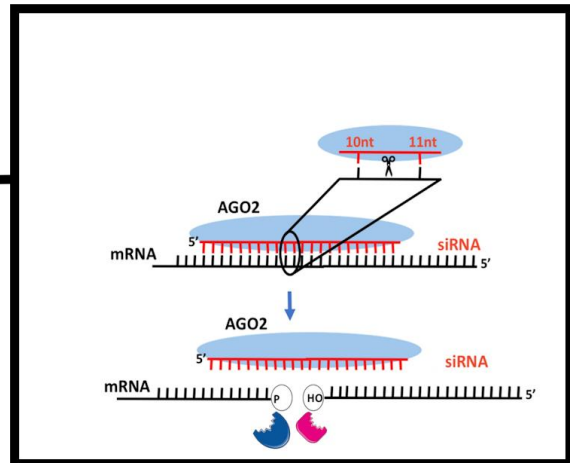
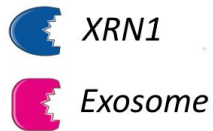
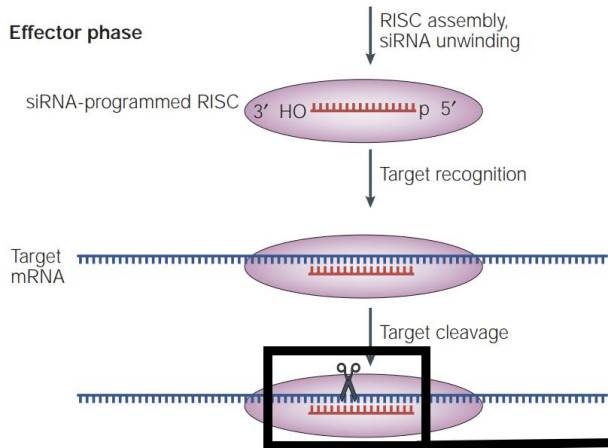


FIGURE 19 : REGULATION POST-TRANSCRIPTIONNELLE PAR LES siRNA

Processus de régulation post-transcriptionnel par les siRNA. (Adapté de Sontheimer *et al.* 2005 (197) et de Gu *et al.* (198))

Ces mécanismes de PTGS semblent conservés à travers le vivant (168,193,199,200).

La régulation par les siRNA peut également se faire par un processus dépendant de la chromatine (répression transcriptionnelle des gènes, TGS). Dans ce processus, le complexe RISC lié au siRNA est transloqué dans le noyau où il pourra reconnaître un ARN en cours de transcription (ARN naissant). La liaison pourra induire un clivage (répression co-transcriptionnelle), mais recrutera aussi des protéines de modification de la chromatine (notamment des histones méthyl-transférases), qui induiront sa compaction au niveau du transcrit naissant (hétérochromatination) (193,201).

La TGS par les siRNA a été décrite chez de multiples espèces et joue des rôles importants dans leur biologie. Chez la levure *Schizosaccharomyces pombe*, la TGS induite par les siRNA a été associée à la formation et au maintien des centromères par l'action de siRNA issus des répétitions centromériques et qui assure leur reconnaissance par complémentarité (193,202). Dans le règne animal les siRNA semblent principalement impliqués dans la défense de l'hôte face à des ARN étrangers double brin comme ceux constituant le génome de certains virus à ARN (191).

II.B. ROLES DES siRNA DANS LES INTERACTIONS HÔTE-PARASITE

Certains parasites produisent des siRNA qu'ils utilisent comme médiateurs dans les interactions avec leurs hôtes. Une des premières descriptions concerne *Botrytis cinerea*, un pathogène de plante, utilisant ses ARNnc similaires à des siRNA au cours de l'infection de son hôte (203).

Botrytis cinerea est un champignon pathogène de plantes capable d'infecter plus de 200 espèces différentes (203). A Weiberg *et al* ont étudié et prédit 73 petits ARN de *B.cinerea* (Bc-sRNAs) ciblant des gènes chez l'hôte (203). La majeure partie de ces Bc-sRNA présentait une taille de 20 à 22 nt et un résidu uracile en 5' terminale. Leur étude a pu démontrer l'implication de trois de ces Bc-sRNAs dans la répression de gènes d'immunité chez les plantes *Arabidopsis thaliana* et *Solanum lycopersicum* (tomate). Ces 3 Bc-sRNA étaient produits par 2 rétrotransposons LTR (2 des Bc-sRNA étaient issus du même transposon LTR mais avec un



décalage de 4 nt entre eux) (203). Les auteurs ont également pu démontrer que ces 3 Bc-sRNAs étaient maturés par la protéine Dicer du parasite, puis qu'ils se liaient à la protéine AGO 1 de l'hôte pour réprimer les gènes de l'immunité des plantes hôtes. Ces résultats représentent une des premières preuves concrètes de l'implication des ARNnc dans les interactions hôte-pathogène (203).

Les plantes infectées par des parasites fongiques comme *B.cinerea* peuvent elles aussi transférer de petits ARNnc aux parasites (204,205). Le transfert des siRNA de la plante vers le pathogène impliquerait des vésicules extracellulaires (EV) (204,205). Les EV sont des structures produites par les cellules vivantes, et constituées d'une bicouche phospholipidique pouvant contenir des protéines, des sucres ou encore des acides nucléiques (206). Le cargo transporté par les EV semble méticuleusement régulé et semble dépendre du contexte biologique (204,206). Ce sont des médiateurs importants pour l'activité des ARNnc dans les interactions intercellulaires (206–208). Lorsque la plante *Arabidopsis thaliana* est infectée par *B.cinerea*, elle libère des vésicules extracellulaires qui s'accumulent aux sites d'infection et sont incorporées par le champignon (204). Les siRNA contenus dans ces vésicules induisent ensuite une répression de gènes clés dans la pathogénicité de *B.cinerea*, notamment le gène Bc-VPS51 (vacuolar protein sorting 51), impliqué dans le trafic vésiculaire du parasite (204).

Ainsi les siRNA peuvent être des médiateurs importants de l'interaction hôte-parasite. D'autres exemples existent comme les endosiRNA du ver parasitaire *Schistosoma japonicus*, qui ont été montrés comme modulant le comportement des macrophages de l'hôte au cours de l'infections (209).

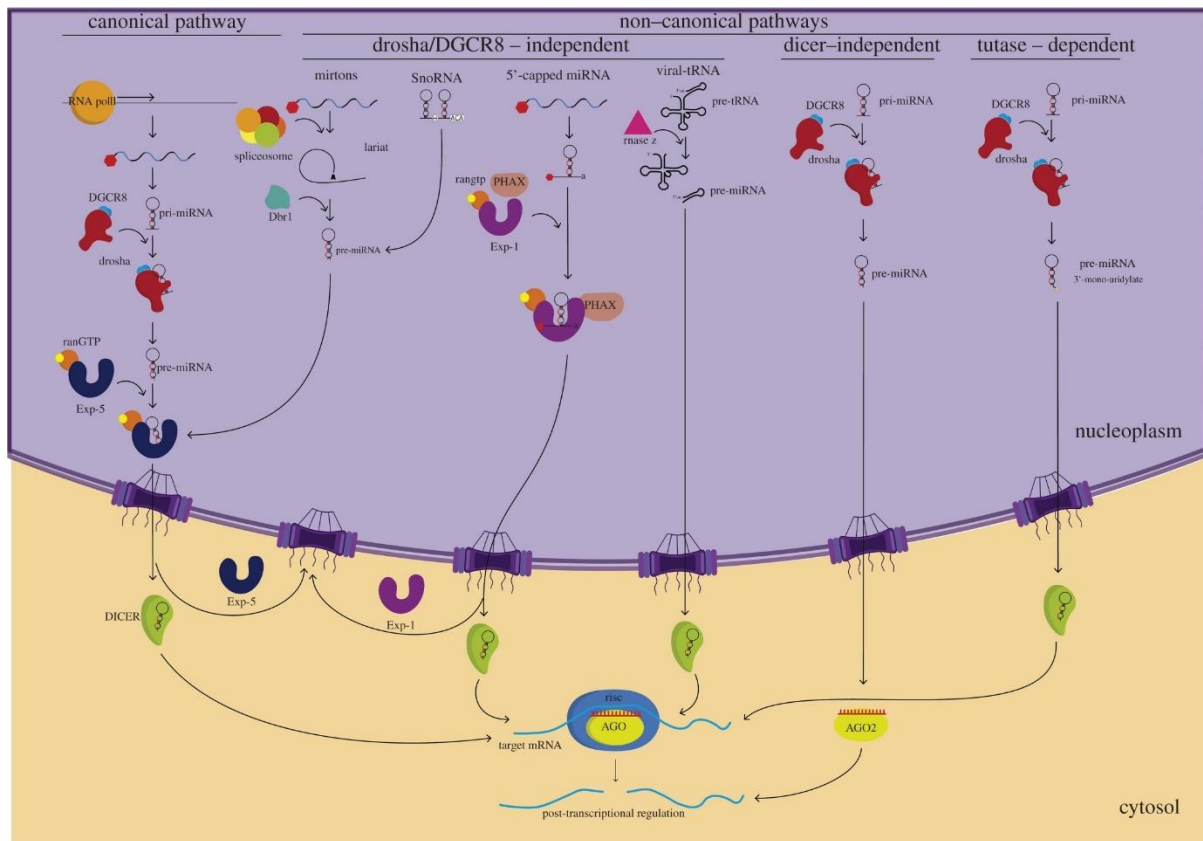


FIGURE 20 : VOIES CANONIQUES ET NON CANONIQUES DE BIOGENESE DES MIRNA CHEZ L'HOMME

(D'après M Rojas-Pirela *et. al* 2022 (210)). Il est important de noter que ces différentes voies ne sont pas observées à la même fréquence ni pour le même nombre de miRNA. La voie indépendante de Dicer n'a par exemple été observée que pour le miR-451. La voie canonique reste majoritaire (168,210).

III. LES MICRORNA

III.A. BIOGENESE ET MODE D'ACTION

Les microRNA (miRNA) sont une classe de petits ARNnc composés de 20 à 25 nt selon les organismes. Leur synthèse est initiée dans le noyau à partir d'un transcrit primaire (primary miRNA ou pri-miRNA) structuré en tige-boucle et transcrit par l'ARN polymérase II (Pol II). Ce pri-miRNA est ensuite mûré de manière canonique ou non canonique (**cf. figure 20**). La maturation canonique se fait classiquement par l'action successive d'endonucléases de type III. Chez l'homme, le complexe Drosha/DGCR8 mature le pri-miRNA le réduisant ainsi à une simple structure tige-boucle (precursor miRNA ou pre-miRNA), qui présente en générale 2 nt non appariés aux extrémités 3'. Le pre-miRNA sera ensuite pris en charge par l'exportine 5 qui assurera son transfert du noyau vers le cytoplasme, au niveau duquel la ribonucléase Dicer clivera la structure tige boucle et laissera un petit ARN double brin chevauchant à la complémentarité généralement incomplète, et présentant des extrémités 5' phosphate et 3' OH (165). Ce petit ARN double brin sera pris en charge par un complexe RISC. Un des brins du miRNA, appelé miRNA « * » ou miRNA star, sera libéré et dégradé. L'autre brin constituera le miRNA mature et servira de guide pour une activité de régulation post-transcriptionnelle (PTGS) (165). Le choix du brin dégradé et de celui conservé semble dicté par des équilibres thermodynamiques et la maturation préalablement effectuée par Drosha (187,200,211). Il existe également des voies de synthèse des miRNA dites « non-canoniques » mais observée chez beaucoup moins de miRNA et beaucoup moins fréquemment (**cf. figure 20**) (168,200,210). Les miRNA peuvent par exemple être produit à partir d'un transcrit coiffé en 5' par une 7-méthyl-guanosine ou encore à partir d'autres ARNnc comme les snoRNA ou les tRNA. Il existe également des voies de maturation des miRNA qui sont indépendantes de Drosha, comme pour les mirtrons, qui sont issus d'introns directement épissés en pre-miRNA, ou encore des voies de maturations qui sont indépendantes de Dicer (exception pour l'instant observée uniquement pour le miR-451) (**cf. Figure 20**) (168,210,212).

La régulation des transcrits par le complexe miRNA RISC nécessite un appariement dans une région en 5' du miRNA appelé région « seed ». Le seed a une longueur d'environ 7 nt et s'étale du second nucléotide au septième, avec un « extended seed » ou « seed region »

		miRNA:mRNA	Site type
A	<pre> 5' -AACUCACAACCAACUCAGGGA-3' O 3' AGUGUGAACUCCAGAGUCCCU 5' </pre>	<i>lin-14</i> 3' UTR <i>lin-4</i> (<i>C. elegans</i>)	Seed site (pairing to miRNA nt 2-8)
B	<pre> 5' -UUUUAUACAACC UUCUGCCUCU-3' O 3' UUGAUUGUUGG AUGAUGGAGU 5' </pre>	<i>lin-41</i> 3' UTR <i>let-7</i> (<i>C. elegans</i>)	3' compensatory site (imperfect seed pairing, extensive 3' pairing)
C	<pre> 5' -CCCAACAACAUGAAACUGCCUA-3' O 3' GGGUUGUUGUACUUUGAUGGGAU 5' </pre>	<i>HOXB8</i> 3' UTR miR-196 (Human)	3' compensatory site (imperfect seed pairing, extensive 3' pairing)
D	<pre> 5' -AGUUUUUCAGUCUGAUUAU-3' O 3' AGUUGUAGUCAGACUAUUCGAU 5' </pre>	<i>GSTM3</i> 3' UTR miR-21 (Human)	Centered site (pairing to miRNA nt 5-16)
E	<pre> 5' -CUCCUCAAGUAGUG CCUUG-3' O 3' CCGUAAGUGGCGCAC GGAAU 5' </pre>	<i>MINK1</i> 3' UTR miR-124 (Human)	Bulged site (1 nt bulge at miRNA nt 5-6)
F	<pre> 5' -GUGG UGCU-CUGGGUGAACCA-3' O 3' GACA AGGACGACUUGACUCGGU 5' </pre>	<i>E2F2</i> 3' UTR miR-24 (Human)	Not defined
G	<pre> 5' -GUCUGAUUCAG-AAGGGUCUA-3' O 3' UGUCCUAACUCCCGGGA 5' </pre>	<i>Nanog</i> ORF miR-296 (Mouse)	Not defined

FIGURE 21 : EXEMPLE D'INTERACTIONS MIRNA-CIBLE

Exemple d'interactions miRNA-cible avec le modèle d'appariement des séquences. La région seed est coloriée en violet. Les liaisons hydrogène sont indiquées par un trait vertical et les liaisons « wobble » (G :U) sont indiquées par un cercle. **A.** Appariement canonique avec la région seed **B. et C.** liaison seed imparfaite mais avec appariement compensatoire en 3' du miRNA **D.** Appariement central avec appariement incomplet de la région seed mais 12 appariements successifs dans la région centrale du miRNA **E.** Appariement complet de la région seed du miRNA mais avec un nucléotide de l'ARNm exclu et formant un renflement dans la structure. **F. et G.** modèles d'appariement validés expérimentalement mais aux structures non canoniques et difficile à généraliser. (D'après Lee et Shin 2012 (213)).

correspondant aux nucléotides 2 à 8 (168,187,196). De nombreuses configurations d'appariement existent (**cf. Figure 21**) et vont influencer l'activité du complexe RISC sur sa cible. Ainsi, un appariement sur de nombreux nucléotides en plus de la région seed (appariement extensif) favorisera un clivage de l'ARN cible (187,196). Ce clivage conduira comme pour les siRNA à la dégradation complète de la cible par l'action de l'exoribonucléase XRN1, et de l'exosome (168,199,200). Dans le cas d'un appariement réduit (non-extensif), le complexe RISC empêchera l'accès à la machinerie de traduction, induisant ainsi une répression traductionnelle et favorisant l'accès à certaines exonucléases (**cf. Figure 22**) (186,214).

Les miRNA sont retrouvés chez les animaux et les plantes. On compte plus de 48 860 miRNA décrits chez 271 eucaryotes dans la base de données miRBase(v22) (215), mais certains miRNA nécessitent encore d'être validés expérimentalement (216). Les miRNA assurent des rôles de PTGS par dégradation directe (activité slicer) ou indirecte (déstabilisation), ou encore par répression traductionnelle des ARN cibles (168,193,200). Ils sont également capables d'agir par TGS mais ce processus semble plus rare (193). Le ciblage des miRNA s'effectue principalement sur la région 3'UTR des ARNm. Une hypothèse pour ce ciblage serait une plus grande accessibilité des séquences. Cependant, il existe également des cas de ciblage des 5'UTR, des séquences introniques ou encore des séquences codantes des ARNm (CDS), mais ces cas restent eux aussi plutôt rares (196,214,217–219).

Les miRNA réguleront une grande partie de l'expression génique chez les mammifères et on estime qu'environ 60% des gènes humains seraient sous contrôle des miRNA (220). Ils seraient impliqués dans de nombreux processus physiopathologiques chez l'homme (165), et leur dérégulation en réponse à des stress, des infections et des pathologies en font de potentiels biomarqueurs d'intérêt médical (221,222). Les miRNA peuvent également être transférés entre cellules par des vésicules extra-cellulaires, et participer à des phénomènes de régulation notamment dans le contexte d'interactions hôte-parasite (207,210,223), ou être considérés comme des molécules hormone-like dans la physiologie des organismes (224).

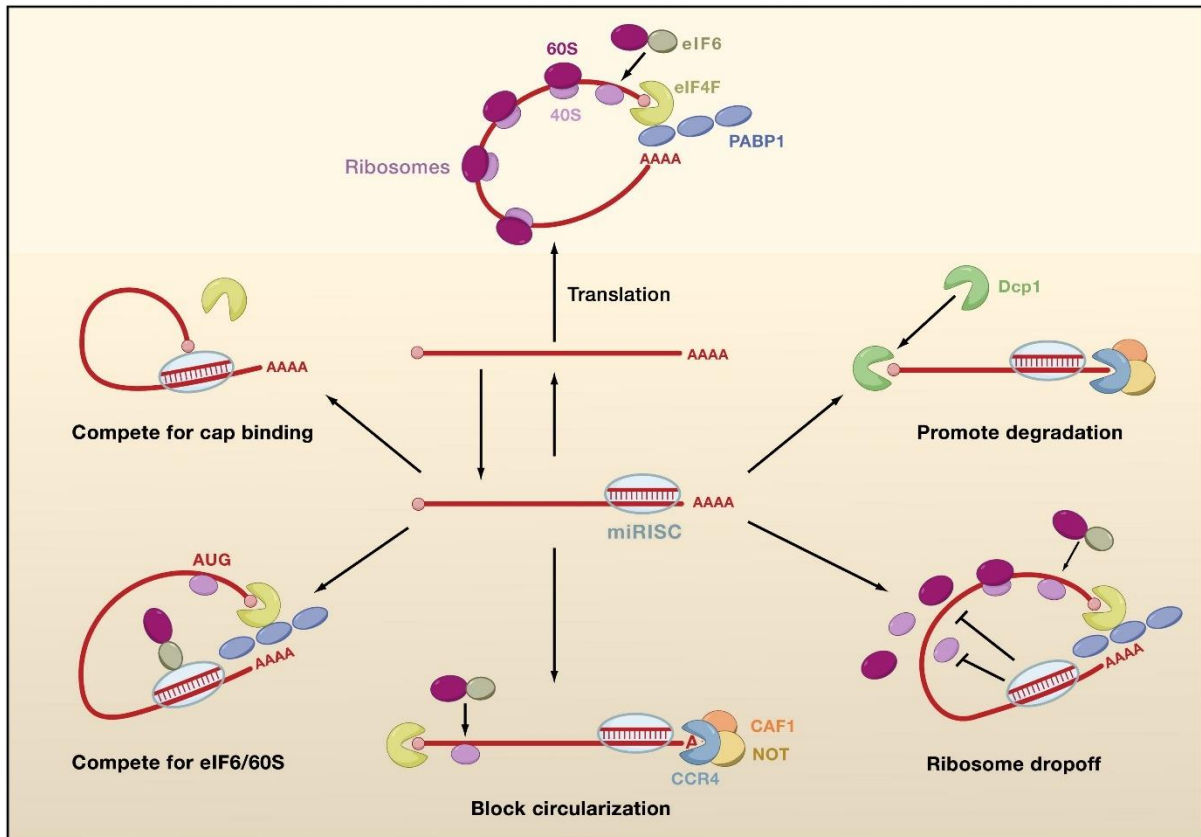


FIGURE 22 : MECANISMES DE REGULATION POST-TRANSCRIPTIONNELLE DES MIRNA

Différents mécanismes d'inhibition traductionnelle par le complexe miRISC. 60S = sous-unité 60S du ribosome; 40S = sous-unité 40S du ribosome, eIF6/eIF4F = elongation initiation factor 6/4F; PABP1 = PolyA binding protein 1; Dcp1 = Decapping enzyme 1; CCR4 = CCR4-Not complex 3'-5'-exoribonuclease subunit Ccr4; NOT = CCR4-Not complex 3'-5'-exoribonuclease subunit Not; CAF1 = Chromatin assembly factor 1. (D'après Carthew et Sontheimer 2009 (225))

III.B. ROLES DES MIRNA DANS LES INTERACTIONS HOTE-PARASITE

Chez l'homme les miRNA sont impliqués dans la régulation de processus immunitaires importants tel que la maturation des lymphocytes, l'activité des macrophages ou encore la régulation fine de processus inflammatoires par leur rôle dans les cellules dendritiques (226). De nombreuses dérégulations de miRNA ont été décrites en réponse à des infections notamment parasitaires (210). Chez l'Homme, l'infection par des protozoaires comme *T. cruzi*, *P. falciparum* ou *T. gondii* induit de nombreuses modulations de l'expression de miRNA qui peuvent participer à la pathogénie du parasite comme à la défense de l'organisme (210,227). Lors d'infections par *T. cruzi*, jusqu'à 133 miRNA ont pu être observés comme dérégulés (210). Quatre d'entre eux voyaient même leur expression corrélée directement avec des paramètres cliniques tels que la parasitémie ou l'intervalle QTc sur les électrocardiogrammes, lié à la fonction cardiaque (210,228). Similairement, au cours d'infections par *Leishmania donovani* de lymphocytes T chez l'Homme, une dérégulation de 208 miRNA a pu être observée (210). De manière intéressante, les miRNA qui étaient surexprimés en réponse à l'infection participaient à diminuer l'activation de processus liés à l'interféron gamma (IFN- γ) (210,229), connus comme importants dans le contrôle des infections à *Leishmania* (210,230).

Cette modulation de l'expression des miRNA au cours d'une infection peut aussi participer à la défense de l'hôte. Par exemple, au cours de l'infection du papillon *Bombyx mori* par la microsporidie *Nosema bombycis*, un des 17 miRNA modulés va réprimer une phosphatase de l'hôte avec pour conséquence une inhibition de la prolifération du parasite (231).

Les miRNA peuvent également être des médiateurs plus directs du dialogue hôte pathogène lorsque le miRNA d'un des deux acteurs cible spécifiquement des gènes de l'autre. Ainsi, les abeilles *Apis cerana* infectées par la microsporidie *V. ceranae* libèreraient des miRNA capables d'interagir et de réguler l'expression de gènes du parasite (232), notamment un gène codant pour une protéine de la paroi sporale de *V. ceranae*, et donc importante dans le développement et la survie du parasite (232). Dans le cadre d'une infection d'érythrocytes humains par *Plasmodium*, le complexe RISC de l'hôte lié à divers miRNA humains se retrouve transféré dans le parasite. Ce complexe RISC réprimera des gènes de *Plasmodium* (transporteurs, gènes de réplication,



gènes de traduction) et diminuera ainsi la virulence et la prolifération du parasite (210,227,233–235).

Réciproquement, il a été montré que certains parasites exprimaient eux aussi des miRNA ciblant des gènes de l'hôte : certains des miRNA exprimés par *T. gondii* (environ 1%) sont des homologues de miRNA trouvés chez l'Homme ou le rongeur et pourraient donc agir sur les mêmes cibles chez l'hôte (236,237). Ils moduleraient de nombreux processus cellulaires, dont l'apoptose (210,237).

Des petits ARNnc « miRNA-like » ont été identifiés chez les microsporidies *V. ceranae* (232,238–240) et *N. bombycis* (241). Chez *V. ceranae* ils ont été prédits comme ciblant des gènes associés principalement au métabolisme cellulaire. Cependant, des gènes associés à des voies de signalisation, au fonctionnement des lysosomes, aux processus d'autophagie, à l'endocytose et à la défense immunitaire seraient également ciblés (232,238–240). Chez *N. bombycis*, les auteurs ont pu démontrer une répression du gène BmPEX16 du peroxyosome de l'hôte par le microRNA-like 8 de *N. bombycis* (Nb-milR8). Ils ont également montré que cette répression de BmPEX16 favorise la prolifération du parasite dans son hôte (241).

Certains petits ARNnc comme les snoRNA les snRNA les rRNA et les tRNA, initialement caractérisés pour leurs rôles structuraux ont récemment été décrits comme étant impliqués dans des phénomènes d'interférence par l'ARN. Les tRNA, rRNA et snoRNA donnent par exemple naissance aux « tRNA-derived small RNA » (tsRNA), aux « rRNA fragments » (rRFs), et aux « snoRNA-derived RNA » (sdRNAs) respectivement. Ces ncRNA peuvent être apparentés aux miRNA en raison de leurs processus de biogénèse (**cf. figure 19**) (210). Le rôle précis de ces nouvelles classes d'ARN interférents reste à définir mais de nombreuses études sont actuellement menées (242–245). Chez *T. cruzi* par exemple, de petits ARNnc supposés dériver des tRNA du parasite ont été décrits comme capables de réguler des gènes de cellules Hela *in vitro* (227,246,247). Similairement, des petits ARNnc (non caractérisés dans l'étude) produits par *Leishmania spp.* ont montré un effet de régulation épigénétique et post-transcriptionnelle sur des macrophages (227,248). Ces ARNnc pourraient en fait être des tsRNA et rRFs, connus comme étant très abondants dans les vésicules de *Leishmania* (249). Des processus d'inhibition de la maturation ou de l'activité des miRNA ont été décrits dans

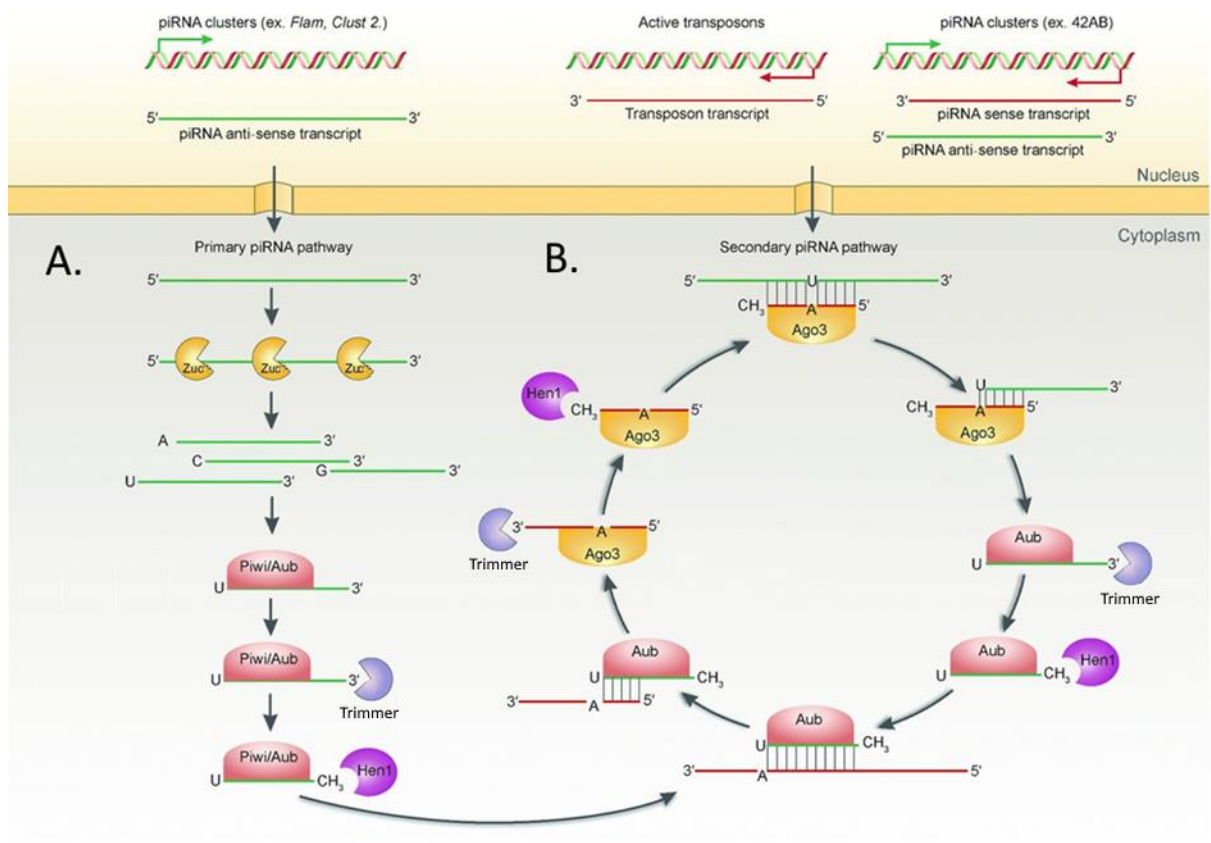


FIGURE 23 : BIOGENESE DES piRNA ET CYCLE PING-PONG

Schéma de la biogénèse des piRNA avec en **A.** la voie primaire et en **B.** la voie secondaire avec le cycle d'amplification ping-pong. (Adapté de Hsueh-Yen et KuHaifan Lin 2014 (250))

plusieurs infections, soulignant encore d'avantage l'enjeu central que peuvent représenter ces molécules dans les interactions entre hôtes et parasites (210).

IV. LES piRNA

IV.A. BIOGENESE ET MODE D'ACTION

Les piRNA (PIWI associated RNA) sont de petits ARNnc nommés ainsi car ils s'associent à une sous famille précise d'Argonautes les Piwi. Les piRNA sont exclusifs au règne animal et présentent un processus de maturation indépendant de la protéine Dicer et donc très différent de ceux des siRNA et des miRNA (**cf. figure 23**) (168,182,251). Les piRNA (168,182) sont d'une taille allant de 24 à 31 nt selon les espèces (les piRNA de *Caenorhabditis elegans* font exception avec 21 nt) (193).

Les piRNA sont issus de de copies isolées de séquences répétées, de gènes piRNA isolés, ou de loci génomiques appelés piRNA clusters (188). La transcription des clusters de piRNA donne un précurseur simple brin ne présentant pas de structure secondaire particulière contrairement aux siRNA et miRNA. Les processus poussant la cellule à utiliser préférentiellement certains transcrits pour la production de piRNA (licencing) demeurent encore mal connus. Le transcrit précurseur sera exporté vers le cytoplasme. La maturation des piRNA a lieu dans un compartiment subcellulaire périnucléaire appelé nuage dans les cellules germinales animale ou dans les Yb-bodies dans les cellules somatiques ovariennes de *Drosophile* (188,252). Les transcrits primaires de piRNA y seront maturés par l'endonucléase Zucchini (Zuc) (**cf. Figure 23 A**) (168,188). Zuc clive le précurseur en portions phasées d'environ 25nt chez la drosophile. *In vivo*, les piRNA issus de la maturation par Zuc semblent présenter un biais pour un résidu d'uracile en position 1. Ce biais n'est pas retrouvé lors d'expériences *in vitro* avec une protéine Zuc recombinante suggérant la présence d'autres facteurs influençant l'affinité et l'activité de Zuc *in vivo*. Les piRNA produits par Zuc sont ensuite pris en charge par une protéine de la sous-famille Piwi dans un complexe RISC (piRISC), l'extrémité 3' du piRNA sera raccourcie par une enzyme appelée Trimmer et sera 2'-O-méthylée par Hen1 (253). Ces étapes de maturation semblent conférer une certaine protection aux piRNA vis-à-vis d'autres nucléases (168,188,253).



Dans les cellules germinales, la quantité de piRNA produite par transcription peut être amplifiée par un processus appelé « ping-pong » (cf. **Figure 23B**). Au cours de ce cycle, des ARN messager d'éléments transposables clivés par le piRISC vont eux même être maturés en piRNA. Chez la Drosophile, un complexe piRISC contenant la protéine Aubergine (Aub) reconnaît et clive un ARN cible (clivage en vis-à-vis des nucléotides 10 et 11 du piRNA comme avec les miRNA et siRNA). Le transcrit clivé sera pris en charge par une protéine AGO3 et mûré en piRNA par une enzyme trimmer. La protéine AGO3, ainsi associée à un piRNA complémentaire au transcrit primaire initial, pourra le cliver. Ce clivage permettra la formation de nouveau piRNA, qui pourront être pris en charge par Aub, complétant ainsi le cycle ping-pong (Cf. **Figure 23B**) (188,198). Ce modèle d'amplification des piRNA permettrait d'expliquer la complémentarité sur 10 nt des piRNA liés à Aub et AGO3 et pourquoi ils présentent respectivement un biais pour un uracile et une adénine en position 1 et 10 du piRNA (caractéristique appelée signature ping-pong) (168,182).

Les complexes piRISC peuvent agir par PTGS ou TGS en fonction de la protéine Argonaute qu'ils contiennent (188). Chez la drosophile, les piRNA liés à Aub et AGO3 (protéines Miwi et Mili respectivement chez la souris), vont participer à une répression post-transcriptionnelle via une activité slicer. Tandis que ceux liés à la protéine Piwi de la drosophile (Miwi 2 chez la souris) vont plutôt induire une translocation vers le noyau du complexe piRISC et une répression transcriptionnelle de leur cible par formation d'hétérochromatine (188).

Les piRNA sont principalement actifs dans les lignées germinales et ont pour action principale la répression des ET (188,193). Les piRNA clusters sont composés de séquences dérivés d'ET agrégées les unes à côté des autres. Ces clusters constitueraient ainsi une sorte de mémoire génétique d'invasions passées de transposons (188,254,255). Les piRNA clusters peuvent être transcrits à partir d'un seul brin (le plus souvent) ou des deux brins (188). Chez la plupart des organismes étudiés, un nombre défini de locus est responsable de la production de la vaste majorité des piRNA de l'organisme (plus de 90% des piRNA seraient issus de ces clusters) (188,256–258).



IV.B.LES piRNA ET LA REGULATION DES ELEMENTS TRANSPOSABLES

On estime qu'environ 44% du génome humain correspond à des ET (259) et environ 20% du génome de *D. melanogaster* (260). Le système des piRNA peut être vu comme un système immunitaire génétique : lorsqu'un ET actif s'intègre dans un piRNA cluster, il permet la production de piRNA le ciblant et pouvant le réguler. Le cycle d'amplification ping-pong permet également de faire face à de fortes activations d'ET, en fournissant une quantité suffisante de piRNA pour les réguler (188).

Dans la lignée germinale et au cours de l'embryogénèse par exemple, des événements de dérépression épigénétique ont lieu à grande échelle conduisant à une plus grande activité des ET (261). L'activation ectopique d'ET lors de la gamétogénèse chez la drosophile a été associée à de la stérilité dans de nombreux cas. Cette stérilité surviendrait suite à une accumulation de mutations dans le génome des cellules germinales. Les piRNA par leur régulation des ET assureraient donc un rôle central dans la reproduction et la survie des organismes (168,182,251).

Chez les mammifères on subdivise les piRNA en piRNA pré-pachytène et piRNA pachytène en fonction du stade de la méiose où ils sont exprimés dans les spermatocytes (168,188,193). Les piRNA pré-pachytènes comme les piRNA chez la drosophile correspondent principalement à des séquences répétées et sont impliqués dans la répression des ET (transposons à ADN (Class I) et rétrotransposons (Class II)) (168,255). Les piRNA pachytènes sont principalement transcrits depuis des régions non condensées du génome (euchromatine) et présentent des caractéristiques communes avec les ARNm classiques (*e.g.* polyadénylation) (168,188). Leurs rôles restent moins bien caractérisés (168,188).

Une expression de piRNA dans les cellules somatiques de certaines espèces a également été décrite, ainsi que des rôles de défense antivirale chez les moustiques, des rôles de régulation de gènes, ou encore des rôle dans la détermination du sexe chez le vers à soie (188,251,262).

Chez les ciliés comme *Paramecium* et *Tetrahymena*, des ARNnc apparentés aux piRNA sont impliqués dans des phénomènes d'élimination d'ADN, durant lesquels des séquences entières d'ET peuvent être excisées du génome au cours de la maturation du zygote (263).

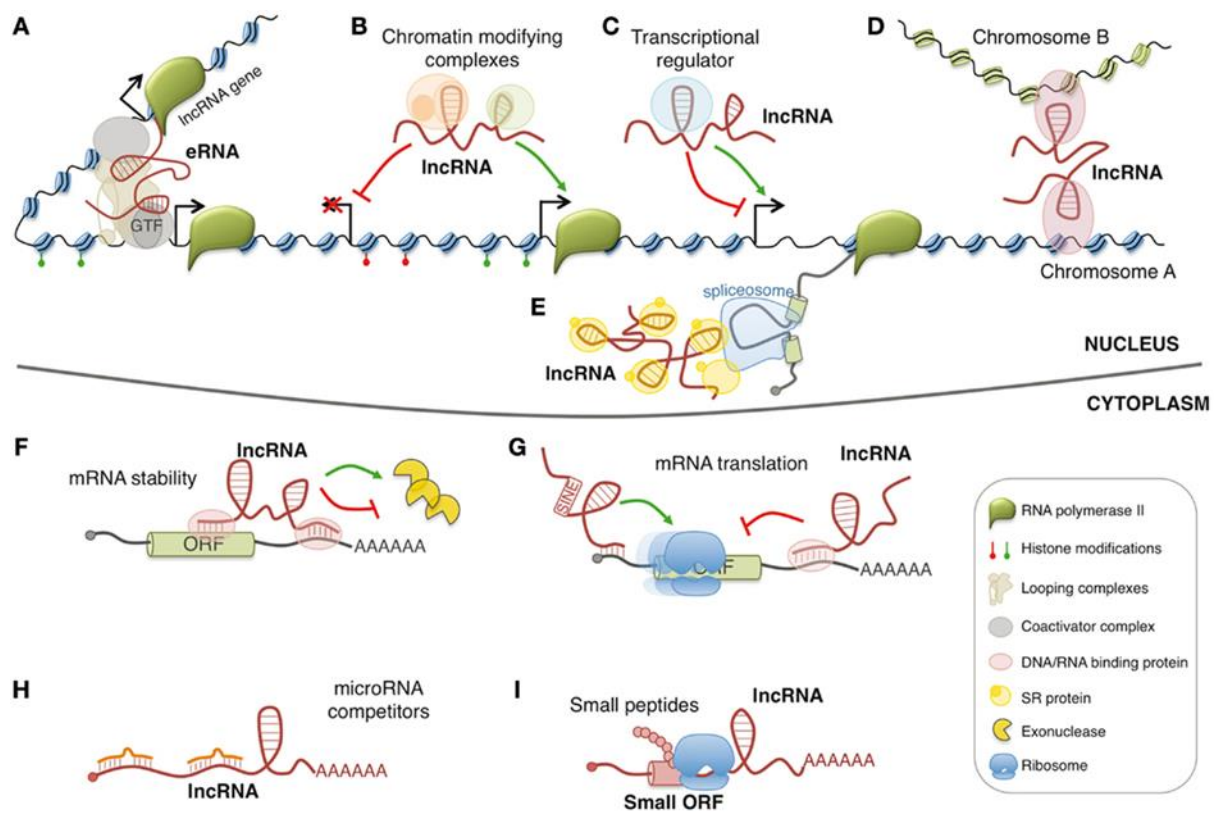


FIGURE 24 : FONCTIONS DES LONGS ARN NON-CODANTS

Les Longs ARN non-codants (LncRNA) peuvent (A) agir comme amplificateurs de l'expression de gènes (enhancer RNAs, eRNA), (B) comme recruteurs de complexes modifiant la chromatine, (C) comme régulateurs de l'activité de facteurs de transcription, (D) comme régulateurs de l'expression génique en agissant sur la conformation spatiale des chromosomes, (E) en influençant l'épissage des pre-mRNA. Les LncRNA cytoplasmiques peuvent (F) aussi réguler la stabilité ou (G) la traduction des mRNA et ainsi moduler leurs expressions ou encore (H) entrer en compétition en tant que cibles de miRN. (I) les LncRNA peuvent enfin contenir de petits cadres de lecture ouverts (Open Reading Frames ORFs), codant des peptides. (Adapté de Morlando et al 2015 (264)).

V. LES LONG ARNS NON-CODANTS

V.A. TYPES DE LNCRNA

Le terme Long ARNs non-codants regroupe l'ensemble des ARNnc de plus de 200nt et de nombreuses fonctions leurs sont associés (**cf. figure 24**) (171). Ils sont moins bien caractérisés que les sncRNA et semblent également plus hétérogènes. On les classifie principalement selon la localisation génomique de leurs gènes en distinguant les LincRNA (Long intergenic non coding RNA) dans les régions intergéniques, les LncRNA antisens, localisés sur l'autre brin des gènes codant pour les protéines, et les eRNA (enhancer RNA) et promotor upstream transcrits, issus de régions promotrices et de régions en amont de promoteurs respectivement (171).

Les lincRNA possèdent des rôles très diversifiés mais semblent souvent impliquer dans une régulation épigénétique de loci précis. Un des premiers à avoir vu son rôle être mis en évidence est XIST, qui est impliqué dans l'hétérochromatinisation et l'inactivation d'un des chromosomes X au cours de l'embryogénèse chez les mammifères femelles. XIST reconnaît certaines séquences du chromosome et recrute divers complexes de manipulation de la chromatine (265,266). Depuis XIST, de nombreux autres LincRNA ont vu leur fonction biologique élucidée (171,265).

Les promotor upstream *transcrits* et les enhancer LncRNA (eLncRNA), sont des ARN respectivement issus de la transcription de régions en amont des promoteurs et de régions promotrices (171,173). Ces transcrits ont été corrélés à une plus forte expression de certains ARNm (171,173). Cependant, leurs rôles et mécanismes d'action potentiels restent à être clarifiés (171,173).

Les *LncRNA* antisense (ou natural antisens transcripts, NAT) résultent de la transcription du brin d'ADN opposé à un ARNm donné. Ces transcrits et l'ARNm correspondant auront tendance à s'apparier et le duplex ainsi formé pourra ensuite être dégradé et donner naissance à des siRNA. Certains gènes sont connus pour être régulés selon ce mode d'action comme le récepteur à l'insuline-like growth factor 2 chez l'homme (IGF2R), régulé par le LncRNA « antisense IGF2R non-protein-coding RNA ». Le mécanisme moléculaire associé



à l'action de ces LncRNA n'est pas encore complètement élucidé (171).

V.B. LES LNCRNA DANS LES INTERACTIONS HÔTE-PARASITE

Certains LncRNA ont été décrits comme acteurs des interactions hôte-parasite. Lors d'une infection par *Cryptosporidium parvum*, les cellules hôtes voient leurs LncRNA dérégulés. Certains de ces LncRNA joueraient des rôles importants dans l'immunité contre le parasite. Ainsi, le LncRNA NR_045064 exprimé dans les cellules épithéliales intestinales suite à une infection par *C. parvum* a été montré comme stimulant l'expression de gènes de défense dans ces cellules (267).

Les LncRNA de *C. parvum* sont eux aussi importants dans ces interactions entre l'hôte et le parasite (268,269). Il a ainsi été décrit qu'au cours d'une infection, le LncRNA Cdg7_FLc_0990 de *C. parvum* était transloqué dans le noyau de la cellule parasitée avec l'aide d'une protéine chaperonne de l'hôte (HSP70). Cdg7_FLc_0990 induit ensuite une répression transcriptionnelle de certains gènes en recrutant des histones méthyltransférases dans leurs régions promotrices (269,270). Un autre LncRNA de *C. parvum* (Cdg7_FLc_1000) aurait lui aussi un rôle de répression, ciblant principalement des gènes impliqués dans la migration et l'adhérence des cellules hôtes, atténuant ainsi leur motilité (269–271). Les mécanismes de ciblage par ces LncRNA seraient médiés par la formation de triplexes ARN ADN (269,270).

Chez *N.bombycis*, des LncRNA ont été identifiés et suggérés comme participant à un réseau de régulations qui ciblerait des gènes de défense et d'apoptose de l'hôte (272).

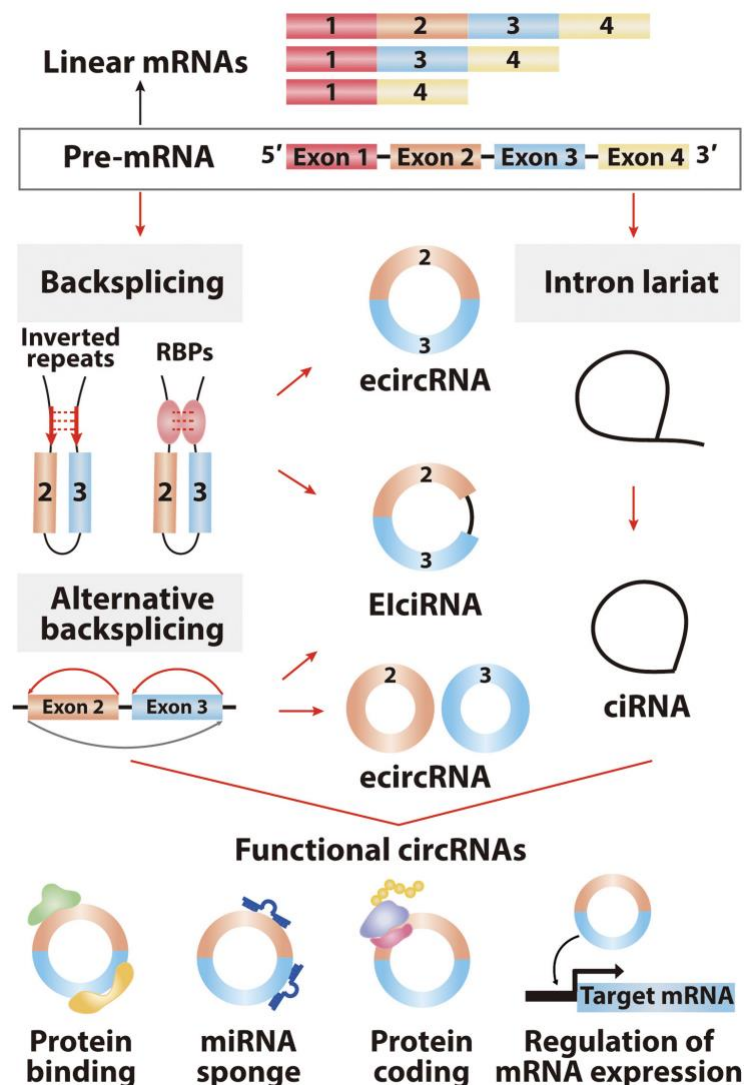


FIGURE 25 : BIOGENESE ET FONCTIONS DES CIRCRNA

Biogénèse et fonction des circRNA. Les circRNA sont issus d'un phénomène de backsplicing des pré-mRNA. Le backsplicing consiste en un site d'épissage donneur qui se lie à un site accepteur en amont de celui-ci. En résulte un ARN simple brin circularisé par une liaison covalente. Les circRNA peuvent également être produits à partir des lassos d'épissage. Les circRNA peuvent être liés à des protéines, jouer un rôle d'éponge à miRNA, être traduits en protéines, ou encore réguler l'expression de certains gènes en se liant aux promoteurs. (D'après Kim *et al* 2021 (273))

VI. LES CIRC RNA

VI.A. BIOGENESE DES CIRC RNA

Les circRNA sont des ARNnc circulaires produits à la suite d'un phénomène de rétro-épissage (back-splicing) de pré-ARNm aboutissant à une liaison covalente entre les deux extrémités du transcrit (cf. **figure 25**). Ils peuvent également être issus des lassos d'épissage, ou des introns libérés lors de la maturation des tRNA (273–276). On distingue trois types de circRNA : les circRNA composés d'exons (exonic circRNA ou ecircRNA), ceux composés d'exons et d'introns (exon-intron circRNA ou EicirRNA) et les circRNA composés uniquement d'introns (intronic circRNA ou ciRNA) (cf. **figure 25**) (273,274).

Les circRNA sont apparus comme des isoformes principaux d'une centaine de gènes lors de plusieurs études s'intéressant aux transcriptomes de divers types cellulaires humains (277). Plusieurs rôles leur sont attribués (274,278), notamment dans la régulation de processus immunitaires (279,280). Le premier mode d'action décrit serait celui d'éponge à miRNA ou à protéines liant l'ARN, permettant dans certaines conditions physiologiques de réduire l'impact de ces molécules (273,274). Ces ARNs peuvent aussi être traduits, ou réguler l'expression d'ARNm en se liant à leurs promoteurs (cf. **figure 25**) (273,274).

VI.B. LES CIRC RNA DANS LES INTERACTIONS HÔTE-PARASITE

Les circRNA ont été décrits ou prédits comme ayant des rôles lors des interactions hôte-parasite. Dans le cadre d'infections de différentes espèces d'abeilles par la microsporidie *V. ceranae*, l'étude de circRNA a révélé chez l'abeille une expression différentielle de plus de 168 circRNA en réponse à l'infection 7 jours après inoculation, et de 306 à 10 jours. Certains de ces circRNA seraient associés à des processus immunitaires (281). Des circRNA ont également été prédits chez *V. ceranae* (282), et chez une autre espèce microsporidienne, *N. bombycis* (parasite du papillon *B.mori*). Ils réguleraient les miRNA du parasite mais également de son hôte (272). Enfin, des circRNA produits par le vers parasitaire *Ascaris suum* et libérés via des vésicules extracellulaires exerceraient une régulation de miRNA impliqués dans l'immunité cellulaire et l'inflammation chez son hôte humain (283).



CHAPITRE 2 : DIVERSITE EVOLUTIVE DES ARNnc ET DES PROCESSUS D'ARN INTERFERENCE

Bien que de nombreux aspects liés aux ARNnc soient conservés à travers le vivant, l'étude de certains organismes modèles a révélé une grande diversité posant la question de la place de l'ARNi dans l'évolution des organismes.

I. PERTE DE L'ARNi : L'EXCEPTION KILLER CHEZ SACCHAROMYCES CEREVISIAE ; LES CENTROMERES DE USTILAGO MAYDIS ; C.GATTI ; T.CRUZI ; ET LES ENCEPHALITOOZON

Aujourd'hui le mécanisme d'ARN interférence a été étudié chez de nombreux organismes et semble être répandu et conservé (284). Il aurait été présent chez le dernier ancêtre commun des eucaryotes (285) et est ainsi largement retrouvé chez les mammifères, les insectes mais aussi chez les champignons (286–288). Cependant, certaines espèces se démarquent, comme la levure *Saccharomyces cerevisiae*, le pathogène de plante *Ustilago maydis*, le sous type VGII de *Cryptococcus gattii*, *T. cruzi*, ou encore les microsporidies du genre *Encephalitozoon*, qui ont toutes perdu le mécanisme d'ARNi (**cf.figure 26**) (284). Ces éléments soulèvent de nombreuses questions concernant la biologie de ces organismes, l'ARNi étant impliqué dans la défense antivirale (185), la régulation de gènes (168,193), la formation des centromères, ou encore la régulation des ET (193,289) chez certaines levures et champignons comme *S. pombe* ou *Cryptococcus* (193,290). La question qui demeure est : comment la perte de l'ARNi a pu constituer un avantage sélectif ou être sans effets délétères pour ces organismes.

Chez *S. cerevisiae*, la perte de l'ARNi serait associée à l'apparition d'un virus à ARN double brin nommé « killer » dans les populations de *S. cerevisiae*. Killer peut être transmis verticalement entre cellules mères et filles, ou horizontalement par infection du zygote lors de la reproduction. Ce virus induit un phénotype où les cellules infectées inhibent la croissance des cellules non infectées aux alentours (284,291,292). Ainsi, l'infection par le virus conférerait un avantage sélectif dans un environnement compétitif où l'activité antivirale de l'ARNi (vérifiée comme efficace contre killer (291)) se présenterait alors comme un

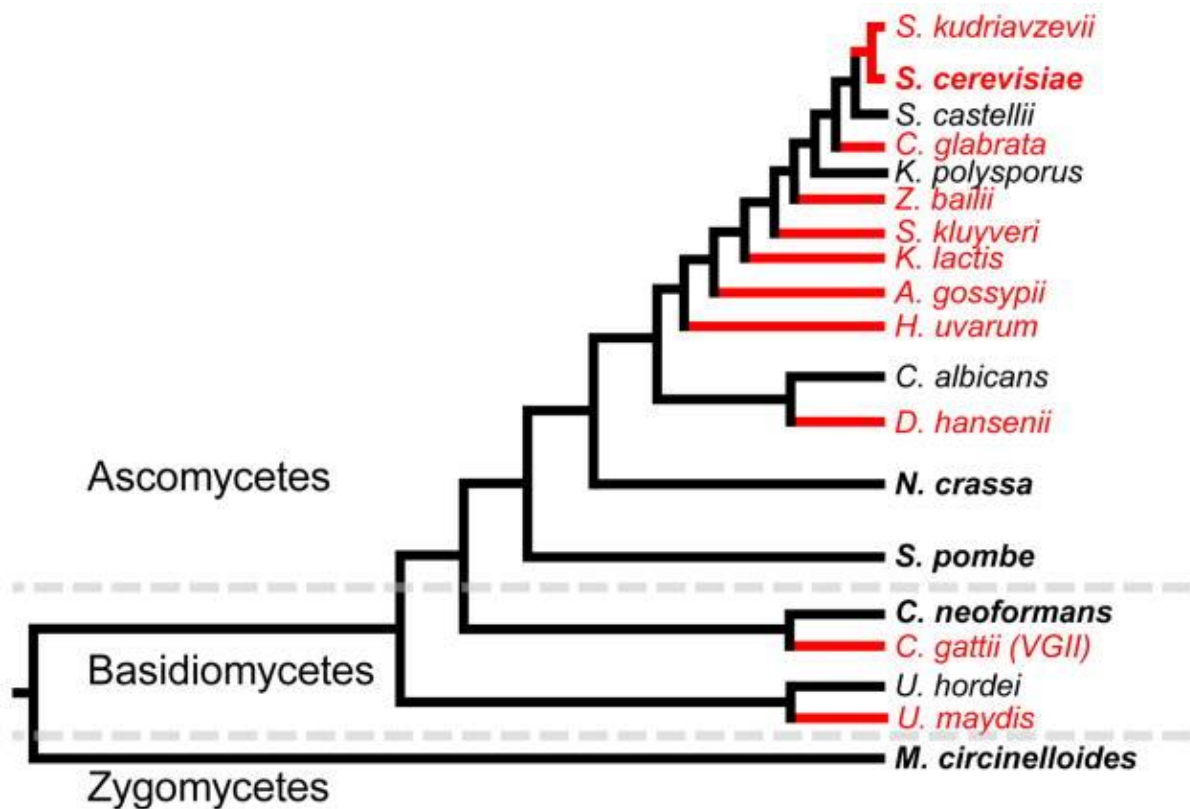


FIGURE 26 : ARBRE PHYLOGENETIQUE DE 19 CHAMPIGNONS ET DE LA PRESENCE D'ARNi CHEZ EUX

Arbre phylogénétique d'espèces de champignons visant à illustrer la perte d'ARNi dans certains groupes. Les espèces dont les noms sont en rouge ne sont pas capables d'ARNi contrairement à celles indiquées en noir. Les espèces qui ont fait l'objet de recherches poussées sur l'ARNi sont en gras (284). (Adapté de Billmyre *et al*, 2013 (284)).

désavantage. Ceci aurait conduit à la disparition de l'ARNi de plusieurs lignées d'organismes chez les champignons (291). Chez *S. cerevisiae*, l'ARNi peut être rétabli par simple transfert de protéines Dicer et Argonaute issues d'ascomycètes proches, indiquant la perte d'un nombre restreint de protéines de la voie (284,293).

Dans le cas du sous-type *VGII* de *C. gattii*, la divergence avec le sous-type *VGI* dont les espèces possèdent encore l'ARNi semble s'être faite il y a environ 12.4 millions d'années (284,294). La perte de l'ARNi aurait ici favorisé l'activité des ET conduisant à un phénotype mutagène (ou Mutator) favorisant l'adaptabilité de l'organisme à des environnements hostiles (284). *C.gattii* est considéré comme tropical et subtropical, mais il a été responsable d'une épidémie dans le nord-ouest américain dans les années 2010 (295). Ce profile Mutator l'aurait aidé à s'adapter à ce nouvel environnement, initialement hostile pour lui. Un parallèle est fait entre ce mécanisme d'adaptation et la réponse SOS chez les procaryotes, où un phénotype mutagène est activé transitoirement en réponse à un stress intense (284,296,297). Si chez les bactéries ce processus est réversible (284,296,297), chez *C. gattii* *VGII* la perte des gènes de l'ARNi ne l'a pas été.

L'ARNi a également été montrée comme impliquée dans la formation et le maintien des centromères chez certains champignons (290). Une étude comparative de la taille des centromères entre des espèces de champignons capables ou non d'ARNi (critère basé sur la présence ou non de protéines d'ARNi (290)), a permis d'observer une corrélation entre la perte d'ARNi et la réduction de la taille des centromères. Ainsi, *Ustilago maydis*, une espèce ne possédant pas les protéines principales d'ARNi présentait des centromères en moyenne 2 à 3 fois plus petits que ceux d'autres espèces du même genre, et capables d'ARNi (12 Kb pour *U. maydis* contre 36 Kb et 27 Kb pour *Ustilago hordei* et *Ustilago bromivora*)(290).

A l'instar d'autres groupes d'être vivants, les trypanosomatidés possèdent un processus d'ARNi, qui a cependant divergé ou été perdu à différent points de leur évolution dans différentes lignées (298,299). La présence et le fonctionnement d'une machinerie ARNi chez ces parasites ont été testés expérimentalement par Lye *et al* qui ont introduit des ARN doubles brins dans le parasite pour en étudier les effets sur des gènes rapporteurs. Ils ont

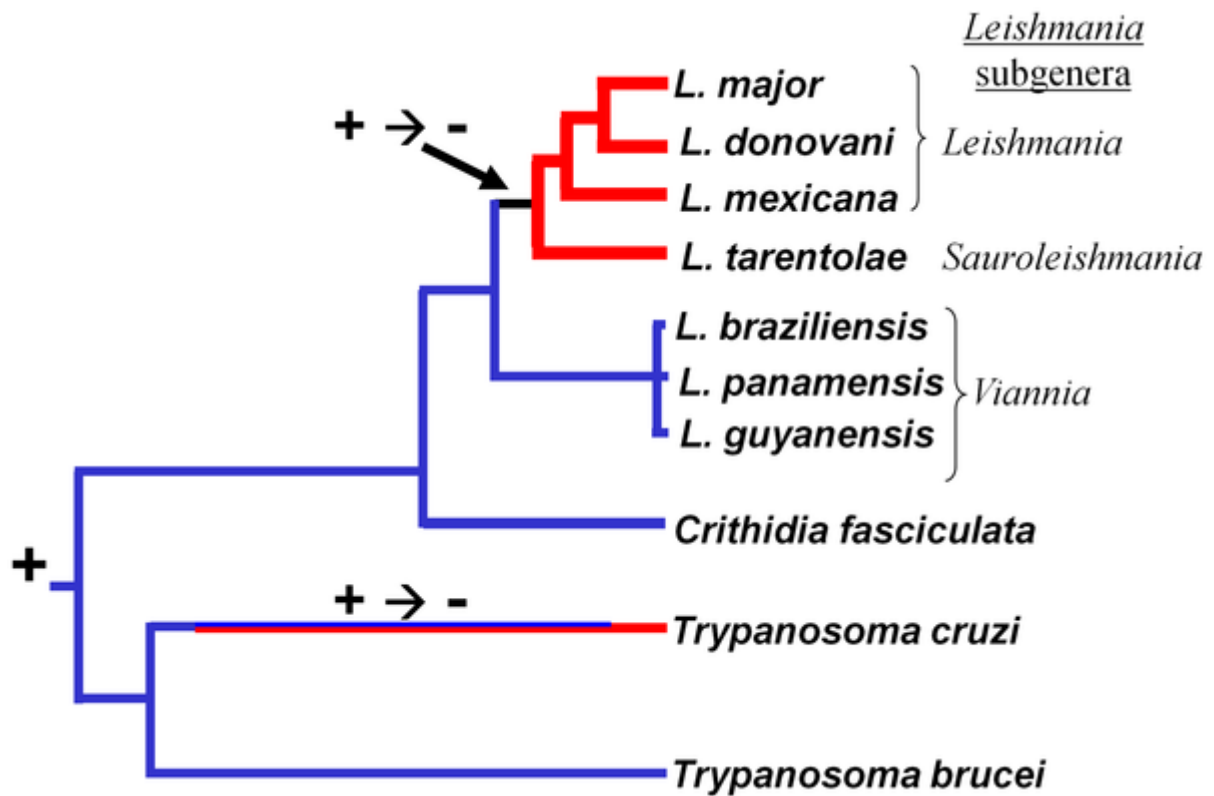


FIGURE 27 : CONSERVATION ET PERTE DE L'ARNi CHEZ LES TRYPANOSOMATIDES

Arbre évolutif consensus des trypanosomatidés. La longueur des branches correspond au degré de divergence évolutive entre les organismes. Les lignées ne présentant pas d'activité ARNi et/ou les gènes associées à l'ARNi sont marquées aux extrémités avec les branches colorées en rouge et celles présentant l'activité ARNi en bleu. La double coloration le long de la branche de *T. cruzi* indique qu'il n'a pas été possible de définir à quel moment de son évolution cette perte a eu lieu. Les symboles + ou - indiquent une présence ou perte de l'ARNi au cours de l'évolution. (Adapté de Lye *et al*, 2010 (298)).

également recherché la présence de protéines tel que Dicer ou Argonaute (298). Les auteurs concluent que malgré la présence d'une machinerie ARNi chez la plupart des trypanosomatidés, elle a été perdue chez une partie des *Leishmania* et chez *T. cruzi* lors d'une bifurcation évolutive (cf. **figure 27**) (298,299). Les trypanomatidés présentent des protéines Argonaute phylogénétiquement à part, définissant leur propre sous-famille, les Trypanosoma-Argonaute (299). Il en existe deux types : les AGO-tryp et les PIWI-tryp (299). De manière intéressante, les deux types ne sont pas systématiquement retrouvés chez toutes les espèces de trypanosomatidés. Les AGO-tryp semblent co-sélectionnées avec la protéine Dicer et semblent associées à l'ARNi en réponse aux ARN double brin exogènes (298–300). Les AGO-tryp ont également été associées à des ARN de 24 à 26 nt dérivés de rétrotransposons et sont constitutivement exprimées chez *T. brucei*, suggérant leurs implications dans la répression d'ET (301). Des expériences de délétion d'AGO-tryp chez *T. brucei* ont montré une ségrégation aberrante des chromosomes et une viabilité réduite, suggérant aussi une association avec les séquences répétées des télomères (290,302). Les PIWI-tryp quant à eux se distinguent par l'absence d'un domaine PAZ, mais la présence d'un repliement protéique qui permettrait tout de même la liaison d'éléments oligonucléotidiques (299,303). De manière surprenante, bien que ne possédant pas d'AGO-tryp, *T. cruzi* et *L. major* possèdent une protéine PIWI-tryp (303). Le rôle des PIWI-tryp et le type d'ARN qui leurs sont associés ne sont pas connus mais le fait que cette protéine soit exprimée à tous les stades de développement de *T. cruzi* suggère un rôle tout au long du cycle de vie du parasite (303). L'expérience de double Knockout de AGO-tryp chez *T. brucei* a montré l'absence d'un rôle compensatoire de la protéine PIWI-tryp en ce qui concerne l'ARNi (302).

Les microsporidies sont un autre groupe où l'on peut observer des disparités dans la conservation de l'ARNi (119,304) (cf. **figure 28**). Cette perte semble s'être produite à plusieurs embranchements de leur phylogénie. Huang (304) a évalué la présence ou l'absence des gènes codant les protéines Argonautes et Dicer dans le génome de 21 microsporidies. Seuls deux cas de figure ont été identifiés : l'absence ou la présence des deux gènes. Ces résultats iraient dans le sens d'une conservation liée à l'ARNi, processus dans lequel elles sont impliquées conjointement (168,193). Notamment, la présence ou l'absence d'ET dans les

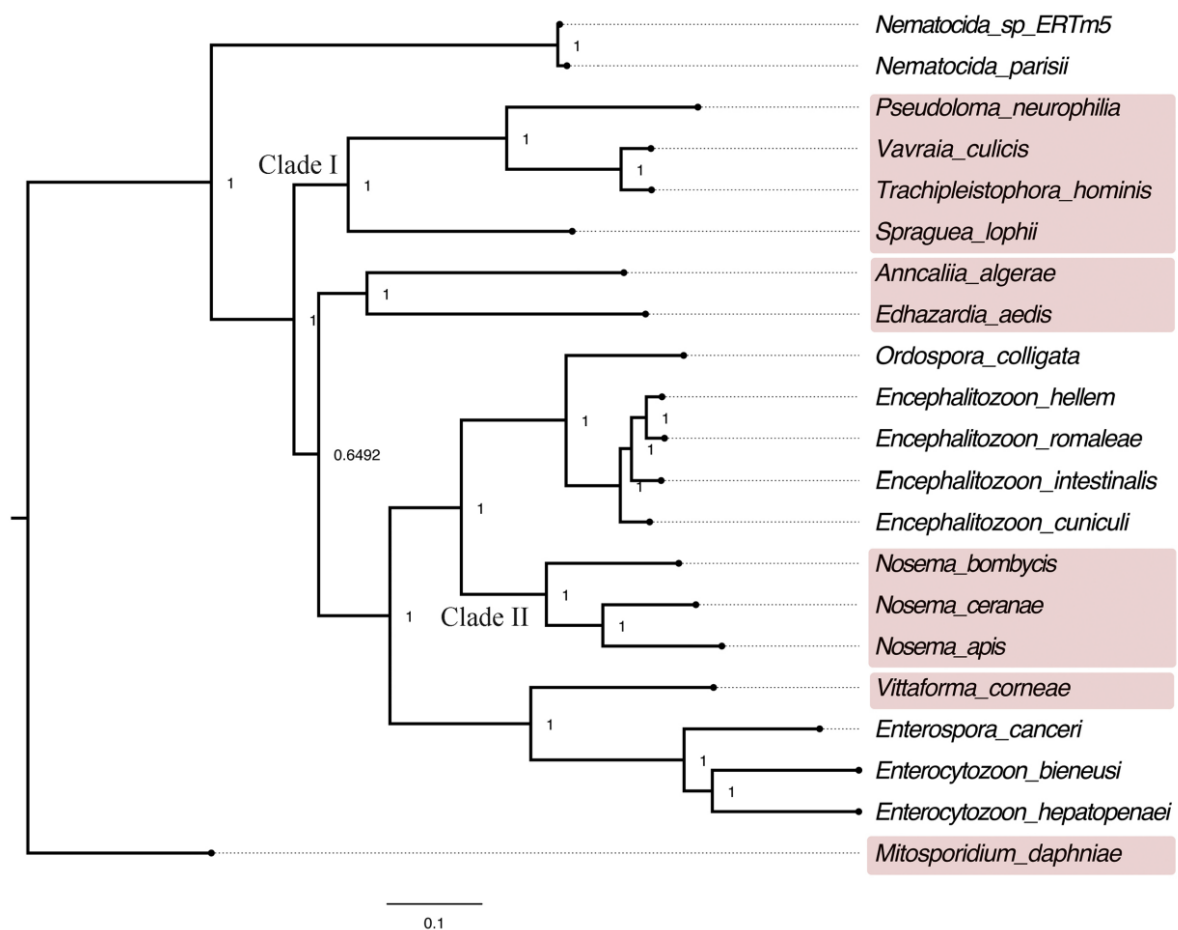


FIGURE 28 : ARBRE PHYLOGENETIQUE DE 21 MICROSPORIDIES ET PRESENCE OU ABSENCE DU COUPLE DICER/ARGONAUTE

Arbre phylogénétique de 21 microsporidies basé sur les séquences de 18 gènes orthologues en copie unique présents dans tous les génomes. Les 11 espèces présentant des orthologues de Dicer et Argonaute sont surlignées. L'arbre a été résolu suivant la crédibilité maximum de clad (Maximum clad credibility, MCC) avec l'outil Mrbayes (305). Les probabilités Bayésiennes postérieures sont indiquées aux nœuds. L'espèce *Mitosporidium daphniae* a été utilisée en tant que groupe externe pour enracer l'arbre. (Adapté de Huang, 2018 (304)).

génomés semblerait jouer un rôle dans la conservation de Dicer et Argonaute car seules les espèces avec ces éléments dans leur génome ont conservé la machinerie d'ARNi (136,138).

II. LES ARNnc DES PARASITES : APICOMPLEXES ET TRYPANOSOMATIDES

Les ARNnc assurent des rôles importants dans la biologie des parasites (269). Alors que *P. falciparum* ne semble pas encoder, ni de siRNA, ni de miRNA, ni de piRNA (269) ; des LncRNA joueraient d'importants rôles dans la régulation de l'expression de ses gènes, particulièrement de ses gènes *var*, impliqués dans sa virulence et dans l'échappement immunitaire chez l'Homme (40,269,306). Les gènes *var* ont une expression mutuellement exclusive. Ils sont composés d'un exon 1 variable et d'un exon 2 conservé, séparés par un intron (40,269). Cet intron contient un promoteur bi-directionnel qui permettrait l'expression de 2 LncRNA. Le premier LncRNA est antisens et complémentaire de l'exon 1 tandis que le second est sur le brin sens et s'étend dans le deuxième exon (307). Ces LncRNA seraient transcrits par l'ARN pol II (308), coiffés d'une 7'-methyl-G mais non polyadénylés (269,307,309). Ils seraient impliqués dans l'activation ou la répression du gène *var* qui les contient (**cf. figure 29**). Des expériences d'expression ectopique du LncRNA antisens ont montré une activation des gènes *var* correspondants, en parallèle de celui naturellement exprimé par le parasite, outre-passant le mécanisme assurant l'expression mutuellement exclusive des gènes *var* (269,309,310).

Similairement, des ARNnc de 136 nt issus de régions riches en GC situées entre les clusters de gène *var* internes aux chromosomes ont montré des rôles de régulation sur l'expression des gènes *var*. (311) La surexpression de plusieurs de ces ncRNA induit l'activation de différents groupes de gènes *var* (311). Réciproquement, la répression transcriptionnelle des régions GC-riches par interférence CRISPR (CRISPRi) conduit à une répression de l'ensemble des gènes *var* (269,312) (**cf. figure 29**). Le mécanisme général par lequel les ARNnc régulent l'expression des gènes *var* n'est pas connu mais une hypothèse suggère que les ARNnc seraient importants dans l'adressage site-spécifique de certains modulateurs épigénétiques (269,306).

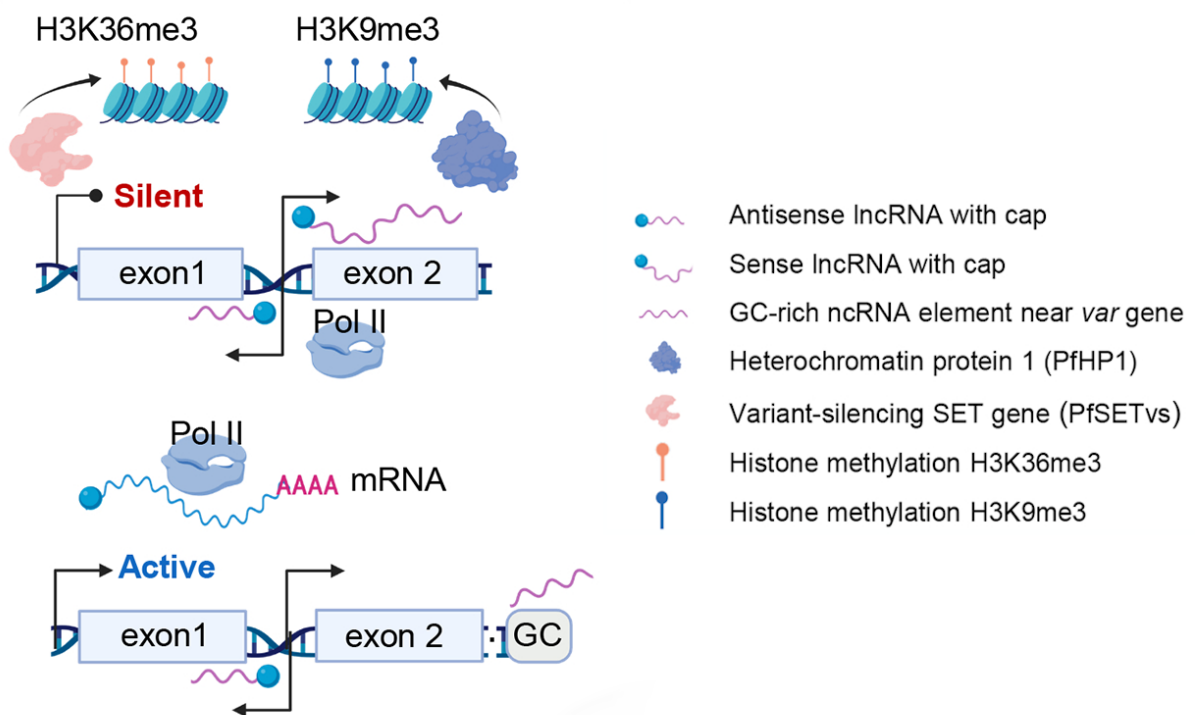


FIGURE 29 : REGULATION DE GENES VAR PAR DES LNCRNA CHEZ P.FALCIPARUM

Régulation de gènes *var* par des lncRNA chez *P.falciparum*. Les lncRNA issus du premier intron des gènes *var* sont associés à leur répression. Les lncRNA issus des régions internes riches en GC des gènes *var* sont associés à leur activation. (D'après Li *et al* 2020 (269))

Chez *T. gondii*, le facteur de réplication Rfa3 est impliqué dans le passage de la forme proliférative tachyzoïte à la forme de quiescence bradyzoïte. La répression du locus *Tg-ncRNA-1*, localisé en amont de Rfa3 et codant deux LncRNA (l'un de 2601 nt et l'autre de 940 nt) a été montrée comme perturbant le passage de tachyzoïte à bradyzoïte (269,313,314). La fonction de ce LncRNA ainsi que sa relation éventuelle avec Rfa3 n'est cependant pas claire. Il est possible que le LncRNA aide au recrutement de complexes de modification des histones pour réguler l'expression de gènes précis au cours de la formation du bradyzoïte (269,315).

T. gondii, contrairement aux *Plasmodium* et *Cryptosporidium* possède les protéines nécessaires à l'ARNi (269,316–318). Cette machinerie est composée d'une protéine Dicer phylogénétiquement apparentée à celle des plantes et champignons, accompagnée d'une protéine Argonaute apparentée à celles des métazoaires (269,318). Les miRNA produits par *T. gondii* sont très variés, avec parfois de fortes similitudes avec certains miRNA de l'Homme et de la souris (210,236,237). La plupart des miRNA de *T.gondii* ne sont pas conservés entre isolats (210,236,237). Les siRNA de *T. gondii* seraient principalement associés à des séquences répétées (269,318).

Chez *T. brucei*, un LncRNA nommé *Grumpy* a été décrit comme jouant un rôle important dans la différenciation du parasite vers sa forme quiescente (319). Cette activité semble se faire grâce à l'action d'un snoRNA présent à l'intérieur du LncRNA, qui ciblerait l'ARNm d'au moins deux gènes liés à la forme quiescente, stimulant leurs expressions (319).



III. LES ARNNC CHEZ LES MICROSPORIDIES

L'existence de petits ARN interférents endogènes a été évaluée chez certaines espèces microsporidiennes présentant la machinerie d'ARNi, notamment *V. ceranae* et *N. bombycis*. Plusieurs petits ARN nommés « miRNA-like » ont pu être identifiés chez ces espèces (238,239,241,272).

La première identification de ces miRNA-like s'est faite sur la base de séquençages haut débit d'ARN en se focalisant sur les molécules de moins de 50 nt (ou small RNAseq (320)). Q Huang et J D Evans (238) ont ainsi pu prédire 6 miRNA-like chez *V. Ceranae* avec l'outil de prédiction de miRNA, miRDeep2. L'expression de ces miRNA a ensuite été confirmée par RT-qPCR (238,321,322). Tirant partie de la mise à jour du génome de *V. ceranae*, trois miRNA-like supplémentaires, toujours prédits avec miRDeep2 ont pu être identifiés puis validés par qPCR et clonage-séquençage (239). L'étude de la corrélation entre l'expression des miRNA-like et l'expression des gènes du parasite et de son hôte a permis l'identification de cibles régulées par ces miRNA-like (239,240). La majorité des cibles identifiées était chez le parasite et concernait des gènes de prolifération et de métabolisme, ce qui suggère que les miRNA-like de *V. ceranae* aient principalement un rôle d'autorégulation du parasite (239,240). Certains de ces miRNA-like présentaient également des cibles chez l'hôte, impliquées dans des processus d'apoptose, de signalisation, et de prolifération cellulaire (240).

Des études fonctionnelles plus générales sur la machinerie RNAi de *V. ceranae* ont montré que la répression par ARNi de la protéine Dicer de *V. ceranae* entraînait une baisse de la charge sporale chez des abeilles infectées (323). Une autre étude a mis en évidence que la répression de la protéine Dicer de *V. ceranae* entraînait également une dérégulation de 10% des gènes codant pour des protéines chez le parasite, et que les abeilles infectées par ces spores présentaient des profils transcriptomiques similaires à des abeilles non-infectées (324). Tous ces éléments laissent penser que l'ARNi de *V. ceranae* joue un rôle important dans l'infection de son hôte (324). Des expériences similaires ciblant la protéine Argonaute du parasite n'ont cependant jamais été effectuées. De nombreux LncRNA, circRNA et miRNA ont été prédits chez *N. bombycis* par des approches basées sur du séquençage d'ARN mais aucune validation biologique n'a été réalisée à ce jour. Des circRNA ont aussi pu être identifiés dans



des spores de *V. cerane* par une approche similaire (282). Les auteurs ont néanmoins validé la prédiction de sept de ces circRNA par une RT-PCR divergente (282). Des études complémentaires demeurent nécessaires afin de confirmer les rôles biologiques associés à ces différents ARNnc.

Dans l'ensemble, les ARN non-codants semblent représenter un pan important de la communication entre un parasite et son hôte. Une grande diversité pour ces molécules existe dans le vivant, et l'étude d'organismes variés permet une compréhension plus complète de leurs rôles. Par leurs caractéristiques génomiques uniques, les microsporidies se présentent comme un modèle de choix chez qui la caractérisation et l'étude des ARNnc dans les interactions hôte-parasite pourraient révéler des aspects fonctionnels et évolutifs encore non décrits, ainsi que des mécanismes physiopathologiques uniques.



PARTIE III : OBJECTIFS DE LA THESE

Dans ce contexte, nous avons établi 4 objectifs pour ce projet de thèse :

OBJECTIF 1 VALIDATION D'UN PIPELINE D'ANNOTATION AUTOMATIQUE SPECIFIQUE DES GENOMES MICROSPORIDIENS

Les génomes microsporidiens présentent des caractéristiques génomiques à part (fort niveau de réduction des gènes y compris des régions UTR ; introns rares et de faibles tailles ; très faible ou très forte densité génique ; nombreux ET avec des séquences divergentes ; forte vitesse d'évolution des séquences). Tout ceci rend l'annotation de leurs génomes difficile et conduit parfois à des annotations peu fiables qui se retrouvent dans les bases de données internationales. Afin de pouvoir mieux caractériser les génomes microsporidiens en vue d'étudier leurs composantes non-codantes, un pipeline d'annotation automatisé spécifique des génomes microsporidiens nommé MicroAnnot a été développé.

OBJECTIF 2 CARACTERISATION GENOMIQUE ET TRANSCRIPTOMIQUE D'ANNCALIIA ALGERAE

Notre modèle d'étude pour ce projet a été la microsporidie *Anncaliia algerae*. Cette espèce présente une forte diversité d'ET dans son génome mais les mécanismes mis en jeu dans leurs régulations sont inconnus. Pour étudier cette question, une première étape a été d'obtenir la séquence complète du génome de cette espèce par l'approche « Highly accurate long-read sequencing » (HiFiSeq) de PacBio qui, en permettant de traverser les régions répétées lors de l'étape d'assemblage, permet d'obtenir les séquences chromosomiques d'un télomère à l'autre. A partir de la séquence du génome ainsi obtenue, une étude génomique a pu être initiée et couplée à une étude de transcriptomique globale du parasite au cours de son cycle infectieux. Ces études ont permis la caractérisation du répertoire en ARN non-codants mais aussi de manière plus spécifique, d'appréhender les mécanismes de régulation des ET chez *A.algerae*.



OBJECTIF 3 ETUDE DU ROLE DE PETITS ARNS NON CODANTS D'ANNCALIIA ALGERAE DANS L'INTERACTION AVEC SA CELLULE HOTE

Le troisième objectif de la thèse visait à se focaliser plus spécifiquement sur les petits ARNnc, afin de voir si des gènes codant ces molécules étaient présents dans le génome d'*A. Algerae* et comment ces petits ARNnc intervenaient dans le dialogue entre hôte et parasite. Pour ce faire, une approche de séquençage spécifique de ces molécules au cours d'un cycle infectieux a été couplée aux données de séquençage Hifi du génome d'*A.algerae*.

OBJECTIF 4 ETUDE DE LA REPOSE TRANSCRIPTOMIQUE DE LA CELLULE HOTE FACE A UNE INFECTION PAR ANNCALIIA ALGERAE

Le dialogue entre hôte et parasite se traduit également par une modulation de l'expression des gènes de l'hôte. Les données de séquençage transcriptomique que nous avons générées nous ont fourni les ressources nécessaires pour évaluer ces changements d'expression chez l'hôte, aussi bien au niveau des gènes codants que des miRNA. L'objectif ici a donc été d'étudier comment l'hôte réagissait à l'infection et comment *A.algerae* modulait l'expression de certains de ses gènes.



Section B : Etudes expérimentales



SECTION B ETUDES EXPERIMENTALES

MODELE D'ETUDE ET METHODES GENERALES DE LA THESE

Pour répondre aux objectifs de ce projet, nous avons opté pour une approche *in vitro*, en vue d'une étude fondamentale des mécanismes d'interaction hôte-parasite médiés par les ARNnc, à l'échelle cellulaire. Un modèle d'infection de fibroblastes humains en monocouche par la microsporidie *A. algerae* a été choisi.

A. algerae est une espèce infectant l'Homme et présentant de nombreuses caractéristiques génomiques et biologiques intéressantes, comme sa capacité à infecter deux espèces à la physiologie très différente (Homme et anophèles) ou encore la forte proportion d'ET présents dans son génome (106,112). *A. algerae* présente également les protéines Argonaute et Dicer dans son génome, qui sont des protéines clés des mécanismes d'ARN interférence (119,304). Enfin, l'infection *in vitro* par *A. algerae* a fait l'objet de plusieurs publications dans la littérature, et fait également l'objet d'une expertise dans notre équipe de recherche. Les spores d'*A. algerae* utilisées sont de la souche Undeen (325).

Les cellules humaines utilisées pour cette étude sont des Human Foreskin Fibroblasts (HFF) obtenues via l'ATCC. Il s'agit de cellules primaires capables de se maintenir *in vitro* sur plusieurs passages (326). L'utilisation de cellules primaires permet d'éviter des biais de réponses cellulaires parfois retrouvés lors de l'utilisation de lignés cellulaires cancéreuses (327). Les HFF sont également un type cellulaire classiquement retrouvé en infectiologie et en parasitologie (328–330), ce qui peut permettre une mise en relation de nos futurs résultats avec ceux issus d'autres expériences d'infectiologie. De plus, le modèle de culture de microsporidies sur HFF a été exhaustivement décrit (331–333).

Point supplémentaire, dans un contexte d'infection microsporidienne naturelle, les fibroblastes seraient probablement le deuxième type cellulaire, après les cellules endothéliales du tractus digestif (97,112), le plus en contact avec le parasite de par leur



ubiquité dans les tissus humains (334). Les fibroblastes possèdent également des rôles charnières dans de nombreux processus physiologiques et pathologiques (335–338).

Enfin, nous avons essayé de choisir un type cellulaire dont les réponses ne seraient pas trop fortement influencées par sa différenciation : les fibroblastes présentent l'avantage d'être des cellules relativement peu différenciées comme peut en attester leur utilisation pour la production de cellules souches pluripotente induite (iPSC) en médecine régénératrice (334,339–342).

Notre modèle d'interaction hôte-parasite établi, il a ensuite été étudié par des approches de biologie cellulaire, biologie moléculaire, séquençage nouvelle génération (NGS), et diverses méthodes d'analyses computationnelles.



PARTIE I : PIPELINE D'ANNOTATION SPECIFIQUE DES GENOMES MICROSPORIDIENS : L'OUTIL MICROANNOT

Pour permettre une étude plus approfondie des gènes d'ARNnc chez *Anncaliia algerae*, des ressources génomiques fiables sont nécessaires. Cela passe notamment par une annotation sensible et spécifique des génomes. L'annotation manuelle bien que fiable, est très chronophage et donc peu adaptée à la quantité de données générée à l'ère génomique (343–345). Des outils d'annotation automatisés ont donc été développés mais ceux-ci, perdent drastiquement en spécificité et en sensibilité lorsqu'il s'agit d'annoter des génomes eucaryotes moins conventionnels comme ceux des microsporidies (346).

Pour faire face à ces problématiques, un pipeline d'annotation dédié aux génomes microsporidiens a été développé et validé : MicroAnnot.

* « Supplementary data » disponibles dans le preprint en ligne

#Article acceptée pour publication dans *International Journal of Molecular Sciences*, preprint :
<https://doi.org/10.20944/preprints202310.1822.v1#>



===ARTICLE ICI===



MicroAnnot représente un apport important dans le domaine de la génomique microsporidienne. Un tel outil permet d'envisager la gestion et la curation des bases de données microsporidiennes et de promouvoir des ressources génomiques aux annotations standardisées et fiables. L'outil est amené à être amélioré encore d'avantage par l'ajout de nouveaux génomes de référence mais également en implémentant de nouvelles approches de machine learning, entraînées sur les données d'annotation haute-fidélité produites par l'approche comparative de l'outil. Dans le cadre de ce projet de thèse, il ouvre la voie à l'étude plus approfondie du génome d'*A.algerae* et de son répertoire en ARNnc mais également en éléments transposables que nous avons ensuite effectuée.



PARTIE II : ETUDE GENOMIQUE ET TRANSCRIPTOMIQUE GENERALE D'*ANNICALIA ALGERAE*

Par rapport à d'autres espèces microsporidiennes, les ressources génomiques et transcriptomiques disponibles pour *A.algerae* étaient relativement parcellaires. En effet, Le génome d' *A.algerae* est très riches en éléments répétés (128), ce qui a rendu sa reconstruction complète difficile avec les approches de séquençage Sanger (138) et Illumina (137). Ces dernières années ont vu l'essor de nouvelles méthodologies de séquençage génomique dont la technologie "Highly accurate long-read sequencing » (HiFi seq) mise au point par la société PacBio. Le HiFi seq est caractérisé par de longs fragments permettant une bonne résolution des séquences, même dans les génomes avec de nombreuses régions répétées ; et des taux d'erreur de seulement 1 nucléotide sur 100 000 (347). Le choix a donc été fait de reséquencer le génome de la souche Undeen d'*A.algerae* en utilisant cette méthodologie afin de permettre une étude plus fine de ses ET.

En parallèle, des données transcriptomiques des petits ARN et des ARN polyadénylés de l'hôte et du parasite tout au long d'un cycle infectieux, ont été générées par séquençage Illumina Novaseq 6000. Le suivi de cette cinétique s'est fait sur 5 points du cycle de développement : 3h, 12h, 24h, 48h et 72h post-infection. Pour chaque temps, trois reliquats biologiques ont été séquencés, ainsi que des échantillons de cellules non infestées (triplicat également) en tant que contrôle.

L'ensemble de ces ressources nous a permis de caractériser plus finement l'organisation, et le répertoire de gènes d'*A.algerae*. Nous avons également pu appréhender les mécanismes de régulation d'ET chez cette espèce.

* « Supplementary data » disponibles dans l'annexe correspondante

#Article en préparation, soumission initiale prévue dans le journal *Genome research*#



LIFE UNDER STRONG TRANSPOSABLE ELEMENT DIVERSITY: GENOMIC AND TRANSCRIPTOMIC LANDSCAPE OF THE MICROSPORIDIAN PARASITE *ANNCALIIA ALGERAE*

Short title: *Anncaliia algerae*'s genomic and transcriptomic landscape

**Authors: Reginald Florian Akossi^{1*}; Ivan Wawrzyniak^{*1}; Emmanuelle Lerat²; Damien Courtine¹;
Jérémy Tournayre³; Frédéric Delbac¹; and Eric Peyretailade^{*1}.**

¹Laboratoire "Microorganismes : Génome et Environnement" (LMGE), UMR 6023 Université Clermont Auvergne
and CNRS, F-63000 Clermont-Ferrand, France.

2

3

E-mails :

Reginald.Akossi@uca.fr / Reginald.F.Akossi@gmail.com

Ivan.WAWRZYNIAK@uca.fr

Eric.PEYRETAILLADE@uca.fr



1. INTRODUCTION

Transposable elements (TEs) are mobile genetic sequences that can spread through and between genomes (1,2). TEs are classified in two main categories: Class I retrotransposons and Class II DNA transposons which can be further divided into several subclasses and numerous families (1). While retrotransposons are transcribed into RNA, exported to the cytoplasm and retrotranscribed by reverse transcriptase to spread in the genome via a “copy-paste” mechanism, DNA transposons can directly expand in the genome as DNA through either a “cut-paste” or “copy-paste” mechanism (1). DNA transposons are also transcribed however, mainly through internal promoters, to produce proteins such as the transposases involved in their (3).

TEs have been described as important for the evolution of genomes and can be found throughout the tree of life (4–6). Through the domestication process implemented by their hosts, they have ensured the acquisition of various regulatory and biological functions in eukaryotes (5,7–10). Despite their roles in genome evolution, TEs can also be the source of genomic instabilities detrimental to the host species. For example, anarchic activation and deregulation of TEs have been associated with enhanced mutation rates, infertility, and cancer (11–15). Thus, TE regulation constitutes an important aspect in the biology of organisms, and accordingly, several TE regulation mechanisms have been described (16–20).

siRNA (small interfering RNA) and piRNA (Piwi interacting RNA) are the main actors of TE regulation in eukaryotes (16). These molecules are small non-coding RNA, typically between 18 and 30 bases long that can repress TE activity in a sequence-specific manner (21). The



regulation can be mediated by heterochromatin formation (for both DNA transposons and retrotransposons) or post-transcriptional repression (mostly affecting retrotransposons), and the small RNAs involved can vary between species (18,21,22). These RNA-dependant post-transcriptional gene silencing (PTGS) or transcriptional gene silence (TGS) fall under the RNA interference mechanisms (RNAi) (21,23). While different proteins may intervene, RNAi typically involves a Dicer-like protein for small RNA maturation, and a protein from the Argonaute family for the regulatory action (21). The proteins from the Argonaute family are defined by the presence of two conserved domains in their sequences, the PAZ domain (PIWI-Argonaute-zwille) involved in oligo-nucleic acid binding and the PIWI domain (P-element-induced wimpy testis) which may display endonuclease activity (24). Argonaute proteins are divided into four clades, the AGO-like family, the PIWI-like family, the WAGO family and the *Trypanosoma Ago* family (24).

As suggested by their name, piRNAs associate with the proteins of the PIWI-like family. piRNAs are absent from plants and fungi and only found in animal species (23). They are mainly expressed in cells of the germline and are associated with TE regulation and silencing (18,25). On the other hand, siRNAs can be found in numerous species including plants, and can be associated with AGO, WAGO or Trypanosome Argonaute proteins, depending on the species and the biological context (21,22,26). siRNAs possess several roles and can be involved in TE regulation in the absence of, or in complement to piRNAs (21,22).

In *Schizosaccharomyces pombe* (fission yeast), TEs are regulated by siRNA generated from the TEs themselves through the action of an RNA dependant RNA polymerase (RdRP) and a Dicer protein. In this case, the transcription products of the TEs are processed by RdRP to form a double-stranded RNA (dsRNA). Next, the Dicer protein recognizes and matures the dsRNA



into siRNAs which, associated with Argonaute and the RISC complex, will regulate TE activity (23,27).

In ciliates such as *Paramecium* and *Tetrahymena*, a macro and a micro nucleus exist (28). In *Tetrahymena*, the maternal macronuclear genome is completely cleansed of TEs, microsatellites, and internal eliminated sequences (IES) during meiosis (IES are thought to be remnants of ancient TEs). This is done through a DNA elimination process mediated by ncRNAs called scanRNAs (scnRNAs). The micro nucleus genome is fully transcribed from both strands in a bidirectional manner by an RNA polymerase II leading to the presence of numerous dsRNAs which are then matured by a Dicer protein and loaded onto the Piwi protein Twi1. The Twi1-scnRNA complex, with the help of several other proteins, is then transferred to the macronucleus and exerts a sequence specific DNA elimination (20,28,29). scnRNAs present strong bias for an Uracil nucleotide at their 5' start and a weaker Adenosine bias 3 bases from their 3' end. These features are derived from base cleavage preferences of the ciliate's Dicer protein (28,30). In *Paramecium*, scnRNAs are processed by two Dicer proteins with complementary functions and then loaded onto two different but highly similar PIWI proteins though to exert similar roles (28).

TEs also present a form of self-regulation in their insertional bias (1). This bias for certain insertion sites is largely influenced by the nuclease protein that catalyses TE integration in genomic sequences. Different nucleases present varying affinities for chromosome sequences, structures, and associated proteins (namely histones). While some TEs can insert randomly, others target specific positions such as intergenic regions (31), promotor regions (32), other TEs (1), or genes where insertions are unlikely to be detrimental to cell function (*i.e.* ribosomal RNA genes (33), unessential genes (34)). Regardless, this would suggest that



smaller eukaryotic genomes with reduced intergenic regions and fewer coding genes such as microsporidia, could be more susceptible to TE induced genomic reorganizations and mutations in the absence of robust TE regulatory mechanisms.

Microsporidia are obligatory intracellular parasites involved in the infection of several animal species of both economic and environmental importance. Their host spectrum includes bees, fish, shrimp, but also mammals including humans, making these parasites relevant in public health concerns (35). Microsporidia are vastly present in food webs and freshwater ecosystems and form a diverse group composed of over 1,600 different species (36,37).

Related to fungi, microsporidia exhibit several genomic and evolutionary traits that set them apart from all other eukaryotes (38,39). First, Microsporidia have the smallest known eukaryotic genomes with the species *Encephalitozoon intestinalis* having a genome of only 2.3Mb (40). This is in part due to the reduction of CDS lengths when compared with orthologues (41). Non-coding genes such as rRNAs are also found shortened (42). Second, the microsporidia phylum exhibits high sequence evolution rates leading to a strong diversity even for proteins involved in conserved mechanisms within the group (38). For example, polar tube proteins, constituents of the microsporidian infectious apparatus, can present strongly divergent primary sequences between orthologous sequences (43). This high sequence evolution rate also makes identifying non-coding RNA genes, such as snoRNA and snRNA, particularly challenging. The identification of the U1 in the *Encephalitozoon* and *Nosema* genera for example, seemed only possible through a dedicated, microsporidian specific, comparative genomic approach (44). Finally, despite sharing several core processes, a large diversity of genomic organisations is found within Microsporidia (37). This includes differing nuclei shapes, numbers and differing ploidy between species (45,46).



This genomic diversity is further enhanced by the numerous TEs found in some microsporidian genomes. While the *Encephalitozoon* genus present no TE (44,47), other microsporidian genera and species can present several hundreds of TE families in their genomes. Some of these TEs seem derived from horizontal transfers with metazoan hosts (48).

The species *Anncaliia algerae* possesses a genome which contrasts with the typical genomic compaction of microsporidian genomes and yields a larger genome, with a light protein coding gene density (49). Indeed, while its haploid genome is currently estimated at about 12 Mb (50), *A. algerae* is thought to harbour only about 2,075 protein coding genes, not much more than in the 2.3 Mb genome of *E. intestinalis* that contains 1,833 predicted protein coding genes (51,52). Despite this relatively small difference, several key molecular functions are found in *A. algerae* but are absent from *E. intestinalis*. This includes, but is not limited to, RNA interference (RNAi), associated proteins Argonaute and Dicer (40,51,53).

The *A. algerae* genome is also marked by the expansion of over 240 TE families (48) and an uncharacterised ncRNA repertoire. Long non-coding RNAs (LncRNAs) are an heterogeneous group of ncRNAs involved in various biological processes (54). These molecules are suspected to have derived from TEs and TE activity (7). As such, their high presence in *A. algerae* seems likely. Taken together, all these factors may explain the size difference between the genomes of *A. algerae* and *E. intestinalis*.

Overall, with its high diversity of TEs, relatively small genome, reduced number of coding genes, and unknown ncRNA repertoire, *A. algerae* is an interesting model for the study of genome evolution and adaptation to a strong TE diversity and the associated regulation processes. Currently available genome assemblies of different *A. algerae* strains were obtained using either Sanger (50) or Illumina sequencing methods (55,56). These assemblies



present numerous contigs with suboptimal assembly of TE copies and non-coding regions. This is most likely due to the high abundance of TEs and low complexity regions, combined with the overall AT richness of the genomes (51). This makes it difficult to determine the ploidy in *A. algerae* even though gene sequence polymorphism has been observed (51). The lack of supporting transcriptomic data is also an issue that makes validation of non-coding RNA genes more difficult.

In this study we applied long read-sequencing, and more specifically PacBio's High Fidelity (HiFi) sequencing approach (57), combined with RNA sequencing (RNAseq) data in the aim of overcoming these constraints and highlight i) The genomic organisation of *A.algerae*, ii) the ploidy of the species, iii) the TE diversity and distribution iv) the complete gene repertoire (ncRNA and genes coding for proteins), and v) potential TE regulation mechanisms.

All in all, this work brings novel insight into the genomic organisation, ploidy, non-coding RNA repertoire and transposable element dynamics in the microsporidian specie *A. algerae* and proposes a TE regulation mechanism at play in the parasite.



2. RESULTS

1. Genomic characteristics of *Anncaliia algerae*

1.1. Novel assembly and annotation of the *A. algerae* genome

The new sequencing effort of the *A. algerae* genome using HiFi sequencing yielded to an assembly of 239 contigs for a total size of 43.96 Mb and a N50 of about 0.50 Mb. A haploid genome of 17 contigs was predicted by Purge Haplotigs with a total size of 11.78 Mb and a N50 of 0.73 Mb (**table 1**).

Telomere repeat regions, identifiable by TTAGGG short tandem repeats, were searched for and found on 16 of the 17 contigs in our haploid assembly suggesting reconstruction at the chromosomal scale (**table 1**).

BUSCO assessed the biological completeness of our complete genomic assembly at 81.9 %. Of the missing 18.1% which corresponded to 109 genes without sufficient HMM profile identity to be kept by BUSCO, 92 presented orthologous sequences in the BUSCO microsporidia database when using BLASTN analysis (e-value < 1e-5). This would allow an actual completeness of about 97% to be expected.

Annotation of the haploid genome assembly using MicroAnnot reported 2,916 protein coding genes (with 73 duplicated and 2,554 single copy genes), 14 complete rRNA genes and 56 tRNA genes (**table 1, supplementary figure 3**). The genome presented a gene density of about 235 genes per Mb (1 gene every 4 Kb). Genes had an average length of about 973 bp (**table 1**).

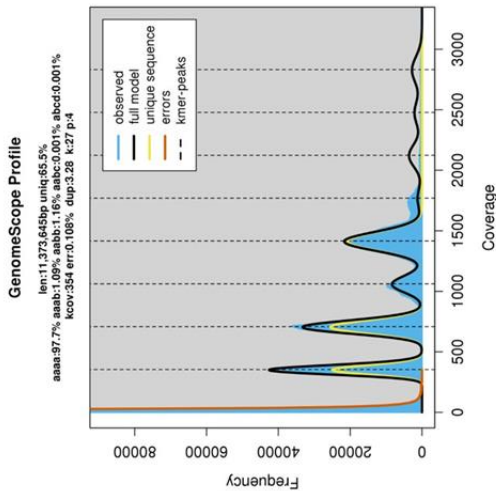
Table 1: Summary of *A. algerae* Undeen genomic DNA assembly and annotation

		A.algerae Undeen assembly	Undeen Haploid genome
Assembly	Number of contigs	239.00	17.00
	Total assembled-genome length (bp)	43,961,032.00	11,783,317.00
	N50 length (bp)	495,181.00	732,356.00
	Longest contig	1,456,444.00	1,456,444.00
	G+C content (%)	23.86	23.96
	Number of N	-	-
Search for telomere reapeat regions	Number of contigs with repeats on both ends		9
	Number of contigs with repeats on one end		7
	Number of contigs with no repeats found		1
Assembly on haploid genome mapping statistics	Percentage of contigs that mapped on the haploid genome		222 (100%)
Microannot annotation	Total number of protein coding genes	10,639.00	2,916.00
	Mean gene length (bp)	964.00	973.00
	Number of genes with introns	-	-
	No of tRNA	210.00	56.00
	No of complete ssU-LSU rRNA genes	34.00	14.00
	Gene density / Mb	237.00	235.00
	Busco completeness (microsporidia_odb10) (%)	81.90	80.60
CDS duplication	Nb of duplicated CDSs		73.00
	Total nb of distinct CDS sequences		2,554.00

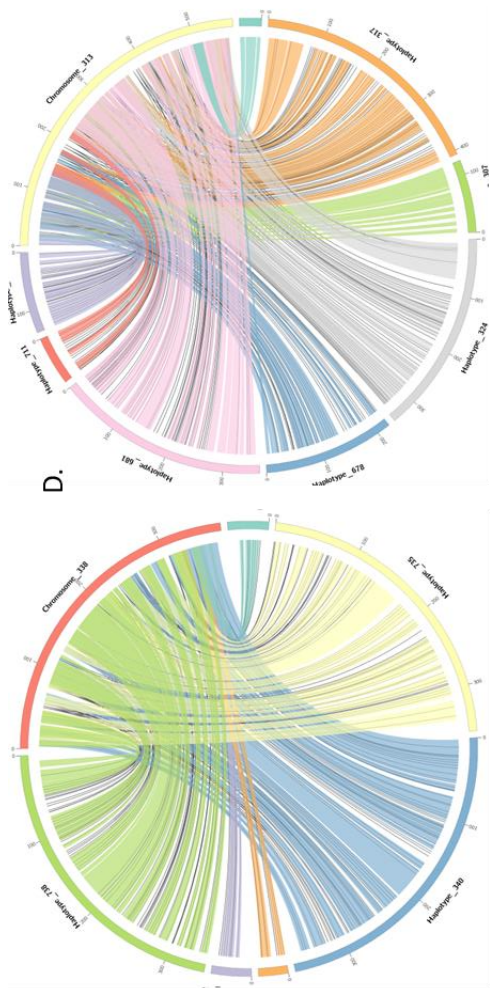
1.2. Comparison with the *A. algerae* reference strain PRA339

The Undeen genome was compared to the *A. algerae* reference strain PRA339 genome assembly (56). This genomic comparison revealed strong conservation of sequences at the nucleic level (**supplementary figure 1A**), which also came with conservation at the level of protein sequences, with high synteny both in terms of CDS organization and orientation (**supplementary figure 1B**). The BUSCO completeness percentage based on the MicroAnnot coding genes annotations in Undeen and PRA339 (haploid genomes) revealed a value of 79% for the Undeen genome and 75.5% for the PRA339 genome, reflecting a gain in genomic information brought by the long read-sequencing approach. Looking at blast homologies against the BUSCO microsporidiaodb10 database, we estimated 95.3% actual completeness for Undeen's haploid assembly and 93.8% for PRA339 (e-value < 1e-5).

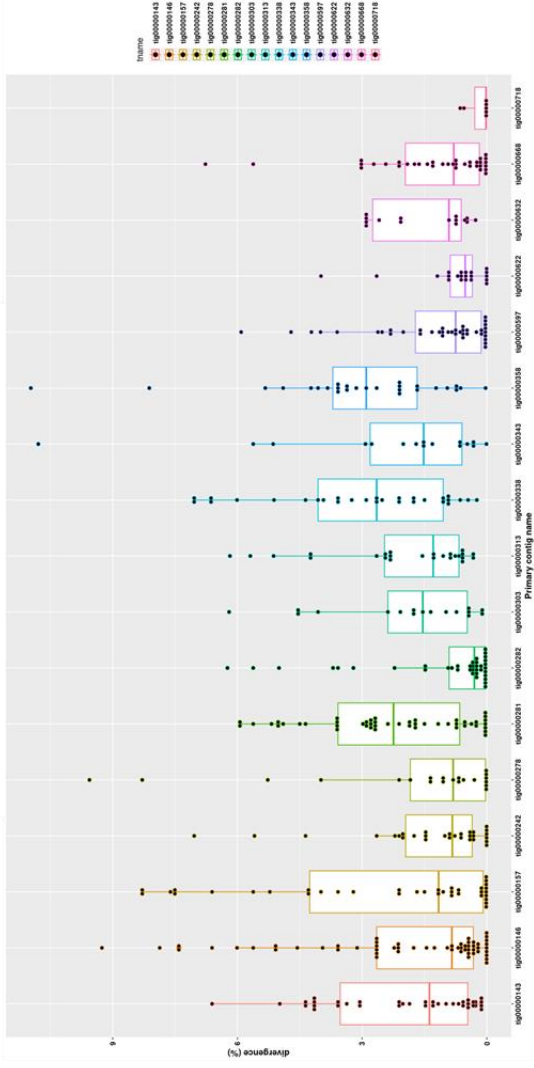
In terms of functional gene composition, gene ontology (GO) annotation did not reveal any major differences between the two strains, whether in terms of cellular component, biological process or molecular function Gene Ontologies (GOs) (**supplementary figure 1C**). However, four, eighteen and seventeen GO terms associated with cellular components, molecular functions and biological processes respectively, were only identified in Undeen (**Supplementary figure 1 C, supplementary table 1**). These were probably due to differences in sequencing methods and depth between the two assemblies and not due to strain differences. Looking at the common PRA339 and Undeen cellular component level GOs, dominant subcategories were "microtubules", "proton transporting V-type ATPase", "NATa complex" and "CCATT-binding factor complex" (**supplementary figure 2A**). For the molecular function



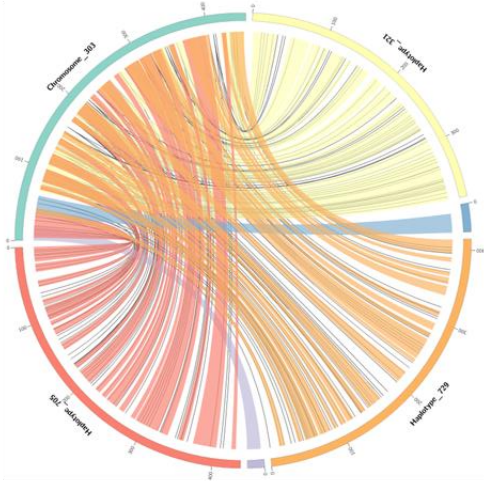
C.



D.



B.



E.

Figure 1: Analysis of the *Anncalia algerae* ploidy and divergence between homologous chromosomes

A. GenomeScope profile of the *A. algerae* assembly data; four peaks can be observed corresponding to tetraploidy in the studied genome B. Per chromosome mean divergence graph defined by comparison of all generated contigs against the primary haploid genome using minimap2 C.D.E. Example synteny graphs between homologous chromosomes. Alignments were made using the CDS and 10 Kb long DNA fragments. Thick coloured lines represent syntenic blocks of >10kb and thin black lines represent CDS alignments.

GO, they were “cation transmembrane transporter activity”, “protein kinase activity” and “GTPase activity” (**supplementary figure 2B**). And finally for biological processes GO, the main terms were “intracellular transport”, “isoprenoid biosynthetic process” and “cation transport” (**supplementary figure 2C**).

1.2. Characteristics of the *A. algerae* ploidy

The ploidy analysis performed with Genomescope2 reported tetraploidy for the *A.algerae* genome (**Fig1 A**). Results were further corroborated by the Smudgeplot and nQuire programs, and by the clustering and analysis of protein sequence homologues (**Supplementary figure 3**).

The four homologues for each *A. algerae* chromosome could not be fully reconstructed from telomere to telomere as shown by the number of contigs and the N50 of the assembly data (**Table 1**).

Sequence divergence and overall synteny between homologous contigs was analysed. This revealed low overall sequence divergence (averages less than 3% divergence), with some chromosomes like tig00000718 showing very strong similarity between homologous sequences (less than 1% divergence). Other chromosomes like tig00000358 exhibited more divergent sequences (about 3% divergence) (**Fig1 B**).

Finally, the evaluation of the synteny between the homologous contigs by an all-vs-all blast approach using CDSs or 10 Kb DNA fragments as a query, both showed large conserved sequence blocks, suggesting high degree of synteny (**Fig1 C**).

Table 2: Summary of the transposable element diversity identified in *A. algaeae* Undeen haploid genome.

Haploid genome			
Family	Copies	Total Bp	% of Genome
DNA/IS630	5.00	1,977.00	0.02
DNA/ISXO2	5.00	1,659.00	0.01
DNA/Mariner-Tc1	280.00	249,736.00	2.12
DNA/Merlin	2,409.00	1,636,698.00	13.89
DNA/Piggy-Bac	8.00	10,946.00	0.09
DNA/Unknown	47.00	17,067.00	0.14
DNA transposons	2,754.00	1,918,083.00	16.28
LINE/Unknown	6.00	4,768.00	0.04
LINE transposons	6.00	4,768.00	0.04
LTR/Unknown	6.00	4,301.00	0.04
LTR/Ty3	2,206.00	881,971.00	7.48
LTR transposons	2,212.00	886,272.00	7.52
Total	4,972.00	2,809,123.00	23.84
Low complexity regions	1,572.00	84,089.00	0.71
Satellite repeats	-	-	-
Simple repeats	3,695.00	234,104.00	1.99
Unknown	81.00	45,815.00	0.39

2. Transposable element landscape: number, families, and distribution

2.1. Transposable elements represent about 25% of the *A. algerae* genome

TE analysis of the *A. algerae* haploid genome identified 4,972 TE copies corresponding to 260 TE families (**table 2**). These TE families account for 23.84% (in base pairs) of the genome, with DNA transposons and LTR-retrotransposons being the most represented subclasses and constituting 16.28% and 7.52 % of the haploid genome respectively. For DNA transposons, while IS630, ISXO2, Mariner-Tc1, Piggy-Bac and other unknown DNA elements were identified, the Merlin DNA transposon family with 2,409 copies (87% of the DNA transposon copies), already accounted for 13.89 % of the genome sequences. For the LTR-retrotransposon elements, the Ty3 family with 2,212 copies represented 7.48% of the genome and 99.7% of the LTR-retrotransposons. Non-LTR retrotransposons were mainly represented by LINE elements but with only six copies identified, accounting for 0.04 % of the genome base pairs.

In addition to these TEs, some low complexity regions and simple repeat regions were also identified. No satellite type repeats were observed however (**table 2**).

A.

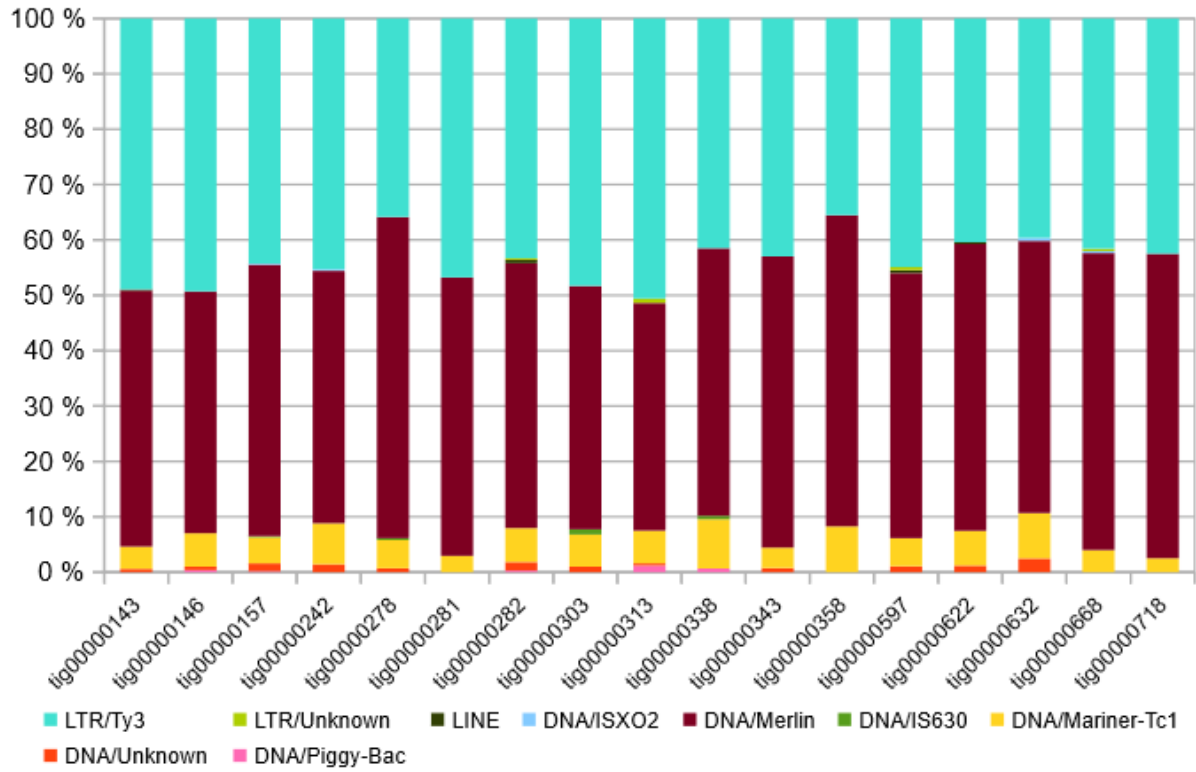


Figure 2: Analysis of transposable element distribution and copy divergence in the haploid genome of *A. algerae*

A. Proportions for each TE class along the 17 chromosomes of the *A. algerae* haploid genome.

2.2. Transposable elements are distributed throughout the *A. algerae* genome and present various divergence levels suggesting past and present activity

TE distribution analysis revealed no insertional bias between chromosomes and/or regions of chromosomes (**Fig 2**). Each chromosome presented similar relative TE family composition (**Fig2 A**), and no specific positional bias within chromosomes was found (**Fig2 B**). The TEs seemed evenly distributed between regions of varying CDS densities, with also some CDSs identified inside TEs and some TEs inserted inside CDSs (**Fig2 B**).

TE activity in the genome was explored through analysis of the sequence divergence of TE copies within each family. Recent (or non-divergent) TEs could be observed in the *A. algerae* genome (**Fig2 C**). No preferential localisation of strongly or weakly divergent TEs was found in our analysis, although a few small stretches of TEs with similar divergence levels could sometimes be observed.

Focusing on the three most represented TE families (*i.e.* Merlin, Ty3 and Mariner-Tc1) and their TE copy divergence profiles, two different dynamics could be observed (**Fig2 D**). For Ty3 the distribution showed few copies with more than 30% divergence, a higher number of copies displaying from 20% to 30% divergence, a plateau between 2 and 20% divergence and then a high number of almost non divergent copies. For the Merlin and Mariner-Tc1 TE families, a bimodal distribution can be observed, with a higher number of copies exhibiting a divergence of more than 26% and between 0 and 6% respectively. A lesser number of TE copies were observed between those intervals. These results suggested different TE dynamics in the *A. algerae* genome depending on the TE family, and potential external constraints and regulations to TE expansion in the genome, inducing non-linear TE divergence and/or elimination from the genome.

B.

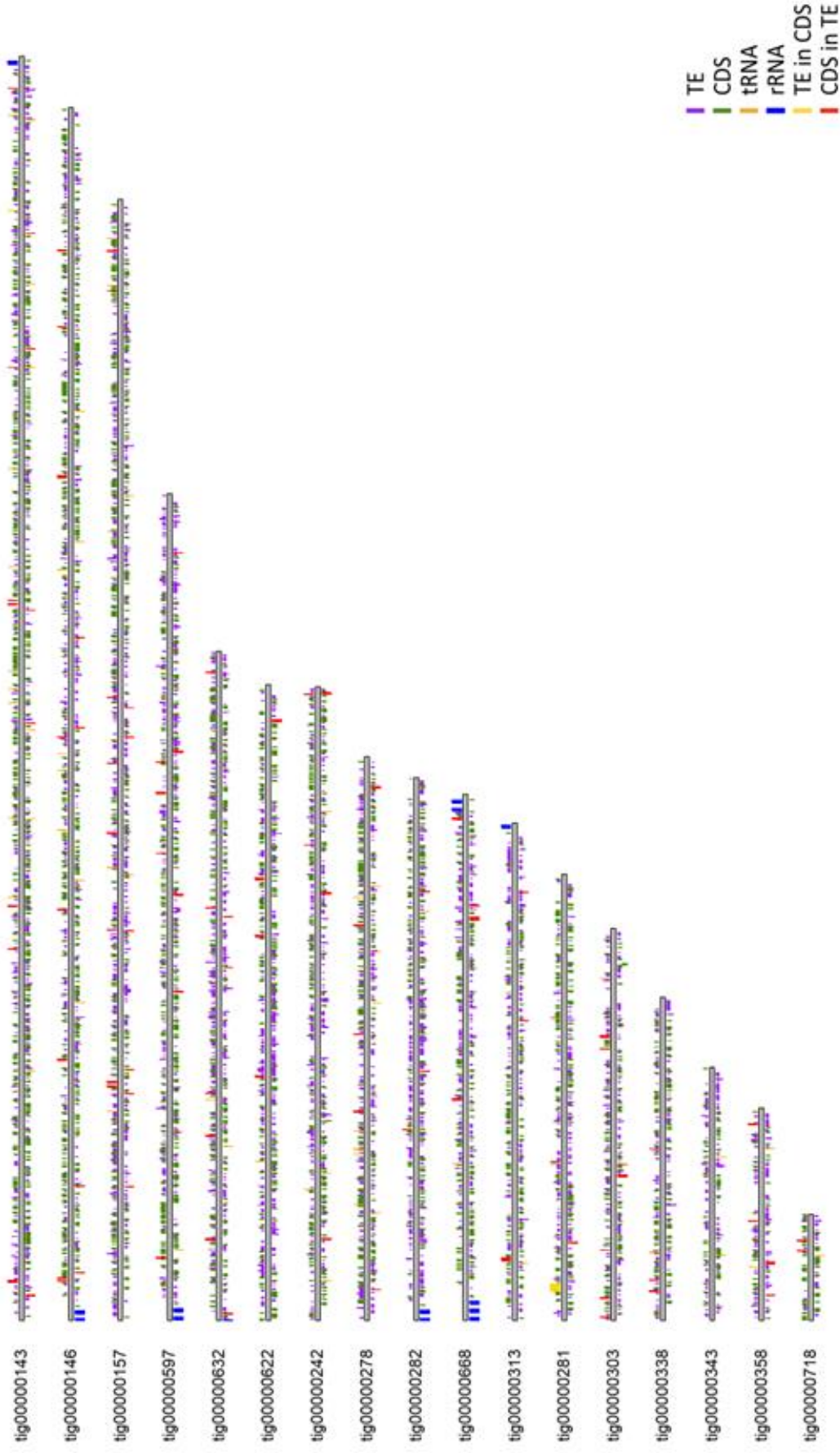


Figure 2: Analysis of transposable element distribution and copy divergence in the haploid genome of *A. algerae* (continued)

B. Physical distribution of TE, CDS, tRNA, and rRNA along each strand of the *A. algerae* haploid genome. TEs are coloured in purple, CDSs in green, tRNA in orange and rRNA in blue. TEs fully contained in CDS are marked in yellow and CDS fully contained in TEs are marked in red. Bars can be stacked on top of one another denoting overlap between regions

C.

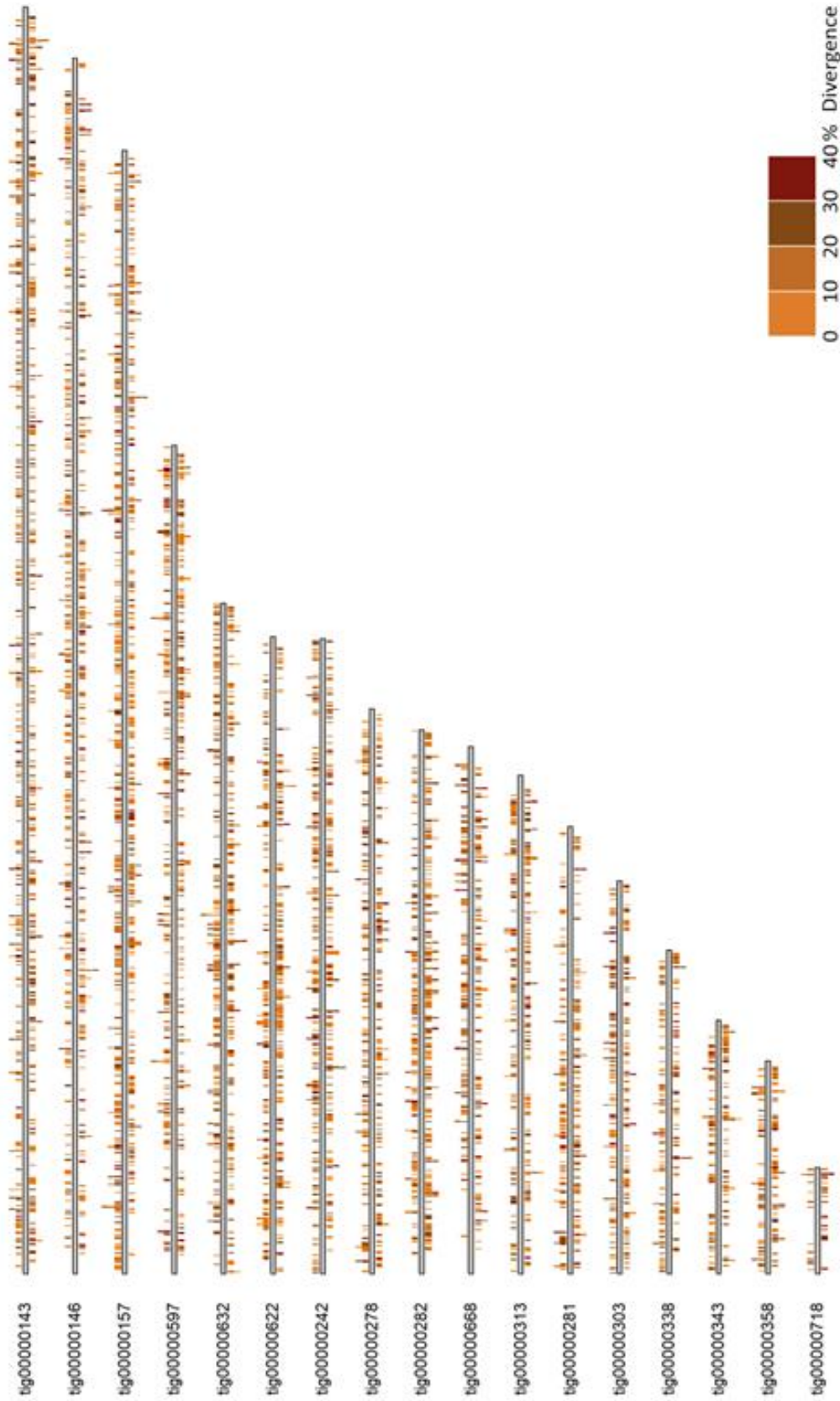


Figure 2: Analysis of transposable element distribution and copy divergence in the haploid genome of *A. algerae* (continued)
 C. Physical distribution of TEs along each strand of the *A. algerae* haploid genome. TEs are coloured in shades of brown depending on their divergence percentage ([0-10%], [10-20%], [20-30%], [30-40%]).

D.

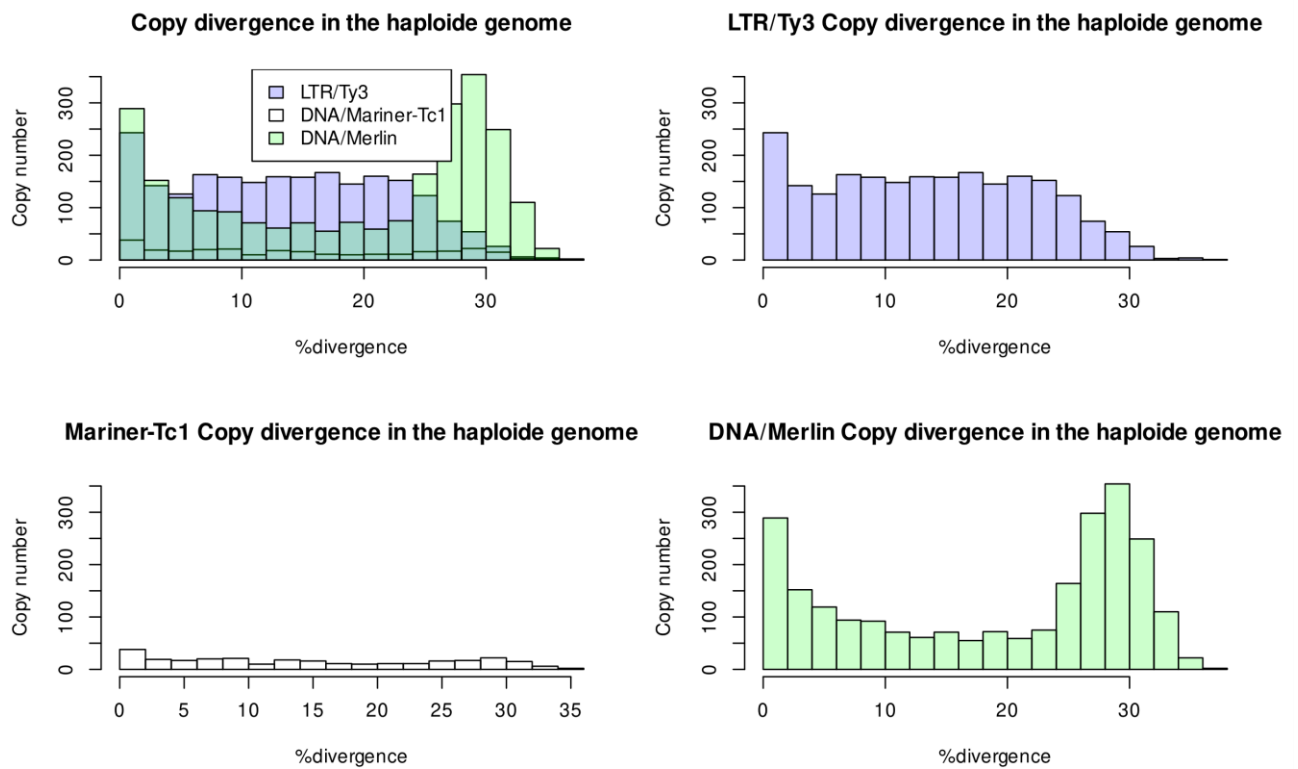


Figure 2: Analysis of transposable element distribution and copy divergence in the haploid genome of *A. algerae* (continued)

D. Divergence profiles for the three most represented TE families (Ty3, Mariner-Tc1, and Merlin).

Table 3: Metrics of the assembled transcriptomes

	Consensus_transcriptome	CDS transcripts	TE_transcripts	tRNA_transcripts	rRNA_transcripts
Number of sequences	8433	2120	263	20	13
Number of sequences between 0-1kb	7017	1728	257	9	12
Number of sequences between 1kb-2kb	1200	341	6	2	1
Number of sequences between 2kb-3kb	162	42	0	0	0
Number of sequences between 3kb-4kb	33	7	0	0	0
Number of sequences between 4kb-5kb	19	2	0	9	0
Minimum length	104	201	200	104	225
Average length	634.6	666.8	374.6	2458.2	591
Maximum length	4835	4644	1965	4764	1142
GC(%)	28.73	28.59	30.46	46.8	28.14
Q1	289	326	230	428	301
Q2	461	522.5	310	1155	641
Q3	774	833.5	431	4719	805

Table 4: Newly identified ncRNA features through the *A. algerae* transcriptome analysis

	Number of transcripts	Number of corresponding Loci in <i>A.algerae</i> genome	non coding transcripts identified by Rfam
Total LncRNA	5894	4777	0
LincRNA	764	670	0
AS_LncRNA	1611	1031	0
eRNA	3519	3076	0
5S rRNA	4	3	4
LSU rRNA fragments	34	31	34
SSU rRNA fragments	9	9	9
tRNA	20	17	20
U2 snRNA	2	2	2
RNase P nuc	2	2	2
RNase MRP	1	1	1
Unclassified < 200nt transcript sequences (sncRNA)	13	13	0

3. Characterization of the *A. algerae* transcriptome

3.1. Presence of a few very long transcripts in the *A. algerae* transcriptome and identification of novel ncRNA genes

The consensus transcriptome assembled from all the transcriptomic data was composed of 8,433 different transcripts (**table 3**). GC percent in the consensus transcriptome was slightly higher although in similar range to that of the genome (28.73% and 23.96% respectively). The average transcript length including CDS, TE and all ncRNA transcripts was of 634.6 nt. CDS associated transcripts displayed a similar average length of 666.8 nt, which is smaller than the average CDS gene length in the genome (973 nt), suggesting the presence of smaller transcripts corresponding to parts of CDSs. Interestingly, the longest transcript in the CDS category displayed a size of 4,644 nt (TCONS_00014202), which is more than four times the average CDS gene length (**table 3**). This transcript started 4150 nt upstream a CDS, covered it completely, and ended 113 nt further downstream. This transcript did not intersect any other annotated element. We evaluated the number of transcripts in size intervals of [0-1 Kb], [1 Kb-2 Kb], [2 Kb-3 Kb], [3 Kb-4 Kb] and [4 Kb-5 Kb] and found that while most transcripts were less than 1000 nt long, several CDS transcripts could be found over this threshold. The number of transcripts seemed to gradually decrease as we looked at larger sizes. Transcripts related to TEs, tRNAs and rRNAs did not account for most of the transcripts in the transcriptome, suggesting the presence of numerous other ncRNA transcripts. NcRNA transcripts were annotated using the Rfam database which allowed us to predict previously

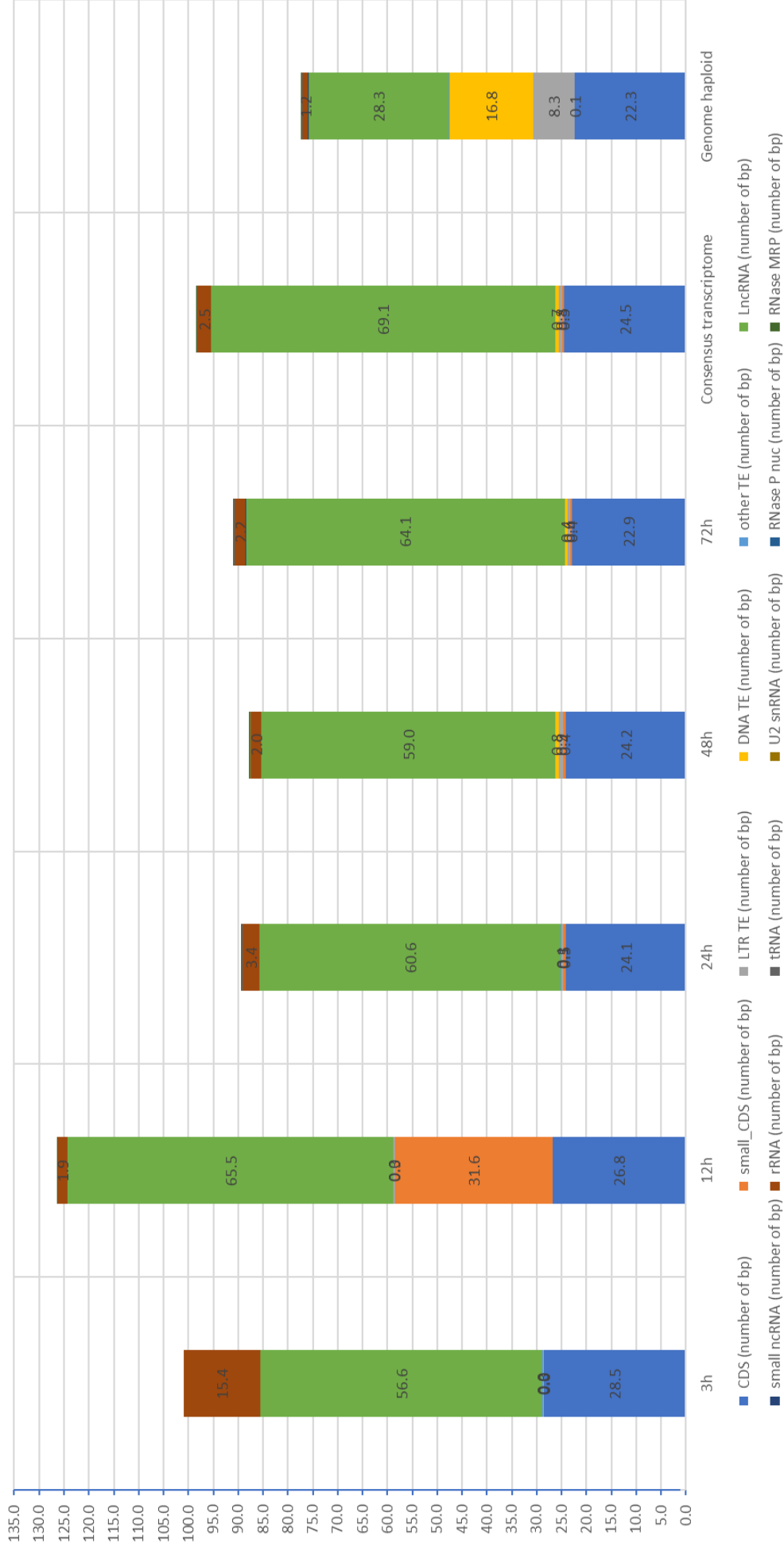


Figure 3: Analysis of the composition of the *A. algerae* genomes and transcriptomes

Stacked histogram showing the number of nucleotide bases associated with various gene types in the genome and transcriptomes of *A. algerae*. Percentages represent the proportion the gene type represents in the given transcriptome or genome.

unannotated ncRNA genes in the *A. algerae* genome (**table 4**). This included two overlapping opposite strand U2 snRNA coding loci, two nuclear RNase P coding loci, one RNase MRP coding loci, and three 5S rRNA loci. Some transfer, rLSU and rSSU RNA fragment transcripts were also identified. In the remaining transcripts Rfam could not annotate, those over 200 nt long were classified as LncRNA and numbered 5,894 from 4,777 different loci. These were subdivided in 764 LincRNAs (670 different loci), 1,611 AS-LncRNA (1,031 different loci) and 3,519 eRNA candidates (3,076 loci) (**table 4**).

3.2. Most of the bases from the *A.algerae* transcriptomes are associated with LncRNA regardless of the infection kinetic timepoint

The gene composition of the *A. algerae* genome, the consensus transcriptome and the timepoint specific transcriptomes were evaluated using all the added annotations (**Fig 3**). About 28.3% of the haploid genome (in base pairs) was associated with LncRNA features. Similarly, most of the bases in the transcriptomes could be associated with LncRNAs regardless of the timepoint. TEs, while representing about a quarter of the genome, only accounted for about 1% of the consensus transcriptome, and similar values were observed in the timepoint specific transcriptomes.

Looking at the timepoint specific transcriptomes sensible variations in the spread of gene type composition could be seen. For example, a higher proportion of rRNA associated transcripts in the 3h transcriptome (15.4% of all nucleotides) was observed. Still another example in the 3h transcriptome, no DNA transposon associated transcript was found, while some LTR-

Argonaute,Dicer and RNA dependant RNA pol transcriptomic expression count graphs

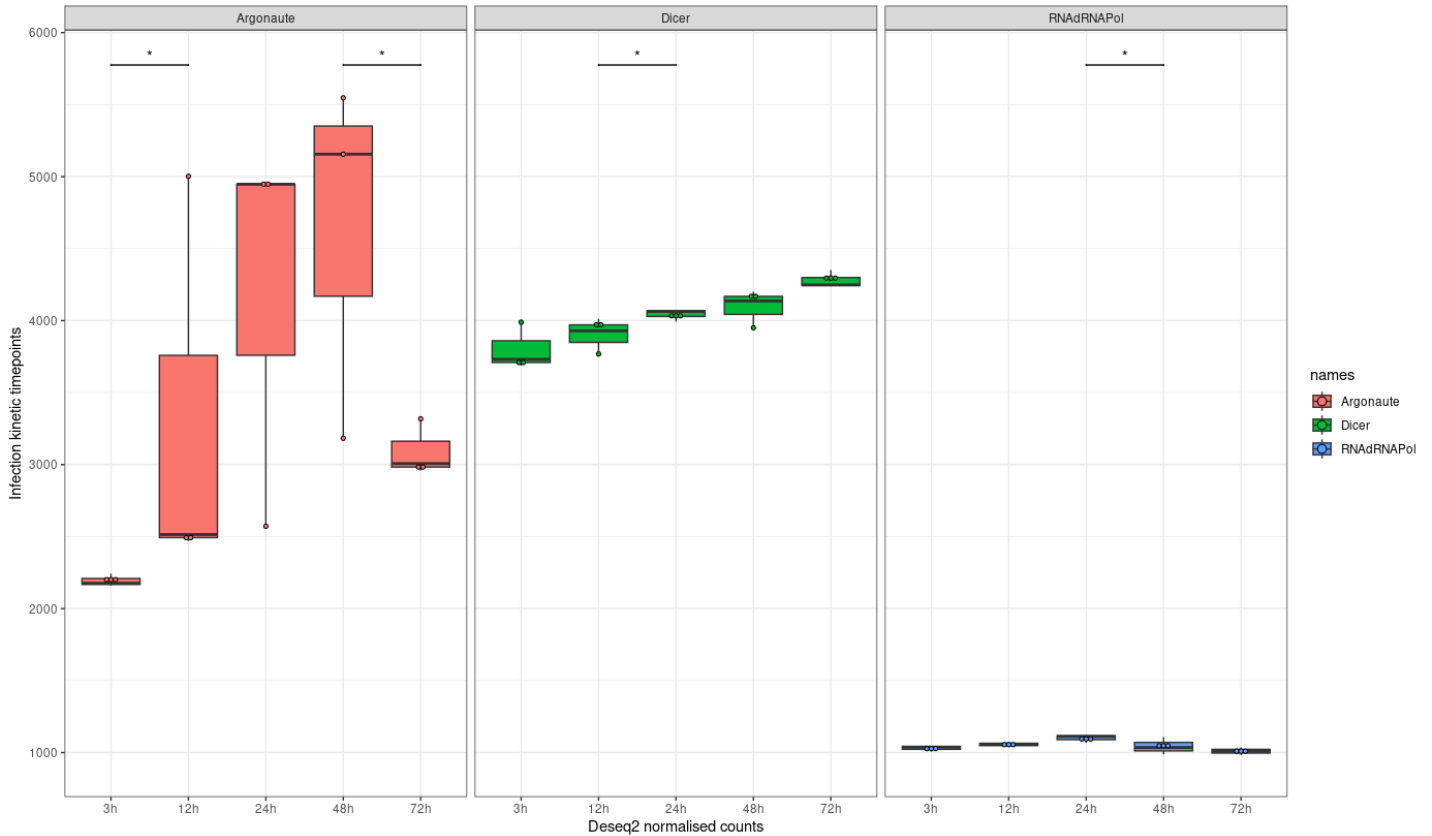


Figure 4: *A. algerae* Argonaute, Dicer and RNA dependant RNA polymerase expression throughout the infection cycle of HFF cells

Gene expression levels of *A. algerae* Argonaute, Dicer and RNA dependant RNA polymerase throughout its infectious cycle. Data is derived from the polyA+ RNAseq data. Counts are normalised using DEseq2. Stars indicate $p_values \leq 0.05$ (Wilcoxon tests). All three genes are expressed throughout *A. algerae* intracellular stages. Differential expression between some timepoints suggests regulation of these genes throughout the parasite development stages.

retrotransposon transcripts could be identified. Other gene types such as CDS, TEs and tRNA seemed to present stable relative representation in the transcriptomes of each timepoint. Values over 100% were observed on our stacked histograms suggesting nucleotides attributed to several genes and thus transcripts fully overlapping multiple genes in our case. Conversely, values less than 100% in some transcriptomes suggested the presence of unassociated nucleotides and so transcripts that did not fully correspond to any annotated gene; such bases were few, representing at most 15% of the transcriptome (48h post infection transcriptome). In the genome, such nucleotides most likely corresponded to unexpressed, intergenic regions.

4. Aspects of transposable element regulation in *A. algerae*

4.1. A minimal set of RNA interference related proteins is present and expressed in *A. algerae*

The search for RNA interference and TE regulation related genes (**supplementary table 2**) confirmed the presence of one Argonaute and one Dicer related protein and revealed the existence of an RNA dependant RNA polymerase protein (RdRP) in the genome of *A. algerae*. Several other proteins involved in canonical miRNA and piRNA biogenesis and activity such as Drosha, DGCR8, Loquacious, GW and Hen1 for miRNA and Zucchini, Vret and Yb for piRNA were found to be absent.

Classification of the Argonaute protein using sequence similarity revealed its association with the AGO clade, rather than with the PIWI clade (**Supplementary figure 4**). The *A. algerae* Argonaute, Dicer and RNA dependant RNA polymerase genes were found expressed throughout the parasitic cycle at variable intensities (**Fig 4**).

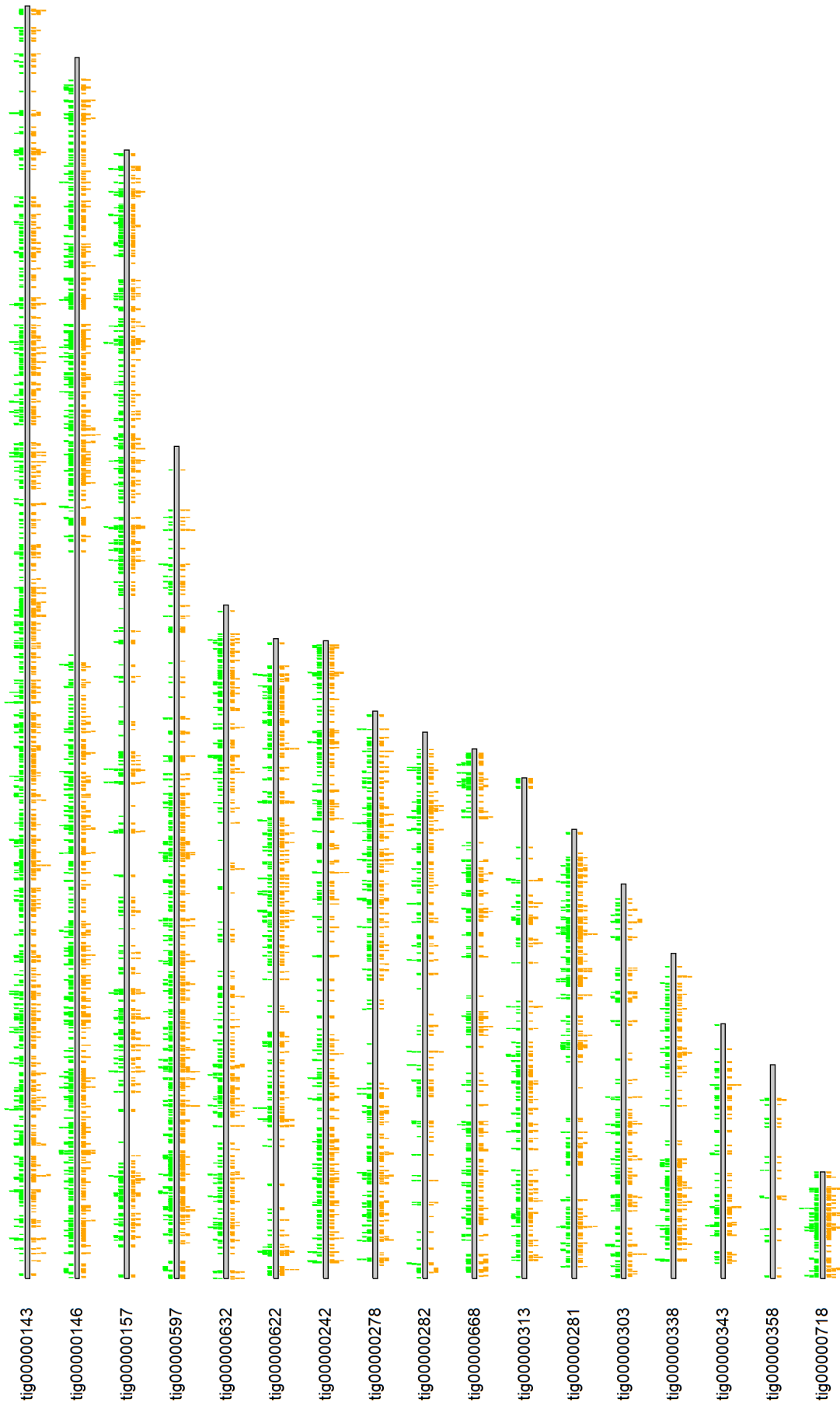


Figure 5: Analysis of strand specific transcript mapping on the *A. algerae* haploid genome

Transcribed regions from one strand are indicated in green on top and those from the opposite strand in orange on the bottom. Overlapping transcripts on the same strand are visible.

Argonaute and Dicer genes were highly expressed throughout the parasite development cycle with some variations between timepoints, while the RdRP gene was found expressed at a rather constant low-level rate throughout the cycle. In both cases this suggested regulation for these genes as well as probable translation and activity during the *A.algerae* life cycle.

4.2. Transcript mapping to both strands can be found on most of the *A. algerae* genome

The strand specific expression profile of the *A. algerae* genome was assessed using the consensus transcriptome, in search of opposite strand transcripts likely to form complementary dsRNAs. Such transcripts were found throughout most of the genome (**figure 5**). This observation suggested either extensive transcription of both strands of the *A.algerae* genome, or the presence of antisens transcripts possibly generated through the action of the RdRP.

4.3. TE expression and number of associated small RNAs are negatively associated for several TE copies

The number of small RNAs derived from transposable elements (TEsRNA) was tested for association with the expression level of the corresponding TE copie throughout the intracellular cycle of *A. algerae*. This association was mostly negative (**Fig 6 A**).

We further tested correlation and most of the TE/TEsRNA groups which significantly correlated were indeed negatively associated (p value ≤ 0.05), regardless of the TE class and family (**Fig 6 B**). When estimating the amount of TEs in each family that had expression levels

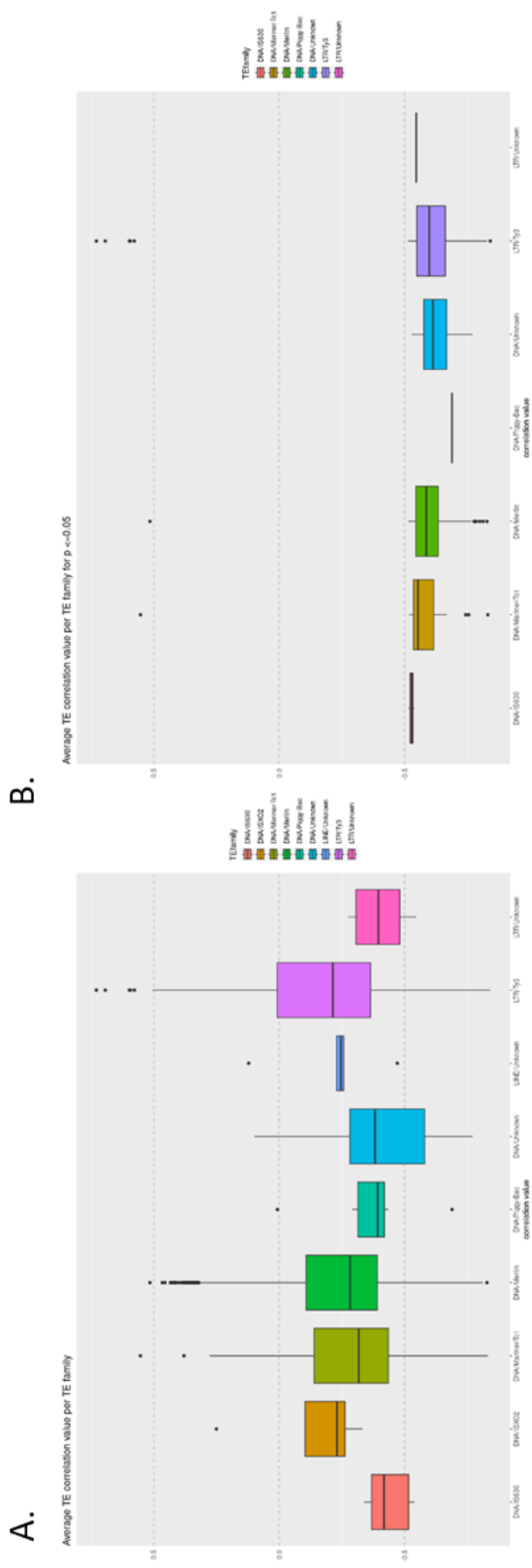
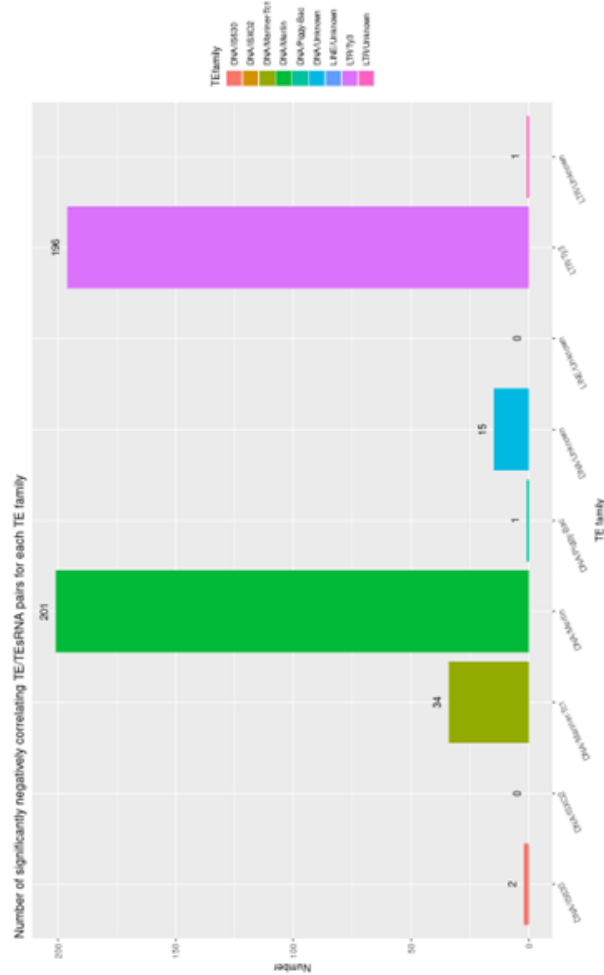


Figure 6: Analysis of the correlation between TE expression and the number of small RNA associated with the TE, throughout the *A. algerae* intracellular stages

A. Average Spearman correlation values of TE/TEsRNA groups by TE family **B.** Average significant correlation values of associated TE/TEsRNA groups after Spearman correlation tests ($p < 0.05$)

C.



D.

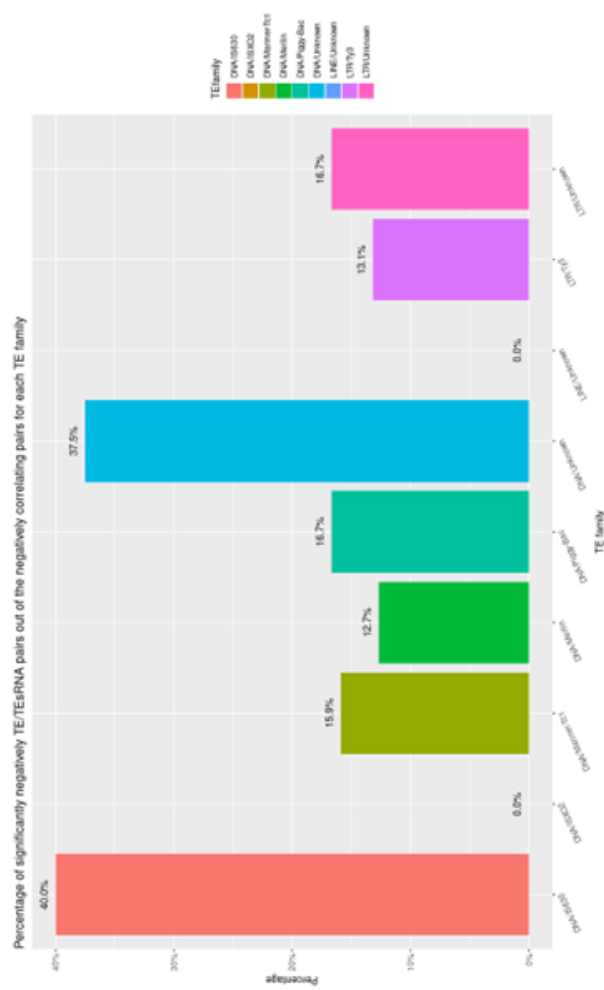


Figure 6: Analysis of the correlation between TE expression and the number of small RNA associated with the TE, throughout the *A. algerae* intracellular stages (continued)
C. Number and D. Percentage of negatively correlated TE/TEsRNA pairs for each TE families ($p < 0.05$).

E.

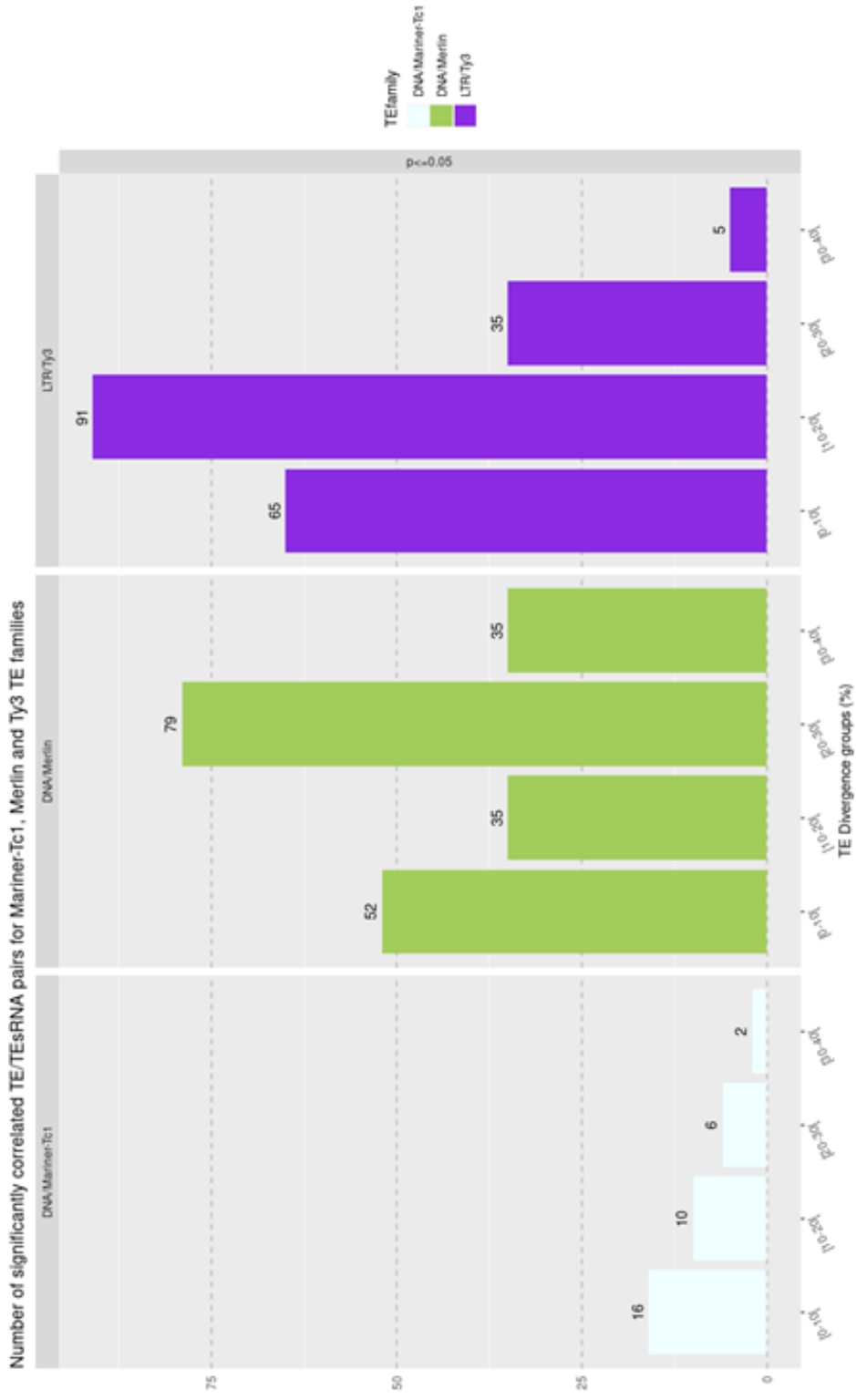


Figure 6: Analysis of the correlation between TE expression and the number of small RNA associated with the TE, throughout the *A. algerae* intracellular stages (continued)

E. Number of negatively correlated TE/TEsRNA pairs per divergence level category for the three main TE families (Mariner-Tc1, Merlin and Ty3). Divergence level categories are [0-10% [divergence, [10-20% [divergence, [20-30% [divergence and [30-40% [divergence.

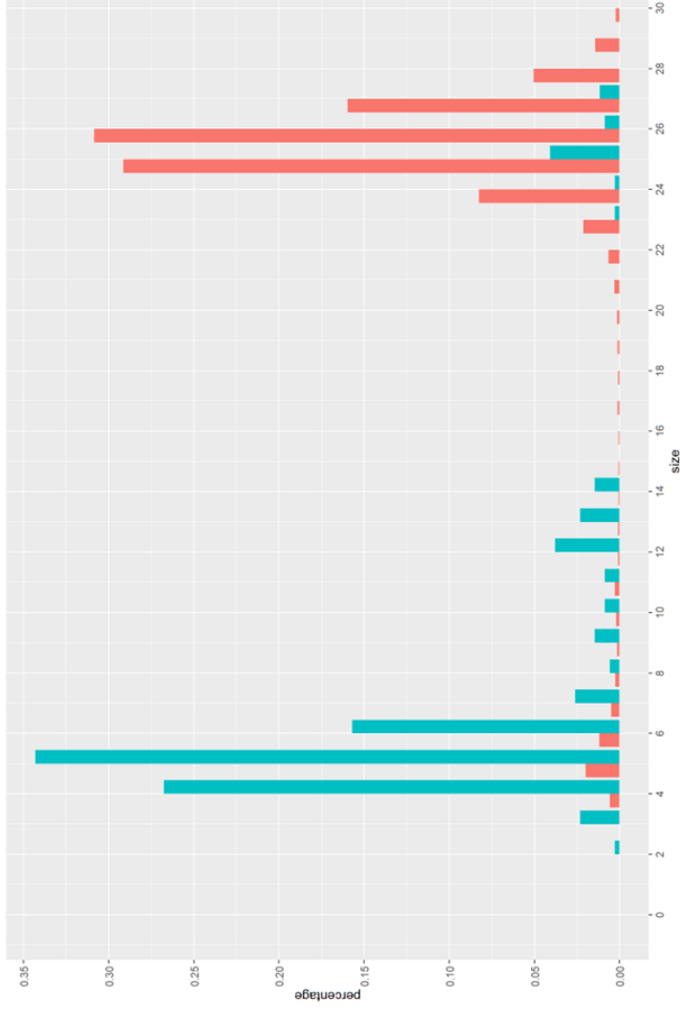
negatively correlating with the number of associated TEsRNA (pvalue ≤ 0.05), disparities between TE families were revealed (**Fig 6 C D**). A similar observation was made when grouping TE copies from each of the three most represented TE families (*i.e.* Mariner, Merlin and Ty3), by divergence percentage (**Fig 6 E**). Overall, these observations hinted at the existence of a small RNA mediated regulation of transposable elements in *A.algerae*, with varying efficacy depending on the TE family and the TE copie's sequence itself (sequence divergence).

4.4. Inversely correlating TEsRNA are mostly between 25 and 27 nucleotide long, and TEsRNA present a bias for a U residu at the 5' end.

The size and sequence composition of the TEsRNAs was investigated (**Fig 7**). Size distribution of the negatively correlating TEsRNAs was compared to that of the positively correlated TEsRNAs ($p < 0.05$) (**Fig 7 A**). Analysis showed a size range between 25 to 27 nt for most of the negatively correlated TEsRNAs, while the positively correlated TEsRNAs were mostly in the 5-7 nt range and may have corresponded to degradation products.

All TEsRNAs (regardless of correlation) presented a strong bias for a Uracil residue (Uracil) as the first nucleotide on the 5' end. This could also be observed for all other TEsRNAs studied (both negatively and positively correlated) (**Fig 7 B C D E F G**). A slight bias for cytosine/uracil at the 3' end was also observed for all 26 nt and 27 nt correlating TEsRNA (**Fig 7 C D, F G**). On the negatively correlated TEsRNAs, a slight bias for adenosine residues around bases 15 of the 27 nt long was also observed along with a TNNANNT pattern between bases 13 and 19 (**Fig 7 D**). For the positively correlating TEsRNAs no specific bias around bases 15 was found (**Fig 7 E F G**), instead, adenosine nucleotides were enriched on bases 11, 20 and 23 of the 27 nt long TEsRNA (**Fig 7 G**).

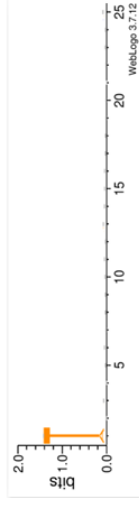
A.



Correlations
■ Positively correlating
■ Negatively correlating

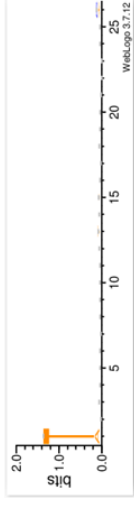
B.

25bp



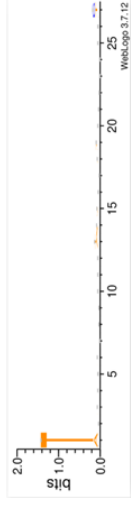
C.

26bp



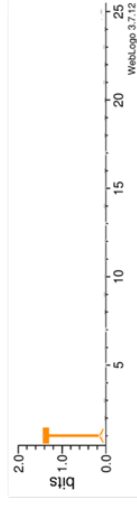
D.

27bp



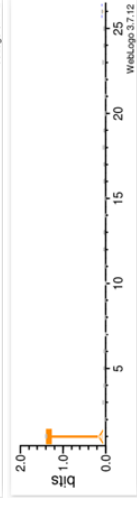
E.

25bp



F.

26bp



G.

27bp

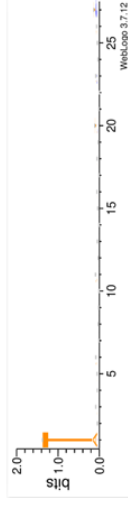


Figure 7 : Size distribution and nucleotide composition of the correlated TESRNAs

A. Size distribution percentages of the correlated TESRNAs. On top TESRNAs correlating with $p < 0.05$. The positively correlated TESRNAs are coloured in blue and the negatively correlated ones in red. Multiple alignment by size and Weblogo of negatively correlating (**B.C.D.**) and positively correlating (**E.F.G.**) for 25 26 and 27nt long TESRNAs.

3. DISCUSSION

In this study, the use of PacBio HiFi sequencing proved crucial to overcome the challenges that came with sequencing and studying an AT rich genome invaded by numerous repetitive elements (38,48,57). The high fidelity long reads generated were able to cross repetitive regions and yielded a new assembly of the genome of *A. algerae* at the chromosome scale. Ploidy seems to be an important aspect of microsporidian biology as a recent study on microsporidian genomes that polyploidy is widespread and variable in the Microsporidian phylum. This study also suggested tetraploidy for *A. algerae* (58). We confirm this observation with our data and interestingly, there doesn't seem to be any specific link between ploidy and the number of nuclei in microsporidia. Indeed, species with diplokaryon like *A. algerae* may be tetraploid but may also be diploid like *Vittaforma corneae*. Similarly, monokaryotic microsporidia like members of the *Glugeida* group may both be tetraploid like *Pseudoloma neurophilia* or diploid like *Cucumispora dikerogammari* (58). Therefore, Ploidy appears to be an evolutionary orientation rather than a consequence of the presence of two nucleus in some microsporidian species (39,58).

The karyotype of *A.algerae* and its genome length have remained a question in the field given difficulties assembling the genome and its repetitive elements (51,55,56). Here, we show a genome assembly size of about 12Mb for *A. algerae* with 17 Haploid Contigs (or chromosomes). Previous work on *A.algerae* tried to assess its karyotype using a Pulse Field gel electrophoresis (PFGE) approach (50). The study suggested 30 chromosomes, with a molecular band over 2.2 Mb for the largest one, which is almost one whole Mb bigger than our longest assembled contig (50). Failing to assemble such a large contig in our study seems surprising given the sequencing method and depth used (57). Also, the presence of TTAGGG



telomere repeats on most of our haploid contigs is evidence of probable full chromosome assembly. Rather, the results from PFGE could be explained by some bands of the molecular karyotype being composed of several chromosomes of similar sizes, that the PFGE was not able to separate.

The BUSCO analysis of our full-length and our haploid genome assemblies first revealed a completion value of only about 80%. However, a comparative approach by BLASTN analysis against the BUSCO microsporidia_odb10 database, showed that this value was in fact an underestimation, most likely due to the high evolutionary rate of microsporidian sequences (40,59). The actual completeness of the genomes was ultimately estimated at about 97%. Several clade specific gain/loss of function and protein adaptations exist in microsporidia (44,46,60). Due to these specific characteristics, assessing true completion of genome assemblies using dedicated software like BUSCO (61) (which still represents a tool of choice for this type of evaluation), is considerably challenging. And despite the large diversity of the microsporidian phylum (more than 200 genera 1,600 species (37)), the current BUSCO database (microsporidia_odb10 database) only contains information from 10 genomes, among which four *Encephalitozoon* species and four *Nematocida* (61). This issue could also be observed in phylogenetic analysis on *Vairimorpha (Nosema) ceranae* where microsporidian genera absent from the database often presented lower BUSCO completion scores (62). Orthologous sequences from the presented *A. algerae* genome will be added to the MicroAnnot tool to further enhance its annotation sensitivity and specificity; it could also be added to the BUSCO database to add more diversity to its pool of genomes. Several other new genomic resources for diverse microsporidian species are also rapidly being generated by colleagues in the field (47,63,64).



In depth annotation of the *A. algerae* assembly was also carried out in this study. The high-value annotation with the MicroAnnot tool, combined with our transcriptomic data allowed the identification of some previously unannotated core ncRNA housekeeping genes in *A. algerae*, namely, two U2 snRNA, two nuclear RNase P, an RNase MRP, and three 5S rRNA loci. Nevertheless, several ncRNAs including snRNA or snoRNAs remain absent from the genome annotation. Their identification remains a very challenging endeavour in microsporidia given the high sequence evolution rate (40,59), and it would require comparative genomic approaches with closely related microsporidian species similar to what has been done to identify the *Encephalytozoon cuniculi* U1 snRNA (44). With *A. algerae*, even using the somewhat close *Nosema* species, this identification was not possible.

This work also identified a large LncRNA repertoire in *A. algerae*. These molecules have been attributed epigenetic, regulatory, and genome integrity related roles (54). *Plasmodium falciparum*'s LncRNA have been shown to regulate the expression of *var* genes, which are crucial genes involved in the parasite's infectious biology (65). Some LncRNA in *Cryptosporidium parvum* have also been attributed important roles, as they can translocate into the host-cell's nucleus and induce epigenetic regulations of specific genes involved in cell motility for example (65,66). The large diversity encoded by *A. algerae* could also suggest important roles both in the parasite's biology, and in host-pathogen interactions.

We also identified transcripts considerably longer than the average CDS size in our transcriptome assemblies, despite the length of 5' and 3' UTR regions in microsporidia typically being short (less than 30 nt, and sometimes only 10 nt long or even completely absent for 5' UTRs) (51). When looking at the longest CDS associated transcripts in detail, we could observe it started 4150 nt upstream a CDS and ended 113 nt downstream that same



CDS. Overrun (read-through, or overlapping) transcription has been described in yeast, plants and humans (73–78) but also in numerous organisms with small genomes like the cryptomonad *Guillardia theta* and the microsporidian species *E. cuniculi* and *Antonospora locustae* (73,79,80). In the Microsporidian phylum, this phenomenon results from stochastic recognition of the polyadenylation signal (AAUAAA) (78). Overrun transcription has been proposed as a stress response to genomic stress (82) which could be consistent with TE activity in the genome of *A. algerae*. It may also help in the expression of sORF-encoded polypeptides (SEPs) (83). Unfortunately, overrun can induce artefactual assembly of individual transcripts, if the sequence of one of them overlaps with the sequence of the gene downstream. Nevertheless, the Trinity tool's assembly method is intended to filter out low level transcriptional read-through and noise when assembling a transcriptome. Microsporidian proteomics studies and quantitative proteomics would greatly help in understanding many of the transcriptional characteristics of these species (81). Direct transcript sequencing approaches using nanopore could also be key in better understanding this overrun phenomenon (84–86).

The TE diversity in the *A. algerae* genome was also analysed and was found amazingly high. This is concordant with previous reports (48). Here, we were able to identify 20 additional TE families in the genome, hence further illustrating this point. Comparing the PRA339 strain of *A. algerae* with our Undeen strain revealed good conservation of coding gene order and nucleotide sequence synteny, also suggesting strong conservation of TE insertion sites and diversity between the strains. This high degree of synteny in microsporidia is not only found between strains of the same species, but can also be found between species from the same



genus, as it was shown in three different species of the *Nosema* genus whose genomes all presented varying degrees of compaction (67,68). Strong coding gene order conservation has also been highlighted in different genomes of the *Nosema/Vairimorpha* clade even though the three species studied presented different levels of TE abundances when comparing (67). Finally, synteny between distantly related species such as *E. cuniculi*, *A. locustae* and *A. algerae* has also been reported, although with some CDS reorientations and intergenic sequences of different lengths (49). *A. algerae* presents a low coding gene density with about 1 gene every 4 Kb but with a homogeneous spread of CDS throughout its genome, which suggests a homogeneous lengthening of its intergenic regions throughout its evolution. TEs have been shown to target accessible sites where insertion does not impair host fitness (1,69), through this targeting, they also participate in the growth of intergenic regions, genome sizes and ncRNA repertoires (5,6,69). Thus, the homogeneous insertion of TEs throughout the genome of *A. algerae* may explain the conservation of synteny paired with the lengthening of intergenic regions.

In view of the large number of TEs, we looked for potential mechanisms of TE regulation in *A. algerae*. The adaptation of *A. algerae* to high TE content in its genome was most likely possible thanks to the acquisition of the Argonaute, Dicer and RdRP proteins: Transcriptomic analysis has shown widespread opposite strand complementary transcripts throughout the *A. algerae* genome. Although some ciliates like *Paramecium tetraurelia* and *Tetrahymena* have been described to almost fully transcribe both strands of their micronucleus genome as a way for TE elimination (28,29), such a process seems to require distinct nuclei, and several other cofactors and processes for the accurate elimination of TEs (28,29). Furthermore, these processes require PIWI family proteins in ciliates, which have not been found in the *A. algerae* genome. On the other hand, genomes like the one of *S. pombe* or *C. elegans* encode an RdRP



protein that generates complementary transcripts (23,70). These double stranded transcripts are then matured into siRNAs by Dicer, and by associating with Argonaute to form the RISC complex, they ensure post-transcriptional regulation and chromatin mediated regulation of TEs, repetitive sequences and coding genes (22,23). Dicer and Argonaute orthologue genes have been identified in *A. algerae* (53). An RdRP gene has now also been identified in the complete genome of *A. algerae* obtained in this study. This gene is present in several microsporidian species according to microsporidiaDB data (71) but it seems absent from genomes without TEs such as the ones of the *Encephalitozoon*. Based on our results and on the absence of the gene encoding this protein in microsporidian genomes lacking ET, we hypothesize that the RdRP protein is involved in the TE regulation process of *A. algerae* (41). We were also able to establish a link between some TE mapping small RNAs (TEsRNAs) and the expression of the corresponding TEs. Indeed, the expression levels of several TE copies were negatively correlated with the number of TEsRNAs associated with that copy reinforcing our general hypothesis.

The size the negatively correlated TEsRNAs centred around 25-27 bp, which is in agreement with the typical small interfering RNA size range (21). Positively correlated TEsRNAs were mostly found to be between 5 and 7 bases long, which could suggest them to be degradation products and/or randomly mapping RNA sequences. Finally, an uracil residue is found at the 5' end extremity of the vast majority of TEsRNA molecules was observed. This bias is typically seen following Dicer dependant maturation and nuclease activity (21,30,72).



4. CONCLUSION

To conclude, we explored the *A. algerae* genomic and transcriptomic landscapes, and investigated gene elements constituting this parasite adaptation to its high TE content. The data generated represents valuable genomic and transcriptomic assets in the fields of microsporidian studies and parasitology, which we hope will contribute to the community effort of investigating microsporidian genomics, TE dynamics, and genome evolution in eukaryotes. The genome assemblies and annotation files are available online as NCBI Bioproject Accession number PRJNA927239.

5. MATERIAL AND METHODS

Cell culture

A. algerae undeveloped isolates were cultured on confluent human foreskin fibroblast cells (HFF) as described previously (87). Released *A. algerae* spores were harvested twice a week by recovering the used media, until cells were completely lysed. The harvested spores were centrifuged 10 minutes at 1 000 g and rinsed once in molecular biological grade water and **twice** in PBS (1X). Lastly, Spores were resuspended in 10 mL of PBS (1X) and stored at 4°C until DNA extraction.



Genomic DNA extraction

Genomic DNA extraction was carried out on 250 million spores. Spores were treated twice with 25 units of Dnase I (Qiagen) 15 minutes at 25°C under agitation to eliminate human host cell DNA. Spores were collected by centrifugating 10 minutes at 1000 g and washed 2 times in PBS (1X). Genomic DNA from clean spores was extracted using the “genomic DNA from soil kit” (Macherey Nagel) according to the manufacturer’s protocol using SL1 buffer lysis and 165 µL of enhancer SX. Extracted DNA was eluted in 50 µL of molecular biological grade water. DNA was quantified in a Qubit 3 Fluorometer Invitrogen and the purity was evaluated with a nanodrop 2000 Thermo. The presence of unfragmented high molecular weight DNA molecules was ascertained on an Agilent Tape Station system 4150 with genomic DNA screen-tapes. The ratio of Human DNA contamination to *A. algerae* genomic DNA was measured by qPCR.

Library preparation and sequencing

Single-molecule Real-time long reads sequencing HiFiseq was performed at Gentyane Sequencing Platform (Clermont-Ferrand, France) with a PacBio Sequel II Sequencer (Pacific Biosciences, Menlo Park, CA, USA). The SMRTBell library was prepared using the SMRTbell prep kit3, following the “procedure and checklist -preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0” protocol. Genomic DNA (3 µg) was not sheared. Single-strand overhangs and any damage that may be present on the DNA backbone was repaired through the kit’s enzymes. An A-tailing reaction followed by the adapter ligation was conducted to generate the SMRT Bell templates. After nuclease treatment, the sample was



purified with 1X SMRTBell Clean Up Beads PB to obtain the final libraries with DNA fragments around 10 Kb. The SMRTBell library was quality inspected and quantified on a FemtoPulse (Agilent Technologies) and a Qubit fluorimeter with Qubit dsDNA HS reagent Assay kit (Life Technologies). A ready-to-sequence SMRTBell Polymerase Complex was created using a Binding Kit 3.2 (PacBio) and the Sequel II primer 3.2. The PacBio Sequel instrument was programmed to load a 90 pM library and sequenced in CCS mode on a PacBio SMRTCell 8M, with the Sequencing Plate 2.0 (Pacific Biosciences), 2 hours of pre-extension time and acquiring one 30-hour film per SMRTCell.

Assembly of the genomes

2.7 million CCS (circular consensus sequences), representing a total of 25 Gb Pacbio HiFi reads were used to perform a *de novo* genome assembly using Hicanu v2.1 with parameters minReadLength 3000, minOverlapLength 1000 (88). The assembled sequences were filtered to exclude contaminating human sequences (GC content greater than 40%). The remaining sequences were processed by Purge Haplotigs with default parameters and coverage thresholds of 200, 550 and 875 (89) to obtain a non-redundant “haploid” genome. Sequences classified as primary and haplotig by Purge Haplotigs were conserved. Sequences classified as repeat were manually curated to eliminate possible chimeric sequences. Finally, a BLASTN analysis against the nr database confirmed the absence of contaminating sequences in the final data set (90).



Genome assessment

The assembly metrics were computed by Quast (91). Complete assembly of *A. algerae* chromosomes was manually investigated by looking for telomeric repeats. Overall genome completeness was assessed using BUSCO (61) on the genome sequences with the microsporidia_odb10 database (61).

Genome annotation

The *A. algerae* genome annotation was performed using the microsporidian specific annotation pipeline MicroAnnot (92) which predicts protein coding genes and their putative functions, ribosomal RNAs, transfer RNAs and CDS corresponding to TEs.

Gene duplication was assessed on the haploid genome assembly using CD-HIT (93) to cluster homologous CDS based on at least 90% identity on 100% of the smaller sequence's length (**Supplementary figure 3A**).

Comparatives genomics quality control

To compare the *A. algerae* Undeen strain genome assembly to the one of *A. algerae* PRA 339 (GCA_000385875.2) considered as the current reference strain (56), annotation of PRA 339 using Microannot was performed. Next, MCscanX was used for synteny analysis between the two genome assemblies (94), and the Gene Ontology enrichment analysis of the two strains was computed and visualized using CirGO (95).

Genome ploidy analysis

The polyploid structure of the genome was estimated using GenomeScope v2 and smudgeplot (96) with jellyfish v 2.2.1 (97) and KMC v 3.2.1(98). For these analyses, only reads that mapped



on our final assembled genome with minimap2 v .22-r1122 were used (99). nQuire v a990a88 was also used as extra validation (100). Cd-hit was also used on all the predicted CDS sequences of the full genome assembly. Sequences with at least 90 % identity on 100% of the smallest sequence length were clustered (**supplementary figure 3B**).

Sequence divergence between homologous contigs was estimated using minimap2's long assembly to reference mapping (parameters -ax asm20). Additional contigs in the full genome assembly were mapped to the primary haploid genome, and the mismatch percentage was calculated between the aligned contigs and the corresponding primary contig using the pafR library (101) on the alignment file in ".paf" format. Boxplots were generated with ggplot2 (102).

Finally, the synteny between the primary haploid chromosomes and the different haplotypes was determined using MCscanX (94) (Blastp e-value $<1e-10$, and a syntenic block ≥ 10 CDS or ≥ 10 Kb for DNA sequence synteny).

Transposable element identification

Despite MicroAnnot allowing identification of predicted CDSs corresponding to TEs, a dedicated approach was implemented to scan the complete genome and identify unambiguously the complete repertoire of these elements. We compared the TEs identified by MicroAnnot to the library of TE consensus previously published for this species (103). Although most of the newly annotated TE sequences were present in the library, we identified 20 sequences corresponding to new TE families that we included in the library. The genome sequence of *A. algerae* was then searched for TE insertions using RepeatMasker (Smit, AFA, Hubley, R. & Green, P "RepeatMasker" at <http://www.repeatmasker.org>) using our custom TE library.



Transposable element divergence analysis

The main output file obtained from RepeatMasker was parsed using One_Code_to_Find_Them_All (Bailly-Béchet, Mobile DNA 2014) to recover each copy and their sequence divergence compared to the TE references, as well as the proportion of each family and class on the different chromosomes.

TE and CDS annotation

The haploid genome assembly gff annotation file from MicroAnnot was converted to BED format using the gff2bed tool (104). To avoid redundancy, TE annotation data generated by MicroAnnot was discarded and only the annotation data from the *de novo* TE identification were kept. TE descriptions (e.g. class, family, name) were included in the BED file.

Karyoplots were generated from the BED annotation file using the GenomicRanges and Rkaryoploter R packages (105,106).

Transcriptome assembly

Transcriptomes were derived of RNA-seq data from NCBI bioproject accession number PRJNA927239 (87). Data in this bioproject derives from a paired-end polyA+ and single-end small RNA sequencing of RNA from HFF cells infected by *A. algerae*. The experiment covered a full intracellular cycle of *A. algerae* with RNA extractions done in triplicates at 3h 12h 24h 48h and 72h post-infection. Detailed experimental design is available in Akossi et al. (87).

Assembly was done sample by sample using Trinity (107) on adaptor trimmed sequencing reads aligned to the *A. algerae* haploid genome. PolyA+ and smallRNA reads were separated.

PolyA+ reads were aligned using “—very-sensitive” bowtie2 in paired-end mode with default settings; and small RNAseq reads were aligned using “—very-sensitive” bowtie2

Trinity parameters used were --genome_guided_bam --genome_guided_max_intron 10.



PolyA+ and smallRNA assemblies were fused for each sample and mapped to the haploid genome once more using bowtie2 for ungapped alignment (same parameters as above were used). The resulting BAM files were converted to BED format using bedtools (108) and processed by gffcompare to generate a consensus transcriptome from all the samples (parameters: `--no-merge -D`, to disable close exon merging and to discard duplicate fragments in the query respectively) (109) This consensus transcriptome was consolidated by running gffread (96) with the parameters `-M` to group together redundant transcripts.

Timepoint specific transcriptomes were generated similarly but by running gffcompare and gffread on samples by timepoint.

The transcriptome metrics were computed using seqkit stats (110) on FASTA files of the different transcriptomes or on the raw assemblies (**Table 3** and **Supplementary table 3** respectively). Timepoint specific transcriptome metrics were computed similarly and are available in **Supplementary table 3**.

LncRNA identification and annotation

LncRNAs were defined as RNA transcripts over 200 nt (54) that do not correspond to coding sequences, or any other type of gene. Thus, in the consensus transcriptome, all transcripts that fully harboured or were fully harboured by any annotated regions in the annotation Microannot+TE BED file and on the same strand were discarded (bedtools intersect, parameters: `-F 1 -e -f 1 -s`).

The remaining transcripts were subjected to an Rfam analysis. Hit sequences were purged from the data set and separately added to the annotation BED file.



Remaining sequences less than 200 nt were discarded as “small ncRNA sequences” and sequences more than 200 nt were considered “Long noncoding RNA” sequences.

Putative LncRNAs were further classified regarding their genomic localisation: transcripts mapped on their whole length on the opposite strand of CDS features were classified as antisens LncRNAs (AS LncRNA, bedtools intersect parameters -f 1 -S). As 5' UTR regions of protein-coding genes are highly reduced (few nucleotides) or even absent in some microsporidia (73), LncRNAs located on the same strand and up to 200 nt upstream of CDS genes, and partially or not overlapping with that CDS gene, were classified as enhancer RNA (eRNA) (111) (bedtools window with -l 200 -r 0 -sw and -u). Remaining transcripts were classified as LincRNA (Long intergenic noncoding RNA).

Genome and transcriptome composition analysis

Overlapping same strand LncRNA coordinates were merged using bedtools merge (108) to avoid redundancy and to identify corresponding loci. The same was done for the Rfam identified transcripts. These annotated regions were then added to the Microannot+TE BED file.

The number of bases of the transcriptomic sequences that were associated with each type of annotation was calculated using bedtools intersect (108) on the reference transcriptome haploid genome alignment bed files. Transcripts had to be fully contained by, or fully contain a feature annotation to be associated with it (parameters: -F 1 -e -f 1 -s); in which case, all the nucleotides of the whole transcript were counted. This was done to account for the fact we do not know the exact function of the transcript, and that it might be involved in several of them.



PiRNA, siRNA and miRNA related protein identification

Orthologues for a custom list of several piRNA, siRNA and miRNA related proteins (17,22,112) were searched for inside *A. algerae* genome based on their interpro conserved domain annotations. (**Supplementary table 3**).

Argonaute, Dicer and RNA dependant RNA polymerase transcript expression analysis

The expression level of the Argonaute, Dicer and RNA dependant RNA polymerase genes from *A. algerae* was estimated using count data, based on the polyA+ paired end RNAseq data mapped to the haploid genome using hisat2 (--very-sensitive). Count matrix was generated using featureCounts (113) and normalised with Deseq2 (114). Multimapping reads were counted. (featureCount parameters: -s 1 -M --countReadPairs) and plots were generated using ggplot2 (102).

Argonaute, Dicer and RNA dependant RNA polymerase phylogenetic analysis

A protein clad specific phylogenetic tree of Argonaute family proteins was generated based on the amino acid sequence multiple alignment conducted using the Muscle algorithm (115) in the MEGA software (116). Using as a query a manually assembled database of AGO and Piwi clade proteins from various species and *A. algerae* Undeen identified Argonaute protein. The phylogenetic tree was generated through MEGA. Other phylogenetic trees were generated from *A. algerae*'s Argonaute, Dicer and RNA dependant RNA polymerase protein sequences, based on amino acid sequence similarities but this time using PSI-Blastp (117) against NCBI's nr database (or the clustered nr database for the RNA dependant RNA polymerase). PSI-blast e-value threshold was set at 0.005 (default). Trees were generated using NCBI blast tools online (results are available as **supplementary figure 4**).



Analysis of the correlation between the expression level of TEs and the number of associated small RNAseq reads.

For all infection kinetic point, the expression level of each TE copy (based on the polyA+ RNAseq data) and their number of associated small RNA reads were normalized using Deseq2 and a Spearman correlation test was performed.

Computation was done in R using the rcorr function and plots were generated using ggplot2 (102)

Correlated TEsRNA size distribution and weblogo

Correlated TEsRNA were separated by size and plotted using ggplot2 (102). WebLogos were generated for each size group using the WebLogo tool (118).



References

1. Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet.* 23 nov 2020;54:539-61.
2. Zhang HH, Peccoud J, Xu MRX, Zhang XG, Gilbert C. Horizontal transfer and evolution of transposable elements in vertebrates. *Nat Commun.* 13 mars 2020;11(1):1362.
3. Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. *Nat Rev Genet.* déc 2020;21(12):721-36.
4. Gasparotto E, Burattin FV, Di Gioia V, Panepuccia M, Ranzani V, Marasca F, et al. Transposable Elements Co-Option in Genome Evolution and Gene Regulation. *Int J Mol Sci.* 30 janv 2023;24(3):2610.
5. Platt RN, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* mars 2018;26(1-2):25-43.
6. Gilbert C, Peccoud J, Cordaux R. Transposable Elements and the Evolution of Insects. *Annu Rev Entomol.* 7 janv 2021;66:355-72.
7. Ali A, Han K, Liang P. Role of Transposable Elements in Gene Regulation in the Human Genome. *Life (Basel).* 4 févr 2021;11(2):118.
8. Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol.* juill 2022;23(7):481-97.
9. Kines KJ, Belancio VP. Expressing genes do not forget their LINES: transposable elements and gene expression. *Front Biosci.* 1 janv 2012;17:1329-44.
10. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet.* 2007;41:331-68.
11. Morales ME, Servant G, Ade C, Roy-Engel AM. Altering Genomic Integrity: Heavy Metal Exposure Promotes Transposable Element-Mediated Damage. *Biol Trace Elem Res.* juill 2015;166(1):24-33.
12. Burns KH. Transposable elements in cancer. *Nat Rev Cancer.* juill 2017;17(7):415-24.
13. Hsu PS, Yu SH, Tsai YT, Chang JY, Tsai LK, Ye CH, et al. More than causing (epi)genomic instability: emerging physiological implications of transposable element modulation. *J Biomed Sci.* 7 août 2021;28(1):58.

14. Nagirnaja L, Mørup N, Nielsen JE, Stakaitis R, Golubickaite I, Oud MS, et al. Variant PNLDC1, Defective piRNA Processing, and Azoospermia. *N Engl J Med.* 19 août 2021;385(8):707-19.
15. Russell SJ, Stalker L, LaMarre J. PIWIs, piRNAs and Retrotransposons: Complex battles during reprogramming in gametes and early embryos. *Reprod Domest Anim.* oct 2017;52 Suppl 4:28-38.
16. Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* juill 2019;20(7):417-31.
17. Onishi R, Yamanaka S, Siomi MC. piRNA- and siRNA-mediated transcriptional repression in *Drosophila*, mice, and yeast: new insights and biodiversity. *EMBO reports.* 5 oct 2021;22(10):e53062.
18. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annual Review of Biochemistry.* 2015;84(1):405-33.
19. Lisch D. Regulation of transposable elements in maize. *Curr Opin Plant Biol.* nov 2012;15(5):511-6.
20. Meyer E, Garnier O. Non-Mendelian inheritance and homology-dependent effects in ciliates. *Adv Genet.* 2002;46:305-37.
21. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* févr 2009;10(2):94-108.
22. Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, et al. An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell.* 2 nov 2017;21(5):694-703.e7.
23. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature.* 22 janv 2009;457(7228):413-20.
24. Wu J, Yang J, Cho WC, Zheng Y. Argonaute proteins: Structural features, functions and emerging roles. *J Adv Res.* 29 avr 2020;24:317-24.
25. Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, et al. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet.* 23 nov 2018;52:131-57.
26. Faehnle CR, Joshua-Tor L. Argonautes confront new small RNAs. *Curr Opin Chem Biol.* oct 2007;11(5):569-77.
27. Goto DB, Nakayama J. RNA and epigenetic silencing: Insight from fission yeast. *Dev Growth Differ.* janv 2012;54(1):129-41.
28. Allen SE, Nowacki M. Roles of Noncoding RNAs in Ciliate Genome Architecture. *J Mol Biol.* 10 juill 2020;432(15):4186-98.

29. Lepère G, Bétermier M, Meyer E, Duhaucourt S. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* 1 juin 2008;22(11):1501-12.
30. Mochizuki K, Kurth HM. Loading and pre-loading processes generate a distinct siRNA population in *Tetrahymena*. *Biochem Biophys Res Commun.* 5 juill 2013;436(3):497-502.
31. Spradling AC, Bellen HJ, Hoskins RA. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A.* 20 sept 2011;108(38):15948-53.
32. Leem YE, Ripmaster T, Kelly F, Ebina H, Heincelman M, Zhang K, et al. Retrotransposon Tf1 is targeted to pol II promoters by transcription activators. *Mol Cell.* 11 avr 2008;30(1):98-107.
33. Eickbush TH, Eickbush DG. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol Spectr.* avr 2015;3(2):10.1128/microbiolspec.MDNA3-0011-2014.
34. Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui MA, Guy J, et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaptation. *Nat Commun.* 31 juill 2019;10:3421.
35. Han B, Pan G, Weiss LM. Microsporidiosis in Humans. *Clin Microbiol Rev.* 34(4):e00010-20.
36. Stentiford GD, Becnel J. J., Weiss LM, Keeling PJ, Didier ES, Williams B. AP, et al. Microsporidia – Emergent Pathogens in the Global Food Chain. *Trends in Parasitology.* 1 avr 2016;32(4):336-48.
37. Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends in Parasitology.* 1 août 2022;38(8):642-59.
38. Peyretilade E, Boucher D, Parisot N, Gasc C, Butler R, Pombert JF, et al. Exploiting the architecture and the features of the microsporidian genomes to investigate diversity and impact of these parasites on ecosystems. *Heredity (Edinb).* mai 2015;114(5):441-9.
39. Han B, Weiss LM. Microsporidia: Obligate Intracellular Pathogens within the Fungal Kingdom. *Microbiol Spectr.* avr 2017;5(2):10.1128/microbiolspec.FUNK-0018-2016.
40. Wadi L, Reinke AW. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLOS Pathogens.* 13 févr 2020;16(2):e1008276.
41. Peyretilade E, El Alaoui H, Diogon M, Polonais V, Parisot N, Biron DG, et al. Extreme reduction and compaction of microsporidian genomes. *Research in Microbiology.* 1 juill 2011;162(6):598-606.
42. Barandun J, Hunziker M, Vossbrinck CR, Klinge S. Evolutionary compaction and adaptation visualized by the structure of the dormant microsporidian ribosome. *Nat Microbiol.* nov 2019;4(11):1798-804.

43. Polonais V, Prensier G, Méténier G, Vivarès CP, Delbac F. Microsporidian polar tube proteins: highly divergent but closely linked genes encode PTP1 and PTP2 in members of the evolutionarily distant Antonospora and Encephalitozoon groups. *Fungal Genet Biol.* sept 2005;42(9):791-803.
44. Belkorchia A, Pombert JF, Polonais V, Parisot N, Delbac F, Brugère JF, et al. Comparative genomics of microsporidian genomes reveals a minimal non-coding RNA set and new insights for transcription in minimal eukaryotic genomes. *DNA Res.* juin 2017;24(3):251-60.
45. Cali A, Takvorian PM. Developmental Morphology and Life Cycles of the Microsporidia. In: *Microsporidia* [Internet]. John Wiley & Sons, Ltd; 2014 [cité 3 juin 2023]. p. 71-133. Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118395264.ch2>
46. Williams TA, Nakjang S, Campbell SE, Freeman MA, Eydal M, Moore K, et al. A Recent Whole-Genome Duplication Divides Populations of a Globally Distributed Microsporidian. *Mol Biol Evol.* août 2016;33(8):2002-15.
47. Mascarenhas dos Santos AC, Julian AT, Liang P, Juárez O, Pombert JF. Telomere-to-Telomere genome assemblies of human-infecting Encephalitozoon species. *BMC Genomics.* 4 mai 2023;24:237.
48. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol Evol.* 28 août 2014;6(9):2289-300.
49. Williams BA, Lee RC, Becnel JJ, Weiss LM, Fast NM, Keeling PJ. Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. *BMC Genomics.* 29 avr 2008;9:200.
50. Belkorchia A, Biderre C, Militon C, Polonais V, Wincker P, Jubin C, et al. 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. *Parasitology International.* 1 mars 2008;57(1):62-71.
51. Peyretilade E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E, et al. Annotation of microsporidian genomes using transcriptional signals. *Nat Commun.* 16 oct 2012;3(1):1-9.
52. Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun.* 21 sept 2010;1:77.
53. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infection, Genetics and Evolution.* 1 nov 2018;65:329-32.

54. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol.* juin 2023;24(6):430-47.
55. Data Set *Anncaliia algerae* PRA109 Genome Sequence and Annotation [Internet]. [cité 4 juin 2023]. Disponible sur: https://microsporidiadb.org/micro/app/record/dataset/NCBITAXON_1240240
56. Data Set *Anncaliia algerae* PRA339 Genome Sequence and Annotation [Internet]. [cité 4 juin 2023]. Disponible sur: https://microsporidiadb.org/micro/app/record/dataset/DS_31f5bcabba
57. Marx V. Method of the year: long-read sequencing. *Nat Methods.* janv 2023;20(1):6-11.
58. Khalaf A, Jaron KS, Blaxter ML, Lawniczak MKN. Polyploidy is widespread in Microsporidia [Internet]. *Evolutionary Biology*; 2023 sept [cité 4 nov 2023]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2023.09.29.560119>
59. Williams BAP, Williams TA, Trew J. Comparative Genomics of Microsporidia. *Exp Suppl.* 2022;114:43-69.
60. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, et al. Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. *Genome Biol Evol.* 2013;5(12):2285-303.
61. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol.* 2019;1962:227-45.
62. Huang Q, Wu ZH, Li WF, Guo R, Xu JS, Dang XQ, et al. Genome and Evolutionary Analysis of *Nosema ceranae*: A Microsporidian Parasite of Honey Bees. *Frontiers in Microbiology* [Internet]. 2021 [cité 14 nov 2023];12. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.645353>
63. Angst P, Pombert JF, Ebert D, Fields PD. Near chromosome-level genome assembly of the microsporidium *Hamiltosporidium tvaerminnensis*. *G3 Genes|Genomes|Genetics.* 1 oct 2023;13(10):jkad185.
64. Xiong X, Geden CJ, Bergstralh DT, White RL, Werren JH, Wang X. New insights into the genome and transmission of the microsporidian pathogen *Nosema muscidifuracis*. *Front Microbiol.* 13 avr 2023;14:1152586.
65. Li Y, Baptista RP, Kissinger JC. Noncoding RNAs in Apicomplexan Parasites: An Update. *Trends Parasitol.* oct 2020;36(10):835-49.
66. Li Y, Baptista RP, Sateriale A, Striepen B, Kissinger JC. Analysis of Long Non-Coding RNA in *Cryptosporidium parvum* Reveals Significant Stage-Specific Antisense Transcription. *Frontiers in Cellular and Infection*

- Microbiology [Internet]. 2021 [cité 21 nov 2023];10. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fcimb.2020.608298>
67. The Genome of *Nosema* sp. Isolate YNPr: A Comparative Analysis of Genome Evolution within the *Nosema/Vairimorpha* Clade | PLOS ONE [Internet]. [cité 14 nov 2023]. Disponible sur: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0162336>
68. Roudel M, Aufauvre J, Corbara B, Delbac F, Blot N. New insights on the genetic diversity of the honeybee parasite *Nosema ceranae* based on multilocus sequence analysis. *Parasitology*. sept 2013;140(11):1346-56.
69. Kannan S, Chernikova D, Rogozin IB, Poliakov E, Managadze D, Koonin EV, et al. Transposable Element Insertions in Long Intergenic Non-Coding RNA Genes. *Frontiers in Bioengineering and Biotechnology* [Internet]. 2015 [cité 4 nov 2023];3. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fbioe.2015.00071>
70. Hoogstrate SW, Volkers RJ, Sterken MG, Kammenga JE, Snoek LB. Nematode endogenous small RNA pathways. *Worm*. 5 mars 2014;3:e28234.
71. Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res*. janv 2011;39(Database issue):D612-619.
72. Hoehener C, Hug I, Nowacki M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell*. 22 mars 2018;173(1):234-247.e7.
73. Peyretailade E, Gonçalves O, Terrat S, Dugat-Bony E, Wincker P, Cornman RS, et al. Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: support for accurate structural genome annotation. *BMC Genomics*. 15 déc 2009;10:607.
74. Maspero D, Dassano A, Pintarelli G, Noci S, De Cecco L, Incarbone M, et al. Read-through transcripts in lung: germline genetic regulation and correlation with the expression of other genes. *Carcinogenesis*. 14 juill 2020;41(7):918-26.
75. Čermák V, Fischer L. Pervasive read-through transcription of T-DNAs is frequent in tobacco BY-2 cells and can effectively induce silencing. *BMC Plant Biol*. 22 oct 2018;18(1):252.
76. Candelli T, Challal D, Briand JB, Boulay J, Porrua O, Colin J, et al. High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *The EMBO Journal*. 15 févr 2018;37(4):e97490.

77. Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, et al. Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription. *Molecular Cell*. 4 déc 2014;56(5):667-80.
78. He Y, Yuan C, Chen L, Lei M, Zellmer L, Huang H, et al. Transcriptional-Readthrough RNAs Reflect the Phenomenon of “A Gene Contains Gene(s)” or “Gene(s) within a Gene” in the Human Genome, and Thus Are Not Chimeric RNAs. *Genes (Basel)*. 16 janv 2018;9(1):40.
79. A high frequency of overlapping gene expression in compacted eukaryotic genomes | PNAS [Internet]. [cité 17 nov 2023]. Disponible sur: <https://www.pnas.org/doi/full/10.1073/pnas.0501321102>
80. Corradi N, Gangaeva A, Keeling PJ. Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics*. avr 2008;91(4):388-93.
81. Rozanova S, Barkovits K, Nikolov M, Schmidt C, Urlaub H, Marcus K. Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol Biol*. 2021;2228:85-116.
82. Vilborg A, Steitz JA. Readthrough transcription: How are DoGs made and what do they do? *RNA Biol*. 9 févr 2016;14(5):632-6.
83. Chu Q, Ma J, Saghatelian A. Discovery and Characterization of A Novel Class of Functional Polypeptides. *Crit Rev Biochem Mol Biol*. 2015;50(2):134-41.
84. Schulz L, Torres-Diz M, Cortés-López M, Hayer KE, Asnani M, Tasian SK, et al. Direct long-read RNA sequencing identifies a subset of questionable exons likely arising from reverse transcription artifacts. *Genome Biology*. 28 juin 2021;22(1):190.
85. Shi X, Qin F, Li H. Confirmation of Transcriptional Read-Through Events by RT-PCR. *Methods Mol Biol*. 2020;2079:177-86.
86. Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. *Nat Methods*. oct 2022;19(10):1160-4.
87. Akossi RF, Delbac F, El Alaoui H, Wawrzyniak I, Peyretailade E. The intracellular parasite *Anncalia algerae* induces a massive miRNA down-regulation in human cells. *Non-coding RNA Research*. 1 sept 2023;8(3):363-75.
88. Nurk S, Walenz BP, Rhie A, Vollger MR, Logsdon GA, Grothe R, et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res*. sept 2020;30(9):1291-305.

89. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 29 nov 2018;19(1):460.
90. BLAST: Basic Local Alignment Search Tool [Internet]. [cité 2 nov 2023]. Disponible sur: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
91. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 15 avr 2013;29(8):1072-5.
92. Tournayre J, Polonais V, Wawrzyniak I, Akossi RF, Parisot N, Lerat E, et al. MicroAnnot: A Dedicated Workflow for Accurate Microsporidian Genome Annotation [Internet]. Preprints; 2023 [cité 30 oct 2023]. Disponible sur: <https://www.preprints.org/manuscript/202310.1822/v1>
93. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences - PubMed [Internet]. [cité 2 nov 2023]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/16731699/>
94. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. avr 2012;40(7):e49.
95. CirGO: an alternative circular way of visualising gene ontology terms | BMC Bioinformatics | Full Text [Internet]. [cité 2 nov 2023]. Disponible sur: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2671-2>
96. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 18 mars 2020;11(1):1432.
97. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 15 mars 2011;27(6):764-70.
98. Kokot M, Długosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*. 1 sept 2017;33(17):2759-61.
99. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 15 sept 2018;34(18):3094-100.
100. Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*. 4 avr 2018;19(1):122.
101. Read, manipulate and visualize Pairwise mapping Format data [Internet]. [cité 2 nov 2023]. Disponible sur: <https://dwinter.github.io/pafr/>
102. ggplot2: Elegant Graphics for Data Analysis | SpringerLink [Internet]. [cité 2 nov 2023]. Disponible sur: <https://link.springer.com/book/10.1007/978-0-387-98141-3>

103. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. *Genome Biology and Evolution*. 1 sept 2014;6(9):2289-300.
104. BEDOPS: high-performance genomic feature operations | Bioinformatics | Oxford Academic [Internet]. [cité 2 nov 2023]. Disponible sur: <https://academic.oup.com/bioinformatics/article/28/14/1919/218826>
105. Software for Computing and Annotating Genomic Ranges | PLOS Computational Biology [Internet]. [cité 2 nov 2023]. Disponible sur: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003118>
106. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. 1 oct 2017;33(19):3088-90.
107. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 15 mai 2011;29(7):644-52.
108. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 15 mars 2010;26(6):841-2.
109. Perteza G, Perteza M. GFF Utilities: GffRead and GffCompare. *F1000Res*. 2020;9:ISCB Comm J-304.
110. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One*. 2016;11(10):e0163962.
111. Kim TK, Hemberg M, Gray JM. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harb Perspect Biol*. 5 janv 2015;7(1):a018622.
112. Simonelig M. Developmental functions of piRNAs and transposable elements. *RNA Biol*. 2011;8(5):754-9.
113. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 1 avr 2014;30(7):923-30.
114. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
115. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 19 août 2004;5(1):113.
116. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol*. mai 2013;30(5):1229-35.

117. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1 sept 1997;25(17):3389-402.
118. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* juin 2004;14(6):1188-90.

A travers cette étude nous avons pu identifier la séquence des 17 chromosomes d'*A.algerae* et montrer la tétraploidie de l'espèce. L'annotation de ce génome et son étude transcriptomique a mis en lumière 260 familles d'ET et un large répertoire de transcrits non-codants. Des petits ARN correspondants à des séquences d'ET ont pu être caractérisés et ils pourraient jouer un rôle dans la régulation de ces éléments. La caractérisation et la confirmation de ce mécanisme de régulation demanderaient une étude complémentaire dédiée.

Dans le cadre de ce projet de thèse et par rapport aux données générées, nous nous sommes ensuite intéressés aux autres types de petits ARNnc qui pourraient être codés par *A.algerae*, et quels rôles ceux-ci pourraient jouer dans les interactions entre le parasite et sa cellule hôte.



PARTIE III : ROLE ET ORIGINE DE PETITS ARNS INTERFERENTS D'*ANNCALIIA ALGERAE* DANS L'INTERACTION AVEC SA CELLULE HOTE

Les petits ARN non codants peuvent être impliqués dans de nombreux processus biologiques et certains d'entre eux ont été caractérisés comme impliqués dans les interactions entre cellules hôtes et parasites (203,210,227,272,348). C'est un groupe hétérogène composé de plusieurs types de molécules aux mécanismes de maturation et d'action variés (168). Afin d'assurer leur prédiction chez *A. algerae* une approche *in silico* basée sur la structure des locus et leurs les profils d'expression a été mise en œuvre. Le rôle potentiel des ARN interférent identifiés a été exploré, en particulier dans le cadre des interactions directes avec la cellule hôte.

Ainsi, leur transfert du parasite vers la cellule hôte a été caractérisé par fluorescence *in situ* hybridization (FISH), et des cibles potentielles ont été prédites en combinant une approche de prédiction *in silico* et l'étude du profil d'expression des cibles tout au long de l'infection.

Ces ARN interférent représenteraient un mécanisme de régulation de fonctions de l'hôte, récemment acquis dans l'évolution du parasite.

* « Supplementary data » disponibles dans les annexes correspondantes

#Article en preparation#



Identification and role of small interfering RNA in the microsporidian parasite *Anncaliia algerae* during an infectious process

Short title: *Anncaliia algerae*'s genomic and transcriptomic landscape

Authors: Reginald Florian Akossi^{1*}; Frédéric Delbac¹; Ivan Wawrzyniak ¹; and Eric Peyretailade^{*1}.

¹Laboratoire "Microorganismes : Génome et Environnement" (LMGE), UMR 6023 Université Clermont Auvergne and CNRS, F-63000 Clermont-Ferrand, France.

E-mails :

Reginald.Akossi@uca.fr / Reginald.F.Akossi@gmail.com

Ivan.WAWRZYNIAK@uca.fr

Eric.PEYRETAILLADE@uca.fr



Introduction

Interfering RNA (iRNA) are small non coding RNA (ncRNA) molecules involved in the epigenetic and post-transcriptional regulation of transcripts in eucaryotes (1). This regulation is a sequence specific mechanism and is exerted through three main types of small ncRNA: small interfering RNA (siRNA), microRNA (miRNA) and PIWI-interacting RNA (piRNA) (1). Each of these molecules follows defined biogenesis and interacts with different protein partners to exert their regulatory function (1,2). These small ncRNAs all require a type III RNase for their biogenesis (Dicer or Dicer-like proteins) and a protein from the Argonaute family to exert their regulatory function. (1,3–5). The type III RNases process the ncRNA precursors into smaller sequences typically between 18 and 30 nucleotides long with about 2 nucleotide overhangs on each side (1,4,6). They present specific affinity for double stranded RNA and preferentially cleave near uridine nucleotides (6). The Argonaute protein family is defined by the presence of a conserved PAZ (PIWI-Argonaute-zwille) and a PIWI (P-element-induced wimpy testis) domain. They link the matured small double stranded RNA through recognition of the 3' end overhang by the PAZ domain, and associate with other proteins to form the RNA-Induced Silencing Complex (RISC) (7). One of the strands of the dsRNA will be released (the star (*) RNA), and the RISC complex will keep the other strand (the mature iRNA) and go on to exert sequence specific post-transcriptional or epigenetic regulation, in the cytoplasm or the nucleus respectively (7,8). The type of small RNA, the Argonaute protein, as well as the other proteins in the RISC complex all influence the regulatory action (4,9). Thus, piRNA RISC complexes are typically involved in regulation of transposable elements (TEs) and repeat elements at the epigenetic and post-transcriptional levels (4). miRNA RISC ensures the fine regulation of numerous messenger RNA (mRNA) genes in various species and intercellular



communication through extracellular vesicles (EVs) are ensured by miRNA (8,10,11). siRNA can either be derived from exogenous or viral RNA or be endogenous (12). In both cases, siRNA RISC can be involved in various epigenetic and post-transcriptional regulation processes (2,5,13,14).

Overlapping roles have been described between siRNA miRNA and piRNA in different biological contexts and species (1). For example, piRNA are animal specific and are typically found expressed only in germline cells (4). In animal somatic cells, and in organisms without piRNA like the yeast *Schizosaccharomyces pombe*, some siRNAs have been observed to regulate TE activity (13,14). Similarly, in species without canonical miRNA, miRNA-like molecules and/or siRNA have been predicted to exert similar regulatory action on mRNA expression (15–21).

In host parasite interactions, cross kingdom regulation by interfering RNAs have also been described (22–24). While *Plasmodium* species do not code miRNA themselves, parasite regulation by host miRNA have been described (25). The parasitic worm *Schistosoma Japonicum* has been shown to release EVs containing small ncRNA with regulatory roles on host macrophages (26,27). Similarly, *Trypanosoma cruzi* ncRNA have been shown to regulate Hela cell *in vitro* (28,29) and *Leishmania spp* small ncRNA have been shown to control host macrophage gene expression (21).

In microsporidia of the *Nosema* genera, miRNA-like molecules have been described (17,30,31). These molecules have been predicted to be involved in host-parasite interaction. In one instance, this interaction has been directly proven, with the miRNA-like molecule Nb-miR8 able to regulate host peroxisome protein BmPEX16 during the infection of the butterfly



Bombyx mori by the microsporidia *Nosema bombycis*. This interaction led to increased parasite proliferation (31).

Microsporidia are obligatory intracellular parasites related to fungi (32). They represent a large group composed of over 1600 different species separated into about 218 genera and 45 families (33). These species are diverse both in their genomic features, and their ecological niches and host spectrums (32,33). Microsporidia exhibit different genomic characteristics between genera and species. Some species like the *Encephalitozoon* have evolved into a reduced and compact genome, exhibiting highly reduced intergenic regions, very few introns, and smaller orthologous coding sequences (CDS) when compared to other unicellular eucaryotes like baker's yeast (34,35). This typically leads to species with smaller genomes and currently, *E.intestinalis* harbours the smallest known eucaryotic haploid genome, with 2.3 Mb (36). Some other microsporidia such as *Anncaliia algerae* or *Edhazardia aedis* contrast with this trend and exhibit larger, gene sparse genomes (12Mb and 51.3Mb respectively) (37). It should also be pointed out that the *A. algerae* genome is tetraploid (cf. previous chapter), unlike the genome of the *Encephalitozoon* species which are diploid (38) difference with the *Encephalitozoons* seems to be linked to the expansion of (TEs but also of non-coding genes (37,39–41). Recently, through an in-depth characterisation of *A. algerae*'s genomic and transcriptomic content, our group highlighted the diversity of TE and LncRNA genes in *A. algerae* (cf. previous chapter). This study reported 260 TE families in *A. algerae* but also a large noncoding RNA (LncRNA) gene repertoire with 5894 LncRNA transcripts identified during an infection kinetic.

The *A. algerae* specie is also remarkable for its adaptive abilities, which enable it to infect distant human and anopheles species (32,42). *In vitro*, It has also shown the ability to



develop in various cell types from various species (43). While some factors governing microsporidian host range, host adaptation and host expansion are known (42, 44) several aspects remain unclear (44). In this work, we hypothesised that some siRNA may exist in *A. algerae* genome, and that some of these noncoding genes may have regulatory effects on host cells and be involved in *A. algerae* adaptation to new hosts.

To better decipher iRNA processes and implications in *A. algerae* host-parasite interactions, iRNA loci from the *A. algerae* genome were predicted using both complete genome sequences obtained by PacBio HiFi sequencing (cf previous paper) and expression data of small RNA molecules produced throughout an infectious cycle. We used a strategy combining differential expression analysis and *in silico* target prediction to identify biologically relevant iRNA-Targets pairs between *A. algerae* coded iRNA and host cell mRNA and verified the localisation and transfer towards the host cell of the iRNA using fluorescence *in situ* hybridisation (FISH).

Table 1 : Summary of iRNA candidates

	Number of precursor loci	Number of different genomic contigs	mature sequence length	mature sequence	mature sequences weblogo	Average star sequence length	Average precursor sequence length	transcript in consensus transcriptomes associated with precursor (mapping)	Other species with blastn sequence hits for the precursor loci (e< 1e-3)
iRNA1	14	13	25	uuuggaccagcgaaggaggaggaggaga		25.4	60.9	TCONS_00004172 (LncRNA) [14]	PRA339 PRA109
iRNA2	4	4	25	uucagatcugaauugcuuucccacc		24.75	75.5	-	PRA339 PRA109
iRNA3	2	2	25	uacguacgagacuuuuuggggcc		25	58	-	PRA339 PRA109
iRNA4	2	2	25	uucucgggacacauuugaaaggaa		25	68	-	PRA339 PRA109
iRNA5	1	1	25	ugauuggagggaauaaaggcaucu		24	82	-	PRA339 PRA109
iRNA6	1	1	25	uacauacacaaagacuaccuc		25	59	-	PRA339 PRA109
iRNA7	1	1	25	uuugaacuaagagcctuggccacc		27	62	-	PRA339 PRA109 Tubulosema_ratisbonensis_Franzen
iRNA8	3	3	25	uagaggaggaggaggaggauacuu		25	56.3	-	Cucumispora_dikergammari_Dv6
iRNA9	3	3	25	ugcauaguuaucgaggauugggc		25.5	66.3	-	PRA339 PRA109
iRNA10	1	1	25	uguuuccuuuguaucucugguacucu		24	81	-	PRA339 PRA109
iRNA11	2	2	25	uacagagaauuuuuaaggaggaga		25	76	-	PRA339 PRA109

Results

Analysis of iRNA precursor loci

Out of the 2025 mirDeep2 hits, 37 iRNA precursor sequences were selected. These precursors encoded 11 different mature iRNA sequences and were named as iRNA-1 to iRNA-11. To further differentiate iRNA precursor loci, a letter was added from A to Z. Three pairs of iRNA precursor loci overlapped over more than 95% of their length (iRNA-9A and iRNA-9E, iRNA-9B and iRNA-9D, iRNA-9C and iRNA-9F). These precursors were fused and renamed iRNA-9A/E, iRNA-9B/D and iRNA-9C/F respectively. This made for a total of 34 different iRNA precursor loci. The precursors presented sizes between 58 and 82 nucleotides, and the corresponding mature iRNA sequences were all 25 nucleotides long, with an associated star sequence between 24 and 27 nucleotides (**table 1**). Of the 11 mature iRNA, precursors for iRNA-1 were by far the most represented in the genome with 14 loci. They were followed by iRNA-9 and iRNA-2 precursors with 4 loci each, iRNA-8 precursors with 3 loci and iRNA-3 and iRNA-4 precursors with 2 loci each. The last 4 iRNA (iRNA-5, iRNA-6, iRNA-7 and iRNA-10) presented a single precursor locus in the genome. All Mature iRNA sequences presented a bias for a uridine residue at the 5' end. Some other biases were also observed on nucleotide positions 3, 6-7, 11, 17-18, 21-22 and 25 but were not present in all sequences (**table 1**). Using a transcriptome of *A. algerae* as reference, we could identify that the iRNA-1 precursor sequences mapped to a single transcript TCONS_00004172, which is 361 nt long and had been annotated as a LncRNA transcript. A comparative analysis showed that all iRNA precursor were found on the genome of the other *A. algerae* strains (PRA339 strain and PRA109 strain). iRNA-7 also presented significant homology with sequences from *Dictyocoela muelleri* strain

A.

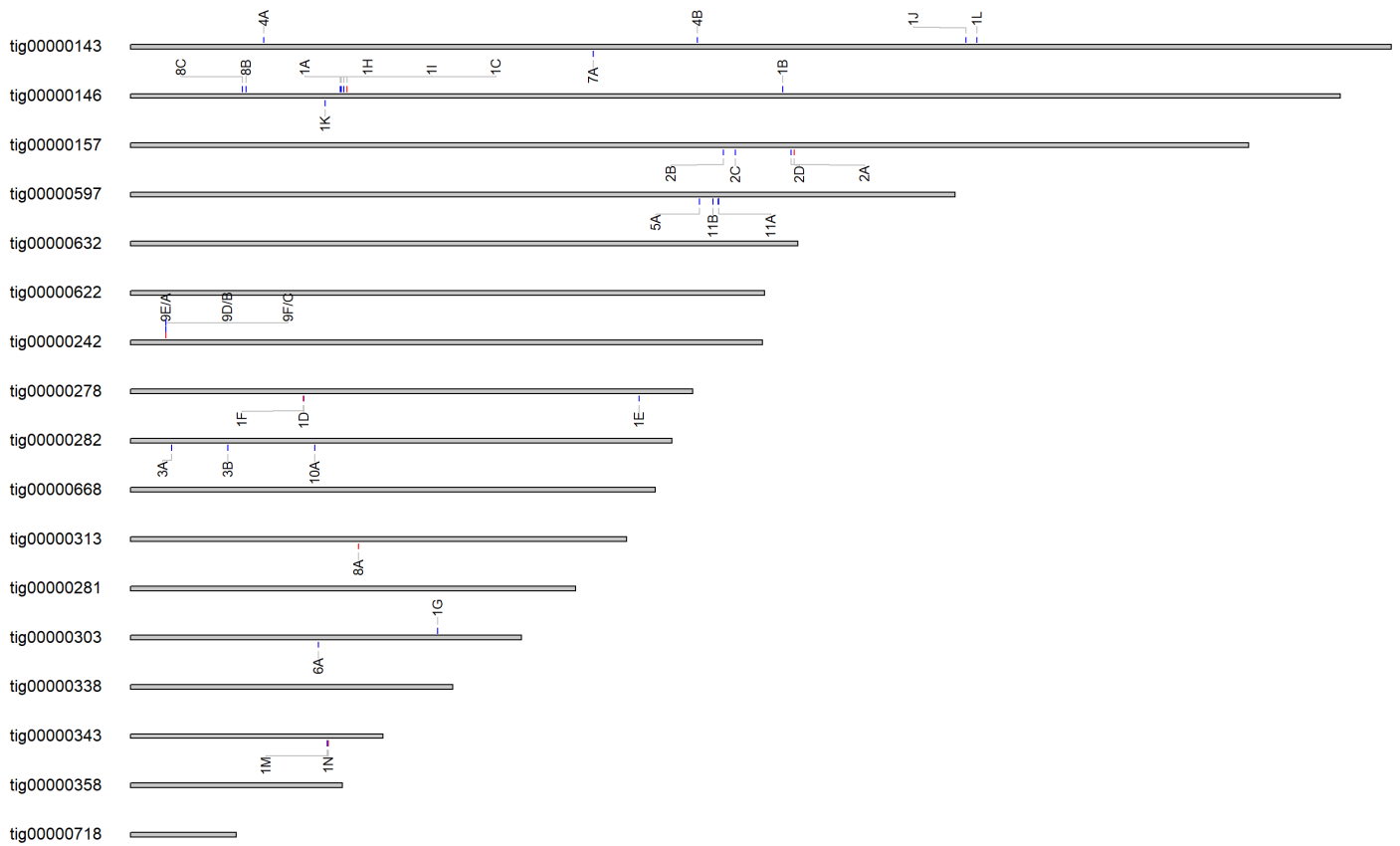


Fig 1 : Structure, genomic organization and expression profile of the 34 small interfering RNA precursor genes

A. Position of the 34 iRNA precursor genes on the *A. algae* haploid genome. Labels correspond to the iRNA identifier with the letter indicating the mature sequence and the letter the locus the precursor was found in. Coloured in red are iRNA precursor loci present on the haploid genome and in blue those that correspond to homologous sequences in the assembly that aligned at the given position.

Ou54, *Tubulinosema ratisbonensis* strain Franzen and *Cucumispora dikerogammari* strain Dv6 (**table 1**). iRNA precursor loci were plotted on the haploid genome of *A. algerae* (**Fig1A**). *A. algerae* is actually tetraploid so iRNA loci found on homologous chromosomes were indicated where they would align on the haploid genome. iRNA precursors presented slight positional shift between the chromosome in the haploid genome and its homologues. Only iRNA 9 precursors was found perfectly aligned between all the homologue chromosomes.

iRNA loci were found on 10 of the 17 different chromosomes in *A. algerae*. Loci coding the same mature iRNA sequence were found solely on the same chromosomes, except for iRNA-1 loci which were found on 5 different chromosomes. Analysis of the gene environment of the RNAi precursors showed that most of them localized in within 500 nt of TEs, with the majority of them (iRNA-1, iRNA-5, iRNA-6, iRNA-7 and 10) even overlapping these TE sequences on the opposite strand. For example, iRNA-1 precursors were found systematically opposite strand of Ty3 retrotransposon elements mainly around the 3' end of the TE. iRNA-7_A and iRNA6_A were each present opposite strand of a strongly divergent Merlin DNA transposon (over 20 % divergence), and in the middle of the TE sequence. All the other TEs identified near iRNA loci had divergence percentages under 15% suggesting relatively recent insertions. Only precursors for iRNA-8, iRNA- 9 and iRNA-11 lacked TE sequences within 500 nt of their loci. We looked at the sequence diversity between the iRNA precursors and found that all the precursors coding the same mature iRNA could be grouped together (**Fig1B**). iRNA-1 sequences still formed two distinct subgroups suggesting slightly higher polymorphism for the iRNA-1 precursors. Finally, the tree did not reveal any clear separation between iRNA precursors that were opposite strand of DNA transposons and the precursors opposite strand retrotransposons (**Fig1B**).

B.

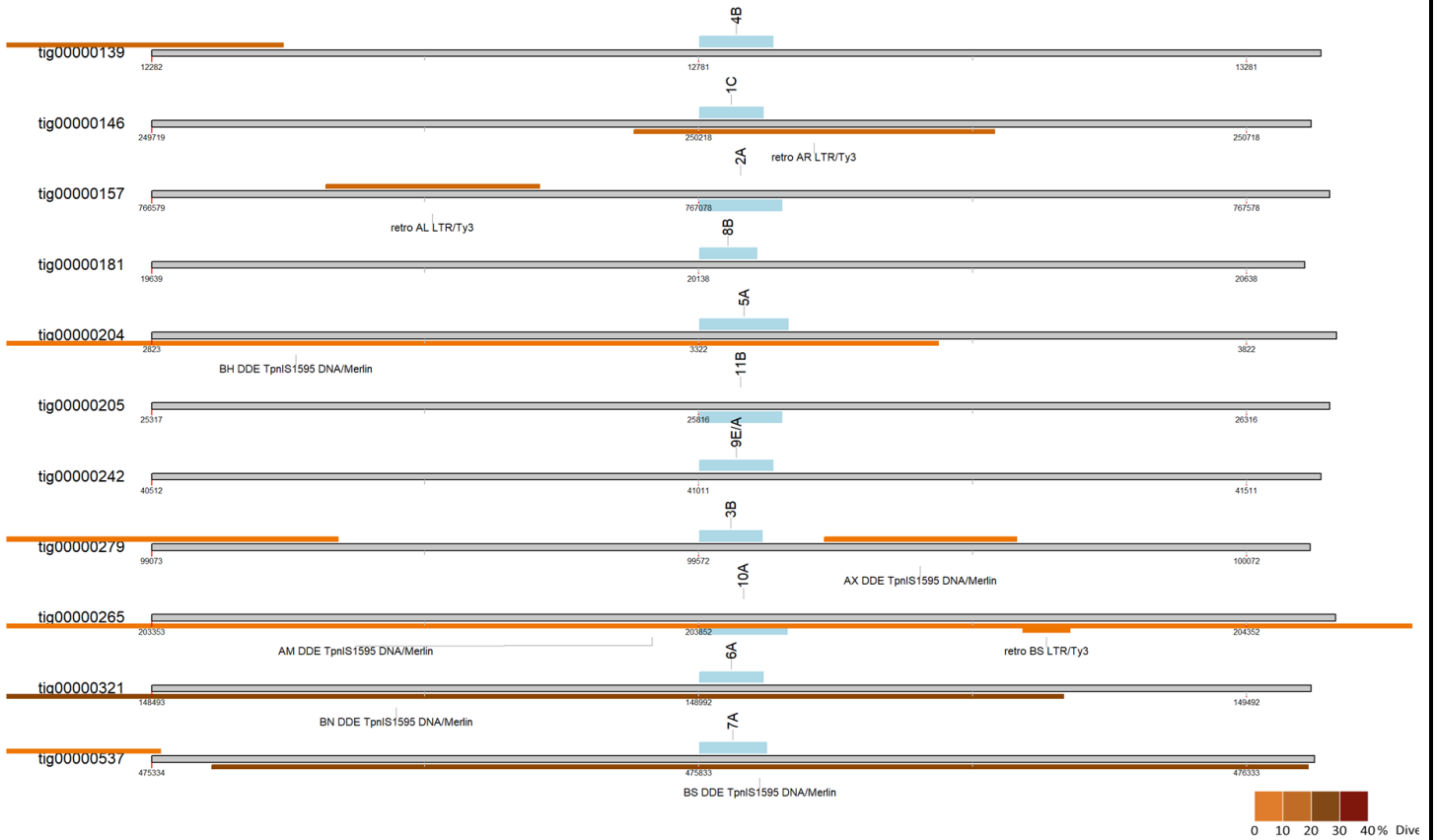


Fig 1 : Structure, genomic organization and expression profile of the 34 small interfering RNA precursor genes (continued)

B. Genomic loci of iRNA precursors with surrounding genes. One representative precursor locus is shown per iRNA. 500 nt upstream and downstream of iRNA precursor loci were plotted and annotated. Annotations include transposable elements, coding sequences, rRNA and tRNA. Transposable elements are coloured in shades of orange corresponding to their divergence percentages. iRNA precursor loci are indicated in light blue.

C.

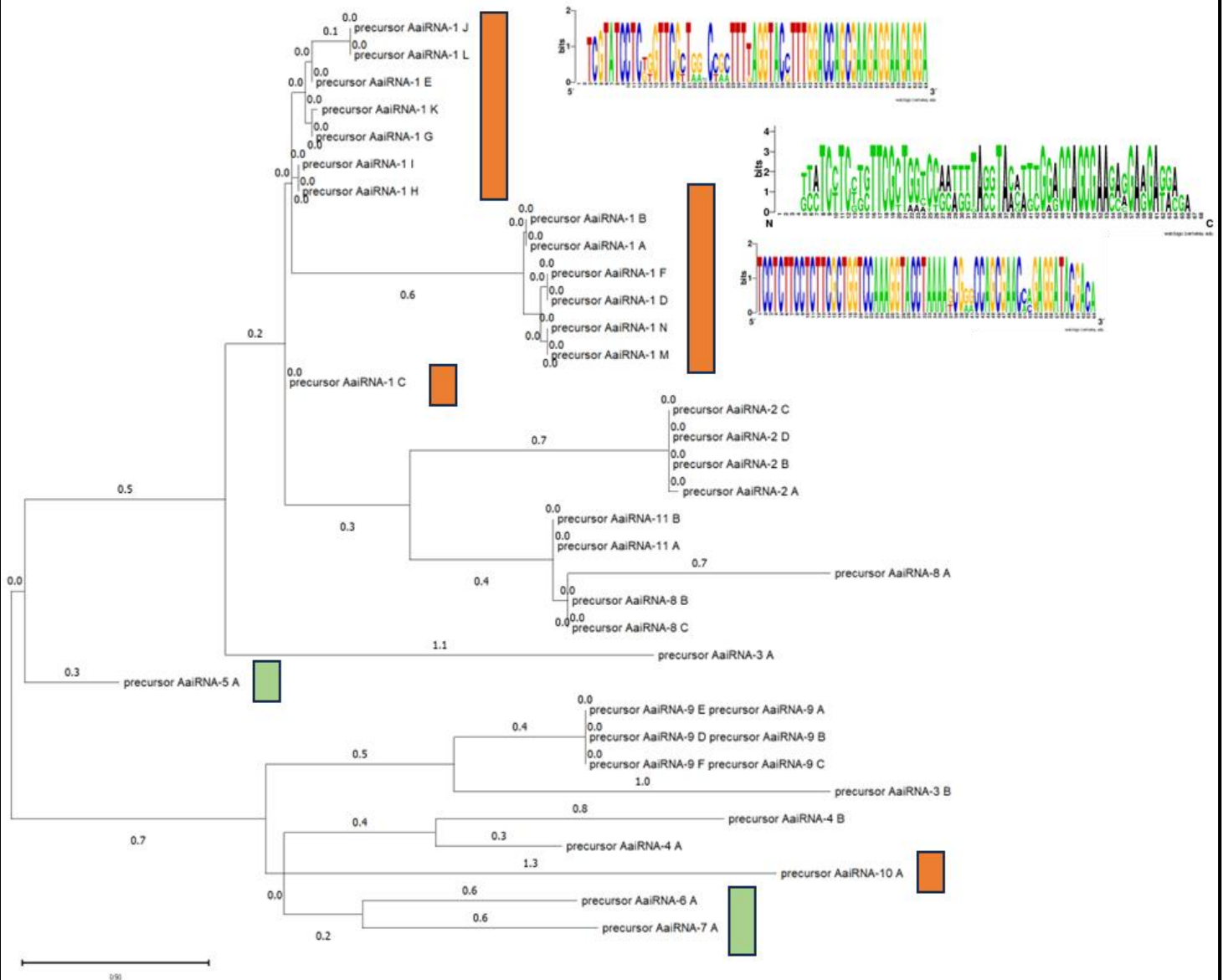


Fig 1 : Structure, genomic organization and expression profile of the 34 small interfering RNA precursor genes (continued)

C. Unrooted, maximum likelihood tree based on sequence diversity of the iRNA precursors. The tree is drawn with the length of the branches representing the number of substitutions per site which was indicated next to the branches. WebLogos for the two groups of iRNA-1 identified are shown along with one of all iRNA-1 precursors. iRNA precursors that were shown to overlap with TEs are indicated by a green rectangle for DNA transposons and an orange rectangle for retrotransposons.

A.

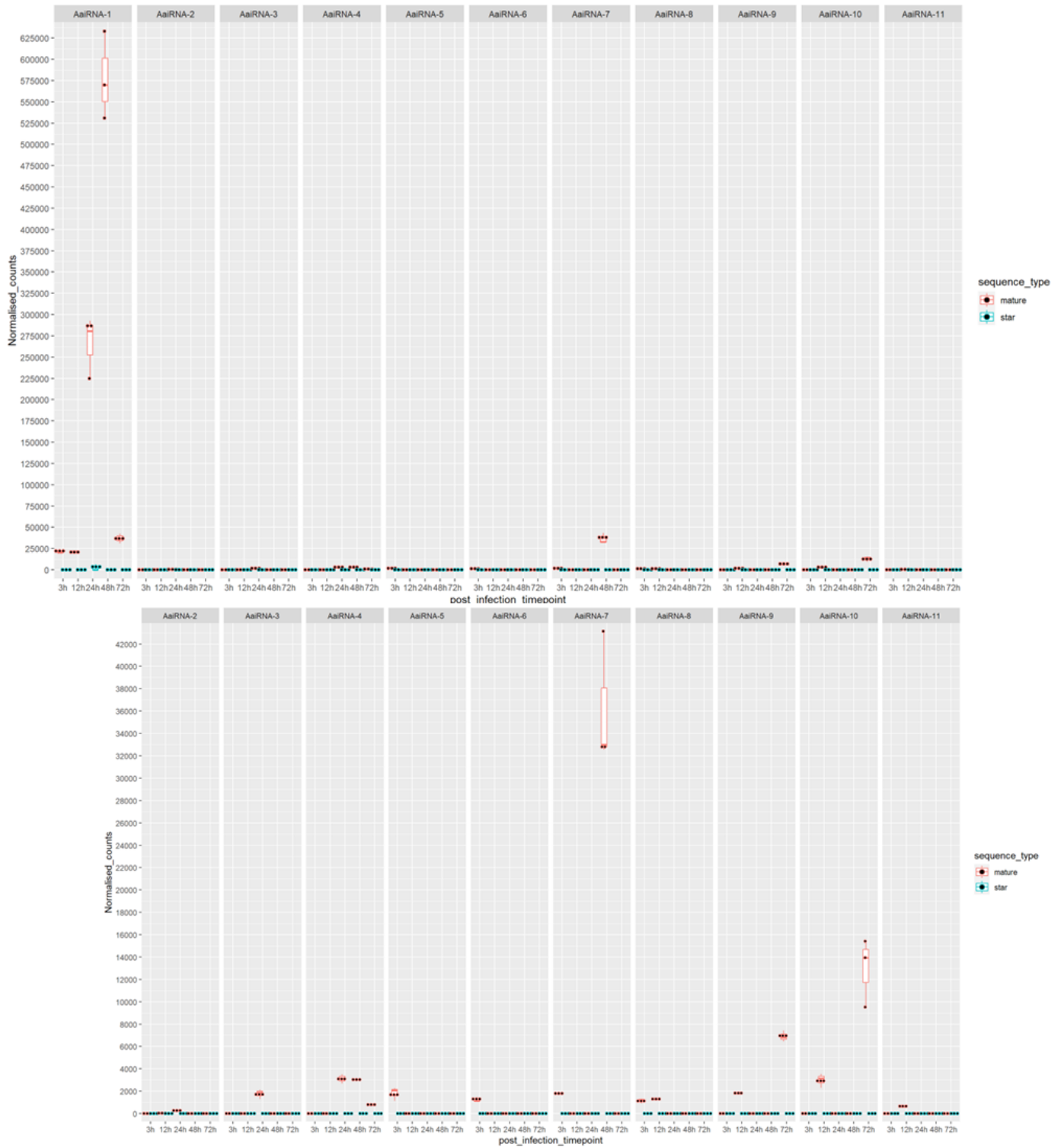


Fig 2: iRNA expression and localization analysis

A. Expression level of each mature *A. algerae* RNAi sequence and its star RNAi sequence assessed at different points in the infection kinetics of HFF cells. counts are derived from miRDeep2 analysis and were normalized using Deseq2. Counts for mature iRNA are represented in red and iRNA star counts are represented in blue. Different scales are used between the upper plot and the lower plot.

Expression and localization of the mature iRNA

To further understand the identified mature iRNA in their biological context, we analysed each of their expression throughout a 72h infection cycle of *A. algerae* in human foreskin fibroblasts (**Fig 2A**).

Regarding expression levels at each kinetic point, iRNA mature sequences were overrepresented compared to their corresponding star sequence. The mature RNAi-1 and its 14 precursor loci was by far the most highly expressed, it presented a peak in expression at 48h post infection, reaching over 500,000 counts before going back down to about 40,000 counts at 72h. iRNA-7 was almost expressed solely at 48h reaching 33,000 counts. and iRNA-9 and iRNA-10 seemed to be upregulated only at 72h post infection.

The remaining 7 RNAi displayed counts under 4,000 through all timepoints but did exhibit timepoint specific expression variations.

Given its strong expression levels, iRNA-1 was selected for FISH experiment to validate its expression and analyse its location during the parasite cycle development (**Fig 2B**). Strong fluorescence could be observed inside parasite stages, especially sporoblasts, meronts and early sporoblasts. Small fluorescent spheres could also be observed and may correspond to vesicles. Furthermore, probe staining was visible inside the perimeter of the host nucleus, and a slight hue staining was visible in the host cytoplasm. No fluorescence was identified in uninfected cells with the same acquisition parameters, except for strong fluorescence spots related to experimental artifacts. Finally, after RNase pre-treatment of infested cells, hybridization with the probe fluorescence staining was destabilized and became more diffuse, suggesting that the probes had indeed hybridized to RNA molecules.

B.

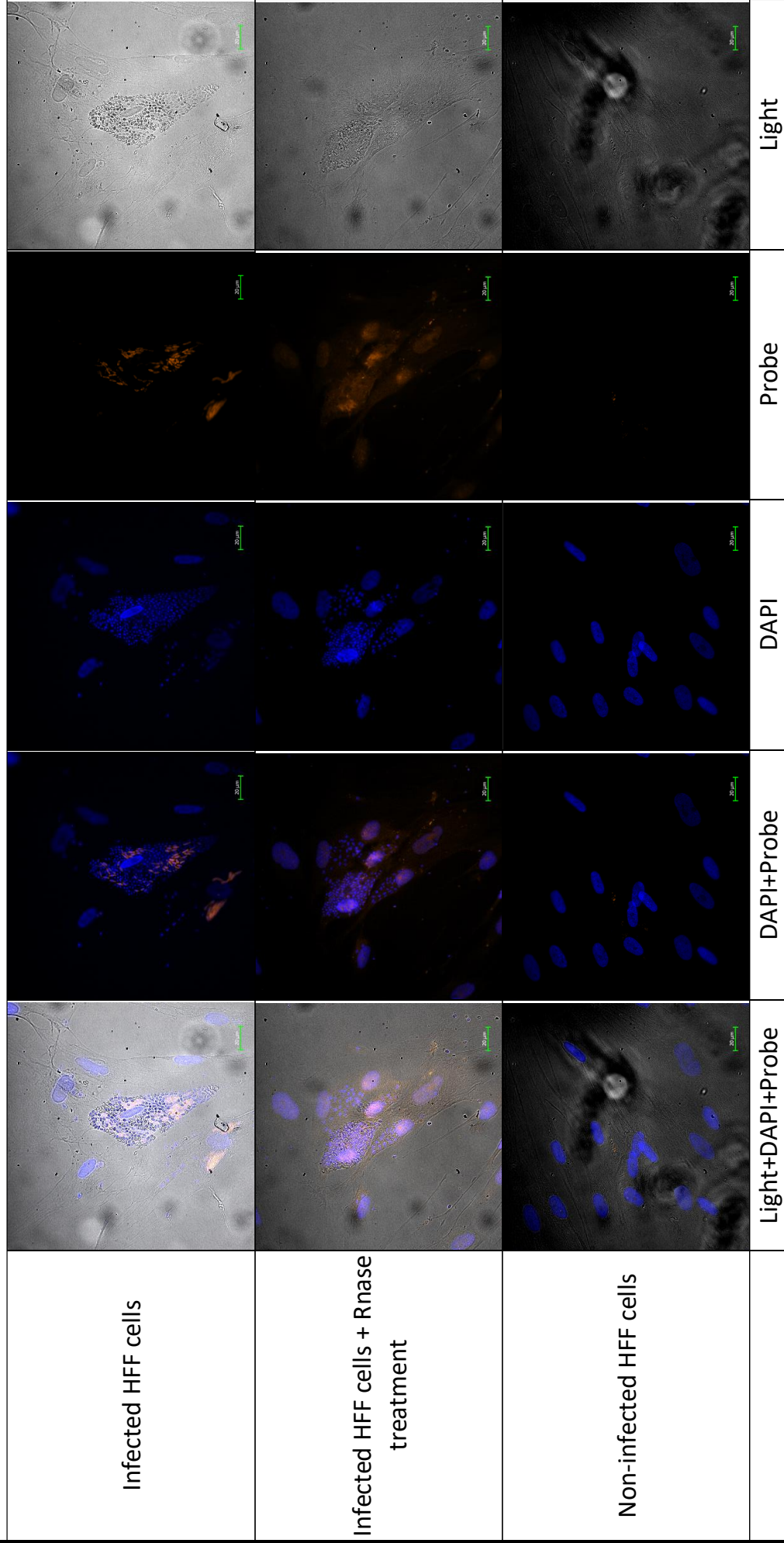


Fig 2: iRNA expression and localization analysis (continued)

B. Fluorescence *in situ* hybridization of iRNA-1 observed under X40 lens. From top to bottom images show infected HFF cells 7 days post infection; the same but with an RNase treatment prior to staining; and finally non-infected HFF cells. From left to right images represent white light + probe fluorescence + DAPI fluorescence; probe + DAPI; probe alone; DAPI alone; and white light alone. *A.algerae* meronts, sporonts and spores are visible inside the HFF cells. DAPI staining colours both HFF and *A.algerae* nuclei. The probe stains iRNA-1 in orange.

Predicted targets of the mature iRNA and experimental validation

iRNA-target pairs were predicted with the miRanda tool in host 3'UTRs, and using the expression data available, we filtered in all targets that showed downregulation every time the corresponding iRNA was upregulated in kinetics (**figure 3**). Targets were successfully identified for iRNA-1, iRNA-5, iRNA-6, iRNA-9, iRNA-10 and iRNA-11. Functional enrichment of targets for each iRNA revealed iRNA-1's CECR2 target to be involved in apoptotic DNA fragmentation processes (GO:0006309), chromatin remodelling (GO:0061640), cytoskeleton-dependent cytokinesis (GO:0043044), and vesicle-mediated transport (GO:0007010). iRNA-5 and iRNA-6 's target TOR1A was associated with positive regulation of clathrin-dependent endocytosis (GO:2000370). iRNA 9 in small GTPase mediated signal transduction (GO:0007264) and MAPK cascade (GO:0000165) and iRNA 11, which was predicted to have the biggest impact on its target GRIK1, was associated with synaptic transmission, glutamatergic (GO:0035249), and iRNA-10 was enriched in targets involved in the activation of protein kinase activity (GO:0032147). 35 *A. algerae* CDSs and 25 *A. algerae* TEs were also predicted as potential iRNA targets (list of targets and values for all iRNA targets available in **supplementary data 1**).

Discussion

In this study we were able to identify 11 interfering RNA loci (iRNA) in the *A. algerae* genome. Some of them could play a role in the control of host cell gene expression. We maintained the term "iRNA" throughout this work as precaution, as the data produced does not

unambiguously make the choice between siRNA and miRNA. Previous work had been done predicting small

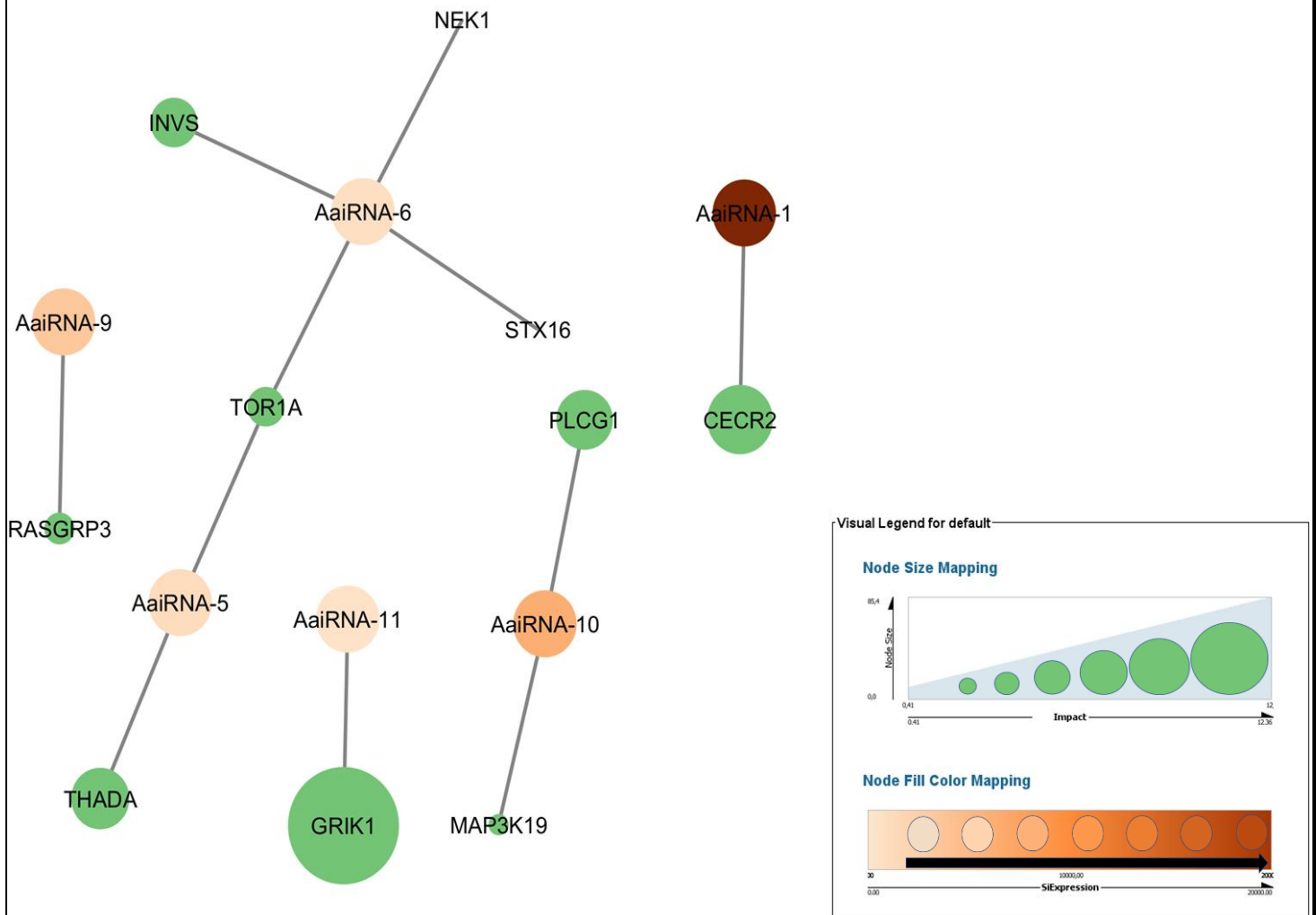


Fig 3 : iRNA-target interaction networks

Interaction network showing iRNA from *A. algerae* and the human target genes predicted by our analysis. iRNA from *A. algerae* are colored in shades of brown. iRNA circles are darker the more the iRNA is expressed over the infection cycle. The predicted target genes are coloured in green. The bigger the circle, the bigger the associated impact score. Impact score was calculated as mean expression value multiplied by fold change

interfering RNA in the microsporidia *Vairimorpha ceranae* (17,30,45) and *Nosema Bombycis* (31). The term given to the small RNA identified in these studies was often “miRNA-like” or “miRNA”, most likely since they had been predicted using the mirDeep2 software (46). We argue that despite mirDeep2 being initially conceived to identify miRNA precursor loci, and despite its proven efficacy for example to differentiate them from piRNA precursors, the mirDeep2 approach allows the prediction of all genomic loci that can form stemloop structures and that harbour a specific expression profile with a RNA strand (mature) over-represented compared to the opposite strand (the star (*) sequence) (46). These characteristics also correspond to endo-siRNA genes which have been shown to sometimes derive from stem-loop structured loci rather than double stranded RNA (12). The question on the nomenclature seems more and more secondary as several studies report overlapping roles between different types of small interfering RNA, thus painting a more complex picture of these molecules (1,27,47).

An important feature we were able to identify is the presence of a uridine residue on the first 5' nucleotide of the predicted iRNA as well as a very standardized 25 nt length. All these features have been described as hallmarks of Dicer protein maturation a key player in most interfering RNA pathways (1,6). The conservation of Argonaute and Dicer proteins in the microsporidian phylum has previously been characterised by Huang 2018 (48). He described that both proteins were either conserved or lost within the microsporidian phylum. This strongly suggests that in microsporidia these proteins were selected in regard to RNAi, as this is the only known pathway involving both of these proteins. Furthermore, functional experiments on *V. ceranae* have shown that knockdown of parasite Dicer protein negatively impacts its intracellular growth and development, leading to reduced spore loads in infected



bees (49,50). These further stresses the biological implications RNAi may have in microsporidian biology.

Despite Argonaute and Dicer having most likely been inherited from a common ancestor, the associated iRNA seem to have different origins and seem to be species specific. Recently, a study on the silkworm microsporidian parasite *Nosema bombycis*, could confirm the role of a parasite small RNA (Nb-milR8) in the sequence specific down-regulation of host peroxisome protein BmPEX16 (31). The study also confirmed the importance of this regulatory process in the development of the parasite: overexpression of the BmPEX16 protein allowed reduced proliferation of *N. bombycis* while a reciprocal knockdown of BmPEX16 resulted in higher parasite proliferation. Work on *V. ceranae* small RNAs highlighted their potential roles in modulating metabolism and immune response in the host, but also showed that 79% of the genes of *V. ceranae* itself were significantly correlated with the expression of one or more parasite miRNA (45). Our work suggests that the apoptosis response of the host may be partially impaired through iRNA activity during *A. algerae* infection. Similarly, Evans and Huang 2018 (45) have identified one miRNA in *V. ceranae* targeting gene Br-c (Broad Complex) which is involved in regulating developmental processes and apoptosis. Apoptosis impairment has been reported in several microsporidian infections, and multiple processes have been described (45,51–54). The parasite effectors initiating these processes are still largely unknown although specific gene dysregulations have been associated with the phenotypes (45,51–54). iRNA-1 seems to be an actor in the apoptosis inhibition during *A. algerae* infection but the localisation of its genes in the genome may also represent a concrete example of a TE derived sequence, conferring selective benefits in a microsporidian.



In addition to iRNA-1, we observed that several of the iRNA precursor loci were located on the opposite strand of transposable elements, from both class I and class II. Those TEs presented varying divergence levels suggesting different ages of insertion. The iRNA loci also presented positional shifts between homologous contigs suggesting previous insertion events. All these insertions appear to be recent in the evolutionary history of microsporidia, visible by the fact that similar iRNA precursor sequences could only be identified in different strains of *A. algerae* but not in other microsporidia (except for iRNA-7). In *Botrytis cinerea* small RNA from the pathogen have been proven to repress genes involved in the host plant's immunity (55). These small RNAs were also reported to derive from TEs for the most part (55).

On top of controlling gene expression level of their host, the different iRNA may also play roles in *A. algerae* gene and TE regulation. Indeed, we identified 35 different proteins coding genes and 25 different TEs potentially targeted by iRNA, but small impacts on their expression level were predicted. A study on *N. bombicys* has aimed to characterise the complete regulatory network of non-coding RNA in the parasite during infection and 284 miRNAs were identified in the parasite, with targets predicted in both the host and the parasite itself (57). However, the exceedingly high number of predicted miRNA and miRNA targets raises the question of biological relevancy (58,59). Indeed, the main question is how many of these identified miRNA and targets hold biological roles in the parasite. These are issues common in RNA interference studies (58,59). It is often difficult to come to a definite conclusion on small RNA biological impact given the high number of predicted interactions (58,59). Thankfully, several solutions have been proposed like CLIP-seq for example, that can allow high throughput identification of RNA binding protein binding sites (60,61). This can help in validating interaction between iRNA and their targets (60,61). Yet, even with this method



further questions remain, namely, 1) is the identified interaction enough to induce repression of the target 2) if it is, does it also have a real biological and phenotypical impact.

To navigate these issues in this work, we used a custom prediction methodology with the aim of reducing the number of potential iRNA targets to only the ones which exhibited downregulation whenever the associated iRNA was upregulated throughout the infection cycle. These targets were thus the most likely to have biological impact on their targets. This was only possible through the presence of triplicates in our sRNA seq and polyA+ seq dataset, as well as a longitudinal study, where we had several timepoints of *A. algerae*'s infection cycle. The roles of iRNA in host-parasite interactions can often involve very punctual activity but that are crucial in the overall process, with this experimental design we could identify such events. An example of this is the parasite *Toxoplasma gondii*, an apicomplexan with the completely functional miRNA machinery (62). These parasites express some miRNA involved in host cell differentiation and development at different phases of their infection (62–64). Some *T. gondii* miRNA are homologous to human and rodent miRNAs (62), but strikingly, others (99% miRNAs) are specific to *T.gondii* isolates (62–64). These miRNAs can represent 5-8% of a given isolates pool of ncRNA (62,63). These further stresses the need for multiple samples and conditions when predicting iRNA for parasites if we aim to identify them globally.

We were also able to reduce the number of iRNA-target pairs to explore thanks to our FISH probing of the iRNA. Expression data gave us a hint, but this allowed us to confirm which iRNA candidates were most interesting for future functional validation.



The FISH experiment revealed transfer of parasite iRNA into host cells cytoplasm and nucleus. This seems rather natural given the literature on host-parasite cross-talks. Dong *et al* (31) also visualized host-parasite exchanges in their work on *N. bombycis* small RNA using a FISH approach (31). They identified Nb-milR8 in both parasite and host cell cytoplasm. They did not report any staining in the host cell nucleus suggesting transfer to the nucleus might not be systematic for microsporidian small RNA (31). The plant parasite *B. cinerea* has also been shown to transfer small RNA into host cells that resulted in repression of plant immunity (55). The host plant as well, has been shown to transfer via extracellular vesicles small interfering RNA that can repress some parasite genes (24). LncRNA from *Cryptosporidium parvum* have been shown to be translocated into host cell nuclei and induce epigenetic repression of genes, further suggesting parasite ncRNA targeting to host nucleus may be a widespread mechanism of host subversion (67,68). In *P. falciparum*, some parasite genes are regulated by the host cell miRNA and Argonaute, not only with detrimental effects to the parasite (65,66).

Studies on *Caenorhabditis* worms infected by microsporidian parasites of the *Nematocida* genus, suggested that microsporidia infecting various species may exhibit similar transcriptional expression profiles (69). Taking this information in the scope of *A. algerae* and its adaptation to human cells, we can wonder if the iRNA we identified could also play a role during *Anopheles* infection by the parasite. It would also be interesting to know whether the targeted sequences belong to similar functional groups. This would require kinetic sequencing of *Anopheles* infected by *A. algerae* and total RNA extractions but would be sure to yield invaluable information on *A. algerae* host adaptation. Looking at all these interactions, small interfering RNA seem without a doubt key player of host-parasite interactions.



Material and methods

Data

Raw data used is available as NCBI Bioproject Accession number PRJNA927239. Detailed experimental design is described in Akossi et al (70). Briefly, Total RNA was extracted in triplicate from *A. algerae* infected HFF cells at 3h, 12h, 24h, 48h, and 72h post infection. The RNA was then used for both small RNAseq and PolyA+ RNAseq on each sample.

Genomic data used is described in previous chapter.

Prediction of iRNA candidates from small RNAseq data

A miRDeep2 analysis was carried out on individual data set (46) with default parameters and using the *A. algerae* genome produced by HiFiseq sequencing from previous chapter. Results were filtered to conserve only mature iRNA candidates predicted in all triplicates of a given timepoint. iRNA precursors were then named with numbers each corresponding to the mature iRNA sequence they encoded. Precursor loci were then further differentiated by adding a different letter to their name if they originated from different genomic coordinates.



Genomic analysis of iRNA candidates and precursors

iRNA loci were plotted using the Karyoplottter R package (71). Loci that were on the haploid genome were directly plotted. For the loci on homologue chromosomes their relative positions were plotted on the haploid genome based on contig-on-contig alignments done using minimap2 (72) with asm20 parameters as described in previous paper.

Annotation data from *A. algerae* genome (cf. previous chapter) were used to identify surrounding genes on either side within 500 nt of the RNAi loci.

Mature sequence WebLogos were generated using the WebLogo internet site from Berkeley [WebLogo - Create Sequence Logos \(berkeley.edu\)](http://weblogo.berkeley.edu).

Mapping of precursor iRNA sequences on the consensus transcriptomes from the previous paper was done using bowtie2 with -very-sensitive-local settings. Search for homologue sequences in other microsporidian genomes was done through BLASTN analysis against genomic sequences available from MicrosporidiaDB (73,74)

Multiple alignments and evolutionary analysis by Maximum Likelihood method

Multiple alignments were generated using the MEGA11 software, and the clustalW algorithm available with default settings (75,76).

The evolutionary history was inferred using the Maximum Likelihood method with the Tamura-Nei model (77). The tree with the highest log likelihood (-1363.32) was kept. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model. The topology with superior log likelihood value then was selected. The tree is drawn with the length of the branches representing the number of substitutions per site which was indicated next to the branches. Evolutionary analyses were conducted in MEGA11 (75).



Mature and star iRNA sequences expression analysis

Expression levels of mature RNAs and their respective star sequences were assessed using MirDeep2 (46) and DESeq2 (78). The number of reads for each mature iRNA sequence was extracted from the mirdeep2 results and normalised using DESeq2 (78). The results were then plotted using ggplot2 (79).

Fluorescence *In situ* hybridisation of mature iRNA and acquisition

Infected and non-infected HFF cells for Fluorescence *In Situ* Hybridisation (FISH) were cultured following the protocol described by Akossi et al (70). Briefly, Human foreskin fibroblasts cells (HFF) (reference ATCC SCRC-1041) were cultured in minimum essential medium (MEM) supplemented with 50 µg/ml Gentamicine, 1% Amphotericine B, 1% Penicillin/Streptomycin, 1% Glutamine and 10% Foetal bovine serum at 30 °C in 5% CO₂ atmosphere. All cells were cultured in 60 mm petri dishes each containing about 10 circular 13 mm diameter glass slides. For infection, cells were incubated at 30 °C for 1h with 1 ml of the supernatant from 7 days *Anncaliia algerae* spore cultures, then rinsed with fresh media.

After 7 days of culture, cells were fixed for 15 mins using 4% Formaldehyde and stored in PBS 1X at 4°C until staining.

Staining was done using custom 3' and 5' digoxigenin labelled probes from Eurogentec complementary to iRNA-1 (Probe sequence : 5'TCCTCTTCCTCTTCGCTGGTCCAAA3'). Slides were first dehydrated by incubating them in 50, 80 and 95% (v/v) ethanol for 5 mins each.

They were then incubated in 1ml of saline-sodium-citrate (SSC) hybridisation buffer for 15 mins at room temperature. Slides were drained and incubated at 55 °C in pre-heated



hybridisation buffer. Before use, the probe was denatured 5 mins at 95°C and later added to the slides at a final concentration of 0.2 µM in hybridisation buffer. Hybridisation was done for 25 mins at 55 °C.

Slides undergoing RNase treatment were treated before staining with in PBS 1X containing bovine pancreas RNase A from Roche at a final concentration of 100 µg/ml (5 U) for 15mins at 55°C.

Anti-digoxin-rhodamine Fab fragments from Roche were used to reveal the digoxigenin labelled probes. Briefly, slides were rinsed twice for 3 mins in probe free hybridisation buffer, then incubated for 2*30 mins in MABT buffer (maleic acid buffer containing Tween 20). Next, the slides were incubated in MABT buffer containing 2% BSA (blocking solution) for 1h.

After 1h, slides were drained and blocking solution containing a 1:200 dilution of the Anti-digoxin-rhodamine Fab fragments from Roche was added. Slides were incubated for 1 h then washed 3*10 mins in MABT buffer.

Finally, all slides were counter stained with 1X DAPI for 5 mins then rinsed 4 times in 1x PBS. Acquisition was done on a ZEISS Axio Imager Apotome at X63 with emersion, and the same laser and microscope parameters for each sample. The Zeiss software was used for post-processing and identical visual parameters were applied to each image.

iRNA-target interaction network

iRNA targets were predicted in host 3' UTRs, parasite TEs and CDS using miRanda and a differential expression analysis strategy (DE). Targets kept were the ones predicted by miRanda and that were always downregulated when the associated iRNA was upregulated throughout the cycle. To account for confounding effects due to human miRNA targeting the same genes as some iRNA candidates, the same target prediction strategy was used with host



miRNAs, using Multimir R package (80) instead of miRanda as described in Akossi et al (70) and all the targets identified were discarded.

Reporter assay validation of iRNA-1 - CECR2 interaction (Planification)

Before going any further into the implications of the iRNA interaction processes between *A. algerae* and its host, we first aim to validate the interaction of iRNA-1 with its CECR2 target. To do so, we plan on using a Luciferase reporter assay or a dual fluorescence reporter assay. In both cases the luciferase/GFP gene would be modified to present the iRNA-1 target site from CECR2 in its 3'UTR. Different versions of the construct would exist, including a control without any modification, one with the CECR2 sequence but with nucleotides in a scrambled order; finally, a construct presenting several back-to-back copies of the CECR2 target site instead of the standard 3'UTR. The constructs would also hold a second fluorescent protein (Red fluorescent protein, RFP for example), this second protein would help estimate the transfection efficacy of the cells. This experiment would be done on non-infected HFF cells and 48h infected HFF cells (timepoint with the highest iRNA-1 expression observed), cultured on 96 well plates. Measures would be done using a plate reader for both GFP and RFP wavelengths; Or could use a fluorescence microscope identifying the shift in colour between yellow (Red+green fluorescence) and red. In case of poor transfection efficacy, Hela cells have been considered as a replacement for the HFF cells. Cells may also be cultured in plates with larger wells (48 or 24) if needed.



References

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* févr 2009;10(2):94-108.
2. Onishi R, Yamanaka S, Siomi MC. piRNA- and siRNA-mediated transcriptional repression in *Drosophila*, mice, and yeast: new insights and biodiversity. *EMBO Rep.* 5 oct 2021;22(10):e53062.
3. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol.* 3 août 2018;9:402.
4. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem.* 2015;84(1):405-33.
5. Obbard DJ, Finnegan DJ. RNA Interference: Endogenous siRNAs Derived from Transposable Elements. *Curr Biol.* 8 juill 2008;18(13):R561-3.
6. Hoehener C, Hug I, Nowacki M. Dicer-like Enzymes with Sequence Cleavage Preferences. *Cell.* 22 mars 2018;173(1):234-247.e7.
7. Wu J, Yang J, Cho WC, Zheng Y. Argonaute proteins: Structural features, functions and emerging roles. *J Adv Res.* 29 avr 2020;24:317-24.
8. Sarshad AA, Juan AH, Muler AIC, Anastasakis DG, Wang X, Genzor P, et al. Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol Cell.* 9 sept 2018;71(6):1040.
9. Iwakawa H, Tomari Y. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol Cell.* 6 janv 2022;82(1):30-43.
10. Kehl T, Backes C, Kern F, Fehlmann T, Ludwig N, Meese E, et al. About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget.* 9 nov 2017;8(63):107167-75.
11. Abels ER, Breakefield XO. Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake. *Cell Mol Neurobiol.* avr 2016;36(3):301-12.
12. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature.* 22 janv 2009;457(7228):413-20.
13. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc B Biol Sci.* 12 janv 2009;364(1513):99-115.

14. Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, et al. An endosRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell*. 2 nov 2017;21(5):694-703.e7.
15. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, et al. RNAi in budding yeast. *Science*. 23 oct 2009;326(5952):544-50.
16. Tschudi C, Shi H, Franklin JB, Ullu E. Small interfering RNA-producing loci in the ancient parasitic eukaryote *Trypanosoma brucei*. *BMC Genomics*. 27 août 2012;13:427.
17. Shao SS, Yan WY, Huang Q. Identification of novel miRNAs from the microsporidian parasite *Nosema ceranae*. *Infect Genet Evol*. 1 sept 2021;93:104930.
18. Gong M, Wang Y, Zhang J, Zhao Y, Wan J, Shang J, et al. Chilling Stress Triggers VvAgo1-Mediated miRNA-Like RNA Biogenesis in *Volvariella volvacea*. *Front Microbiol*. 2020;11:523593.
19. Sahoo GC, Ansari MY, Dikhit MR, Gupta N, Rana S, Das P. Computational Identification of microRNA-like Elements in *Leishmania major*. *MicroRNA Shariqah United Arab Emir*. 2014;2(3):225-30.
20. Marin FR, Dávalos A, Kiltshewskij D, Crespo MC, Cairns M, Andrés-León E, et al. RNA-Seq, Bioinformatic Identification of Potential MicroRNA-like Small RNAs in the Edible Mushroom *Agaricus bisporus* and Experimental Approach for Their Validation. *Int J Mol Sci*. 28 avr 2022;23(9):4923.
21. Afrin F, Khan I, Hemeg HA. *Leishmania*-Host Interactions—An Epigenetic Paradigm. *Front Immunol*. 22 mars 2019;10:492.
22. Zeng J, Gupta VK, Jiang Y, Yang B, Gong L, Zhu H. Cross-Kingdom Small RNAs Among Animals, Plants and Microbes. *Cells*. 23 avr 2019;8(4):371.
23. Chi X, Wang Z, Wang Y, Liu Z, Wang H, Xu B. Cross-Kingdom Regulation of Plant-Derived miRNAs in Modulating Insect Development. *Int J Mol Sci*. 28 avr 2023;24(9):7978.
24. Gualtieri C, Leonetti P, Macovei A. Plant miRNA Cross-Kingdom Transfer Targeting Parasitic and Mutualistic Organisms as a Tool to Advance Modern Agriculture. *Front Plant Sci*. 2020;11:930.
25. Lodde V, Floris M, Muroli MR, Cucca F, Idda ML. Non-coding RNAs in malaria infection. *Wiley Interdiscip Rev RNA*. 2022;13(3):e1697.
26. Liu J, Zhu L, Wang J, Qiu L, Chen Y, Davis RE, et al. *Schistosoma japonicum* extracellular vesicle miRNA cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathog*. 4 juin 2019;15(6):e1007817.

27. Hao L, Cai P, Jiang N, Wang H, Chen Q. Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics*. 21 janv 2010;11:55.
28. Bayer-Santos E, Lima FM, Ruiz JC, Almeida IC, da Silveira JF. Characterization of the small RNA content of *Trypanosoma cruzi* extracellular vesicles. *Mol Biochem Parasitol*. févr 2014;193(2):71-4.
29. Garcia-Silva MR, Cabrera-Cabrera F, Cura das Neves RF, Souto-Pradrón T, de Souza W, Cayota A. Gene Expression Changes Induced by *Trypanosoma cruzi* Shed Microvesicles in Mammalian Host Cells: Relevance of tRNA-Derived Halves. *BioMed Res Int*. 2014;2014:305239.
30. Huang Q, Evans JD. Identification of microRNA-like small RNAs from fungal parasite *Nosema ceranae*. *J Invertebr Pathol*. 1 janv 2016;133:107-9.
31. Dong Z, Zheng N, Hu C, Deng B, Fang W, Wu Q, et al. *Nosema bombycis* microRNA-like RNA 8 (Nb-miR8) Increases Fungal Pathogenicity by Modulating BmPEX16 Gene Expression in Its Host, *Bombyx mori*. *Microbiol Spectr*. 31 oct 2021;9(2):e0104821.
32. Wadi L, Reinke AW. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLOS Pathog*. 13 févr 2020;16(2):e1008276.
33. Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol*. 1 août 2022;38(8):642-59.
34. Peyretailade E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E, et al. Annotation of microsporidian genomes using transcriptional signals. *Nat Commun*. 16 oct 2012;3(1):1137.
35. Mascarenhas dos Santos AC, Julian AT, Liang P, Juárez O, Pombert JF. Telomere-to-Telomere genome assemblies of human-infecting *Encephalitozoon* species. *BMC Genomics*. 4 mai 2023;24:237.
36. Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun*. 21 sept 2010;1:77.
37. Williams BA, Lee RC, Becnel JJ, Weiss LM, Fast NM, Keeling PJ. Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. *BMC Genomics*. 29 avr 2008;9:200.
38. Khalaf A, Jaron KS, Blaxter ML, Lawniczak MKN. Polyploidy is widespread in Microsporidia [Internet]. *Evolutionary Biology*; 2023 sept [cité 4 nov 2023]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2023.09.29.560119>
39. Haag KL, Pombert JF, Sun Y, de Albuquerque NRM, Batliner B, Fields P, et al. Microsporidia with Vertical Transmission Were Likely Shaped by Nonadaptive Processes. *Genome Biol Evol*. 1 janv 2020;12(1):3599-614.

40. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. *Genome Biol Evol.* 1 sept 2014;6(9):2289-300.
41. de Albuquerque NRM, Ebert D, Haag KL. Transposable element abundance correlates with mode of transmission in microsporidian parasites. *Mob DNA.* 23 juin 2020;11:19.
42. Seatamanoch N, Kongdachalart S, Sunantaraporn S, Siriyasatien P, Brownell N. Microsporidia, a Highly Adaptive Organism and Its Host Expansion to Humans. *Front Cell Infect Microbiol* [Internet]. 2022 [cité 30 août 2023];12. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.924007>
43. Monaghan SR, Rumney RL, Vo NTK, Bols NC, Lee LEJ. In vitro growth of microsporidia *Anncaliia algerae* in cell lines from warm water fish. *Vitro Cell Dev Biol - Anim.* 1 févr 2011;47(2):104-13.
44. Willis AR, Reinke AW. Factors That Determine Microsporidia Infection and Host Specificity. In: Weiss LM, Reinke AW, éditeurs. *Microsporidia: Current Advances in Biology* [Internet]. Cham: Springer International Publishing; 2022 [cité 2 nov 2023]. p. 91-114. (Experientia Supplementum). Disponible sur: https://doi.org/10.1007/978-3-030-93306-7_4
45. Evans JD, Huang Q. Interactions Among Host–Parasite MicroRNAs During *Nosema ceranae* Proliferation in *Apis mellifera*. *Front Microbiol.* 10 avr 2018;9:698.
46. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* janv 2012;40(1):37-52.
47. Simonelig M. piRNAs, master regulators of gene expression. *Cell Res.* juill 2014;24(7):779-80.
48. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infect Genet Evol.* 1 nov 2018;65:329-32.
49. Huang Q. Effective Silencing of Dicer Decreases Spore Load of the Honey Bee Parasite *Nosema ceranae*. *Fungal Genomics Biol.* 8 août 2016;06.
50. Huang Q, Li W, Chen Y, Retschnig-Tanner G, Yanez O, Neumann P, et al. Dicer regulates *Nosema ceranae* proliferation in honeybees. *Insect Mol Biol.* 2019;28(1):74-85.
51. del Aguila C, Izquierdo F, Granja AG, Hurtado C, Fenoy S, Fresno M, et al. Encephalitozoon microsporidia modulates p53-mediated apoptosis in infected cells. *Int J Parasitol.* juill 2006;36(8):869-76.
52. He X, Fu Z, Li M, Liu H, Cai S, Man N, et al. *Nosema bombycis* (Microsporidia) suppresses apoptosis in BmN cells (*Bombyx mori*). *Acta Biochim Biophys Sin.* sept 2015;47(9):696-702.

53. Higes M, Juarranz Á, Dias-Almeida J, Lucena S, Botías C, Meana A, et al. Apoptosis in the pathogenesis of *Nosema ceranae* (Microsporidia: Nosematidae) in honey bees (*Apis mellifera*). *Environ Microbiol Rep*. août 2013;5(4):530-6.
54. Kurze C, Le Conte Y, Dussaubat C, Eler S, Kryger P, Lewkowski O, et al. *Nosema* Tolerant Honeybees (*Apis mellifera*) Escape Parasitic Manipulation of Apoptosis. *PLoS ONE*. 7 oct 2015;10(10):e0140174.
55. Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I, et al. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*. 4 oct 2013;342(6154):118-23.
56. Peyretilade E, El Alaoui H, Diogon M, Polonais V, Parisot N, Biron DG, et al. Extreme reduction and compaction of microsporidian genomes. *Res Microbiol*. 1 juill 2011;162(6):598-606.
57. Shen Z, Yang Q, Luo L, Li T, Ke Z, Li T, et al. Non-coding RNAs identification and regulatory networks in pathogen-host interaction in the microsporidia congenital infection. *BMC Genomics*. 26 juill 2023;24:420.
58. Hansen TB, Kjems J, Bramsen JB. Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis. *RNA Biol*. 2011;8(3):378-83.
59. Pinzón N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, et al. The number of biologically relevant microRNA targets has been largely overestimated. *Genome Res [Internet]*. 15 nov 2016 [cité 6 nov 2023]; Disponible sur: <https://genome.cshlp.org/content/early/2016/11/07/gr.205146.116>
60. Hafner M, Katsantoni M, Köster T, Marks J, Mukherjee J, Staiger D, et al. CLIP and complementary methods. *Nat Rev Methods Primer*. 4 mars 2021;1(1):1-23.
61. Stoute J, Liu KF. CLIP-Seq to identify targets and interactions of RNA binding proteins and RNA modifying enzymes. *Methods Enzymol*. 2021;658:419-34.
62. Rojas-Pirela M, Andrade-Alviárez D, Medina L, Castillo C, Liempi A, Guerrero-Muñoz J, et al. MicroRNAs: master regulators in host-parasitic protist interactions. *Open Biol*. 15 juin 2022;12(6):210395.
63. Wang J, Liu X, Jia B, Lu H, Peng S, Piao X, et al. A comparative study of small RNAs in *Toxoplasma gondii* of distinct genotypes. *Parasit Vectors*. 3 sept 2012;5(1):186.
64. Saçar MD, Bağcı C, Allmer J. Computational Prediction of MicroRNAs from *Toxoplasma gondii* Potentially Regulating the Hosts' Gene Expression. *Genomics Proteomics Bioinformatics*. oct 2014;12(5):228-38.
65. Wang Z, Xi J, Hao X, Deng W, Liu J, Wei C, et al. Red blood cells release microparticles containing human Argonaute 2 and miRNAs to target genes of *Plasmodium falciparum*. *Emerg Microbes Infect*. 23 août 2017;6(8):e75.

66. Mueller AK, Hammerschmidt-Kamper C, Kaiser A. RNAi in Plasmodium. *Curr Pharm Des.* 2014;20(2):278-83.
67. Wang Y, Gong AY, Ma S, Chen X, Strauss-Soukup JK, Chen XM. Delivery of Parasite Cdg7_Flc_0990 RNA Transcript into Intestinal Epithelial Cells during *Cryptosporidium parvum* Infection Suppresses Host Cell Gene Transcription through Epigenetic Mechanisms. *Cell Microbiol* [Internet]. nov 2017 [cité 21 nov 2023];19(11). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5638686/>
68. Li Y, Baptista RP, Kissinger JC. Noncoding RNAs in Apicomplexan Parasites: An Update. *Trends Parasitol.* oct 2020;36(10):835-49.
69. Wan YC, Troemel ER, Reinke AW. Conservation of Nematocida microsporidia gene expression and host response in *Caenorhabditis* nematodes. *PLoS One.* 2022;17(12):e0279103.
70. Akossi RF, Delbac F, El Alaoui H, Wawrzyniak I, Peyretailade E. The intracellular parasite *Anncaliia algerae* induces a massive miRNA down-regulation in human cells. *Non-Coding RNA Res.* 1 sept 2023;8(3):363-75.
71. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinforma Oxf Engl.* 1 oct 2017;33(19):3088-90.
72. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma Oxf Engl.* 15 sept 2018;34(18):3094-100.
73. Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res.* janv 2011;39(Database issue):D612-619.
74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 5 oct 1990;215(3):403-10.
75. MEGA11: Molecular Evolutionary Genetics Analysis Version 11 | Molecular Biology and Evolution | Oxford Academic [Internet]. [cité 16 nov 2023]. Disponible sur: <https://academic.oup.com/mbe/article/38/7/3022/6248099>
76. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. - PMC [Internet]. [cité 16 nov 2023]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308517/>
77. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* mai 1993;10(3):512-26.

78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
79. ggplot2: Elegant Graphics for Data Analysis | SpringerLink [Internet]. [cité 2 nov 2023]. Disponible sur: <https://link.springer.com/book/10.1007/978-0-387-98141-3>
80. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res.* 2014;42(17):e133.



Nous avons pu mettre en évidence l'expression de petits ARN interférents par *A.algerae*. Ces petits ARN sont codés par des gènes propres à cette espèce et non conservés chez les autres microsporidies. Ils seraient impliqués dans des mécanismes d'auto-régulation transcriptionnelle mais également dans la régulation de la cellule hôte, notamment par la régulation de ses voies apoptotiques. Ces ARN ont pu être observés comme transférés dans le cytoplasme et le noyau de la cellule hôte, et cette régulation participerait au phénotype des cellules infectées et à la physiopathologie des infections à l'échelle cellulaire. Ces travaux sont en cours et une validation biologique par gène rapporteur est prévue.

Ceci étant, le phénotype final des cellules infectées est dicté par l'action d'*A.algerae* mais également par la réponse de la cellule hôte. Nous nous sommes donc ensuite intéressés à cette réponse cellulaire de l'hôte, en nous focalisant sur son transcriptome, en particulier ses ARNm et ses miRNA.



PARTIE IV : ETUDE DE LA REPOSE TRANSCRIPTOMIQUE DE LA CELLULE HOTE FACE A UNE INFECTION PAR *ANNICALIA ALGERAE*

Chez l'homme, les microsporidies ont été décrites comme étant capables de moduler les réponses immunitaires de leurs hôtes (107). Cela semble parfois conduire à des infections persistantes, bien que classiquement elles soient éliminées par le système immunitaire chez les individus immunocompétents (349). La part que joue la réponse cellulaire dans ces interactions n'est pas claire. Les miRNA humains pourraient également constituer un enjeu important de ces interactions par leurs rôles de régulateurs immuns et cellulaires (348,350).

Dans cette dernière partie de la thèse, nous avons cherché à mieux caractériser la réponse cellulaire de l'hôte vis-à-vis de l'infection par *A. algerae*, en se concentrant sur la réponse transcriptomique. La dynamique d'expression des miRNA et ARN messagers humains a été étudiée tout au long du cycle infectieux d'*A. algerae*.

* « Supplementary data » disponibles dans la version en ligne

#Article publié dans *Non-coding RNA research*, Volume 8, Issue 3, September 2023, Pages 363-375 <https://doi.org/10.1016/j.ncrna.2023.05.003> #



===ARTICLE ICI===



A travers cette étude nous avons pu observer les modulations transcriptomiques induites par l'infection d'une cellule hôte par *A. algerae*. Ces modulations vont dans le sens de ce qui est connu dans la littérature sur les infections microsporidiennes à savoir l'altération de processus immunitaires et métaboliques.

Nous avons toutefois pu mettre en évidence une répression d'une grande quantité de miRNA de l'hôte tout au long de l'infection par *A. algerae*. Ce phénomène n'avait encore jamais été décrit et souligne l'importance que peuvent avoir les miRNA dans les interactions hôte-parasite. Ceci soulève également la question du mécanisme sous-jacent à cette répression généralisée, et les conséquences associées dans un contexte *in vivo* ou dans un contexte de co-infection.



**SECTION C : DISCUSSION
GENERALE, CONCLUSIONS ET
PERSPECTIVES**



SECTION C DISCUSSION GENERALE, CONCLUSIONS ET PERSPECTIVES

PARTIE I DISCUSSION GENERALE ET CONCLUSIONS

Les objectifs de ce travail de thèse étaient : i) La validation d'un pipeline d'annotation automatique spécifique des génomes microsporidiens ii) La caractérisation génomique et transcriptomique de la microsporidie *A. algerae* iii) L'étude du rôle et de l'origine de petits ARNs non codants d'*A. algerae* dans l'interaction avec sa cellule hôte et iv) L'étude de la réponse transcriptomique de la cellule hôte face à une infection par *A. algerae* avec comme objectif plus global d'explorer et de décrypter le rôle des ARNnc dans les interactions hôte-parasite.

Concernant l'outil d'annotation MicroAnnot (346); le développement et la baisse du coût des technologies de séquençage haut débit a conduit à la production massive de données de génomique, y compris pour les microsporidies (351). Cet immense flux de données nécessite des méthodologies de traitement et d'annotation dédiées en raison de la spécificité que l'on peut retrouver dans chaque phylum du vivant (343,345). En effet, l'usage parfois trop systématique et non standardisé d'outils d'annotation « généralistes », ainsi que la difficulté à paramétrer ces outils pour chaque espèce, a conduit à de nombreuses erreurs que l'on retrouve dans les bases de données (352–354). Corriger ces erreurs et minimiser leur survenue lors de l'annotation d'un génome représente un enjeu majeur des études génomiques contemporaines. Quelle que soit la spécificité d'un outil d'annotation, aucun n'est fiable à 100%. Néanmoins, pour tendre vers cette valeur, la combinaison d'une annotation automatique s'appuyant sur les caractéristiques propres de chaque organisme et d'une curation manuelle demeure le « gold standard » (344,345,355,356). MicroAnnot, par la méthodologie utilisée présente une très forte spécificité et sensibilité sur les génomes microsporidiens. La réalisation lors d'une première étape d'annotation d'une approche comparative assure la construction d'un jeu de données qui peut être considéré comme fiable



et permet d'assurer le meilleur apprentissage possible au logiciel Glimmer pour l'étape suivante d'annotation *de novo* (357,358). Cependant, l'approche comparative nécessite des génomes aussi fiables que possible en référence, et c'est pour cette raison que les bases de données utilisées par MicroAnnot ont été manuellement nettoyées avant d'être implémentées dans l'outil. Par la suite, de nouveaux génomes microsporidiens annotés par MicroAnnot et curés manuellement pourront à leur tour être ajoutés comme base de référence à l'outil. Ainsi, de manière dynamique la qualité des annotations s'améliorera au cours du temps. C'est notamment ce qui sera fait avec l'ajout des données issues du nouveau séquençage du génome d'*A.algerae*. En effet, les données concernant cette espèce étaient parcellaires en raison des problématiques d'assemblage liées aux régions répétées du génome (138). Elles seront maintenant plus complètes et participeront à enrichir les bases de données.

En plus de gènes codants, MicroAnnot assure également la prédiction des CDS retrouvés au sein des séquences d'ET ainsi que certains gènes non-codants tels que les ARNt et les ARNr. Pour les ARNt, la prédiction se fait via l'outil tRNAscan-SE qui tire parti de modèles de covariance basés sur les connaissances sur la structure des ARNt (359). En ce qui concerne la prédiction des ET et des ARNr, celle-ci s'appuie là encore sur une approche comparative utilisant des bases de données spécifiques et propres à MicroAnnot. Pour les ARNr la base de données a été construite en utilisant des séquences des 10 clades microsporidiens (110), permettant d'avoir une représentation de la diversité phylogénétique des microsporidies. Concernant les ET, bien que les séquences microsporidiennes soient parfois très divergentes, la base de données que nous avons construite manuellement à partir de séquences issues de différentes bases de données (128) s'est avérée particulièrement efficace pour identifier ces ET dans le nouveau génome d'*A. Algerae*.

L'emploi de MicroAnnot sur le nouveau génome d'*A.algerae* Undeen généré par l'approche HiFi seq de PacBio a permis d'identifier 2916 gènes codant au sein du génome haploïde, ce qui est cohérent avec le nombre moyen de gènes jusqu'à présent retrouvés chez les microsporidies (119,137,138,360). Cette annotation nous a ensuite également permis de confirmer la faible densité en gène codants de son génome (127), ainsi que d'étudier le polymorphisme et la synténie entre chromosomes homologues. Bien que MicroAnnot puisse identifier les CDS correspondant aux ET dans un génome, il n'est pas capable d'assurer



l'identification de ceux ne contenant pas de CDS définis et notre base bien qu'étant la plus exhaustive possible ne contenait pas l'ensemble des séquences d'ET retrouvées dans les génomes microsporidiens. Nous avons donc opté pour une approche combinée entre l'annotation MicroAnnot, et une prédiction *de novo* dédiée aux ET afin de caractériser plus efficacement d'éventuelles nouvelles familles d'ET. Cette approche s'est avérée particulièrement pertinente en assurant l'identification de 20 nouvelles familles lesquelles pourront également être ajoutées à la base de données de MicroAnnot.

L'annotation des autres types de gènes (ARNnc) est une problématique encore plus complexe qui ne peut pas être résolue en s'appuyant uniquement sur de l'analyse du fait de la diversité de ces séquences et de leur faible degré de conservation entre espèces (168,361). Cette annotation est d'autant plus difficile chez les organismes les moins caractérisés et présentant de fortes vitesses d'évolution des séquences, comme c'est le cas chez les microsporidies. Nous avons donc combiné les ressources génomiques aux données de transcriptomique générées sur le cycle infectieux du parasite pour l'annotation des gènes d'ARNnc. Ceci nous a permis de mettre en évidence un grand répertoire de long ARN non codants (LncRNA) chez *A.algerae*. Les rôles décrits pour les LncRNA sont très variés et certains demeurent encore peu clairs (171,362). Cependant, un lien entre la présence d'éléments transposables et les LncRNA a déjà été suggéré (363). A l'issue de notre étude, nous avons également proposé un mécanisme de régulation des éléments transposables chez *A.algerae*. Ce mécanisme serait basé sur des ARNnc produits par l'action successive de trois protéines : une RNA dépendante RNA polymérase (Rdrpl), une protéine Dicer, et une protéine Argonaute. Ce processus tirerait parti de l'overrun transcriptionnel observé chez *A.algerae* pour réguler les éléments transposables insérés suffisamment proche de ces gènes, y compris les éléments transposables à ADN. Ce phénomène d'overrun avait déjà été décrit chez les microsporidies, mais ses rôles dans un contexte de régulation des ET n'avaient pas encore été envisagés (131,364).

Un autre point intéressant concerne cette fois l'origine des protéines Argonaute, Dicer et Rdrpl qui ne sont pas retrouvées chez toutes les microsporidies (119,304). En effet, l'étude de leur phylogénie respective chez les espèces où elles sont retrouvées semble décrire trois origines distinctes, mais elles semblent tout de même avoir été héritées de manière



ancestrale chez *A.algerae* (119,304) Cet aspect est particulièrement intéressant lorsqu'on intègre le fait que les ET chez *A. algerae* semblent en partie issus de transferts horizontaux provenant de métazoaires (128). Ces trois protéines pourraient avoir été maintenues uniquement chez les microsporidies présentant des ET. Ces ET seraient possiblement dérivés d'hôtes abritant eux même une forte diversité d'ET actifs. Cette caractéristique est retrouvée chez de nombreux insectes dont les moustiques (162). Cette hypothèse est d'autant plus séduisante qu'un lien a été établi entre le mode de transmission des microsporidies et l'abondance d'ET dans leur génome. Ainsi, les microsporidies capables d'être transmises verticalement (de parents parasités directement vers leurs progénitures) présenteraient une plus grande proportion d'ET dans leurs génomes (129). Ceci est à mettre en lien avec le fait que les ET soient connus pour être particulièrement actifs dans les cellules de la lignée germinale (144,365). Dans ce modèle, la conservation des protéines Argonaute, Dicer et Rdrpl, ainsi que la production d'ARNnc associée, constitueraient une adaptation directe à la forte présence d'ET dans le génome d'*A.algerae* et de certaines autres microsporidies.

Dans la suite de notre étude, l'analyse des données transcriptomiques a également permis l'identification de 34 gènes à l'origine d'ARN interférents impliqués dans la régulation de transcrits chez *A.algerae*, mais aussi chez son hôte humain. Le transfert vers la cellule hôte de l'un de ces ARNi a pu être mis en évidence par une approche d'hybridation *in situ* du transcrit, révélant de petits amas sphériques marqués dans le cytoplasme de la cellule hôte, pouvant rappeler des vésicules. Une libération directement dans le cytoplasme de la cellule hôte est également une hypothèse plausible. De plus, les marquages observés dans le noyau des cellules hôtes soulèvent la question d'une activité sur l'épigénétique de l'hôte (192).

De manière intéressante, ces ARN interférents réprimerait des gènes associés à l'apoptose ainsi qu'au trafic intracellulaire. L'inhibition de l'apoptose suite à des infections microsporidiennes a été décrite chez plusieurs espèces, et plusieurs mécanismes ont été proposés pour celle-ci (107,366–368). Une validation biologique de l'effet anti-apoptotique de l'irRNA-1 d'*A.algerae* reste nécessaire afin de pouvoir valider notre hypothèse. La méthodologie que nous avons mise en place, avait pour but de sélectionner les ARNi et leurs



cibles potentielles chez l'hôte, impliquées dans des voies cellulaires aux phénotypes visibles, en vue de cette validation. Si le processus de régulation inter-espèce que nous suggérons s'avérait réel, la question de son acquisition se poserait alors. En effet, Les gènes encodant les iRNA identifiés chez *A.algerae* ne sont pas retrouvés chez d'autres espèces microsporidiennes. Ces gènes sont pour la plupart retrouvés à proximité d'ET pouvant témoigner d'un lien probable entre les ET et l'apparition de ces gènes. Chez le pathogène de plantes *B. cinerea* des ARNi issus d'ET avaient été montrés comme ciblant et réprimant des gènes de l'immunité de la plante (203). Ces observations pourraient figurer une acquisition récente de ces gènes de manière fortuite par évolution de séquences non codantes et d'ET. Ces gènes auraient alors été conservés par rapport au bénéfice qu'ils apporteraient au parasite (369,370). Les ET représenteraient dans cette hypothèse un moyen d'évolution et d'acquisition rapide de nouvelles séquences non-codantes impliquées dans les interactions hôte-parasite (168).

Enfin, nous nous sommes intéressés à la réponse transcriptionnelle de cellules hôtes humaines infectées par *A.algerae* (371). Cette étude a révélé une surprenante répression généralisée des miARN humains. Au cours du cycle d'infection, le nombre de miRNA sous-exprimé semblait corrélér avec un grand nombre de gènes du parasite. Cette dérégulation pourrait en fait être corrélée à l'activité transcriptionnelle même d'*A.algerae*, ou à sa prolifération et ainsi être une conséquence directe de son développement. *A.algerae* pourrait donc utiliser les petits ARNnc de la cellule hôte comme ressource énergétique, soit en important et dégradant ces molécules, soit en stimulant leurs catabolismes au niveau du cytoplasme de l'hôte et en utilisant ensuite les nucléotides produits à son profit. Ce dernier point pourrait être en accord avec une étude ayant décrit une stimulation métabolique des cellules hôte infectées par des microsporidies mais avec maintien d'une homéostasie de l'ATP (372). Les auteurs observent une augmentation de métabolites liés à la glycolyse et au cycle de Krebs. Ils suggèrent également que les métabolites identifiés puissent stimuler la synthèse de nucléotides, d'acides gras et d'acides aminés par l'hôte. Tout ceci avec un maintien de l'homéostasie de l'hôte en ATP. Les microsporidies possèdent de nombreux transporteurs à ATP conservés dans le phylum (119,123), et sont dépourvues de réelles mitochondries même si un reliquat subsiste (100). Cela suggère une dépendance vis-à-vis de l'hôte pour l'ATP (373).



De plus, les microsporidies expriment des protéines NTT à leur surface (NucleoTide Transport proteins) par lesquels elles importeraient divers nucléotides (373,374). Cette stimulation métabolique pourrait ainsi être une conséquence de l'import de ressources par le parasite qui entraînerait une surproduction d'ATP compensatrice, utilisant peut-être les miRNA comme substrats.

En ce qui concerne les ARNm au cours de l'infection par *A.algerae*, aucune dérégulation de l'expression de gènes associés à la maturation ou à l'action des miRNA n'a été observée. Cependant, des dérégulations de gènes métaboliques et de nombreux gènes de sumoylation ont été identifiées. La sumoylation est un processus de modification post-traductionnelle des protéines qui assure de nombreux rôles physiologiques chez l'homme (375). Un de ces rôles concerne la mise en cargo des miRNA dans les transports intra et intercellulaires par des vésicules (376). En plus de celle des gènes de sumoylation, nous avons pu mettre en évidence une dérégulation d'autres transcrits associés au transport intracellulaire. Certains iRNA d'*A.algerae* ont également été prédits comme ciblant des gènes associés au trafic intracellulaire. Ces problématiques de transport semblent particulièrement importantes dans la biologie infectieuse d'*A.algerae* et il est possible que suite à ces perturbations, des vésicules produites par la cellule hôte et contenant des miRNA soient mal adressées : vers l'extérieur de la cellule ; vers des systèmes de dégradation comme les lysosomes ; ou encore vers le parasite. Dans les trois cas, une réduction de la quantité de miRNA dans les cellules hôtes serait alors induite. Plusieurs mécanismes de subversion (ou Hacking) parasitaire des ressources de l'hôte ont été décrits chez les microsporidies (105,377). Dans le cadre de notre dernière hypothèse, les miRNA humains, loin de leur rôle de régulation transcriptionnelle, représenteraient une ressource énergétique ou un substrat de base pour permettre au parasite de réaliser son cycle de développement intracellulaire.

Pour conclure, l'étude du rôle des ARNnc dans les interactions hôte-parasite via un modèle d'interaction entre la microsporidie *Anncaliia algerae* et des cellules humaines a permis : i) d'illustrer et caractériser d'avantage la diversité des ARNnc parasitaires et leurs implications dans les interactions avec l'hôte ii) de proposer un mécanisme de régulation transcriptomique inter-espèce chez les microsporidies, ainsi qu'un modèle de son acquisition au cours de l'évolution, et enfin, iii) de souligner encore d'avantage la place des ARNnc dans l'étude et la compréhension des interactions entre hôtes et parasites.



PARTIE II PERSPECTIVES

Les différents travaux menés et les résultats obtenus ouvrent des perspectives qu'il serait pertinent d'étudier afin de continuer à élucider les nombreux mécanismes mis en jeu par les microsporidies lors de leurs cycles infectieux.

Tout d'abord, le large répertoire de LncRNA identifiés chez *A.algerae* pose la question de leurs rôles. De plus, les LncRNA sont de plus en plus décrits comme pouvant coder de petits peptides et l'identification de petites protéines ou smORF-encoded polypeptide (SEPS) appartenant jusqu'alors au « dark proteome » serait primordial pour considérer l'ensemble des acteurs moléculaire de ce dialogue hôte-pathogène. En effet, à l'heure actuelle seule une fraction de ces petits peptides, représentant la pointe de ce nouvel iceberg de la biologie, a donné lieu à des études visant à appréhender leur diversité et leur hétérogénéité fonctionnelle. Ces petits peptides seraient des régulateurs importants de l'activité d'autres protéines influençant leurs conformations et leurs interactions (378). Dans ce sens, un projet de recherche a été initié dans l'équipe IHP, et viserait à identifier et caractériser ces peptides au cours d'infections, par des approches de protéomiques.

Concernant les processus d'ARNi chez *A.algerae*, la validation biologique de l'interaction entre l'ARNi-1 d'*A.algerae* et de son gène cible prédit CECR2 dans des cellules humaines est en cours. Si cette expérience conforte les prédictions, elle validerait un exemple fort d'interaction hôte-parasite médiée par les ARNnc et ayant des conséquences physiologiques probantes. Cela inviterait à tester l'action des autres ARNi identifiés sur les gènes cibles prédits par notre approche, mais également à étendre la recherche de nouveaux ARNi à d'autres espèces microsporidiennes.

La réalisation d'anticorps spécifiques des protéines Argonaute, Dicer et RdRP d'*A. algerae* a également été initiée. Pour cela, des fragments de ces trois protéines ont été clonés et exprimés dans un système bactérien afin d'assurer la production de peptides recombinants en vue d'une immunisation de souris. La fabrication d'hybridomes a également été considérée afin d'avoir une source pérenne et reproductible de ces anticorps. A terme, ces anticorps permettraient une étude plus poussée des mécanismes d'ARNi mis en jeu dans l'interaction entre *A. algerae* et sa cellule hôte. Cette approche permettrait de répondre à de



nombreuses questions : lorsque *A. algerae* transfère de petits ARNi vers sa cellule hôte, ces ARNi sont-ils libres ou complexés à des protéines ? Sont-ils liés à une protéine Argonaute, celle du parasite ou une de l'hôte ? La protéine Dicer d'*A. algerae* présente-t-elle des affinités de clivage particulières ? La RdRP nécessite-t-elle une signature spécifique pour compléter un transcrit ?

Il serait également intéressant de caractériser plus précisément les complexes associés à ces trois protéines en termes de taille, de diversité et d'abondance. Cela pourrait être effectué par des approches d'immunoprécipitation (IP) et de spectrométrie de masse (379,380). Ces deux approches pourraient également être combinées à du séquençage RNAseq afin de mieux caractériser les ARNnc liés à ces trois protéines, favorisant ainsi l'identification de motifs communs ou de modifications post-transcriptionnelles propres à ces ARNnc.

Le développement d'approches d'hybridation *in situ* et de CLIP-seq (381) serait également pertinent afin de localiser ces protéines au cours du cycle d'*A. algerae* et de caractériser des interactions entre ARN interférent et cible.

Pour une étude approfondie des processus génomiques et post-génomiques chez *A. algerae*, de nombreuses autres méthodes de séquençage seraient intéressantes à développer tel que le Riboseq, le Chipseq ou encore l'ATACseq (382–384), afin de mieux caractériser l'organisation de la chromatine et d'appréhender certaines modifications épigénétiques chez *A. algerae* voire de caractériser la diversité de ces modifications. Le riboseq pourrait également aider dans la meilleure caractérisation du phénomène d'overrun transcriptionnel, ainsi que dans la détection des petits ORFs. Des technologies de séquençage directe de l'ARN se développent également de plus en plus ces dernières années (385).

Des approches de séquençage de cellule isolée (single-cell) ouvriraient la voie à des études plus précises du polymorphisme chez *A. algerae*, et permettraient une étude plus fine des dynamiques d'ET au sein du génome de ces parasites. Sur l'aspect transcriptomique, ces approches en cellule isolée offriraient des suivis de dynamiques populationnelles et de diversité physiologique et fonctionnelle du parasite. Ces approches demanderaient une mise au point considérable et des méthodologies dédiées, *A. algerae* étant un parasite intracellulaire. Cependant elle permettrait de s'affranchir de nombreux biais liés à l'étude d'une culture entière, où toutes les cellules ne sont pas forcément infectées ni toujours au



même stade de développement. Les méthodes computationnelles disponibles permettraient de différencier les stades et même d'identifier les phénotypes de transition entre stades.

Les nouvelles ressources de génomique et de transcriptomique produites au cours de ce projet de thèse pourraient enrichir les connaissances sur les microsporidies et peuvent servir de base pour promouvoir des études comparatives entre espèces microsporidiennes. Cela conduirait à une meilleure connaissance des espèces retrouvées dans ce phylum et permettrait de mettre en évidence les mécanismes clés derrière son évolution et sa diversité (132,351).

De nombreux autres mécanismes d'interaction existent n'impliquant pas les ARNnc. Des approches intégrées « multi-omics » clarifieraient sûrement d'avantage la place qu'occupent les ARNnc et les autres types de molécules au sein des interactions entre hôte et parasite.

Enfin, il a été décrit des interactions antagonistes lors de co-infections entre des microsporidies comme *microsporidia* MB et des *Plasmodium* dans l'anophèle (386). Les mécanismes sous-jacents à cette interaction demeurent inconnus. Nous pouvons nous demander s'ils sont médiés par le métabolisme, l'immunité, d'autres process, ou encore, par des ARNnc.

La diversité des modes d'interactions du vivant continue de se révéler à la science et aux Hommes. L'émerveillement que procure l'observation et la compréhension du monde qui nous entoure continuera, je l'espère, à nous guider vers une paix et une harmonie joyeuse avec ce qui nous entoure.



SECTION D : RÉFÉRENCES



SECTION D RÉFÉRENCES

1. Jheeta S. The Landscape of the Emergence of Life. *Life*. 16 juin 2017;7(2):27.
2. Chen IA. The Emergence of Cells During the Origin of Life. *Science*. 8 déc 2006;314(5805):1558-9.
3. Nair A, Chauhan P, Saha B, Kubatzky KF. Conceptual Evolution of Cell Signaling. *Int J Mol Sci*. 4 juill 2019;20(13):3292.
4. Drew GC, Stevens EJ, King KC. Microbial evolution and transitions along the parasite–mutualist continuum. *Nat Rev Microbiol*. oct 2021;19(10):623-38.
5. Barraclough TG. How Do Species Interactions Affect Evolutionary Dynamics Across Whole Communities? *Annu Rev Ecol Evol Syst*. 2015;46(1):25-48.
6. Horta MF, Andrade LO, Martins-Duarte ÉS, Castro-Gomes T. Cell invasion by intracellular parasites – the many roads to infection. *J Cell Sci*. 20 févr 2020;133(4):jcs232488.
7. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet*. févr 2021;22(2):71-88.
8. Brücher BLD, Jamall IS. Cell-Cell Communication in the Tumor Microenvironment, Carcinogenesis, and Anticancer Treatment. *Cell Physiol Biochem*. 8 juill 2014;34(2):213-43.
9. Berumen Sánchez G, Bunn KE, Pua HH, Rafat M. Extracellular vesicles: mediators of intercellular communication in tissue injury and disease. *Cell Commun Signal*. 16 oct 2021;19(1):104.
10. Zhang Y, Liu Y, Liu H, Tang WH. Exosomes: biogenesis, biologic function and clinical potential. *Cell Biosci*. 15 févr 2019;9(1):19.
11. Chen Y, Zhao Y, Yin Y, Jia X, Mao L. Mechanism of cargo sorting into small extracellular vesicles. *Bioengineered*. 12(1):8186-201.
12. Olano JP, Weller PF, Guerrant RL, Walker DH. Principles of Parasitism. *Trop Infect Dis Princ Pathog Pract*. 2011;1-7.
13. Theel ES, Pritt BS. Parasites. *Microbiol Spectr*. 12 août 2016;4(4):10.1128/microbiolspec.dmih2-0013-2015.
14. Torgerson PR. One world health: Socioeconomic burden and parasitic disease control priorities. *Vet Parasitol*. 1 août 2013;195(3):223-32.

15. Rist CL, Garchitorena A, Ngonghala CN, Gillespie TR, Bonds MH. The Burden of Livestock Parasites on the Poor. *Trends Parasitol.* 1 nov 2015;31(11):527-30.
16. Rohr JR, Barrett CB, Civitello DJ, Craft ME, Delius B, DeLeo GA, et al. Emerging human infectious diseases and the links to global food production. *Nat Sustain.* juin 2019;2(6):445-56.
17. Flores JM, Gámiz V, Jiménez-Marín Á, Flores-Cortés A, Gil-Lebrero S, Garrido JJ, et al. Impact of *Varroa destructor* and associated pathologies on the colony collapse disorder affecting honey bees. *Res Vet Sci.* mars 2021;135:85-95.
18. Walker DM, Oghumu S, Gupta G, McGwire BS, Drew ME, Satoskar AR. Mechanisms of cellular invasion by intracellular parasites. *Cell Mol Life Sci CMLS.* avr 2014;71(7):1245-63.
19. Thomson S, Hamilton CA, Hope JC, Katzer F, Mabbott NA, Morrison LJ, et al. Bovine cryptosporidiosis: impact, host-parasite interaction and control strategies. *Vet Res.* 11 août 2017;48(1):42.
20. Texier C, Vidau C, Viguès B, El Alaoui H, Delbac F. Microsporidia: a model for minimal parasite–host interactions. *Curr Opin Microbiol.* 1 août 2010;13(4):443-9.
21. Han B, Weiss LM. Microsporidia: Obligate Intracellular Pathogens within the Fungal Kingdom. *Microbiol Spectr.* avr 2017;5(2):10.1128/microbiolspec.FUNK-0018-2016.
22. Manuels MSD pour le grand public [Internet]. [cité 25 août 2023]. Présentation des infections parasitaires - Infections. Disponible sur: <https://www.msdmanuals.com/fr/accueil/infections/infections-parasitaires-pr%C3%A9sentation/pr%C3%A9sentation-des-infections-parasitaires>
23. Seeber F, Steinfeld S. Recent advances in understanding apicomplexan parasites. *F1000Research.* 14 juin 2016;5:F1000 Faculty Rev-1369.
24. Dos Santos Pacheco N, Tosetti N, Koreny L, Waller RF, Soldati-Favre D. Evolution, Composition, Assembly, and Function of the Conoid in Apicomplexa. *Trends Parasitol.* août 2020;36(8):688-704.
25. Librairie Lavoisier [Internet]. [cité 29 août 2023]. Parasitologie et mycologie médicales MOULINIER Claude. Disponible sur: <https://www.lavoisier.fr/livre/medecine/parasitologie-et-mycologie-medicales/moulinier/descriptif-9782743004880>
26. Beck HP, Blake D, Dardé ML, Felger I, Pedraza-Díaz S, Regidor-Cerrillo J, et al. Molecular approaches to diversity of populations of apicomplexan parasites. *Int J Parasitol.* 15 janv 2009;39(2):175-89.

27. Tomazic ML, Marugan-Hernandez V, Rodriguez AE. Next-Generation Technologies and Systems Biology for the Design of Novel Vaccines Against Apicomplexan Parasites. *Front Vet Sci* [Internet]. 2022 [cité 29 août 2023];8. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fvets.2021.800361>
28. Su X zhuan, Wu J. Zoonotic Transmissions and Host Switches of Malaria Parasites. *Zoonoses Burlingt Mass.* 2021;1(1):11.
29. Ayala SC. Checklist, Host Index, and Annotated Bibliography of Plasmodium from Reptiles*. *J Protozool.* 1978;25(1):87-100.
30. Sato S. Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. *J Physiol Anthropol.* 7 janv 2021;40:1.
31. PNLPCOTEDIVOIRE. PALUDISME | Paludisme Généralités [Internet]. [cité 3 sept 2023]. Disponible sur: <https://www.pnlpcotedivoire.org/paludisme-generalites/>
32. Prevention CC for DC and. CDC - Malaria - About Malaria - Biology [Internet]. 2020 [cité 3 sept 2023]. Disponible sur: <https://www.cdc.gov/malaria/about/biology/index.html>
33. Paludisme [Internet]. [cité 29 août 2023]. Disponible sur: <https://www.who.int/fr/news-room/fact-sheets/detail/malaria>
34. Laurens MB. Novel malaria vaccines. *Hum Vaccines Immunother.* 17(11):4549-52.
35. Chinappi M, Via A, Marcatili P, Tramontano A. On the mechanism of chloroquine resistance in Plasmodium falciparum. *PloS One.* 19 nov 2010;5(11):e14064.
36. Su X zhuan, Lane KD, Xia L, Sá JM, Wellems TE. Plasmodium Genomics and Genetics: New Insights into Malaria Pathogenesis, Drug Resistance, Epidemiology, and Evolution. *Clin Microbiol Rev.* 31 juill 2019;32(4):e00019-19.
37. Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* janv 2009;37(Database issue):D539-543.
38. Cunningham D, Lawton J, Jarra W, Preiser P, Langhorne J. The pir multigene family of Plasmodium: Antigenic variation and beyond. *Mol Biochem Parasitol.* 1 avr 2010;170(2):65-73.
39. Singh V, Gupta P, Pande V. Revisiting the multigene families: Plasmodium var and vir genes. *J Vector Borne Dis.* juin 2014;51(2):75-81.
40. Real E, Nardella F, Scherf A, Mancio-Silva L. Repurposing of Plasmodium falciparum var genes beyond the blood stage. *Curr Opin Microbiol.* 1 déc 2022;70:102207.

41. Dzikowski R, Deitsch KW. Genetics of Antigenic Variation in *Plasmodium falciparum*. *Curr Genet. avr* 2009;55(2):103-10.
42. Jabeena CA, Rajavelu A. Epigenetic Players of Chromatin Structure Regulation in *Plasmodium falciparum*. *Chembiochem Eur J Chem Biol.* 15 mai 2019;20(10):1225-30.
43. Sanchez SG, Besteiro S. The pathogenicity and virulence of *Toxoplasma gondii*. *Virulence.* 12(1):3095-114.
44. Lehmann T, Marcet PL, Graham DH, Dahl ER, Dubey JP. Globalization and the population structure of *Toxoplasma gondii*. *Proc Natl Acad Sci U S A.* 25 juill 2006;103(30):11423-8.
45. Attias M, Teixeira DE, Benchimol M, Vommaro RC, Crepaldi PH, De Souza W. The life-cycle of *Toxoplasma gondii* reviewed using animations. *Parasit Vectors.* 23 nov 2020;13(1):588.
46. S. Al-Malki E. Toxoplasmosis: stages of the protozoan life cycle and risk assessment in humans and animals for an enhanced awareness and an improved socio-economic status. *Saudi J Biol Sci.* janv 2021;28(1):962-9.
47. Diagnostic et traitement de la toxoplasmose [Internet]. [cité 29 août 2023]. Disponible sur: <https://www.ameli.fr/assure/sante/themes/toxoplasmose/diagnostic-traitement>
48. Dubey JP. *Toxoplasma Gondii*. In: Baron S, éditeur. *Medical Microbiology* [Internet]. 4th éd. Galveston (TX): University of Texas Medical Branch at Galveston; 1996 [cité 29 août 2023]. Disponible sur: <http://www.ncbi.nlm.nih.gov/books/NBK7752/>
49. VIDAL [Internet]. [cité 29 août 2023]. Pyriméthamine : substance active à effet thérapeutique. Disponible sur: <https://www.vidal.fr/medicaments/substances/pyrimethamine-2980.html>
50. Lau YL, Lee WC, Gudimella R, Zhang G, Ching XT, Razali R, et al. Deciphering the Draft Genome of *Toxoplasma gondii* RH Strain. *PLoS ONE.* 29 juin 2016;11(6):e0157901.
51. Massive invasion of organellar DNA drives nuclear genome evolution in *Toxoplasma* | PNAS [Internet]. [cité 8 nov 2023]. Disponible sur: <https://www.pnas.org/doi/abs/10.1073/pnas.2308569120>
52. Yucesan B, Guldemir D, Babur C, Kilic S, Cakmak A. Whole-genome sequencing of a *Toxoplasma gondii* strain from a Turkish isolate using next-generation sequencing technology. *Acta Trop.* juin 2021;218:105907.
53. Harb OS, Roos DS. ToxoDB: Functional Genomics Resource for *Toxoplasma* and Related Organisms. *Methods Mol Biol Clifton NJ.* 2020;2071:27-47.
54. Kim K, Weiss LM. *Toxoplasma*: the next 100 years. *Microbes Infect Inst Pasteur.* juill 2008;10(9):978-84.
55. Hosseini SA, Amouei A, Sharif M, Sarvi Sh, Galal L, Javidnia J, et al. Human toxoplasmosis: a systematic review for genetic diversity of *Toxoplasma gondii* in clinical samples. *Epidemiol Infect.* 5 nov 2018;147:e36.

56. Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants | Genome Biology | Full Text [Internet]. [cité 8 nov 2023]. Disponible sur: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-5-r53>
57. CDC - DPDx - Cryptosporidiosis [Internet]. 2022 [cité 3 sept 2023]. Disponible sur: <https://www.cdc.gov/dpdx/cryptosporidiosis/index.html>
58. Feng Y, Ryan UM, Xiao L. Genetic Diversity and Population Structure of *Cryptosporidium*. Trends Parasitol. 1 nov 2018;34(11):997-1011.
59. Hassan EM, Örmeci B, DeRosa MC, Dixon BR, Sattar SA, Iqbal A. A review of *Cryptosporidium* spp. and their detection in water. Water Sci Technol. 28 oct 2020;83(1):1-25.
60. Thompson RCA, Olson ME, Zhu G, Enomoto S, Abrahamsen MS, Hijjawi NS. *Cryptosporidium* and Cryptosporidiosis. In: Baker JR, Muller R, Rollinson D, éditeurs. Advances in Parasitology [Internet]. Academic Press; 2005 [cité 29 août 2023]. p. 77-158. Disponible sur: <https://www.sciencedirect.com/science/article/pii/S0065308X0559002X>
61. Clevers H. The Intestinal Crypt, A Prototype Stem Cell Compartment. Cell. 18 juill 2013;154(2):274-84.
62. CNR cryptosporidioses [Internet]. [cité 29 août 2023]. Diagnostic et traitement des cryptosporidioses. Disponible sur: <https://cncryptosporidioses.chu-rouen.fr/espace-public/les-cryptosporidioses/diagnostic-traitement-crypto/>
63. Nitazoxanide [Internet]. [cité 29 août 2023]. Disponible sur: <https://go.drugbank.com/drugs/DB00507>
64. Rueckert S, Pipaliya SV, Dacks JB. Evolution: Parallel Paths to Parasitism in the Apicomplexa. Curr Biol. 9 sept 2019;29(17):R836-9.
65. Ryan U, Zahedi A, Feng Y, Xiao L. An Update on Zoonotic *Cryptosporidium* Species and Genotypes in Humans. Anim Open Access J MDPI. 19 nov 2021;11(11):3307.
66. Baptista RP, Cooper GW, Kissinger JC. Challenges for *Cryptosporidium* Population Studies. Genes. 10 juin 2021;12(6):894.
67. Fan Y, Feng Y, Xiao L. Comparative genomics: how has it advanced our knowledge of cryptosporidiosis epidemiology? Parasitol Res. 1 déc 2019;118(12):3195-204.
68. Owusu CK, Bennett HM. How apicomplexans became free-riders. Nat Rev Microbiol. oct 2015;13(10):603-603.

69. Mathur V, Kolísko M, Hehenberger E, Irwin NAT, Leander BS, Kristmundsson Á, et al. Multiple Independent Origins of Apicomplexan-Like Parasites. *Curr Biol.* 9 sept 2019;29(17):2936-2941.e5.
70. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michálek J, et al. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. Nordborg M, éditeur. *eLife.* 15 juill 2015;4:e06974.
71. Assembly and Motility of Eukaryotic Cilia and Flagella. Lessons from *Chlamydomonas reinhardtii* | *Plant Physiology* | Oxford Academic [Internet]. [cité 8 nov 2023]. Disponible sur: <https://academic.oup.com/plphys/article/127/4/1500/6103647?login=false>
72. Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, et al. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc Natl Acad Sci.* 18 août 2015;112(33):10200-7.
73. Swapna LS, Parkinson J. Genomics of Apicomplexan Parasites. *Crit Rev Biochem Mol Biol.* juin 2017;52(3):254-73.
74. Kissinger JC, DeBarry J. Genome cartography: charting the apicomplexan genome. *Trends Parasitol.* août 2011;27(8):345.
75. Wasmuth J, Daub J, Peregrín-Alvarez JM, Finney CAM, Parkinson J. The origins of apicomplexan sequence innovation. *Genome Res.* juill 2009;19(7):1202-13.
76. Boisard J, Florent I. Why the -omic future of Apicomplexa should include gregarines. *Biol Cell.* juin 2020;112(6):173-85.
77. Mallik P, Bandyopadhyay PK. Observation on *Monocystis kuidongae* sp. nov. (Apicomplexa: Eugregarinida) from an Indian earthworm *Perionyx excavatus* Perrier (Annelida: Oligochaeta). *J Parasit Dis Off Organ Indian Soc Parasitol.* juin 2017;41(2):361-3.
78. Rueckert S, Betts EL, Tsaousis AD. The Symbiotic Spectrum: Where Do the Gregarines Fit? *Trends Parasitol.* 1 sept 2019;35(9):687-94.
79. CDC - DPDx - American Trypanosomiasis [Internet]. 2021 [cité 3 sept 2023]. Disponible sur: <https://www.cdc.gov/dpdx/trypanosomiasisamerican/index.html>
80. Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, et al. Kinetoplastids: related protozoan pathogens, different diseases. *J Clin Invest.* 1 avr 2008;118(4):1301-10.

81. Cavalcanti DP, de Souza W. The Kinetoplast of Trypanosomatids: From Early Studies of Electron Microscopy to Recent Advances in Atomic Force Microscopy. *Scanning*. 19 juin 2018;2018:9603051.
82. Martín-Escolano J, Marín C, Rosales MJ, Tsaousis AD, Medina-Carmona E, Martín-Escolano R. An Updated View of the *Trypanosoma cruzi* Life Cycle: Intervention Points for an Effective Treatment. *ACS Infect Dis*. 10 juin 2022;8(6):1107-15.
83. Zuma AA, Dos Santos Barrias E, de Souza W. Basic Biology of *Trypanosoma cruzi*. *Curr Pharm Des*. 2021;27(14):1671-732.
84. CDC - DPDx - Leishmaniasis [Internet]. 2019 [cité 3 sept 2023]. Disponible sur: <https://www.cdc.gov/dpdx/leishmaniasis/index.html>
85. Bartholomeu DC, Teixeira SMR, Cruz AK. Genomics and functional genomics in *Leishmania* and *Trypanosoma cruzi*: statuses, challenges and perspectives. *Mem Inst Oswaldo Cruz*. 2021;116:e200634.
86. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 15 juill 2005;309(5733):404-9.
87. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 15 juill 2005;309(5733):416-22.
88. Faria JRC. A nuclear enterprise: zooming in on nuclear organization and gene expression control in the African trypanosome. *Parasitology*. sept 2021;148(10):1237-53.
89. Alsford NS, Navarro M, Jamnadass HR, Dunbar H, Ackroyd M, Murphy NB, et al. The identification of circular extrachromosomal DNA in the nuclear genome of *Trypanosoma brucei*. *Mol Microbiol*. janv 2003;47(2):277-89.
90. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The Genome of the Kinetoplastid Parasite, *Leishmania major*. *Science*. 15 juill 2005;309(5733):436-42.
91. Herreros-Cabello A, Callejas-Hernández F, Gironès N, Fresno M. *Trypanosoma Cruzi* Genome: Organization, Multi-Gene Families, Transcription, and Biological Implications. *Genes*. 14 oct 2020;11(10):1196.
92. Sanchez MCH, Zapata HSO, Suarez BD, Ospina C, Barbosa HJ, Martinez JCC, et al. A nearly complete and phased genome assembly of a Colombian *Trypanosoma cruzi* Tc1 strain and the evolution of gene families [Internet]. *bioRxiv*; 2023 [cité 19 nov 2023]. p. 2023.07.17.549441. Disponible sur: <https://www.biorxiv.org/content/10.1101/2023.07.17.549441v1>

93. Wang W, Peng D, Baptista RP, Li Y, Kissinger JC, Tarleton RL. Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. *PLoS Pathog.* 28 janv 2021;17(1):e1009254.
94. Ribeiro YC, Robe LJ, Veluza DS, dos Santos CMB, Lopes ALK, Krieger MA, et al. Study of VIPER and TATE in kinetoplastids and the evolution of tyrosine recombinase retrotransposons. *Mob DNA.* 5 août 2019;10:34.
95. Pita S, Díaz-Viraqué F, Iraola G, Robello C. The Tritryps Comparative Repeatome: Insights on Repetitive Element Evolution in Trypanosomatid Pathogens. *Genome Biol Evol.* 4 févr 2019;11(2):546-51.
96. Franzen C. Microsporidia: how can they invade other cells? *Trends Parasitol.* juin 2004;20(6):275-9.
97. Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol.* 1 août 2022;38(8):642-59.
98. Weiss LM, Reinke AW, éditeurs. *Microsporidia: Current Advances in Biology* [Internet]. Cham: Springer International Publishing; 2022 [cité 30 août 2023]. (Experientia Supplementum; vol. 114). Disponible sur: <https://link.springer.com/10.1007/978-3-030-93306-7>
99. Peyretilade E, Biderre C, Peyret P, Duffieux F, Méténier G, Gouy M, 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.* 1 août 1998;26(15):3513-20.
100. Burri L, Williams BAP, Bursac D, Lithgow T, Keeling PJ. Microsporidian mitosomes retain elements of the general mitochondrial targeting system. *Proc Natl Acad Sci U S A.* 24 oct 2006;103(43):15916-20.
101. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature.* nov 2001;414(6862):450-3.
102. Peyretilade E, Broussolle V, Peyret P, Méténier G, Gouy M, Vivarès CP. Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Mol Biol Evol.* juin 1998;15(6):683-9.
103. Han B, Takvorian PM, Weiss LM. Invasion of Host Cells by Microsporidia. *Front Microbiol.* 18 févr 2020;11:510280.
104. Dynamics of polar filament discharge and sporoplasm expulsion by microsporidian spores - Frixione - 1992 - Cell Motility - Wiley Online Library [Internet]. [cité 30 août 2023]. Disponible sur: <https://onlinelibrary.wiley.com/doi/10.1002/cm.970220105>
105. Jarkass HTE, Reinke AW. The ins and outs of host-microsporidia interactions during invasion, proliferation and exit. *Cell Microbiol.* 1 nov 2020;22(11):e13247.

106. Weiss LM, Takvorian PM. *Anncaliia algerae*. Trends Parasitol. 1 août 2021;37(8):762-3.
107. Han Y, Gao H, Xu J, Luo J, Han B, Bao J, et al. Innate and Adaptive Immune Responses Against Microsporidia Infection in Mammals. Front Microbiol [Internet]. 2020 [cité 4 sept 2023];11. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01468>
108. Carriere E, Abdul Hamid AI, Feki I, Dubuffet A, Delbac F, Gueirard P. A mouse ear skin model to study the dynamics of innate immune responses against the microsporidian *Encephalitozoon cuniculi*. Front Microbiol. 2023;14:1168970.
109. Stentiford GD, Becnel J., Weiss LM, Keeling PJ, Didier ES, Williams B., et al. Microsporidia – Emergent Pathogens in the Global Food Chain. Trends Parasitol. 1 avr 2016;32(4):336-48.
110. Dubuffet A, Chauvet M, Moné A, Debroas D, Lepère C. A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. Environ Microbiol. 2021;23(8):4344-59.
111. Jespersen N, Monroy L, Barandun J. Impact of Genome Reduction in Microsporidia. In: Weiss LM, Reinke AW, éditeurs. Microsporidia: Current Advances in Biology [Internet]. Cham: Springer International Publishing; 2022 [cité 3 sept 2023]. p. 1-42. (Experientia Supplementum). Disponible sur: https://doi.org/10.1007/978-3-030-93306-7_1
112. Han B, Pan G, Weiss LM. Microsporidiosis in Humans. Clin Microbiol Rev. 34(4):e00010-20.
113. Ghosh K, Weiss LM. T cell response and persistence of the microsporidia. FEMS Microbiol Rev. 1 mai 2012;36(3):748-60.
114. Redondo F, Hurtado-Marcos C, Izquierdo F, Cuéllar C, Fenoy S, Sáez Y, et al. Latent Microsporidia Infection Prevalence as a Risk Factor in Colon Cancer Patients. Cancers. janv 2022;14(21):5342.
115. Ruan Y, Xu X, He Q, Li L, Guo J, Bao J, et al. The largest meta-analysis on the global prevalence of microsporidia in mammals, avian and water provides insights into the epidemic features of these ubiquitous pathogens. Parasit Vectors. 1 avr 2021;14(1):186.
116. VIDAL [Internet]. [cité 30 août 2023]. Albendazole : substance active à effet thérapeutique. Disponible sur: <https://www.vidal.fr/medicaments/substances/albendazole-3934.html>
117. VIDAL [Internet]. [cité 30 août 2023]. Fumagilline : substance active à effet thérapeutique. Disponible sur: <https://www.vidal.fr/medicaments/substances/fumagilline-22818.html>

118. Peyretailade E, El Alaoui H, Diogon M, Polonais V, Parisot N, Biron DG, et al. Extreme reduction and compaction of microsporidian genomes. *Res Microbiol.* 1 juill 2011;162(6):598-606.
119. Wadi L, Reinke AW. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLOS Pathog.* 13 févr 2020;16(2):e1008276.
120. Reinke AW, Balla KM, Bennett EJ, Troemel ER. Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. *Nat Commun.* 9 janv 2017;8:14023.
121. Luallen RJ, Reinke AW, Tong L, Botts MR, Félix MA, Troemel ER. Discovery of a Natural Microsporidian Pathogen with a Broad Tissue Tropism in *Caenorhabditis elegans*. *PLoS Pathog.* 30 juin 2016;12(6):e1005724.
122. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, et al. Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. *Genome Biol Evol.* 2013;5(12):2285-303.
123. He Q, Vossbrinck CR, Yang Q, Meng XZ, Luo J, Pan GQ, et al. Evolutionary and functional studies on microsporidian ATP-binding cassettes: Insights into the adaptation of microsporidia to obligated intracellular parasitism. *Infect Genet Evol.* 1 mars 2019;68:136-44.
124. Corradi N. Microsporidia: Eukaryotic Intracellular Parasites Shaped by Gene Loss and Horizontal Gene Transfers. *Annu Rev Microbiol.* 2015;69(1):167-83.
125. Khalaf A, Jaron KS, Blaxter ML, Lawniczak MKN. Polyploidy is widespread in Microsporidia [Internet]. *Evolutionary Biology*; 2023 sept [cité 4 nov 2023]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2023.09.29.560119>
126. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun.* 13 mai 2015;6(1):7121.
127. Williams BA, Lee RC, Becnel JJ, Weiss LM, Fast NM, Keeling PJ. Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. *BMC Genomics.* 29 avr 2008;9:200.
128. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. *Genome Biol Evol.* 1 sept 2014;6(9):2289-300.

129. de Albuquerque NRM, Ebert D, Haag KL. Transposable element abundance correlates with mode of transmission in microsporidian parasites. *Mob DNA*. 23 juin 2020;11:19.
130. Chen L, Gao X, Li R, Zhang L, Huang R, Wang L, et al. Complete genome of a unicellular parasite (*Antonospora locustae*) and transcriptional interactions with its host locust. *Microb Genomics*. 12 août 2020;6(9):mgen000421.
131. Peyretailade E, Gonçalves O, Terrat S, Dugat-Bony E, Wincker P, Cornman RS, et al. Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: support for accurate structural genome annotation. *BMC Genomics*. 15 déc 2009;10:607.
132. Belkorchia A, Pombert JF, Polonais V, Parisot N, Delbac F, Brugère JF, et al. Comparative genomics of microsporidian genomes reveals a minimal non-coding RNA set and new insights for transcription in minimal eukaryotic genomes. *DNA Res Int J Rapid Publ Rep Genes Genomes*. juin 2017;24(3):251-60.
133. Corradi N, Gangaeva A, Keeling PJ. Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics*. avr 2008;91(4):388-93.
134. Lee RCH, Gill EE, Roy SW, Fast NM. Constrained Intron Structures in a Microsporidian. *Mol Biol Evol*. sept 2010;27(9):1979-82.
135. Whelan TA, Lee NT, Lee RCH, Fast NM. Microsporidian Introns Retained against a Background of Genome Reduction: Characterization of an Unusual Set of Introns. *Genome Biol Evol*. 29 nov 2018;11(1):263-9.
136. Mascarenhas dos Santos AC, Julian AT, Liang P, Juárez O, Pombert JF. Telomere-to-Telomere genome assemblies of human-infecting *Encephalitozoon* species. *BMC Genomics*. 4 mai 2023;24:237.
137. Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res*. janv 2011;39(Database issue):D612-619.
138. Peyretailade E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E, et al. Annotation of microsporidian genomes using transcriptional signals. *Nat Commun*. 16 oct 2012;3(1):1137.
139. Seatamanoch N, Kongdachalert S, Sunantaraporn S, Siriyasatien P, Brownell N. Microsporidia, a Highly Adaptive Organism and Its Host Expansion to Humans. *Front Cell Infect Microbiol* [Internet]. 2022 [cité 30 août 2023];12. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.924007>

140. Monaghan SR, Rumney RL, Vo NTK, Bols NC, Lee LEJ. In vitro growth of microsporidia *Anncaliia algerae* in cell lines from warm water fish. *Vitro Cell Dev Biol - Anim.* 1 févr 2011;47(2):104-13.
141. Belkorchia A, Biderre C, Militon C, Polonais V, Wincker P, Jubin C, et al. 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. *Parasitol Int.* mars 2008;57(1):62-71.
142. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* déc 2007;8(12):973-82.
143. Makałowski W, Gotea V, Pande A, Makałowska I. Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In: Anisimova M, éditeur. *Evolutionary Genomics: Statistical and Computational Methods* [Internet]. New York, NY: Springer; 2019 [cité 20 nov 2023]. p. 177-207. (Methods in Molecular Biology). Disponible sur: https://doi.org/10.1007/978-1-4939-9074-0_6
144. Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet.* 23 nov 2020;54:539-61.
145. Kapitonov VV, Jurka J. Self-synthesizing DNA transposons in eukaryotes. *Proc Natl Acad Sci U S A.* 21 mars 2006;103(12):4540-5.
146. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. - PMC [Internet]. [cité 9 nov 2023]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC365346/>
147. The role of retrotransposable elements in aging and age-associated diseases - PMC [Internet]. [cité 20 nov 2023]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8600649/>
148. Ribet D, Harper F, Dupressoir A, Dewannieux M, Pierron G, Heidmann T. An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* avr 2008;18(4):597-609.
149. Zhang L, Yan L, Jiang J, Wang Y, Jiang Y, Yan T, et al. The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. *Virulence.* 15 août 2014;5(6):655-64.
150. Verma AS, Singh IG, Bansal R, Singh A. Chapter 9 - HIV and Antiretroviral Drugs. In: Verma AS, Singh A, éditeurs. *Animal Biotechnology* [Internet]. San Diego: Academic Press; 2014 [cité 20 nov 2023]. p. 155-76. Disponible sur: <https://www.sciencedirect.com/science/article/pii/B9780124160026000092>

151. A new group of tyrosine recombinase-encoding retrotransposons - PubMed [Internet]. [cité 9 nov 2023]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/14963102/>
152. Kojima KK, Jurka J. Crypton transposons: identification of new diverse families and ancient domestication events. *Mob DNA*. 19 oct 2011;2:12.
153. Arkhipova IR. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mob DNA*. 6 déc 2017;8:19.
154. Transposases are the most abundant, most ubiquitous genes in nature - PMC [Internet]. [cité 9 nov 2023]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910039/>
155. Fricker AD, Peters JE. Vulnerabilities on the lagging-strand template: opportunities for mobile elements. *Annu Rev Genet*. 2014;48:167-86.
156. Spradling AC, Bellen HJ, Hoskins RA. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A*. 20 sept 2011;108(38):15948-53.
157. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinforma*. mars 2009;Chapter 4:4.10.1-4.10.14.
158. Huang CRL, Burns KH, Boeke JD. Active Transposition in Genomes. *Annu Rev Genet*. 2012;46:651-75.
159. Pritham EJ, Putliwala T, Feschotte C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene*. 1 avr 2007;390(1-2):3-17.
160. Pritham EJ, Feschotte C, Wessler SR. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. *Mol Biol Evol*. sept 2005;22(9):1751-63.
161. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet*. 2007;41:331-68.
162. Gilbert C, Peccoud J, Cordaux R. Transposable Elements and the Evolution of Insects. *Annu Rev Entomol*. 7 janv 2021;66:355-72.
163. Kines KJ, Belancio VP. Expressing genes do not forget their LINES: transposable elements and gene expression. *Front Biosci J Virtual Libr*. 1 janv 2012;17:1329-44.
164. Alexander WG, Wisecaver JH, Rokas A, Hittinger CT. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. *Proc Natl Acad Sci*. 12 avr 2016;113(15):4116-21.

165. Saliminejad K, Khorram Khorshid HR, Soleymani Fard S, Ghaffari SH. An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *J Cell Physiol.* mai 2019;234(5):5451-65.
166. Ramón y Cajal S, Segura MF, Hümmer S. Interplay Between ncRNAs and Cellular Communication: A Proposal for Understanding Cell-Specific Signaling Pathways. *Front Genet* [Internet]. 2019 [cité 1 sept 2023];10. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00281>
167. Li C, Ni YQ, Xu H, Xiang QY, Zhao Y, Zhan JK, et al. Roles and mechanisms of exosomal non-coding RNAs in human health and diseases. *Signal Transduct Target Ther.* 10 nov 2021;6(1):1-31.
168. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* févr 2009;10(2):94-108.
169. Zhang P, Wu W, Chen Q, Chen M. Non-Coding RNAs and their Integrated Networks. *J Integr Bioinforma.* 13 juill 2019;16(3):20190027.
170. Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning. In: Slaby O, Calin GA, éditeurs. *Non-coding RNAs in Colorectal Cancer* [Internet]. Cham: Springer International Publishing; 2016 [cité 2 sept 2023]. p. 3-17. (Advances in Experimental Medicine and Biology). Disponible sur: https://doi.org/10.1007/978-3-319-42059-2_1
171. Bhan A, Mandal SS. Long noncoding RNAs: emerging stars in gene regulation, epigenetics and human disease. *ChemMedChem.* sept 2014;9(9):1932-56.
172. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature.* déc 2008;456(7224):921-6.
173. Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol.* 2021;22(2):96-118.
174. van der Krol AR, Mur LA, Beld M, Mol JN, Stuitje AR. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell.* avr 1990;2(4):291-9.
175. Napoli C, Lemieux C, Jorgensen R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell.* avr 1990;2(4):279-89.
176. Matzke AJ, Neuhuber F, Park YD, Ambros PF, Matzke MA. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. *Mol Gen Genet MGG.* 2 août 1994;244(3):219-29.

177. Mittelsten Scheid O, Paszkowski J, Potrykus I. Reversible inactivation of a transgene in *Arabidopsis thaliana*. *Mol Gen Genet MGG*. août 1991;228(1-2):104-12.
178. Baulcombe DC. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol Biol*. oct 1996;32(1-2):79-88.
179. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. 29 oct 1999;286(5441):950-2.
180. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 19 févr 1998;391(6669):806-11.
181. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. janv 2001;409(6818):363-6.
182. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem*. 2015;84(1):405-33.
183. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, et al. The *rde-1* Gene, RNA Interference, and Transposon Silencing in *C. elegans*. *Cell*. 15 oct 1999;99(2):123-32.
184. Sen GL, Blau HM. A brief history of RNAi: the silence of the genes. *FASEB J Off Publ Fed Am Soc Exp Biol*. juill 2006;20(9):1293-9.
185. Martinez J, Patkaniowska A, Urlaub H, Lührmann R, Tuschl T. Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. *Cell*. 6 sept 2002;110(5):563-74.
186. Iwakawa H, Tomari Y. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol Cell*. 6 janv 2022;82(1):30-43.
187. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*. août 2014;15(8):509-24.
188. Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, et al. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet*. 23 nov 2018;52:131-57.
189. Micheel J, Safrastyan A, Wollny D. Advances in Non-Coding RNA Sequencing. *Non-Coding RNA*. 30 oct 2021;7(4):70.
190. Sun YM, Chen YQ. Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. *J Hematol Oncol J Hematol Oncol*. 10 août 2020;13(1):109.

191. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc B Biol Sci.* 12 janv 2009;364(1513):99-115.
192. Wei JW, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics (Review). *Oncol Rep.* janv 2017;37(1):3-9.
193. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature.* 22 janv 2009;457(7228):413-20.
194. Alshaer W, Zureigat H, Al Karaki A, Al-Kadash A, Gharaibeh L, Hatmal MM, et al. siRNA: Mechanism of action, challenges, and therapeutic approaches. *Eur J Pharmacol.* 15 août 2021;905:174178.
195. Chapman EJ, Carrington JC. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet.* nov 2007;8(11):884-96.
196. Gorski SA, Vogel J, Doudna JA. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol.* avr 2017;18(4):215-28.
197. Sontheimer EJ. Assembly and function of RNA silencing complexes. *Nat Rev Mol Cell Biol.* févr 2005;6(2):127-38.
198. Gu K, Mok L, Chong MMW. Regulating gene expression in animals through RNA endonucleolytic cleavage. *Heliyon.* 1 nov 2018;4(11):e00908.
199. ORBAN TI, IZAURRALDE E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA.* avr 2005;11(4):459-69.
200. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol.* 3 août 2018;9:402.
201. Gonzalez S, Pisano DG, Serrano M. Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. *Cell Cycle Georget Tex.* 15 août 2008;7(16):2601-8.
202. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science.* 13 sept 2002;297(5588):1833-7.
203. Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I, et al. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science.* 4 oct 2013;342(6154):118-23.
204. Cai Q, Qiao L, Wang M, He B, Lin FM, Palmquist J, et al. Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. *Science.* 8 juin 2018;360(6393):1126-9.

205. Huang CY, Wang H, Hu P, Hamby R, Jin H. Small RNAs – Big Players in Plant-Microbe Interactions. *Cell Host Microbe*. 14 août 2019;26(2):173-82.
206. Abels ER, Breakefield XO. Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake. *Cell Mol Neurobiol*. avr 2016;36(3):301-12.
207. O'Brien K, Breyne K, Ughetto S, Laurent LC, Breakefield XO. RNA delivery by extracellular vesicles in mammalian cells and its applications. *Nat Rev Mol Cell Biol*. oct 2020;21(10):585-606.
208. Urzi O, Gasparro R, Ganji NR, Alessandro R, Raimondo S. Plant-RNA in Extracellular Vesicles: The Secret of Cross-Kingdom Communication. *Membranes*. 23 mars 2022;12(4):352.
209. Hao L, Cai P, Jiang N, Wang H, Chen Q. Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics*. 21 janv 2010;11:55.
210. Rojas-Pirela M, Andrade-Alviárez D, Medina L, Castillo C, Liempi A, Guerrero-Muñoz J, et al. MicroRNAs: master regulators in host–parasitic protist interactions. *Open Biol*. 15 juin 2022;12(6):210395.
211. Wu H, Ye C, Ramirez D, Manjunath N. Alternative Processing of Primary microRNA Transcripts by Drosha Generates 5' End Variation of Mature microRNA. *PLOS ONE*. 27 oct 2009;4(10):e7566.
212. Kretov DA, Walawalkar IA, Mora-Martin A, Shafik AM, Moxon S, Cifuentes D. Ago2-Dependent Processing Allows miR-451 to Evade the Global MicroRNA Turnover Elicited during Erythropoiesis. *Mol Cell*. 16 avr 2020;78(2):317-328.e6.
213. Lee D, Shin C. MicroRNA–target interactions: new insights from genome-wide approaches. *Ann N Y Acad Sci*. 2012;1271(1):118-28.
214. Iwakawa HO, Tomari Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol*. nov 2015;25(11):651-65.
215. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 8 janv 2019;47(D1):D155-62.
216. Hansen TB, Kjems J, Bramsen JB. Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis. *RNA Biol*. 2011;8(3):378-83.
217. Stavast CJ, Erkeland SJ. The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. *Cells* [Internet]. nov 2019 [cité 3 sept 2023];8(11). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6912820/>

218. Sarshad AA, Juan AH, Muler AIC, Anastasakis DG, Wang X, Genzor P, et al. Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Mol Cell*. 9 sept 2018;71(6):1040.
219. Meng Y, Shao C, Ma X, Wang H. Introns targeted by plant microRNAs: a possible novel mechanism of gene regulation. *Rice*. 15 avr 2013;6(1):8.
220. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. janv 2009;19(1):92-105.
221. Huang W. *MicroRNAs: Biomarkers, Diagnostics, and Therapeutics*. Methods Mol Biol Clifton NJ. 2017;1617:57-67.
222. Tribolet L, Kerr E, Cowled C, Bean AGD, Stewart CR, Dearnley M, et al. MicroRNA Biomarkers for Infectious Diseases: From Basic Research to Biosensing. *Front Microbiol* [Internet]. 2020 [cité 3 sept 2023];11. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01197>
223. Liu J, Zhu L, Wang J, Qiu L, Chen Y, Davis RE, et al. Schistosoma japonicum extracellular vesicle miRNA cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathog*. 4 juin 2019;15(6):e1007817.
224. Pardini B, Calin GA. MicroRNAs and Long Non-Coding RNAs and Their Hormone-Like Activities in Cancer. *Cancers*. 17 mars 2019;11(3):378.
225. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. *Cell*. 20 févr 2009;136(4):642-55.
226. MicroRNA in Immune Regulation - PMC [Internet]. [cité 20 nov 2023]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8476048/>
227. Bayer-Santos E, Marini MM, da Silveira JF. Non-coding RNAs in Host-Pathogen Interactions: Subversion of Mammalian Cell Functions by Protozoan Parasites. *Front Microbiol* [Internet]. 2017 [cité 3 sept 2023];8. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmicb.2017.00474>
228. Navarro IC, Ferreira FM, Nakaya HI, Baron MA, Vilar-Pereira G, Pereira IR, et al. MicroRNA Transcriptome Profiling in Heart of Trypanosoma cruzi-Infected Mice: Parasitological and Cardiological Outcomes. *PLoS Negl Trop Dis*. 18 juin 2015;9(6):e0003828.
229. Leishmania donovani infection induce differential miRNA expression in CD4+ T cells | Scientific Reports [Internet]. [cité 20 nov 2023]. Disponible sur: <https://www.nature.com/articles/s41598-020-60435-2>
230. Frontiers | Interferon Gamma in Leishmaniasis [Internet]. [cité 20 nov 2023]. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fimmu.2013.00156/full>

231. Hu C, Dong Z, Deng B, Wu Q, Chen P, Lu C, et al. MicroRNA-6498-5p Inhibits *Nosema bombycis* Proliferation by Downregulating BmPLPP2 in *Bombyx mori*. *J Fungi*. 8 déc 2021;7(12):1051.
232. Fan X, Zhang W, Zhang K, Zhang J, Long Q, Wu Y, et al. In-depth investigation of microRNA-mediated cross-kingdom regulation between Asian honey bee and microsporidian. *Front Microbiol* [Internet]. 2022 [cité 3 sept 2023];13. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1003294>
233. Dandewad V, Vindu A, Joseph J, Seshadri V. Import of human miRNA-RISC complex into *Plasmodium falciparum* and regulation of the parasite gene expression. *J Biosci*. 16 avr 2019;44(2):50.
234. LaMonte G, Philip N, Reardon J, Lacsina JR, Majoros W, Chapman L, et al. Translocation of sickle cell erythrocyte microRNAs into *Plasmodium falciparum* inhibits parasite translation and contributes to malaria resistance. *Cell Host Microbe*. 16 août 2012;12(2):187-99.
235. Kataria P, Surela N, Chaudhary A, Das J. MiRNA: Biological Regulator in Host-Parasite Interaction during Malaria Infection. *Int J Environ Res Public Health*. 19 févr 2022;19(4):2395.
236. Wang J, Liu X, Jia B, Lu H, Peng S, Piao X, et al. A comparative study of small RNAs in *Toxoplasma gondii* of distinct genotypes. *Parasit Vectors*. 3 sept 2012;5(1):186.
237. Saçar MD, Bağcı C, Allmer J. Computational Prediction of MicroRNAs from *Toxoplasma gondii* Potentially Regulating the Hosts' Gene Expression. *Genomics Proteomics Bioinformatics*. oct 2014;12(5):228-38.
238. Huang Q, Evans JD. Identification of microRNA-like small RNAs from fungal parasite *Nosema ceranae*. *J Invertebr Pathol*. 1 janv 2016;133:107-9.
239. Shao SS, Yan WY, Huang Q. Identification of novel miRNAs from the microsporidian parasite *Nosema ceranae*. *Infect Genet Evol*. 1 sept 2021;93:104930.
240. Evans JD, Huang Q. Interactions Among Host-Parasite MicroRNAs During *Nosema ceranae* Proliferation in *Apis mellifera*. *Front Microbiol*. 10 avr 2018;9:698.
241. Dong Z, Zheng N, Hu C, Deng B, Fang W, Wu Q, et al. *Nosema bombycis* microRNA-like RNA 8 (Nb-miR8) Increases Fungal Pathogenicity by Modulating BmPEX16 Gene Expression in Its Host, *Bombyx mori*. *Microbiol Spectr*. 31 oct 2021;9(2):e0104821.
242. Rosace D, López J, Blanco S. Emerging roles of novel small non-coding regulatory RNAs in immunity and cancer. *RNA Biol*. 17(8):1196-213.

243. Lambert M, Benmoussa A, Provost P. Small Non-Coding RNAs Derived From Eukaryotic Ribosomal RNA. Non-Coding RNA. 4 févr 2019;5(1):16.
244. Liu B, Cao J, Wang X, Guo C, Liu Y, Wang T. Deciphering the tRNA-derived small RNAs: origin, development, and future. Cell Death Dis. 21 déc 2021;13(1):24.
245. Weng Q, Wang Y, Xie Y, Yu X, Zhang S, Ge J, et al. Extracellular vesicles-associated tRNA-derived fragments (tRFs): biogenesis, biological functions, and their role as potential biomarkers in human diseases. J Mol Med Berl Ger. 2022;100(5):679-95.
246. Garcia-Silva MR, Cabrera-Cabrera F, Cura das Neves RF, Souto-Padrón T, de Souza W, Cayota A. Gene Expression Changes Induced by Trypanosoma cruzi Shed Microvesicles in Mammalian Host Cells: Relevance of tRNA-Derived Halves. BioMed Res Int. 2014;2014:305239.
247. Bayer-Santos E, Lima FM, Ruiz JC, Almeida IC, da Silveira JF. Characterization of the small RNA content of Trypanosoma cruzi extracellular vesicles. Mol Biochem Parasitol. févr 2014;193(2):71-4.
248. Afrin F, Khan I, Hemeg HA. Leishmania-Host Interactions—An Epigenetic Paradigm. Front Immunol. 22 mars 2019;10:492.
249. Lambertz U, Oviedo Ovando ME, Vasconcelos EJ, Unrau PJ, Myler PJ, Reiner NE. Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. BMC Genomics. 5 mars 2015;16(1):151.
250. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression | National Science Review | Oxford Academic [Internet]. [cité 21 nov 2023]. Disponible sur: <https://academic.oup.com/nsr/article/1/2/205/1400440?login=false>
251. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. févr 2019;20(2):89-108.
252. Hirakata S, Ishizu H, Fujita A, Tomoe Y, Siomi MC. Requirements for multivalent Yb body assembly in transposon silencing in Drosophila. EMBO Rep. juill 2019;20(7):e47708.
253. Izumi N, Shoji K, Suzuki Y, Katsuma S, Tomari Y. Zucchini consensus motifs determine the mechanism of pre-piRNA production. Nature. févr 2020;578(7794):311-6.
254. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, et al. piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell. 6 juill 2012;150(1):88-99.

255. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*. 4 mai 2007;316(5825):744-7.
256. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 23 mars 2007;128(6):1089-103.
257. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*. 13 juill 2006;442(7099):203-7.
258. A germline-specific class of small RNAs binds mammalian Piwi proteins | *Nature* [Internet]. [cité 11 nov 2023]. Disponible sur: <https://www.nature.com/articles/nature04917>
259. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? *Trends Genet TIG*. avr 2007;23(4):183-91.
260. Mérel V, Boulesteix M, Fablet M, Vieira C. Transposable elements in *Drosophila*. *Mob DNA*. 3 juill 2020;11(1):23.
261. Maupetit-Mehouas S, Vaury C. Transposon Reactivation in the Germline May Be Useful for Both Transposons and Their Host Genomes. *Cells*. 8 mai 2020;9(5):1172.
262. Wang X, Ramat A, Simonelig M, Liu MF. Emerging roles and functional mechanisms of PIWI-interacting RNAs. *Nat Rev Mol Cell Biol*. févr 2023;24(2):123-41.
263. Allen SE, Nowacki M. Roles of Noncoding RNAs in Ciliate Genome Architecture. *J Mol Biol*. 10 juill 2020;432(15):4186-98.
264. Morlando M, Ballarino M, Fatica A. Long Non-Coding RNAs: New Players in Hematopoiesis and Leukemia. *Front Med* [Internet]. 2015 [cité 3 sept 2023];2. Disponible sur: <https://www.frontiersin.org/articles/10.3389/fmed.2015.00023>
265. Yang Z, Jiang X, Jiang X, Zhao H. X-inactive-specific transcript: A long noncoding RNA with complex roles in human cancers. *Gene*. 30 déc 2018;679:28-35.
266. Loda A, Collombet S, Heard E. Gene regulation in time and space during X-chromosome inactivation. *Nat Rev Mol Cell Biol*. avr 2022;23(4):231-49.
267. Li M, Gong AY, Zhang XT, Wang Y, Mathy NW, Martins GA, et al. Induction of a long non-coding RNA transcript, NR_045064, promotes defense gene transcription and facilitates intestinal epithelial cell responses against *Cryptosporidium* infection. *J Immunol Baltim Md 1950*. 15 déc 2018;201(12):3630-40.

268. Wang Y, Gong AY, Ma S, Chen X, Li Y, Su CJ, et al. Delivery of Parasite RNA Transcripts Into Infected Epithelial Cells During Cryptosporidium Infection and Its Potential Impact on Host Gene Transcription. *J Infect Dis.* 2 févr 2017;215(4):636.
269. Li Y, Baptista RP, Kissinger JC. Noncoding RNAs in Apicomplexan Parasites: An Update. *Trends Parasitol.* oct 2020;36(10):835-49.
270. Wang Y, Gong AY, Ma S, Chen X, Strauss-Soukup JK, Chen XM. Delivery of Parasite Cdg7_Flc_0990 RNA Transcript into Intestinal Epithelial Cells during Cryptosporidium parvum Infection Suppresses Host Cell Gene Transcription through Epigenetic Mechanisms. *Cell Microbiol [Internet].* nov 2017 [cité 21 nov 2023];19(11). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5638686/>
271. Ming Z, Gong AY, Wang Y, Zhang XT, Li M, Mathy NW, et al. Involvement of Cryptosporidium parvum Cdg7_Flc_1000 RNA in the Attenuation of Intestinal Epithelial Cell Migration via Trans-Suppression of Host Cell SMPD3. *J Infect Dis.* 1 janv 2018;217(1):122-33.
272. Shen Z, Yang Q, Luo L, Li T, Ke Z, Li T, et al. Non-coding RNAs identification and regulatory networks in pathogen-host interaction in the microsporidia congenital infection. *BMC Genomics.* 26 juill 2023;24:420.
273. Kim E, Kim YK, Lee SJV. Emerging functions of circular RNA in aging. *Trends Genet.* 1 sept 2021;37(9):819-29.
274. Zhou WY, Cai ZR, Liu J, Wang DS, Ju HQ, Xu RH. Circular RNA: metabolism, functions and interactions with proteins. *Mol Cancer.* 14 déc 2020;19(1):172.
275. Schmidt CA, Matera AG. tRNA introns: Presence, processing, and purpose. *Wiley Interdiscip Rev RNA.* mai 2020;11(3):e1583.
276. Tang X, Ren H, Guo M, Qian J, Yang Y, Gu C. Review on circular RNAs and new insights into their roles in cancer. *Comput Struct Biotechnol J.* 22 janv 2021;19:910-28.
277. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS ONE.* 1 févr 2012;7(2):e30733.
278. Zhao X, Zhong Y, Wang X, Shen J, An W. Advances in Circular RNA and Its Applications. *Int J Med Sci.* 27 mai 2022;19(6):975-85.
279. Li I, Chen YG. Emerging Roles of Circular RNAs in Innate Immunity. *Curr Opin Immunol.* févr 2021;68:107.
280. Li Z, Cheng Y, Wu F, Wu L, Cao H, Wang Q, et al. The emerging landscape of circular RNAs in immunity: breakthroughs and challenges. *Biomark Res.* 10 juill 2020;8(1):25.

281. Chen H, Fan X, Zhang W, Ye Y, Cai Z, Zhang K, et al. Deciphering the CircRNA-Regulated Response of Western Honey Bee (*Apis mellifera*) Workers to Microsporidian Invasion. *Biology* [Internet]. sept 2022 [cité 21 nov 2023];11(9). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9495892/>
282. Guo R, Chen D, Chen H, Xiong C, Zheng Y, Hou C, et al. Genome-Wide Identification of Circular RNAs in Fungal Parasite *Nosema ceranae*. *Curr Microbiol*. déc 2018;75(12):1655-60.
283. Minkler SJ, Loghry-Jansen HJ, Sondjaja NA, Kimber MJ. Expression and Secretion of Circular RNAs in the Parasitic Nematode, *Ascaris suum*. *Front Genet*. 2022;13:884052.
284. Billmyre RB, Calo S, Feretzaki M, Wang X, Heitman J. RNAi Function, Diversity, and Loss in the Fungal Kingdom. *Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol*. déc 2013;21(0):10.1007/s10577-013-9388-2.
285. Shabalina SA, Koonin EV. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol*. oct 2008;23(10):578-87.
286. Cogoni C, Macino G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature*. 13 mai 1999;399(6732):166-9.
287. Wang X, Hsueh YP, Li W, Floyd A, Skalsky R, Heitman J. Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. *Genes Dev*. 15 nov 2010;24(22):2566-82.
288. Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM. Two classes of small antisense RNAs in fungal RNA silencing triggered by non-integrative transgenes. *EMBO J*. 1 août 2003;22(15):3983-91.
289. Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, et al. An endosRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell*. 2 nov 2017;21(5):694-703.e7.
290. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, et al. RNAi is a critical determinant of centromere evolution in closely related fungi. *Proc Natl Acad Sci U S A*. 20 mars 2018;115(12):3108-13.
291. Drinnenberg IA, Fink GR, Bartel DP. Compatibility with Killer explains the Rise of RNAi-deficient fungi. *Science*. 16 sept 2011;333(6049):1592.
292. Magliani W, Conti S, Gerloni M, Bertolotti D, Polonelli L. Yeast killer systems. *Clin Microbiol Rev*. juill 1997;10(3):369-400.
293. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, et al. RNAi in budding yeast. *Science*. 23 oct 2009;326(5952):544-50.

294. D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, et al. Genome Variation in *Cryptococcus gattii*, an Emerging Pathogen of Immunocompetent Hosts. *mBio*. 8 févr 2011;2(1):e00342-10.
295. Byrnes EJ, Marr KA. The Outbreak of *Cryptococcus gattii* in Western North America: Epidemiology and Clinical Issues. *Curr Infect Dis Rep*. juin 2011;13(3):256-61.
296. Fijalkowska IJ, Dunn RL, Schaaper RM. Genetic requirements and mutational specificity of the *Escherichia coli* SOS mutator activity. *J Bacteriol*. déc 1997;179(23):7435-45.
297. Oliver A, Cantón R, Campo P, Baquero F, Blázquez J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science*. 19 mai 2000;288(5469):1251-4.
298. Lye LF, Owens K, Shi H, Murta SMF, Vieira AC, Turco SJ, et al. Retention and Loss of RNA Interference Pathways in Trypanosomatid Protozoans. *PLOS Pathog*. 28 oct 2010;6(10):e1001161.
299. Batista TM, Marques JT. RNAi pathways in parasitic protists and worms. *J Proteomics*. 24 août 2011;74(9):1504-14.
300. Ngô H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*. 8 déc 1998;95(25):14687-92.
301. Djikeng A, Shi H, Tschudi C, Ullu E. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA N Y N*. nov 2001;7(11):1522-30.
302. Durand-Dubief M, Bastin P. TbAGO1, an Argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC Biol*. 12 déc 2003;1(1):2.
303. Garcia Silva MR, Tosar JP, Frugier M, Pantano S, Bonilla B, Esteban L, et al. Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene*. 15 oct 2010;466(1-2):26-35.
304. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infect Genet Evol*. 1 nov 2018;65:329-32.
305. MRBAYES: Bayesian inference of phylogenetic trees - PubMed [Internet]. [cité 12 nov 2023]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/11524383/>
306. Vembar SS, Scherf A, Siegel TN. Noncoding RNAs as emerging regulators of *Plasmodium falciparum* virulence gene expression. *Curr Opin Microbiol*. août 2014;20:153-61.

307. Epp C, Li F, Howitt CA, Chookajorn T, Deitsch KW. Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA*. janv 2009;15(1):116-27.
308. MILITELLO KT, PATEL V, CHESSLER AD, FISHER JK, KASPER JM, GUNASEKERA A, et al. RNA polymerase II synthesizes antisense RNA in *Plasmodium falciparum*. *RNA*. avr 2005;11(4):365-70.
309. Amit-Avraham I, Pozner G, Eshar S, Fastman Y, Kolevzon N, Yavin E, et al. Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*. 3 mars 2015;112(9):E982-91.
310. Jing Q, Cao L, Zhang L, Cheng X, Gilbert N, Dai X, et al. *Plasmodium falciparum* var Gene Is Activated by Its Antisense Long Noncoding RNA. *Front Microbiol*. 18 déc 2018;9:3117.
311. Guizetti J, Barcons-Simon A, Scherf A. Trans-acting GC-rich non-coding RNA at var expression site modulates gene counting in malaria parasite. *Nucleic Acids Res*. 16 nov 2016;44(20):9710-8.
312. Barcons-Simon A, Cordon-Obras C, Guizetti J, Bryant JM, Scherf A. CRISPR Interference of a Clonally Variant GC-Rich Noncoding RNA Family Leads to General Repression of var Genes in *Plasmodium falciparum*. *mBio*. 21 janv 2020;11(1):e03054-19.
313. Patil V, Lescault PJ, Lirussi D, Thompson AB, Matrajt M. Disruption of the Expression of a Non-Coding RNA Significantly Impairs Cellular Differentiation in *Toxoplasma gondii*. *Int J Mol Sci*. 28 déc 2012;14(1):611-24.
314. Lescault PJ, Thompson AB, Patil V, Lirussi D, Burton A, Margarit J, et al. Genomic Data Reveal *Toxoplasma gondii* Differentiation Mutants Are Also Impaired with Respect to Switching into a Novel Extracellular Tachyzoite State. *PLoS ONE*. 30 déc 2010;5(12):e14463.
315. Behnke MS, Radke JB, Smith AT, Sullivan WJ, White MW. The transcription of bradyzoite genes in *Toxoplasma gondii* is controlled by autonomous promoter elements. *Mol Microbiol*. juin 2008;68(6):1502-18.
316. Mueller AK, Hammerschmidt-Kamper C, Kaiser A. RNAi in *Plasmodium*. *Curr Pharm Des*. 2014;20(2):278-83.
317. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 16 avr 2004;304(5669):441-5.

318. Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, et al. A Complex Small RNA Repertoire Is Generated by a Plant/Fungal-Like Machinery and Effected by a Metazoan-Like Argonaute in the Single-Cell Human Parasite *Toxoplasma gondii*. *PLoS Pathog.* 27 mai 2010;6(5):e1000920.
319. Guegan F, Rajan KS, Bento F, Pinto-Neves D, Sequeira M, Gumińska N, et al. A long noncoding RNA promotes parasite differentiation in African trypanosomes. *Sci Adv.* 17 juin 2022;8(24):eabn2706.
320. Benesova S, Kubista M, Valihrach L. Small RNA-Sequencing: Approaches and Considerations for miRNA Analysis. *Diagn Basel Switz.* 27 mai 2021;11(6):964.
321. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* janv 2012;40(1):37-52.
322. Mackowiak SD. Identification of novel and known miRNAs in deep-sequencing data with miRDeep2. *Curr Protoc Bioinforma.* déc 2011;Chapter 12:12.10.1-12.10.15.
323. Huang Q. Effective Silencing of Dicer Decreases Spore Load of the Honey Bee Parasite *Nosema ceranae*. *Fungal Genomics Biol.* 8 août 2016;06.
324. Huang Q, Li W, Chen Y, Retschnig-Tanner G, Yanez O, Neumann P, et al. Dicer regulates *Nosema ceranae* proliferation in honeybees. *Insect Mol Biol.* 2019;28(1):74-85.
325. *Nosema algerae* n. sp. (Cnidospora, Microsporida) a pathogen in a laboratory colony of *Anopheles stephensi* Liston (Diptera, Culicidae) - PubMed [Internet]. [cité 21 nov 2023]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/4915459/>
326. HFF-1 - SCRC-1041 | ATCC [Internet]. [cité 3 sept 2023]. Disponible sur: <https://www.atcc.org/products/scrc-1041>
327. Irfan Maqsood M, Matin MM, Bahrami AR, Ghasroldasht MM. Immortality of cell lines: challenges and advantages of establishment. *Cell Biol Int.* 2013;37(10):1038-45.
328. Wu L, Zhang QX, Li TT, Chen SX, Cao JP. [In vitro culture of *Toxoplasma gondii* tachyzoites in HFF and HeLa cells]. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi.* juin 2009;27(3):229-31.
329. Jung H, Han S, Lee Y. Transcriptome analysis of alternative splicing in the pathogen life cycle in human foreskin fibroblasts infected with *Trypanosoma cruzi*. *Sci Rep.* 15 oct 2020;10(1):17481.
330. Aguiar A, Galinato M, Bradley Silva M, Toth B, McVoy MA, Hertel L. Human Cytomegalovirus Replication and Infection-Induced Syncytia Formation in Labial, Foreskin, and Fetal Lung Fibroblasts. *Viruses.* 24 nov 2021;13(12):2355.

331. Molestina R, Becnel JJ, Weiss LM. Culture and Propagation of Microsporidia. In: Microsporidia [Internet]. John Wiley & Sons, Ltd; 2014 [cité 3 sept 2023]. p. 457-67. Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118395264.ch18>
332. LALLO MA, VIDOTO DA COSTA LF, ALVARES-SARAIVA AM, ROCHA PRD, SPADACCI-MORENA DD, KONNO FT de C, et al. Culture and propagation of microsporidia of veterinary interest. *J Vet Med Sci.* févr 2016;78(2):171-6.
333. Visvesvara GS. In Vitro Cultivation of Microsporidia of Clinical Importance. *Clin Microbiol Rev.* juill 2002;15(3):401-13.
334. The human Fibroblasts - The Human Protein Atlas [Internet]. [cité 3 sept 2023]. Disponible sur: <https://www.proteinatlas.org/humanproteome/tissue+cell+type/Fibroblasts>
335. Buechler MB, Turley SJ. A short field guide to fibroblast function in immunity. *Semin Immunol.* févr 2018;35:48-58.
336. Neidleman JA, Chen JC, Kohgadai N, Müller JA, Laustsen A, Thavachelvam K, et al. Mucosal stromal fibroblasts markedly enhance HIV infection of CD4+ T cells. *PLoS Pathog.* févr 2017;13(2):e1006163.
337. Bukholm G, Berdal BP, Haug C, Degré M. Mouse fibroblast interferon modifies Salmonella typhimurium infection in infant mice. *Infect Immun.* juill 1984;45(1):62-6.
338. Caligiuri G, Tuveson DA. Activated fibroblasts in cancer: Perspectives and challenges. *Cancer Cell.* 13 mars 2023;41(3):434-49.
339. Ghazimoradi MH, Hasegawa K, Zolghadr E, Montazeri S, Farivar S. Reprogramming of fibroblast cells to totipotent state by DNA demethylation. *Sci Rep.* 20 janv 2023;13(1):1154.
340. Lendahl U, Muhl L, Betsholtz C. Identification, discrimination and heterogeneity of fibroblasts. *Nat Commun.* 14 juin 2022;13(1):3409.
341. Castro-Viñuelas R, Sanjurjo-Rodríguez C, Piñeiro-Ramil M, Hermida-Gómez T, Rodríguez-Fernández S, Oreiro N, et al. Generation and characterization of human induced pluripotent stem cells (iPSCs) from hand osteoarthritis patient-derived fibroblasts. *Sci Rep.* 6 mars 2020;10(1):4272.
342. Shamis Y, Hewitt KJ, Bear SE, Alt-Holland A, Qari H, Margvelashvili M, et al. iPSC-derived fibroblasts demonstrate augmented production and assembly of extracellular matrix proteins. *In Vitro Cell Dev Biol Anim.* févr 2012;48(2):112-22.
343. Pallen MJ. Microbial bioinformatics 2020. *Microb Biotechnol.* 29 juill 2016;9(5):681-6.

344. McDonnell E, Strasser K, Tsang A. Manual Gene Curation and Functional Annotation. In: de Vries RP, Tsang A, Grigoriev IV, éditeurs. *Fungal Genomics: Methods and Protocols* [Internet]. New York, NY: Springer; 2018 [cité 5 nov 2023]. p. 185-208. (Methods in Molecular Biology). Disponible sur: https://doi.org/10.1007/978-1-4939-7804-5_16
345. Baumgartner WA Jr, Cohen KB, Fox LM, Acquah-Mensah G, Hunter L. Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics*. 1 juill 2007;23(13):i41-8.
346. Tournayre J, Polonais V, Wawrzyniak I, Akossi RF, Parisot N, Lerat E, et al. MicroAnnot: A Dedicated Workflow for Accurate Microsporidian Genome Annotation [Internet]. *Preprints*; 2023 [cité 30 oct 2023]. Disponible sur: <https://www.preprints.org/manuscript/202310.1822/v1>
347. Marx V. Method of the year: long-read sequencing. *Nat Methods*. janv 2023;20(1):6-11.
348. Paul S, Ruiz-Manriquez LM, Serrano-Cano FI, Estrada-Meza C, Solorio-Diaz KA, Srivastava A. Human microRNAs in host–parasite interaction: a review. *3 Biotech*. 5 nov 2020;10(12):510.
349. Sak B, Kváč M. Chronic Infections in Mammals Due to Microsporidia. *Exp Suppl* 2012. 2022;114:319-71.
350. Acuña SM, Floeter-Winter LM, Muxel SM. MicroRNAs: Biological Regulators in Pathogen–Host Interactions. *Cells*. janv 2020;9(1):113.
351. Williams BAP, Williams TA, Trew J. Comparative Genomics of Microsporidia. *Exp Suppl* 2012. 2022;114:43-69.
352. Salzberg SL. Next-generation genome annotation: we still struggle to get it right. *Genome Biol*. 16 mai 2019;20(1):92.
353. Brenner SE. Errors in genome annotation. *Trends Genet*. avr 1999;15(4):132-3.
354. Gene annotation errors are common in the mammalian mitochondrial genomes database | *BMC Genomics* | Full Text [Internet]. [cité 5 nov 2023]. Disponible sur: <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-5447-1>
355. Clarke ZA, Andrews TS, Atif J, Pouyabahr D, Innes BT, MacParland SA, et al. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat Protoc*. juin 2021;16(6):2749-64.
356. Petty NK. Genome annotation: man versus machine. *Nat Rev Microbiol*. nov 2010;8(11):762-762.
357. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinforma Oxf Engl*. 15 mars 2007;23(6):673-9.

358. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*. 1 nov 2004;20(16):2878-9.
359. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol Clifton NJ*. 2019;1962:1-14.
360. Angst P, Pombert JF, Ebert D, Fields PD. Near chromosome-level genome assembly of the microsporidium *Hamiltosporidium tvaerminnensis*. *G3 GenesGenomesGenetics*. 1 oct 2023;13(10):jkad185.
361. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*. mars 2014;1840(3):1063-71.
362. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol*. juin 2023;24(6):430-47.
363. Ali A, Han K, Liang P. Role of Transposable Elements in Gene Regulation in the Human Genome. *Life Basel Switz*. 4 févr 2021;11(2):118.
364. Vilborg A, Steitz JA. Readthrough transcription: How are DoGs made and what do they do? *RNA Biol*. 9 févr 2016;14(5):632-6.
365. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*. 2 nov 2007;318(5851):761-4.
366. He X, Fu Z, Li M, Liu H, Cai S, Man N, et al. *Nosema bombycis* (Microsporidia) suppresses apoptosis in BmN cells (*Bombyx mori*). *Acta Biochim Biophys Sin*. sept 2015;47(9):696-702.
367. Martín-Hernández R, Higes M, Sagastume S, Juarranz Á, Dias-Almeida J, Budge GE, et al. Microsporidia infection impacts the host cell's cycle and reduces host cell apoptosis. *PLoS ONE*. 2 févr 2017;12(2):e0170183.
368. del Aguila C, Izquierdo F, Granja AG, Hurtado C, Fenoy S, Fresno M, et al. Encephalitozoon microsporidia modulates p53-mediated apoptosis in infected cells. *Int J Parasitol*. juill 2006;36(8):869-76.
369. Hledík M, Barton N, Tkačik G. Accumulation and maintenance of information in evolution. *Proc Natl Acad Sci U S A*. 6 sept 2022;119(36):e2123152119.
370. Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, et al. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature*. févr 2022;602(7895):101-5.
371. Akossi RF, Delbac F, El Alaoui H, Wawrzyniak I, Peyretaillade E. The intracellular parasite *Anncaliia algerae* induces a massive miRNA down-regulation in human cells. *Non-Coding RNA Res*. 1 sept 2023;8(3):363-75.

372. Luo J, He Q, Xu JZ, Xu C, Han YZ, Gao HL, et al. Microsporidia infection upregulates host energy metabolism but maintains ATP homeostasis. *J Invertebr Pathol*. 1 nov 2021;186:107596.
373. Dean P, Hirt RP, Embley TM. Microsporidia: Why Make Nucleotides if You Can Steal Them? *PLoS Pathog*. 17 nov 2016;12(11):e1005870.
374. Major P, Sendra KM, Dean P, Williams TA, Watson AK, Thwaites DT, et al. A new family of cell surface located purine transporters in Microsporidia and related fungal endoparasites. *eLife*. 8:e47037.
375. Sahin U, de Thé H, Lallemand-Breitenbach V. Sumoylation in Physiology, Pathology and Therapy. *Cells*. 26 févr 2022;11(5):814.
376. Ptak C, Wozniak RW. SUMO and Nucleocytoplasmic Transport. *Adv Exp Med Biol*. 2017;963:111-26.
377. Panek J, El Alaoui H, Mone A, Urbach S, Demettre E, Texier C, et al. Hijacking of host cellular functions by an intracellular parasite, the microsporidian *Anncaliia algerae*. *PloS One*. 2014;9(6):e100791.
378. Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. *Brief Bioinform*. 3 juin 2019;20(5):1853-64.
379. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 14 avr 2006;312(5771):212-7.
380. Unsihuay D, Mesa Sanchez D, Laskin J. Quantitative Mass Spectrometry Imaging of Biological Systems. *Annu Rev Phys Chem*. 20 avr 2021;72:307-29.
381. Uhl M, Houwaart T, Corrado G, Wright PR, Backofen R. Computational analysis of CLIP-seq data. *Methods San Diego Calif*. 15 avr 2017;118-119:60-72.
382. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc*. juin 2022;17(6):1518-52.
383. Ma S, Zhang Y. Profiling chromatin regulatory landscape: insights into the development of ChIP-seq and ATAC-seq. *Mol Biomed*. 2020;1(1):9.
384. Andreev DE, O'Connor PBF, Loughran G, Dmitriev SE, Baranov PV, Shatsky IN. Insights into the mechanisms of eukaryotic translation gained with ribosome profiling. *Nucleic Acids Res*. 25 janv 2017;45(2):513-26.
385. Jain M, Abu-Shumays R, Olsen HE, Akeson M. Advances in nanopore direct RNA sequencing. *Nat Methods*. oct 2022;19(10):1160-4.

386. Herren JK, Mbaisi L, Mararo E, Makhulu EE, Mobegi VA, Butungi H, et al. A microsporidian impairs *Plasmodium falciparum* transmission in *Anopheles arabiensis* mosquitoes. *Nat Commun.* 4 mai 2020;11(1):2187.
387. CirGO: an alternative circular way of visualising gene ontology terms | BMC Bioinformatics | Full Text [Internet]. [cité 2 nov 2023]. Disponible sur: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2671-2>



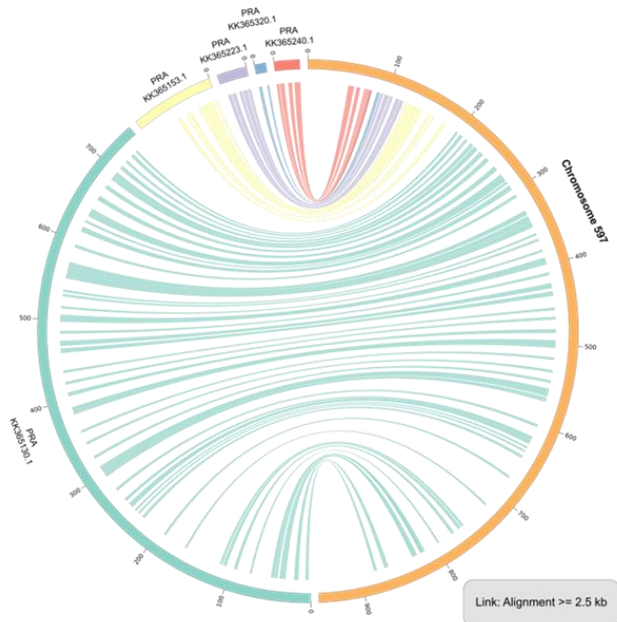
SECTION E : ANNEXES



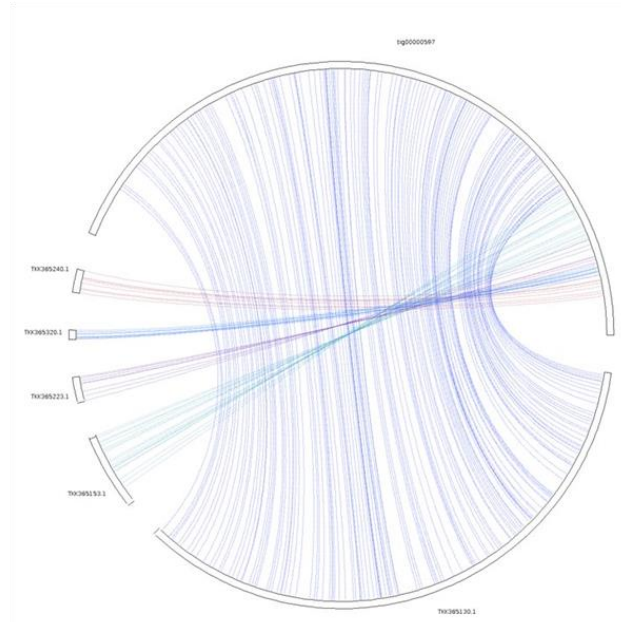
SECTION E ANNEXES

ANNEXES SECTION B PARTIE II

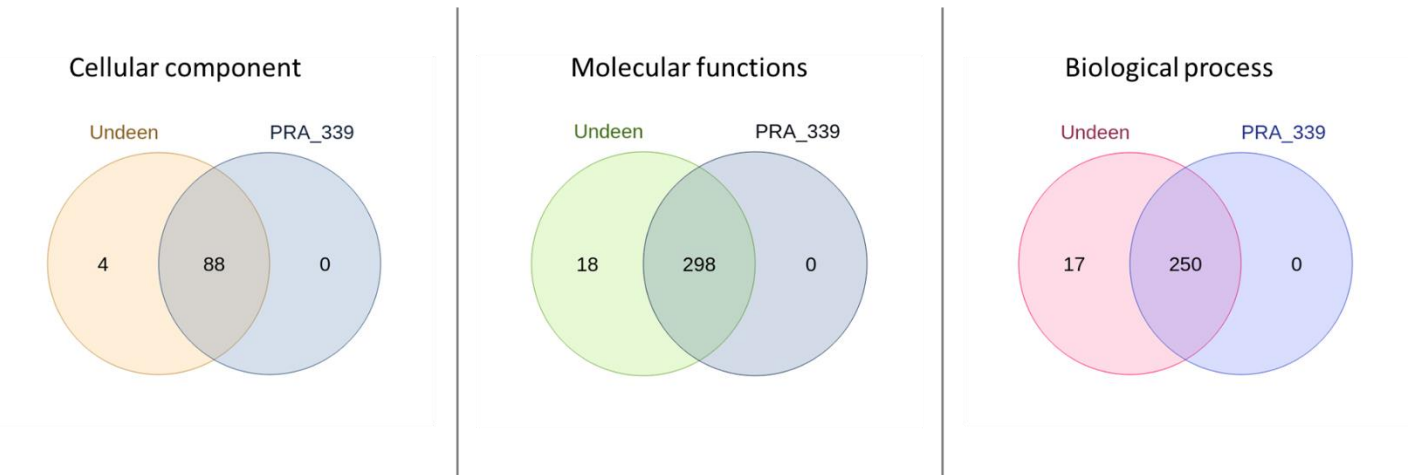
A.



B.



C.

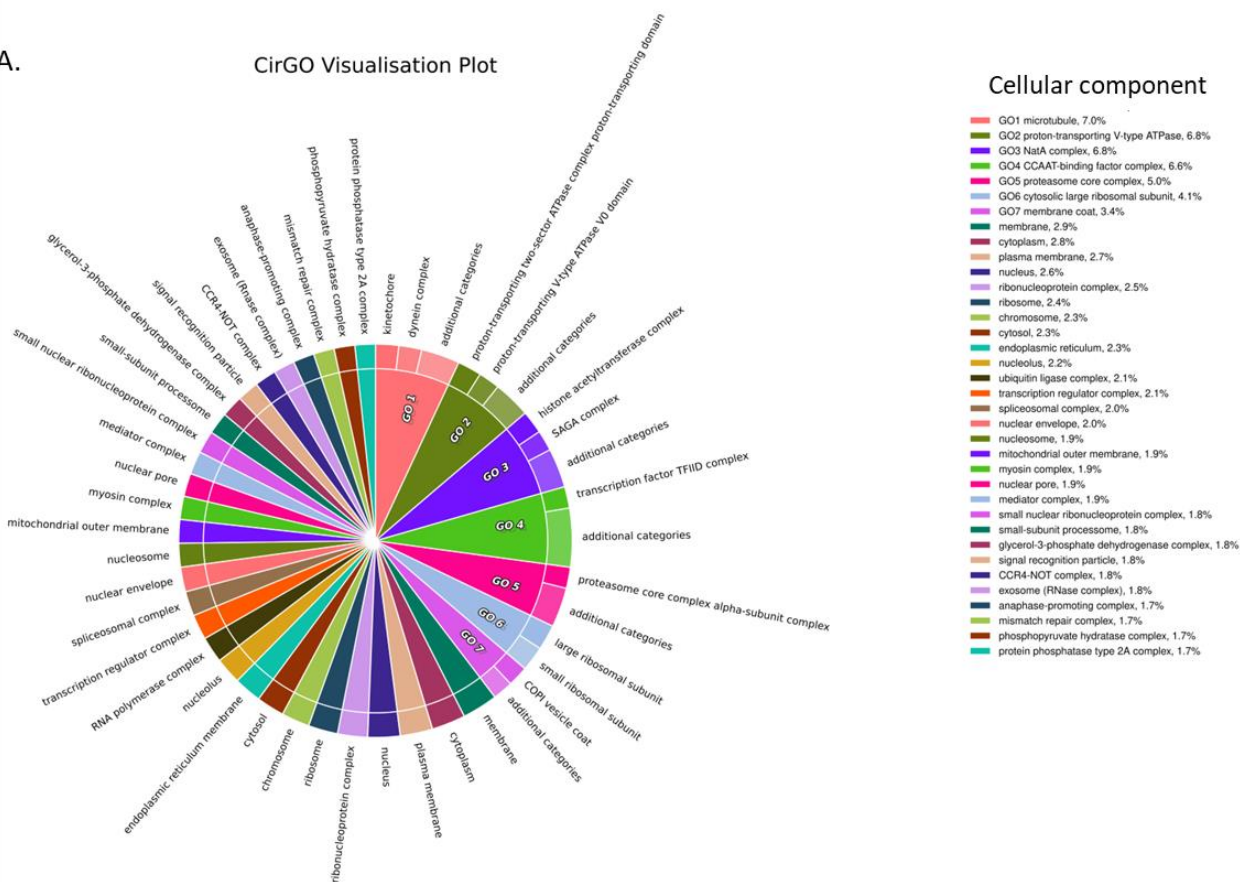


Supplementary figure 1: Contig correspondence, synteny and similar completion between the PRA339 the Undeen genomes

Comparison between *A. algerae* Undeen Haploid genome assembly and *A. algerae* PRA 339 reference assembly: Synteny graph results after All versus All blastn analysis (results for two of Undeen’s contigs are shown as an example) A. Nucleotide sequence synteny (alignments $\geq 2.5\text{Kb}$ are shown). B. CDS synteny. C. Graphical representation of the distribution of the different GO terms assigned to genes, in the three categories, cellular component, molecular function, biological process.

A.

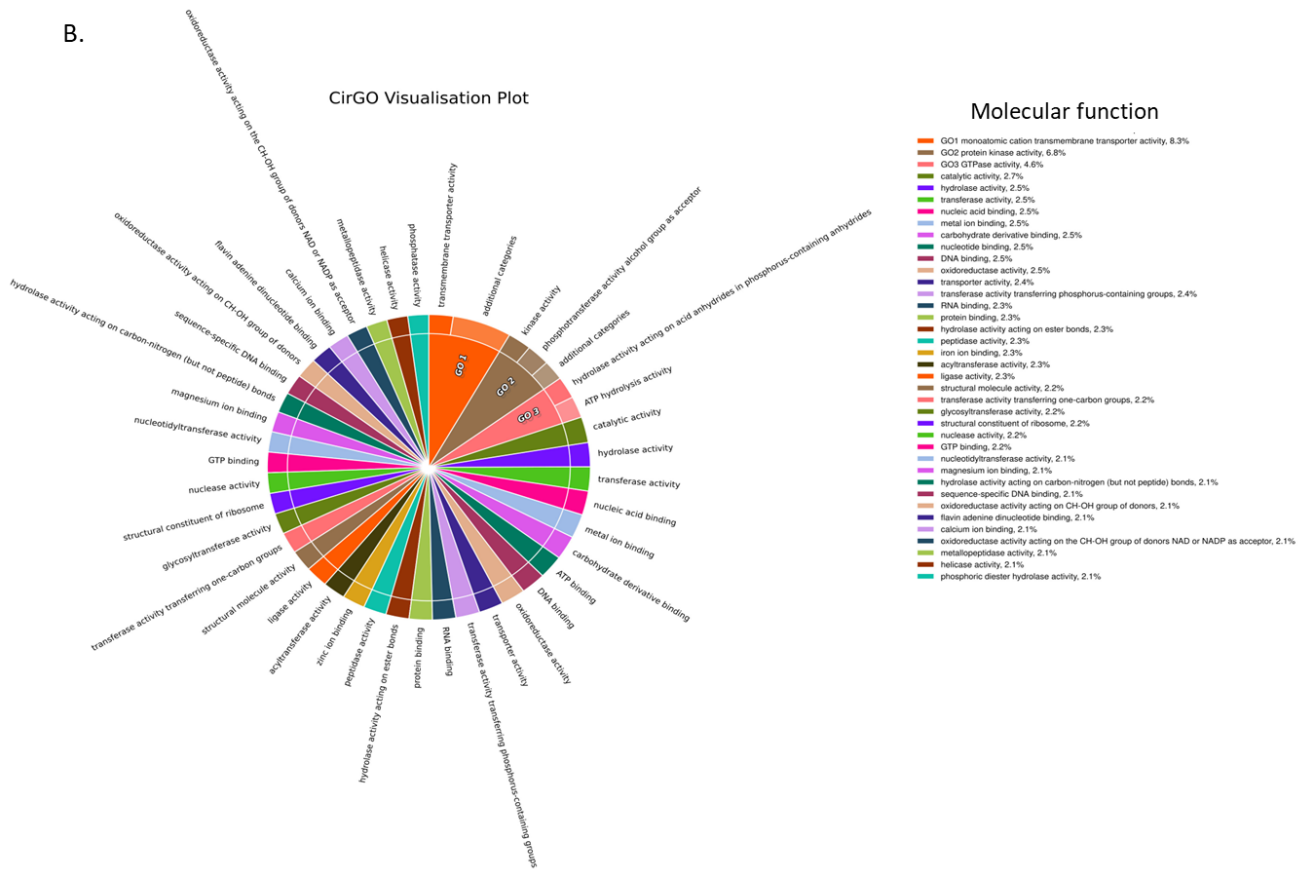
CirGO Visualisation Plot



Supplementary figure 2: GO term annotation comparison between Undeen and PRA 339 strains

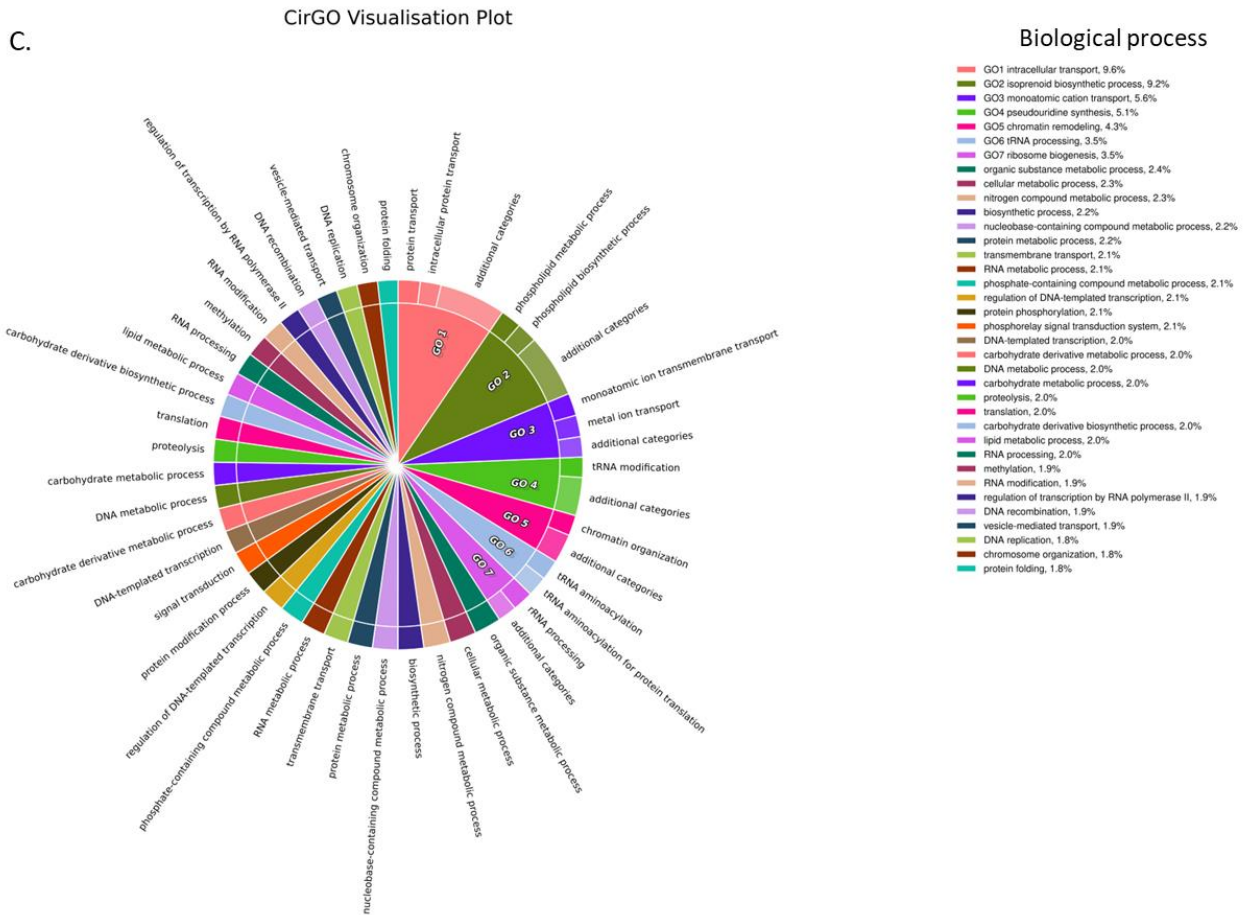
A. Pie chart representing the common 30 first categories Go term between PRA 339 and Undeen strains at the cellular component level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. B. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strains at the molecular functions level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. C. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strain at the biological process level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups.

B.



Supplementary figure 2: GO term annotation comparison between Undeen and PRA 339 strains

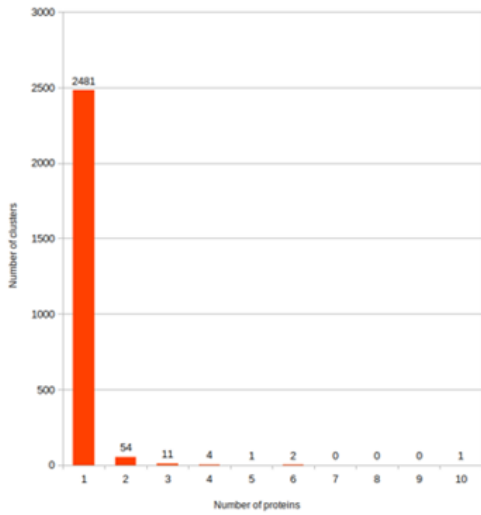
A. Pie chart representing the common 30 first categories Go term between PRA 339 and Undeen strains at the cellular component level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. B. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strains at the molecular functions level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. C. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strain at the biological process level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups.



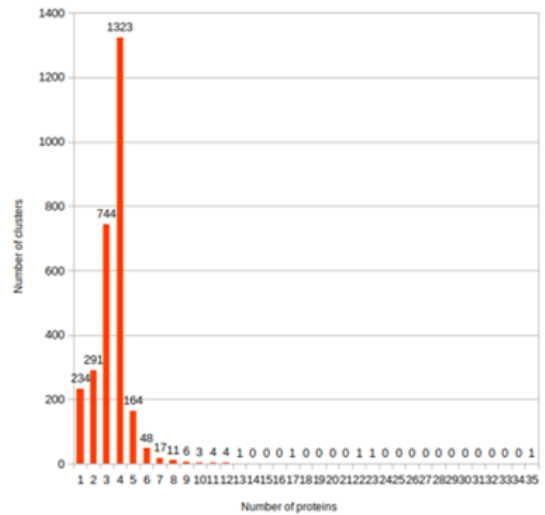
Supplementary figure 2: GO term annotation comparison between Undeen and PRA 339 strains

A. Pie chart representing the common 30 first categories Go term between PRA 339 and Undeen strains at the cellular component level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. B. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strains at the molecular functions level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups. C. Pie chart representing the common 40 first categories Go term between PRA339 and Undeen strain at the biological process level, generated using CirGO software (387). The inner circle contains the main GO group, while the outer represents the subordinate GO groups.

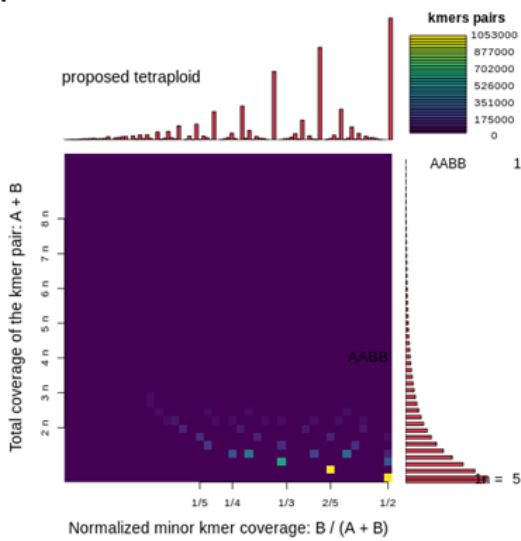
A.



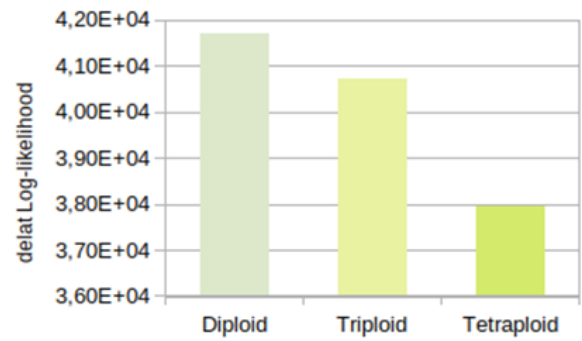
B.



C.



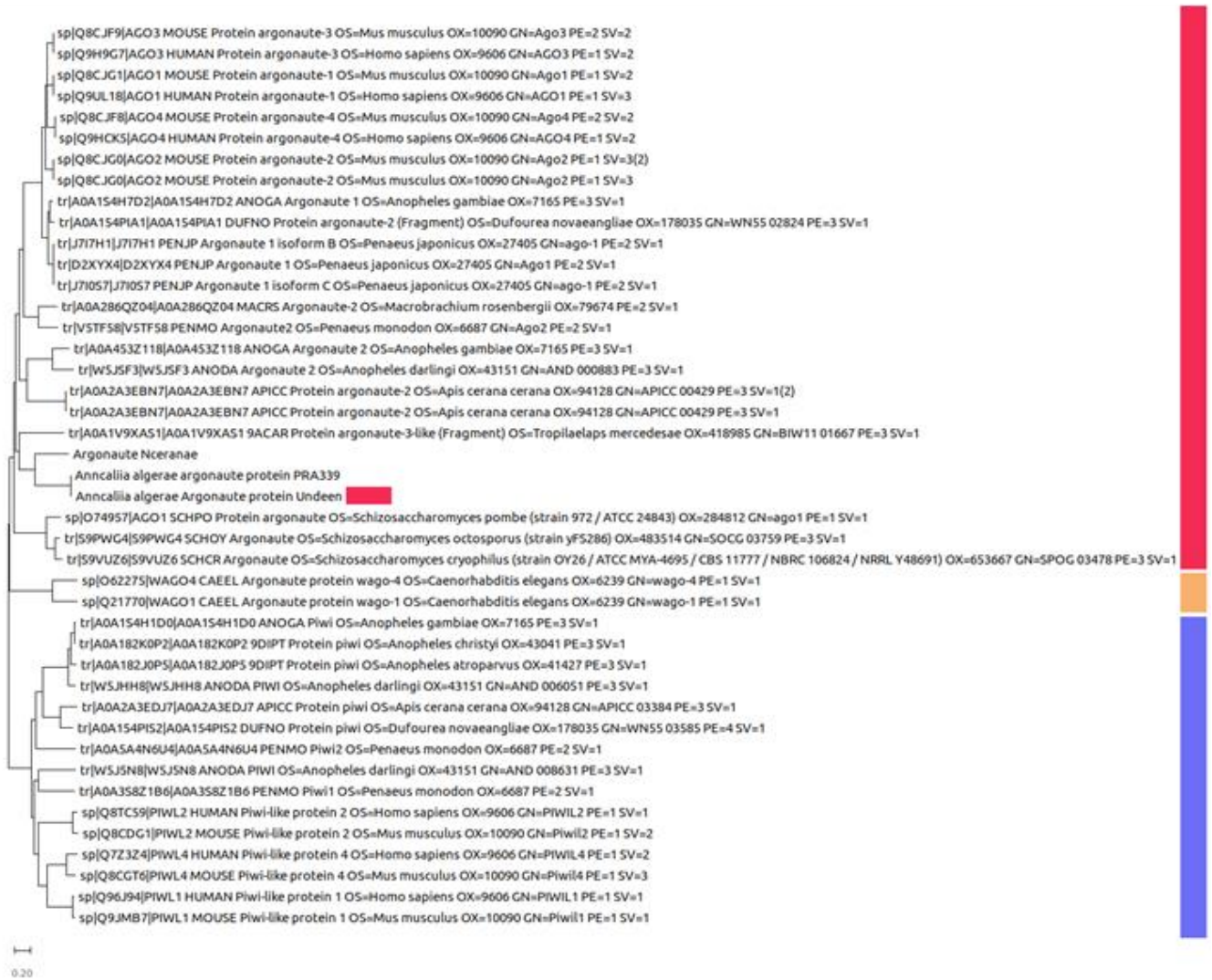
D.



Supplementary figure 3: Elements further supporting *A.algerae* tetraploidy

Graph of the number and size of clusters formed after the clustering of CDS sequences using cd-hit with 90% homology on 100% of the length of the smaller sequence. A. Results for *A.algerae* haploid genome assembly. B. Results for the full genome assembly. C. Smudgeplot and D. nQuire analysis results, the ploidy of the genome is the ploidy with the lowest delta log-likelihood value.

A.



Supplementary figure 4: *A. algerae* Argonaute, Dicer, and RdRP of *A. algerae* phylogenetic trees

A. Phylogenetic tree of *A. algerae*'s Argonaute protein with AGO and Piwi clade proteins from various species. multiple alignments of Amino acid sequences were carried out with Muscle algorithm using MEGA). AGO clade proteins, WAGO proteins and Piwi clade proteins were represented in red, orange, and purple respectively. *A. algerae* Argonaute protein clusters with the AGO clade of Argonaute proteins. **B.** Phylogenetic trees of *A. algerae* Argonaute, Dicer and RNA dependant RNA polymerase based on PSI-BLASTP against NCBI nr database or clustered nr database for the RNA dependant RNA polymerase. PSI-blast e-value threshold used was 0.005.

Supplementary table 1: List of Undeen specific GO term annotations

Cellular component		Molecular functions		Biological Process	
GO terms	component	GO terms	function	GO terms	process
GO:0005847	mRNA cleavage and polyadenylation specificity factor complex	GO:0001054	RNA polymerase I activity	GO:0006032	chitin catabolic process
GO:0016035	zeta DNA polymerase complex	GO:0001056	RNA polymerase III activity	GO:0006261	DNA-templated DNA replication
GO:0005685	U1 snRNP	GO:0003950	NAD+ ADP-ribosyltransferase activity	GO:0006376	mRNA splice site recognition
GO:0033178	proton-transporting two-sector ATPase complex, catalytic domain	GO:0003896	DNA primase activity	GO:0000290	deadenylation-dependent decapping of nuclear-transcribed mRNA
		GO:0004105	choline-phosphate cytidyltransferase activity	GO:0006657	CDP-choline pathway
		GO:0004568	chitinase activity	GO:0006895	Golgi to endosome transport
		GO:0004748	ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor	GO:0042147	retrograde transport, endosome to Golgi
		GO:0004809	tRNA (guanine-N2-)-methyltransferase activity	GO:0006233	dTDP biosynthetic process
		GO:0008047	enzyme activator activity	GO:0044571	[2Fe-2S] cluster assembly
		GO:0008374	O-acyltransferase activity	GO:0016998	cell wall macromolecule catabolic process
		GO:0004844	uracil DNA N-glycosylase activity	GO:0018342	protein prenylation
		GO:0004865	protein serine/threonine phosphatase inhibitor activity	GO:0019985	translesion synthesis
		GO:0016799	hydrolase activity, hydrolyzing N-glycosyl compounds	GO:0032515	negative regulation of phosphoprotein phosphatase activity
		GO:0030623	U5 snRNA binding	GO:0043085	positive regulation of catalytic activity
		GO:0051087	protein-folding chaperone binding	GO:0071786	positive regulation of catalytic activity
		GO:0004017	adenylate kinase activity	GO:1902751	positive regulation of cell cycle G2/M phase transition
		GO:0004798	thymidylate kinase activity	GO:1990426	mitotic recombination-dependent replication fork processing
		GO:0019829	ATPase-coupled monoatomic cation transmembrane transporter activity		

Supplementary table 2: Summary of the Transposable element diversity identified in *A.algerae* Undeen's culture metagenome assembly

Full genome assembly

Family	Copies	Total Bp	% of Genome
DNA/IS630	14.00	5,531.00	0.01
DNA/ISXO2	19.00	6,301.00	0.01
DNA/Mariner-Tc1	1,054.00	933,846.00	2.12
DNA/Merlin	9,084.00	6,175,063.00	14.05
DNA/Piggy-Bac	27.00	39,689.00	0.09
DNA/Unknown	180.00	67,557.00	0.15
DNA transposons	10,378.00	7,227,987.00	16.44
LINE/Unknown	22.00	30,985.00	0.07
LINE transposons	22.00	30,985.00	0.07
LTR/Unknown	26.00	19,083.00	0.04
LTR/Ty3	8,266.00	3,122,226.00	7.10
LTR transposons	8,292.00	3,141,309.00	7.15
Total	18,692.00	10,400,281.00	23.66
Low complexity regions	5,839.00	313,816.00	0.71
Satellite repeats	-	-	-
Simple repeats	13,762.00	903,698.00	2.06
Unknown	279.00	143,311.00	0.33

Supplementary table 3: Metrics of raw Trinity assembly of transcripts

	Aa3hms1	Aa3hms2	Aa3hms3	Aa12hms1	Aa12hms2	Aa12hms3	Aa24hms1	Aa24hms2	Aa24hms3	Aa48hms1	Aa48hms2	Aa48hms3	Aa72hms1	Aa72hms2	Aa72hms3
Sample	Aa3hms1	Aa3hms2	Aa3hms3	Aa12hms1	Aa12hms2	Aa12hms3	Aa24hms1	Aa24hms2	Aa24hms3	Aa48hms1	Aa48hms2	Aa48hms3	Aa72hms1	Aa72hms2	Aa72hms3
Number of sequences	11020	9642	6295	150138	172606	210031	120489	111115	134321	108946	97345	85312	39415	43598	42868
Min length	201	180	201	201	103	201	104	105	201	194	201	201	111	112	113
Average length	355.3	346.4	348	430.8	426.8	430	408.9	414.8	405.4	396.7	398.2	396.9	383.3	381.6	388.4
Max length	3153	3429	3549	8917	8894	9089	11589	10911	9586	9196	11238	9396	6469	5369	5580
Q1	230	229	228	242	242	243	239	240	239	237	237	237	233	233	233
Q2	279	275	276	315	315	317	305	308	303	299	301	300	286	287	291
Q3	386	379	377	480	474	483	453	459	449	437	442	440	413	414	418
N50	354	342	344	470	462	471	434	443	429	414	418	417	391	390	398
GC(%)	42.12	42.67	43.36	44.04	43	43.61	42.05	40.71	41.92	42.71	44.2	46.68	39.71	40.27	40.45

Supplementary table 4: Metrics of timepoint specific transcriptomes

	Number of sequences	Total length	Min length	Average length	Max length	Q1	Q2	Q3	GC(%)
Consensus_3h_transcriptome	705	319320	200	452.9	2193	250	336	487	36.21
Consensus_12h_transcriptome	3953	2182280	105	552.1	4330	282	415	648	30.25
Consensus_24h_transcriptome	4567	3054963	104	668.9	4764	310	504	827	29.46
Consensus_48h_transcriptome	5103	3208030	189	628.7	4724	295	479	775	29.7
Consensus_72h_transcriptome	6096	3916035	135	642.4	4835	297	476.5	794	28.78
Consensus_transcriptome	8433	5351464	104	634.6	4835	289	461	774	28.73

Supplementary table 5 : Table of a few miRNA, piRNA and siRNA associated proteins ,their Interpro conserved domains, and their presence in *A.algerae*'s haploid genome assembly
 Domains where mostly taken from *C.elegans* homologue when possible, in parenthesis are the names of said homologues in the species used

Common name (and name in species used)	Interpro domains/families associated	CDS found in <i>A.algerae</i> 's genome and identified interpro domains/families
Drosha (drsh-1 <i>C.elegans</i>)	IPR000999 Ribonuclease III domain IPR014720 Double-stranded RNA-binding domain IPR044442 RNase III, double-stranded RNA binding domain, animal IPR011907 Ribonuclease III	
DGCR8 (Pash-1 <i>C.elegans</i>)	IPR001202 VWW domain IPR014720 Double-stranded RNA-binding domain IPR040375 Microprocessor complex subunit DGCR8	
Loquacious (rde-4 <i>C.elegans</i>)	IPR014720 Double-stranded RNA-binding domain	
Dicer (dcr-1 <i>C.elegans</i>)	IPR005034 Dicer dimerisation domain IPR014001 Helicase superfamily 1/2, ATP-binding domain IPR003100 PAZ domain IPR000999 Ribonuclease III domain IPR001650 Helicase, C-terminal IPR044441 Dicer, double-stranded RNA-binding domain IPR006935 Helicase/UvrB, N-terminal IPR048512 Dicer, platform domain IPR014720 Double-stranded RNA-binding domain IPR048513 Dicer, partner-binding domain IPR011907 Ribonuclease III	tig00000597_27 IPR003100 PAZ domain IPR000999 Ribonuclease III domain IPR036389 Ribonuclease III, endonuclease domain superfamily
Argonaute (Wago-1 <i>C.elegans</i>)	IPR003100 PAZ domain IPR003165 Piwi domain	tig00000338_68 IPR003165 Piwi domain IPR036085 PAZ domain superfamily IPR036397 Ribonuclease H superfamily IPR014811 Argonaute, linker 1 domain
Exportin (Exportin-1 <i>S.cerevisiae</i>)	IPR045065 Exportin-1/5	tig00000146_224 IPR045065 Exportin-1/5
Vret (<i>D.melanogaster</i>)	IPR002999 Tudor domain IPR047396 Protein vreteno, second Tudor domain	
Yb (YBX1 Human)	IPR002059 Cold-shock protein, DNA-binding IPR011129 Cold shock domain IPR019844 Cold-shock (CSD) domain	
GW (gawky <i>D.melanogaster</i>)	IPR015940 Ubiquitin-associated domain IPR041971 Gawky, UBA domain IPR019486 Argonaute hook domain IPR033503 Gawky, RNA recognition motif IPR026805 GW182, middle domain	
RdRpol (ego-1 <i>C.elegans</i>)	IPR007855 RNA-dependent RNA polymerase, eukaryotic-type	tig00000157_163 IPR007855 RNA-dependent RNA polymerase, eukaryotic-type
Hen1 (henn-1 <i>C.elegans</i>)	IPR026610 3'-RNA ribose 2'-O-methyltransferase, Hen1	
Zucchini (Zuc <i>D.melanogaster</i>)	IPR025202 Phospholipase D-like domain	

ANNEXES SECTION B PARTIE III

Supplementary data 1: iRNA-target Interaction network tables

*NbLinkCases indicates the number of times in the kinetics where the iRNA was upregulated and the target was accordingly downregulated. Only targets downregulated everytime the iRNA was upregulated were kept.

A. iRNA targets in host cell

TARGET	baseMean	log2FoldChar	padj	Impact	SOURCE	SiExpression	NbLinkCases
INVS	16.3919462	-1.742077	0.01025022	2.87115134	AaiRNA-6	809.933333	4
TOR1A	14.2459771	-1.9473824	0.01373734	2.03214558	AaiRNA-6	809.933333	4
STX16	7.94494521	-3.6690212	0.01986864	0.20260737	AaiRNA-6	809.933333	4
NEK1	5.28597632	-3.8134042	0.01474324	0.11667688	AaiRNA-6	809.933333	4
THADA	43.0112765	-2.4122018	9.52E-07	3.85457421	AaiRNA-5	1099.06667	4
TOR1A	14.2459771	-1.9473824	0.01373734	2.03214558	AaiRNA-5	1099.06667	4
GRIK1	31.6782792	-1.3155587	0.01441169	8.50005408	AaiRNA-11	374.533333	3
RASGRP3	7.08392454	-1.6278438	0.00208472	1.39094631	AaiRNA-9	3369.4	2
PLCG1	7.38947357	-0.6878954	0.03769676	3.71419181	AaiRNA-10	6221.2	2
MAP3K19	3.19609908	-1.8389165	0.00801077	0.50814652	AaiRNA-10	6221.2	2
CECR2	12.2895398	-1.0025329	0.01571908	4.50963192	AaiRNA-1	84016.7333	3

Supplementary data 1: iRNA-target Interaction network tables (suite)

B. iRNA targets in *A.algerae* CDS

TARGET	baseMean	log2FoldChar	padj	Impact	SOURCE	SiExpression	NbLinkCases
tig00000668:177520-178483_CDS	2004.31137	-0.2557271	0.01618128	1552.04501	AaiRNA-5	1099.06667	4
tig00000358:112925-113747_CDS	1685.21394	-0.4002079	0.00017608	1129.39786	AaiRNA-5	1099.06667	4
tig00000281:320439-321492_CDS	1762.14653	-0.2425348	0.00097522	1382.64436	AaiRNA-8	1468.2	6
tig00000278:73739-74909_CDS	1603.633	-0.1430281	0.00820063	1389.9164	AaiRNA-11	374.533333	3
tig00000157:90724-91606_CDS	1234.27989	-0.145141	0.0129851	1067.5291	AaiRNA-11	374.533333	3
tig00000281:335347-336772_CDS	1102.32025	-0.1562327	0.01511744	942.880766	AaiRNA-11	374.533333	3
tig00000718:7983-9099_CDS	1608.67837	-0.1182207	0.00636429	1429.31058	AaiRNA-9	3369.4	2
tig00000313:425568-426030_CDS	449.901088	-0.1639891	0.02436602	381.854131	AaiRNA-9	3369.4	2
tig00000157:825700-827155_CDS	1917.18232	-0.1030416	0.02509058	1729.46997	AaiRNA-10	6221.2	2
tig00000146:583629-584448_CDS	701.37999	-0.1763563	0.03762014	587.980341	AaiRNA-10	6221.2	2
tig00000313:425568-426030_CDS	449.901088	-0.1639891	0.02436602	381.854131	AaiRNA-10	6221.2	2
tig00000157:1268258-1273109_CDS	4456.15118	-0.1174719	0.00784794	3962.25602	AaiRNA-2	31.4666667	4
tig00000343:69167-69704_CDS	596.390104	-0.2476079	0.00704345	465.581475	AaiRNA-2	31.4666667	4
tig00000143:332694-333165_CDS	3111.00031	-0.2690478	0.00021574	2377.13626	AaiRNA-3	89.8666667	2
tig00000146:65222-66950_CDS	2282.8516	-0.1726889	0.00275937	1920.78976	AaiRNA-3	89.8666667	2
tig00000157:711305-712532_CDS	2066.98182	-0.1479313	0.00244782	1782.75182	AaiRNA-3	89.8666667	2
tig00000146:138240-139377_CDS	1965.32204	-0.1562216	0.00387203	1681.07673	AaiRNA-3	89.8666667	2
tig00000146:62372-63518_CDS	1013.10511	-0.2129943	0.00018096	818.751814	AaiRNA-3	89.8666667	2
tig00000146:679133-680183_CDS	921.871542	-0.1528448	0.01880291	791.208195	AaiRNA-3	89.8666667	2
tig00000597:219125-219539_CDS	464.64438	-0.2438399	0.02305093	364.101407	AaiRNA-3	89.8666667	2
tig00000278:134261-135440_CDS	1843.19009	-0.1496127	0.0254053	1587.06293	AaiRNA-4	592.666667	2
tig00000313:410533-411637_CDS	1533.23115	-0.1571266	0.00902344	1310.29306	AaiRNA-4	592.666667	2
tig00000143:160352-161285_CDS	1137.87332	-0.1530534	0.01035105	976.390776	AaiRNA-4	592.666667	2
tig00000668:238295-239219_CDS	1110.74716	-0.1653798	0.00671767	941.437897	AaiRNA-4	592.666667	2
tig00000278:523622-524330_CDS	988.766388	-0.2224206	0.00179714	791.585197	AaiRNA-4	592.666667	2
tig00000668:188373-189381_CDS	843.674706	-0.2224261	0.00991214	675.424178	AaiRNA-4	592.666667	2
tig00000718:66115-66958_CDS	792.775414	-0.2288441	0.00724732	630.615186	AaiRNA-4	592.666667	2
tig00000718:78614-79265_CDS	594.347497	-0.1528739	0.03614567	510.091572	AaiRNA-4	592.666667	2
tig00000282:215290-215698_CDS	547.25936	-0.201048	0.03069907	447.588758	AaiRNA-4	592.666667	2
tig00000278:410207-410735_CDS	824.213556	-0.6811287	2.10E-06	417.089543	AaiRNA-4	592.666667	2
tig00000146:339439-339868_CDS	466.731998	-0.2126569	0.04299616	377.321781	AaiRNA-4	592.666667	2
tig00000303:174444-174837_CDS	412.103479	-0.2622762	0.00497437	317.03058	AaiRNA-4	592.666667	2
tig00000143:619232-620498_CDS	2605.63869	-0.1316942	0.00907665	2284.12627	AaiRNA-1	84016.7333	2
tig00000278:134261-135440_CDS	1843.19009	-0.1496127	0.0254053	1587.06293	AaiRNA-1	84016.7333	2
tig00000313:410533-411637_CDS	1533.23115	-0.1571266	0.00902344	1310.29306	AaiRNA-1	84016.7333	2
tig00000143:160352-161285_CDS	1137.87332	-0.1530534	0.01035105	976.390776	AaiRNA-1	84016.7333	2
tig00000668:238295-239219_CDS	1110.74716	-0.1653798	0.00671767	941.437897	AaiRNA-1	84016.7333	2
tig00000281:391623-392466_CDS	968.421252	-0.1896944	0.00557747	801.089537	AaiRNA-1	84016.7333	2
tig00000278:523622-524330_CDS	988.766388	-0.2224206	0.00179714	791.585197	AaiRNA-1	84016.7333	2
tig00000718:66115-66958_CDS	792.775414	-0.2288441	0.00724732	630.615186	AaiRNA-1	84016.7333	2
tig00000718:78614-79265_CDS	594.347497	-0.1528739	0.03614567	510.091572	AaiRNA-1	84016.7333	2
tig00000157:68279-68813_CDS	563.993733	-0.2217699	0.02129461	451.815209	AaiRNA-1	84016.7333	2
tig00000632:386026-386329_CDS	511.266814	-0.2464445	0.017693	399.593226	AaiRNA-1	84016.7333	2
tig00000146:339439-339868_CDS	466.731998	-0.2126569	0.04299616	377.321781	AaiRNA-1	84016.7333	2

C. iRNA targets in *A.algerae* TEs

TARGET	baseMean	log2FoldChar	padj	Impact	SOURCE	SiExpression	NbLinkCases
tig00000278:563506-570832_mobile_element_retro_BA_LTR/Ty3	17698.3694	-1.2147907	0.00012005	5252.38242	AaiRNA-8	1468.2	6
tig00000718:15864-17381_mobile_element_DDE3_F_DNA/Mariner-Tc1	1559.06498	-0.1464221	0.01650765	1346.70939	AaiRNA-11	374.533333	3
tig00000157:612663-613845_mobile_element_DP_DDE_TpnlS1595_DNA/Merlin	1238.65883	-0.1655142	0.00343615	1049.71115	AaiRNA-9	3369.4	2
tig00000157:440512-441566_mobile_element_CX_DDE_TpnlS1595_DNA/Merlin	10709.1559	-1.0763196	0.00049059	3650.19101	AaiRNA-3	89.8666667	2
tig00000282:280681-281723_mobile_element_CX_DDE_TpnlS1595_DNA/Merlin	942.502008	-0.2273548	0.02475951	750.832971	AaiRNA-3	89.8666667	2
tig00000303:398289-398574_mobile_element_DDE3_J_DNA/Mariner-Tc1	389.858127	-0.3955212	0.01531983	262.502793	AaiRNA-3	89.8666667	2
tig00000303:417026-417301_mobile_element_DDE3_J_DNA/Mariner-Tc1	319.233139	-0.31971	0.00380561	231.878065	AaiRNA-3	89.8666667	2
tig00000157:718951-719162_mobile_element_retroB_LTR/Ty3	291.963423	-0.3535719	0.00144608	205.009563	AaiRNA-3	89.8666667	2
tig00000343:273107-273159_mobile_element_M_DDE_TpnlS1595_DNA/Merlin	57.3790376	-0.5704704	0.00860625	32.4340455	AaiRNA-3	89.8666667	2
tig00000143:1377242-1377781_mobile_element_CK_DDE_TpnlS1595_DNA/Merlin	2714.98415	-0.4278565	0.01839816	1769.91163	AaiRNA-4	592.666667	2
tig00000278:293805-295080_mobile_element_AL_DDE_TpnlS1595_DNA/Merlin	1240.75652	-0.1974882	0.00052741	1018.40031	AaiRNA-4	592.666667	2
tig00000278:420038-421228_mobile_element_AO_DDE_TpnlS1595_DNA/Merlin	1020.74132	-0.1761315	0.01809556	855.899433	AaiRNA-4	592.666667	2
tig00000622:16763-17577_mobile_element_AU_DDE_TpnlS1595_DNA/Merlin	805.119996	-0.1839134	0.01209423	669.866189	AaiRNA-4	592.666667	2
tig00000146:1290662-1290947_mobile_element_retro_AR_LTR/Ty3	810.752643	-0.273724	0.00041062	616.611388	AaiRNA-4	592.666667	2
tig00000313:316351-317130_mobile_element_DDE3_E_DNA/Mariner-Tc1	672.597225	-0.1904644	0.0247716	555.95215	AaiRNA-4	592.666667	2
tig00000303:22678-22961_mobile_element_retro_BJ_LTR/Ty3	424.372	-0.3549731	0.00862126	297.566387	AaiRNA-4	592.666667	2
tig00000668:574915-575239_mobile_element_DC_DDE_TpnlS1595_DNA/Merlin	447.352005	-0.472052	0.00088109	279.022862	AaiRNA-4	592.666667	2
tig00000632:531868-532128_mobile_element_retro_BJ_LTR/Ty3	320.856074	-0.2547468	0.01840886	248.699616	AaiRNA-4	592.666667	2
tig00000622:3394-3581_mobile_element_retroM_LTR/Ty3	176.19646	-0.3140556	0.0263409	128.707706	AaiRNA-4	592.666667	2
tig00000313:86611-86857_mobile_element_retro_AR_LTR/Ty3	171.561668	-0.4107347	0.00636921	113.773327	AaiRNA-4	592.666667	2
tig00000143:1377242-1377781_mobile_element_CK_DDE_TpnlS1595_DNA/Merlin	2714.98415	-0.4278565	0.01839816	1769.91163	AaiRNA-1	84016.7333	2
tig00000278:293805-295080_mobile_element_AL_DDE_TpnlS1595_DNA/Merlin	1240.75652	-0.1974882	0.00052741	1018.40031	AaiRNA-1	84016.7333	2
tig00000622:16763-17577_mobile_element_AU_DDE_TpnlS1595_DNA/Merlin	805.119996	-0.1839134	0.01209423	669.866189	AaiRNA-1	84016.7333	2
tig00000146:1290662-1290947_mobile_element_retro_AR_LTR/Ty3	810.752643	-0.273724	0.00041062	616.611388	AaiRNA-1	84016.7333	2
tig00000278:157970-158200_mobile_element_retroJ_LTR/Ty3	230.368079	-0.3796225	0.01874009	157.59932	AaiRNA-1	84016.7333	2
tig00000622:3394-3581_mobile_element_retroM_LTR/Ty3	176.19646	-0.3140556	0.0263409	128.707706	AaiRNA-1	84016.7333	2
tig00000622:696730-696979_mobile_element_retroG_LTR/Ty3	171.718446	-0.311627	0.02202769	125.741629	AaiRNA-1	84016.7333	2
tig00000242:386921-387133_mobile_element_retroB_LTR/Ty3	134.326702	-0.3209636	0.04208067	97.4472555	AaiRNA-1	84016.7333	2
tig00000143:523612-523696_mobile_element_tig00000597:333642-334374_DNA/U	84.4376792	-0.5379627	0.01287185	49.3062597	AaiRNA-1	84016.7333	2
tig00000668:183529-183581_mobile_element_retroW_LTR/Ty3	35.6037257	-0.9098597	0.0045661	14.3333726	AaiRNA-1	84016.7333	2



Summary

Microsporidia are a family of obligate intracellular parasites that can infect a wide variety of hosts ranging from invertebrates like bees to vertebrates such as fish or even humans. Despite their innocuity in most immuno-competent hosts, their ability to infect various cell types and species has placed them as potential emergent threats.

Given their obligate intracellular lifestyle, microsporidia constantly and vastly interact with their host cells to assure their survival, proliferation, and nutrition. As parasites, they use various mechanisms such as signalling interference and metabolic pathway hijacking to alter host behaviour and maintain their presence.

In recent years, thanks to the advent of next generation sequencing techniques, the role of noncoding RNA as major cell regulatory molecules has emerged. They have important roles in normal cell behaviour, immunity, physiopathology, and cell-pathogen interactions. This raises the question of their potential roles in host-parasite crosstalk during infection by a parasite.

Microsporidia are a great model to study this. On top of their large host spectra, this family has evolved in a genome reducing pattern which can be seen by their small genomes (from 2.3Mb to 51.3 Mb) ; small number of protein genes (~2000-3000); the reduction of all coding regions when compared to orthologs in other eukaryotes; and the high level of gene compaction with intergenic spaces that can be reduced to a few nucleotides. Although the low number of genes comes with the loss of numerous cellular and metabolic pathways in microsporidia, an RNA interference machinery has been identified in more than half of the sequenced microsporidian genomes. In all the genomes harboring this machinery, transposable elements have also been highlighted suggesting a role of the RNA interference process in the regulation of TE in microsporidia.

In this work we focus on the specie *Anncaliia algerae* which can infect both Humans and the phylogenetically distant anopheles, and which presents a striking diversity of transposable elements inside its genome.

We show that noncoding RNA play important roles in microsporidian infection and biology, both from the host, and the parasite's perspectives. This knowledge serves to better understand interactions between parasites and their hosts and may contribute to the conceptualisation of novel strategies against intracellular parasite infections, in humans, but also animals.

Keywords : Host-parasite interactions ; non-coding RNA ; Microsporidia ; *Anncaliia algerae* ;

Résumé

Les microsporidies sont une famille de parasites intracellulaires obligatoires qui peuvent infecter une grande variété d'hôtes, allant des invertébrés comme les abeilles, aux vertébrés comme les poissons ou même l'Homme. Malgré leur innocuité chez la plupart des hôtes immunocompétents, leur capacité à infecter divers types de cellules les présente comme un risque émergent.

Compte tenu de leur mode de vie intracellulaire obligatoire, les microsporidies interagissent constamment avec leurs cellules hôtes pour assurer leur survie, leur prolifération et leur nutrition. En tant que parasites, elles utilisent divers mécanismes de subversion afin de maintenir leur présence dans leurs hôtes.

Ces dernières années, grâce à l'avènement des techniques de séquençage de nouvelle génération, un rôle majeur des ARN non codant (ARNnc) dans les processus d'interactions cellulaires a émergé. Les ARNnc jouent un rôle important dans la physiologie cellulaire, l'immunité, la physiopathologie et les interactions entre cellules et pathogènes. Cela soulève la question de leur rôle potentiel dans les interactions hôte-parasite lors d'une infection.

Les microsporidies sont un excellent modèle pour étudier ces aspects. En plus de leur large spectre d'hôtes, ces organismes ont évolué selon un schéma de compaction génomique visible par leurs petits génomes (de 2,3 Mo à 51,3 Mo) ; leur nombre réduit de gènes codants des protéines (~ 2 000 à 3 000) ; la réduction de la taille de ces gènes par rapport à d'autres eucaryotes ; et leurs espaces intergéniques pouvant être réduits à quelques nucléotides.

Ces réductions s'accompagnent de la perte de nombreuses voies cellulaires et métaboliques dans les microsporidies, cependant, une machinerie d'interférence d'ARN a été identifiée dans plus de la moitié des génomes séquencés. Dans tous les génomes hébergeant cette machinerie, des éléments transposables (ET) ont été mis en évidence suggérant un rôle du processus d'interférence ARN dans la régulation des ET chez les microsporidies.

Dans ce travail nous nous sommes concentrés sur l'espèce *Anncaliia algerae* qui peut infecter aussi bien l'Homme que les anophèles, et qui présente une diversité frappante d'éléments transposables à l'intérieur de son génome.

Nous montrons que les ARN non codants jouent un rôle important dans la biologie des microsporidies et dans leur processus infectieux, tant au niveau de l'hôte que du parasite. Ces connaissances permettent de mieux comprendre les interactions entre les parasites et leurs hôtes et pourraient contribuer à la conceptualisation de nouvelles stratégies contre les infections parasitaires intracellulaires, chez l'Homme, mais aussi chez l'animal.

Mots clés : Interactions hôte-parasite ; ARN non-codants ; Microsporidies ; *Anncaliia algerae* ;
