Genomic and transcriptomic landscape of the microsporidian parasite *Anncaliia algerae*

Ivan WAWRZYNIAK[1], Reginald Florian AKOSSI[1], Emmanuelle LERAT[2], Frederic DELBAC[1], Eric PEYRETAILLADE[1]

1 Laboratoire Microorganismes: Génome et Environnement, UMR 6023, CNRS, Université Clermont Auvergne, 63000 Clermont-Ferrand, France
2 Laboratoire de Biométrie et Biologie Evolutive, UMR5558, CNRS, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

Corresponding Authors : ivan.wawrzyniak@uca.fr, eric.peyretaillade@uca.fr

Acknowledgments

This work was supported by a grant from the sfbi

We are grateful to the AuBi platform and the Mésocentre of Clermont Auvergne University for providing computing resources

Introduction :. *Anncaliia algerae* is a eukaryotic parasite belonging to Microsporidia, which are obligate intracellular pathogens related to Fungi. Initially identified as a mosquitoe pathogen, *A.algerae* has emerged as a rare opportunistic human pathogen in immunocompromised patients. Compared to other microsporidian species, genomic and transcriptomic data for *A. algerae* are limited. Although three strains of *A. algerae* have been sequenced using both Sanger and Illumina methods, complete genome reconstruction remains challenging due to the high abundance of repetitive elements and a low GC content of around 20%. We therefore resequenced the genome of the Undeen strain of *A. algerae* using PacBio HiFiSeq technology, a long-read sequencing method. Additionally, we re-annotated this genome using the MicroAnnot pipeline [1] recently developed in our team and specifically designed for microsporidian genome annotation. In parallel, we performed a comprehensive transcriptomic analysis by sequencing small RNAs and polyadenylated RNAs from *A. algerae*-infected human cells at five time points throughout the infection cycle (3h, 12h, 24h, 48h, and 72h) using Illumina Novaseq 6000 sequencing.

Experimental design

Genome sequencing & assembly

DNA extraction (*A .algerae* spores) → **Long-read sequencing** (PacBio Sequel II HiFiseq) → **Assembly** (HiCanu V 2.1) → **Sequence filtering** (GC 40%) → **Curated ploidy** (Purge Haplotigs) → **Contamination analysis** (Blast) → **Genome polyploidy** (Genoscope & Smudgeplot)

Genome annotation

***A. algerae* genome** → **Transposable Elements identification** (Repeatmasker) → **Gene prediction** (MicroAnnot pipeline) → **CDS Validation** (use of transcriptomics data) → **Functional annotation** (InterProscan)

Transcriptomic

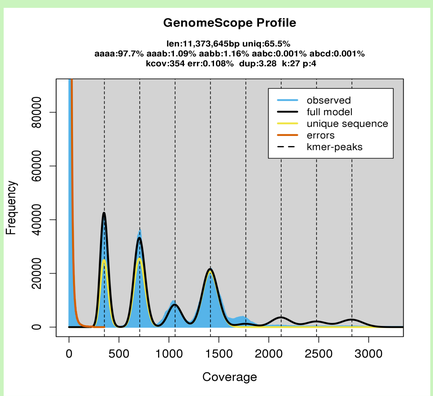
Kinetic of HFF Cells infection by *A. algerae* → **Total RNA extraction** (5 times post-infection (3h, 12h, 24h, 48, 72h) in triplicate) → **Dual RNA-seq** (Poly A & small RNA, Illumina Novaseq 6000) → **Identification of non coding RNA** (microRNA, LncRNA) → **HFF & *A. algerae* transcriptomics analysis**

Genome analysis

- Pac Bio sequencing : 2742535 CCS with a mean size of 11,9 kb and a total size of 25 Gb
- Assembly confirming a tetraploid genome [2]
- Haploid genome assembled from telomere to telomere

Table of summary of *A. algerae* assembly

	Tetraploid	Haploid
Number of contigs	239	17
Total length	43.9 Mb	11.7 Mb
N 50 length	0.5 Mb	0.73 Mb



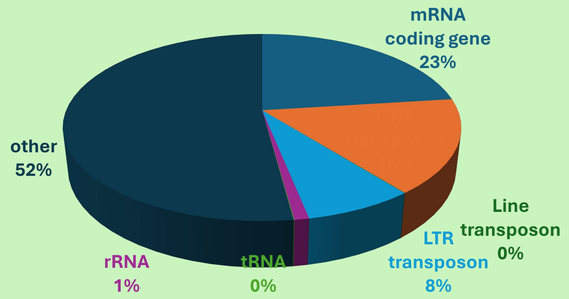
GenomeScope validation of the tetraploidy

Genome annotation

Table of summary of *A. algerae* annotation

	Tetraploid	Haploid
Number of protein coding genes	10369	2916
Number of tRNA	210	56
Number of complete SSU/ LSU genes	34	14
Gene density / Mb	237	235
Busco completeness (%)	81.9	80.6

High conservation of the gene synteny between homologous chromosomes

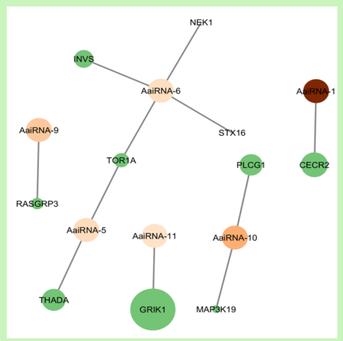


4972 Transposable elements copies identified in the haploid genome corresponding to 260 families and representing 25% of the genome sequence

Transcriptomic

- Poly A RNAseq runs yielded between 1000 to 29000 Mb, with an average of 6000 Mb per sample. The percentage *of A. algerae* reads range from 9.9% to 24.6%.
- Small RNAseq runs yielded between 2000 to 8000 Mb, with an average of 2700 Mb per sample.
- RNAseq data analysis revealed a massive miRNA downregulation in human cells [3] induced by *A. algerae*. This could be beneficial to the parasite.

MirDeep analysis revealed 11 different mature *A. algerae* miRNA overexpressed during the infection kinetic. Targets were identified for 6 of them and were involved in apoptotic, DNA fragmentation processes or activation of protein kinase activity.



Interaction network showing miRNA from *A. algerae* and the predicted human gene targets.

Conclusions

- The sequencing identified 17 chromosomes and the tetraploid nature of *A. algerae*. Genome annotation and transcriptomic data analysis revealed 260 transposon families and a large set of non-coding transcripts.
- Our study revealed that *A. algerae* expresses small non-coding RNAs, which may play a role in regulating host cell function, particularly apoptosis. This phenomenon highlights the crucial role of miRNAs in host-parasite interactions
- Our observations revealed that *A. algerae* infection induces transcriptomic, caracterised by a high repression of host miRNAs throughout the infection.

References : 1 - Tournayre J et al : A Dedicated Workflow for Accurate Microsporidian Genome Annotation. Int J Mol Sci. 2024 Janv 10;25(2):880
2 - Khalaf A et al : Polyploidy is widespread in Microsporidia. Microbiol Spectr. 2024 Feb; 12(2): e03669-23
3 - Akossi RF et al : The intracellular parasite *A. algerae* induces a massive miRNA down-regulation in human cells. Noncoding RNA Res 2023 May 18;8(3):363-375