# Genome Sequencing of the Relevant Zebrafish-Infecting Microsporidian *Pseudoloma neurophilia* Reveals Atypical Genome Dynamics

By
Steve Ndikumana

Supervisor
Dr. Nicolas Corradi

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
In partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology

Thèse soumise à la
Faculté des études supérieures et Postdoctorales
Université d'Ottawa
En vue de l'obtention de la maîtrise en science à
L'Institut de biologie d'Ottawa-Carleton

# Abstract

Since their first discovery in the 19th century, microsporidian species have been found to be successful obligate intracellular parasites capable of infecting a wide variety of hosts including economically and ecologically important organisms as well as model organisms for biomedical research. Recently, cases of infection of the widely used animal model *Danio rerio*, commonly known as the zebrafish, by the newly described microsporidium *Pseudoloma neurophilia* have been reported in an increasing number of research facilities. Current knowledge of the biology of this parasite found in 75% of the Zebrafish Resource Center facilities is limited to microscopic analyses on its lifecycle as well as its physical, behavioral and psychic impact on its hosts. Despite the growing relevance of this parasite in biomedical research no current data is available on its genome. In this dissertation, I provide additional knowledge on the basic biology of *P. neurophilia* by acquiring and exploring the content and structure of the first genome draft of the zebrafish parasite. My findings reveal that the 5.25 Mb genome of *P. neurophilia* harbors an unusually high amount of transposable elements as well as numerous inserts found in coding regions typically conserved in microsporidia and other organisms. This peculiar obligate parasite demonstrates strong phylogenetic and genetic relationships with other fish-mosquito microsporidia. Similar to what is observed in closely related species, intra-genomic analyses of *P. neurophilia*'s genome suggest that it is diploid and possesses a large repertoire of over a thousand putative genes unique to this specie. Overall, my findings provide new insights into the basic biology of this parasite and represent a milestone in the understanding of *P. neurophilia* and *D. rerio* host-parasite interaction and ultimately in the development of treatments against this parasite that has been infecting the zebrafish research industry for the past decades.

# Résumé

Les espèces de microsporidies sont des parasites obligatoirement intracellulaires ayant la capacité d'infecter une grande variété d'hôtes, y compris certains organismes importants d'un point de vue économique et écologique ainsi que certains organismes modèles utilisés dans de nombreux centres de recherche. Récemment, plusieurs cas d'infections de l'organisme modèle *Danio* rerio (poisson-zèbre) par la microsporidie *Pseudoloma neurophilia* ont été signalés dans un nombre croissant de centres de recherche. Présentement, les connaissances actuelles sur la biologie de ce parasite trouvé dans 75% des installations du Centre de ressources de poisson-zèbre sont limitées à la description de son cycle de vie ainsi que l'impact physique, comportemental et psychique sur ses hôtes. Malgré l'importance croissante de ce parasite dans la recherche biomédicale, aucune donnée actuelle n'existe sur son génome. Dans cette thèse, je fournis des connaissances supplémentaires sur la biologie de base de *P. neurophilia* en acquérant et explorant le contenu et la structure de la première ébauche du génome du parasite du poisson-zèbre. Mes résultats démontrent que le génome de 5,25 Mb de *P. neurophilia* possède une quantité inhabituellement élevée d'éléments transposables ainsi que de nombreuses insertions dans des régions codantes conservées chez les microsporidies et d'autres organismes. Ce parasite démontre de fortes relations phylogénétiques et génétiques avec d'autres microsporidies infectant les poissons et moustiques. Similaire à ce qui est observé chez ses espèces proches, les analyses intra-génomiques du génome de *P. neurophilia* suggèrent qu'il est diploïde et possède un vaste répertoire de plus d'un millier de gènes putatifs et uniques à cette espèce. Dans l'ensemble, mes résultats fournissent de nouveaux aperçus sur la biologie fondamentale de ce parasite et représentent une première étape importante dans la compréhension de l'interaction hôte-parasite entre *P. neurophilia* et *D. rerio* et finalement, dans le développement de traitements contre ce

parasite qui affecte de façon non-négligeable la recherche sur le poisson-zèbre depuis plusieurs décennies.

# Table of Contents

## Acknowledgment

First and foremost, I would like to thank my thesis supervisor Dr. Nicolas Corradi for all his support, time and energy invested in me and my projects. For more than three years, he has guided and motivated me during the most challenging time. I am most grateful for your unparalleled devotion and will cherish all the moments that we have shared together. I would also like to extend this thanks to my committee members Dr. Marc Ekker and Dr. Ashkan Golshani for their invaluable advices and feedbacks.

I would like to take this opportunity to acknowledge and thank current and former lab members for all their help and guidance since day one; Adrian Pelin, Philippe Charron, Timea Marton, Jeanne Ropars, Denis Beaudet, Eric Chen, Jessica Noel, Rohan Riley, Mohammed Selman, as well as newer members of the Corradi lab. My appreciation also goes to Stefan Amyotte for his guidance and support through my debut in research as well as to Benoit Tremblay for his invaluable friendship as a fellow Master' student. Additionally, I would like to thank my collaborators Justin Sanders and Michael Kent for providing us with spores from infected zebrafish.

Last but certainly not least, a special thanks goes to my mother, father, siblings and friends for their unconditional encouragement and moral support. Encore une fois, merci!

# List of tables



# List of figures

# Chapter 1 – Introduction

## Overview on Microsporidians

Microsporidia are highly successful obligate intracellular parasites capable of infecting almost all known animal lineages. Their successful invasive strategies have been linked with mortality or development of infectious diseases in many ecological and economical important eukaryotes. One notorious example includes the honey bee colony collapse disorder, which has been associated with the presence of two species of this intracellular parasitic group; namely *Nosema apis* and *N. ceranae* (Higes *et al.*, 2013). Other species of microsporidia can also infect humans (*Encephalitozoon cuniculi*), as well as some important aquatic models (i.e. the zebrafish *Danio rerio*) used in studies of different aspects of the human health (Lobo *et al.*, 2012; Matthews *et al.*, 2001).

Recent genome sequencing projects have unraveled the dynamics of microsporidian genomes as well as demonstrated that an adaptation for an obligate parasitic life style have led to massive reductions in gene content and a shrinkage of intergenic regions and protein lengths (Corradi *et al.*, 2010; Corradi & Slamovits, 2011). These reductive forces led to the absence or reduction of many biochemical and cellular features usually present in other eukaryotes, such as the *de novo* biosynthesis of amino acids and nucleotides and the tricarboxylic acids cycle pathways (Corradi & Slamovits, 2011). Microsporidia also lack a conventional Golgi apparatus, and possess reduced mitochondria that have lost their genomes, known as mitosomes. These characteristics have led to the hypothesis that they were ancient eukaryotes, however, recent

phylogenetic analyses provided rigorous support suggesting their relationship as a sister group to the fungal kingdom (James *et al.*, 2013).

## Microsporidia: The taxonomic nomads

### *Early classification of microsporidia*

Since their first discovery in 1857 by Karl Wilhelm von Nägeli (Nageli, 1857), the diversity and our knowledge of microsporidia have greatly improved primarily due to the development of new molecular and microscopic techniques as well as bioinformatics tools. Since then, approximately 1500 species in almost 200 genera have been identified (Keeling & Fast, 2002). Many of these species demonstrate the potential to be host non-specific while a significant portion are presently considered host specific (Ardila-Garcia & Fast, 2012; Corradi, 2015; Desjardins *et al.*, 2015).

Despite the initial characterization of microsporidia as parasites of immuno-compromised patients, several studies have identified microsporidia in immune-competent individuals (Lobo *et al.*, 2012). The first case of microsporidiosis in humans was detected in an HIV positive patient in 1985 in the United States of America (Franzen & Müller, 2001; Katinka *et al.*, 2001); a century after the first microsporidian was identified. The latter microsporidia (*Nosema bombycis*) was a notorious parasite found to be mainly responsible for the Pébrine disease, a disease that destroyed the silkworm industry in the 20$^{th}$ century (Ma *et al.*, 2013).

First described as a member of the Schizomycetes by Nägeli using microscopy, this family of intracellular obligate parasites has since been reclassified several times, and their

taxonomy has been a source of major debates in this field of study (Corradi & Keeling, 2009; Nageli, 1857). Indeed, in 1987, the first phylogenetic analysis of microsporidia, based on SSU rRNA genes led the scientific community to hypothesize that microsporidia were part of the Archezoa group; a kingdom that included all unicellular eukaryotes with no apparent mitochondria (Vossbrinck *et al.*, 1987). This classification of microsporidia as ancient eukaryotes was supported by the absence of typical mitochondrial organelles, peroxisomes, centrioles, conventional Golgi apparatus, as well as the presence of 70S prokaryote-like ribosomes (Keeling & Fast, 2002). The notion that microsporidia are ancient eukaryotes was further supported by a phylogeny based on EF-1 alpha orthologues (Kamaishi *et al.*, 1996).

***Fungal identity struggle***

In the mid-90s, the deep position of microsporidians in the tree of life was proposed to be resulting from long branch attractions, an artefactual placement typical of highly diverged sequences. New phylogenetic analyses taking into consideration these parameters were later performed and suggested a relationship between Microsporidia and Fungi. Indeed, in 2000, the first phylogeny proposing a relationship between the Fungal kingdom and microsporidia was based on beta-tubulin orthologs (Keeling *et al.*, 2000). In this study, Keeling *et al*. demonstrated that microsporidia branched as a sister-group to the chytrids groups (basal fungi). Other phylogenies based on protein-encoding genes (alpha and beta-tubulin genes) further supported the hypothesis that microsporidia branched with Fungi, but particularly with zygomyces, a fungal group with many parasites (Keeling, 2003). In 2006, a 6 genes phylogeny also provided evidence supporting the initial fungal hypothesis (James *et al.*, 2006). However, in this study, the authors reconstructed the fungal tree demonstrating that microsporidia branched with the chytrids.

More recent phylogenetic analyses based on 53 genes revealed that microsporidia represent an early divergent clade of Fungi (Capella-Gutiérrez *et al.*, 2012). This hypothesis was later supported by independent phylogenies, all of which exposed microsporidia as being basal sister to the fungal tree of life, and as a sister-group to *Rozella allomycis* (James *et al.*, 2013). The latter phylogeny included sequences from the first sequenced Cryptomycotan *R. allomycis* and genome analyses demonstrated that this specie and microsporidians share many genomic features that are absent in other eukaryotes. These include chitinase genes, nucleotide phosphate transporters as well as nucleoside H+ symporters; which are all thought to have been acquired by horizontal gene transfers. Additionally, *R. allomycis* possesses mitochondria with a very reduced gene content and an extremely high AT content (90% AT). These latter features, combined with the basal position of these two groups in the Fungal tree, provide a strong evolutionary link between Microsporidia and Cryptomycota.

This phylogenetic positioning was recently supported by the discovery of the first microsporidian-like specie with a mitochondrial organelle, *Mitosporidum daphniae* (Haag *et al.*, 2014). The basal positioning of *M. daphnia* in the microsporidian group, which lies in between Rozella and Microsporidia, suggests that the loss of a functional mitochondrion in microsporidia is a recent event that appeared after the extreme genome content reduction observed in members of this group. This discovery represents additional strong evidence supporting the position of microsporidia as sister group of Fungi, close to the Cryptomycota.

## A simple yet successful parasitic lifecycle.

### *The microsporidian extracellular life stage*

Microsporidia are successful intracellular obligate parasites that possess a unique infectious mechanism as well as the means to survive under extreme conditions when outside of their hosts (Keohane & Weiss, 1998). In extracellular environment, the microsporidium is found under the form of spores of 1-20 µm in diameter. Spores have a cell wall composed of chitin that envelop the microsporidian sporoplasm and polar tube (Fig. 1.1). The polar tube is a hallmark of microsporidia and is approximately up to 25 fold (50-500 µm) the size of its spores. This apparatus is enfolded within the spore and is fully extruded under appropriate conditions (Franzen, 2004). Indeed, in the presence of a potential host cell, the initial step of the microsporidian infectious cycle, named the germination, is initiated (Franzen, 2004; Xu & Weiss, 2005). While the stimulus that triggers this step has not yet been fully described, it is now accepted that this germination results in an increase in osmotic pressure within the spore, and an inflation of the posterior vacuole that leads to the expulsion of the polar tube out of the microsporidian spore and into the extracellular environment.

To date, several hypotheses have been proposed to explain this observed increase in osmotic pressure. The first, consists on the intake of extracellular water into the parasite vacuoles hence resulting in an increase in osmotic pressure (Franzen, 2004; Ghosh *et al.*, 2006; Xu & Weiss, 2005). Another consists of the release of calcium by the microsporidian which could induce the parasite water flow and hence have an impact on its osmotic pressure (Keohane *et al.*, 1996). Nevertheless, despite the different processes resulting in this increase in osmotic pressure, once it reaches an ideal level, the 500 µm polar tube is ejected out the host into the external

environment. Interestingly, the polar tube is a fully synthesized apparatus that is present and ready to be used during this extracellular lifecycle stage (Franzen, 2004). The ejection of this apparatus is considered successful if the polar tube pierces a cell. Under this condition, the content of the spore's sploroplasm is transferred from the parasite into its new host cell.

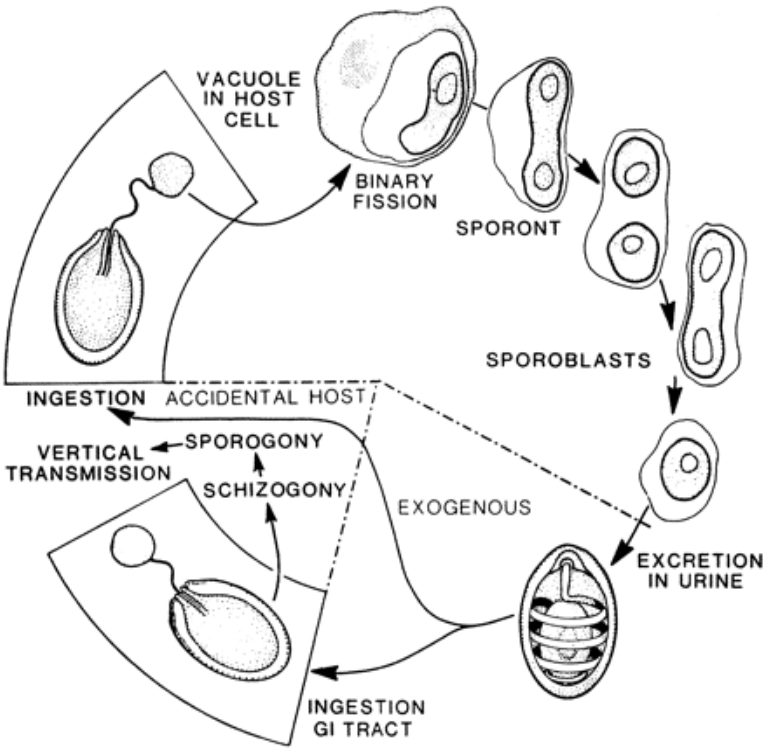



**Figure 1.1** Schematic representation of the microsporidian lifecycle (Garcia, 2002)

***The microsporidian intracellular life stage***

Once inside the host, the microsporidian sporoplasm undergoes the proliferative stage (Fig. 1.1), called merogony (Keeling & Fast, 2002). The microsporididan meronts rapidly replicate through binary or multiple fissions. These meronts are surrounded by host's organelles, including the nucleus, endoplasmic reticulum and mitochondria. This particular cellular characteristic is an important evolutionary feature that facilitates the acquisition of the host's ATP by the parasite. Indeed, since microsporidian species lack the necessary components to produce ATP, they rely on their host's source of energy to survive (Katinka *et al.*, 2001). This peculiar feature was made possible by the acquisition by horizontal gene transfer ADP/ATP transporter of bacterial origins (Tsaousis *et al.*, 2008).

Following merogony, the parasite undergoes sporogony, which includes multiple physical changes such as the reduction of its size and the synthetization of a chitinous endospore (Keeling & Fast, 2002; Vávra & Ronny Larsson, 2014). This results in the maturation of meronts into spores and their release in the environment (Ardila-Garcia & Fast, 2012; Sanders *et al.*, 2013). These spores represent the invasive stage of microsporidia and are responsible for the horizontal propagation of the parasite between species and individuals (Ardila-Garcia & Fast, 2012; Sanders *et al.*, 2013). Microsporidian infections can be the result of the ingestion of contaminated water by an organism, hence supporting the hypothesis that they may be waterborne pathogens, but it has been shown that microsporidian can also propagate through vertical and zoonotic transmission (Jamshidi *et al.*, 2012; Mathis *et al.*, 2005). For instance, cases of cats and dogs being potential microsporidian reservoirs for human infections have been recently proposed.

## A highly dynamic streamed genome

The first microsporidian genome sequencing project (the vertebrate parasite *Encephalitozoon cuniculi*) was an opening into the apparent simplicity of these microorganisms. The genome is only 2.9 Mb and codes for a mere 2000 proteins (Corradi *et al.*, 2010). This is three times smaller than the genome content of the yeast *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996).

Additional microsporidian genome sequencing projects revealed the presence of a similar genome reduction resulting in the absence or presence of reduced biochemical and cellular features usually present in other eukaryotes; such as the *de novo* biosynthesis of amino acids and nucleotides, as well as the tricarboxylic acid pathways (Corradi & Slamovits, 2011). In addition to a massive gene content reduction, the genome of microsporidians demonstrates the presence of reduced intergenic regions and gene length, which seems to be the result of an adaptation to their intracellular obligate lifestyle.

Although all microsporidian genomes have a similar biochemical repertoire (Corradi *et al.*, 2009; Cuomo *et al.*, 2012), their size range from 2.3 (*Encephalitozoon intestinalis*) (Corradi *et al.*, 2010) to 51.3 Mb (*Edhazardia aedis*) (Desjardins *et al.*, 2015) resulting from variations in gene density and in the number of transposable elements. Surprisingly, while microsporidians gene content ranges from approximately 2,000 to 3,500 genes, roughly 20 to 50% of these genes are microsporidian specific (Nakjang *et al.*, 2013) (possess no homologs outside the microsporidian phylum). As microsporidian have undergone a substantial gene content reduction, the observation of the presence of such a significant repertoire of microsporidian specific genes suggest that these proteins might play a crucial role in the survival and/or pathogenicity of these parasites.

## RNAi, a common feature of microsporidian genomes

Recent studies have shown that multiple species of microsporidia have retained the three essential eukaryotic key components of the RNA interference (RNAi) pathway (Campbell *et al.*, 2013; Heinz *et al.*, 2012; Paldi *et al.*, 2010); hence suggesting the functionality of this machinery in microsporidia. These latter components (i.e. the argonaute/PIWI protein, the Dicer protein (or Dicer-like protein) and the RNA-dependent RNA polymerase) are involved with the RNA-induced silencing complex; a complex that cleaves targeted sequences through the use of small regulatory RNAs (Chacko & Lin, 2013; Dang *et al.*, 2011; Shabalina & Koonin, 2008). In eukaryotes, the RNAi pathway plays a role in the regulation of cellular processes, defense mechanisms and host immune responses (Buck *et al.*, 2015; Chacko & Lin, 2013; Ding & Voinnet, 2007). Interestingly, all these three components are present in multiple microsporidian species (Table 2.2).

***RNAi in microsporidia, the future of host-parasite interaction studies?***

While such mechanism is still unknown in microsporidia, the functionality of this pathway has been proposed in one microsporidian species which infects honeybees (*i.e. Nosema cerenae,* Paldi *et al.*, 2010). Indeed, Paldi *et al.* were able to successfully silence the expression of ADP/ATP transporters of *N. cerenae* with the use of double-stranded RNAs providing evidences that double-stranded RNA ingested by the host can silence the expression of genes in *N. cerenae*. Furthermore, the recent discovery of microRNA-like sequences in *N. cerenae* provides additional evidences supporting the functionality of the RNAi pathway in this group (Huang & Evans, 2016).

Microsporidia are intracellular obligate parasites that actively interact with their hosts using a variety of secreted effectors. Small non-coding RNA effectors such as microRNA have been proposed to play a role in cross-kingdom communication between hosts and their respective parasites (Knip *et al.*, 2014). In animal host-parasite interactions, this apparent translocation of gene silencing signals has been described between nematodes (i.e. *Heligmosomoides polygyrus*, *Litomosoides sigmodontis* and *Schistosoma japonicum*) and their respective hosts (Buck *et al.*, 2015; Cheng *et al.*, 2013). To date, two methods have been described for the secretion of microRNA effectors (via active and passive transports) (Chen *et al.*, 2012). The first one requires the microRNAs to be transported out of the parasite cell into the host cell inside macrovesicles or via the intermediate of RNA-binding proteins such as the Argonaute protein, while the second method consists on the circulation of microRNA by passive transport of broken cells due to cell apoptosis or tissue injury.

While RNAi is still a novel theme in microsporidian studies, evidence for the functionality of this pathway in its major sister group as well as the recent discovery of

microRNA-like sequences in a *N. ceranae*, suggest that regulatory RNAs could play a crucial role in gene regulation in these parasites.

## Transposable elements

While evidence of the functionality of the RNAi pathway in microsporidia has recently been obtained (Huang & Evans, 2016; Paldi *et al.*, 2010), the notion of a plausible RNAi pathway in these parasites was already known. A number of studies have proposed that RNAi in microsporidia is a defense mechanism against the proliferation of transposable elements (TE) (Campbell *et al.*, 2013; Heinz *et al.*, 2012). The presence of TEs is rare in the group, and is only observed in a few species and in a limited amount. When present, however, TEs can be numerous, particularly in microsporidia with large genomes, and have been shown to be involved in horizontal gene transfers (Parisot *et al.*, 2014a).

In this study, the two major families of TEs were retrieve (LTR retrotransposons and DNA transposons) and classified in different classes, namely Non-LTR Retrotransposons (LINE), LTR Retrotransposons, as well as four DNA transposons classes (Helitron, Mariner/Tc1, Merlin and Piggybac). While the first family requires the use of a reverse transcriptase and a RNA intermediate for its mobilization, the second family does not involve an RNA intermediate and is characterized by a 'cut and paste' mechanism allowing the relocation of these mobile elements within the genome (Levin & Moran, 2011). As LTR retrotransposons are first transcribed into mRNA by the RNA polymerase II, then double stranded DNA by the reverse transcriptase and finally inserted into the genome by an integrase protein, their mobilization mechanism could enable rapid duplication of their number within their host genome. Indeed, members of this family can be observed in large number in other fungal lineages (i.e *Postia*

*placenta* (3108), *Magnaporthe grisea* (677), maize (~75% of their genome); (Levin & Moran, 2011; Muszewska *et al.*, 2011)).

TEs are DNA sequences that possess the ability to relocate and duplicate at virtually any location within a genome with potential positive or negative repercussions on the organism. Indeed, the insertions of TEs within a coding region can potentially pseudogenize a gene and have a negative repercussion on the associated biochemical pathway (Biémont, 2010a). However, TEs have been shown to generate genomic diversity in regions coding for antigens and effectors hence creating genome plasticity (Raffaele & Kamoun, 2012), so their presence can also be beneficial.

## *Pseudoloma neurophilia*, the zebrafish most common pathogen in research facilities

Microsporidia have well-known impacts on key important sectors (health, ecological and economical), but also on animal research models. Within this context, a recent increase in microsporidian infections of the powerful animal model for biomedical research, the zebrafish (*Danio rerio*) (Matthews *et al.*, 2001), have been reported. In the past decades, this tropical freshwater fish has gained popularity as a vertebrate model in numerous research fields due to its short and rapid embryonic development stages, extensive regenerative capacity, the availability of its complete genome as well as an immune system similar to humans (Goldshmit *et al.*, 2012; Howe *et al.*, 2013; Meeker & Trede, 2008; Spence *et al.*, 2008).

While these characteristics have favored its use as an aquatic model for vertebrate studies, it may also have led to an increase in diagnosed cases of infection of *D. rerio* by the microsporidium *P. neurophilia* (Murray *et al.*, 2011; Sanders *et al.*, 2012). Indeed, cases of

microsporidiosis by *P. neurophilia* have been reported in up to 74% of the facilities of the zebrafish international resource center (Murray *et al.*, 2011); a center dedicated in providing scientists with wild-type and mutant strains of zebrafish.

Since its discovery in 2001, studies on *P. neurophilia* have mainly revolved on studying its lifecycle, transmission methods, microscopic features, stress and behavior changes in host (Cali *et al.*, 2012; Murray *et al.*, 2011; Sisson *et al.*, 2006). Current methods of diagnosis are limited to histopathologic analyses and SSU amplification by PCR, so prevention of its transmission has so far been difficult (Murray *et al.*, 2011; Sanders & Kent, 2011). Indeed, *P. neurophilia* spores are capable of surviving a 25-50 ppm bleaching treatment for 10 minutes, making it difficult to development efficient prevention procedures. The proposed *P. neurophilia* transmission methods include horizontal contamination by ingestion of spores found in aquarium water or fish carcasses, as well as vertical contamination by maternal transmission (Sanders *et al.*, 2013). Interestingly, recent studies have suggested a higher prevalence of infection in male zebrafish when compared to female individuals under similar conditions (Chow *et al.*, 2015). A similar pattern is observed for the count of parasite clusters found in zebrafish, severe myelitis, menixitis and encephalitis (Spagnoli *et al.*, 2015b).

Furthermore, studies have revealed the capacity of this microsporidian to infect nearly all body tissues of the zebrafish but to be preferentially found in neural tissue such as the nerve roots, spinal cord ventral white matter and the hindbrain (Sanders *et al.*, 2014). Once infected, the most common physical observed symptoms in the zebrafish are the presence of spinal deformities (Supplementary figure 9) and emaciation (Matthews *et al.*, 2001). Additional symptoms include encephalitis, myelitis, myositis, granulomatous inflammation and parasite clusters (Murray *et al.*, 2011; Spagnoli *et al.*, 2015b).

## Research Justification and Goals

The present thesis aims to acquire greater insights into the basic biology of the zebrafish parasite, *P. neurophilia*. Despite the growing prevalence of this pathogen in zebrafish research facilities, there are currently no or poor methods of treatment and prevention against microsporidiosis caused by *P. neurophilia* in zebrafish, which could potentially be the result of restricted information on the parasite biology. To date, our current knowledge of *P. neurophilia* is limited to microscopic analyses of stages of its lifecycle and studies on the physical, psychological and behavior impact of this parasite on *D. rerio*. Surprisingly, there are currently no genome data of *P. neurophilia* on any publically available databases. The goal of this thesis research project was to obtain additional knowledge of *P. neurophilia* biology by analyzing the content and structure of its genome. To this end, we have successfully acquired and assembled the first genome draft of *P. neurophilia*.

# Chapter 2 - Genome analysis of Pseudoloma neurophilia: a microsporidian parasite of Zebrafish (Danio rerio)

Steve Ndikumana [1], Adrian Pelin [1], Alex Williot [1], Justin L. Sanders [2], Michael Kent [2], and Nicolas Corradi [1*]

[1] Center for Advanced Research in Environment Genomic, Department of Biology, University of Ottawa, ON, Canada.

[2] Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA

[*] To whom correspondence should be addressed. E-mail: ncorradi@uottawa.ca

**Data deposition:** *Pseudoloma neurophilia* MK1 (LGUB00000000.1)

**Comments:**
This work is being submitted for publication in *Journal of Eukaryotic Microbiology*. My contributions to this work include all aspects of this study. Contributions from other authors:

- AP helped with preliminary sequence analyses and DNA extraction.
- AW helped with preliminary analyses of TEs.
- JLS and MK provided the spores of *P. neurophilia* from infected zebrafish.

## Abstract

Microsporidia are highly successful parasites that infect virtually all known animal lineages, including the model *Danio rerio* (zebrafish). The widespread use of this aquatic model for biomedical research has resulted in an unexpected increase in infections from the microsporidium *Pseudoloma neurophilia*, which can lead to significant physical, behavioral and immunological modifications resulting in non-protocol variation during experimental procedures. Here, we seek to obtain insights into the biology of *P. neurophilia* by investigating its genome content, which was obtained using the Miseq technology and paired-end Illumina sequencing. We found that the genome of *P. neurophilia* is phylogenetically and genetically related to other fish-microsporidians, but features unique to this intracellular parasite are also found. Its small 5.25Mb genome includes 1139 unique open reading frames and an unusually high amount of transposable elements. Investigations of intragenomic diversity also provided strong indications that the mononucleate nucleus of this species is diploid. Overall, our study provides insights into the dynamics of microsporidian genomes, and represents a solid sequence reference to be used in future studies of host-parasite interactions using the zebrafish *Danio rerio* and *P. neurophilia* as a model.

**Keywords:** Microsporidia, Transposable element, Zebrafish, host-parasite interactions, Parasite, Genomics, Genetics, RNAi

## Introduction

Microsporidia are highly successful obligate intracellular parasites that infect many ecologically and economically important terrestrial animal lineages, such as honey bees (Higes *et al.*, 2013), humans (Lobo *et al.*, 2012), and aquatic research models like the zebrafish (Matthews *et al.*, 2001; Sanders *et al.*, 2012; Sanders & Kent, 2014; Spagnoli *et al.*, 2015a). These parasites have been proposed to propagate *via* both vertical and horizontal modes of transmission (Ardila-Garcia & Fast, 2012; Izquierdo *et al.*, 2011; Sanders *et al.*, 2013), and have been recently included in the phylum Cryptomycota (James *et al.*, 2013). Microsporidia are best known for their reductive and derived features, which include a genome-less mitochondrion (the mitosome), prokaryote-sized ribosomal RNA (rRNA) genes, and an unconventional Golgi apparatus (Beznoussenko *et al.*, 2007; Vossbrinck *et al.*, 1987; Williams *et al.*, 2002).

Genome sequencing projects of these successful parasites have provided essential insights into their basic biology and infection mechanisms, and have also unraveled the evolutionary dynamics of their genomes (Corradi, 2015). As an adaptation to their obligate parasitic life cycle, these have undergone massive reductions in their genome content, particularly for genes involved in nucleic and amino acids metabolism (*de novo* biosynthesis of amino acids and nucleotides and the tricarboxylic acids cycle pathways (Corradi & Slamovits, 2011; Higes *et al.*, 2013)). In some species, intergenic regions and even proteins have been shortened in size, resulting at the extreme in genomes that are not only gene-poor but also extremely streamlined (Corradi *et al.*, 2010; Corradi & Slamovits, 2011). Evidently, the miniaturization of many cellular activities has rendered all microsporidia very dependent on their hosts, even for energy

production (Keeling *et al.*, 2010). Although the biochemical repertoire of known microsporidian species is always reduced, their genome sizes can vary substantially within the group, ranging from only 2.3 (*Encephalitozoon intestinalis*) to a known maximum of 51.3 Mb (*Edhazardia aedis)*(Corradi *et al.*, 2010; Desjardins *et al.*, 2015). These drastic variations in genome size have been correlated with gene density and the number of transposable elements (TE) - *i.e.* larger genomes harbor more TEs, and much larger intergenic regions. To date, the genomes of 20 known species have been sequenced. This wealth of data has enabled a broad understanding of the microsporidian's biology, but as more species get sequenced it becomes clear that the definition of a typical microsporidian genome often differs from the original description.

The present paper aims to obtain first-hand knowledge of a notorious parasite of the zebrafish (*Danio rerio*); namely *Pseudoloma neurophilia*. This parasite mostly targets the central nervous system and peripheral nerves of its host (Kent *et al.*, 2012), and significant increases in infection cases caused by this particular microsporidian species have been described in recent years (Murray *et al.*, 2011; Sanders *et al.*, 2012). Such infections have been detected in approximately 75% of the facilities of the Zebrafish International Resource Centre and their presence has been demonstrated to result in significant physical, behavioral and neurological (Sisson *et al.*, 2006) modifications; all of which are likely to result in non-protocol sources of variation during experimental procedures. Here, we show that the genome size of *P. neurophilia* is on the lower end compared to relative species in the group, and we describe several of its atypical features. These include an unusual expansion of TEs, and the presence of large numbers of inserts located in many protein-coding genes that are otherwise highly conserved in sequence among eukaryotes (including microsporidian species). Besides

improving our knowledge of microsporidian biology, this sequence data will also be helpful for future analyses of host-parasite interactions in the zebrafish, particularly gene expression studies based on next generation sequencing, such as RNA-seq.

## Results and Discussion

### *A typical, small and likely diploid microsporidian genome*

A total of 19.1 nanograms of DNA was extracted from spores isolated from the neural system of infected *Danio rerio* individuals (Fig. 1). This small amount of DNA was subjected to paired-end Illumina sequencing using the Miseq platform, resulting in 7,230,699 paired-end reads with an average length of 250 bp. Reads were assembled using Spades v3.0.0, resulting in an assembly of 31,043 contigs. Because of the obligate intracellular lifestyle of *P. neurophilia*, a large number of contigs were of obvious contaminant origin. Plotting coverage against GC content revealed a large number of highly covered contigs, as well as many others with low coverage and high GC content (Supplementary figure S1-2). The vast majority of contaminant contigs harbored a lower coverage and GC content higher than what is typically observed in microsporidia (average GC: 38.16, (Selman *et al.*, 2013)). Interestingly, low covered contaminants of microsporidian origin were also found (*e.g. Encephalitozoon cuniculi*. *E. romaleae*, *Nosema ceranae,*). Although it is possible that these represent bona-fide microsporidian infection that are less frequent (much lower coverage), they most probably originate from low-amounts of contaminant DNA from sequencing projects performed in parallel in our lab. Non-microsporidian contaminations with high coverage included unsurprisingly animal mitochondrial genomes (e.g. the *D. rerio* mtDNA), as well as nuclear sequences affiliated with aquatic animals (the water snail *Lottia gigantea*, *Hydra vulgaris*, *Branchiostoma sp*).

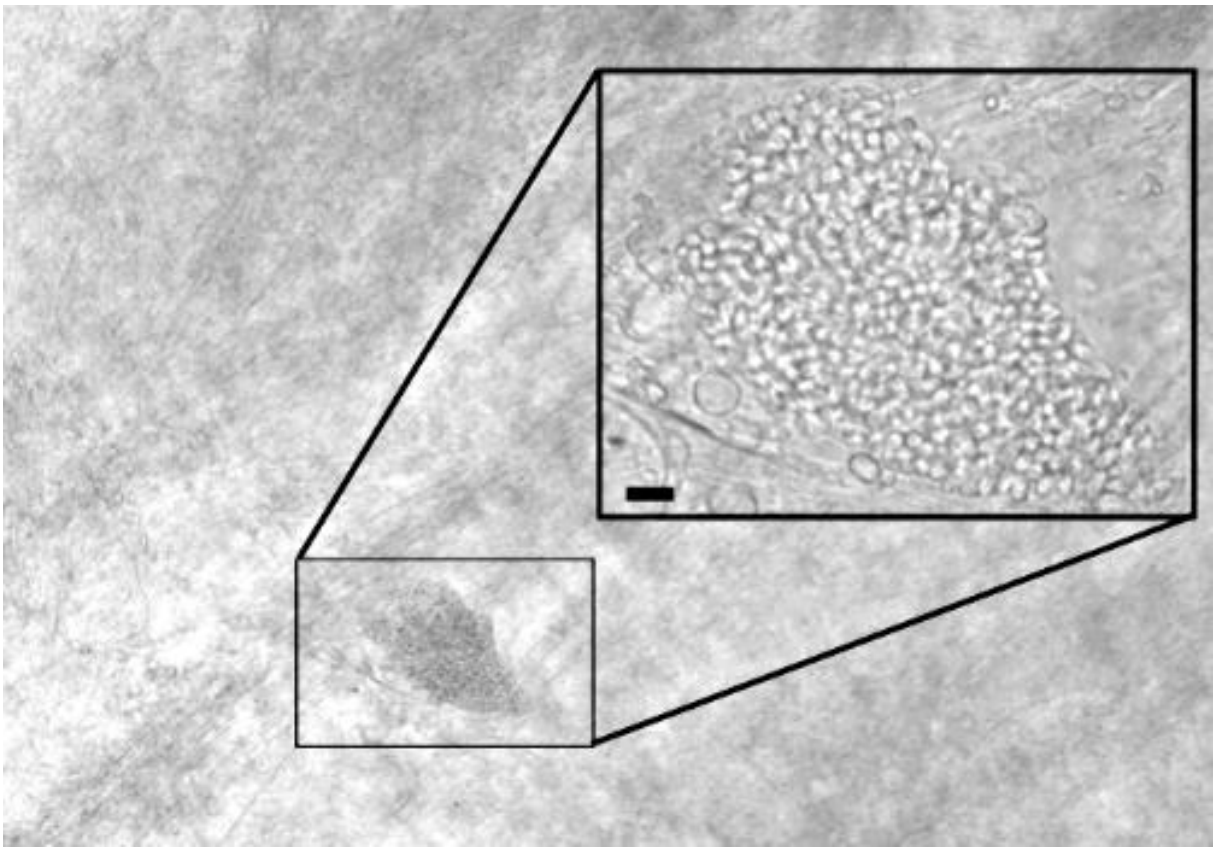


**Figure 2.2.** Wet mount of an aggregate of *Pseudoloma neurophilia* spores in the hindbrain of an infected adult zebrafish. *P. neurophilia* generally forms small pseudocysts in the neural tissue of the posterior brain and spinal cord of adult zebrafish. Pseudocysts are often focal and contain several small spores (inset). Bar = 10 µm.

Highly covered contigs with expected GC content were manually inspected for their microsporidian origin and were chosen as representative of the *P. neurophilia* genome. The total genome assembly of *P. neurophilia* is 5.25 Mb (Table 2.1). The genome is gene-dense, with 3633 predicted genes and coding regions encompassing 51% of the genome. Comparing Open Reading Frames (ORFs) counts and biological categories with publicly available microsporidian genome reveals that the *P. neurophilia* contigs harbor a typical microsporidian protein repertoire, and demonstrates that most of the genome space has been sampled (Fig. 2, Supplementary table S1). Interestingly, a third of the predicted ORFs (1139) have no known homologues in publicly available database. Protein domain analyses identified 196 proteins with signal peptide cleavage sites, indicating that these are likely secreted (Supplementary table S2) (Petersen *et al.*, 2011), more than half of which (111) are in *Pseudoloma*-specific genes. These unique *P. neurophilia* homologues may represent candidate effectors (Campbell *et al.*, 2013), whose function could be tested in the future using this host-parasite system (*P. neurophilia- Danio rerio*).

**Table 2.1.** Genome statistics of *Pseudoloma neurophilia*

| **Genome statistics of *P. neurophilia*** | |
|---|---|
| Haploid genome size (Mb) | 5.3 |
| Contigs | 1603 |
| Average genome coverage | 138X |
| GC content (%) | 29.6 |
| GC content coding (%) | 31.4 |
| Coding regions (Mb) | 2.9 |
| Genes | 3633 |
| SNPs frequency (SNPs/kb) | 1.98 |
| Candidate effectors* | 111 |

*Defined as predicted secreted proteins of unknown function

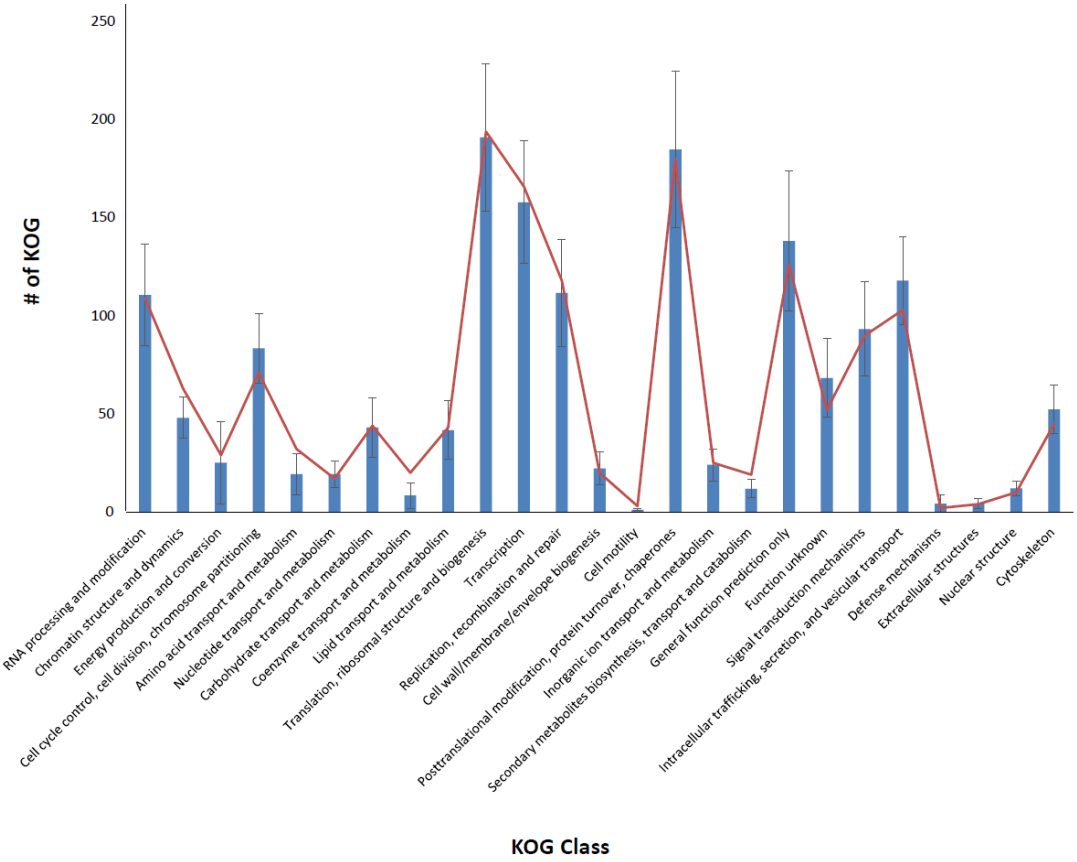



**Figure 2.3.** Mean KOG Class distribution among microsporidian genomes (blue; with standard deviation) with observed KOG domain in *Pseudoloma neurophilia* (red)

To date, genome analyses of microsporidia have suggested that the genomes of these intracellular parasites are mainly diploid, although evidence of polyploidy has also been observed in species with diplokaryons (two unfused nuclei bound together by anendomembrane) (Cuomo *et al.*, 2012; Pelin *et al.*, 2015). Here, we tested the ploidy levels of *P. neurophilia* by plotting the allele frequency of single nucleotide polymorphisms (SNPs) along all contigs. These analyses identified 7,246 SNPs with a 0.5 frequency which is strongly indicative of diploidy (Fig. 3-A), and were independently confirmed using a K-mer coverage distribution (Fig. 3-B). The remaining 3,162 SNPs were found to have frequencies between 0.1-0.4. Sanger sequencing of selected regions confirmed the heterozygous nature of SNPs with a 0.5 frequency and the demonstrated that lower sequencing variants could represent a mixture of sequencing errors and individual polymorphisms (Supplementary figure S3). The overall SNP frequency in *P. neurophilia* is 1.98 SNP/Kb, in line to what has been reported from other microsporidian species (Supplementary table S3; (Cuomo *et al.*, 2012; Desjardins *et al.*, 2015; James *et al.*, 2013; Pelin *et al.*, 2015; Selman *et al.*, 2013)). With 58.85% of SNPs occurring in coding regions of this gene-dense genome, the SNPs distribution also reflects what is observed in other microsporidian species (Cuomo *et al.*, 2012).

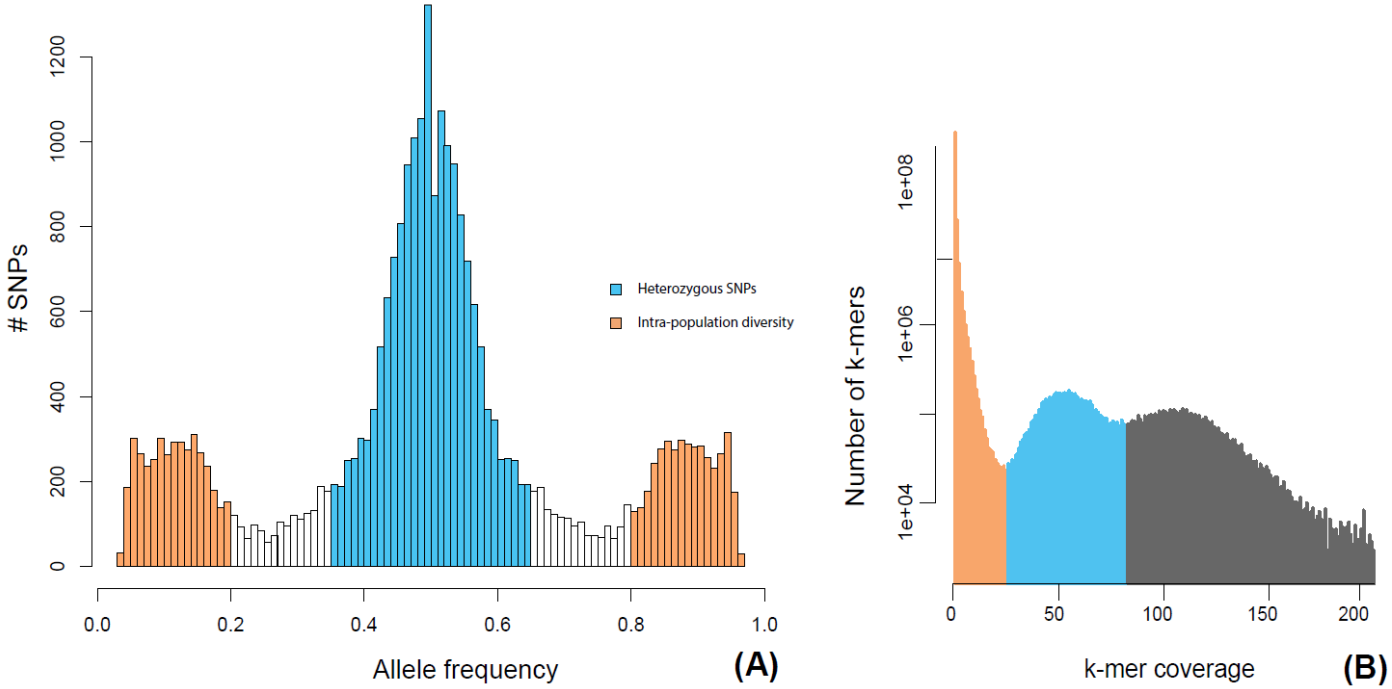



**Figure 2.4.** A. SNP allele frequency distribution within the genome of *Pseudoloma neurophilia*. B. K-mer calculated for k= 71. Proposed heterozygous variants (blue) and homozygous variants (orange) are highlighted.

Finally, our genome analyses also revealed regions with loss of heterozygosity (LOH, Fig. 4). LOH has been hypothesized to play a role in pathogenicity of distant pathogens, such as the yeast *Candida albicans* (Diogo *et al.*, 2009) or the stramenopile *Phytophtora capsici* (Lamour *et al.*, 2012). In these species, LOH has been proposed to facilitate adaptation to external negative stimuli (new host or chemical modification) through allele fixation, but its effect on microsporidia is unknown. In *P. neurophilia*, these regions affected by LOH can attain a maximum of 52 kb and harbor allele frequencies 30 times lower than the genome average. Genes affected by LOH have a variety of functions, and include 7 hypothetical highly similar ORFs with secretion signals that may have originated from successive duplication events (Supplementary figure S4, Supplementary table S4-S5)

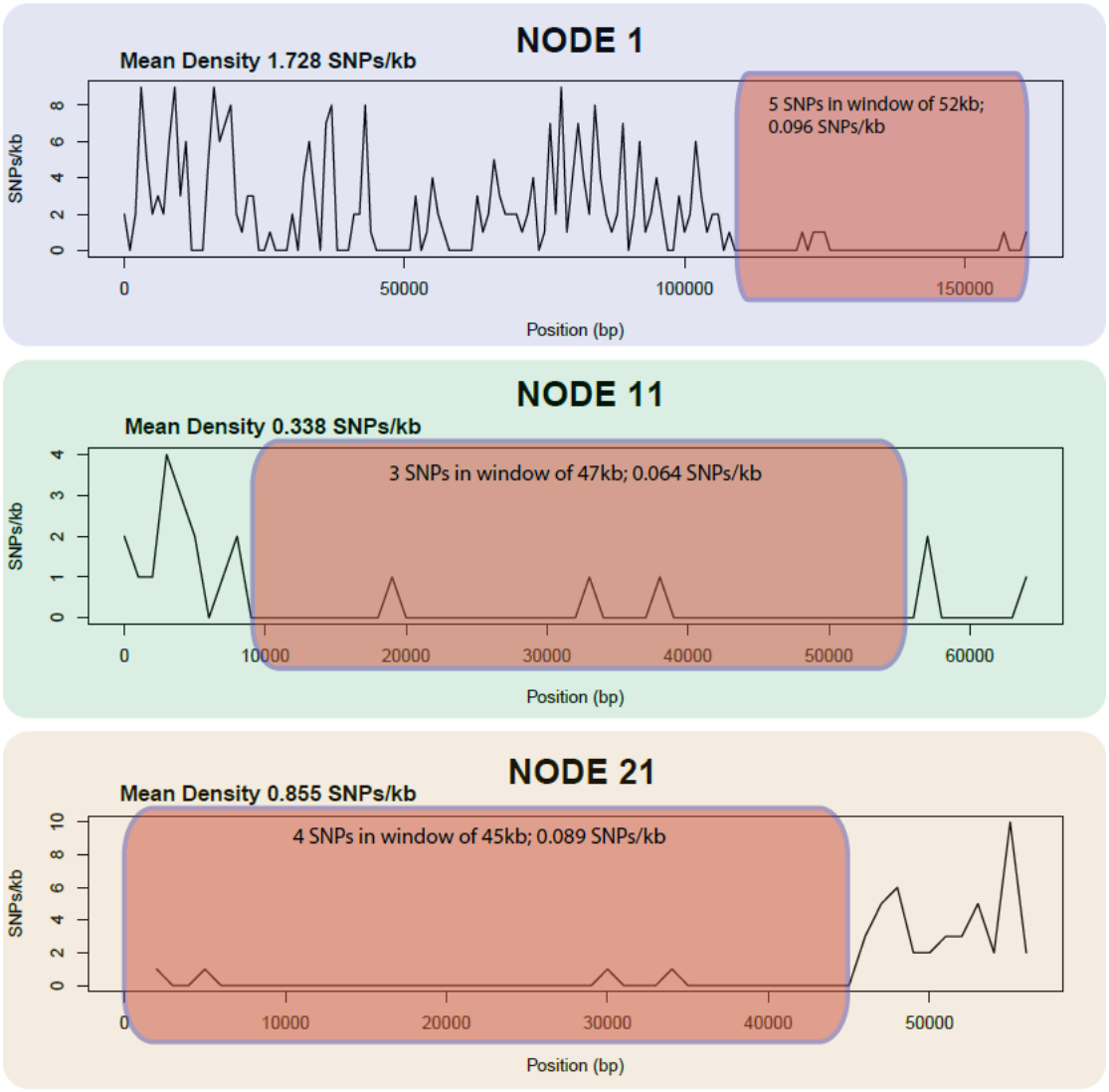
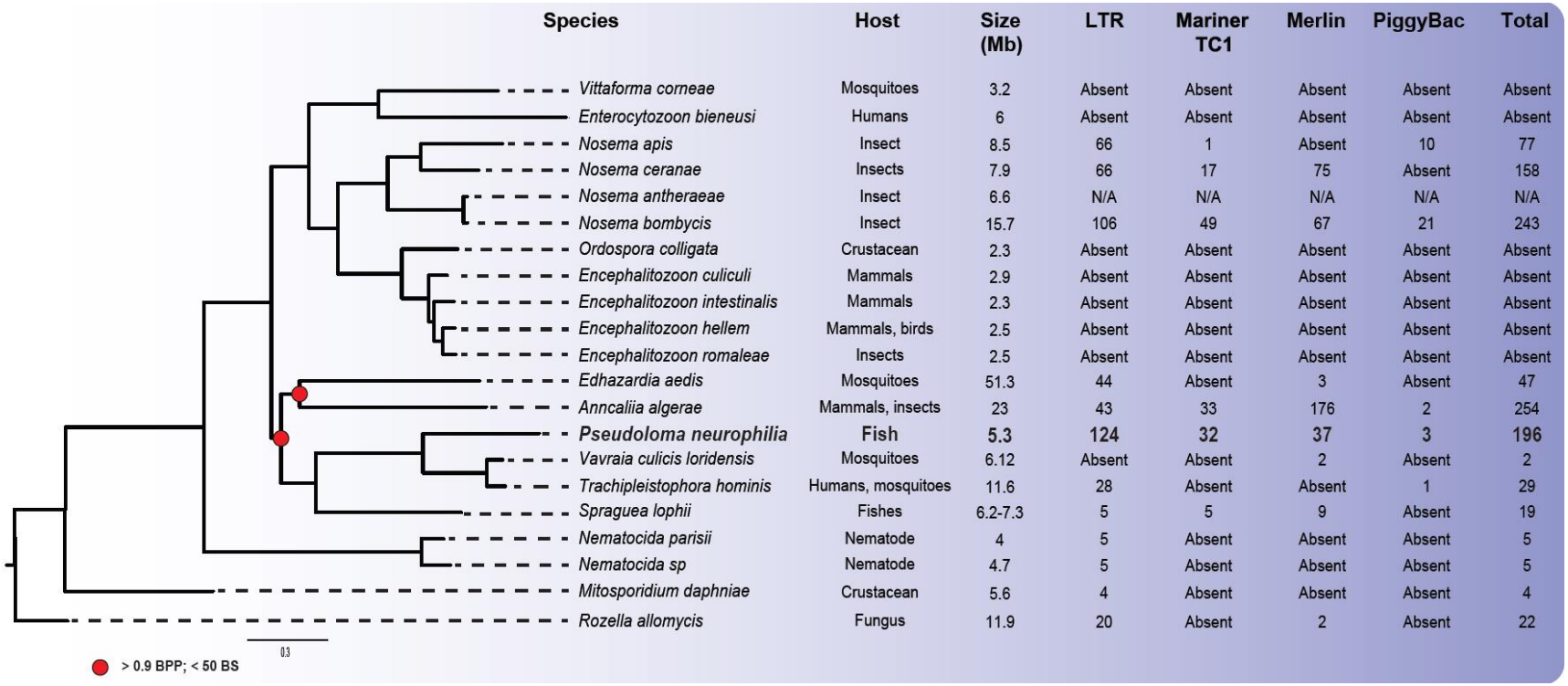



**Figure 2.5.** Large contigs harboring regions of LOH (red). SNP count within region is shown.

### *Unique and lineage specific genes in P. neurophilia*

Comparing gene content between *P. neurophilia* and 21 publicly available microsporidian genomes (Supplementary table S6) allowed us to identify 62 single copy orthologues shared by all sequenced species. These were used to reconstruct a reliable phylogeny of the group, and showed that *P. neurophilia* clusters within a supported clade that contains fish/mosquitoes microsporidians (Fig. 5). The tree topology is consistent with published phylogenies, and demonstrates that *P. neurophilia* is closely related to *Trachipleistophora hominis*, *Vavraia culicis loridensis*, *Spraguea lophii* (Vossbrinck & Debrunner-Vossbrinck, 2005). Consistent with this, our analysis revealed that the genome of *P. neurophilia* shares a substantially higher number of gene families with its most closely related species (*T. hominis* and *V. culicis)* than with any other member of the group (n=1256 out of 1558 orthologous groups present in *P. neurophilia*). Notable families shared by these species but absent in distant lineages include some with similarities to metabolically relevant proteins, such as the N-arginine dibasic convertase NRD1 and the glutaredoxin protein, as well as several ORFs with unknown function (Supplementary table S7). All species in this subclade also share the machinery required for the RNA interference pathway, as well as several putative effector proteins (Campbell *et al.*, 2013; Heinz *et al.*, 2012; Paldi *et al.*, 2010). One example of effector proteins includes the ricin b lectin proteins, which were previously proposed to play a role in the pathogenicity of the microsporidia by facilitating its adhesion on the host cell during infection (Campbell *et al.*, 2013). Consistent with what has been reported in the microsporidium S. *lophii*, signal peptide cleavage sites are also present in *P. neurophilia*'s ricin b lectin genes**,** suggesting that these proteins are probably also secreted into the host cell and play a role in the host-parasite interaction.




| Species | Host | Size (Mb) | LTR | Mariner TC1 | Merlin | PiggyBac | Total |
|---|---|---|---|---|---|---|---|
| *Vittaforma corneae* | Mosquitoes | 3.2 | Absent | Absent | Absent | Absent | Absent |
| *Enterocytozoon bieneusi* | Humans | 6 | Absent | Absent | Absent | Absent | Absent |
| *Nosema apis* | Insect | 8.5 | 66 | 1 | Absent | 10 | 77 |
| *Nosema ceranae* | Insects | 7.9 | 66 | 17 | 75 | Absent | 158 |
| *Nosema antheraeae* | Insect | 6.6 | N/A | N/A | N/A | N/A | N/A |
| *Nosema bombycis* | Insect | 15.7 | 106 | 49 | 67 | 21 | 243 |
| *Ordospora colligata* | Crustacean | 2.3 | Absent | Absent | Absent | Absent | Absent |
| *Encephalitozoon culiculi* | Mammals | 2.9 | Absent | Absent | Absent | Absent | Absent |
| *Encephalitozoon intestinalis* | Mammals | 2.3 | Absent | Absent | Absent | Absent | Absent |
| *Encephalitozoon hellem* | Mammals, birds | 2.5 | Absent | Absent | Absent | Absent | Absent |
| *Encephalitozoon romaleae* | Insects | 2.5 | Absent | Absent | Absent | Absent | Absent |
| *Edhazardia aedis* | Mosquitoes | 51.3 | 44 | Absent | 3 | Absent | 47 |
| *Anncaliia algerae* | Mammals, insects | 23 | 43 | 33 | 176 | 2 | 254 |
| ***Pseudoloma neurophilia*** | **Fish** | **5.3** | **124** | **32** | **37** | **3** | **196** |
| *Vavraia culicis loridensis* | Mosquitoes | 6.12 | Absent | Absent | 2 | Absent | 2 |
| *Trachipleistophora hominis* | Humans, mosquitoes | 11.6 | 28 | Absent | Absent | 1 | 29 |
| *Spraguea lophii* | Fishes | 6.2-7.3 | 5 | 5 | 9 | Absent | 19 |
| *Nematocida parisii* | Nematode | 4 | 5 | Absent | Absent | Absent | 5 |
| *Nematocida sp* | Nematode | 4.7 | 5 | Absent | Absent | Absent | 5 |
| *Mitosporidium daphniae* | Crustacean | 5.6 | 4 | Absent | Absent | Absent | 4 |
| *Rozella allomycis* | Fungus | 11.9 | 20 | Absent | 2 | Absent | 22 |

**Figure 2.6.** Consensus phylogenetic tree of microsporidia (maximum likelihood and Bayesian approach) and TEs distribution. Support values for nodes are omitted when over 85 bootstraps and 1 Bayesian posterior probability for Maximum Likelihood and Bayesian algorithms respectively

The conservation of synteny among *P. neurophilia* and closely related species is also very high (Fig. 6). At the extreme, up to 13 genes encompassing 37 kb can be found in highly conserved order among *P. neurophilia*, *T. hominis* and *V. culicis*. This confirms that the maintenance of gene order is common in microsporidia, and is in sharp contrast with their notoriously elevated sequence divergence (Nakjang *et al.*, 2013). As in other microsporidia, several meiosis-related genes appear to be absent in *P. neurophilia*, another indication that this universal sexual process has been heavily streamlined in these parasites. When compared to its closely related species, *P. neurophilia* is the microsporidian species where the least meiosis-related genes (MRG) have been found (Supplementary table 8), and targeted searches for additional meiosis-specific motifs within sequence reads and contigs has been unsuccessful, so the slightly reduced MRG numbers are unlikely to result from missing genomic regions. To date, only three microsporidian assemblies contain less MRGs than *P. neurophilia* (*Nosema apis, Mitosporidium daphniae* and *Enterocytozoon bieneusi*).

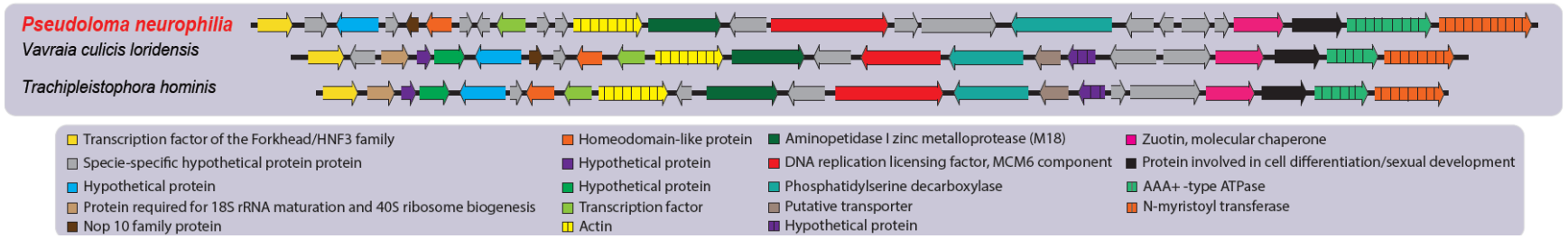



**Figure 2.7.** Schematic representation of gene order conservation of thirteen proteins shared between *P. neurophilia* and closely related microsporidians.

### *RNA interference and in-frame inserts in conserved protein encoding genes*

Interestingly, *P. neurophilia* harbors homologues of three key proteins involved in RNA-induced silencing complex. These include the argonaute, the dicer protein homologues and the RNA-dependent RNA polymerase. These homologues are also found in closely related species and other fungi (Campbell *et al.*, 2013; Dang *et al.*, 2011; Heinz *et al.*, 2012) and have been hypothesised to be involved in transposon silencing through RNA-interference (Heinz *et al.*, 2012; Obbard *et al.*, 2009). The presence of these orthologues in almost all microsporidian genera (Table 2.2) suggests that such defense mechanism may be common to most microsporidians. While components of the RNA-induced silencing complex have not been demonstrated to be directly involved in gene regulation or affect transposons in microsporidia, their role in RNA interference pathway has been recently proposed in the bee pathogen *Nosema cerenae* (Paldi *et al.*, 2010).

**Table 2.2.** Distribution of the RNA interference key components within the microsporidian clade.

| Micropsoridian | Argonaute | Dicer | RNA-dependent RNA polymerase |
|---|---|---|---|
| *Pseudoloma neurophilia* | ✓ | ✓ | ✓ |
| *Vavraia culics loridensis* | ✓ | ✓ | ✓ |
| *Trachipleistophora homins* | ✓ | ✓ | ✓ |
| *Spraguea lophii* | ✓ | ✓ | ✓ |
| *Anncaliia algerae* | ✓ | ✓ | ✓ |
| *Edhazardia aedis* | ✓ | ✓ | ✓ |
| *Nosema apis* | ✓ | ✓ | ✓ |
| *Nosema ceranae* | ✓ | ✓ | ✓ |
| *Vittaforma corneae* | ✓ | ✓ | ✓ |
| *Rozella allomycis* | ✓ | - | ✓ |
| *Mitosporidium daphniae* | ✓ | ✓ | - |
| *Nosema bombycis* | ✓ | ✓ | - |
| *Nosema antheraeae* | - | ✓ | - |
| *Nematocida parisii* | - | - | - |
| *Nematocida sp.* | - | - | - |
| *Encephalitozoon culiculi* | - | - | - |
| *Encephalitozoon hellem* | - | - | - |
| *Encephalitozoon intestinalis* | - | - | - |
| *Encephalitozoon romaleae* | - | - | - |
| *Enterocytozoon bieneusi* | - | - | - |
| *Ordospora colligata* | - | - | - |

Surprisingly, we found that the argonaute protein harbors a 99 bp sequence insert within a region that is essential for its function in model organisms (the PIWI domain (Höck & Meister, 2008)) and that is conserved among microsporidian orthologues (Fig. 7-8). Genome explorations exposed many similar inserts elsewhere in the genome. In all cases, these inserts are found in-frame (*i.e.* none lead to stop codons), and do not demonstrate similarities with known introns (*i.e.* no obvious splicing sites, always in frame). These inserts are probably analogous to inserts first reported in several genes of *Hamiltosporidium tvaerminnensis* (Corradi *et al.*, 2009); a distant microsporidium that infects the crustacean *Daphnia magna*. In *P. neurophilia*, 30 conserved protein-encoding genes have been observed to harbor inserts with variable lengths (Fig. 8, Supplementary table S9). These insertions occur in amino-acid sequences that are otherwise conserved among microsporidia and other eukaryotes, so there is a possibility that these affect the function of these proteins. Searching for similar inserts in other microsporidian lineages reveals that these intriguing genomic signatures are ubiquitous in this group of parasites, although their sequence and location varies among species (Supplementary table S9). Indeed, protein encoding genes affected by these insertions are always different, with no clear functional pattern emerging from our analyses.

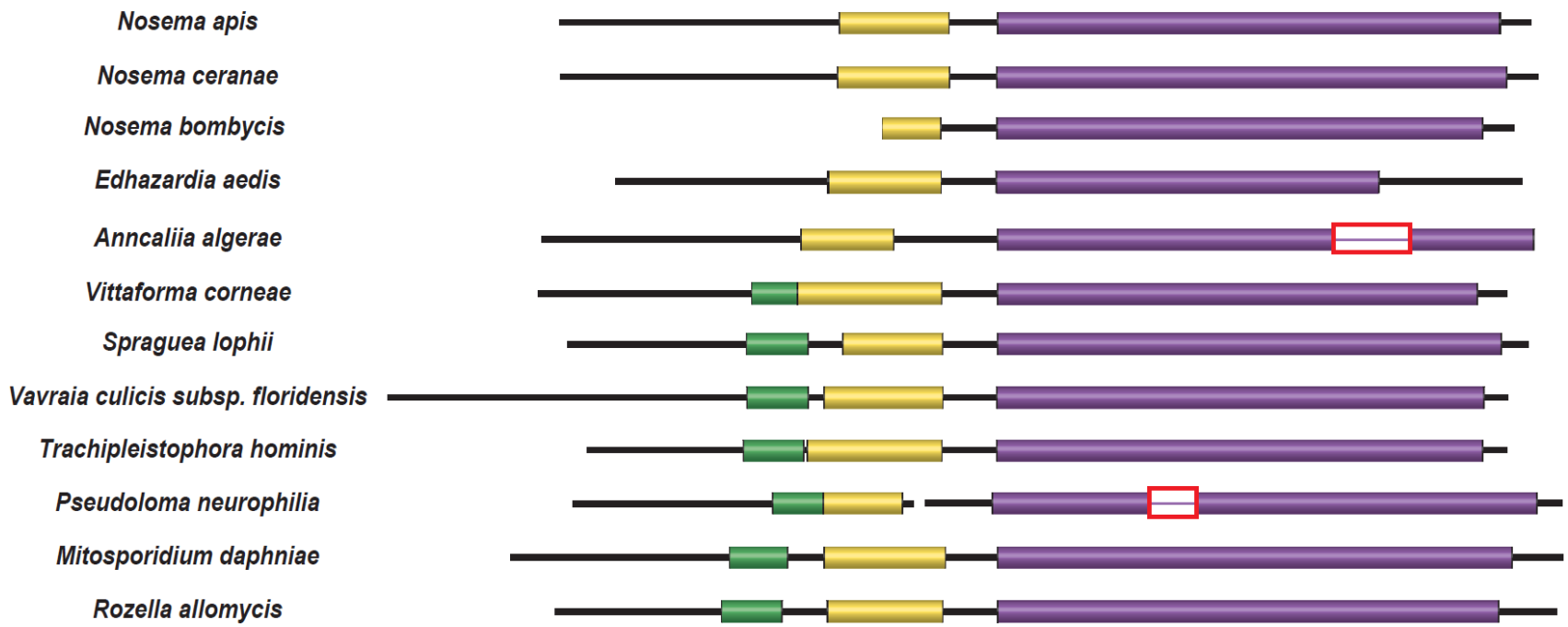



**Figure 2.8.** Schematic representation of conserved domains of the argonaute protein in Microsporidia; the PIWI-like domain (purple) and their respective inserts (red); the PAZ argonaute-like domain (yellow); and domain with unknown function (green) are shown.

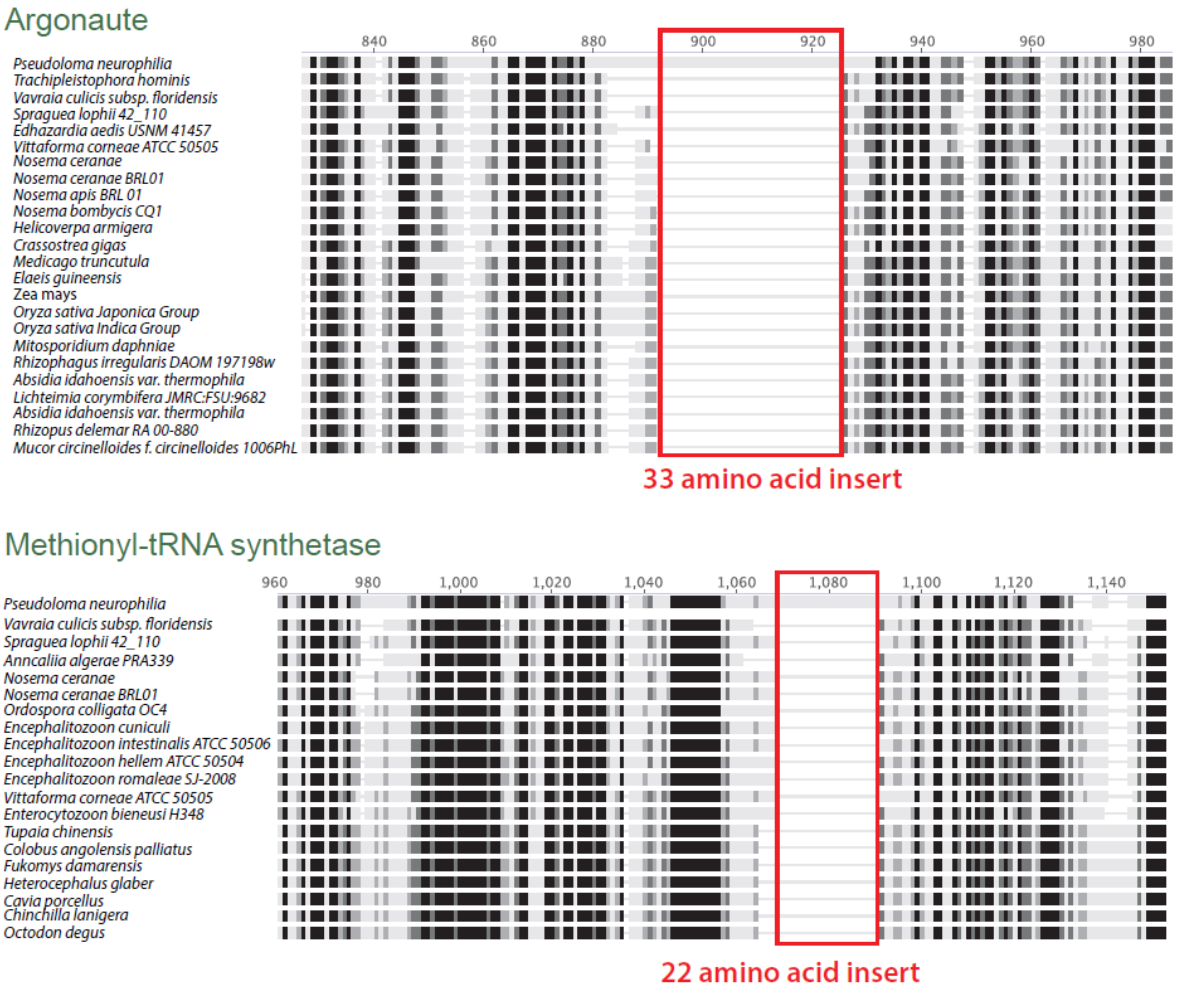



**Figure 2.9.** Protein alignments demonstrating the presence of large inserts (red) in regions typically conserved in microsporidia and other phylum.

***Transposable element family expansion in a reduced genome***

Sequence homology searches showed that the putative biochemical repertoire of *P. neurophilia* is very similar to that of other sequenced microsporidian species (Fig. 2, Supplementary table S1). However, multiple gene family expansions within *P. neurophilia* of notable interest are also found, the most compelling including transposable families. Indeed, the genome of *P. neurophilia* harbors a total of 196 ORFs whose sequences encode for proteins with conserved microsporidian transposable element domains, including DNA and retrotransposons (Parisot *et al.*, 2014b)**.** The number of TEs in *P. neurophilia* is much more elevated than in closely related species with which it shares high sequence and gene content similarities (Fig. 5).

While TEs are numerous in a few select microsporidia with very large genomes (Parisot *et al.*, 2014b), the presence of TEs are otherwise rare in many members of the group; including the basal *Nematocida sp.* and the more derived *Encephalitozoon sp*. (Fig. 5). In this context, *P. neurophilia* appears to be quite notable, as approximately 3% of its genome sequence (or 196 ORFs) are occupied by TEs. Most of these transposons have homologs in distant members of the group, and none of these appear to have originated from horizontal gene transfer (Parisot *et al.*, 2014b; Watson *et al.*, 2015). In fact, there is evidence that some TE families may have been fuelled by numerous, and probably recent, duplication events, as highlighted by the high sequence similarity of several members of this family (Supplementary figure S5-S8).

The very high numbers of TEs in *P. neurophilia* (compared to most species in the group) almost certainly affect the biology of this parasite. Indeed, TEs possess the ability to relocate and/or duplicate along genome and can have potential negative repercussion on the organism by creating pseudogenes (Biémont, 2010b), so ensuring that their number do not inflate beyond control is an essential defense mechanism for any organism. One gene involved in gene silencing

of TEs is the Argonaute gene (McCue *et al.*, 2015; Thomson & Lin, 2009), but the abovementioned insertion in this gene does not seem to be associated with the expansion of TE in *P. neurophilia*, as TEs are also abundant in species with more conventional Argonaute genes (Figure 5,7). TE proliferation is also linked with the existence of sexual reproduction (Arkhipova & Meselson, 2000; Lee *et al.*, 2014; Wright & Finnegan, 2001), a notion that is supported by the presence of diploidy and MRGs in *P. neurophilia*.

## Conclusions

Acquiring a genome reference for P. neurophilia is a first step in understanding its biology. These sequences also pave the road for future analyses of host-parasite molecular interactions, using RNA-seq or other NGS tools. We demonstrated that this relevant parasite of zebrafish harbors a small haploid genome relative to most members of the group, and that it is closely related in both content and structure with those of *T. hominis* and *V. culicis*. Despite these similarities, a third of the P. neurophilia genome is specific to this species, with many of these unique genes harboring signal peptide cleavage sites that indicate their putative secretion and involvement in host manipulation (e.g. effectors).

Our analyses also support *P. neurophilia* as a diploid organism that harbors genomic signatures that are linked with both sexual (ploidy, TE abundance, MRG) and clonal modes of evolution (LOH). Comparative analysis between P. neurophilia and other sequenced microsporidia demonstrated the presence of an unusual high amount of TEs inside this small genome. This is surprising as all sequenced microsporidian genomes with similar sizes harbor no, or low amounts of TE. TEs abundance does not seem to be affected by the presence of key components of the RNAi pathways (Dicer and Argonaute proteins) that silence the expression of TEs and

subsequently and slow-down their expansion (Obbard *et al.*, 2009). This machinery has been demonstrated to be functioning in the bee pathogen N. cerenae, and future studies on P. neurophilia will hopefully provide more insight into the regulatory role of RNA interference in other microsporidians. Lastly, the discovery of in-frame inserts into coding sequence of many genes that are usually highly conserved, and their ubiquitous presence in this group may also be of primary relevance for understanding these parasites. In particular, their universal presence in microsporidia, coupled with their localization in many homologues with known cellular function, suggest that these may have important effects on the biology of these parasites. Hopefully, recent advances in the development of genetic techniques in intracellular parasites may soon be applicable to study microsporidia (Vinayak *et al.*, 2015). Such techniques will be essential to determine the function of any microsporidia genomic region, including those atypical inserts.

## Acknowledgment

The authors declare no competing interest. Nicolas Corradi is a Fellow of the Canadian Institute for Advanced Research. This work was supported by the Discovery program from the Natural Sciences and Engineering Research Council of Canada (NSERC-Discovery) and an Early Researcher Award from the Ontario Ministry of Research and Innovation.

## Material and Method

### *Sample collection and DNA extraction.*

Spores of *Pseudoloma neurophilia* strain MK1were extracted from 50 infected adult *Danio rerio* at Oregon State University (Fig 1, Supplementary figure S9). Fish were euthanized using an overdose of MS222, brains and spinal cords were removed and placed in 5 ml of deionized water with antibiotics (pen-strep). This solution was passed through progressively smaller-gauge needles using a 5 ml syringe. An additional 5 ml of deionized water and antibiotics was added and the mixture was vortexed and allowed to sit overnight at room temperature. The following day, the mixture was again vortexed and passed through a 40 µm cell strainer into a 50 ml conical tube. Water and antibiotics were added to bring the total volume to 20 ml then 20 ml of filter-sterilized Percoll was added. The mixture was vortexed and centrifuged at 1200 g for 1 hr. The pellet was recovered and placed in a 1.5 ml tube where it was washed two times with 1% SDS followed by two washes in PBS. The spores were resuspended in 200 µl of TE buffer and quantified using a hemocytometer.

DNA was extracted from approximately 1.5 x $10^6$ spores using the MasterPure DNA purification kit (Epicentre). Spores were first pelleted by centrifugation; resuspended in 300 µl of Tissue & Cell lysis buffer (Epicentre) with 1µl of Proteinase K and vortexed thoroughly. Glass beads (150-212 microns) were added to the solution and incubated at 65$^o$C for 15 minutes and shaken at 2500 rpm for 30 seconds every 5 minutes. Following the incubation, the sample was pelleted by centrifugation; cooled to 37$^o$C and incubated for 30 min at the same temperature after the addition of 1µl of 5µg/µl of RNAse A (Epicentre). The sample was then put on ice for 5 minute and 150 µl of MPC Protein Precipitation Reagent (Epicentre) were added. The new solution was vortex vigorously for 10 seconds and the debris were pelleted by centrifugation for 15 minutes at

≥10,000 × g at 4ºC. Following the centrifugation, the supernatant was transferred to a clean microcentrifuge tube and 500 µl of isopropanol were added to the sample and incubated at -20ºC for 30 minutes. The incubated sample was then centrifuged at 4ºC for 20 min at ≥10,000 × g. The pelleted DNA was rinsed twice with cold 70% ethanol and resuspended in Tris-EDTA buffer. DNA extraction resulted in a total of 29 ng of DNA.

***DNA sequencing and de novo assembly.***

Extracted DNA of *P. neurophilia* spores was sent to Illumina sequencing using the MiSeq technology by Fasteris S.A. (Geneva, Switzerland); which resulted in a library of 7,230,699 paired-end reads for a total of 14,461,398 reads of 250 bp length. Paired-end reads were trimmed using Trimal V1.2 (Kofler *et al.*, 2011) to remove adaptors and used to assemble the genome draft using Spades v3.0.0 (Bankevich *et al.*, 2012) with k-mers ranging from 23 to 113 (23,33,43,53,63,73,83,93,103,113); giving 31,043 scaffolds totalling 30 Mb. Resulting contigs were screened for contaminant based on their GC content and average read coverage (Supp. Fig 2). Highly supported contigs (average coverage above 44) were retained and used to perform a BLAST analysis of all ORFs (e-value cut-off of $1xE^{-5}$) against the nr database in order to identify contigs demonstrating obvious microsporidian origin. Here, contigs with at least one ORF demonstrating homology with sequenced microsprodian species were retained. The genome sequences have been deposited at DDBJ/EMBL/GenBank under the accession LGUB00000000.

***Genome annotation.***

Potential open reading frames (ORFs) were predicted using using an in-house script that combines Glimmer's ab initio gene prediction algorithm and with the detection of CCC and

GGG motifs found in close proximity of microsporidian transcription initiation sites (Peyretaillade *et al.*, 2012). The gene function of each ORF was predicted using homology, which we inferred by performing BLAST searches against the NCBI nr database with an e-value cut-off of $1xE^{-10}$. ORFs were annotated using Geneious R8 (Kearse *et al.*, 2012). To further describe the function of *P. Neurophilia*'s predicted genes, a search for secretion signals and subcellular location prediction was performed using SignalP (Petersen *et al.*, 2011) and TargetP (Emanuelsson *et al.*, 2000), respectively. KOG gene enrichments analyses were further conducted to categorize and identify orthologues and paralogs among microsporidian genomes (Wu *et al.*, 2011).

***Genome data used in this study***

The following microsporidian genomes were retrieved: *Anncaliia algerae* PRA109 (AOMV00000000.2), *Edhazardia aedis* USNM 41457 (AFBI00000000.3), *Encephalitozoon cuniculi* GB-M1 (AL391737.2), *Encephalitozoon hellem* ATCC 50504 (CP002713.1), *Encephalitozoon intestinalis* ATCC 50506 (CP001942.1), *Encephalitozoon romaleae* SJ-2008 (CP003518.1), *Enterocytozoon bieneusi* H348 (NZ_ABGB00000000.1), *Mitosporidium daphniae* (JMKJ00000000.1), *Nematocida parisii* ERTm1 (AEFF00000000.2), *Nematocida* sp. 1 ERTm2 (AERB00000000.1), *Nosema antheraeae* YY (http://silkpathdb.swu.edu.cn/silkpathdb/ftpserver, last accessed July 1, 2014), *Nosema apis* BRL 01 (ANPH00000000.1), *Nosema bombycis* CQ1 (ACJZ00000000.1), *Nosema ceranae* BRL01 (NZ_ACOL00000000.1), *Ordospora colligata* OC4 (JOKQ00000000.1), *Spraguea lophii* 42_110 (ATCN00000000.1), *Trachipleistophora hominis* (ANCC00000000.1), *Vavraia culicis* subsp. floridensis (AEUG00000000.1), *Vittaforma corneae* ATCC 50505 (AEYK00000000.1), and *Rozella allomycis* (ATJD00000000.1)

***Gene clustering and phylogenetic tree reconstruction.***

OrthoMCL v.2 (Li *et al.*, 2003) was used to conduct comparative genomic analyses of *P. neurophilia* predicted proteome with 20 other published microsporidian proteomes by identifying orthologous proteins within microsporidian genomes. Families were reconstructed using OrthoMCL with an e-value cut-off of 1e-10. Only families with more than one member were kept for further analyses. In house script was then used to convert the OrthoMCL standard output into a numeric table (Supp Table 4). 62 families were represented by a single member from each genome and were retained for phylogenetic analysis. To this end, members of each family were aligned using Muscle V3.8.31 (Edgar, 2004) and the resulting 62 alignments were concatenated using the bioinformatics software Geneious R9 (Kearse *et al.*, 2012). Concatenated alignments were trimmed using TrimAl V1.2 (Capella-Gutierrez *et al.*, 2009) and the best-fit model for phylogenetic analysis was estimated using ProtTest V3.4 (Darriba *et al.*, 2011) based on the Aikaike Information Criterion. To reconstruct the microsporidian phylogenetic tree, the best phylogenetic model was implemented in MrBayes V3.1.2(Ronquist *et al.*, 2012) and PhyloBayes-MPI V1.5(Lartillot *et al.*, 2009) using posterior probabilities, 100 bootstraps branch support for the Bayesian and Maximum Likelihood analyses respectively.

***Detection of putative transposable elements.***

Transposable elements from P. neurophilia were obtained following BLASTx search against the NCBI nr database to identify candidates with similarities with published TEs (evalue cut-off of 1e-10). Only candidates demonstrating the presence of previously described microsporidian TEs conserved domains (Parisot *et al.*, 2014b) within putative ORF were retained for further analyses. Based on the presence of specific conserved domains, *P. neurophilia's* transposable

elements were classified among 4 major families; LTR retrotransposon, Piggyback, Mariner and Merlin family. For comparison purpose, TEs in other microsporidians were identified using the same methodology. Bayesian phylogenetic analyses were performed using previously described transposable elements (Parisot *et al.*, 2014b) and reconstructed using the previously described phylogenetic tree reconstruction methodology.

***Heterozygosity in P. neurophilia.***

Polymorphic sites were identified by two independent variant callers FreeBayes v0.9.18-3 and PoPoolation V1.2.2 following the alignment of the reads to the reference genome assembly of *P. neurophilia* using the Burrows-Wheeler Alignment tool BWA V0.7.12 with the BWA-ALN algorithm. The resulting SAM file was then converted to the FreeBayes required input sorted BAM file using SAMtools. Polymorphic sites were then identified using the variant caller Freebayes. Variants were filtered using vcffilter from the C++ library vcflib to retain SNPs (*TYPE* = snp) found within a 25% interval of the average genome coverage (0.75 X > *DP* < 1.25 X genome coverage) that possessed only one alternative allele (*NUMALT* = 1). Resulting SNPs were retrieved and plotted to evaluate their frequency as well as their distribution within the genome using environment software for statistical computing and graphics R V3.2.1. K-mer coverage distribution was also performed to infer *P. neurophilia*'s ploidy using K-mergenie V1.6982

***Gene order conservation***

Region with potential gene order conservation were identify using SynMap tool from the CoGe platform (Lyons & Freeling, 2008). To this end, previously published annotated microsporidian

genomes were transferred into the platform and analyzed using the Quota Align Merge Algorithm with the DAGChainer Option "Relative Gene Order". Flanking regions of identified regions were then manually inspected for potential expansion of the predicted region with gene synteny.

### *PCR and Sanger sequencing*

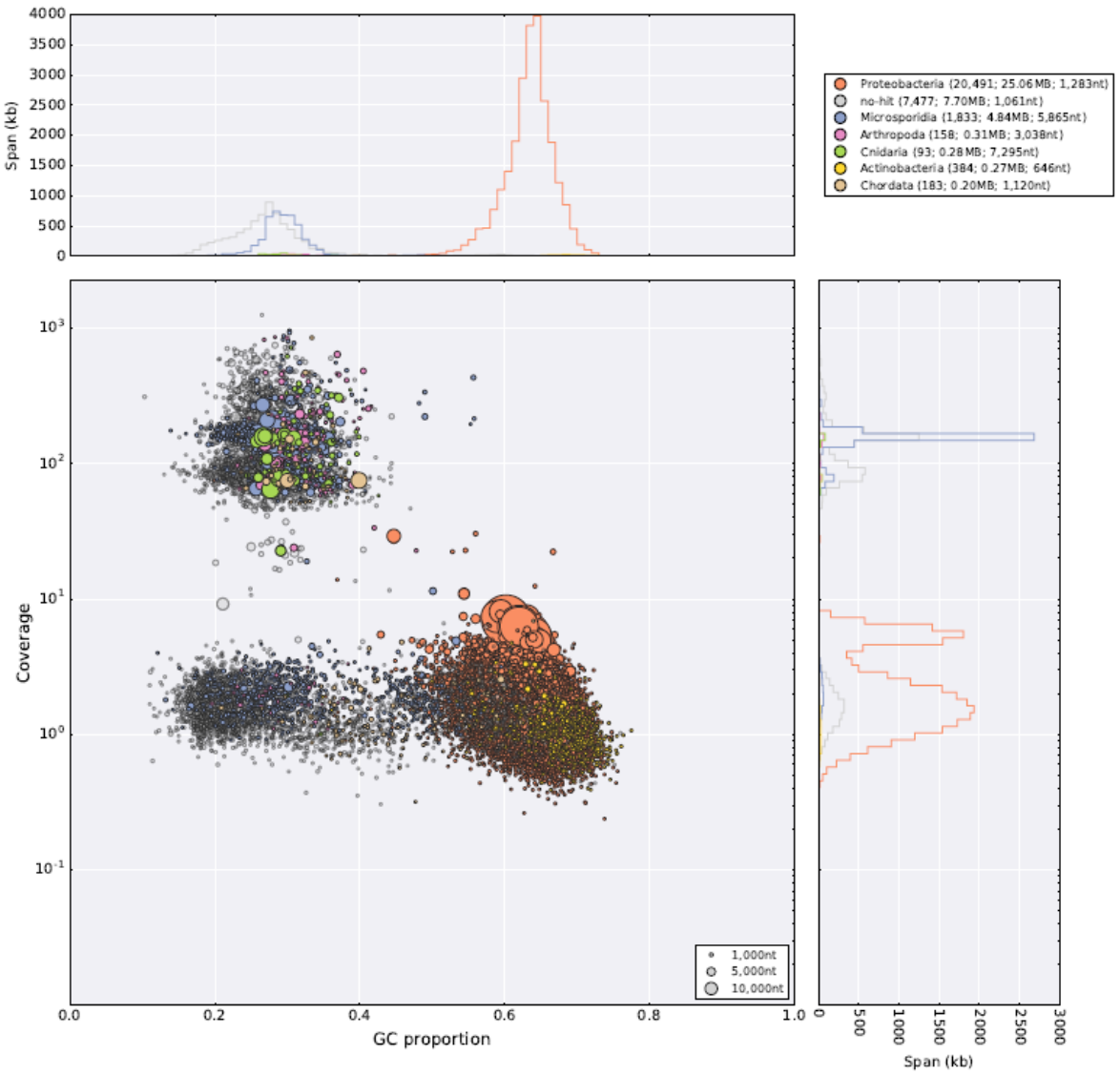
Selected regions with variable single nucleic polymorphic allele frequencies were amplified by PCR (Supplementary table 10) using standard parameters and sequenced using Sanger Sequencing (McGill University and Génome Québec Innovation, Montréal, Canada).

### *MRG distribution in microsporidia*

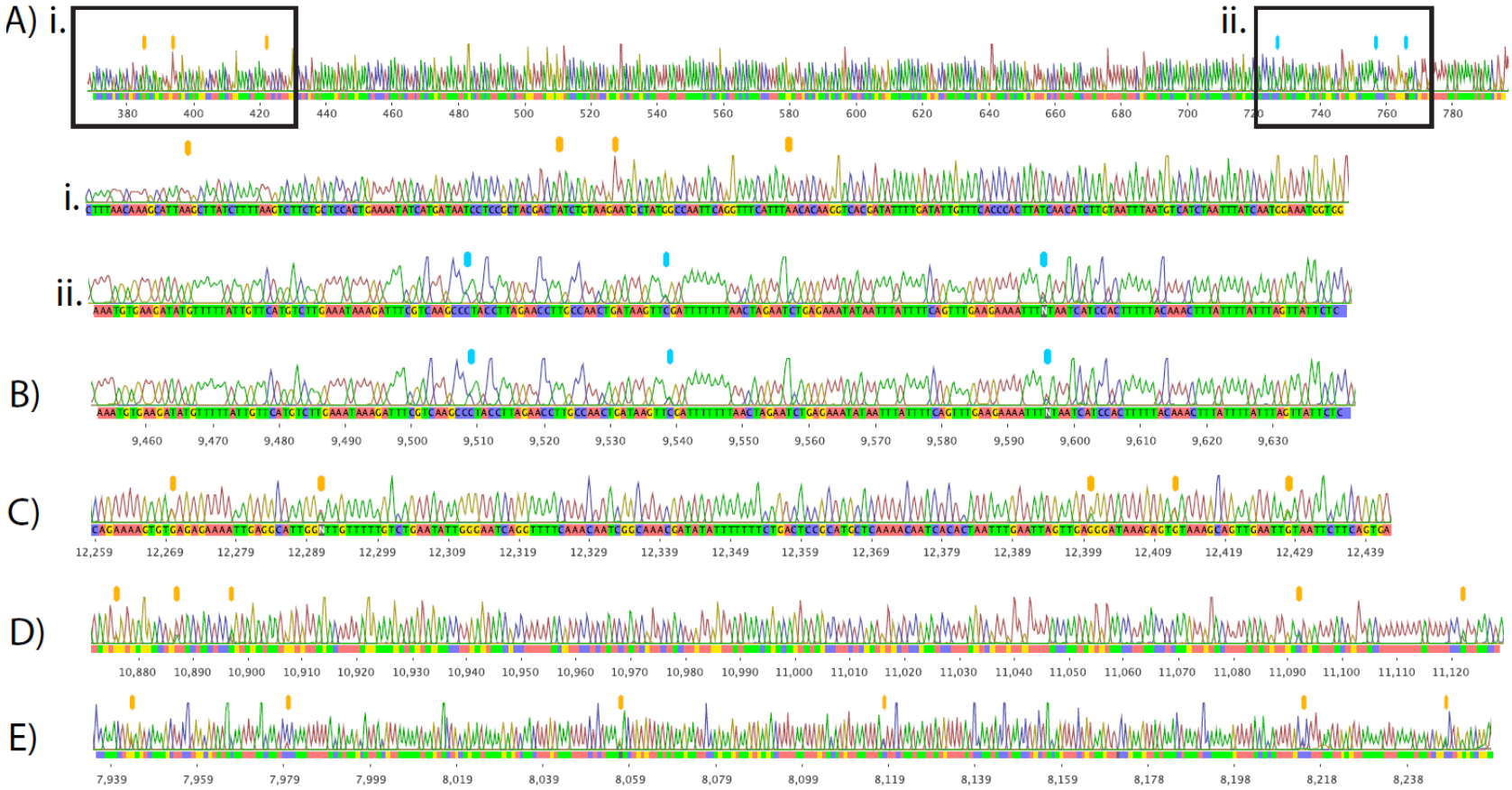
Meiosis specific-specific genes, as described in Lee et al. (Lee *et al.*, 2014), were retrieved from the NCBI database and used as queries against the *Pseudoloma neurophilia*'s genome (tBlastn; E-value cut-off of 1x $e^{-20}$). True homology was confirmed by reciprocal blast. Search of MRGs in other microsporidian genomes were conducted using the same methodology.

## Supplementary Data

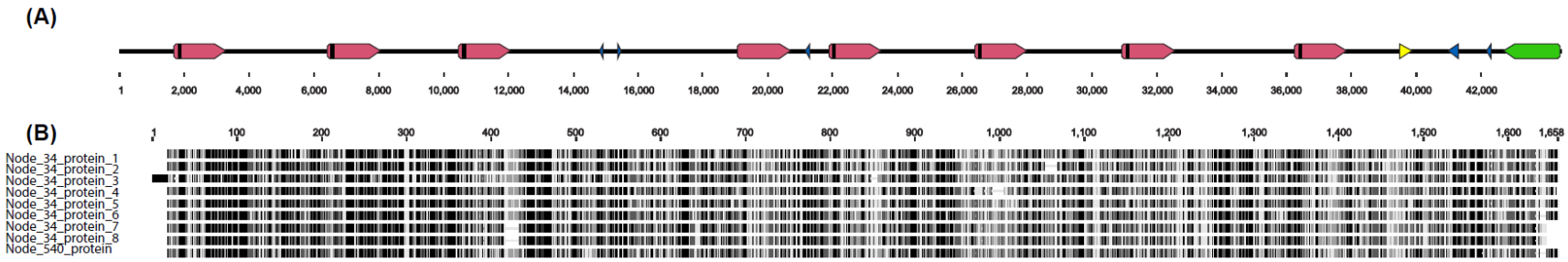

**Supplementary figure 1.** Graphical representation of *Pseudoloma neurophilia* (orange) contigs and proposed contaminants (green) contigs from the initial assembly. Overall G-C content and coverage of respective contigs is plotted.



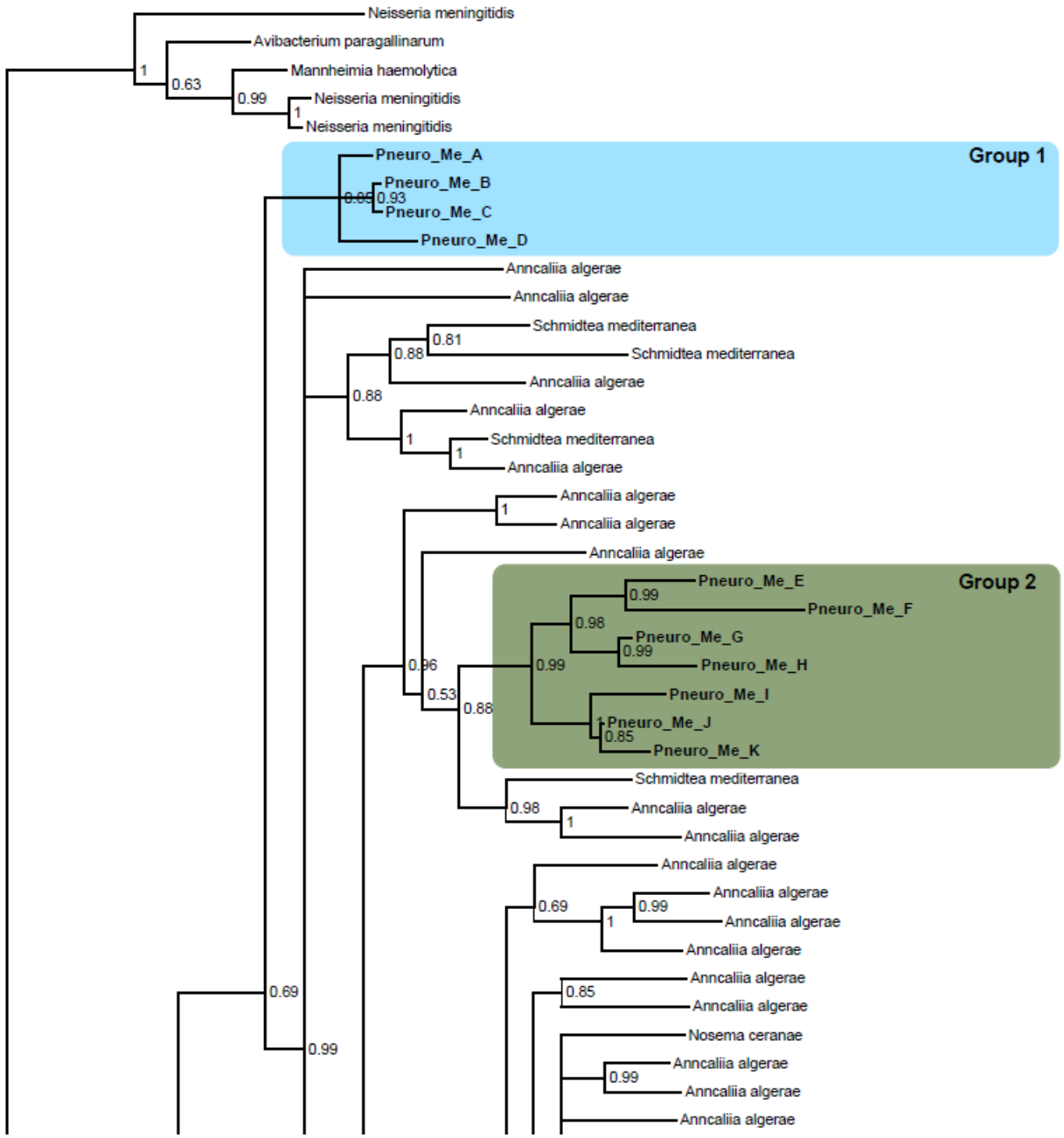

**Supplementary figure 2.** Graphical representation of contigs from the initial assembly and their respective putative origin. Seven predominant contigs homology hits are presented. Overall G-C content and coverage of contigs is plotted as well as their respective density plot.



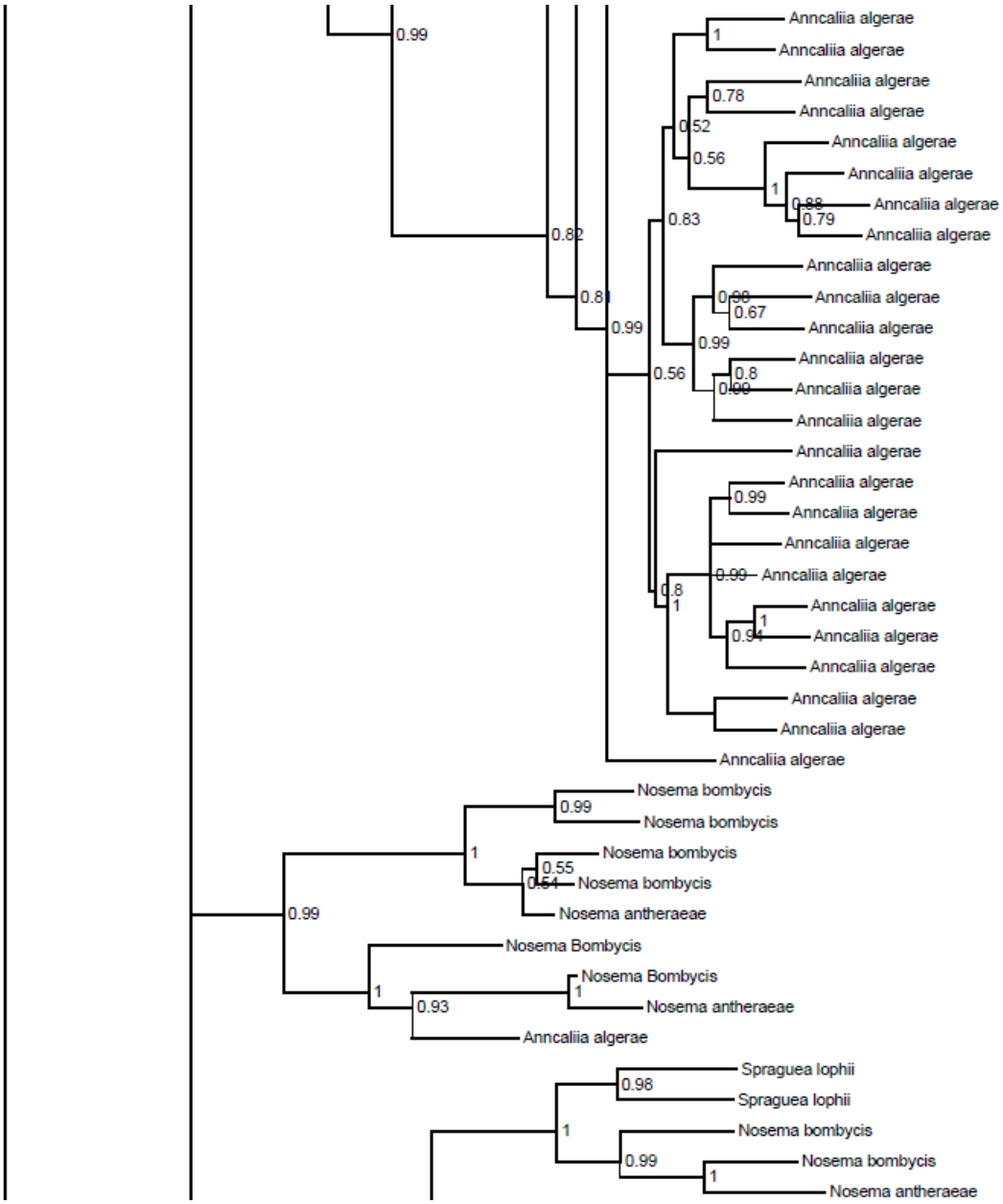

**Supplementary figure 3.** Sanger sequencing chromatograms of five loci with variable SNPs allele frequencies. Argued heterozygous variants (blue) and variant with low allelic frequencies (orange) are shown.



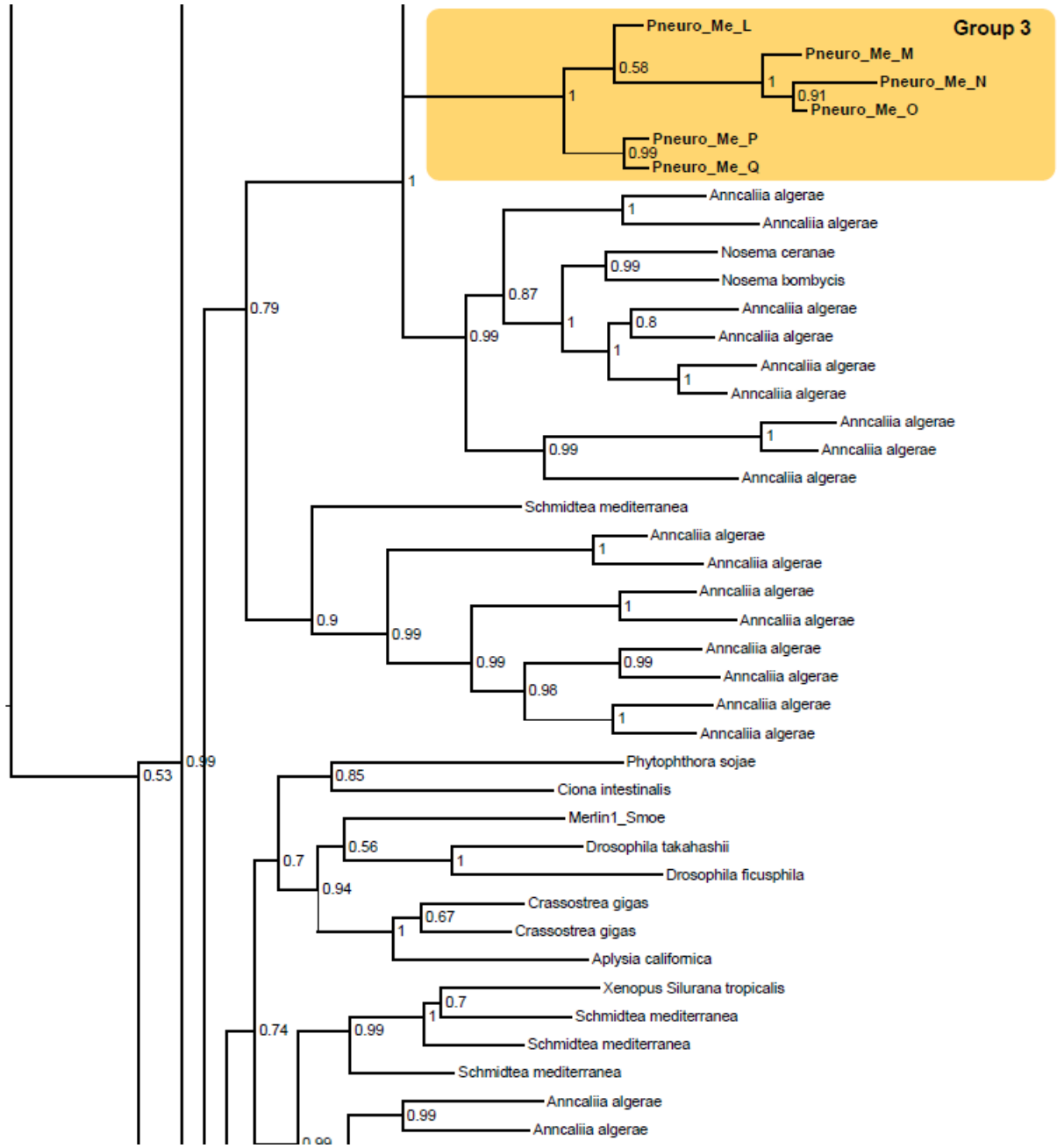

**Supplementary figure 4.** (A) Schematic representation of a 44 kb region with LOH harboring 8 genes unique to P. neurophilia resulting from multiple duplication events (pink) and their respective secretory signal sites. Putative secretory signal sites are shown (black). (B) Nucleotide pairwise alignment of genes is shown (pairwise identity 76.6%).



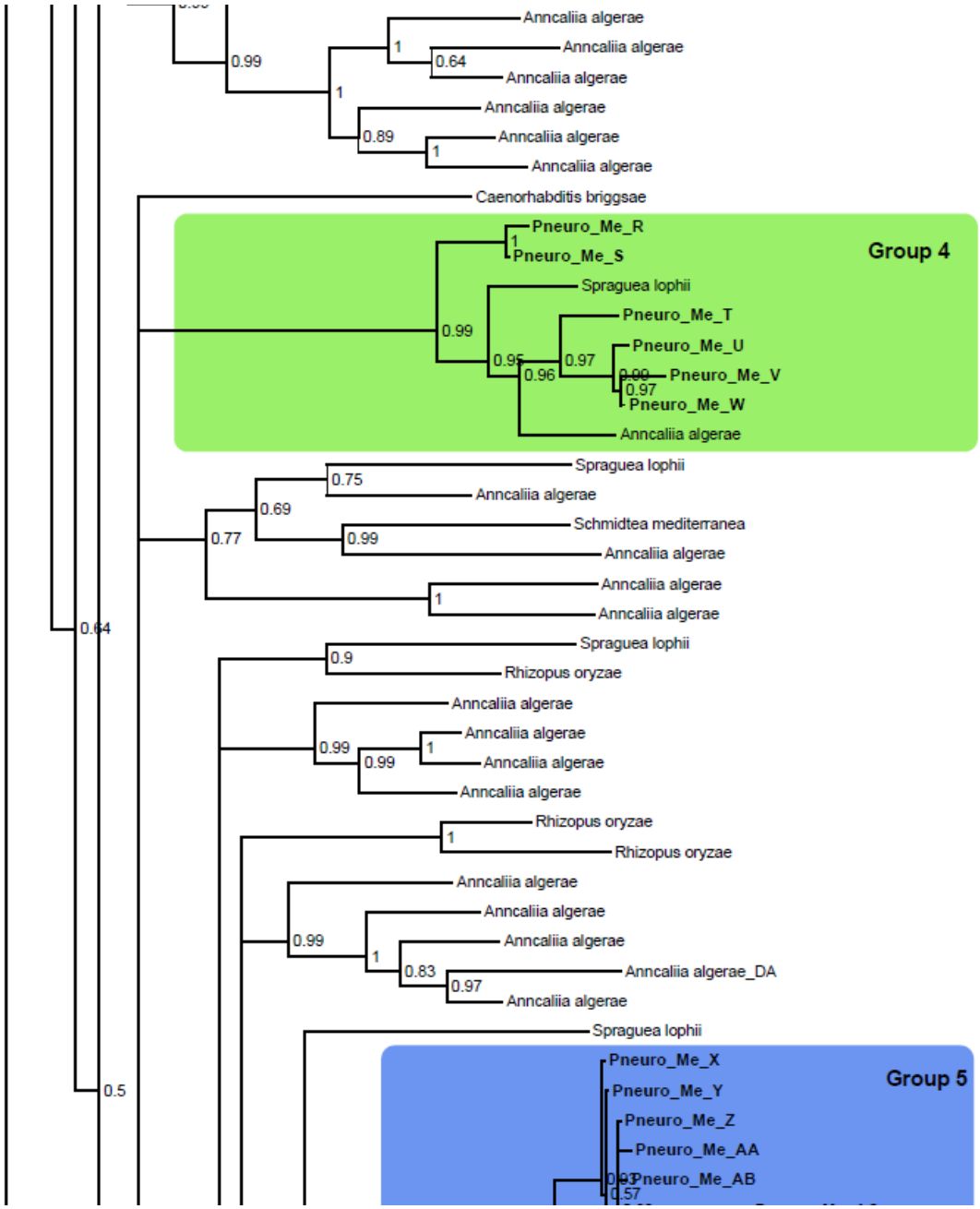

**Supplementary figure 5.** Phylogenetic reconstruction of the Merlin transposable family found in *Pseudoloma neurophilia* (in bold). Posterior probabilities for each node are shown.



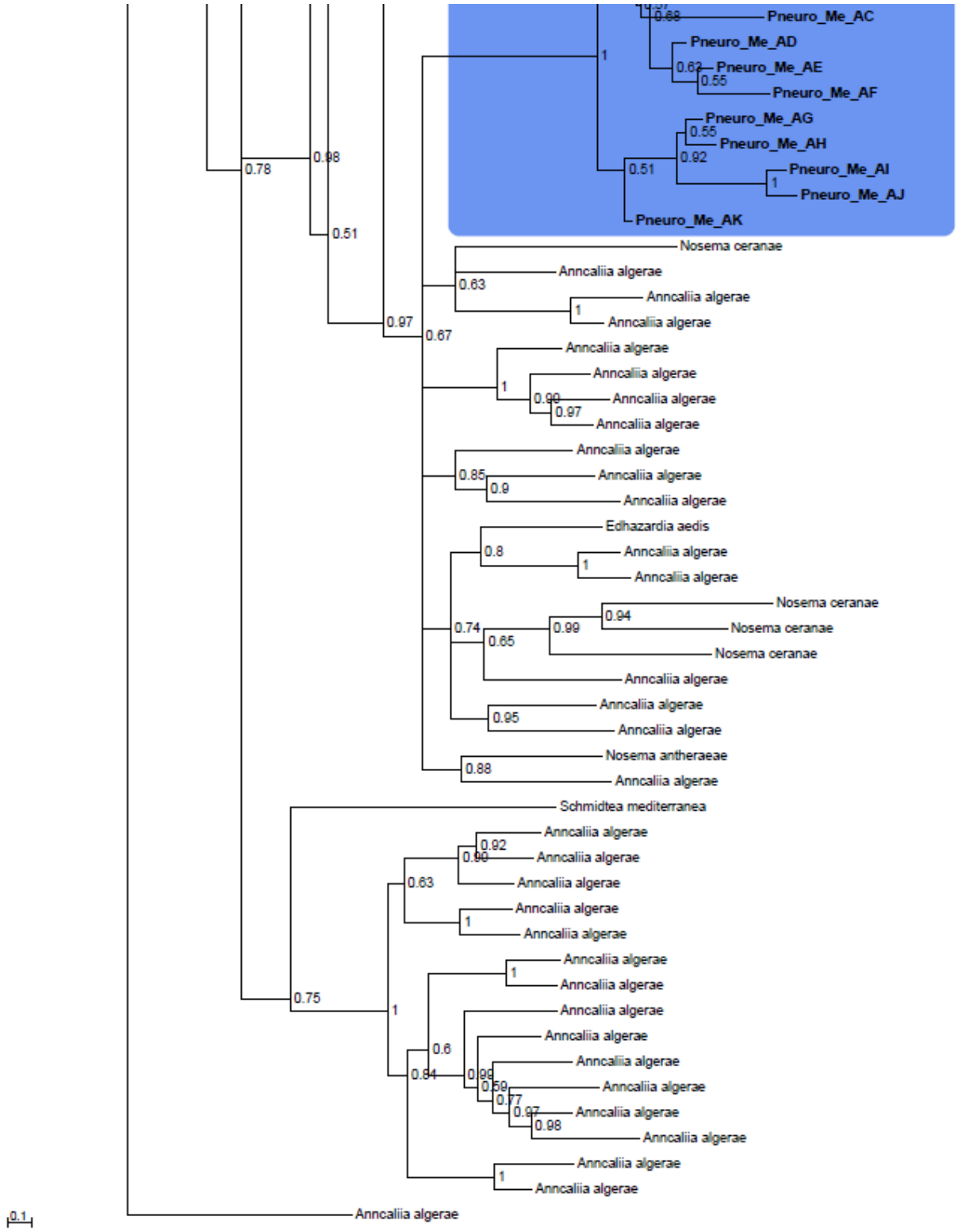

**Supplementary figure 5** (continued).



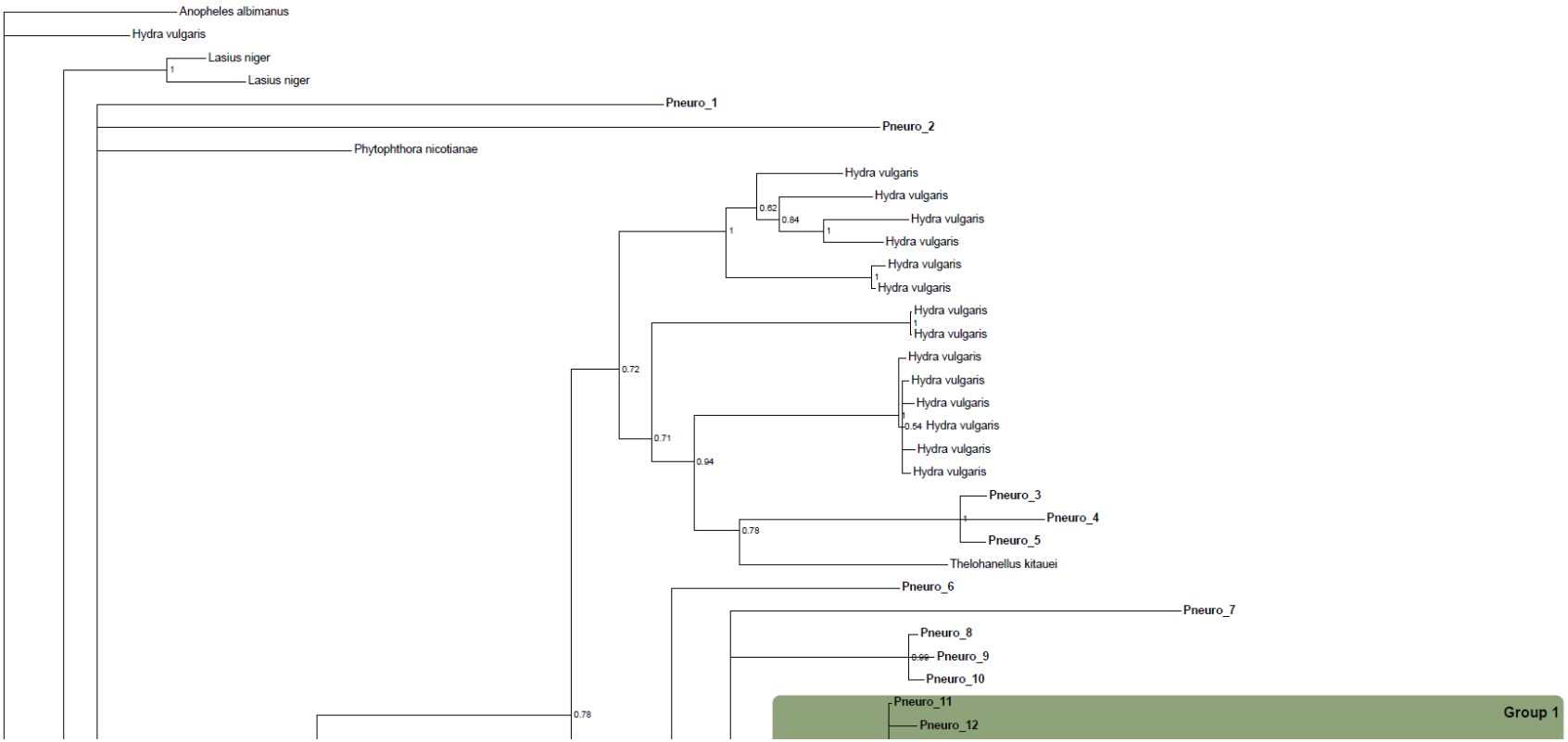

**Supplementary figure 5** (continued).



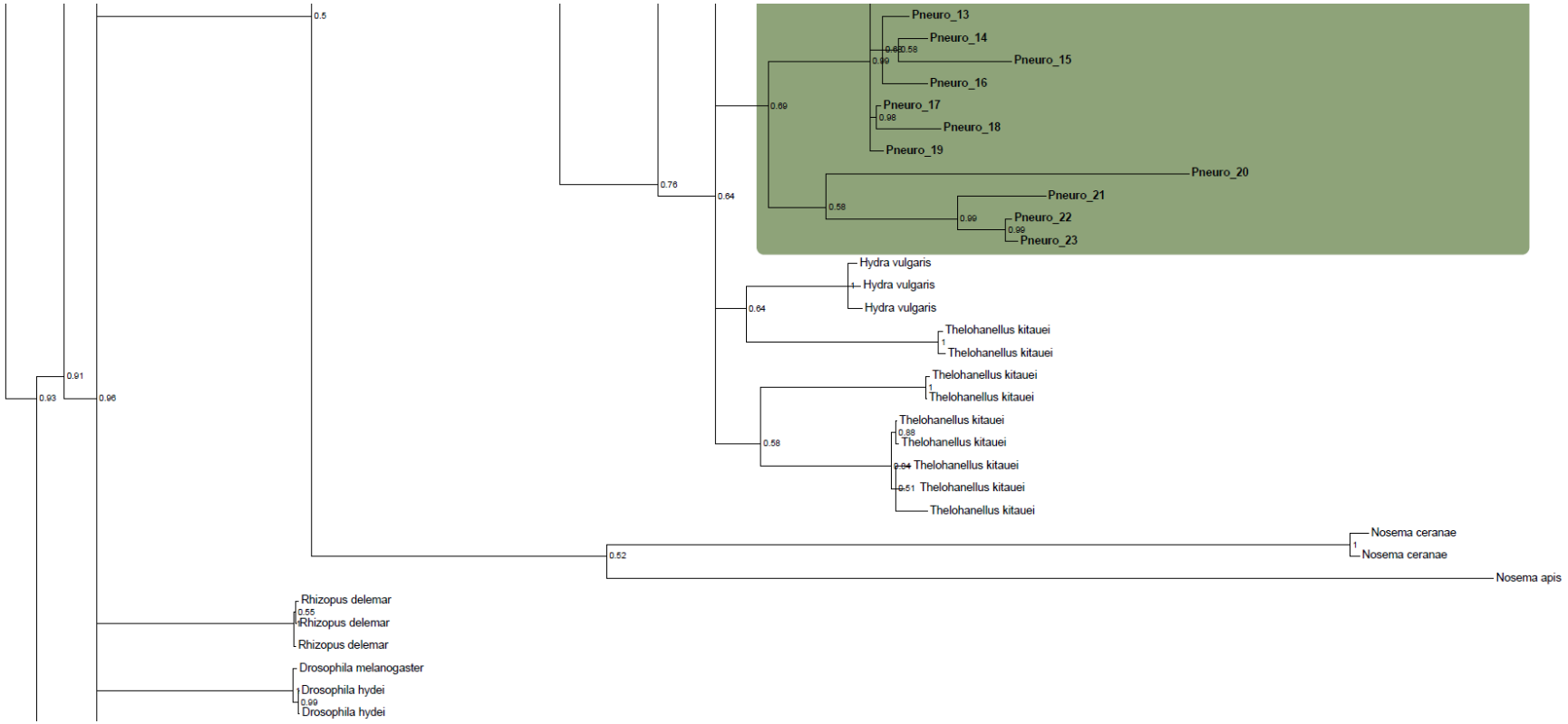

**Supplementary figure 5** (continued).



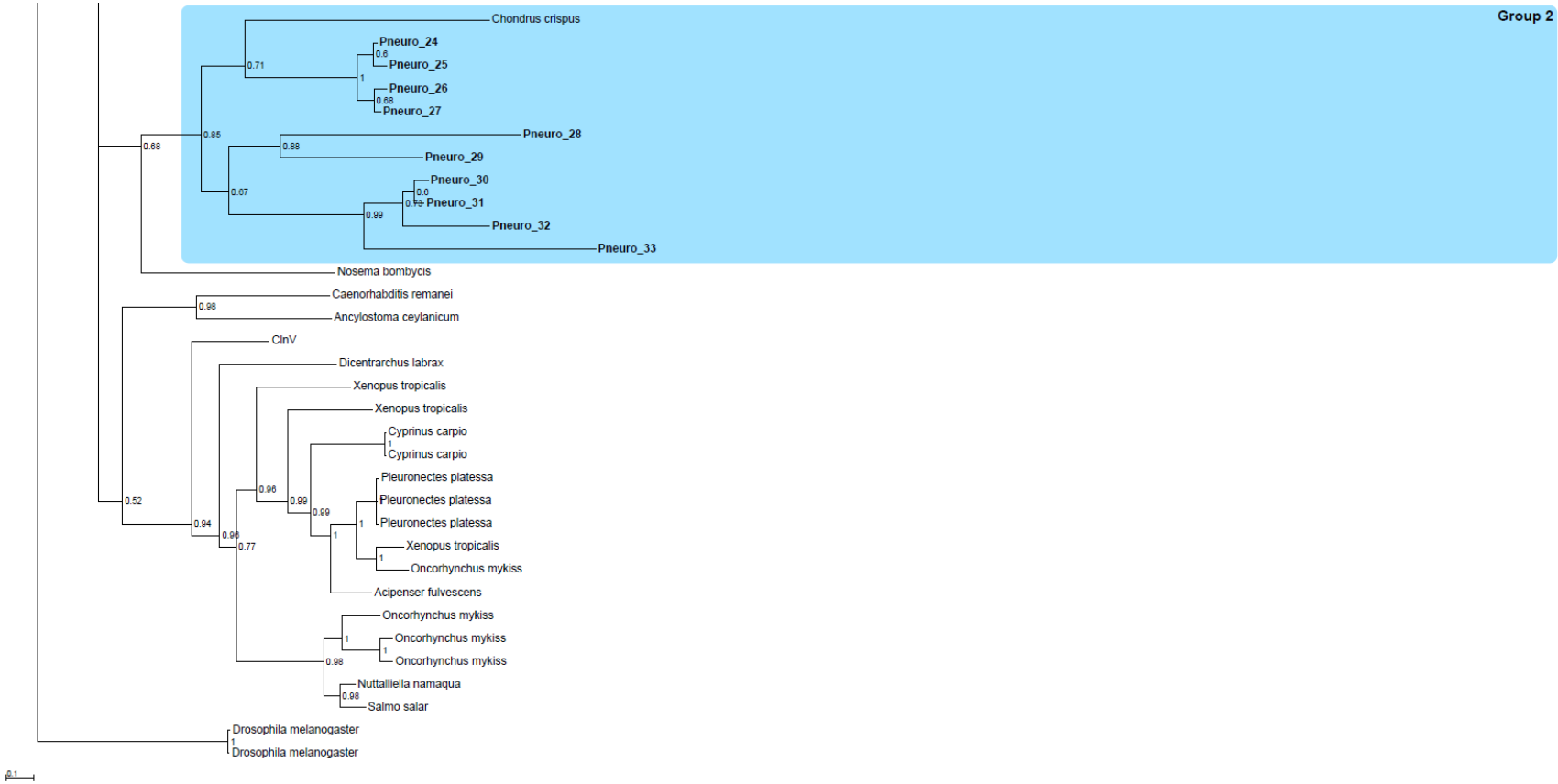

**Supplementary figure 5** (continued).



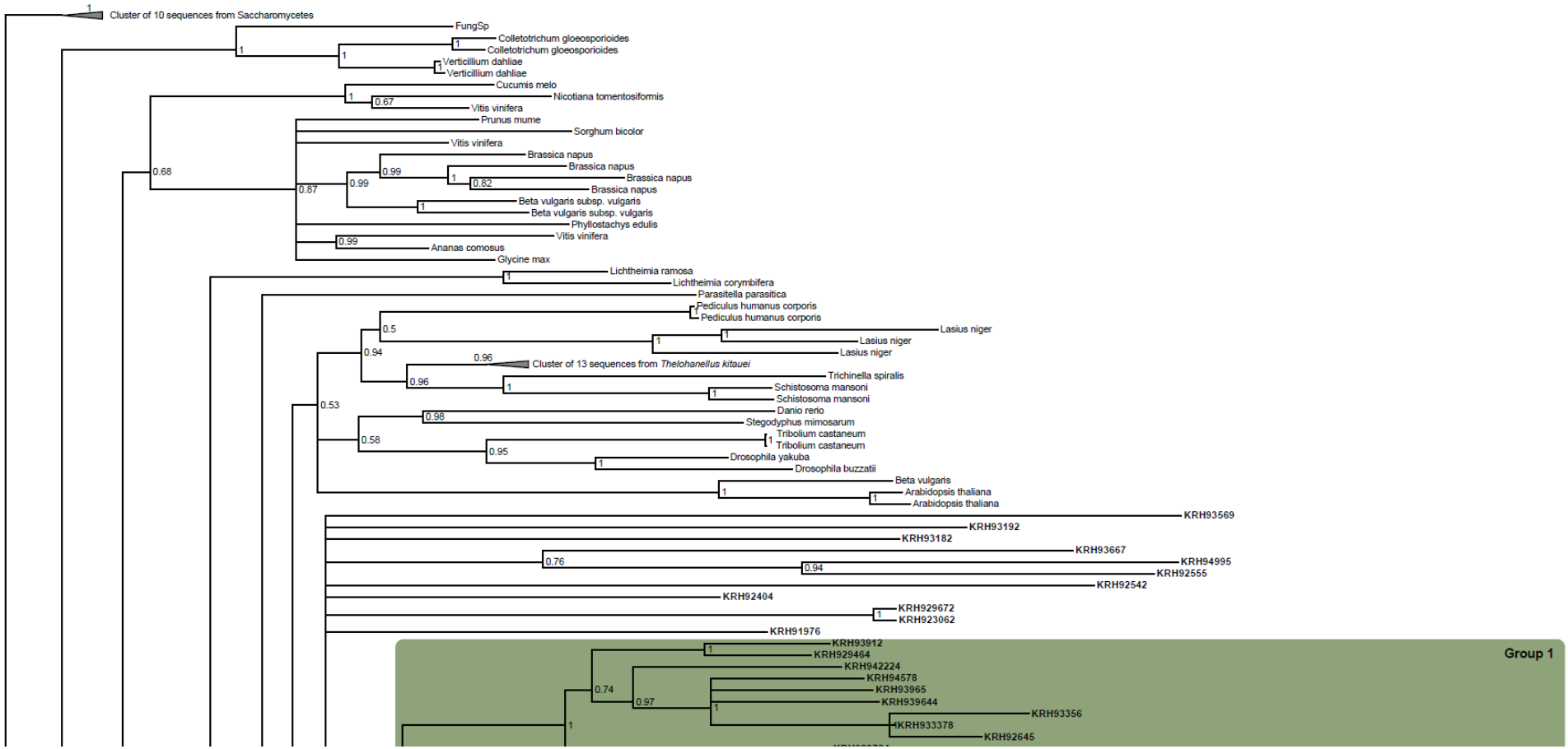

**Supplementary figure 6.** Phylogenetic reconstruction of the Mariner transposable family found in *P. neurophilia* (in bold). Posterior probabilities for each node are shown.



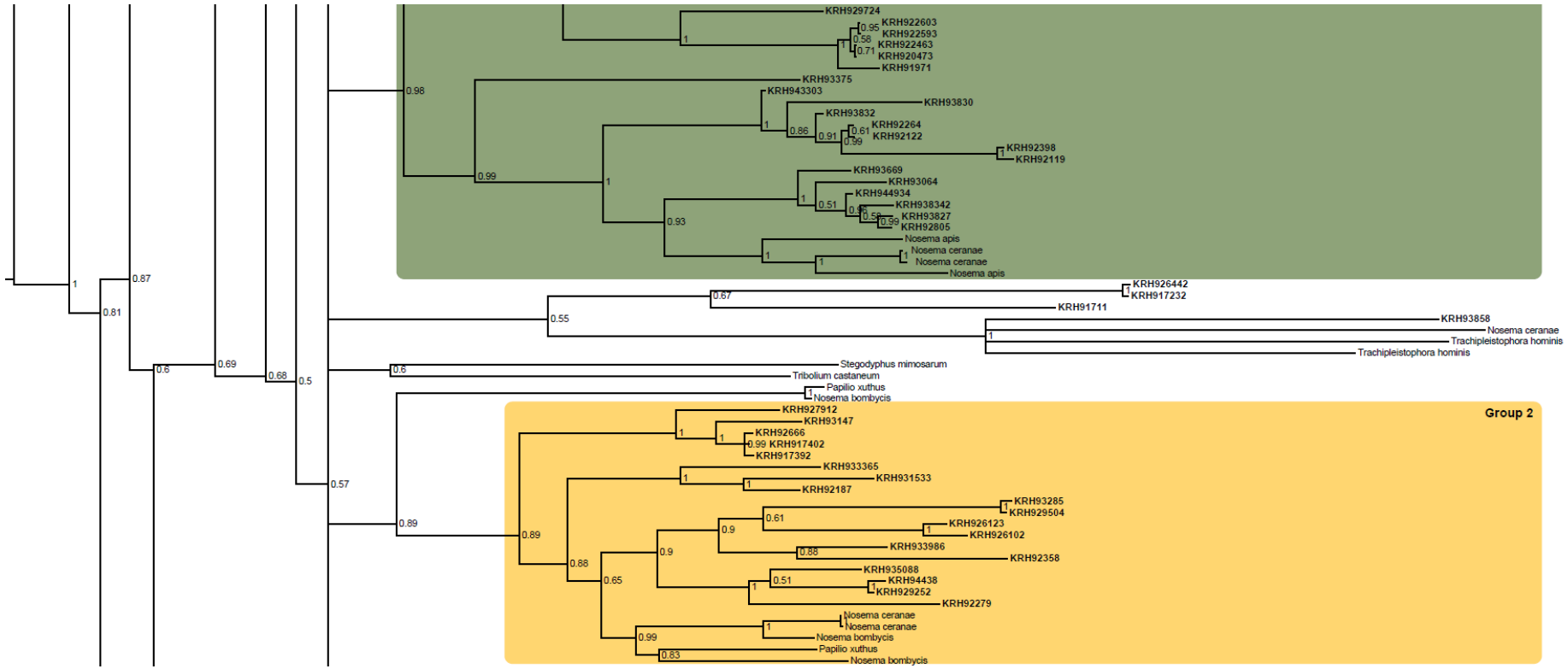

**Supplementary figure 6** (continued).



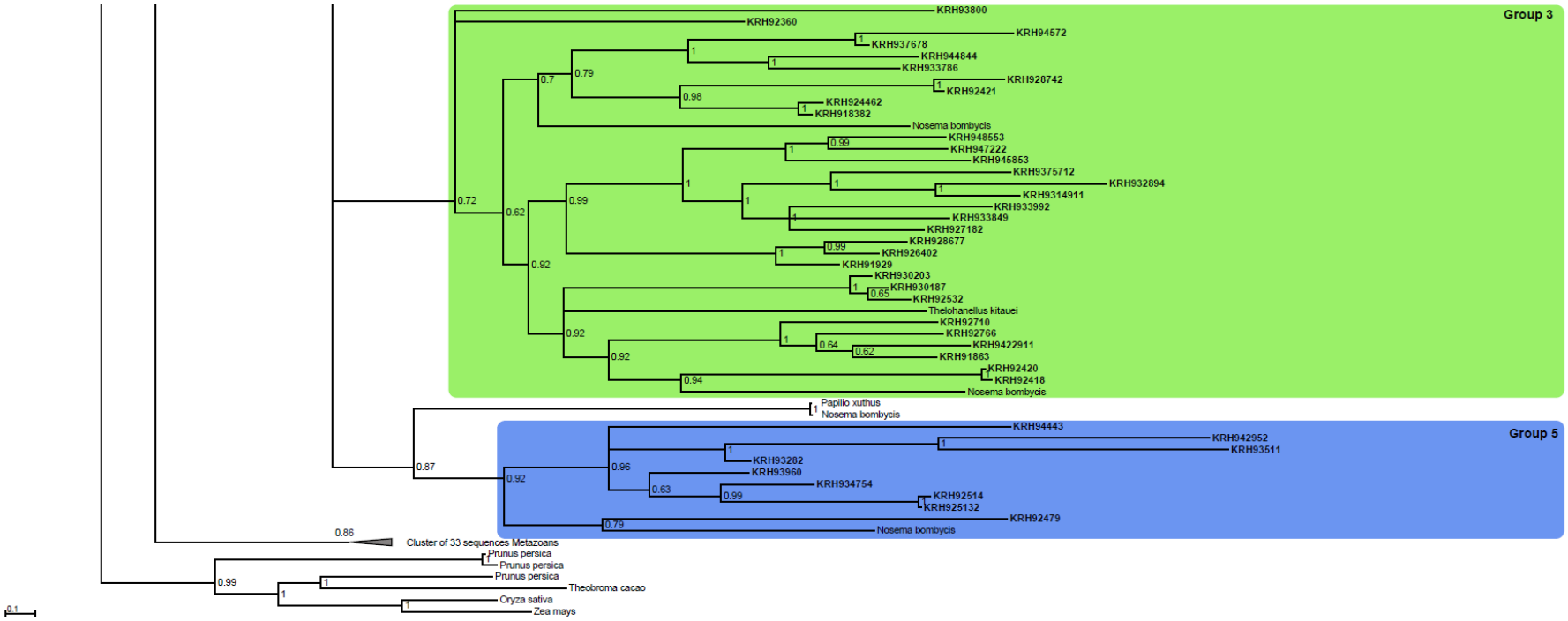

**Supplementary figure 6** (continued).



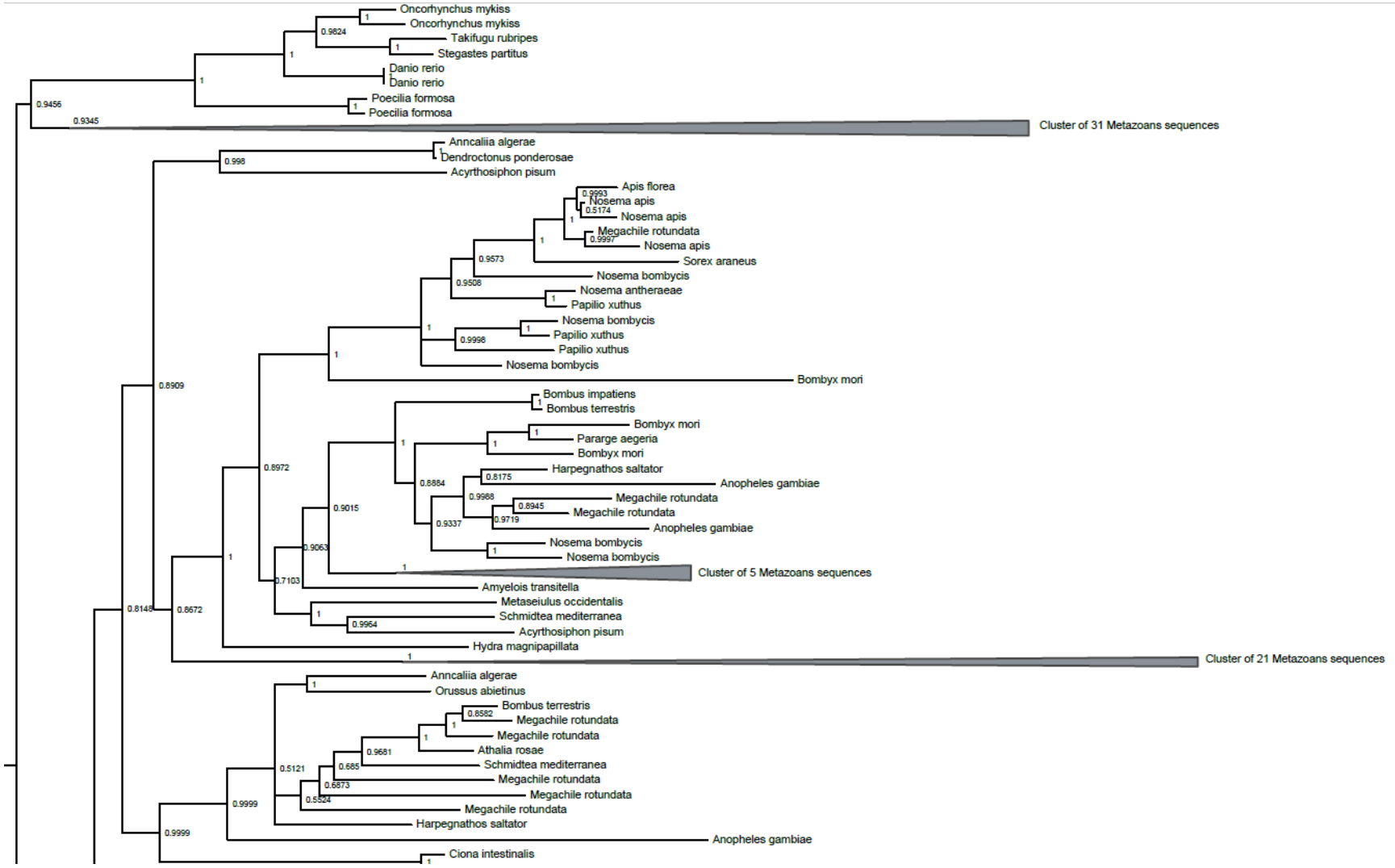

**Supplementary figure 7.** Phylogenetic reconstruction of the LTR transposable family found in *P. neurophilia* (in bold). Posterior probabilities for each node are shown.



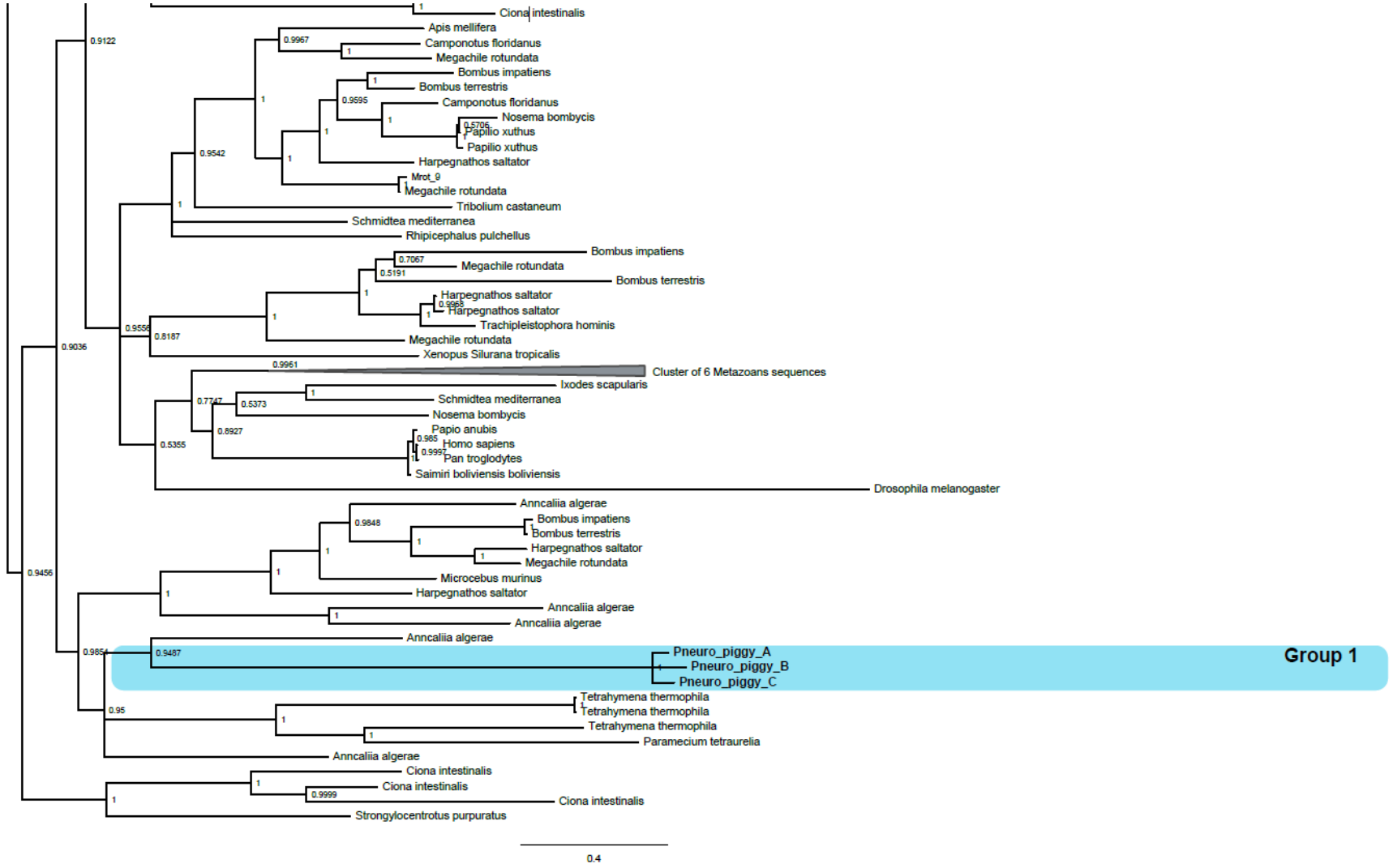

**Supplementary figure 7** (continued).




**Supplementary figure 7** (continued).




**Supplementary figure 8.** Phylogenetic reconstruction of the Piggybac transposable family found in *P. neurophilia*. Posterior probabilities for each node are shown.




**Supplementary figure 8** (continued).

**Supplementary figure 9.** Adult zebrafish, *Danio rerio,* infected with *Pseudoloma neurophilia*. This fish shows grossly visible signs of infection with associated spinal curvature. Most *P. neurophilia* infections are chronic and fish show little to no signs of infection, confounding the detection of this parasite. Bar = 1 mm.

**Supplementary table S1**. KOG family distribution among sequenced microsporidian genomes.

The supplementary table S1 contains over 3000 rows, and was not included in this document since it would occupy over 100 pages. The data presented in this table can be acquired by contacting the corresponding author Nicolas Corradi at ncorradi [AT] uottawa.ca as well as will be available in the online version of the paper following its publication.

**Supplementary table S2.** List of putative secreted genes found in *Pseudoloma neurophilia*. Targeted predicted location is shown: S, secretory pathway; M, mitochondrial; -, any other location. Reliability class is also shown; where lower the value, higher the support for the target prediction is. * represent predicted ORF unique to *P. neurophilia*.

| Sequence Description | Prediction of localization | Reliability class |
|---|---|---|
| Hypothetical protein * | S | 1 |
| hypothetical protein | S | 3 |
| Hypothetical protein * | M | 4 |
| hypothetical protein | S | 1 |
| u1 zinc finger family protein | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 5 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| polar tube protein ptp2 | S | 4 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| endoplasmic reticulum membrane protein | S | 2 |
| membrane integral protein | S | 2 |
| Hypothetical protein * | S | 1 |
| chitinase | S | 4 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| aminopeptidase n | S | 1 |
| endoplasmic reticulum translocation complex subunit sec66 | S | 1 |
| Hypothetical protein * | S | 3 |
| hypothetical protein | S | 1 |
| peptidyl-trna hydrolase | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 3 |
| sugar transporter | S | 3 |
| hypothetical protein | S | 3 |
| membrane protein | S | 1 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 2 |

**Supplementary table S2.** (continued)

| | | |
|---|---|---|
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| hypothetical protein | S | 1 |
| hypothetical protein | S | 3 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 4 |
| Hypothetical protein * | S | 3 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 4 |
| hypothetical protein | S | 4 |
| hypothetical protein | S | 4 |
| Hypothetical protein * | S | 5 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 4 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 3 |
| spore wall and anchoring disk complex protein | S | 4 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 3 |
| transcriptional antiterminator | S | 1 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 4 |
| scp pr1 domain-containing protein | S | 1 |
| Hypothetical protein * | S | 2 |
| lpxtg-domain-containing protein | S | 3 |
| af142406_1200 kda antigen p200 | S | 1 |
| hypothetical protein | S | 1 |
| hypothetical protein | S | 2 |
| hypothetical protein | S | 2 |
| hypothetical protein | M | 2 |
| ribosome-binding protein 1 | S | 2 |
| threonyl-trna synthetase | S | 1 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 5 |
| Hypothetical protein * | S | 2 |
| nucleoside phosphatase | S | 3 |
| dnaj-class molecular chaperone | S | 2 |
| hypothetical protein | S | 3 |
| sulfate transporter family protein | S | 1 |
| f-box domain containing protein | S | 5 |
| Hypothetical protein * | S | 3 |
| ubiquitin peptidase c19 | S | 3 |
| molecular chaperone | S | 1 |
| lpxtg-motif cell wall anchor domain protein | S | 3 |

**Supplementary table S2.** (continued)

| | | |
|---|---|---|
| lectin vip36 | S | 2 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 2 |
| hypothetical protein | S | 3 |
| lysyl-trna synthetase | S | 3 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 1 |
| Hypothetical protein * | S | 1 |
| aminopeptidase 2 | S | 3 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | _ | 5 |
| hypothetical protein | S | 4 |
| Hypothetical protein * | S | 2 |
| peptidyl-prolyl cis-trans isomerase | S | 3 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| sad1 unc domain protein | S | 2 |
| hypothetical protein | S | 1 |
| Hypothetical protein * | S | 5 |
| drug metabolite transporter superfamily | S | 1 |
| Hypothetical protein * | S | 2 |
| drug metabolite transporter superfamily | S | 1 |
| gpi-anchor transamidase | S | 4 |
| glycosyltransferase family 15 protein | S | 1 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| ricin b lectin protein | M | 5 |

**Supplementary table S2.** (continued)

| | | |
|---|---|---|
| ubiquitin hydrolase | S | 1 |
| translation initiation factor if-2 | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| carbohydrate-binding protein | S | 1 |
| hypothetical protein | S | 5 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 5 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| ricin b lectin protein | M | 5 |
| acid phosphatase precursor | S | 2 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 3 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 2 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 1 |
| endoplasmic reticulum membrane-associated oxidoreductin | S | 3 |
| hypothetical protein | S | 3 |
| Hypothetical protein * | S | 1 |
| gethr pentapeptide repeat (5 copies) family | S | 1 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| spermatogenesis-associated protein 32 | S | 1 |
| projectin twitchin | S | 3 |
| hypothetical protein | S | 4 |

**Supplementary table S2.** (continued)

| | | |
|---|---|---|
| Hypothetical protein * | S | 1 |
| trehalase | S | 2 |
| Hypothetical protein * | S | 1 |
| pyrimidine 5 -nucleotidase | S | 1 |
| vesicle transport v-snare protein | S | 1 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 5 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 1 |
| Hypothetical protein * | S | 1 |
| hypothetical protein | S | 1 |
| hypothetical protein | S | 2 |
| Hypothetical protein * | S | 4 |
| Hypothetical protein * | S | 1 |
| hypothetical protein | S | 4 |
| Hypothetical protein * | S | 2 |
| protein disulfide isomerase | S | 4 |
| endoplasmic oxidoreductin-1 | S | 3 |
| hypothetical protein | S | 4 |
| Hypothetical protein * | S | 2 |
| hypothetical protein | S | 1 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 3 |
| Hypothetical protein * | S | 5 |
| y215_enccu ame: full=spore wall protein ecu02_0150 flags: precursor | S | 1 |
| ricin b lectin protein | S | 1 |
| hypothetical protein | S | 3 |

**Supplementary table S2.** (continued)

| **Description of secreted genes** | **Count** |
|---|---|
| Hypothetical protein * | 111 |
| hypothetical protein | 35 |
| ricin b lectin protein | 3 |
| drug metabolite transporter superfamily | 2 |
| Other | 46 |
| Total | 197 |
| **Targeted predicted location of secreted genes** | **Count** |
| Secretory Pathway | 192 |
| Mitochondrial | 4 |
| Any other location | 1 |

**Supplementary table S3**. Overall SNP frequency of several microsporidian genomes

| **Specie** | **SNP frequency (SNP/Kb)** | **Reference** |
|---|---|---|
| *Encephalitozoon cuniculi* | 0.01 | Selman et al. |
| *Rozella allomycis* | 0.33 | James et al. |
| *Vavraia culics loridensis* | 1.43 | Desjardins et al. |
| *Nematocida parisii ERTm3* | 1.79 | Cuomo et al. |
| *Nematocida parisii ERTm1* | 1.94 | Cuomo et al. |
| *Pseudoloma neurophilia* | 1.98 | This study |
| *Edhazardia aedis* | 3.52 | Desjardins et al. |
| *Nematocida sp. ERTm2* | 8.97 | Cuomo et al. |
| *Nosema ceranae* | 9.91 | Pelin et al. |

**Supplementary table S4.** List of genes found in regions with loss of heterozygosity present in the genome assembly's largest contigs.

| Location | Description |
|---|---|
| NODE_1 | 26S proteasome subunit P45 family protein |
| NODE_1 | 40S ribosomal protein S2/30S ribosomal protein S5 |
| NODE_1 | anaphase-promoting complex subunit 10 |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein |
| NODE_1 | hypothetical protein with similarity to ARID/BRIGHT DNA binding domain protein |
| NODE_1 | hypothetical protein with similarity to WD40-repeat-containing subunit of the 18S rRNA processing complex |
| NODE_1 | Microtubule-associated protein essential for anaphase spindle elongation |
| NODE_1 | putative RNA methylase involved in rRNA processing |
| NODE_1 | Regulator of arginine metabolism, MADS box-containing transcription factor |
| NODE_1 | RNA polymerase III transcription factor (TF)IIIC subunit |
| NODE_1 | Transcription elongation factor SPT6 |
| NODE_11 | 20S proteasome, regulatory subunit beta type PSMB2/PRE1, partial |
| NODE_11 | 26S proteasome regulatory complex, subunit RPN8/PSMD7 |
| NODE_11 | 40S ribosomal protein SA (P40)/Laminin receptor 1 |
| NODE_11 | 60S acidic ribosomal protein P0 |
| NODE_11 | DEAH-box RNA helicase |
| NODE_11 | DNA polymerase delta catalytic subunit |
| NODE_11 | Glycosyltransferase |
| NODE_11 | Histone acetyltransferase complex SAGA/ADA, subunit ADA2 |
| NODE_11 | Histone acetyltransferase PCAF/SAGA, subunit SUPT3H/SPT3, partial |
| NODE_11 | hypothetical protein, partial |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |

**Supplementary table S4** (continued).

| | |
|---|---|
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | hypothetical protein |
| NODE_11 | lysine-tRNA ligase |
| NODE_11 | putative CT20 protein |
| NODE_11 | putative DHHC-type Zn-finger protein, partial |
| NODE_11 | putative DHHC-type Zn-finger protein |
| NODE_11 | putative E3 ubiquitin ligase |
| NODE_11 | putative EKC/KEOPS complex, subunit Pcc1 protein |
| NODE_11 | putative Metallo-hydrolase/oxidoreductase, Beta-lactamase-like protein, partial |
| NODE_11 | putative transporter |
| NODE_11 | putative Zn-finger protein |
| NODE_11 | Regulator of spindle pole body duplication |
| NODE_11 | RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, subunit SSL1 |
| NODE_11 | RNA polymerase II transcription initiation/nucleotide excision repair factor TFIIH, subunit TFB2 |
| NODE_11 | Structural maintenance of chromosome |
| NODE_11 | U3 small nucleolar ribonucleoprotein (snoRNP) component |
| NODE_11 | Zinc finger C3HC4 type protein |
| NODE_21 | 40S ribosomal protein S21 |
| NODE_21 | 60S ribosomal protein L4 |
| NODE_21 | 60S ribosomal protein L18A |
| NODE_21 | Arginyl-tRNA synthetase |
| NODE_21 | Ca2+:Cation Antiporter (CaCA) Family |
| NODE_21 | Dolichol-phosphate mannosyltransferase |
| NODE_21 | hypothetical protein, partial |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein |
| NODE_21 | hypothetical protein VCUG_01782 |
| NODE_21 | hypothetical protein with similarity to Lectin VIP36 |
| NODE_21 | Mismatch repair ATPase MSH2 (MutS family) |

**Supplementary table S4** (continued).

| | |
|---|---|
| NODE_21 | mRNA deadenylase subunit |
| NODE_21 | Nucleotide excision repair complex XPC-HR23B, subunit XPC/DPB11 |
| NODE_21 | putative E3 ubiquitin ligase containing RING finger |
| NODE_21 | RNA polymerase III transcription factor TFIIIC |
| NODE_21 | Signal peptidase subunit |
| NODE_21 | Signal recognition particle, subunit Srp19 |
| NODE_21 | Spindle pole body protein |
| NODE_21 | triosephosphate isomerase |
| NODE_21 | Willebrand factor type A domain-containing AAA ATPase |

**Supplementary table S5**. Description and presence of secretion signal in duplicated genes found in a large LOH region.

| Description | Secreted | Duplicated gene |
|---|---|---|
| SNF2 family DNA-dependent ATPase | N | N |
| hypothetical protein | Y | Y |
| hypothetical protein | Y | Y |
| hypothetical protein | N | Y |
| hypothetical protein | Y | Y |
| hypothetical protein | Y | Y |
| hypothetical protein | Y | Y |
| hypothetical protein | Y | Y |
| hypothetical protein | Y | Y |
| aminomethyltransferase domain protein | N | N |
| hypothetical protein | N | N |
| hypothetical protein | N | N |
| hypothetical protein | N | N |
| hypothetical protein | N | N |
| hypothetical protein | N | N |

**Supplementary table S6.** Gene family count and distribution within microsporidian genomes.

The supplementary table S6 contains over 3000 rows, and was not included in this document since it would occupy over 100 pages. The data presented in this table can be acquired by contacting the corresponding author Nicolas Corradi at ncorradi [AT] uottawa.ca as well as will be available in the online version of the paper following its publication.

**Supplementary table S7.** List of predicted genes only found in *Pseudoloma neurophilia* and its closely related species *Vavraia culicis foridensis* and *Trachipleistophora hominis*. Presence of predicted secretion signal is shown.

| **Predicted function** | **secreted** |
|---|---|
| hypothetical protein | Yes |
| hypothetical protein | No |
| hypothetical protein | No |
| hypothetical protein | No |
| Glutaredoxin | No |
| hypothetical protein | Yes |
| hypothetical protein | No |
| hypothetical protein | No |
| N-arginine dibasic convertase NRD1, Zn2+-dependent endopeptidase, insulinase superfamily | No |
| Nucleoside phosphatase | Yes |
| hypothetical protein | Yes |
| ricin b lectin protein | Yes |
| hypothetical protein | Yes |
| hypothetical protein | No |
| putative Ribonuclease P/MRP protein subunit protein | No |
| hypothetical protein | No |

**Supplementary Table S8**. Presence (✓) and absence (-) of conventional meiosis genes within microsporidian species.

| Gene | Function | *Anncaliia algerae* | *Edhazardia aedis* | *Encephalitozoon cuniculi* | *E. hellem* | *E. intestinalis* | *E. romaleae* | *E. bieneusi* |
|---|---|---|---|---|---|---|---|---|
| dmc1 | DNA repair | ✓ | - | - | - | - | - | - |
| hop1 | meiosis specific DNA binding protein | ✓ | - | ✓ | ✓ | - | ✓ | - |
| hop2 | meiotic double strand DNA break repair | | - | - | - | - | - | - |
| mlh1/mutl/pms2 | mismatch repair during mitosis and meiosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - |
| mlh3 | mismatch repair during mitosis and meiosis | - | - | - | - | - | - | - |
| mnd1 | recombination and meiotic nuclear division | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| mre11 | DNA repair | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| msh2 | mismatch repair during meiosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| msh4 | meiotic DNA crossover | ✓ | ✓ | - | - | - | - | - |
| msh5 | meiotic DNA crossover | ✓ | ✓ | - | - | - | - | - |
| msh6 | mismatch repair during mitosis and meiosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| pds5 | sister chromatid condensation and cohesion | - | - | - | - | - | - | - |
| pms1/mlh2 | mismatch repair during meiosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| rad18 | post replication repair | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| rad50 | meiotic double strand DNA break | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| rad51 | recombination DNA repair | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| rad52/22 | double strand DNA break repair | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| rec8 | subunit of sister chromatid cohesion complex | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| smc1 | chromosome and double strand DNA break repair | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| smc2 | sister chromatid cohesion | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| smc3 | sister chromatid cohesion and recombination during meiosis | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| smc4 | subunit of cohesion complex | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| smc5 | DNA repair | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - |
| spo11 | Meiosis specific double strand DNA break formation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - |

**Supplementary Table S8**. (continued)

| *Mitosporidium daphniae* | *Nematocida sp.* | *N. parisii 1* | *Nosema apis* | *N. bombycis* | *N. ceranae* | *Ordospora colligata* | *Pseudoloma neurophilia* | *Rozella allomycis* | *Spraguea lophii* | *Trachipleistophora homins* | *Vavraia culics loridensis* | *Vittaforma corneae* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | - | ✓ | - | - | - | - | - | - | ✓ | ✓ | - |
| - | - | - | ✓ | ✓ | ✓ | - | - | - | ✓ | ✓ | ✓ | - |
| - | - | - | - | - | - | - | - | - | - | - | - | - |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| - | - | - | - | - | - | - | - | - | - | - | - | - |
| - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | - | - | - | ✓ | ✓ | - | - | - | ✓ | ✓ | ✓ | - |
| - | - | - | - | - | - | - | - | - | ✓ | | ✓ | - |
| ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| - | - | - | - | - | - | - | - | - | - | - | - | - |
| | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | - | - | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| - | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| - | ✓ | ✓ | - | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Supplementary table S9.** Genes in *Pseudoloma neurophilia* with large inserts in regions usually conserved in other organisms. Presence of inserts in other microsporidian species is also shown.

| | | Presence in Microsporidia | | | | |
|---|---|---|---|---|---|---|
| **Function** | **Total length of all inserts (aa)** | *P. neurophilia* | *T. hominis* | *V. culicis* | *S. lophii* | *E. aedis* |
| 20S proteasome, regulatory subunit alpha type PSMA7/PRE6 | 51 | ✓ | - | - | - | - |
| 26S proteasome regulatory complex | 30 | ✓ | - | - | - | - |
| 6-phosphogluconate dehydrogenase | 22 | ✓ | - | - | - | ✓ |
| Argonaute | 30 | ✓ | - | - | - | - |
| ATP-binding Cassette (ABC) Superfamily | 138 | ✓ | ✓ | ✓ | - | - |
| ATP-binding Cassette (ABC) Superfamily | 79 | ✓ | - | - | - | - |
| ATP-dependent RNA helicase | 44 | ✓ | - | - | - | - |
| Beta-lactamase fold containing | 231 | ✓ | - | - | - | ✓ |
| Ca2+-binding actin-bundling protein (actinin), alpha chain (EF-Hand protein superfamily), partial | 20 | ✓ | - | - | - | - |
| Chromatin remodeling complex WSTF-ISWI, small subunit | 27 | ✓ | - | - | - | - |
| DNA polymerase (pol2) | 22 | ✓ | - | - | - | - |
| DNA replication licensing factor, MCM7 component | 70 | ✓ | - | - | - | - |
| DNA topoisomerase type II, partial | 68 | ✓ | ✓ | ✓ | - | - |
| Dual specificity, serine/threonine and tyrosine kinase, partial | 97 | ✓ | - | - | - | - |
| Helicase of the dead superfamily | N/A | - | - | - | - | ✓ |
| Histone deacetylase complex, SIN3 component | N/A | - | - | - | - | ✓ |
| Kinesin-like protein | 45 | ✓ | - | - | - | - |
| Lecithin:cholesterol acyltransferase (LCAT)/Acyl-ceramide synthase | 144 | ✓ | - | - | - | - |
| Methionine-tRNA ligase | 45 | ✓ | - | - | - | - |
| Molecular chaperones HSP70/HSC70, HSP70 superfamily | 70 | ✓ | - | - | - | - |
| Phenylalanine-tRNA synthetase, beta subunit | 65 | ✓ | - | - | - | - |
| Phosphatidylinositol transfer protein | 15 | ✓ | - | - | - | - |
| Phosphoglucomutase/phosphomannomutase, partial | 139 | ✓ | - | ✓ | - | - |
| P-type ATPase (P-ATPase) Superfamily | N/A | - | - | ✓ | - | - |
| putative phosphate acyltransferase | 81 | ✓ | - | - | - | - |
| RNA binding/translational regulation protein of the SUA5 family | 48 | ✓ | - | - | - | - |
| RNA polymerase II, second largest subunit | 19 | ✓ | ✓ | ✓ | - | - |
| RRM domain protein | 90 | ✓ | ✓ | ✓ | ✓ | - |

**Supplementary table S9** (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| Serine/threonine protein kinase | 279 | ✓ | - | - | - | - |
| T-complex protein 1, epsilon subunit | 35 | ✓ | - | - | - | - |
| Trehalose-6-phosphate synthase component TPS1 subunit, partial | 90 | ✓ | ✓ | ✓ | - | - |
| Zn-finger protein | 14 | ✓ | - | - | - | - |

**Supplementary table S10.** Primers used for PCR in the identification of low variants

| Region amplified in Supplementary Fig 3 | Sequence |
|---|---|
| Region_A_F | TGCCAAACACTCAATCTGTCAAG |
| Region_A_R | TTCTGTTCAATGTGCGGCAC |
| Region_B_F | TTGTGAAACCTCCTCTCTGAGA |
| Region_B_R | GCAACTACAAAGATAAAAGGCTACGA |
| Region_C_F | GCTCAATACACATGTCGAGGC |
| Region_C_R | CGGAGAATATCTGCTCAAGTTTAAA |
| Region_D_F | ACCAGTTCCATCACAAGAGCA |
| Region_D_R | GAAGTGGTGGTGGGAACGAT |
| Region_E_F | ATTTGAAAGCACGTGCCTGTC |
| Region_E_R | ACAGGACCAATGAATCTGTACGTA |

# Chapter 3 – Discussion, Future Directions and Concluding Remarks

## Importance of Research

*Pseudoloma neurophilia* is an intracellular obligate parasite of the powerful animal model *Danio rerio* (the zebrafish). The latter is widely used as a vertebrate animal model in biomedical research. Since its first discovery in 2001, this parasite gained growing scientific interest and was described as one of the zebrafish most relevant parasite, with cases of infection diagnosed in 75% of facilities by the Zebrafish International Resource Center (Matthews *et al.*, 2001; Murray *et al.*, 2011).

However, despite this spike in relevance and interest, to date, less than twenty articles are found in PubMed that provide insights on this parasite. Most of these studies consist on research projects describing *P. neurophilia* lifecycle, method of detection and response of the host following its infection. Gene content and genomic information on this relevant zebrafish parasite remains unknown. The present thesis aimed to tackle this issue by acquiring and performing genome-scale analyses of the first genome draft of *P. neurophilia*. The novel findings presented here are key stones for future genomic, transcriptomic, proteomic, bioinformatics, and host-parasite interaction studies of this parasite that has been infecting the research industry for far too long.

## Summary of novel findings

Acquiring a better understanding of the basic biology of *P. neurophilia* is the first step into the development of treatment and prevention against this parasite that has been affecting biomedical research all around the world. To this end, we acquired and assembled, analyzed the first genome draft of *P. neurophilia*.

## The first genome draft of *P. neurophilia*

The newly assembled genome of *P. neurophilia* is closely related to that of its phylogenetic sister-species *V. culics and T. hominis*. While demonstrating high similarity in genome content with other sequenced microsporidia, a third of *P. neurophilia* predicted genes are specific to these species. Signal peptide cleavage sites are found in 111 of these *P. neurophilia* specific genes, suggesting that these are secreted proteins and potential putative candidate effectors. As microsporidian are intracellular obligate parasite, the identification of putative mode of pathogenicity is a crucial step into the potential development of treatment against microsporidiosis. Here, we demonstrated that the genome of *P. neurophilia* possesses a wide variety of microsporidian proteins hypothesized or known to be involved in such host manipulation mechanisms.

### *P. neurophilia mode of reproduction*

The putative mode of reproduction of microsporidia is currently the source of a debate as no clear evidences have demonstrated the presence of a clonal or sexual mode of reproduction for these parasites. Fortunately, genomic analyses can be used to infer the putative mode of reproduction of members of this group. These methods include demonstrating the presence of

meiosis-related genes (MRG) as well as the level of ploidy of the organism, and show evidence for recombination.

Because microsporidia have undergone a genome compression resulting in the loss of multiple cellular features, the presence of MRG within multiple microsporidian genomes highlights the importance of these genes for the lifecycle of these intracellular parasites. Here, nine MRGs could not be found in *P. neurophilia,* but this apparently severe lack of MRGs is not unusual in the group. Altogether, our analysis of MRGs and their comparison with other species in the group suggests two possible scenarios. One possibility is that these genes are not involved in meiosis, but in other cellular mechanisms such as replication. Such a scenario would imply that microsporidia are clonal organisms that cannot undergo a conventional meiosis. A second, and more likely explanation, is that microsporidia have evolved a more efficient and streamlined meiotic machinery, so that many MRGs present in other eukaryotes are not required anymore to initiate and terminate this sexual process.

The latter hypothesis is supported by evidence of diploidy in *P. neurophila* and many other species of microsporidia. Indeed, our SNP analyses revealed that most SNPs are present at a 0.5 frequency that suggests that *P. neurophilia* is a diploid organism. Additionally, independent Sanger sequencing and K-mer analyses confirmed these findings.

***Putative impact of large inserts on the RNAi and transposable elements amount***

When first sequenced, the genome of microsporidians demonstrated no or low amount of TEs, but it was later found that certain microsporidian with large genomes could harbor a considerable amount of TEs. From these results, it was concluded that TEs are only present and abundant in large microsporidian genomes. Despite this apparent positive correlation, the small

genome of *P. neurophilia* harbors an unusually high amount of TEs. Indeed, a total of 196 ORFs containing conserved domains demonstrating similarities with known TEs domains were found in the zebrafish parasite. This amount of TEs is 100 times more than the amount found in genomes closely related species with similar genome sizes.

Unexpectedly, despite the presence of a high number of TEs, the presence of key component of the RNAi pathways (the RNA-dependent RNA polymerase, argonaute and dicer protein) has been observed in *P. neurophilia*. Indeed, the RNAi pathway can silence or regulate the expression of genes, and has been demonstrated to also silence the expression of TEs. While the functionality of this pathway was confirmed in *N. ceranae,* further studies would provide more insight into the regulatory role of non-coding RNA in microsporidian but also as their potential role in pathogenicity. Comparative genomic analysis of the RNAi component revealed the presence of a 99 bp insert in the PIWI domain of the argonaute protein of *P. neurophilia*. Perhaps, that the presence of this insert in a highly conserved domain of the main component of the RNAi pathway has an impact in the functionality of this pathway and that a correlation could exist between its presence and the unusually high amount of TEs found in *P. neurophilia*. Similar occurrences are found in other *P. neurophilia* genes suggesting that these insert could play direct or indirect roles in biochemical or cellular processes of this parasite. While these inserts seemed to demonstrate no similarities with known introns, RT-PCR or other RNA-based analyses will be crucial to fully understand if these inserts are part of the coding sequence. In any case, these inserts provide novel insight to our current knowledge of the microsporidian gene content and evolution.

## Further implications and directions

The present study provided insights into the basic biology of *P. neurophilia* and paved the way for future analyses that will complement the findings of this thesis. Such studies may focus on the acquisition of more samples from these parasites, as a mean to better understand their population genetics and more of propagation. Recently, Sanders et al. (private communicate) diagnosed microsporidiosis from *P. neurophilia* in several different laboratory fish species, not only in *D. rerio*. Sequencing the *P. neurophilia* genome and transcriptome from different host may reveal how these parasites adapt to different species. Ideally, a significant host-specific variation in genome content and structure would be observed between these parasites. Such study would also provide information on the SNP distribution between and within different hosts (i.e. are the SNPs found in this study common to isolates of *P. neurophilia* infecting all host or are most of them host-specific?). Finally, the acquisition of more samples may complement the genome data we provided, and possibly reveal genomic regions that were missed in the present study. The former genome could be used as a reference genome for new genome assembly and facilitate the assembly of such genomes as well as population-based analyses.

As previously mentioned, obtaining transcriptomic data from this parasite will be essential to better understand the biology and infectious mechanism of this microsporidium. In this study, 3633 putative genes have been predicted to exist in *P. neurophilia*, but their exact function is solely based on sequence homology. Similarly, the effect of the large inserts we identified on gene function is unknown. Transcriptomic data of different life stages of this parasite would provide much needed information on potential roles and functions of some of

these genes and those inserts. Particular attention should be put towards the analyses of putative effector proteins observed in the *P. neurophilia* genome, as these are likely to play a key role in host-pathogen interactions. Finally, RNA sequencing data could be used to validate our genome annotation.

Similarly, deep RNA sequencing could be performed to identify small non-coding RNA in *P. neurophilia*. Recently, six micro-RNA like RNAs were discovered in *N. ceranae*, a microsporidian harboring all three components of the RNAi pathway. Perhaps, because all three components are found in *P. neurophilia*, these non-coding RNAs could very well be present in this microsporidium. If that's the case, this pathway could be exploited as a tool for functional gene analyses via gene silencing of targeted sequences. Within this context, several fungal parasites possess secreted small non-coding RNA that play a role in host-parasite interactions. The identification of such sequences would provide information on *P. neurophilia*'s host-parasite interaction.

## Concluding remarks

My investigation of the genome structure and content of the zebrafish most relevant pathogen in research facilities demonstrated that *P. neurophilia* genome was similar to other microsporidia but demonstrated the presence of several unique features. These included the presence of an unusual amount of transposable element in relatively small microsporidian genome, the presence of large inserts in regions conserved among several organisms and over a thousand putative unique ORFs not found anywhere else. On the other hand, similar to what is observed in its closely related species, *P. neurophilia* genome demonstrates evidence of ploidy, harbors several MRG and putative effector genes (ricin b lectin and secreted putative unique genes) as well as all

three components of the RNAi pathway. As previous microsporidian genomes have provided insights into the basic biology and genome structure of these parasites, the acquisition and analysis of the relevant parasite *P. neurophilia* was of critical importance. Overall, the findings presented in this fundamental research brings us a step closer to understanding *P. neurophilia*'s biology and the host-parasite interaction that it has with the widely used animal model *D. rerio*.

# References

**Ardila-Garcia, A. M. & Fast, N. M. (2012).** Microsporidian Infection in a Free-Living Marine Nematode. *Eukaryot Cell* **11**, 1544–1551.

**Arkhipova, I. & Meselson, M. (2000).** Transposable elements in sexual and ancient asexual taxa. *Proc Natl Acad Sci U S A* **97**, 14473–14477.

**Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Son Pham & other authors**. **(2012).** SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**, 455–477.

**Beznoussenko, G. V., Dolgikh, V. V., Seliverstova, E. V., Semenov, P. B., Tokarev, Y. S., Trucco, A., Micaroni, M., Giandomenico, D. D., Auinger, P. & other authors**. **(2007).** Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. *J Cell Sci* **120**, 1288–1298.

**Biémont, C. (2010a).** A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* **186**, 1085–1093.

**Biémont, C. (2010b).** A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* **186**, 1085–1093.

**Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., Bihan, T. L., Kumar, S., Abreu-Goodger, C., Lear, M. & other authors**. **(2015).** Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat Commun* **5**.

**Cali, A., Kent, M., Sanders, J., Pau, C. & Takvorian, P. M. (2012).** Development, Ultrastructural Pathology, and Taxonomic Revision of the Microsporidial Genus, Pseudoloma and its Type Species Pseudoloma neurophilia, in Skeletal Muscle and Nervous Tissue of Experimentally Infected Zebrafish Danio rerio. *J Eukaryot Microbiol* **59**, 40–48.

**Campbell, S. E., Williams, T. A., Yousuf, A., Soanes, D. M., Paszkiewicz, K. H. & Williams, B. A. P. (2013).** The Genome of Spraguea lophii and the Basis of Host-Microsporidian Interactions. *PLoS Genet* **9**, e1003676.

**Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. (2009).** trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.

**Capella-Gutiérrez, S., Marcet-Houben, M. & Gabaldón, T. (2012).** Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol* **10**, 47.

**Chacko, N. & Lin, X. (2013).** Non-coding RNAs in the development and pathogenesis of eukaryotic microbes. *Appl Microbiol Biotechnol* **97**, 7989–7997.

**Cheng, G., Luo, R., Hu, C., Cao, J. & Jin, Y. (2013).** Deep sequencing-based identification of pathogen-specific microRNAs in the plasma of rabbits infected with Schistosoma japonicum. *Parasitology* **140**, 1751–1761.

**Chen, X., Liang, H., Zhang, J., Zen, K. & Zhang, C.-Y. (2012).** Secreted microRNAs: a new form of intercellular communication. *Trends Cell Biol* **22**, 125–132.

**Chow, F. W., Xue, L. & Kent, M. L. (2015).** Retrospective study of the prevalence of Pseudoloma neurophilia shows male sex bias in zebrafish Danio rerio (Hamilton–Buchanan). *J Fish Dis* n/a–n/a.

**Corradi, N. (2015).** Microsporidia: Eukaryotic Intracellular Parasites Shaped by Gene Loss and Horizontal Gene Transfers. *Annu Rev Microbiol* **69**, 167–183.

**Corradi, N. & Keeling, P. J. (2009).** Microsporidia: a journey through radical taxonomical revisions. *Fungal Biol Rev* **23**, 1–8.

**Corradi, N. & Slamovits, C. H. (2011).** The intriguing nature of microsporidian genomes. *Brief Funct Genomics* **10**, 115–124.

**Corradi, N., Haag, K. L., Pombert, J.-F., Ebert, D. & Keeling, P. J. (2009).** Draft genome sequence of the Daphnia pathogen Octosporea bayeri: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. *Genome Biol* **10**, R106–R106.

**Corradi, N., Pombert, J.-F., Farinelli, L., Didier, E. S. & Keeling, P. J. (2010).** The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. *Nat Commun* **1**, 77–77.

**Cuomo, C. A., Desjardins, C. A., Bakowski, M. A., Goldberg, J., Ma, A. T., Becnel, J. J., Didier, E. S., Fan, L., Heiman, D. I. & other authors**. **(2012).** Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res* **22**, 2478–2488.

**Dang, Y., Yang, Q., Xue, Z. & Liu, Y. (2011).** RNA Interference in Fungi: Pathways, Functions, and Applications ▿. *Eukaryot Cell* **10**, 1148–1155.

**Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2011).** ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165.

**Desjardins, C. A., Sanscrainte, N. D., Goldberg, J. M., Heiman, D., Young, S., Zeng, Q., Madhani, H. D., Becnel, J. J. & Cuomo, C. A. (2015).** Contrasting host–pathogen

interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun* **6**.

**Ding, S.-W. & Voinnet, O. (2007).** Antiviral Immunity Directed by Small RNAs. *Cell* **130**, 413–426.

**Diogo, D., Bouchier, C., d'Enfert, C. & Bougnoux, M.-E. (2009).** Loss of heterozygosity in commensal isolates of the asexual diploid yeast Candida albicans. *Fungal Genet Biol* **46**, 159–168.

**Edgar, R. C. (2004).** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

**Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000).** Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol* **300**, 1005–1016.

**Franzen, C. (2004).** Microsporidia: how can they invade other cells? *Trends Parasitol* **20**, 275–279.

**Franzen, C. & Müller, A. (2001).** Microsporidiosis: human diseases and diagnosis. *Microbes Infect* **3**, 389–400.

**Garcia, L. S. (2002).** Laboratory Identification of the Microsporidia. *J Clin Microbiol* **40**, 1892–1901.

**Ghosh, K., Cappiello, C. D., McBride, S. M., Occi, J. L., Cali, A., Takvorian, P. M., McDonald, T. V. & Weiss, L. M. (2006).** Functional characterization of a putative aquaporin from Encephalitozoon cuniculi, a microsporidia pathogenic to humans. *Int J Parasitol* **36**, 57–62.

**Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C. & other authors**. **(1996).** Life with 6000 genes. *Science* **274**, 546+.

**Goldshmit, Y., Sztal, T. E., Jusuf, P. R., Hall, T. E., Nguyen-Chi, M. & Currie, P. D. (2012).** Fgf-Dependent Glial Cell Bridges Facilitate Spinal Cord Regeneration in Zebrafish. *J Neurosci* **32**, 7477–7492.

**Haag, K. L., James, T. Y., Pombert, J.-F., Larsson, R., Schaer, T. M. M., Refardt, D. & Ebert, D. (2014).** Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proc Natl Acad Sci* **111**, 15480–15485.

**Heinz, E., Williams, T. A., Nakjang, S., Noël, C. J., Swan, D. C., Goldberg, A. V., Harris, S. R., Weinmaier, T., Markert, S. & other authors**. **(2012).** The Genome of the Obligate Intracellular Parasite Trachipleistophora hominis: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. *PLoS Pathog* **8**.

**Higes, M., Meana, A., Bartolome, C., Botias, C. & Martin-Hernandez, R. (2013).** Nosema ceranae (Microsporidia), a controversial 21st century honey bee pathogen. *Environ Microbiol Rep* **5**, 17–29.

**Höck, J. & Meister, G. (2008).** The Argonaute protein family. *Genome Biol* **9**, 210.

**Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K. & other authors**. **(2013).** The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503.

**Huang, Q. & Evans, J. D. (2016).** Identification of microRNA-like small RNAs from fungal parasite Nosema ceranae. *J Invertebr Pathol* **133**, 107–109.

**Izquierdo, F., Castro Hermida, J. A., Fenoy, S., Mezo, M., Gonzalez-Warleta, M. & del Aguila, C. (2011).** Detection of microsporidia in drinking water, wastewater and recreational rivers. *Water Res* **45**, 4837–4843.

**James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., Celio, G., Gueidan, C., Fraker, E. & other authors**. **(2006).** Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818–822.

**James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N. & Stajich, J. E. (2013).** Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Curr Biol* **23**, 1548–1553.

**Jamshidi, S., Tabrizi, A. S., Bahrami, M. & Momtaz, H. (2012).** Microsporidia in household dogs and cats in Iran; a zoonotic concern. *Vet Parasitol* **185**, 121–123.

**Kamaishi, T., Hashimoto, T., Nakamura, Y., Nakamura, F., Murata, S., Okada, N., Okamoto, K., Shimizu, M. & Hasegawa, M. (1996).** Protein phylogeny of translation elongation factor EF-1α suggests microsporidians are extremely ancient eukaryotes. *J Mol Evol* **42**, 257–263.

**Katinka, M. D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P. & other authors**. **(2001).** Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature* **414**, 450–453.

**Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. & other authors**. **(2012).** Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

**Keeling, P. J. (2003).** Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of microsporidia. *Fungal Genet Biol* **38**, 298–309.

**Keeling, P. J. & Fast, and N. M. (2002).** Microsporidia: Biology and Evolution of Highly Reduced Intracellular Parasites. *Annu Rev Microbiol* **56**, 93–116.

**Keeling, P. J., Luker, M. A. & Palmer, J. D. (2000).** Evidence from Beta-Tubulin Phylogeny that Microsporidia Evolved from Within the Fungi. *Mol Biol Evol* **17**, 23–31.

**Keeling, P. J., Corradi, N., Morrison, H. G., Haag, K. L., Ebert, D., Weiss, L. M., Akiyoshi, D. E. & Tzipori, S. (2010).** The Reduced Genome of the Parasitic Microsporidian Enterocytozoon bieneusi Lacks Genes for Core Carbon Metabolism. *Genome Biol Evol* **2**, 304–309.

**Kent, M. L., Harper, C. & Wolf, J. C. (2012).** Documented and Potential Research Impacts of Subclinical Diseases in Zebrafish. *Ilar J* **53**, 126–134.

**Keohane, E. M. & Weiss, L. M. (1998).** Characterization and function of the microsporidian polar tube: a review. *Folia Parasitol (Praha)* **45**, 117–127.

**Keohane, E. M., Takvorian, P. M., Cali, A., Tanowitz, H. B., Wittner, M. & Weiss, L. M. (1996).** Identification of a Microsporidian Polar Tube Protein Reactive Monoclonal Antibody. *J Eukaryot Microbiol* **43**, 26–31.

**Knip, M., Constantin, M. E. & Thordal-Christensen, H. (2014).** Trans-kingdom Cross-Talk: Small RNAs on the Move. *PLoS Genet* **10**, e1004602.

**Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C. & Schlötterer, C. (2011).** PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE* **6**, e15925.

**Lamour, K. H., Mudge, J., Gobena, D., Hurtado-Gonzales, O. P., Schmutz, J., Kuo, A., Miller, N. A., Rice, B. J., Raffaele, S. & other authors**. **(2012).** Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen Phytophthora capsici. *Mol Plant-Microbe Interact MPMI* **25**, 1350–1360.

**Lartillot, N., Lepage, T. & Blanquart, S. (2009).** PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288.

**Lee, S. C., Heitman, J. & Ironside, J. E. (2014).** Sex and the Microsporidia. In *Microsporidia*, pp. 231–243. Edited by L. M. Weiss & J. J. Becnel. John Wiley & Sons, Inc.

**Levin, H. L. & Moran, J. V. (2011).** Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**, 615–627.

**Li, L., Stoeckert, C. J. & Roos, D. S. (2003).** OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* **13**, 2178–2189.

**Lobo, M. L., Xiao, L., Antunes, F. & Matos, O. (2012).** Microsporidia as emerging pathogens and the implication for public health: A 10-year study on HIV-positive and -negative patients. *Int J Parasitol* **42**, 197–205.

**Lyons, E. & Freeling, M. (2008).** How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* **53**, 661–673.

**Mathis, A., Weber, R. & Deplazes, P. (2005).** Zoonotic potential of the microsporidia. *Clin Microbiol Rev* **18**, 423–+.

**Matthews, J. L., Brown, A. M. V., Larison, K., Bishop-Stewart, J. K., Rogers, P. & Kent, M. L. (2001).** Pseudoloma neurophilia n. g., n. sp., a New Microsporidium from the

Central Nervous System of the Zebrafish (Danio rerio). *J Eukaryot Microbiol* **48**, 227–233.

**Ma, Z., Li, C., Pan, G., Li, Z., Han, B., Xu, J., Lan, X., Chen, J., Yang, D. & other authors. (2013).** Genome-Wide Transcriptional Response of Silkworm (Bombyx mori) to Infection by the Microsporidian Nosema bombycis. *PLoS ONE* **8**.

**McCue, A. D., Panda, K., Nuthikattu, S., Choudury, S. G., Thomas, E. N. & Slotkin, R. K. (2015).** ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J* **34**, 20–35.

**Meeker, N. D. & Trede, N. S. (2008).** Immunology and zebrafish: Spawning new models of human disease. *Dev Comp Immunol* **32**, 745–757.

**Murray, K. N., Dreska, M., Nasiadka, A., Rinne, M., Matthews, J. L., Carmichael, C., Bauer, J., Varga, Z. M. & Westerfield, M. (2011).** Transmission, Diagnosis, and Recommendations for Control of Pseudoloma neurophilia Infections in Laboratory Zebrafish (Danio rerio) Facilities. *Comp Med* **61**, 322–329.

**Muszewska, A., Hoffman-Sommer, M. & Grynberg, M. (2011).** LTR Retrotransposons in Fungi. *PLOS ONE* **6**, e29425.

**Nageli, N. (1857).** Uber die neue Krankheit der Seidenraupe und verwandte Organismen. *Bot Z* **15**, 760–761.

**Nakjang, S., Williams, T. A., Heinz, E., Watson, A. K., Foster, P. G., Sendra, K. M., Heaps, S. E., Hirt, R. P. & Martin Embley, T. (2013).** Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. *Genome Biol Evol* **5**, 2285–2303.

**Obbard, D. J., Gordon, K. H. J., Buck, A. H. & Jiggins, F. M. (2009).** The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc B Biol Sci* **364**, 99–115.

**Paldi, N., Glick, E., Oliva, M., Zilberberg, Y., Aubin, L., Pettis, J., Chen, Y. & Evans, J. D. (2010).** Effective Gene Silencing in a Microsporidian Parasite Associated with Honeybee (Apis mellifera) Colony Declines. *Appl Environ Microbiol* **76**, 5960–5964.

**Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., Alaoui, H. El, Biron, D. G., Brasset, E. & other authors**. **(2014a).** Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. *Genome Biol Evol* **6**, 2289–300.

**Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., El Alaoui, H., Biron, D. G., Brasset, E. & other authors**. **(2014b).** Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. *Genome Biol Evol* **6**, 2289–300.

**Pelin, A., Selman, M., Aris-Brosou, S., Farinelli, L. & Corradi, N. (2015).** Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen Nosema ceranae. *Environ Microbiol* n/a–n/a.

**Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011).** SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786.

**Peyretaillade, E., Parisot, N., Polonais, V., Terrat, S., Denonfoux, J., Dugat-Bony, E., Wawrzyniak, I., Biderre-Petit, C., Mahul, A. & other authors**. **(2012).** Annotation of microsporidian genomes using transcriptional signals. *Nat Commun* **3**.

**Raffaele, S. & Kamoun, S. (2012).** Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* **10**, 417–430.

**Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012).** MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* **61**, 539–542.

**Sanders, J. L. & Kent, M. L. (2011).** Development of a sensitive assay for the detection of Pseudoloma neurophilia in laboratory populations of the zebrafish Danio rerio. *Dis Aquat Organ* **96**, 145–156.

**Sanders, J. L. & Kent, M. L. (2014).** The Zebrafish as a Model for Microsporidiosis. In *Microsporidia*, pp. 357–370. Edited by L. M. Weiss & J. J. Becnel. John Wiley & Sons, Inc.

**Sanders, J. L., Watral, V. & Kent, M. L. (2012).** Microsporidiosis in Zebrafish Research Facilities. *Ilar J* **53**, 106–113.

**Sanders, J. L., Watral, V., Clarkson, K. & Kent, M. L. (2013).** Verification of Intraovum Transmission of a Microsporidium of Vertebrates: Pseudoloma neurophilia Infecting the Zebrafish, Danio rerio. *PLoS ONE* **8**.

**Sanders, J. L., Peterson, T. S. & Kent, M. L. (2014).** Early Development and Tissue Distribution of Pseudoloma neurophilia in the Zebrafish, Danio rerio. *J Eukaryot Microbiol* **61**, 238–246.

**Selman, M., Sak, B., Kváč, M., Farinelli, L., Weiss, L. M. & Corradi, N. (2013).** Extremely Reduced Levels of Heterozygosity in the Vertebrate Pathogen Encephalitozoon cuniculi. *Eukaryot Cell* **12**, 496–502.

**Shabalina, S. & Koonin, E. (2008).** Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* **23**, 578–587.

**Sisson, M., Cawker, J., Buske, C. & Gerlai, R. (2006).** Fishing for genes of vertebrate behavior: zebrafish as an upcoming model system. In *Lab Anim*, pp. 33–39.

**Spagnoli, S., Xue, L. & Kent, M. L. (2015a).** The common neural parasite Pseudoloma neurophilia is associated with altered startle response habituation in adult zebrafish (Danio rerio): Implications for the zebrafish as a model organism. *Behav Brain Res* **291**, 351–360.

**Spagnoli, S. T., Xue, L., Murray, K. N., Chow, F. & Kent, M. L. (2015b).** Pseudoloma neurophilia: A Retrospective and Descriptive Study of Nervous System and Muscle Infections, with New Implications for Pathogenesis and Behavioral Phenotypes. *Zebrafish* **12**, 189–201.

**Spence, R., Gerlach, G., Lawrence, C. & Smith, C. (2008).** The behaviour and ecology of the zebrafish, Danio rerio. *Biol Rev* **83**, 13–34.

**Thomson, T. & Lin, H. (2009).** The Biogenesis and Function PIWI Proteins and piRNAs: Progress and Prospect. *Annu Rev Cell Dev Biol* **25**, 355–376.

**Tsaousis, A. D., Kunji, E. R. S., Goldberg, A. V., Lucocq, J. M., Hirt, R. P. & Embley, T. M. (2008).** A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi. *Nature* **453**, 553–556.

**Vávra, J. & Ronny Larsson, J. I. (2014).** Structure of Microsporidia. In *Microsporidia*, pp. 1–70. Edited by L. M. Weiss & J. J. Becnel. John Wiley & Sons, Inc.

**Vinayak, S., Pawlowic, M. C., Sateriale, A., Brooks, C. F., Studstill, C. J., Bar-Peled, Y., Cipriano, M. J. & Striepen, B. (2015).** Genetic modification of the diarrhoeal pathogen Cryptosporidium parvum. *Nature* **523**, 477–480.

**Vossbrinck, C. R. & Debrunner-Vossbrinck, B. A. (2005).** Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol (Praha)* **52**, 131–142; discussion 130.

**Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. (1987).** Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326**, 411–414.

**Watson, A. K., Williams, T. A., Williams, B. A. P., Moore, K. A., Hirt, R. P. & Embley, T. M. (2015).** Transcriptomic profiling of host-parasite interactions in the microsporidian Trachipleistophora hominis. *BMC Genomics* **16**.

**Williams, B. A. P., Hirt, R. P., Lucocq, J. M. & Embley, T. M. (2002).** A mitochondrial remnant in the microsporidian Trachipleistophora hominis. *Nature* **418**, 865–869.

**Wright, S. & Finnegan, D. (2001).** Genome evolution: Sex and the transposable element. *Curr Biol* **11**, R296–R299.

**Wu, S., Zhu, Z., Fu, L., Niu, B. & Li, W. (2011).** WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12**.

**Xu, Y. & Weiss, L. M. (2005).** The microsporidian polar tube: A highly specialised invasion organelle. *Int J Parasitol* **35**, 941–953.