Genomic Analysis of *Encephalitozoon* species

by Mohammed Selman

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 ès science à
L'Institut de biologie d'Ottawa-Carleton

Abstract

Microsporidia are obligate intracellular pathogens of medical and ecological importance whose genomes have been studied extensively over the last decade. Their parasitic lifestyle has lead them to lose a great number of genes and, thus, biochemical pathways capacities, but these reductive processes have been often offset by the acquisition of several genes by means of horizontal gene transfer (HGT). First, in this thesis, we will describe the complete genomes of *Encephalitozoon hellem* and *Encephalitozoon romaleae*. Both species also were found to harbor a number of protein-coding genes absent in other microsporidia, which products assembled complete metabolic pathways. All these genes are functionally related to DNA and folate metabolism, and all appear to have been acquired from HGT events from different eukaryotic and prokaryotic donors. Interestingly in *E. romaleae* genes involved in *de novo* synthesis of folate are all pseudogenes, highlighting the transient nature of transferred genes. Secondly, we took a closer look at the ploidy and sexual status of *Encephalitozoon cuniculi*, a vertebrate pathogen, by mapping Illumina sequence reads against the genomes of four strains of this species. We identified the presence of low level of heterozygosity in all strains investigated; a feature that revealed the diploid nuclear state of the species. This reductive intra-individual genetic diversity could result from the long-term propagation of these strains under laboratory conditions, but we propose that it could also reflect an intrinsic capacity of these vertebrate pathogens to self-reproduce. Overall, the work presented in this thesis resulted in a much greater understanding of the genome evolution of a medically and economically important group of parasites.

Résumé

Les microsporidies sont des pathogènes intracellulaires obligatoires d'importance médicale et écologique dont les génomes ont fait l'objet de nombreuses études au cours de la dernière décennie. Leur mode de vie parasitaire a eu pour conséquence une perte d' un grand nombre de gènes et de mécanisme biochimique, mais ces processus réducteurs ont souvent été compensés par l'acquisition de plusieurs gènes par transfert horizontal (HGT). Dans cette thèse, nous décrirons les génomes séquencés d'*Encephalitozoon hellem* et *Encephalitozoon romaleae*. Les deux espèces contiennent un certain nombre de gènes codant pour des protéines, qui ne se retrouvent chez aucune autre microsporidie. Tous ces gènes sont fonctionnellement liés au métabolisme du folate et de l'ADN, acquis par de multiples événements de transfert de gène indépendants, d'origine eucaryote et procaryote. De plus, chez *E. romaleae*, les gènes impliqués dans la synthèse *de novo* du folate sont tous des pseudo-gènes, ce qui souligne le caractère transitoire des HGTs. Deuxièmement, nous avons examiné de plus près la ploïdie et le mode de reproduction chez *Encephalitozoon cuniculi*, en alignant les séquences, obtenue par la technologie de séquençage à haut débit Illumina, sur le génome de quatre souches d'*E. cuniculi* . Nous avons identifié la présence d'un faible niveau d'hétérozygosité dans toutes les souches étudiées, ce qui montre que l'espèce est diploïde. Cette faible diversité génétique intra-individuelle pourrait résulter de la propagation à long terme de ces souches dans des conditions de laboratoire, mais nous émettons l'hypothèse que cela pourrait aussi refléter une capacité intrinsèque de ces pathogènes à s'auto-reproduire. Dans l'ensemble, le travail présenté dans cette thèse nous a conduit à une meilleure compréhension de l'évolution génomique d'un groupe de parasites médicalement et économiquement important .

Acknowledgments

This thesis would not have been possible without the aid and support of several people. First, I would like to express my sincere and profound gratitude toward Dr. Nicolas Corradi for is guidance, motivation and continuous support during the duration of my Masters. I would also like to extend this thanks to my committee members, Dr.Stéphane Aris-Brosou and Dr. Ashkan Golshani for their helpful advices throughout the course of the program.

I would like to thank the many collaborators for their work and valuable input: Bohumil Sak, Martin Kvác and Louis M. Weiss for providing us with the *E. cuniculi* samples. Jean-François Pombert and Patrick J. Keeling for the sequencing and analysis of *E. hellem*. Fabien Burki for the phylogenic analysis. Laurent Farinelli for the next generation sequencing services provided by FASTERIS. Leellen F. Solter and Douglas W. Whitman for providing us with *E. romaleae* samples. Naomi Fast and Cameron Grisdale for verifying the splicing of the new introns in *Encephalitozoon*. Christina Cuomo and the Broad Institute for the use of unpublished *Encephalitozoon cuniculi* genome data from the "Microsporidian Comparative Genomes Consortium". Soo-Chan Lee and Timothy James for their comments on a previous version of the *Eukaryotic cell* manuscript.

Finally, thanks to all current and former lab members for their ongoing support. Special mention to my friend and colleague, Adrian Pelin, and my little sister, Sara Selman.

Table of Contents

Abstract 2

Résumé 3

Acknowledgments 4

List of Figures 7

List of Tables 7

Chapter 1 : Introduction 8

- Biology of Microsporidia 8
- Cellular biology and molecular biology of Microsporidia 9
- Microsporidia Evolution 11
- Microsporidia Genomics 13
- Goals and objectives 16
- Reference 18

Chapter 2: Microsporidia: Horizontal gene transfers in vicious parasites. 21

- Abstract 22
- Introduction 23
- Role of Horizontal Gene Transfers in the Evolution of Microsporidia 24
- An Animal Gene in a Microsporidia 26
- A Novel HGT Candidate in the Genome of *Encephalitozoon* Romaleae 27
- An Insect Gene in a Parasite of Humans: Underpinning the Frequency of "Host-Switch" in Microsporidian Parasites 30
- Conclusion 31
- Reference 32

Chapter 3: Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. 35

- Abstract 36
- Introduction 37
- Results 38
  - General Characteristics of the *E. hellem* and *E. romaleae* Genomes. 38
  - *E. hellem* and *E. romaleae* Contain Functionally Related Genes That Likely Are Derived from Multiple HGT Events from Different Donors. 43
  - *E. romaleae* Folate Biosynthesis Genes Are Nonfunctional Pseudogenes. 49

Discussion 49
Materials and Methods 53
References 58
Chapter 4: Latest progress in microsporidian genome research. 61
Abstract 62
Main text 63
Smallest Microsporidian Genomes: It is not All about Gene Loss 65
What Makes a Microsporidia a Microsporidia 67
Microsporidian Genomics: Future Directions 69
Reference 71
Chapter 5: Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. 73
Abstract 74
Introduction 75
Material &Methods 76
Results and Discussion 79
Identification of low levels of heterozygosity in four genetically different strains of Encephalitozoon cuniculi 79
Strains of Encephalitozoon are likely to be diploid 88
Opportunistic sex in Encephalitozoon cuniculi? 89
References 93
Chapter 6 : Conclusion 96
Reference 99
Appendix A : Acquisition of an animal gene by microsporidian intracellular parasites. 101
Abstract 102
Main text 103
Supplemental Information 107
Appendix B : Supplemental information for Chapter 3 113
Appendix C : Supplemental information for Chapter 5 123
Appendix D : Copyrights 126

List of Figures

Figure 2.1 Events of horizontal gene transfer (HGT) known to have occurred in microsporidia based on published sequence data. 25

Figure. 3.1. Physical characteristics of *E. hellem*, *E. romaleae*, *E. intestinalis*, and *E. cuniculi* .42

Figure. 3.2. Genomic context of HGT-acquired genes in *E. hellem* and *E. romaleae*. 44

Figure 3.3. Hybrid origin of the folate metabolic pathways in *E. hellem* and *E. romaleae*. 45

Figure 3.4. Bayesian phylogenetic trees of three proteins, PRT (A), GTPCH (B), and FPGS (C), involved in the folate metabolic pathways in *E. hellem* and *E. romaleae* (shown in white on black) but absent from other microsporidians. 48

Figure 4.1. Phylogeny of the newly sequenced microsporidian species based on their respective 16S rRNA and tubulin genes. 64

Figure 5.1. Amount, location and nature of heterozygous loci in *E. cuniculi* and comparison with *Nematocida*. 81

Figure 5.2. A. Selfing, inbreeding and outcrossing in *Encephalitozoon cuniculi.* 91

List of Tables

Table 1.1. Taxonomic classification of Microsporidia over the years.....12

Table 1.2. Microsporidia Genomic 14

Table 2.1.Protein with closest homology to FPGS gene from *E. romaleae* 29

Table 3.1. General characteristics of *Encephalitozoon* genomes 41

Table 5.1. List of heterozygous loci in *E. cuniculi* genotype I, II and III.....82

Table 5.2. Substitutions Variation between two independent isolates of *E. cuniculi* genotype II; EcII and EcII-CZ.....87

Chapter 1 : Introduction

Biology of Microsporidia

The microsporidia phylum is very diverse; it contains more than 170 genera and approximately 1300 species(1). These organisms are ubiquitous and known to infect a wide array of hosts ranging from protists to insects and vertebrates, including humans; they are basically found in all animals(1, 2).

The transmission of microsporidia is still poorly understood, but is thought to be most commonly transmitted by either by ingestion or inhalation. More recently, evidence of possible zoonotic transmission between wild animals, and domestic animals has also been observed (3-6). A number of microsporidian species have been detected in drinking water, wastewater and recreational rivers, hence, microsporidia have also been suggested to represent waterborne pathogen (7-9). Furthermore, soil, sand, compost and fecal exposure as also been found as another environmental sources of microspodian transmissions (10, 11).

The distribution of Microsporidia is truly global, they are found in both developing and developed nations. The first case of human microsporidia infection was first reported in 1985 in a HIV positive patient in the USA(12). From that point on, microsporidia species were studied more extensively, especially those infecting humans. In human, microsporidia can infect most tissue or organs of the body, including skin, lungs, kidney, muscle, intestine, liver, eyes, brain, etc...(13) Microsporidiosis (the disease associated with microsporidian infection) is most

commonly present in immuno-incompetent patients, like AIDS patients, transplant patients, or patients undergoing chemotherapy; but severe infections of apparently healthy individuals have also been observed. Untreated infections in those patients are usually fatal (14). Chronic diarrhea is the most common symptom, but different species invade different tissue, hence the symptoms due to microsporidiosis varies depending on the site of infection. For instance, when it infects the ocular tissues, symptoms can range from eye pain to blurred vision; while brain infections can cause headache or seizures (13, 14).

Cellular biology and molecular biology of Microsporidia

The microspodian cell is quite unique compared to the typical eukaryotic cell. For instance, their cells lack a conventional mitochondria intrigued many scientists for decades. However, we now know, that remains of a mitochondria are present in their cells under a very derived form; called the mitosome (15, 16). The role of the mitosome is unclear, but it was shown to play an important role in the biosynthesis of cytosolic Fe–S proteins, including *Nar1* and *Rli1*, that are essential for the cell survival(16, 17).

Microsporidia have adapted to effectively disperse among hosts and disseminate within their host. Their infection process starts with the transfer of the spore content into the host, via a highly specific host invasion apparatus; the polar tube. The presence of this apparatus is a hallmark of all members of the group. The life cycles of microsporidia in humans consist of three stages: infective, proliferative, and sporogonic. The infective stage, occurs when the spore triggers the injection of its polar tube, which will penetrate a host cell and allow the spore's

content (sporoplasm) to be transferred into the host cell via the polar tube. Following infection, the sporoplasm grows and divides producing many cells (called meronts) within the host-cell cytoplasm (proliferative stage, or merogony) within a parasitous vacuole. During the final stage, sporogony, a thick wall is formed around the spore, which provides resistance to adverse environmental conditions. The production of spores is followed by the lysis of the host cell; and the spores are released to the surroundings. These free mature spores can infect new cells thus continuing the cycle (1, 2, 13).

The ploidy of microsporidia species is still unclear – i.e. whether they are haploid, diploid or even tetraploid, is still unknown – and the sexual state of this phylum is still poorly understood. Are they capable to undergo a meiosis cycle and to mate, or are they purely clonal? Claims of sexual reproduction in microsporidia have been published, but these have been solely based on the presence of sex-related genes in their genomes; including a potential mating locus (18, 19), and many genes required for meiosis (20) .

Microsporidian have also been found to harbor unusual characteristics in some species. For instance mRNA transcripts in *Antonospora locustae* (microsporidia infecting grasshopper) have been shown to be quite large in size, and sometimes capable to encode the sequence of up to four different genes. Importantly, the encoded genes were not always found to be in the same transcriptional direction, suggesting the presence of a very atypical post-transcriptional mechanism in this species (21, 22).

The presence of a upstream motif of the start codon and TATA-like promoters found in *Nosema ceranae* and *Encephalitozoon cuniculi* suggest a that a common regulatory mechanism exists among the microsporidia(23-26).

Microsporidia Evolution

Over the years the taxonomic classification of microsporidia phylum has drastically changed (1), (Table1.1). First described as a yeast-like fungi in the 19th century, the microsporidia were later thought to represent very primitive eukaryote lineages (the Archezoa), that predated the origin a of mitochondria; based on the lack of many typical eukaryotic features (lack of mitochondria, golgi bodies, peroxisomes), and early phylogenies based on rRNA subunits and elongation factor (EF) gene(27-29). Later on, microsporidia were re-classified as a fungi based on a more sophisticated phylogenetic studies, complete gene data (i.e. *Encephalitozoon cuniculi*) (23), and the discovery of microsporidian mitosomes(15). Still, their exact position in the fungal tree remained debated for decades (i.e. sister group to *dikarya*, *chytrid*(30-32) or *zygomycota*(33)), and only recently, microsporidia were shown to represent an the early offshoot of the fungal clade(31, 34), and possibly a member of the recently proposed phylum *Cryptomycota*(31, 35).

Table 1.1. Taxonomic classification of Microsporidia over the years.

Date	Classification of microsporidia	Data based on	Reference
1987	Ancient eukaryotes	Ribosomal RNA sequence	(27)
1996	Ancient eukaryotes	EF-1 alpha phylogeny	(28, 29)
2000	Branches with fungi (with *chytrids*)	Beta-tubulin phylogeny	(32)
2001	Branches with fungi	Genome of *E. cuniculi*	(23)
2003	Branches with fungi (with *Zygomycete*)	Alpha and beta-tubulin phylogeny	(33)
2006	Branches with fungi (with *chytrids*)	6 gene phylogeny	(30)
2008	Evolved from ancestral sexual fungi	Genome structure	(18)
2012	Earliest-diverging clade of sequenced fungi	53 genes phylogeny	(34)
2013	*Rozella* and microsporidia are closely related early fungi	Full Genome, 200 genes phylogeny	(31)

Microsporidia Genomics

The first complete microsporidian genome was sequenced in 2001 (that of the vertebrate pathogen, *Encephalitozoon cuniculi*), and its acquisition revealed the extreme adaptive nature of this parasite. Indeed, the genome was found to be small in size and simplistic in content (approximately 2000 genes in a 2,9 Mb genome) (23). The smallest genome in the group, however, is that of *Encephalitozoon intestinalis,* with a genome size of 2.3Mb; which makes it the smallest eukaryotic genomes currently known (36) (Table 1.1). Not all microsporidian genomes are small, however, and can range 10 fold in size. Nevertheless, all are characterized by a reduced gene content (23, 36-38) (Table 1.1).

The lack of many genes means that many biochemical pathways and key enzymes are absent in these parasites. For instance, enzymes in the Krebs cycle, as well as those involved in the metabolism of amino acids, nucleotides and fatty acids are mostly absent(23, 39). This lack of metabolic capacities is compensated with different ingenious ways to obtain these basic compounds from their host, which microsporidia acquire thanks to several transporters that locate on the cellular membrane(20, 40, 41), (i.e. ATP transporters; which allow them to acquire essential energy and nutrients from their hosts(40)). These transporters appear to have been transferred to the microsporidia genomes by means of horizontal gene transfer (HGT) from a number of prokaryotic donors(40).

Table 1.2. Microsporidia Genomic research .

Date	Specie	Data	Features	Reference
2001	*Encephalitozoon cuniculi*	Full genome	Small reduced genome Mitosome	(23)
2002	*Vittaforma corneae*	Genomic survey	Presence of topoisomerase IV	(42)
2004	*Antonospora locustae*	Genomic survey	Genome compaction	(43)
2008	*Brachiola algerae* & *Edhazardia aedis*	Genomic survey	Genome architecture varies greatly between microsporidia	(44)
2008	*Edhazardia aedis*	EST	Suggests that extreme genome reduction is lineage specific and not typical of all microsporidia	(45)
2009	*Nosema ceranae*	Full genome	Presence of transposable-elements ; Motif upstream of start codons	(24)
2009	*Enterocytozoon bieneusi*	Genomic survey	Genome compaction	(37)
2009	*Octosporea bayeri*	Draft genome	Less biochemically dependent on its host	(46)
2010	*Encephalitozoon intestinalis*	Full genome	Smallest eukaryotic genome	(36)
2012	*Nematocida parisii* & *Nematocida sp1*	Full genome & Transcriptome	Lost the tumor-suppressor gene retinoblastoma Acquired nucleosides transporters Secretion signal in hexokinase Evidence for a sexual cycle	(20)
2012	*Trachipleistophora hominis*	Full genome & Proteome	Loss of metabolic capacity Presence of machinery for RNAi, transposable elements	(41)
2013	*Nosema apis*	Full genome	Similar to *N. ceranae*	(25)
2013	*Nosema bombycis* & *Nosema antheraeae*	Full genome	Genome expansion: Tandem duplications, HGT, segmental duplications, Transposable elements	(47)

A horizontal gene transfer (HGT) is the nonsexual movement of genetic information between two organisms. HGTs are widespread in prokaryotes, however, little is known about them in eukaryotes(48). This prevalence of HGT among prokaryotes is probably due to their prevalence in many ecosystems and their intimate interactions. In some cases, HGTs bring novel function to the recipient of the transferred gene; including the acquisition of drug resistance (49).

In addition to the ATP transporters, other HGTs have also been observed in a few microsporidian lineages. For instance, species in the *Antonospora*, *Nosema* and *Octosporea* genus where found to have acquired genes, which provides them with additional protection for their cells (46, 50, 51). Genes that encodes for superoxide dismutase and catalase were acquired; both genes are critical in the detoxification of cytoplasmic environment. Hence, the sequencing of new microsporidia genomes may reveal new acquisition of HGTs in different species.

Goals and objectives

The present thesis will have a particular interest in *Encephalitozoon* species. To date, a total of four species of microsporidia in the genus *Encephalitozoon* species have been described, which have long been proposed to mainly infect vertebrates. For instance, *E.. cuniculi and E. intestinalis* infect a wide range of mammals, including humans, dogs, and rabbits (23, 36). *Encephalitozoon hellem*, has also been found in humans and a wide variety of birds. Recently, a new species in the genus was isolated from grasshoppers. This species, which was named *Encephalitozoon romaleae* (52), represents the very first *Encephalitozoon* species found to infect an invertebrate. Interestingly, phylogenetic analysis among *Encephalitozoon sp.* showed a close relationship between *E.hellem* and *E.romaleae*, even though the two species infect radically different hosts, suggesting a recent switch in host.

Here, the complete genome sequence of *E.romaleae* will be presented and investigated, in order to determine the genome content and shape of the first members of the genus that is known to infect non-vertebrates (Chapter 3). Comparative genomics and molecular approaches will be used to obtain a better understanding of the genomic composition of this microsporidia species compared to other ones with sequenced genomes, and to obtain a better view of their gene content and potential correlations with the host specificity of this particular parasite. The presence, origin and frequency of horizontal gene transfer in this microsporidian species will also be investigated (Chapter 2 & 3).

Another portion of the thesis aimed to take a closer look at the sexual status of these fungi, as this could provide essential insights into the mechanisms that are used by microsporidian species to propagate, create new genetic information, speciate and perhaps become more pathogenic, since mating appears to be important for the generation of diversity and virulence in other pathogenic fungi . To this end, new isolates of *Encephalitozoon cuniculi* have been acquired, sequenced and analyzed, and their mode of reproduction (sexual and asexual) investigated using bioinformatics techniques (Chapter 5).

Reference

1. Keeling, P.J. and N.M. Fast, *Microsporidia: biology and evolution of highly reduced intracellular parasites.* Annu Rev Microbiol, 2002. **56**: p. 93-116.
2. Vavra, J. and J. Lukes, *Microsporidia and 'the art of living together'.* Adv Parasitol, 2013. **82**: p. 253-319.
3. Dado, D., et al., *Detection of zoonotic intestinal parasites in public parks of Spain. Potential epidemiological role of microsporidia.* Zoonoses Public Health, 2012. **59**(1): p. 23-8.
4. Jamshidi, S., et al., *Microsporidia in household dogs and cats in Iran; a zoonotic concern.* Vet Parasitol, 2012. **185**(2-4): p. 121-3.
5. Mathis, A., R. Weber, and P. Deplazes, *Zoonotic potential of the microsporidia.* Clin Microbiol Rev, 2005. **18**(3): p. 423-45.
6. Sak, B., et al., *Diversity of microsporidia (Fungi: Microsporidia) among captive great apes in European zoos and African sanctuaries: evidence for zoonotic transmission?* Folia Parasitol (Praha), 2011. **58**(2): p. 81-6.
7. Graczyk, T.K., et al., *Bather density and levels of Cryptosporidium, Giardia, and pathogenic microsporidian spores in recreational bathing water.* Parasitol Res, 2007. **101**(6): p. 1729-31.
8. Hoffmann, R., et al., *Natural infection with microsporidian organisms (KW19) in Vannella spp. (Gymnamoebia) isolated from a domestic tap-water supply.* Parasitol Res, 1998. **84**(2): p. 164-6.
9. Slodkowicz-Kowalska, A., et al., *Microsporidian species known to infect humans are present in aquatic birds: implications for transmission via water?* Appl Environ Microbiol, 2006. **72**(7): p. 4540-4.
10. Widmer, G., et al., *Frequent Occurrence of Mixed Enterocytozoon bieneusi Infections in Humans.* Appl Environ Microbiol, 2013. **79**(17): p. 5357-62.
11. Ardila-Garcia, A.M., et al., *Microsporidian Diversity in Soil, Sand, and Compost of the Pacific Northwest.* J Eukaryot Microbiol, 2013.
12. Desportes, I., et al., *Occurrence of a new microsporidan: Enterocytozoon bieneusi n.g., n. sp., in the enterocytes of a human patient with AIDS.* J Protozool, 1985. **32**(2): p. 250-4.
13. Wittner, M. and L.M. Weiss, *The microsporidia and microsporidiosis*. 1999, Washington, D.C.: ASM Press. xvii, 553 p.
14. Didier, E.S., *Microsporidiosis: an emerging and opportunistic infection in humans and animals.* Acta Trop, 2005. **94**(1): p. 61-76.
15. Williams, B.A., et al., *A mitochondrial remnant in the microsporidian Trachipleistophora hominis.* Nature, 2002. **418**(6900): p. 865-9.
16. Burri, L., et al., *Microsporidian mitosomes retain elements of the general mitochondrial targeting system.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15916-20.
17. Goldberg, A.V., et al., *Localization and functionality of microsporidian iron-sulphur cluster assembly proteins.* Nature, 2008. **452**(7187): p. 624-8.
18. Lee, S.C., et al., *Microsporidia evolved from ancestral sexual fungi.* Curr Biol, 2008. **18**(21): p. 1675-9.
19. Lee, S.C., L.M. Weiss, and J. Heitman, *Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer.* Commun Integr Biol, 2009. **2**(5): p. 414-7.
20. Cuomo, C.A., et al., *Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth.* Genome Res, 2012. **22**(12): p. 2478-88.
21. Williams, B.A., et al., *A high frequency of overlapping gene expression in compacted eukaryotic genomes.* Proc Natl Acad Sci U S A, 2005. **102**(31): p. 10936-41.

22. Corradi, N., L. Burri, and P.J. Keeling, *mRNA processing in Antonospora locustae spores.* Mol Genet Genomics, 2008. **280**(6): p. 565-74.
23. Katinka, M.D., et al., *Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi.* Nature, 2001. **414**(6862): p. 450-3.
24. Cornman, R.S., et al., *Genomic analyses of the microsporidian Nosema ceranae, an emergent pathogen of honey bees.* PLoS Pathog, 2009. **5**(6): p. e1000466.
25. Chen, Y., et al., *Genome sequencing and comparative genomics of honey bee microsporidia, Nosema apis reveal novel insights into host-parasite interactions.* BMC Genomics, 2013. **14**: p. 451.
26. Peyretaillade, E., et al., *Identification of transcriptional signals in Encephalitozoon cuniculi widespread among Microsporidia phylum: support for accurate structural genome annotation.* BMC Genomics, 2009. **10**: p. 607.
27. Vossbrinck, C.R., et al., *Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes.* Nature, 1987. **326**(6111): p. 411-4.
28. Kamaishi, T., et al., *Complete nucleotide sequences of the genes encoding translation elongation factors 1 alpha and 2 from a microsporidian parasite, Glugea plecoglossi: implications for the deepest branching of eukaryotes.* J Biochem, 1996. **120**(6): p. 1095-103.
29. Kamaishi, T., et al., *Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes.* J Mol Evol, 1996. **42**(2): p. 257-63.
30. James, T.Y., et al., *Reconstructing the early evolution of Fungi using a six-gene phylogeny.* Nature, 2006. **443**(7113): p. 818-22.
31. James, T.Y., et al., *Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia.* Curr Biol, 2013.
32. Keeling, P.J., M.A. Luker, and J.D. Palmer, *Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi.* Mol Biol Evol, 2000. **17**(1): p. 23-31.
33. Keeling, P.J., *Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia.* Fungal Genet Biol, 2003. **38**(3): p. 298-309.
34. Capella-Gutierrez, S., M. Marcet-Houben, and T. Gabaldon, *Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi.* BMC Biol, 2012. **10**: p. 47.
35. Jones, M.D., et al., *Validation and justification of the phylum name Cryptomycota phyl. nov.* IMA Fungus, 2011. **2**(2): p. 173-5.
36. Corradi, N., et al., *The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis.* Nat Commun, 2010. **1**: p. 77.
37. Akiyoshi, D.E., et al., *Genomic survey of the non-cultivatable opportunistic human pathogen, Enterocytozoon bieneusi.* PLoS Pathog, 2009. **5**(1): p. e1000261.
38. Pombert, J.F., et al., *Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12638-43.
39. Keeling, P.J., et al., *The reduced genome of the parasitic microsporidian Enterocytozoon bieneusi lacks genes for core carbon metabolism.* Genome Biol Evol, 2010. **2**: p. 304-9.
40. Tsaousis, A.D., et al., *A novel route for ATP acquisition by the remnant mitochondria of Encephalitozoon cuniculi.* Nature, 2008. **453**(7194): p. 553-6.
41. Heinz, E., et al., *The genome of the obligate intracellular parasite Trachipleistophora hominis: new insights into microsporidian genome dynamics and reductive evolution.* PLoS Pathog, 2012. **8**(10): p. e1002979.
42. Mittleider, D., et al., *Sequence survey of the genome of the opportunistic microsporidian pathogen, Vittaforma corneae.* J Eukaryot Microbiol, 2002. **49**(5): p. 393-401.

43. Slamovits, C.H., et al., *Genome compaction and stability in microsporidian intracellular parasites.* Curr Biol, 2004. **14**(10): p. 891-6.
44. Williams, B.A., et al., *Genome sequence surveys of Brachiola algerae and Edhazardia aedis reveal microsporidia with low gene densities.* BMC Genomics, 2008. **9**: p. 200.
45. Gill, E.E., J.J. Becnel, and N.M. Fast, *ESTs from the microsporidian Edhazardia aedis.* BMC Genomics, 2008. **9**: p. 296.
46. Corradi, N., et al., *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, 2009. **10**(10): p. R106.
47. Pan, G., et al., *Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation.* BMC Genomics, 2013. **14**: p. 186.
48. Keeling, P.J. and J.D. Palmer, *Horizontal gene transfer in eukaryotic evolution.* Nat Rev Genet, 2008. **9**(8): p. 605-18.
49. Dzidic, S. and V. Bedekovic, *Horizontal gene transfer-emerging multidrug resistance in hospital bacteria.* Acta Pharmacol Sin, 2003. **24**(6): p. 519-26.
50. Xiang, H., et al., *A tandem duplication of manganese superoxide dismutase in Nosema bombycis and its evolutionary origins.* J Mol Evol, 2010. **71**(5-6): p. 401-14.
51. Fast, N.M., et al., *Bacterial catalase in the microsporidian Nosema locustae: implications for microsporidian metabolism and genome evolution.* Eukaryot Cell, 2003. **2**(5): p. 1069-75.
52. Lange, C.E., et al., *A new Encephalitozoon species (Microsporidia) isolated from the lubber grasshopper, Romalea microptera (Beauvois) (Orthoptera: Romaleidae).* J Parasitol, 2009. **95**(4): p. 976-86.

Chapter 2: Microsporidia: Horizontal gene transfers in vicious parasites.

Mohammed Selman[a] , Nicolas Corradi[a]

[a] Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, K1N1H7, Canada

Published in November 2011 in Mobile Genetic Elements. This is a follow up to : Selman, M., *et al*. Acquisition of an animal gene by microsporidian intracellular parasites. Current Biology (See Appendix A), which was published prior to the start of my Masters.

Abstract

Microsporidia are obligate intracellular parasites whose genomes have been shaped by an extreme lifestyle. Specifically, their obligate intracellular parasitism has resulted in the loss of many genes and biochemical pathways, but these reductive processes have been often offset by the acquisition of several genes by means of horizontal gene transfer (HGT). Until recently, these HGTs were all found to have derived from prokaryotic donors, but a recent study suggests that some species took advantage of this mechanism to acquire one gene from an animal, which they maintained in their genome for metabolic purposes. The gene encodes for a purine nucleoside phosphorylase, and shows a strong phylogenetic signal of arthropod origin. Here, we briefly review our current knowledge of HGTs discovered across microsporidian genomes and discuss the implications of the most recent findings in this research area for understanding the origin and evolution of this highly adapted group of intracellular parasites. A novel gene potentially transferred by means of HGT to one microsporidia is also reported.

Introduction

Microsporidia represent a group of ultra-adapted, obligate intracellular parasites that are found to infect virtually every known animal lineage; from worms to humans. Microsporidian cells are very simplistic in form, lacking conventional mitochondria, and harboring an atypical Golgi apparatus and "prokaryote-like" rRNA molecules (rRNA) (1). The presence of these unconventional features in microsporidia have long been thought to reflect their primitive eukaryotic nature (2-5). However, it is now well recognized that these result from their adaptation for an atypical lifestyle, and these curious unicellular organisms are currently best described as representing one of many offshoots of the fungal kingdom (6-10).

The genomes of microsporidia are also compelling mirrors of their specialized way of living—they are small, and simplistic in both form and content. Until now, sequencing efforts have revealed that the genomes of all members of the group are characterized by at most 2,500 genes, which are usually involved in a few, reduced metabolic pathways.11 Obviously, this reduction in many biochemical pathways has resulted in microsporidia being unable to produce a number of cellular compounds, so members of this lineage strongly depend on their host's metabolism for many cytoplasmic supplies; including nucleotides and amino acids.12,13 In the most extreme case, one microsporidia appears to have lost the capacity to produce ATP altogether;14 resulting in its dependence on the host for even the most basic source of energy.

Undergoing massive gene losses may not appear as an ideal solution to evolve a successful lifestyle, but microsporidian parasites have managed to offset the miniaturization of their proteome by acquiring a battery of genomic tools; which they now use to protect themselves from environmental insults and to profit more fully from their hosts.

Role of Horizontal Gene Transfers in the Evolution of Microsporidia

The compelling reductive processes that characterize microsporidia have often overshadowed what's abundant in their genomes. Indeed, all species with surveyed genomes are all surprisingly rich in "transporter" proteins (e.g., ATP transporters, Folate transporters), that are typically used by these parasites to steal metabolites (13-15) and energy from the hosts they invade16 (Figure 2.1A). In the case of ATP/ADP transporters, these appear to have been acquired by microsporidian genomes from a number of donors by means of HGT, and it is now increasingly accepted that such transfers have likely originated from co-infecting bacteria; such as Chlamydia (15,17,18).

HGTs have also affected microsporidian lineages in other ways. For instance, species in genus *Antonospora*, *Nosema* and Octosporea have experienced additional gene transfers that provided a welcomed protection for their cells (15,19,20) One of these genes encodes for a superoxide dismutase that catalyzes the conversion of superoxide anions into oxygen and hydrogen peroxide, while the other encodes for a catalase; a protein responsible for the decomposition of hydrogen peroxide to water and oxygen. Both genes are, therefore, critical in the detoxification of cytoplasmic environment from reactive oxygen species. Interestingly, the catalase is found in the peroxysome in other organisms, an organelle that is lacking from microsporidian cells; and both sequences harbor a strong prokaryotic signal (19,20). Some of these latter genera have also gained protection by acquiring a "photolyase," which is required for the repair of UV-induced DNA damage; an essential tool to avoid cell death as a consequence of mutagenic factors.21 These genes are all absent in more diverged lineages of the group (i.e., *Encephalitozoon* spp), leading to the prediction that spores from *Antonospora*, *Nosema* and Octosporea may be better protected from environmental factors than those of other species (21).

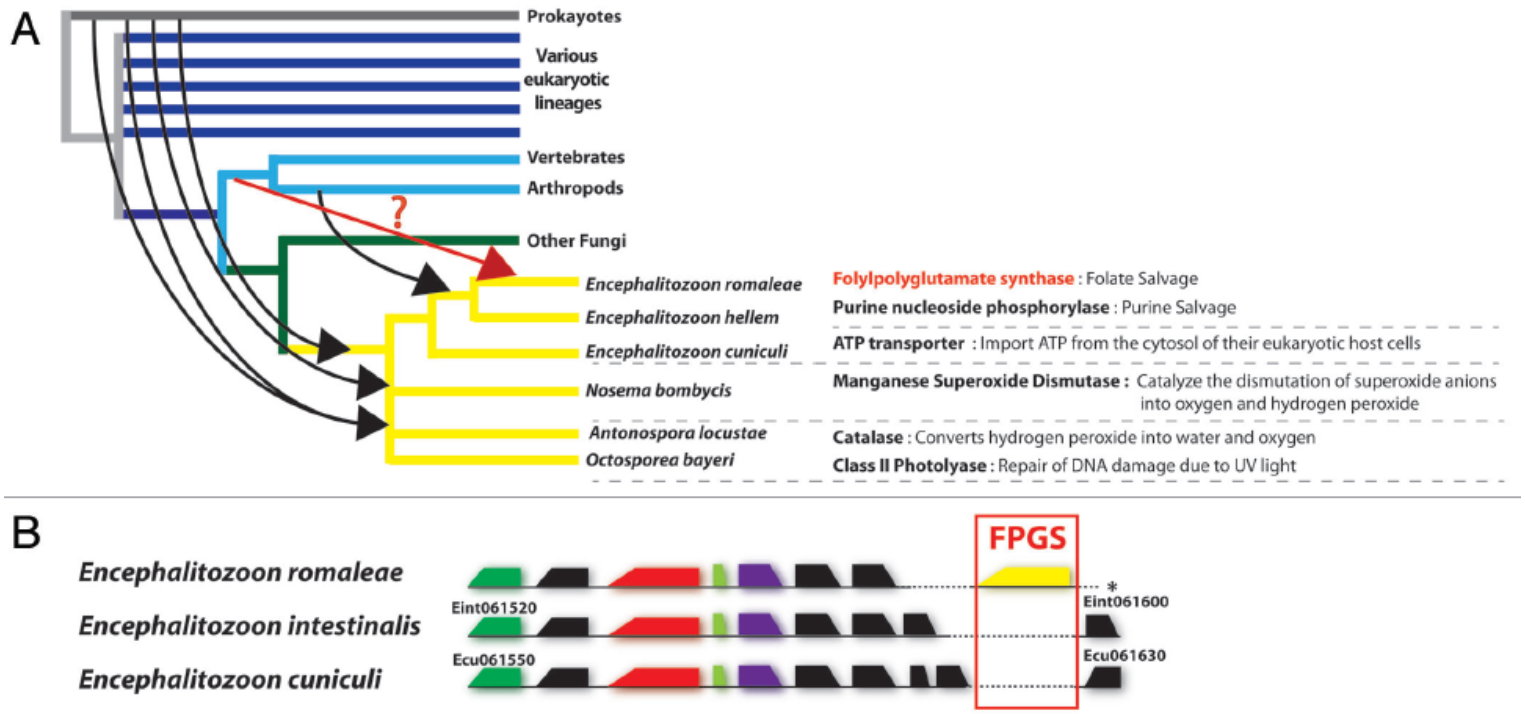


Figure 2.1 (A) Events of horizontal gene transfer (HGT) known to have occurred in microsporidia based on published sequence data. A schematic representation of major evolutionary lineages is shown. Arrows depicts the evolutionary origin (e.g., gene donor) of different HGTs acquired by different microsporidian genera. The red arrow highlights the gene transfer of the FPGS gene, whose origin is currently unclear. (B) Genomic location of the FPGS gene in the genome of *E. romaleae*. Alignment of a homologous region of approximately 15kb region between the from *E. romaleae*, *E. intestinalis* and *E. cuniculi* . This region is located on the chromosome 6 of *E. intestinalis* and *E. cuniculi* . The location of the folylpolyglutamate synthase (FPGS) genes in *E. romaleae* is shown in yellow. Conserved hypothetical proteins are shown as black rectangles, whereas genes encoding for proteins with known functions are shown as colored rectangles. Dashed lines indicate portions of the genome sequence that are missing a particular gene. * end of the *E. romaleae* contig currently available.

An Animal Gene in a Microsporidia

For some time, questions have remained as to whether microsporidia have also acquired genes from more closely related donors and, perhaps, used these HGTs to rebuild broken biochemical pathways. Interestingly, a recent study based on ongoing genome sequencing projects suggests that these processes are likely to have happened in two species.

Inspections of the genome draft of *Encephalitozoon romaleae*, an insect-infecting microsporidia, have revealed the presence of a protein encoding gene that was previously unknown from other members of this group; a purine nucleoside phosphorylase (PNP) (22). This enzyme is known to be involved in the salvage of purines (23-25), so its presence is likely to benefit the metabolic power of *E. romaleae*. Surprisingly, the amino acid sequence of the *E. romaleae* PNP gene did not show significant similarities with homologs from prokaryotes, or from closely related unicellular eukaryotes, but shared instead a strong identity with homologs from arthropods. This unexpected finding was further confirmed using a variety of methods and models for phylogenetic analyses, and resulted in the discovery of the first compelling case of HGT between a microsporidian parasite and an animal (Figure 2.1A). Because *E. romaleae* is an obligate intracellular parasite of insects, this HGT seems better explained if it had derived from its host, rather than from another unknown source. It is also notable that the PNP appears to be somehow amenable to HGTs, since similar gene acquisitions have been reported in other parasites, such as *Giardia lamblia*(26) *Borrelia hermsii*(27) and *Cryptosporidium parvum*(28); although never from eukaryotic donors. Certainly, the unusual frequency at which PNP has been moved across many distantly related lineages suggests that the acquisition of this gene represents a strong selective advantage for many parasites. One may, therefore expect other metabolically relevant genes to have been moved around in the same way.

A Novel HGT Candidate in the Genome of *Encephalitozoon* Romaleae

Interestingly, more recent inspections of the genome of *E. romaleae* in our lab have identified the presence of another HGT candidate. This latter genome has been sequenced using the Illumina technology, and assembled and annotated as described in (22) As for PNP, this gene is absent from other relatives with published genomes(12,13), and encodes for a metabolically important protein; the folylpolyglutamate synthase (FPGS). This protein represents an ATP-dependent enzyme that plays a key role in the salvage of folate and in its cellular retention,29 so its presence in the genome is another strong indication that the metabolism of this species may be more elaborated than that of other microsporidia. The gene is located on a large contig that resembles subtelomeric regions of sibling species (Figure 2.1B), and its presence in the genome was confirmed by PCR and conventional DNA sequencing. Verification of the location of folylpolyglutamate synthase within *E. romaleae* was performed using the following set of primers; 5′-FPGS_F1: GGATCGATGTTCGTGACTAAAAGGGT and FPGS_R1: 5′ TTCCATCTTCAAAGCGCCTTAGATCCT-3′. PCR reactions were performed in 25µl containing a final concentration of 1X EconoTaq® DNA Polymerase (Lucigen, WI, USA), 0.5mM of each primer and 0.3µl of DNA template; and the products were sequenced using conventional Sanger sequencing.

Amino acid sequences of publicly available FPGS's homologs from other taxa were acquired from RefSeq GenBank, ESTdb, as well as from complete eukaryotic genome databases deposited in the Broad institute and DOE Joint Genome Institute databases, and aligned using Muscle 3.7 (30). to carry phylogenetic analyses as described in (22). BLAST searches (Table 2.1) and

phylogenetic analyses all pointed toward an animal origin for this gene but, as opposed to the PNP gene, statistical support for most analyses was low and phylogenetic reconstructions sometimes failed to support the monophyly of natural lineages (e.g., the fungi). So, while the FPGS gene has undoubtedly been acquired by *E. romaleae* by means of HGT, its donor cannot be identified with certainty. Whatever its origin, however, FPGS represents another outstanding addition to the biochemical repertoire of *E. romaleae,* and highlights how remarkably well this species has benefited from these stochastic events compared with any other member of the group. The contig containing the FPGS gene is available in GenBank under the following accession number JN859606.

Table 2.1 Protein with closest homology to FPGS gene from *E. romaleae*

E-value	% Pair wise Identity	Query coverage	Accession	Organism
1.5E-100	0.43	0.9524	NP_998602	*Danio rerio*
1.03E-98	0.432	0.9524	XP_003440209	*Oreochromis niloticus*
9.84E-96	0.427	0.9714	EFX89284	*Daphnia pulex*
1.51E-94	0.4	0.9643	XP_003230625	*Anolis carolinensis*
2.77E-91	0.42	0.969	XP_003440208	*Oreochromis niloticus*
3.89E-91	0.388	0.969	XP_003470799	*Cavia porcellus*
6.21E-91	0.388	0.9714	AAH05484	*Mus musculus*
1.49E-90	0.408	0.9714	XP_001365989	*Monodelphis domestica*
3.6E-90	0.387	0.969	NP_001230938	*Cricetulus griseus*
8.49E-90	0.385	0.9667	NP_001019651	*Bos taurus*
2.58E-89	0.385	0.969	XP_851481	*Canis lupus*
4.68E-89	0.374	0.969	XP_002915221	*Ailuropoda melanoleuca*
5.78E-89	0.42	0.9548	XP_002109877	*Trichoplax adhaerens*
8.69E-89	0.423	0.9619	XP_002430516	*Pediculus humanus*
1E-88	0.395	0.9238	DAA24139	*Bos Taurus*

An Insect Gene in a Parasite of Humans: Underpinning the Frequency of "Host-Switch" in Microsporidian Parasites

The identification of an arthropod gene in *E. romaleae* (PNP) was certainly surprising, but not inconceivable given that this species is an intimate intracellular parasite of insects. More intriguing, however, was the fact that the gene could also be found in the genome of a notorious human pathogen, *Encephalitozoon hellem* (31). So, how can an insect gene move into the genome of parasite of vertebrates? This question was partly answered using phylogenetic analyses, which demonstrated that *E. hellem* and *E. romaleae* are closely related, sister species22,32 sharing a recent common ancestor from which they have likely both acquired the PNP gene. Whether *E. hellem* can also infect insects is currently unknown, but the discovery of an insect gene in this species indicates that "host-switching" must have occurred at one point in this particular lineage. Importantly, infection of both insects and vertebrate hosts by microsporidia has already been documented for other species, e.g., *Anncaliia algerae* (33), *Trachipleistophora hominis* (34) and *Trachipleistophora extenrec* (35); so having an ancestor that could infect both mammals and insects is a likely event.

Perhaps, the ability of some microsporidia to infect drastically different lineages (i.e. arthropods and vertebrates) is correlated with the metabolic complexity of some species. In particular, an increased metabolism may reduce the dependency of intracellular parasites for supplies from their host, facilitating their autonomy and their capacity to invade new environments (e.g., new hosts). This hypothesis is consistent with the recently discovered, improved metabolic capabilities of *E. romaleae* and *E. hellem* (i.e., the FPGS was also recently found in the

unpublished genome of *E. hellem*; J.F. Pombert and Patrick Keeling, personal communication), but genome data from other species that are prone to radical host switches (i.e. *A. algerae*, *T. hominis*, *T. extenrec*) will be required to fully test this hypothesis.

Conclusion

HGT can be very profitable for their recipients, enabling parasites to invade new environments and adapt to conditions prevailing within animals (21),(36-39). Microsporidia seem to have mastered this type of gene acquisitions, having received and maintained a number of genes from different donors by HGT that they now use for protection, or to fuel their metabolism. Microsporidian genomes have been studied for some time, but only recently sequencing efforts have started to focus on species that are not exclusively medically or ecologically relevant. Such efforts just started to pay off with the recent identification of an animal gene in these parasites. This outstanding finding has many implications for the field of evolutionary biology, and may result in an increased interest by many researchers to search for similar events in other, currently overlooked microsporidian species.

Reference

1. Keeling PJ, Fast NM. Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 2002; 56: 93 - 116
2. Cavalier-Smith T. Eukaryote kingdoms: seven or nine?. Biosystems 1981; 14: 461 - 81
3. Cavalier-Smith T. Eukaryotes with no mitochondria. Nature 1987; 326: 332 - 3
4. Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 1987; 326: 411 - 4
5. Brown JR, Doolittle WF. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc Natl Acad Sci USA 1995; 92: 2441 - 5
6. Gill EE, Lee RC, Corradi N, Grisdale CJ, Limpright VO, Keeling PJ. Splicing and transcription differ between spore and intracellular life stages in the parasitic microsporidia. Mol Biol Evol 2010; 27: 1579 - 84
7. Lee SC, Corradi N, Byrnes EJ, Torres-Martinez S, Dietrich FS, Keeling PJ. Microsporidia evolved from ancestral sexual fungi. Curr Biol 2008; 18: 1675 - 9
8. Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ, Heitman J. Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLoS ONE 2010; 5: e10539
9. Corradi N, Keeling PJ. Microsporidia: a journey through radical taxonomical revisions. Fungal Biol Rev 2009; 23: 1 - 8
10. Gill EE, Fast NM. Assessing the microsporidia-fungi relationship: Combined phylogenetic analysis of eight genes. Gene 2006; 375: 103 - 9
11. Corradi N, Slamovits CH. The intrigiung nature of microsporidian genomes. Brief Funct Genomics 2011; 10: 115 - 24
12. Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon* intestinalis. Nat Commun 2010; 1: 77
13. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 2001; 414: 450 - 3
14. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM. The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biol Evol 2010; 2: 304 - 9
15. Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ. 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 2009; 10: R106
16. Tsaousis AD, Kunji ER, Goldberg AV, Lucocq JM, Hirt RP, Embley TM. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature 2008; 453: 553 - 6
17. Lee SC, Weiss LM, Heitman J. Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. Commun Integr Biol 2009; 2: 414 - 7
18. Richards TA, Hirt RP, Williams BA, Embley TM. Horizontal gene transfer and the evolution of parasitic protozoa. Protist 2003; 154: 17 - 32
19. Fast NM, Law JS, Williams BA, Keeling PJ. Bacterial catalase in the microsporidian *Nosema* locustae: implications for microsporidian metabolism and genome evolution. Eukaryot Cell 2003; 2: 1069 - 75

20. Xiang H, Pan G, Vossbrinck CR, Zhang R, Xu J, Li T. A tandem duplication of manganese superoxide dismutase in *Nosema* bombycis and its evolutionary origins. J Mol Evol 2010; 71: 401 - 14
21. Slamovits CH, Keeling PJ. Class II photolyase in a microsporidian intracellular parasite. J Mol Biol 2004; 341: 713 - 21
22. Selman M, Pombert JF, Solter L, Farinelli L, Weiss LM, Keeling P. Acquisition of an animal gene by microsporidian intracellular parasites. Curr Biol 2011; 21: R576 - 7
23. Montgomery JA. Purine nucleoside phosphorylase: a target for drug design. Med Res Rev 1993; 13: 209 - 28
24. Taylor EA, Rinaldo-Matthis A, Li L, Ghanem M, Hazleton KZ, Cassera MB. Anopheles gambiae purine nucleoside phosphorylase: catalysis, structure, and inhibition. Biochemistry 2007; 46: 12405 - 15
25. Bzowska A, Kulikowska E, Shugar D. Purine nucleoside phosphorylases: properties, functions, and clinical aspects. Pharmacol Ther 2000; 88: 349 - 425
26. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 2007; 317: 1921 - 6
27. Barbour AG, Putteet-Driver AD, Bunikis J. Horizontally acquired genes for purine salvage in Borrelia spp. causing relapsing fever. Infect Immun 2005; 73: 6165 - 8
28. Striepen B, Pruijssers AJ, Huang J, Li C, Gubbels MJ, Umejiego NN. Gene transfer in the evolution of parasite nucleotide biosynthesis. Proc Natl Acad Sci USA 2004; 101: 3154 - 9
29. Sun X, Bognar AL, Baker EN, Smith CA. Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase. Proc Natl Acad Sci USA 1998; 95: 6647 - 52
30. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32: 1792 - 7
31. Didier ES, Didier PJ, Friedberg DN, Stenson SM, Orenstein JM, Yee RW. Isolation and characterization of a new human microsporidian, *Encephalitozoon hellem* (n. sp.), from three AIDS patients with keratoconjunctivitis. J Infect Dis 1991; 163: 617 - 21
32. Johny S, Larson TM, Solter LF, Edwards KA, Whitman DW. Phylogenetic characterization of *Encephalitozoon* romaleae (Microsporidia) from a grasshopper host: relationship to *Encephalitozoon* spp. infecting humans. Infect Genet Evol 2009; 9: 189 - 95
33. Coyle CM, Weiss LM, Rhodes LV, Cali A, Takvorian PM, Brown DF. Fatal myositis due to the microsporidian Brachiola algerae, a mosquito pathogen. N Engl J Med 2004; 351: 42 - 7
34. Weidner E, Canning EU, Rutledge CR, Meek CL. Mosquito (Diptera: Culicidae) host compatibility and vector competency for the human myositic parasite *Trachipleistophora hominis* (Phylum Microspora). J Med Entomol 1999; 36: 522 - 5
35. Vávra J, Kamler M, Modry D, Koudela B. Opportunistic nature of the mammalian microsporidia: experimental transmission of Trachipleistophora extenrec (Fungi: Microsporidia) between mammalian and insect hosts. Parasitol Res 2011; 108: 1565 - 73
36. de Koning AP, Brinkman FS, Jones SJ, Keeling PJ. Lateral gene transfer and metabolic adaptation in the human parasite Trichomonas vaginalis. Mol Biol Evol 2000; 17: 1769 - 73
37. Huang J, Mullapudi N, Sicheritz-Ponten T, Kissinger JC. A first glimpse into the pattern and scale of gene transfer in Apicomplexa. Int J Parasitol 2004; 34: 265 - 74
38. Rogers M, Keeling PJ. Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae. J Mol Evol 2004; 58: 367 - 75

39. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. Curr Biol 2006; 16: 1857 - 64

Chapter 3: Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites.

Jean-François Pombert[a,1], Mohammed Selman[b,1], Fabien Burki[a], Floyd T. Bardell[a], Laurent Farinelli[c], Leellen F. Solter[d], Douglas W. Whitman[e], Louis M. Weiss[f,g], Nicolas Corradi[b,2], and Patrick J. Keeling[a,2]

[a] Department of Botany, Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC, V6T 1Z4 Canada;

[b] Department of Biology, Canadian Institute for Advanced Research, University of Ottawa, Ottawa, ON, K1N 1H7 Canada;

[c] Fasteris SA, CH-1228 Plan-les-Ouates, Geneva, Switzerland;

[d] Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820;

[e] School of Biological Sciences, Illinois State University, Normal, IL 61790; and

[f] Division of Parasitology, Department of Pathology, and

[g] Division of Infectious Diseases, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461

[1]J.-F.P. and M.S. contributed equally to this work.

Published in July 2012 in the Proceedings of the National Academy of Sciences.

Abstract

Microsporidia of the genus *Encephalitozoon* are widespread pathogens of animals that harbor the smallest known nuclear genomes. Complete sequences from *Encephalitozoon* intestinalis (2.3 Mbp) and *Encephalitozoon cuniculi* (2.9 Mbp) revealed massive gene losses and reduction of intergenic regions as factors leading to their drastically reduced genome size. However, microsporidian genomes also have gained genes through horizontal gene transfers (HGT), a process that could allow the parasites to exploit their hosts more fully. Here, we describe the complete sequences of two intermediate-sized genomes (2.5 Mbp), from *Encephalitozoon hellem* and *Encephalitozoon* romaleae. Overall, the *E. hellem* and *E. romaleae* genomes are strikingly similar to those of *Encephalitozoon cuniculi* and *Encephalitozoon* intestinalis in both form and content. However, in addition to the expected expansions and contractions of known gene families in subtelomeric regions, both species also were found to harbor a number of protein-coding genes that are not found in any other microsporidian. All these genes are functionally related to the metabolism of folate and purines but appear to have originated by several independent HGT events from different eukaryotic and prokaryotic donors. Surprisingly, the genes are all intact in *E. hellem*, but in *E. romaleae* those involved in de novo synthesis of folate are all pseudogenes. Overall, these data suggest that a recent common ancestor of *E. hellem* and *E. romaleae* assembled a complete metabolic pathway from multiple independent HGT events and that one descendent already is dispensing with much of this new functionality, highlighting the transient nature of transferred genes.

Introduction

Microsporidia are highly derived relatives of fungi that are obligate intracellular parasites of virtually all animal lineages and which lead to a number of economically and medically important diseases, particularly in sericulture and apiculture (1). To date, more than 1,200 microsporidian species have been described, and at least 13 of these species infect humans; many are opportunistic pathogens found in immune-compromised patients (2, 3). The group is distinguished by a number of cellular characteristics, including the presence of a specialized host-invasion apparatus (the polar tube), an unconventional Golgi apparatus, and highly reduced mitochondria called "mitosomes" (4, 5). Many other cellular features common to other eukaryotes are missing; because microsporidia now are recognized as being related to fungi (4), this simplicity is interpreted as extreme reduction that also extends to the molecular and genomic levels. Genome reduction in microsporidia has followed a number of routes, including losses of entire metabolically relevant pathways, the shortening of proteins, and in some species the shrinking of intergenic regions (6). In the most extreme cases these trends have resulted in major effects on cellular function (7) or genome evolution (8).

The smallest (nonorganellar) nuclear genomes currently known are those of microsporidian species in the genus *Encephalitozoon*, making them a model for extreme reductive forces in nuclear genome evolution. Complete genome sequences from two species (9, 10) were found to encode about 2,000 genes making up a reduced set of sometimes simplified molecular and biochemical pathways. Their high degree of host dependence also is reflected in the relatively large number of transporters encoded in these genomes (e.g., ATP transporters), which allow them to acquire essential energy and nutrients from their hosts. Some of these transporters are

thought to have originated by horizontal gene transfer (HGT), possibly from coexisting bacterial pathogens (11), and the recent finding of an animal-derived gene in both *Encephalitozoon* romaleae and *Encephalitozoon hellem* (12) also raised the intriguing possibility that microsporidia can acquire genes from their hosts.

Here we describe the complete nuclear genomic sequences from *E. hellem* and *E. romaleae* to examine the extent of HGT in these lineages. With these two genomes we now have complete sequences for four of the five described species of *Encephalitozoon*, so relatively detailed analyses of gene presence/absence can be performed. Although both genomes are extremely similar in form and content to those of other *Encephalitozoon* species, we identified a number of genes coding for products that are involved in metabolic pathways that either are absent or are substantially reduced in all other members of the group. Overall, these genes make up nearly complete pathways for de novo folate and purine biosynthesis. Their absence in all other sequenced members of the group suggests that these genes have been acquired by HGT. In most cases this hypothesis is supported by robust phylogenetic evidence, suggesting that the genes have been acquired from different donors including both prokaryotes and eukaryotes. Curiously, the majority of these genes now are mutated to render them nonfunctional in *E. romaleae*, suggesting that the gain of function made possible through multiple independent HGT events has been maintained in one branch of this lineage but was transient in another.

Results

General Characteristics of the *E. hellem* and *E. romaleae* Genomes.

E. hellem (American Type Culture Collection (ATCC) 50504) DNA was isolated from spores grown in RK13 (rabbit kidney) cells. *E. romaleae* (SJ-2008) DNA was isolated from spores purified from infected captive male and female *Romalea microptera* grasshoppers. In both cases, total DNA was used for Illumina sequencing, and the resulting reads were assembled de novo. For *E. hellem* this process resulted in an assembly of 2,251,784 bp distributed among 12 contigs (53× average coverage) and for *E. romaleae* resulted in an assembly of 2,138,148 bp distributed among 13 contigs (300× average coverage). In both cases, single contigs corresponded to chromosomes I–VIII, X, and XI, whereas chromosome IX contained one gap in *E. hellem* and two gaps in *E. romaleae* (Figure 3.1). The subtelomeric regions of *Encephalitozoon* chromosomes typically encode the ribosomal RNA (rRNA) operons, and seven of these operons were linked physically to the ends of *E. hellem* chromosome assemblies (compared with five in *Encephalitozoon intestinalis* and three in *Encephalitozoon cuniculi*). Both the *E. hellem* and *E. romaleae* rRNA operon sequences are represented 25-fold in excess of the rest of their respective genomes, suggesting that all their 22-chromosome ends also are likely capped by rRNA operon subtelomeres. We tested whether the gaps in chromosome IX could represent actual fragmentation by long-range PCR from the assembly ends to rRNA operons. No such a link could be demonstrated, and therefore we conclude that this region is problematic for assembly (it encodes multiple paralogous genes) until direct evidence indicates otherwise.

The *E. hellem* and *E. romaleae* genomic GC contents are different from those of *E. intestinalis* and *E. cuniculi* (Table 3.1). This variation is correlated with different codon-use biases among the four species, with *E. romaleae* favoring AT bases and *E. cuniculi* favoring GC bases in third codon positions . This bias is uniform throughout all *Encephalitozoon* genomes with one notable

exception: In all species chromosome I displays an increase of about 1.5–3% in GC content relative to the rest of the genome (Figure 3.1). The four *Encephalitozoon* genomes share a total of 38 introns that are inserted at cognate sites in orthologous genes. This number includes two tRNA introns, 34 previously reported spliceosomal introns (13), and two additional spliceosomal introns identified in the present study (Table S1). The orthologous microsporidian introns are relatively well conserved, with an average nucleotide identity of 72.1% across the four species. Gene order also is extremely conserved across all *Encephalitozoon* genomes, with 1,824 collinear genes located within 55 completely syntenic blocks. Only three of these blocks were found to be inverted or translocated in one or more genomes .

Table 3.1. General characteristics of *Encephalitozoon* genomes

Characteristic	*E. intestinalis* ATCC 50506	*E. hellem* ATCC 50504	*E. romaleae* SJ-2008	*E. cuniculi* GB-M1
Chromosomes (no.)	11	11	11	11
Genome size (Mbp)	2.3	2.5	2.5	2.9
Assembled (Mbp)	2.2	2.3	2.2	2.5
Genome coverage (%)	96	92	88	86
G+C content (%)	41.4	43.4	40.3	47
Gene density (gene/kbp)	0.86	0.86	0.84	0.83
Mean gene length	1,041 bp	1,080 bp	1,061 bp	1,041 bp
Mean intergenic length*	120 bp	124 bp	130 bp	166 bp
Overlapping genes present	Yes	Yes	Yes	Yes
SSU-LSU rRNA genes	22	22†	22†	22
5S rRNA genes	3	3	3	3
tRNAs	46	46	46	46
tRNA synthetases	21	21	21	21
tRNA introns (size)	2 (12, 41 bp)	2 (12, 41 bp)	2 (12, 41 bp)	2 (12, 41 bp)
Splic. introns (size)	36 (23–76 bp)	36 (23–76 bp)	36 (23–76 bp)	36 (23–76 bp)
Predicted ORFs	1,848	1,848	1,835	2,010

LSU, large subunit; SSU, small subunit.

*The values presented here were recalculated after correction of the start methionines; therefore the intergenic lengths differ slightly from the values presented in Corradi et al. (10).

†Minimum estimated from coverage.

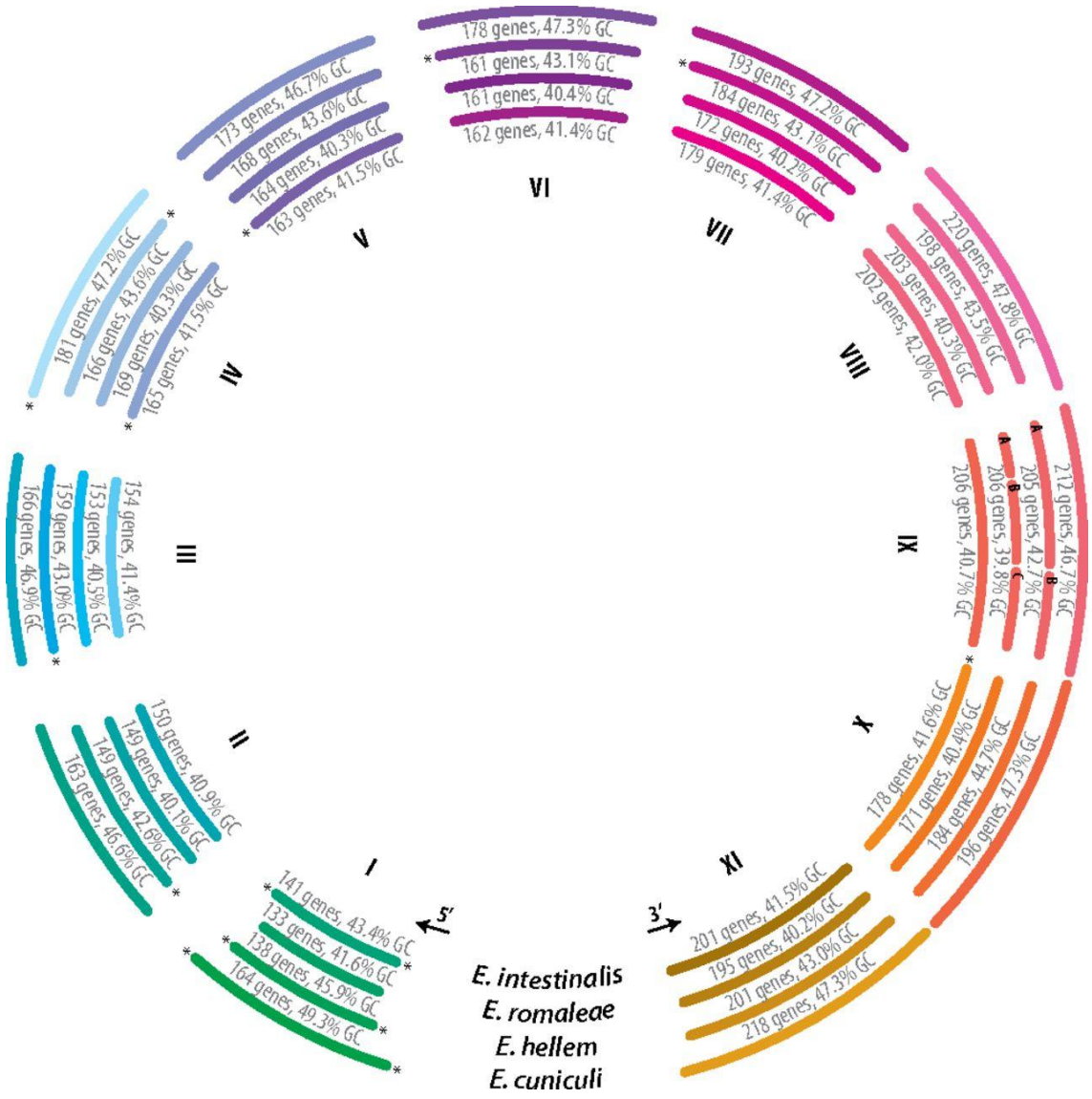


Figure. 3.1 Physical characteristics of *E. hellem*, *E. romaleae*, *E. intestinalis*, and *E. cuniculi* genomes. Chromosomes I–XI are numbered according to their respective sizes in *E. cuniculi* (9) and are shown to scale. Chromosome IX is fragmented in assemblies of *E. hellem* and *E. romaleae* (as indicated by IXa and IXb in *E. hellem* and IXa–IXc in *E. romaleae*), but there is no evidence supporting an actual physical fragmentation of this chromosome in either species.

E. hellem and *E. romaleae* Contain Functionally Related Genes That Likely Are Derived from Multiple HGT Events from Different Donors.

The gene contents of the *E. hellem* and *E. romaleae* genomes are virtually identical to those of *E. intestinalis* and *E. cuniculi* : They contain an identical set of tRNA genes, and only a handful of genes varies between species (Figure S1, Appendix B). However, one suite of genes shared by *E. hellem* and *E. romaleae* stands out. These six genes from four chromosomal regions are absent in all other microsporidia, including other *Encephalitozoon* species (Figure 3.2). They are significant because they are functionally related: Other microsporidia encode a few genes related to folate salvage and purine metabolism, but in *E. hellem* and *E. romaleae* these six genes contribute to making up intact pathways for folate salvage, folate de novo biosynthesis, and purine metabolism (Figure 3.3).

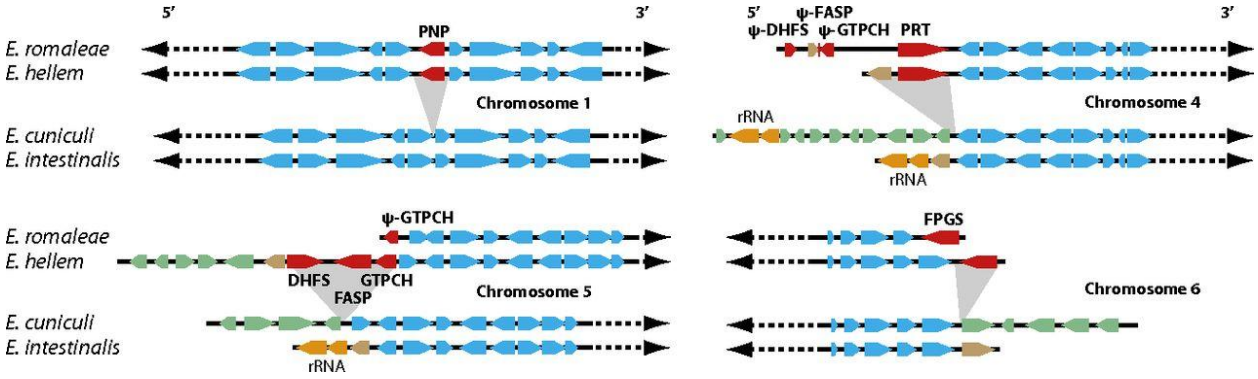


Figure 3.2. Genomic context of HGT-acquired genes in *E. hellem* and *E. romaleae*. The genes in question that are unique to *E. hellem* and *E. romaleae* are shown in red, and syntenic genes that are common to all four completed *Encephalitozoon* genomes are shown in cyan. The rRNA genes are shown in orange, unknown ORFs displaying no synteny are shown in beige, and subtelomeric genes unique to *E. cuniculi* are shown in green. Decayed folate-related pseudogenes in *E. romaleae* are indicated by a ψ-sign.

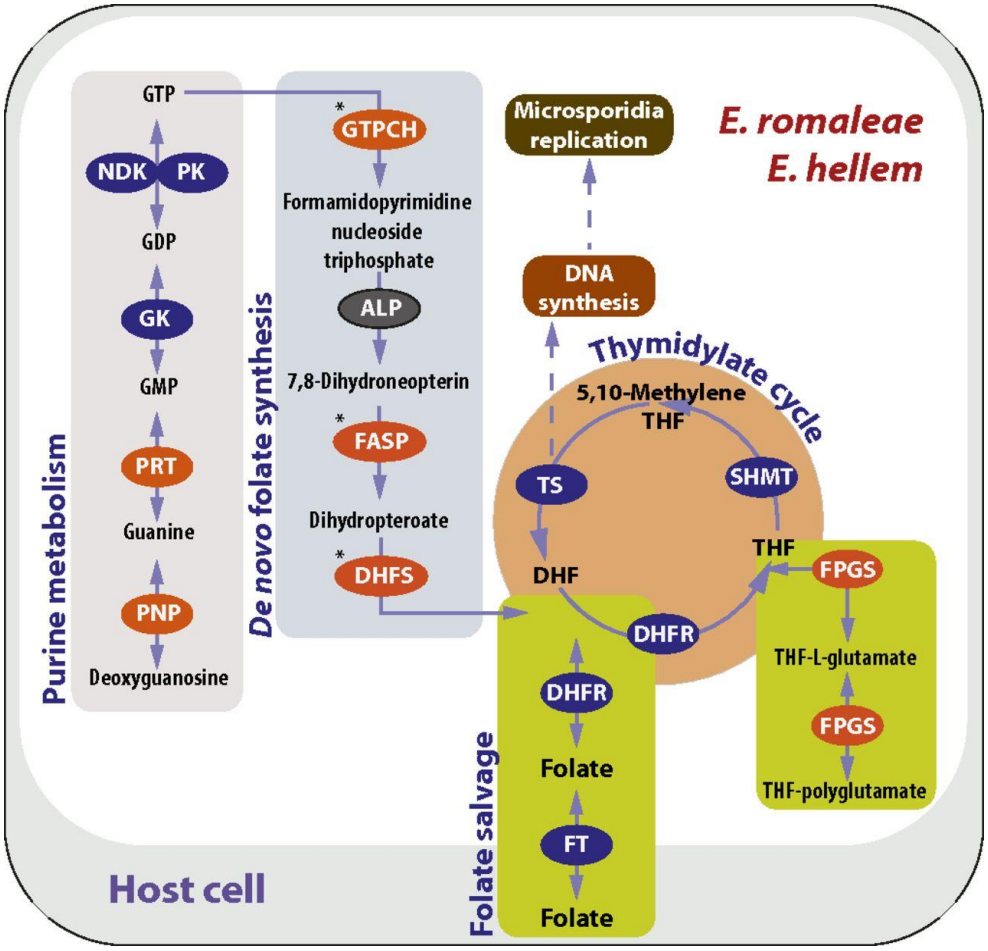


Figure 3.3. Hybrid origin of the folate metabolic pathways in *E. hellem* and *E. romaleae*. Genes vertically inherited in *E. hellem*, *E. romaleae*, *E. intestinalis*, and *E. cuniculi* are shown in blue. Genes acquired by HGT in the lineage leading to *E. hellem* and *E. romaleae* are shown in orange. Functions for which specific genes have not been attributed in *E. hellem* or *E. romaleae* are shown in dark gray. Arrows denote the directions of the enzymatic reactions. Asterisks indicate genes that underwent pseudogenization in *E. romaleae*. FT, folate transporter; GK, guanylate kinase; PK, pyruvate kinase; SHMT, serine hydroxymethyltransferase. Note that the putative FT function attributed to homologs of ECU11_1600 (unique to *Encephalitozoon* species) is uncertain, and that folate transport often is carried out by miscellaneous ATP transporters. The presence of all these genes in the *E. hellem* and *E. romaleae* genomes was confirmed using PCR and sequencing.

The extremely narrow distribution of these genes within microsporidia is highly suggestive that they were acquired by HGT: They are found only in the two closely related sister-species, *E. hellem* and *E. romaleae*, and are absent from all other microsporidia examined to date, including other members of the same genus, *E. intestinalis* and *E. cuniculi* (Figs. 3.2 and 3.3). Phylogenetic analyses support the conclusion that these genes were derived by HGT but, surprisingly, not from a single source. Instead, the genes appear to be derived from multiple donor lineages, including both prokaryotes and eukaryotes (Figure 3.4, Figure S2 Appendix B, and ref. 12). The phylogenetic relationships in these trees frequently are complex or poorly resolved (especially within bacteria), but in several cases the overall position of microsporidia is resolved sufficiently to conclude that their evolutionary histories are not congruent. Specifically, the GTP cyclohydrolase I (GTPCH) is related to *Gammaproteobacteria*, folic acid synthase (FASP) is related to *Firmicutes*, phosphoribosyltransferase (PRT) is related to *spirochetes*, folylpolyglutamate synthase (FPGS) is related to either Metazoa or fungi, dihydrofolate synthase (DHFS) is related to fungi, and a previously reported purine nucleotide phosphorylase (PNP) is related to animals (12). The phylogenetic support for the horizontal acquisition of FPGS and DHFS is weaker, because these genes potentially are related to fungi, and microsporidia also are related to fungi; however, a horizontal acquisition of these genes still is favored by their distribution, because neither has been found in any other microsporidian genome, and their vertical transmission followed by independent losses in all lineages except *E. hellem* and *E. romaleae* would require a large number of independent losses.

In contrast to the diverse phylogenetic origins of the six potentially horizontally transferred genes described above, the phylogenies of all folate- and purine-related genes that are common to other microsporidia are consistent with vertical transmission (Figure S3, Appendix B).

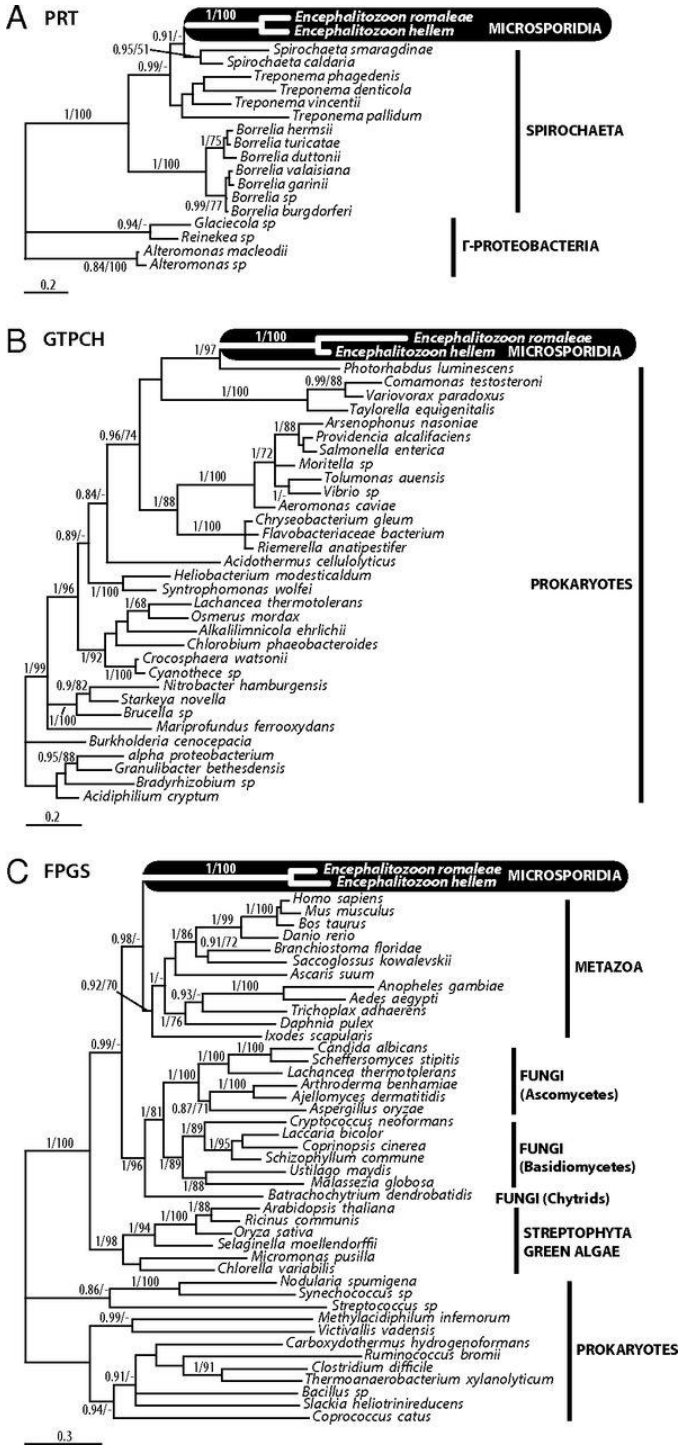


Figure 3.4. Bayesian phylogenetic trees of three proteins, PRT (A), GTPCH (B), and FPGS (C), involved in the folate metabolic pathways in *E. hellem* and *E. romaleae* (shown in white on black) but absent from other microsporidians. Numbers at nodes representing Bayesian posterior probabilities (Left) and bootstrap proportions (Right) are indicated when higher than 0.8 and 70%, respectively. The scale bar corresponds to the estimated number of amino acid substitutions per site. All trees are shown unrooted.

E. romaleae Folate Biosynthesis Genes Are Nonfunctional Pseudogenes.

All six HGT-acquired folate- or purine-related genes are intact and appear functional in *E. hellem* but, surprisingly, not in *E. romaleae*. Indeed, three of the six genes show obvious signs of pseudogenization: GTP cyclohydrolase contains multiple frameshift mutations resulting in premature stop codons, whereas FASP and DHFS both sustained large-scale deletions, all of which were verified by PCR and sequencing. Strikingly, the functional distribution of these genes within the folate and purine pathways is not random: The three genes involved in the salvage of purines and in folate synthesis are all intact, whereas none of the genes functioning in the de novo synthesis of folate are functional, and no intact copy of any of these genes exists elsewhere in the genome.

Discussion

Horizontal gene transfer is known to have had an impact on a variety of eukaryotic genomes and on the functional versatility of their proteomes, even in the otherwise highly reduced and metabolically simple microsporidia (11, 14, 15). In most cases, this impact comes from either a single gene or a single donor, but *E. hellem* and *E. romaleae* apparently have acquired several functionally related genes from several distantly related donors. This mode of acquisition raises several questions about how and why this pathway was assembled and why it then was partially lost in *E. romaleae*.

In most organisms in which they are found, these folate-related pathways play an essential housekeeping role by producing tetrahydrofolate (THF), a compound used as cofactor by thymidylate synthase (TS) for the synthesis of DNA (16). Folate is composed of linked pterin and p-aminobenzoate rings attached to a glutamate moiety and feeds into the one carbon core (C1) metabolism (17–20). Although plants, most fungi, and most protists are capable of synthesizing folate, animals are not, and nearly all microsporidia infect animals. So even if a microsporidian can take up folate from its host (all microsporidia have a single folate transporter gene), infecting a folate-deficient animal could have serious deleterious effects on the parasite, and facultative folate synthesis could be beneficial. All four *Encephalitozoon* species investigated are capable of scavenging folate in the form of THF because of the presence of the ubiquitous dihydrofolate reductase (DHFR) gene and folate transporters (Figure 3.3). Both *E. hellem* and *E. romaleae* also contain FPGS, an enzyme used for folate homeostasis, but apparently only *E. hellem* can synthesize folate from its most basic constituents (e.g., GDP, GMP, and guanine, deoxyguanosine). There is one caveat: We did not identify an alkaline phosphatase (ALP), which removes the pyrophosphate in the second step of the pathway (21). However, other parasites bypass the need for a specific ALP in various ways (22), and it is possible that *E. hellem* does so similarly, using a different enzyme or by using a host ALP.

The de novo biosynthesis of folate requires at least four proteins—GTPCH, FASP, DHFS, and FPGS—in addition to generic ALPs and/or hydrolases. Given the genes common to all *Encephalitozoon* species, the addition of FPGS alone would have a clear beneficial role in folate retention, but the immediate benefits accrued through the acquisition of any other of these genes by itself is not so obvious. The lack of an individual benefit raises questions about how the

pathway could have been assembled from multiple independent HGT events, because a stepwise construction would result in intermediates encoding partial pathways without an obvious functional advantage. One possibility is that the pathway was assembled in some other genome and then was acquired as a whole by a single HGT in the ancestor of *E. hellem* and *E. romaleae*. This possibility limits the otherwise seemingly rare HGT to microsporidia but ultimately does not explain how the pathway was assembled. Alternatively, a complete pathway (e.g., a bacterial operon) may have been acquired by a single HGT in the ancestor of *E. hellem* and *E. romaleae*, and then individual genes were replaced by subsequent HGT events. This process requires no intermediates with partial pathways but requires a lot of HGT to microsporidia. Perhaps the most likely explanation is a combination: A complete pathway existed in a donor lineage in which HGT is relatively common, leading to the replacement of several individual genes, and this mosaic pathway was acquired by the ancestor of *E. hellem* and *E. romaleae* through a single HGT.

Other important questions are why this pathway was partially lost in *E. romaleae* and whether this loss is related to host range. *E. hellem* and *E. romaleae* are closely related sisters, and the overall phylogeny of *Encephalitozoon* suggests that their ancestors were vertebrate parasites (although the arthropod origin of PNP also could suggest that their ancestor infected arthropods). If *E. romaleae* moved to insects recently, it is possible that some aspects of biochemistry of the new host made synthesizing folate less advantageous or, alternatively, that this pathway is less sustainable in insects. For example, if ALP activity is supplied by the host, then a lower or more tissue-specific ALP activity in insects could make the parasite folate synthesis impossible (23, 24). Conversely, the new host environment might have made de novo folate synthesis

unnecessary, leading to its loss. Additional data from *Encephalitozoon* parasites of vertebrate and invertebrate hosts doubtless would be very helpful in distinguishing these possibilities, but expanding the catalog of *Encephalitozoon* genomes will require additional sampling of their natural diversity, because genomes for all four cultivated species are now available.

It also is noteworthy that all but one of these genes is found in subtelomeric regions of *E. hellem* and *E. romaleae*. These poorly studied genomic regions usually evolve faster than the chromosome cores and often are associated with rapid biological innovation, such as parasites coevolving with their hosts' immune systems (25, 26), and in other cases have been observed to be rich in genes derived from HGT (27). Given the compactness of *Encephalitozoon* genomes, it also is more likely that recombination into the chromosome cores would disrupt vital sequence information (e.g., any random insertion is about 10 times more likely to hit a gene than to hit an intergenic region). Subtelomeric regions harbor many repeated genes and gene families, so any one copy can be disrupted with less likelihood of a negative effect. The apparently high rate of recombination between subtelomeres also might generate multiple copies of a new gene and consequently increase its chances of being fixed in the genome. Indeed, the folate-related genes acquired by *E. hellem* and *E. romaleae* that are located in the subtelomeric regions are present at other chromosome ends, as indicated by their 2× to 4× sequencing coverage relative to the single-copy genes from the chromosomal cores; therefore it is likely that they were multiplied in this way following their acquisition by HGT.

Materials and Methods

Cultivation and Collection of *E. hellem* and *E. romaleae* Material. Spores from *E. hellem* (ATCC 50504; originally isolated from humans (28)) were grown in the rabbit kidney fibroblast cell line RK 13 (ATCC CCL-37) with RPMI 1640 (Sigma-Aldrich) supplemented with 5% (vol/vol) foetal bovine serum (FBS), 2 mM L-glutamine, and antibiotics (penicillin, 100 U mL, and streptomycin, 100 μg/mL). T75 flasks were incubated at 37 °C with 5% CO2, and culture medium was replaced two or three times per week. Supernatants containing spores were stored at 4 °C until extraction of DNA. To enrich spores from host-cell debris, the collected culture supernatants were subjected to sequential washes at 400 × g each with distilled H2O, Tris-buffered saline (TBS)-Tween 20 (0.3%), and TBS. The final pellet was resuspended in TBS and mixed with an equal volume of 100% Percoll (Sigma-Aldrich), followed by centrifugation at 400 × g for 45 min at 4 °C. Host-cell debris in the top 75% volume of Percoll was removed. The lower 25% volume of Percoll and the pellet were transferred to a new tube, resuspended in TBS, and washed several times. Because of continued adherence of the host cell (that is, rabbit) nucleic acid to the spores, an additional series of washes was performed with TBS-SDS (0.1%) followed by three washes with TBS.

Spores from *E. romaleae* (SJ-2008; originally isolated from Lubber Grasshopper, *R. microptera* (29)) were produced in infected captive male and female R. microptera grasshoppers. Spores were collected, as previously described (30), from the alimentary canal by homogenizing the midgut and gastric caeca with sterile distilled water. The homogenate was filtered through nylon mesh cloth (sieve size ~200 μm). The collected spores were washed with sterile distilled water five times by centrifugation at 2,700 × g for 10 min and were purified on a 1:1 sterile water:Ludox HS-40 colloidal silica gradient (Sigma-Aldrich), totaling 35 mL, in 50-mL plastic

centrifuge tubes. The purified spores were cleaned again by centrifugation in sterile water. A 1-mL suspension of spores was pelleted by centrifugation, resuspended in 150 mL of buffer (40 mM Tris-Acetate, 1 mM EDTA) in 1.5-mL microfuge tubes, and shaken with 150 mg of 0.5-mm glass beads in a Mini-Beadbeater (Biospec Products). Supernatants containing spores were stored at 4 °C until the extraction of DNA.

DNA Extraction. Genomic DNA for each species was extracted from spores with the MasterPure DNA purification kit (Epicentre). For each sample, spores were pelleted by centrifugation, resuspended in 300 μL of lysis solution (Epicentre) containing Proteinase K, and mixed thoroughly using a vortex. Glass beads (200 μL, 150–212 μm in diameter) were added to the samples, which were incubated immediately at 65 °C for 15 min and bead-beaten at 2,500 rpm in a Mini-Beadbeater (Biospec Products) for 30 s every 5 min. The samples then were cooled to 37 °C and incubated for 30 min at the same temperature with the addition of RNase A (10 μg total) (Epicentre). After treatment with RNase, the samples were placed on ice for 5 min, 150 μL of MPC Protein Precipitation Reagent (Epicentre) was added to each sample, and the solutions were vortexed vigorously for 10 s. Protein debris were pelleted at 4 °C for 10 min at a speed of ≥10,000 × g, and the supernatants were transferred to clean microcentrifuge tubes. DNA then was precipitated using isopropanol, rinsed twice using 70% (vol/vol) ethanol, and finally was suspended in Tris-EDTA buffer.

Genome Sequencing and de Novo Assembly. The *E. hellem* and *E. romaleae* deep-sequencing shotgun libraries, averaging insert sizes of 327 and 337 bp, respectively, were prepared as described by Corradi et al. (10) with Fasteris-modified bar-coded adapters (AAAGT and

ACTTGA, respectively) added to the beginning of the forward and reverse reads for multiplexing. The *E. hellem* library was subjected to two rounds of deep sequencing, each using half a channel on the GA-IIx instrument (Illumina). In the first round, a total of 3,487,666 paired-end reads (101 bp) were generated, resulting in 334,815,936 bp of unique sequences. The 101 bp reads were trimmed to remove the barcodes and were assembled using Velvet 0.7.54 (31) with a hash value of 19, generating a total of 260 contigs with an average size of 8,510 bp and coverage of 53×. The reads were mapped on contigs using the addSolexaReads Perl script from Consed 20 (32), and the internal breaks within the scaffolds were polished using Consed in combination with PCR, cloning and Sanger sequencing. Subtelomeric regions were linked to the chromosome internal segments by PCR as described by Corradi et al. (10). A second round of sequencing (38-bp-long reads, 36,754,350 reads, 1,212,893,550 bp total) was performed in parallel to resolve potentially ambiguous regions from the first round. The resulting reads were mapped on the *E. hellem* contigs with Consed, and the ambiguous regions were curated manually. De novo assemblies also were performed independently with Ray 1.4.0 (33) using a k-mer of 21 and reads from both sequencing rounds for cross-validation. The *E. romaleae* library was subjected to one round of deep sequencing using half a channel of the GA-IIx instrument (Illumina), resulting in 1,901,807,856 bp of unique DNA sequence. Reads were assembled using Velvet with a hash value of 27, resulting in 165 scaffolds with an average size of 13,355 bp and an average coverage of 40×. The resulting contigs were polished as described for the *E. hellem* genome.

Genome Annotation and Analysis. ORFs in the *E. hellem* and *E. romaleae* genomes were predicted using Artemis 12.0 (34) and were identified by BLAST homology searches (35) against the National Center for Biotechnology Information nonredundant database; tRNAs were mapped on the chromosomes with tRNAscan-SE 1.21 (36). The start codons of the *E. hellem*, *E.*

intestinalis, and *E. cuniculi* ORFs were assessed/reassessed by (i) orthologous alignment between the three *Encephalitozoon* species, (ii) cross-checking with the predicted translation initiation codons from Peyretaillade et al. (37), and (iii) adding the intron positions identified in Lee et al. (13). The start codons in the later-sequenced *E. romaleae* genome were annotated afterward. The introns previously identified in the *E. cuniculi* and *E. intestinalis* genomes were positioned on the *E. hellem* and *E. romaleae* genomes with DREG from the EMBOSS 6.3.1 package (38). Putative novel introns were searched for with DREG using the regular expression GTA(AG)GT(ACGT){5,30}TT(ACGT){0,3}AG derived from the microsporidian introns reported in Lee et al. (13) and were curated manually. Nucleotide sequence identity between orthologous introns inserted at the cognate site was calculated from their L-INS-i alignment computed with MAFFT 6.847b (39). The two additional introns reported in this study were ascertained to be spliced at the mRNA level in the sister species *E. cuniculi* using the 5′Race dataset from Lee et al. (13).

The overall GC content of each chromosome was determined using Artemis built-in tools. Codon use tables for the complete set of ORFs from each *Encephalitozoon* species were calculated with Artemis. Metabolic pathways were investigated using the KEGG PATHWAY database (40) and a biochemical atlas (41). The *E. hellem* and *E. romaleae* chromosomes are available in GenBank under the individual accession numbers CP002713–CP002724 and CP003518–CP003530, respectively.

Phylogenetic Analyses. Homologs of the *E. hellem* and *E. romaleae* proteins were identified by BLASTP searches against GenBank and were retrieved and automatically aligned with the L-

INS-i method of the MAFFT package (39). Poorly aligned positions were eliminated with Gblocks 0.91b (42), with half the gapped positions allowed, the minimum number of sequences for a conserved and a flank position set to 50% of the number of taxa plus one, the maximum of contiguous nonconserved positions set to 12, and the minimum length of a block set to 4, followed by manual inspection of the alignments using SeaView 4 (43). Bayesian analyses using the WAG +Γ +F model (four gamma categories) were performed with MrBayes 3.2 (44). Each inference consisted of two independent runs starting from a random tree and four Metropolis-coupled Markov Chain Monte Carlo (MCMCMC), initially for 1,000,000 generations with sampling every 100 generations. The average SD of split frequencies was used to assess the convergence of the two runs after the initial 1,000,000 generations (< 0.01) and proved sufficient in all genes except nucleoside-diphosphate kinase (NDK) and TS, for which 1,000,000 additional generations were required to reach convergence. Bayesian posterior probabilities were calculated from the majority rule consensus of the tree sampled after the initial burn-in period, which corresponded to 25% of the total generations. Maximum Likelihood (ML) analyses were performed using RAxML 7.2.8 (45), with the rapid hill-climbing algorithm and the LG +Γ +F model of evolution (-m PROTGAMMALGF, four discrete rate categories). The best-scoring ML trees were determined in multiple searches using 20 randomized, stepwise-addition, parsimony starting trees. Statistical support was evaluated with nonparametric bootstrapping using 100 replicates.

References

1. Texier C, Vidau C, Viguès B, El Alaoui H, Delbac F (2010) Microsporidia: A model for minimal parasite-host interactions. Curr Opin Microbiol 13:443–449.
2. Weber R, Bryan RT (1994) Microsporidial infections in immunodeficient and immunocompetent patients. Clin Infect Dis 19:517–521.
3. Corradi N, Slamovits CH (2011) The intriguing nature of microsporidian genomes. Brief Funct Genomics 10:115–124.
4. Keeling PJ (2009) Five questions about microsporidia. PLoS Pathog 5:e1000489. CrossRefMedline
5. Didier ES, Weiss LM (2008) Overview of microsporidia and microsporidiosis. Protistology 5(4):243–255. Search Google Scholar
6. Peyretaillade E, et al. (2011) Extreme reduction and compaction of microsporidian genomes. Res Microbiol 162:598–606. MedlineWeb of Science
7. Keeling PJ, et al. (2010) The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biol Evol 2:304–309.
8. Corradi N, Gangaeva A, Keeling PJ (2008) Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. Genomics 91:388–393.
9. Katinka MD, et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414:450–453. CrossRefMedline
10. Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon* intestinalis. Nat Commun 1:77. Medline
11. Lee SC, Weiss LM, Heitman J (2009) Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. Commun Integr Biol 2:414–417. CrossRefMedline
12. Selman M, et al. (2011) Acquisition of an animal gene by microsporidian intracellular parasites. Curr Biol 21:R576–R577. CrossRefMedline
13. Lee RCH, Gill EE, Roy SW, Fast NM (2010) Constrained intron structures in a microsporidian. Mol Biol Evol 27:1979–1982.
14. Tsaousis AD, et al. (2008) A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature 453:553–556.
15. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605–618.
16. Carreras CW, Santi DV (1995) The catalytic mechanism and structure of thymidylate synthase. Annu Rev Biochem 64:721–762.
17. Henderson GB, Huennekens FM (1986) Membrane-associated folate transport proteins. Methods Enzymol 122:260–269. MedlineWeb of Science

18. Cossins EA, Chen L (1997) Folates and one-carbon metabolism in plants and fungi. Phytochemistry 45:437–452.
19. Hyde JE (2005) Exploring the folate pathway in Plasmodium falciparum. Acta Trop 94:191–206. MedlineWeb of Science
20. Bermingham A, Derrick JP (2002) The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery. Bioessays 24:637–648.
21. Suzuki Y, Brown GM (1974) The biosynthesis of folic acid. J Biol Chem 249:2405–2410.
22. Hyde JE, et al. (2008) Plasmodium falciparum: A paradigm for alternative folate biosynthesis in diverse microorganisms? Trends Parasitol 24:502–508.
23. Jurat-Fuentes JL, et al. (2011) Reduced levels of membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to Cry toxins from Bacillus thuringiensis. PLoS ONE 6:e17606. CrossRefMedline
24. Eguchi M (1995) Alkaline phosphatase isozymes in insects and comparison with mammalian enzyme. Comp Biochem Physiol B Biochem Mol Biol 111:151–162.
25. Hall AR, Scanlan PD, Morgan AD, Buckling A (2011) Host-parasite coevolutionary arms races give way to fluctuating selection. Ecol Lett 14:635–642. CrossRefMedline
26. Allen DE, Little TJ (2009) Exploring the molecular landscape of host-parasite coevolution. Cold Spring Harb Symp Quant Biol 74:169–176.
27. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. Science 320:1210–1213.
28. Didier ES, et al. (1991) Isolation and characterization of a new human microsporidian, *Encephalitozoon hellem* (n. sp.), from three AIDS patients with keratoconjunctivitis. J Infect Dis 163:617–621.
29. Lange CE, Johny S, Baker MD, Whitman DW, Solter LF (2009) A new *Encephalitozoon* species (Microsporidia) isolated from the Lubber grasshopper, Romalea microptera (Beauvois) (Orthoptera: Romaleidae) J Parasitol 95:976–986.
30. Johny S, Larson TM, Solter LF, Edwards KA, Whitman DW (2009) Phylogenetic characterization of *Encephalitozoon* romaleae (Microsporidia) from a grasshopper host: Relationship to *Encephalitozoon* spp. infecting humans. Infect Genet Evol 9:189–195. CrossRefMedline
31. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829.
32. Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. Genome Res 8:195–202.
33. Boisvert S, Laviolette F, Corbeil J (2010) Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17:1519–1533. CrossRefMedline
34. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. Bioinformatics 16:944–945.

35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410.
36. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.
37. Peyretaillade E, et al. (2009) Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: Support for accurate structural genome annotation. BMC Genomics 10:607. CrossRefMedline
38. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276–277.
39. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26:1899–1900.
40. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38(Database issue):D355–D360.
41. Michal G ed. (1999) Biochemical Pathways. An Atlas of Biochemistry and Molecular Biology (John Wiley and Sons, New York) 277 pp.
42. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552.
43. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224.
44. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542.
45. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Chapter 4: Latest progress in microsporidian genome research.

Nicolas Corradi[a] , Mohammed Selman[a]

[a] Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, K1N1H7, Canada

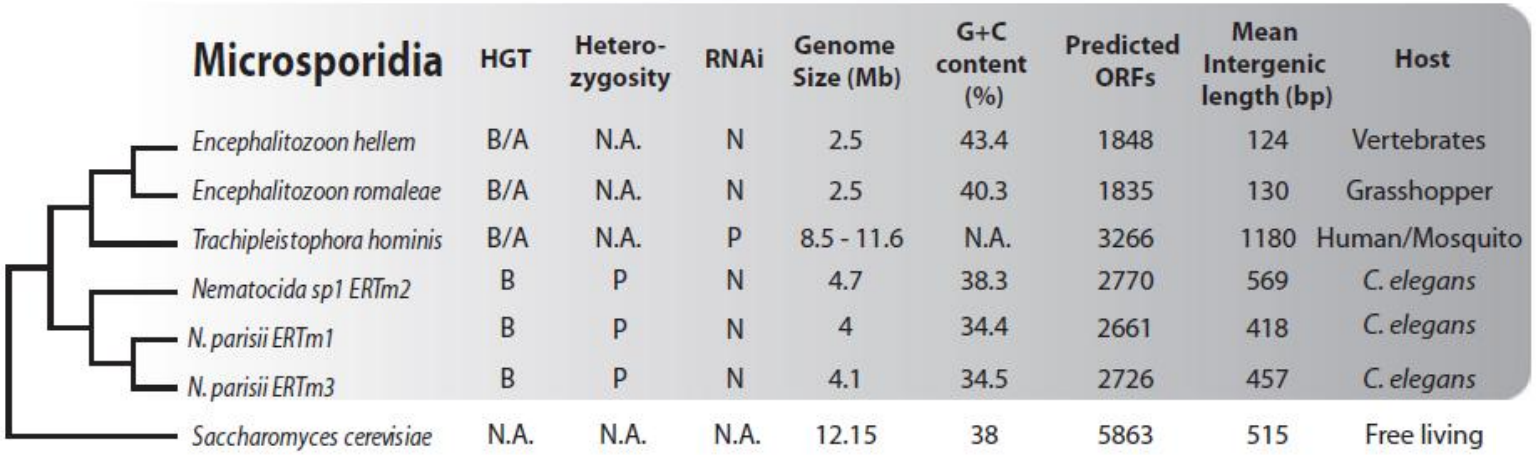
Published in February 2013 in the Journal of Eukaryotic Microbiology.

Abstract

Microsporidia are obligate intracellular pathogens of medical and ecological importance whose genomes have been studied extensively over the last decade. Such studies have focused on the remarkably reduced gene content that characterizes all known species, and some have unraveled the mechanisms that are involved in their extreme genome compaction. In the last year, a large number of new genome sequences from several divergent members of the group have been finally released and analyzed, and these have revealed the presence of many features that were previously unsuspected to exist within the group. This study aims to shortly review the most recent progress in the field of microsporidian genomics, highlighting the importance of the most recently released genome data for our understanding of the biology and evolution of this important group of parasites.

Main text

Microsporidia are “poster-children” of reduction at all levels. The phylum Microsporidia is composed of a diverse group of obligate intracellular parasites, which all share reduced cellular features and a highly effective host invasion apparatus; called the polar tube. Species in this group are known to infect a wide range of eukaryotes, including protists and virtually all known animal lineages, and their extreme parasitic intracellular lifestyle has resulted in an overall simplification of the content and form of their cells (5, 14, 23). For instance, these organisms are known to lack a conventional mitochondrion and Golgi apparatus, and their ribosomal RNAs are considerably reduced in size compared with those of most eukaryotes. Not surprisingly, these cellular features were long thought to represent the primitive nature of Microsporidia (23), but all recent phylogenetic analyses have pointed toward a close affiliation of these parasites with the fungal Kingdom (6, 13, 15); with most recent studies suggesting that Microsporidia represents the earliest branch of the fungal tree of life (4, 9) (Figure 4.1).



Microsporidia	HGT	Hetero-zygosity	RNAi	Genome Size (Mb)	G+C content (%)	Predicted ORFs	Mean Intergenic length (bp)	Host
Encephalitozoon hellem	B/A	N.A.	N	2.5	43.4	1848	124	Vertebrates
Encephalitozoon romaleae	B/A	N.A.	N	2.5	40.3	1835	130	Grasshopper
Trachipleistophora hominis	B/A	N.A.	P	8.5 - 11.6	N.A.	3266	1180	Human/Mosquito
Nematocida sp1 ERTm2	B	P	N	4.7	38.3	2770	569	*C. elegans*
N. parisii ERTm1	B	P	N	4	34.4	2661	418	*C. elegans*
N. parisii ERTm3	B	P	N	4.1	34.5	2726	457	*C. elegans*
Saccharomyces cerevisiae	N.A.	N.A.	N.A.	12.15	38	5863	515	Free living

Figure 4.2 Phylogeny of the newly sequenced microsporidian species based on their respective 16S rRNA and tubulin genes. The presence of HGTs (B—Bacterial; A—Animal origin), heterozygosity (P—Present), and RNAi machinery (P—Present, N—Absent) is shown. Known genome size, GC content, predicted ORF, intergenic length, and parasite's host are also indicated.

Smallest Microsporidian Genomes: It is not All about Gene Loss

Microsporidia have been long distinguished for their genome reduction, but recent large-scale genome studies have also provided new views on the evolutionary processes that have shaped the genomes of these parasites. In particular, the complete sequences of several Microsporidia have revealed the presence of many genes that were gained by means of horizontal gene transfers (HGT). This evolutionary mechanism has been known to be an integral part of the evolution of microsporidian species for quite some time, but its frequency and diversity within the group has long been shadowed by a lack of genome sequence data from poorly studied species.

Horizontal gene transfers and rise of new metabolic pathways in derived Microsporidia. In the last couple of years, compelling events of HGTs have started to emerge from the genomes of four distantly related species with variable genomes sizes, including those of *Encephalitozoon* romaleae, its sister-species *E. hellem*, *Trachipleistophora hominis*, and the nematode pathogen *Nematocida parisii* (9, 10, 18, 19, 20). Among these genomes, those of *E. romaleae* and *E. hellem* were found to harbor the most notable and widespread HGTs, which involved genes whose products are essential for the metabolism of purines and folate. These two pathways are functionally linked and essential in other organisms, but they are otherwise disrupted in other Microsporidia. In addition to their relevance for the metabolism of these two species, these HGTs also stood out because of their surprising evolutionary origin. Specifically, some of these genes were found to have been transferred from animals, while others, although functionally related, were found to have been brought into their genomes from bacteria or from other fungi; ultimately resulting in "mosaic" pathways that are fully functional (18).

Among these HGTs, those of animal origin were particularly interesting for two main reasons. First, their presence in the genome of *E. romaleae* and *E. hellem* suggests the capacity of some Microsporidia to acquire genes directly from their hosts (i.e., *E. romaleae* infects an insect), and these events were previously unknown to occur between eukaryotic hosts and their eukaryotic intracellular parasites. Secondly, the presence of insect-related genes in *E. hellem*, a sister-species to *E. romaleae* that is currently only known to infect vertebrates, underpins the capacity of some Microsporidia to infect hosts that are evolutionary unrelated. Specifically, the presence of an insect gene in a vertebrate parasite suggests that the common ancestor of *E. romaleae* and *E. hellem* must have jumped between an arthropods and a vertebrate at one point in time. Interestingly, *E. hellem* is notorious for infecting many species of birds; so, it is literally possible that this host switch has occurred through avian means.

Additional examples of horizontal gene transfers in the Microsporidia. The notion that radical host switches have been frequent throughout the evolution of Microsporidia is also highlighted by the recent discovery of a transposable element of insect origin in the genome of the human pathogen, *T. hominis* (10); a microsporidium that is capable of infecting both humans and mosquitoes.

Other potential HGTs in Microsporidia discovered within the last year include a nucleoside H+ symporter that is shared among all members of the group, and which plays a role in the acquisition of nucleosides from the surrounding environment (i.e., the host), as well as an

endomembrane metalloprotease that typically provides resistance against bacteriocin-mediated attacks in bacteria (9, 10). Both proteins are missing from other opistokhonts, so their presence in the Microsporidia may represent yet another compelling example of how microsporidia can take advantage of events of HGTs to use their host and fight against environmental pressures more efficiently.

What Makes a Microsporidia a Microsporidia

In addition to providing important insights into the evolution of genome structure, the most recent explorations of microsporidian genomes have also allowed the identification of those biochemical pathways that are shared among all members of this group, i.e., the Microsporidia-specific proteome. This was done by comparing the gene set of several species with variable genome sizes (i.e., *Nematocida* spp., A. locustae, *E. cuniculi* , *E. intestinalis*, *Vittaforma cornea*, *Enterocytozoon bieneusi*, *Vavraia culicis*, and *Antonospora locustae* (2, 9), an analysis that resulted in the identification of 882 orthologous genes that are essential for the survival of all Microsporidia with sequenced genomes (9).

Secreted hexokinases in the Microsporidia. Among these latter gene families, the one encoding for hexokinases stood out for harboring a "secretion" signal that is not present in any homolog from other eukaryotes, including other parasites, but this signal was found sufficient to provoke the secretion of these enzymes in orthologous systems (i.e., *S. cerevisiae*). The presence of such signals in this gene family is highly surprising, because these enzymes are only known to

function in glycolysis. This latter pathway is universally found among Microsporidia, and the hexokinases are specifically required for the production of glucose-6-phosphate that ultimately enters the trehalose and pentose phosphate pathways. As a result, these enzymes play a pivotal role in the production of energy in all Microsporidia parasites, and it is difficult to conceive why these parasites would bother secreting hexokinases when they could use them to propel their own energy production.

One hypothesis that has been put forward suggests that these secreted enzymes could be directed toward the glycolysis of the host, triggering the production of more energy (under the form of ATP); a mechanism that would ultimately benefit the Microsporidia themselves (i.e., Microsporidia can pump ATP from their surrounding cytoplasmic environment) (9). If true, this mechanism would represent an evolutionary innovation of outstanding importance, and its existence in these vicious parasites is currently well supported by a number of in vitro experiments performed with appropriate negative and positive controls (9).

Loss of the Retinoblastoma tumor repressor gene and the presence of RNA interference in the Microsporidia. Furthermore, the annotation of some microsporidian genomes, most notably those of *Nematocida* spp. and *T. hominis*, provides additional insights into the biology of these parasites by revealing, respectively, the loss of one gene essential for the control of the cell cycle in most eukaryotes and the presence of cellular machineries that were previously not known to exist in microsporidian parasites. The gene found to be absent is commonly referred to as “the retinoblastoma tumor repressor gene” (RB), which produces a protein involved in the inhibition

of cell proliferation. This gene is found in almost all eukaryotes (i.e., it has been replaced by analogous proteins in some lineages), so its absence in all microsporidians is notable. It is currently being speculated that the loss of RB in the microsporidian lineage could trigger the very rapid proliferation of these parasites that is characteristic of the group, and ultimately result in the notoriously extreme nucleotide sequence divergence of their genomes (9).

In parallel, the annotation of the *T. hominis* genome reveals other surprises, including the presence of genes that function in RNA interference, such as the Argonaute or DICER genes. Mapping their presence along the microsporidian phylogeny suggests their early evolution within the group, but also demonstrates that a number of lineages, including those of important vertebrate pathogens (i.e., *Encephalitozoon* spp.), have lost them altogether. Although the exact role of RNA interference in *T. hominis* is currently unknown, its identification in microsporidian parasites opens up interesting new ways to use these organisms as models for parasite biology in a broad sense, and could provide new opportunities to fight against these re-emergent pathogens (10; 16).

Microsporidian Genomics: Future Directions

Within the least 2 yr, the wealth of sequence data that has been acquired on Microsporidia has dramatically increased our knowledge of their overall biology (i.e., presence of RNA interference, absence of essential cell cycle components, (9, 10) and evolutionary origin (i.e., identification of Microsporidia as the earliest fungal lineage (4), and have also revealed the

unsuspected metabolic malleability of even the most reduced genomes of the group (i.e., acquisition of new pathways by HGT in the smallest eukaryotic genomes known (18, 20). These data, however, have yet to be fully exploited to understand other aspects of the microsporidian life cycle that are still rather obscure, including their overall genetic diversity within and across their natural populations, or the mode of reproduction and the ploidy level that characterize many species of the group.

The latter two aspects have started to be addressed by some researchers, for instance, by exploring the genome of three strains of the microspoiridium *Nematocida parisii* for the presence of heterozygosity (9). Interestingly, these latter investigations have revealed the presence of a large number of heterozygous sites in all strains investigated, a compelling genetic signature of diploidy in *Nematocida* spp. Furthermore, the number of heterozygous sites was found to be quite elevated, reaching a total of 42,175 for one strain, and these levels are generally linked with the presence of a sexual cycle (3, 11). It remains unknown, however, whether the presence of diploidy is widespread across the group, and if extensive heterozygosity is a feature that is common to all microsporidian parasites. Consequently, similar investigations along the genomes of more derived members of the group are certainly warranted, particularly for those species that are known to infect humans and a number of economically important vertebrates and invertebrates worldwide (i.e., *Encephalitozoon* spp., *Nosema cerenae*). Reference genomes from several strains are currently available for some of these latter species; a feature that will turn out to be very handy when performing targeted genetic studies of their natural diversity, by using next-generation sequencing or other approaches.

Reference

1. Akiyoshi D. E., Morrison H. G., Lei S., Feng X., Zhang Q., Corradi N., Mayanja H., Tumwine J. K., Keeling P. J., Weiss L. M. & Tzipori S. 2009. Genomic survey of the non-cultivatable opportunistic human pathogen. *Enterocytozoon bieneusi*. PLoS Pathog., 5:e1000261.
2. Aurrecoechea C., Brestelli J., Brunk B. P., Fischer S., Gajria B., Gao X., Gingle A., Grant G., Harb O. S., Heiges M., Innamorato F., Iodice J., Kissinger J. C., Kraemer E. T., Li W., Miller J. A., Nayak V., Pennington C., Pinney D. F., Roos D. S., Ross C., Srinivasamoorthy G., Stoeckert C. J. Jr, Thibodeau R., Treatman C. & Wang H. 2010. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res., 38:D415–D419.
3. Balloux F., Amos W. & Coulson T. 2004. Does heterozygosity estimate inbreeding in real populations? Mol. Ecol., 13:3021–3031.
4. Capella-Gutierrez S., Marcet-Houben M. & Gabaldon T. 2012. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol., 10:47.
5. Corradi N. & Slamovits C. H. 2011. The intriguing nature of microsporidian genomes. Brief. Funct. Genomics, 10:115–124.
6. Corradi N. & Keeling P. J. 2009. Microsporidia: a journey through radical taxonomical revisions. Fungal Biol. Rev., 23:1–8.
7. 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.
8. Corradi N., Akiyoshi D. E., Morrison H. G., Feng X., Weiss L. M., Tzipori S. & Keeling P. J. 2007. Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antonospora locustae* and *Enterocytozoon bieneusi*. PLoS ONE, 2:e1277.
9. 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., Levin J. Z., Young S., Zeng Q. & Troemel E. R. 2012. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res., 22:2478–2488
10. Heinz E., Williams T. A., Nakjang S., Noel C. J., Swan D. C., Goldberg A. V., Harris S. R., Weinmaier T., Markert S., Becher D., Bernhardt J., Dagan T., Hacker C., Lucocq J. M., Schweder T., Rattei T., Hall N., Hirt R. P. & Embley T. M. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathog., 8:e1002979.
11. Heitman J. 2010. Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host Microbe, 8:86–99.
12. Katinka M. D., Duprat S., Cornillot E., Metenier G., Thomarat F., Prensier G., Barbe V., Peyretaillade E., Brottier P., Wincker P., Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M., Weissenbach J. & Vivares C. P. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature, 414:450–453.
13. Keeling P. J. 2003. Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genet. Biol., 38:298–309.

14. Keeling P. J. & Fast N. M. 2002. Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol., 56:93–116.
15. Lee S. C., Corradi N., Byrnes E. J. 3rd, Torres-Martinez S., Dietrich F. S., Keeling P. J. & Heitman J. 2008. Microsporidia evolved from ancestral sexual fungi. Curr. Biol., 18:1675–1679.
16. 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.
17. Peyretaillade E., Parisot N., Polonais V., Terrat S., Denonfoux J., Dugat-Bony E., Wawrzyniak I., Biderre-Petit C., Mahul A., Rimour S., Goncalves O., Bornes S., Delbac F., Chebance B., Duprat S., Samson G., Katinka M., Weissenbach J., Wincker P. & Peyret P. 2012. Annotation of microsporidian genomes using transcriptional signals. Nat. Commun., 3:1137.
18. Pombert J. F., Selman M., Burki F., Bardell F. T., Farinelli L., Solter L. F., Whitman D. W., Weiss L. M., Corradi N. & Keeling P. J. 2012. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. Proc. Natl. Acad. Sci. U.S.A., 109:12638–12643.
19. Selman M. & Corradi N. 2011. Microsporidia: horizontal gene transfers in vicious parasites. Mob. Genet. Elements, 1:251–255.
20. Selman M., Pombert J. F., Solter L., Farinelli L., Weiss L. M., Keeling P. & Corradi N. 2011. Acquisition of an animal gene by microsporidian intracellular parasites. Curr. Biol., 21:R576–R577.
21. Slamovits C. H., Fast N. M., Law J. S. & Keeling P. J. 2004. Genome compaction and stability in microsporidian intracellular parasites. Curr. Biol., 14:891–896.
22. Texier C., Vidau C., Vigues B., El Alaoui H. & Delbac F. 2010. Microsporidia: a model for minimal parasite-host interactions. Curr. Opin. Microbiol., 13:443–449.
23. Vossbrinck C. R., Maddox J. V., Friedman S., Debrunner-Vossbrinck B. A. & Woese C. R. 1987. Ribosomal RNA sequence suggests microspo4ridia are extremely ancient eukaryotes. Nature, 326:411–414

Chapter 5: Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*.

Mohammed Selman [a], Bohumil Sak [b], Martin Kvác [b], Laurent Farinelli [c], Louis M. Weiss [d] and Nicolas Corradi [a]

[a] Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, ON, K1N 6N5, Canada

[b] Institute of Parasitology, Biology Centre of the Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic.

[c] FASTERIS S.A., Ch. du Pont-du-Centenaire 109, PO Box 28, CH-1228 Plan-les-Ouates, Switzerland

[d] Department of Pathology and Medicine, Albert Einstein College of Medicine, Bronx, New York, USA

Published in April 2013 in Eukaryotic Cell.

Abstract

The genomes of microsporidia in the genus *Encephalitozoon* have been extensively studied for their minimalistic features, but they have seldom been used to investigate basic characteristics of their biology; such as their ploidy or their mode of reproduction. In the present study, we aimed to tackle this issue by mapping Illumina sequence reads against the genomes of four strains of *E. cuniculi* . This approach, combined with more conventional molecular biology techniques, resulted in the identification of heterozygosity in all strains investigated; a typical signature of a diploid nuclear state. In sharp contrast with similar studies recently performed on a distant microsporidian lineage (*Nematocida* spp.), the level of heterozygosity we identified across the *E. cuniculi* genomes was found to be extremely low. This reductive intra-individual genetic variation could result from the long-term propagation of these strains under laboratory conditions, but we propose that it could also reflect an intrinsic capacity of these vertebrate pathogens to self-reproduce.

Introduction

Microsporidia are a medically and economically important group of obligate intracellular parasites that are known to infect a wide array of hosts, including protists, invertebrates, humans and many other vertebrates. The cells of these eukaryotic organisms are atypical in that they lack conventional mitochondria and Golgi apparatus, and harbor prokaryote-like ribosomal RNA molecules (1). For some time, these simple cellular features were thought to reflect the primitive nature of microsporidian cells (2), but all recent phylogenetic studies now strongly support the association of microsporidia with the fungal kingdom - with current scientific disputes focusing on their exact phylogenetic placement within the kingdom itself (i.e. earliest branch *vs.* more derived lineage; (3-7)).

Besides their intriguing evolutionary origin, microsporidia have drawn scientific attention for their reductive genome features. Specifically, their obligate intracellular parasitic lifestyle has resulted in dramatic gene losses in the biochemical repertoires of all known species, making them ideal models to study the processes involved in genome adaptation to obligate intracellular parasitism (8-12). The most compelling examples of genome reduction in microsporidia are known from species in the genus *Encephalitozoon,* where gene losses have been paralleled by a massive compression of the genomes themselves (13). To date, complete genomes from four species in the genus *Encephalitozoon* have been sequenced (i.e. *Encephalitozoon cuniculi*, *E. intestinalis*, *E. hellem* and *E. romaleae;* (11-13)), all of which have been found to be strikingly small and very similar in size, form and content; encoding approximately 1900 genes with a limited metabolic capacity. The few differences that subsist between these four genomes are notable, however, and include a significant variability in the number of genes acquired by means of horizontal gene transfer (HGT) (11, 14, 15).

To date, the acquisition of genome data from several *Encephalitozoon* spp. have helped us understand how these intracellular parasites have managed to offset their reduced metabolism, and ultimately benefited more efficiently from their hosts. However, this sequence information could have been better exploited to comprehend other aspects of their biology that are still poorly understood but vital for their survival and propagation; including their mode of reproduction or the ploidy of their nuclei (16, 17). In particular, claims of sexual reproduction in *Encephalitozoon spp.* have so far only been based on the presence of sexually-related genomic regions (i.e. meiosis genes, sex-related locus (12, 18-20)), and the presence of a diploid state for their nuclei has been only inferred from limited and unrelated observations (e.g. one heterozygous allele in *E. cuniculi* , and potential homologous chromosomes detected by pulse gel electrophoresis; (12, 18-20).

Here, we confirm the presence of a diploid state and the potential for sexual reproduction in the vertebrate pathogen *Encephalitozoon cuniculi,* by identifying the presence of heterozygosity along the re-sequenced genomes of four strains of this species. Similar investigations recently performed on distant microsporidian relatives with small genomes (i.e. *Nematocida* spp.) have resulted in the identification of extensive heterozygosity (3); a typical hallmark of diploidy that also suggests the presence of sex (3, 21, 22). However, it remains, to be seen if diploidy and elevated heterozygosity are genetic features that are also present in more derived members of the group.

Material &Methods

Genome acquisition, sequence analysis and polymorphisms detection

Strains of *E. cuniculi* are generally classified within three genotypes based on the number of "GTTT" repeats found within the internal transcribed spacer (ITS) of their RNA operon (23). In the present paper, these isolates are designated as Ec-I (rabbit isolate, ATCC 50503, (24)), -II (mouse isolate, (25) , and –III (canine isolate, ATCC 50502,(26)). 100 bp-long Illumina sequence reads from representatives of these genotypes have been mapped against their respective reference genomes to identify the presence of intra-individual genetic variation. In total, 7,805,454; 1,898,812; 13,704,622 reads were found to map against the reference genomes of ECI, II, and III; respectively. Three of these Illumina read sets have been obtained from the NCBI SRA database (Accession numbers SRX002289, SRX002287, SRX002285; sequenced at the Broad Institute), and mapped against their respective reference genomes available from http://microsporidiadb.org/micro/; (27, 28) and NCBI (EcI: AEWD00000000; EcII: AEWQ0100000 and EcIII: AEWR00000000). Genome assembly procedures for the strains ECI, II and III are provided elsewhere (28)). In parallel, reads were also newly obtained from one *E. cuniculi* strain of genotype II isolated from Czech Republic (referred herein as EcII-CZ (29)), and deposited in the NCBI SRA database under the following accession numbers (SRA059442).

The mapping procedure was performed as follows: low quality reads were either removed or trimmed prior to alignment in order to eliminate biases induced by sequencing errors. Variation was subsequently searched using the “Find Variations/SNPs" function available in the Geneious software package. Heterozygosity was scored in all cases where reference and alternate alleles were found to map one genomic region at a similar frequency (in between 35% to 65% respective frequency; n loci = 89). All these heterozygous loci were verified using PCR, followed by direct Sanger sequencing of the resulting products and manual inspection of the sequencing chromatograms (Supplemental Figure 1). We found that heterozygous loci located

within highly covered regions (i.e. within the 35/65% cutoff, but within regions with coverage twice as large as the average coverage of the genome) all represented false positives following PCR and direct Sanger sequencing (i.e. variation was due to paralogy rather than allelic variation). For this reason, these covered genome regions were discarded for downstream analyses (i.e. no evidence of heterozygosity *in-vitro;* loci tested = 23; data not shown). These highly covered regions typically include the subtelomeres of *Encephalitozoon* spp and the rRNA operons of these strains. Finally, variation that deviated from the 35/65% "cutoff" *in-silico* was always found to be the result of sequencing or assembly errors following PCR and Sanger sequencing (no evidence of heterozygosity both *in-silico* and *in vitro;* loci tested = 14; data not shown).

PCR reactions were performed in 25µl containing a final concentration of 1X EconoTaq® DNA Polymerase (Lucigen, WI, USA), 0.5mM of each primer, 0.3µl of DNA. The thermal cycling conditions included an initial step of 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, and final step of 72°C for 12 min.

E. cuniculi II-CZ genome acquisition and annotation

Spores from one strain of *E. cuniculi* isolated from a wild mouse in the Czech Republic (29) were grown in VERO E6 cell line, and the resulting spores subjected to deep-sequencing to further confirm the presence of reduced heterozygosity in *Encephalitozoon cuniculi*. The strain belongs to the genotype II, as suggested by a specific tandem repeat present in its internal transcribed spacer (23) and is referred in the present paper as the strain EcII-CZ. Both strains of genotype II investigated in the present study (i.e. ECII, and ECII-CZ) have been originally isolated from mice within a close geographical region in Czech Republic (25, 29), and our genome analyses confirmed that both isolates indeed represent two very closely related strains of

the same genotype. However, the exact time at which ECII-CZ has been isolated and the overall history of their propagation under laboratory environment are currently unknown (29). For these reasons, we will refrain to speculate about their evolutionary relationship, and will exclusively use their genome sequences to confirm the methodology we have used to detect heterozygous sites (i.e. same methodology used on two independently acquired genomes of one genotype lead to very similar results).

Genomic DNA was extracted from EcII-CZ spores and deep-sequencing shotgun libraries were prepared as previously described (13). The EcII-CZ library was subjected to one round of deep sequencing using $1/10^{th}$ of a channel on the Illumina GA-IIx instrument, resulting in 331'449'750 bp of unique DNA sequence (50 bp-long sequence reads). Reads were assembled using the reassembly tool available in the Geneious package and the EcII genome as a reference (AEWQ0100000), resulting in 54 scaffolds, with an average size of 13,355 bp and an average coverage of 46X. A total of 2,224,725 Reads mapped against this newly acquired genome sequence using the same procedure described above for other strains. The EcII-CZ reference genome has been annotated as previously described (13). The scaffolds have been deposited in Genbank under the following accession numbers (GenBank Submissions: KC513604-KC513657).

Results and Discussion

Identification of low levels of heterozygosity in four genetically different strains of Encephalitozoon cuniculi

Aligning Illumina reads against the reference genomes of EcI, II and III resulted in the detection of a total of 66 regions that were mapped by both reference and alternate reads at similar

frequency (i.e. within the expected threshold). The amount of heterozygosity identified is far lower than that recently reported for distant microsporidia (Figure 5.1, Table 5.1, (3)), with a maximum of 23 nucleotide sites (less than 0.00001 % of the total genome sequence) being affected by heterozygosity in *E. cuniculi* , compared to over 42,175 (1.035% of the complete genome) for more basal species such as *Nematocida* spp. (Figure 5.1, (3)).

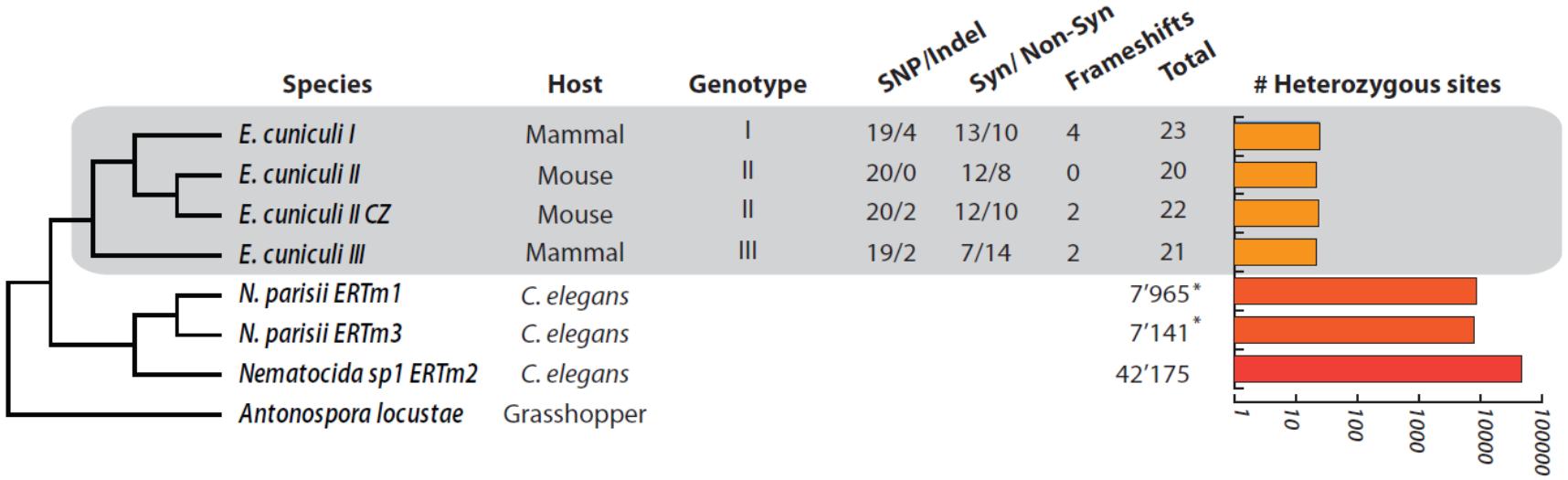


Figure 5.1. Amount, location and nature of heterozygous loci in *E. cuniculi* and comparison with *Nematocida*. Phylogenetic reconstruction of the studied microsporidian species based on the nucleotide sequence of actin and alpha tubulin genes. Known genotype, parasite's host, amount of variation and nature of variation are indicated. SNPs and indels are included in the count of (non)synonymous mutations. *Antonospora locustae* used as a outgroup. Total number of heterozygous site is presented in a bar graph on a logarithmic (log10) scale. Asterisk (*) indicates that variation was found using 454 sequencing. The *Encephalitozoon* genus is represented by a grey square. Detailed list of heterozygous loci is shown in table 1.

Table 5.1. List of heterozygous loci in *E. cuniculi* genotype I, II and III.

Genotype	Gene	Product	Polymorphism Type	Protein Effect	Nucleotide Change	Amino Acid Change	Position in CDS
I	ECU02_0050	hypothetical protein	SNP (transversion)	None	C/A		1738
I	ECU02_0050	hypothetical protein	SNP (transition)	Substitution	G/A	R -> W	1736
I	ECU02_0260	gata zinc finger transcription factor 3	SNP (transition)	Truncation	C/T		60
I	ECU02_0720	histone h2a	SNP (transition)	None	A/G		221
I	ECU05_1190	belongs to the abc transporter superfamily	SNP (transition)	None	A/G		381
I	ECU06_0470	nuclear pore complex protein nup155	SNP (transition)	None	A/G		2656
I	intergenic_ch6		SNP (transversion)		A/C		
I	ECU06_0590	hypothetical protein	Indel	Frame Shift	C/-		655
I	ECU06_0700	hypothetical protein	Indel	Frame Shift	CC/--		429
I	ECU06_0730	ribonucleoside diphosphate reductase small chain	SNP (transition)	None	G/A		719
I	ECU06_0830	hypothetical protein	SNP (transversion)	Substitution	C/A	M -> I	1711
I	ECU07_1005	60s ribosomal protein l37a	SNP (transition)	None	A/G		252
I	ECU11_1160	possible protein of nuclear scaffold	SNP (transition)	None	C/T		609
I	ECU11_1330.6	transcription factor (forkhead domain)	Indel	Frame Shift	A/-		692
I	ECU11_1330.1	transcription factor (forkhead domain)	SNP (transition)	None	G/A		511
I	ECU11_1330.2	transcription factor (forkhead domain)	SNP (transition)	None	T/C		457
I	ECU11_1330.2	transcription factor (forkhead domain)	SNP (transition)	Substitution	C/T	D -> N	452
I	ECU11_1330.3	transcription factor (forkhead domain)	SNP (transition)	None	A/G		368
I	ECU11_1330.4	transcription factor (forkhead domain)	SNP (transition)	None	A/G		232
I	ECU11_1330.5	transcription factor (forkhead domain)	SNP (transition)	Substitution	C/T	A -> T	108
I	intergenic_ch11		SNP (transition)		C/T		
I	ECU11_1340	abc transporter-like protein	Indel	Frame Shift	-/A		1890
I	ECU11_1340	abc transporter-like protein	SNP (transversion)	Substitution	T/A	I -> F	1020
II	ECU04_0440	metal-dependent protease of the pad1 jab1	SNP (transition)	None	T/C		908
II	ECU05_0720	splicing factor u2af large subunit b	SNP (transition)	Substitution	A/G	? -> C	1012
II	ECU06_0700	hypothetical protein	SNP (transition)	None	C/T		413
II	ECU07_0680	chromosome segregation atpase	SNP (transition)	None	G/A		564
II	ECU07_0770	npl4 family protein	SNP (transition)	Substitution	A/G	? -> Y	433

II	ECU07_1490	copper amine oxidase n-terminal domain protein	SNP (transition)	None	C/T		267
II	ECU07_1500	calcium-transporting atpase	SNP (transition)	None	A/G		1857
II	ECU08_0080	dihydrofolate reductase	Indel	Frame Shift	T/-		94
II	ECU08_0230	cell division protein kinase	SNP (transition)	None	T/C		186
II	ECU08_1010	proteophosphoglycan ppg4	SNP (transition)	Substitution	G/A	I -> M	768
II	ECU08_1730	hypothetical protein	SNP (transversion)	Substitution	T/G	? -> N	241
II	ECU09_0360	similarity to hypothetical proteins (upf0129 family)	SNP (transversion)	None	A/C		272
II	ECU09_0410	dna excision repair protein ercc-6	SNP (transition)	Substitution	C/T	? -> L	1417
II	ECU09_0730	bem46 family protein	SNP (transition)	Substitution	A/G	? -> D	738
II	ECU09_1120	hypothetical protein	SNP (transition)	None	C/T		35
II	ECU09_1820	snrnp-like protein	Indel	Frame Shift	G/-		149
II	ECU10_1210	dna polymerase catalytic subunit a	SNP (transition)	Substitution	A/G	? -> F	857
II	ECU11_0410	hypothetical protein	SNP (transition)	Substitution	C/T	N -> S	461
II	Intergenic_CH11		SNP (transition)		C/T		
II	intergenic_ch8.1		SNP (transition)		G/A		
II	Intergenic_CH8.2		SNP (transversion)		T/C		
II	intergenic_ch10		SNP (transition)		A/G		
III	ECU01_0410	hypothetical protein	SNP (transition)	Substitution	A/G	R -> G	674
III	ECU01_0750	hypothetical protein	SNP (transition)	Substitution	T/C	F -> S	333
III	ECU03_0320	60s ribosomal protein l13	SNP (transition)	None	C/T		264
III	ECU03_1180	subtilisin-related endopeptidase k	SNP (transition)	Substitution	G/A	P -> S	1109
III	ECU04_0470	hypothetical protein	SNP (transition)	None	A/G		754
III	ECU04_1460	cyclin b-like guanine nucleotide binding protein	Indel	Frame Shift	T/-		2100
III	ECU04_1470	na+/h+ antiporter	SNP (transversion)	None	T/G		1073
III	ECU04_1490	hypothetical protein	SNP (transversion)	Substitution	C/A	D -> Y	1237
III	ECU05_0700	hypothetical protein	SNP (transition)	Substitution	G/A	G -> R	762
III	ECU05_0850	hypothetical protein	SNP (transition)	Substitution	C/T	G -> S	41
III	intergenic_ch7.1		SNP (transition)		T/C		
III	intergenic_ch7.2		SNP (transition)		A/G		
III	ECU08_1150	oxidoreductase short chain dehydrogenases (sdr) family	SNP (transversion)	Substitution	T/G	W -> G	549
III	ECU08_1680	hypothetical protein	SNP (transition)	Substitution	A/G	E -> G	597

III	ECU10_0540	similarity to adp/atp carrier protein	SNP (transition)	Substitution	C/T	R -> K	5
III	ECU10_0960	similarity to hypothetical proteins ycd7_human	SNP (transition)	Substitution	A/G	S -> P	68
III	ECU10_1030	hypothetical protein	Indel	Frame Shift	G/-		1938
III	ECU11_1120	translation elongation factor 2	SNP (transition)	Substitution	G/A	V -> M	433
III	ECU11_1410	hypothetical protein	SNP (transition)	None	T/C		632
III	intergenic_ch11		SNP (transversion)		G/T		
III	ECU06_0630	hypothetical protein	SNP (transition)	Substitution	G/A	C->Y	1472

In all cases, heterozygous loci were confirmed using PCR and direct Sanger sequencing, with the resulting chromatograms showing the presence of double peaks of similar intensity irrespective of the genotype (Supplemental Figure 1). Additional confirmation for the presence of reduced heterozygosity in these strains was also obtained by mapping reads onto the genome sequence from second strain of genotype II (EcII-CZ) that was independently sequenced in our lab. Specifically, ECII and ECII-CZ were found to differ by only 9 homozygous substitutions (Figure 5.1; Table 5.2), but were otherwise found to harbor an almost identical pattern of heterozygosity (ECII-CZ has two heterozygous indels that are absent in ECII; Table 5.1, 5.2). Given the elevated divergence that is generally present between strains of *E. cuniculi* (28), the great sequence similarity between ECII and ECII-CZ suggests that they represent two very closely related strains.

On average, about 86% of the total heterozygous variation was found to affect coding regions of the genome, possibly because of the elevated gene density of these genomes (i.e. average density of approximately 0.9 genes per Kb; (12)). Between 22% and 57% of substitutions resulted in an amino acid change at one allele, affecting a total of 31 genes with known function and 19 hypothetical proteins (Table 5.1). Several heterozygous deletions and insertions events (indels) were also identified, all of which were found to affect coding regions. In total, 9 genes were found to have undergone pseudogenization at one allele, with ECI showing the largest amount of heterozygous frameshifts mutations (Figure 5.1; Table 5.1). The slightly elevated heterozygosity of ECI mirrors an unusually elevated number of substitutions (n=7) at the locus ECU11_1130, which encodes for a transcription factor. While it is intriguing to speculate about a potential for this locus to represent a mating type locus *in E. cuniculi* (i.e. allelic variation at within transcription factors is linked with sexual identity in other fungi; (30-32)), it seems that the

elevated number of mutations along this gene results instead from the pseudogenization of one allele (i.e. relaxation of selective pressures at one allele).

Intriguingly, all strains investigated harbor similarly low levels of heterozygosity, but this genetic variation was always found to affect different portions of the proteome among genotypes. Finally, sequence coverage was never found to vary among different chromosomes in all genotypes, suggesting the absence of aneuploidy in all strains of this species.

Table 5.2. Substitutions Variation between two independent isolates of *E. cuniculi* genotype II; EcII and EcII-CZ.

Gene	Product	Protein ID	Polymorphism Type	Protein Effect	Nucleotide Change (CZ->II)	Amino Acid Change
Intergenic-CH1			SNP (transversion)		G -> C	
ECU02_0080	hypothetical protein	CAD25039.1	SNP (transversion)	Substitution	T -> A	R -> S
ECU02_0080	hypothetical protein	CAD25039.1	SNP (transition)	Substitution	A -> G	F -> L
ECU02_0650	hypothetical protein	CAD25095.1	SNP (transition)	Substitution	T -> C	D -> G
ECU03_0450	hypothetical protein	CAD26191.1	SNP (transition)	None	C -> T	
ECU03_0710	60S Ribosomal protein L34	CAD26217.1	SNP (transversion)	None	T -> A	
ECU08_1210	hypothetical protein	CAD26427.1	SNP (transition)	Substitution	G -> A	S -> N
ECU08_2120	hypothetical protein	CAD26514.1	SNP (transversion)	Substitution	T -> A	I -> N
ECU08_2120	hypothetical protein	CAD26514.1	SNP (transversion)	Substitution	C -> A	L -> I

Strains of Encephalitozoon are likely to be diploid

Species in the genus *Encephalitozoon* have been proposed to be diploid on a few occasions, most notably following observations of potential homologous chromosomes using hybridization procedures (18-20), and reports of potential heterozygosity at one specific locus (i.e. reported for the CTP synthase, but data not shown (12)). However, these reports have long remained speculative, warranting further explorations across available genomes of these pathogens.

By identifying the presence of several heterozygous loci across the genomes of four genetically divergent strains of *E. cuniculi* , the present study fills an important gap in our understanding of the cellular biology of these vertebrate-infecting parasites. In particular, the identification of heterozygosity in all strains analyzed strongly supports the notion that members of this lineage are diploid; harboring pairs of homologous chromosomes that almost certainly help these obligate intracellular parasites offset the deleterious mutational load harbored by their nuclei. Specifically, an average of 3 heterozygous loci per genome were found to harbor one allele with frameshifts mutations, so having one functional allele at these loci (i.e. resulting from diploidy) was certainly essential for the survival of these parasites with ultra-reduced genome contents.

The presence of two sets of chromosomes per nucleus may also have additional beneficial effects for *E. cuniculi* , particularly within natural populations of these parasites, as diploidy has been linked with the capacity of some vertebrate parasites to escape the immune system of their host (33). This speculative hypothesis, however, is not currently supported by our data, as most heterozygous loci did not appear to encode for proteins known to be involved in such evading mechanisms (Table 5.1 and 5.2). Certainly, future analyses based on natural populations of *E. cuniculi* will be key to understand whether the presence of heterozygosity is relevant for epidemiological studies of these important pathogens of vertebrates; or not.

Opportunistic sex in Encephalitozoon cuniculi?

Recently, mapping of Illumina reads against the genomes of different *Nematocida* spp strains (a basal microsporidian lineage that is evolutionary very distant from *Encephalitozoon* spp.) has resulted in the identification of extremely high levels of heterozygosity. Such elevated levels typically result from the combination of a diploid nuclear state, frequent recombinational events (i.e. genetic variation resulting from meiosis) and genetic exchange. These are hallmarks of sexual reproduction in eukaryotes, and their identification in *Nematocida* has been obviously linked with the presence of sexuality in this latter species (3).

Here, we propose that sexual reproduction may also occur in the more divergent microsporidian species *E. cuniculi ;* for two main reasons. First, regardless of its frequency, the presence of heterozygosity represents compelling evidence for a diploid nuclear state in this latter species, and an essential pre-requisite for the production of haploid cells through meiosis. This mechanism could allow *E. cuniculi* to complete a sexual cycle similar to those found in many other eukaryotes (Figure 5.2A and B), or other diploid fungi (e.g. *Candida albicans*) (Figure 5.2C) (3, 21, 22). Secondly, the poor levels of heterozygosity we have identified in *E. cuniculi* do not typically mirror the absence of sexual reproduction, because exclusively clonal diploids are expected to harbor many, highly divergent alleles in the absence of a mechanism that allows gene shuffling (i.e. meiosis) (22, 34-39). For this reason, the over 1500-fold difference in the level of heterozygosity between strains of *E. cuniculi* and *Nematocida* spp cannot be simply explained by a different mode of reproduction between those lineages (i.e. exclusively clonal *vs.* exclusively sexual).

So where are these sharp differences in heterozygosity between both lineages coming from? One possibility is that the extremely low heterozygosity found in *E. cuniculi* is the result of self-

reproduction (i.e. selfing); a mechanism that is known to heavily reduce the amount of heterozygosity in many other vertebrate parasites (22, 34, 35, 40) and to produce clones that are optimally-adapted to certain hosts (i.e. highly homozygous individuals), but that can still maintain the capacity to undergo outcrossing and create new variation following rare changes in environmental conditions (i.e. opportunistic sexuals). However, given the source of the *E. cuniculi* strains we have analyzed, the extreme reduction in heterozygosity we have observed may have been exacerbated by the methodology used to originate and propagate these under laboratory conditions. Specifically, if those strains were originated from few individuals, then founder effects combined with selfing or inbreeding (e.g. fusion of haploid cells produced by one individual; or by genetically similar individuals) would quickly eradicate most genetic variation from those cultures; resulting in the highly homozygous individuals here (Figure 5.2).

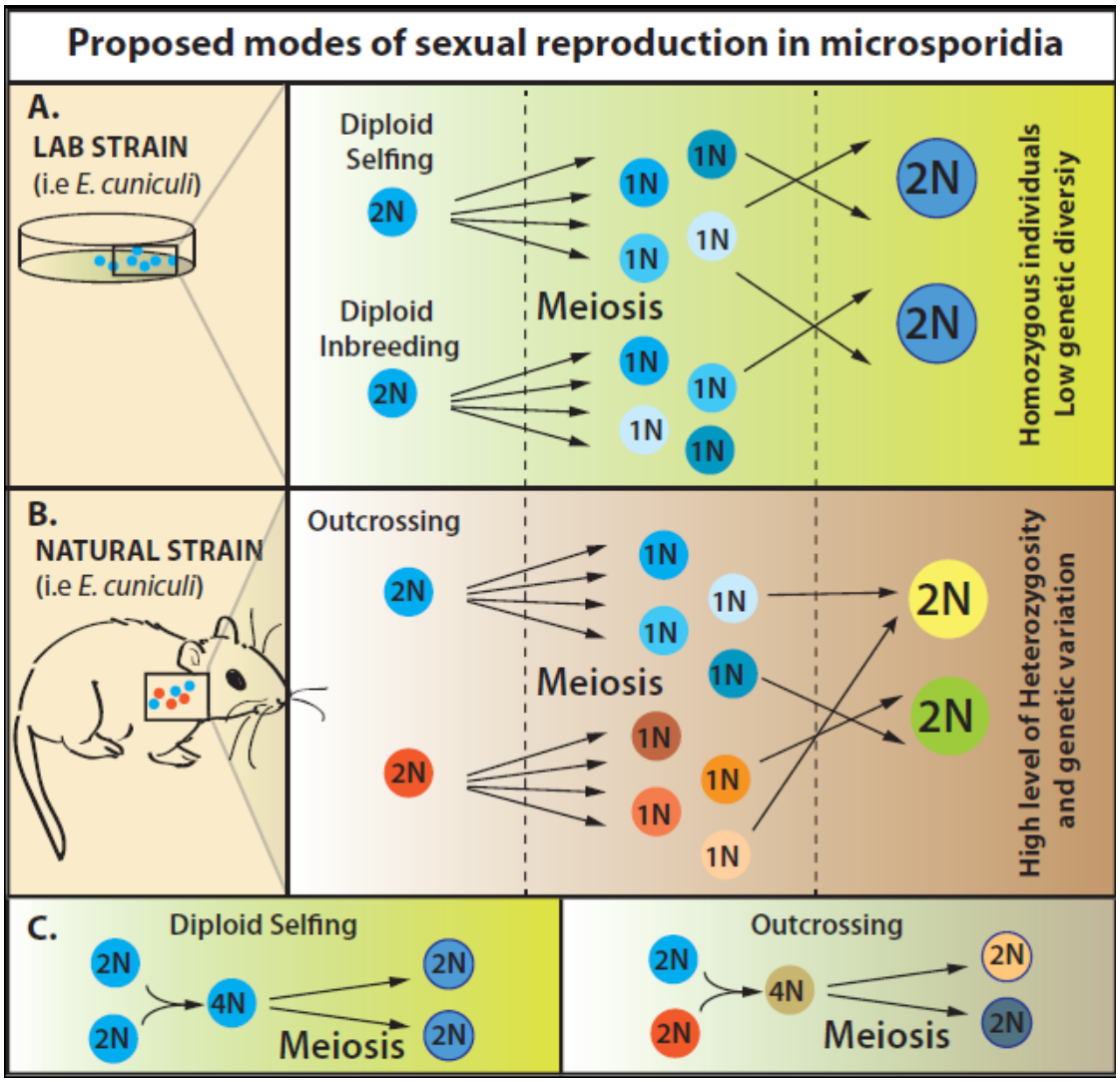


Figure 5.2. A. Selfing, inbreeding and outcrossing in *Encephalitozoon cuniculi.* A. Proposed modes of reproduction and mating systems present within laboratory strains of *Encephalitozoon cuniculi*. **B.** Hypothesis regarding the modes of reproduction and mating systems present within natural populations of *Encephalitozoon cuniculi*. **C.** Same mechanisms shown in panels A and B, but involving the fusion of two diploid mating type cells to increase ploidy, followed by meiosis to generate the progeny.

E. cuniculi could also possibly be clonal and yet find a way to produce genetic variation following a number of different mechanisms that are rarely observed, including the "asexual ploidy cycle" that is found in some amoebozoans and radiolarians (41) or through frequent mitotic recombination. In the asexual ploidy cycle, the ploidy level is increased by endomitosis (i.e. replication without cellular division), and subsequently reduced through cell divisions. This alternation of ploidy levels (2N → 4N → 2N in this case) could mirror a sexual cycle without necessitating the fusion of gametes.

In any case, a number of additional studies are still required to determine whether selfing represents the major cause of reduced heterozygosis in *E. cuniculi* or not, and if gene exchange can actually occur between genetically different members of one species in the field. One such experiment could involve co-infection of the same host cells with different genotypes (co-infection with Ec I with Ec II, or Ec I and Ec III) , followed by genomic analysis of the resulting spores after a number of generations in search of gene exchanges. Certainly, the results from such analyses would provide important clues about the sexual mechanisms that are used by these organisms to produce genetic variation in the first place (i.e. frequent versus *vs.* rare sex).

Similarly, investigations of heterozygosity and presence of recombination along the genomes of *E. cuniculi* strains isolated from the field (i.e. natural populations) will be essential to determine whether selfing is common in these vertebrate pathogens, or whether it is simply a side-effect of their long-term propagation under laboratory conditions. In particular, the identification of elevated number of heterozygous sites within natural populations of this species would be more in line with those recently observed natural in strains of other species such as *Nematocida* (3); a feature that would reflect their capacity to outcross on a more regular basis. Studying natural

populations of *E. cuniculi* could also provide insights into the apparent genetic disparity that is found within and among genotypes of this species. Specifically, the genomes of the ECI, II and III have been recently found to diverge extensively at the sequence level (i.e. elevated homozygous divergence between individuals) (Pombert et al.), while being able to maintain a very reduced intra-genomic diversity. Importantly, the absence of shared heterozygosity between ECI, II and III is consistent with the idea that these strains are rather diverged, and are unlikely to have recently experienced genetic exchange between each other. It now remains to be fully understood whether selfing has allowed these strains to accumulate so many homozygous mutations over time without leaving many traces of heterozygosity behind.

References

1. Keeling PJ and Fast NM: Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 56: 93-116, 2002.
2. Corradi N and Keeling PJ: Microsporidia: a journey through radical taxonomical revisions. Fungal Biology Reviews 23: 1-8, 2009.
3. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, Didier ES, Fan L, Heiman DI, Levin JZ, Young S, Zeng Q and Troemel ER: Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res: 2012.
4. Capella-Gutierrez S, Marcet-Houben M and Gabaldon T: Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol 10: 47, 2012.
5. Lee SC, Corradi N, Byrnes EJ, 3rd, Torres-Martinez S, Dietrich FS, Keeling PJ and Heitman J: Microsporidia evolved from ancestral sexual fungi. Curr Biol 18: 1675-9, 2008.
6. Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ and Heitman J: Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLoS One 5: e10539, 2010.
7. Keeling PJ: Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genet Biol 38: 298-309, 2003.
8. Akiyoshi DE, Morrison HG, Lei S, Feng X, Zhang Q, Corradi N, Mayanja H, Tumwine JK, Keeling PJ, Weiss LM and Tzipori S: Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. PLoS Pathog 5: e1000261, 2009.
9. Corradi N, Haag KL, Pombert JF, Ebert D and Keeling PJ: 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 Biology 10: 2009.

10. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM, Akiyoshi DE and Tzipori S: The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biol Evol 2: 304-9, 2010.
11. Pombert JF, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF, Whitman DW, Weiss LM, Corradi N and Keeling PJ: Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. Proc Natl Acad Sci U S A: 2012.
12. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J and Vivares CP: Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450-3, 2001.
13. Corradi N, Pombert JF, Farinelli L, Didier ES and Keeling PJ: The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon* intestinalis. Nat Commun 1: 77, 2010.
14. Selman M, Pombert JF, Solter L, Farinelli L, Weiss LM, Keeling P and Corradi N: Acquisition of an animal gene by microsporidian intracellular parasites. Curr Biol 21: R576-7, 2011.
15. Selman M and Corradi N: Microsporidia: Horizontal gene transfers in vicious parasites. Mob Genet Elements 1: 251-255, 2011.
16. Schottelius J, Schmetz C, Kock NP, Schuler T, Sobottka I and Fleischer B: Presentation by scanning electron microscopy of the life cycle of microsporidia of the genus *Encephalitozoon*. Microbes Infect 2: 1401-6, 2000.
17. Bigliardi E and Sacchi L: Cell biology and invasion of the microsporidia. Microbes Infect 3: 373-9, 2001.
18. Biderre C, Mathis A, Deplazes P, Weber R, Metenier G and Vivares CP: Molecular karyotype diversity in the microsporidian *Encephalitozoon cuniculi*. Parasitology 118 (Pt 5): 439-45, 1999.
19. Biderre C, Pages M, Metenier G, Canning EU and Vivares CP: Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. Mol Biochem Parasitol 74: 229-31, 1995.
20. Brugere JF, Cornillot E, Metenier G, Bensimon A and Vivares CP: *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. Nucleic Acids Res 28: 2026-33, 2000.
21. Ni M, Feretzaki M, Sun S, Wang X and Heitman J: Sex in fungi. Annu Rev Genet 45: 405-30, 2011.
22. Billiard S, Lopez-Villavicencio M, Hood ME and Giraud T: Sex, outcrossing and mating types: unsolved questions in fungi and beyond. J Evol Biol 25: 1020-38, 2012.
23. Didier ES, Vossbrinck CR, Baker MD, Rogers LB, Bertucci DC and Shadduck JA: Identification and characterization of three *Encephalitozoon cuniculi* strains. Parasitology 111 (Pt 4): 411-21, 1995.
24. Shadduck JA: *Nosema* cuiculi: in vitro isolation. Science 166: 516-7, 1969.
25. Vavra J, Bedrnik P and Cinatl J: Isolation and in vitro cultivation of the mammalian microsporidian *Encephalitozoon cuniculi*. Folia Parasitol (Praha) 19: 349-54, 1972.
26. Shadduck JA, Bendele R and Robinson GT: Isolation of the causative organism of canine *Encephalitozoon*osis. Vet Pathol 15: 449-60, 1978.

27. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Iodice J, Kissinger JC, Kraemer ET, Li W, Nayak V, Pennington C, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Jr., Treatman C and Wang H: AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. Nucleic Acids Res 39: D612-9.
28. Pombert JF, Xu J, Smith DR, Heiman D, Young S, Cuomo CA, Weiss LM and Keeling PJ: Complete genome sequences from three genetically distinct strains reveal a high intra-species genetic diversity in the microsporidian *Encephalitozoon cuniculi*. Eukaryot Cell: 2013.
29. Koudela B, Vitovec J, Kucerova Z, Ditrich O and Travnicek J: The severe combined immunodeficient mouse as a model for *Encephalitozoon cuniculi* microsporidiosis. Folia Parasitol (Praha) 40: 279-86, 1993.
30. Gerke J, Lorenz K and Cohen B: Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. Science 323: 498-501, 2009.
31. Idnurm A: Sex determination in the first-described sexual fungus. Eukaryot Cell 10: 1485-91, 2011.
32. Wong Sak Hoi J and Dumas B: Ste12 and Ste12-like proteins, fungal transcription factors regulating development and pathogenicity. Eukaryot Cell 9: 480-5, 2010.
33. Xu X, Sun J, Zhang J, Wellems D, Qing X, McCutchan T and Pan W: Having a pair: the key to immune evasion for the diploid pathogen Schistosoma japonicum. Sci Rep 2: 346, 2012.
34. Heitman J: Sexual reproduction and the evolution of microbial pathogens. Curr Biol 16: R711-25, 2006.
35. Heitman J: Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host Microbe 8: 86-99, 2010.
36. Barraclough TG, Birky CW, Jr. and Burt A: Diversification in sexual and asexual organisms. Evolution 57: 2166-72, 2003.
37. Balloux F, Lehmann L and de Meeus T: The population genetics of clonal and partially clonal diploids. Genetics 164: 1635-44, 2003.
38. Balloux F, Amos W and Coulson T: Does heterozygosity estimate inbreeding in real populations? Mol Ecol 13: 3021-31, 2004.
39. De Meeus T, Lehmann L and Balloux F: Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. Infect Genet Evol 6: 163-70, 2006.
40. Billiard S, Lopez-Villavicencio M, Devier B, Hood ME, Fairhead C and Giraud T: Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol Rev Camb Philos Soc 86: 421-42, 2011.
41. Kondrashov AS: The Asexual Ploidy Cycle and the Origin of Sex. Nature 370: 213-216, 1994.

Chapter 6 : Conclusion

In the last 8 years, the field of genomics was revolutionized by next-generation sequencing (NGS), and its application on microsporidian parasites have unveiled the unique aspects of their genomes, and their biology.

In the present thesis, this technology have allowed me to revealed many previously unknown aspect of the microsporidian evolution, by acquiring sequence data from two previously overlooked species ; *E. hellem* and *E. romaleae*. Indeed, the genome of these latter organisms were found to harbor many genes that were not known to exist in other members of the group , and which we found originated from multiple HGTs evens from different donors. Even more surprising, the genes acquired by HGTs were found to provide essential , new biochemical capacities to those parasites; some of which may have resulted in a lesser dependency on their hosts(1-3).

Our investigation also revealed that, in addition to being quite rare, events of HGTs can also be rapidly lost. Indeed, we found that not all the genes acquired by *E. romaleae* have been maintained in its genome. Specifically, the genes related to *de novo* synthesis of folate were all found to be non-functional (i.e. pseudogenes). Why was the pathway lost in *E. romaleae* but maintained in *E. hellem* is unclear, but one possible explanation could related to the difference in their hosts (vertebrates vs. arthropods). Perhaps, some aspects of *E. romaleae*'s host biochemistry could made synthesizing folate less advantageous or, alternatively, this latter pathway is less sustainable in insects due to abundance of folate available to the parasite(4, 5) (no need for produce it anymore; so let's just steal it); or perhaps *E. hellem* is also in the process of losing these genes too.

Overall, the genome of *E. hellem* and *E. romaleae* revealed that HGTs are more indeed rare, but still common than previously thought in this phylum, and that the can be an important part of metabolic capacities of these parasites. This observation suggests that other microsporidian species could also have acquired or will acquire one or multiple HGTs within their genomes. HGTs could be excepted to bring new metabolic capacities to make the parasite less dependent on its host, or increase or expand the functional pathways already present in the species.

Could we used the HGTs as drug targets ? Currently, anti-fungal drugs are used for the treatment of microsporidian infections. However there is a lack of alternative treatments when the anti-fungal drugs fail. Several parasitic diseases such as malaria, toxoplasmosis, and bacterial infections are commonly treated with anti-folate drugs. Anti-folate drugs are designed to stop folate synthesis by inhibiting enzymes found in folate biosynthesis pathway. These drugs were previously tested on microsporidia infections by *E. cuniculi* , and showed no effect. However, we now know that *E. cuniculi* is unable to produce its own folate(6), hence anti-folate drugs would be ineffective, as observed(7). Although *E. hellem* is very similar to *E. cuniculi* , their folate pathways largely differ due to the acquisition of HGTs. *E. hellem* contains the genes to synthesis folate, and could be inhibited by the different anti-folate drugs. Hence future testing of anti-folate drugs on *E. hellem* infection would lead to alternative treatment of microsporidiosis due to *E. hellem* infection.

Another important contribution of the present thesis relates to the mode of reproduction of these parasites. Reproduction in fungi is quite complex and diverse; with some lineages being sexual, while others are purely clonal, and many can switch from sexuality to clonality depending on a number of environmental factors (8-10). To date, microsporidian reproduction

was not well understood, but our sequencing efforts targeted at different strains of *E. cuniculi* revealed interesting insights into this specific question.

First, we found that the level of heterozygosity along the genomes of several strains of *E. cuniculi* was extremely low in all cases, which we suggested to be due to the occurrence of self-reproduction (i.e. selfing); a mechanism that is known to heavily reduce the amount of heterozygosity in many other vertebrate parasites. The levels of heterozygosity are not due to asexual reproduction; if the *E. cuniculi* strains were clonal, we would have expected highly divergent alleles and high levels of heterozygosity in the absence of a mechanism that allows gene shuffling. Furthermore, the diploid state of the species is essential for the production of haploid cells through meiosis, which would allow *E. cuniculi* to complete a sexual cycle similar to those found in other eukaryotes.

The main pros for selfing is reproductive assurance, hence not being dependant on other individuals to propagate. The decrease in genetic variability decreases the chances of acquiring deleterious mutation, and also decreases the chances for the parasite to be recognized by its host. The main cons due to selfing is reduced in genetic variability. The loss of heterozygosity and hence the high levels of homozygosity, could lead to the expression of deleterious recessive alleles . Furthermore, lower genetic variability in selfing populations would make them less capable to respond or to adapt to environmental change(11).

Recently, we analyzed a number of natural microsporidia isolates from *E. cuniculi* and closely related species *Nosema ceranae*. Preliminary analysis of those genome both show a much higher level of heterozygosity. This show that natural sample have in fact higher level of heterozygosity, due to the higher genetic diversity found within one population, and the

occurrence of outcrossing, compared to selfing that occurs when propagated under laboratory conditions, as suggested in *E. cuniculi* .

Different experiment could be conducted to experimentally study evolution and sexual reproduction in microsporidia. For instance: studying mating by looking at recombination between the genetically diverse *E. cuniculi* strains; looking at the increase (if asexual) or decrease (if selfing) in the levels of heterozygosity of a natural population by the serial passaging of the parasites in cell culture.

Overall, the work presented in this thesis had lead us to a better understanding of the genome evolution of a medically and economically important group of parasites.

Reference

1. Pombert, J.F., et al., *Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12638-43.
2. Selman, M. and N. Corradi, *Microsporidia: Horizontal gene transfers in vicious parasites.* Mob Genet Elements, 2011. **1**(4): p. 251-255.
3. Selman, M., et al., *Acquisition of an animal gene by microsporidian intracellular parasites.* Curr Biol, 2011. **21**(15): p. R576-7.
4. Blatch, S.A., K.W. Meyer, and J.F. Harrison, *Effects of dietary folic acid level and symbiotic folate production on fitness and development in the fruit fly Drosophila melanogaster.* Fly (Austin), 2010. **4**(4): p. 312-9.
5. Allen, R.H., et al., *Metabolic abnormalities in cobalamin (vitamin B12) and folate deficiency.* FASEB J, 1993. **7**(14): p. 1344-53.
6. Katinka, M.D., et al., *Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi.* Nature, 2001. **414**(6862): p. 450-3.
7. Beauvais, B., et al., *In vitro model to assess effect of antimicrobial agents on Encephalitozoon cuniculi.* Antimicrob Agents Chemother, 1994. **38**(10): p. 2440-8.
8. Dyer, P.S. and C.M. O'Gorman, *Sexual development and cryptic sexuality in fungi: insights from Aspergillus species.* FEMS Microbiol Rev, 2012. **36**(1): p. 165-92.
9. Whittle, C.A., K. Nygren, and H. Johannesson, *Consequences of reproductive mode on genome evolution in fungi.* Fungal Genet Biol, 2011. **48**(7): p. 661-7.
10. Ni, M., et al., *Sex in fungi.* Annu Rev Genet, 2011. **45**: p. 405-30.
11. Agrawal, A.F. and C.M. Lively, *Parasites and the evolution of self-fertilization.* Evolution, 2001. **55**(5): p. 869-79.

Appendix A : Acquisition of an animal gene by microsporidian intracellular parasites.

Mohammed Selman[a], Jean-François Pombert[b], Leellen Solter[c], Laurent Farinelli[d], Louis M. Weiss[d], Patrick Keeling[b] and Nicolas Corradi[a]

[a] Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, K1N1H7, Canada

[b] Canadian Institute for Advanced Research, Botany Department, University of British Columbia, Vancouver, BC, Canada

[c] Illinois Natural History Survey, University of Illinois, 1816 S. Oak St., Champaign, IL 61820, USA

[d] FASTERIS S.A., Ch. du Pont-du-Centenaire 109, P.O. Box 28, CH-1228 Plan-les-Ouates, Geneva, Swizerland

[e] Department of Medicine, Albert Einstein College of Medicine, Bronx, New York, USA

Abstract

Parasites have adapted to their specialized way of life by a number of means, including the acquisition of genes by horizontal gene transfer. These newly acquired genes seem to come from a variety of sources, but seldom from the host, even in the most intimate associations between obligate intracellular parasite and host (1). Microsporidian intracellular parasites have acquired a handful of genes, mostly from bacteria, that help them take energy from their hosts or protect them from the environment (2,3). To date, however, no animal genes have been documented in any microsporidian genome. Here, we have surveyed the genome of the microsporidian *Encephalitozoon romaleae*, which parasitises arthropods for evidence of animal genes. We found one protein-encoding gene that is absent from publicly available sequence data from other microsporidia. The gene encodes a component of the purine salvage pathway, and has been independently acquired by other parasites through horizontal gene transfer from other donors. In this case, however, the gene shows a very strong phylogenetic signal for arthropod origin.

Main text

We created a 20-fold coverage survey of the *E. romaleae* genome, resulting in 165 contigs, with an average length of 13,350 bp. Search for genes of potential animal origin revealed the presence of only one candidate, a purine nucleotide phosphorylase (PNP). Interestingly, this gene is absent from any other publicly available microsporidian sequence data, including complete genomes from other members of the genus *Encephalitozoon*(4). *Encephalitozoon* genomes share a high level of co-linearity, and the *E. romaleae* PNP gene is flanked by genes with high sequence similarity and gene order conservation from regions of chromosome 1 of *E. cuniculi* and *E. intestinalis*, respectively (Supplemental information). This protein is involved in a pathway that is notoriously reduced in other members of the lineage, but otherwise essential for salvaging purines in other eukaryotes (4), and its inclusion in the genome of *E. romaleae* was confirmed by PCR and conventional DNA sequencing.

The origin of the PNP gene was assessed using a variety of models and methods for phylogenetic reconstruction. The phylogeny consistently showed the microsporidia to cluster not just with animals, but specifically with arthropods with high support (Figure 1). The exclusion of the more divergent arthropod sequences (i.e., crustaceans and Pediculus) had no effect on either tree topology or support (Supplemental information). *E. romaleae* is unusual in that it is the first described species or *Encephalitozoon* isolated from an insect (5); all other members of the genus are only known to infect vertebrates. The arthropod origin of its PNP might suggest a recent, insect host origin, so we also searched an ongoing genome project from a putative sister species, the human parasite *E. hellem*, for the presence of PNP. Interestingly, the arthropod PNP is also found in the same genomic context in *E. hellem* (Figure 1), and we confirmed that these two species are indeed sister-species using a multigene phylogeny (Supplemental information).

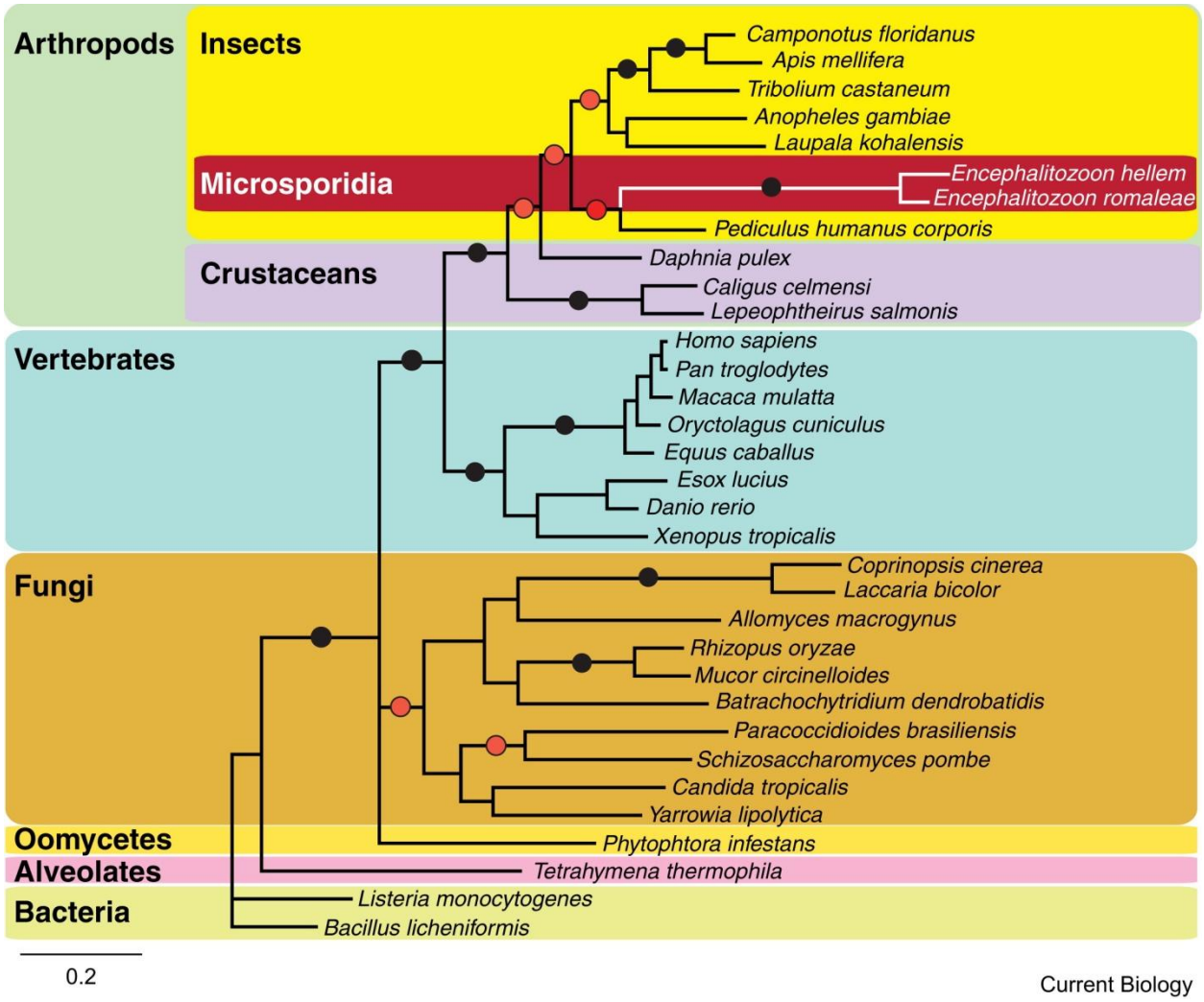


Figure 1.1. Phylogeny of the PNP genes.

Phylogenetic relationships between the PNP genes based on 240 amino acid positions from a broad diversity of eukaryotes and prokaryotes. Major lineages are indicated by coloured boxes, while black circles indicate branches with bootstrap support of over 95% from Maximum likelihood analyses (WAG model of evolution) and over 0.95 posterior probabilities obtained using Mr Bayes (WAG model of evolution) and Phylobayes (CAT and LG models of evolution). Red circles indicate branches with posterior probabilities of 1 using Mr Bayes, but with bootstrap support and posterior probabilities sometimes below 95% and 0.95 for, either, maximum likelihood analyses, or for Bayesian analyses performed under the CAT and LG models of evolution implemented in Phylobayes. Phylogenetic relationships between the PNP genes of several eukaryotic and prokaryotic lineages based on 240 amino acid positions after removal of sequences corresponding to *Pediculus humanus* and Crustaceans (i.e. the longest branches) are shown in Figure S1.

Overall, these data indicate that the PNP gene was acquired from an insect in the ancestor of E. romalea and *E. hellem*, which raises the question: was this insect the host? The exceedingly narrow distribution of this gene in the sister species *E. hellem* and *E. romaleae* is most consistent with a recent gain of the gene. But *E. hellem*, like all other described members of this genus, is a parasite of vertebrates. It is possible that our current understanding of host-range in *Encephalitozoon* species is limited by sampling bias, or ancestral types had broader host-ranges. Indeed, infection of both insects and vertebrate hosts by microsporidia has been documented in Anncalia algerae(6), *Trachipleistophora hominis*(7) and Trachipleistophora extenrec(8). This is particularly plausible given that *E. romaleae* is an insect parasite, so some host switching must have occurred in the ancestor of *E. romaleae* and *E. hellem*. The alternative explanation — that an ancestral intracellular parasite that specifically infected vertebrates somehow acquired an insect gene — is difficult to imagine since exposure of the parasite to insect genes would presumably be very limited.

The function of the PNP gene in parasite biology is also of interest because many parasites depend on salvage pathways for their nucleotides. In the apicomplexan Cryptosporidium, the pyrimidine salvage enzyme thymidine kinase was acquired from a bacterium (9), as was the PNP itself in the diplomonad *Giardia*(10). These three lineages acquired similar functions in parallel by acquiring new genes through HGT, but only in microsporidia was it apparently derived from the host. The genome-level data from microsporidia now available also raise the interesting question of why some species of *Encephalitozoon* get by without PNP while these two species have retained it, despite their otherwise highly reduced gene repertoire. Neither the long-term fate of such genes acquired by HGT, nor the short-term implications of their integration into

cellular pathways are well understood, but the relatively tractable genomes of *Encephalitozoon* make this an appealing genus in which to address such questions.

Reference

1 Anderson, M.T., and Seifert, H.S. (2011). Opportunity and means: horizontal gene transfer from the human host to a bacterial pathogen. mBio 2, .

2 Chan, K.W., Slotboom, D.J., Cox, S., Embley, T.M., Fabre, O., van der Giezen, M., Harding, M., Horner, D.S., Kunji, E.R., Leon-Avila, G., et al. (2005). A novel ADP/ATP transporter in the mitosome of the microaerophilic human parasite Entamoeba histolytica. Curr. Biol. 15, 737–742.

3 Fast, N.M., Law, J.S., Williams, B.A., and Keeling, P.J. (2003). Bacterial catalase in the microsporidian *Nosema* locustae: implications for microsporidian metabolism and genome evolution. Eukaryot. Cell 2, 1069–1075.

4 Corradi, N., Pombert, J.F., Farinelli, L., Didier, E.S., and Keeling, P.J. (2010). The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon* intestinalis. Nat. Comm. 1, . doi:10 1038/ncomms1082

5 Lange, C.E., Johny, S., Baker, M.D., Whitman, D.W., and Solter, L.F. (2009). A new *Encephalitozoon* species (Microsporidia) isolated from the lubber grasshopper, Romalea microptera (Beauvois) (Orthoptera: Romaleidae). J. Parasitol. 95, 976–986.

6 Coyle, C.M., Weiss, L.M., Rhodes, L.V., Cali, A., Takvorian, P.M., Brown, D.F., Visvesvara, G.S., Xiao, L., Naktin, J., Young, E., et al. (2004). Fatal myositis due to the microsporidian Brachiola algerae, a mosquito pathogen. New Eng. J. Med. 351, 42–47.

7 Weidner, E., Canning, E.U., Rutledge, C.R., and Meek, C.L. (1999). Mosquito (Diptera: Culicidae) host compatibility and vector competency for the human myositic parasite *Trachipleistophora hominis* (Phylum Microspora). J. Med. Entomol. 36, 522–525.

8 Vavra, J., Kamler, M., Modry, D., and Koudela, B. (2010). Opportunistic nature of the mammalian microsporidia: experimental transmission of Trachipleistophora extenrec (Fungi: Microsporidia) between mammalian and insect hosts. Parasitol. Res. 108, 1565–1573.

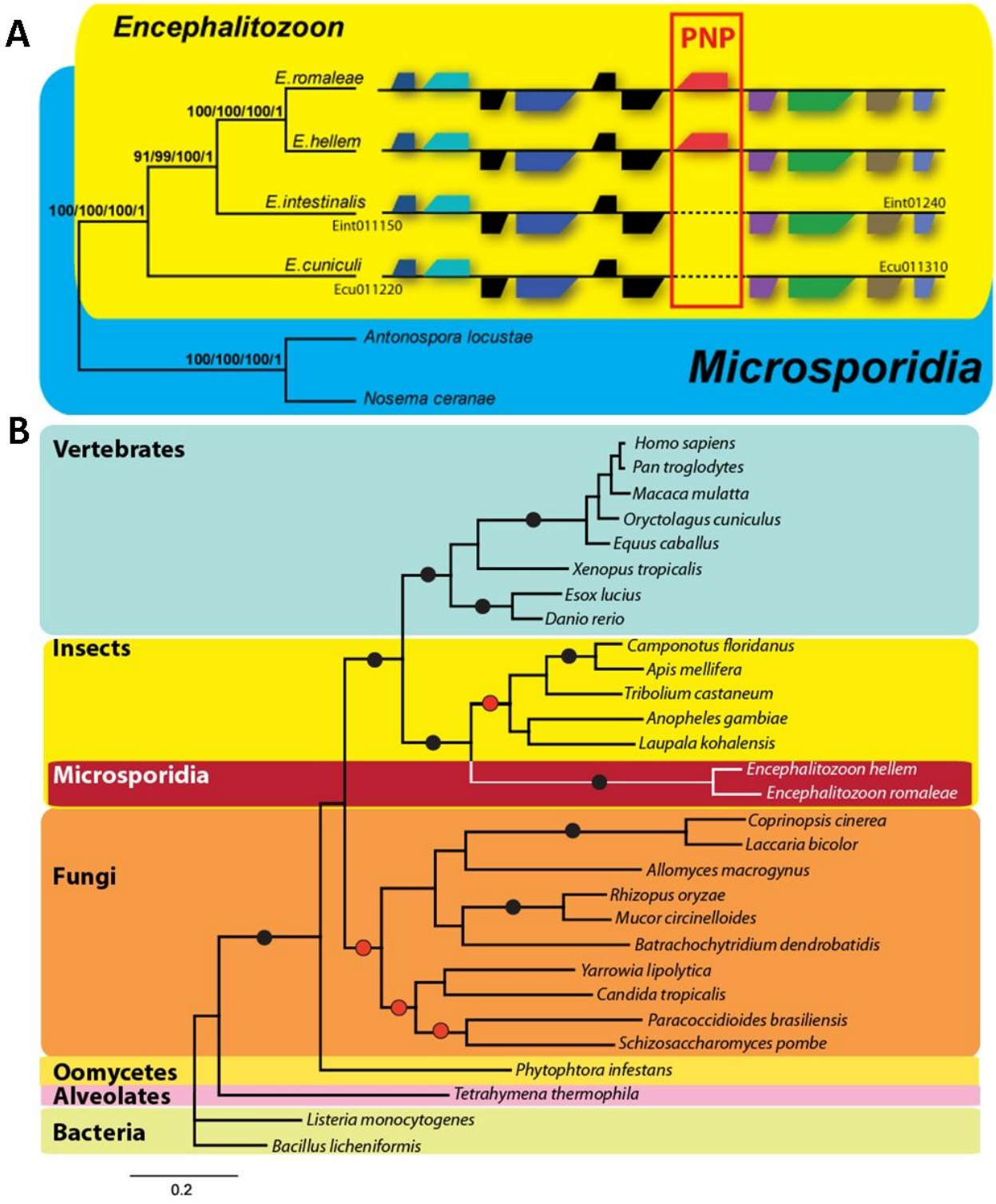
9 Striepen, B., Pruijssers, A.J., Huang, J., Li, C., Gubbels, M.J., Umejiego, N.N., Hedstrom, L., and Kissinger, J.C. (2004). Gene transfer in the evolution of parasite nucleotide biosynthesis. Proc. Natl. Acad. Sci. USA 101, 3154–3159.

10 Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., et al. (2007). Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. Science 317, 1921–1926.

Supplemental Information

Figure S1 (Related to Figure 1): A. Phylogeny of the genus *Encephalitozoon* and presence of a new gene in *Encephalitozoon romaleae* and *Encephalitozoon hellem.* B. Phylogenies of the PNP genes from *E. romaleae* and *E. hellem* after removal of the longest branches from Arthropods;

A. Alignment of highly collinear sections of approximately 13kb from *E. romaleae* and *E. hellem* that are homologous to regions of chromosome 1 of *Encephalitozoon cuniculi* and *E. intestinalis*, and phylogeny of the genus *Encephalitozoon* based on the amino acid sequence of 20 conserved genes (8155 amino acids in total). Values at nodes represent bootstrap and posterior probabilities support obtained using Minimum evolution (JTT model of evolution), as well as PhyML, RAxML, and Mr bayes (WAG model of evolution). The location of the purine nucleoside phosphorylase (PNP) genes in *E. romaleae* and *E. hellem* are shown in red. Conserved hypothetical proteins are shown as black rectangles, whereas genes encoding for proteins with known functions are shown as coloured rectangles. Dashed lines indicate portions of the genome sequence that are missing a particular gene. The GenBank accession number for the *E. romaleae* contig containing the PNP gene is JF808666. **B.** Phylogenetic relationships between the PNP genes of several eukaryotic and prokaryotic lineages based on 240 amino acid positions after removal of sequences corresponding to *Pediculus humanus* and Crustaceans. Major lineages are indicated by coloured boxes, while black circles indicate branches with bootstrap support of over 95% from Maximum Likelihood analyses (WAG model of evolution) and posterior probabilities over 0.95 obtained using Mr bayes (WAG model of evolution) and Phylobayes (CAT and LG models of evolution). Red circles indicate branches with posterior probabilities of 1 using Mr bayes, but with bootstrap support and posterior probabilities sometimes below 95% and 0.95 for either Maximum Likelihood analyses (WAG model of evolution), or for the Bayesian analyses performed under the CAT and LG models of evolution implemented in Phylobayes.


A
Encephalitozoon
PNP
E.romaleae
100/100/100/1
E.hellem
91/99/100/1
E.intestinalis
Eint011150
Eint01240
100/100/100/1
E.cuniculi
Ecu011220
Ecu011310
Antonospora locustae
100/100/100/1
Microsporidia
Nosema ceranae
B
Vertebrates
Homo sapiens
Pan troglodytes
Macaca mulatta
Oryctolagus cuniculus
Equus caballus
Xenopus tropicalis
Esox lucius
Danio rerio
Insects
Camponotus floridanus
Apis mellifera
Tribolium castaneum
Anopheles gambiae
Laupala kohalensis
Microsporidia
Encephalitozoon hellem
Encephalitozoon romaleae
Fungi
Coprinopsis cinerea
Laccaria bicolor
Allomyces macrogynus
Rhizopus oryzae
Mucor circinelloides
Batrachochytridium dendrobatidis
Yarrowia lipolytica
Candida tropicalis
Paracoccidioides brasiliensis
Schizosaccharomyces pombe
Oomycetes
Phytophtora infestans
Alveolates
Tetrahymena thermophila
Bacteria
Listeria monocytogenes
Bacillus licheniformis
0.2

Supplemental Experimental Procedures

Identification of putative animal-derived genes in the E. romaleae genome

Open reading frames (ORFs) were annotated across the genome survey of *Encephalitozoon romaleae* using Artemis (1). Potential LGTs of animal origin were searched, at first, across annotated ORFs using a number blast procedures (Blastp, Blastx, tBlastn, tBlastx) (2) against available microsporidian genome data and the 'non-redundant' repository at NCBI. One gene found to be absent from any other available microsporidian sequence data (PNP), was also found to show strong sequence similarities against a number of metazoan sequences following Blast searches. Its metazoan origin was further tested using a number of phylogenetic methodologies, as explained in the section below**.**

Phylogenetic analyses

Amino acid sequences of PNP's eukaryotic orthologs (best reciprocal blast hits) were acquired from RefSeq GenBank, ESTdb, and from complete eukaryotic genome databases from the Broad institute and DOE Joint Genome Institute. GenBank accession numbers used for phylogenetic reconstruction can be found in the table below. Protein sequences were aligned using MUSCLE 3.7(3), with a maximum number of iterations of 16. Poorly aligned positions and divergent regions of the alignment were removed using GBlocks 0.91b and default settings (4), resulting in a data set of 174 amino acids. Trimal (5), a less stringent trimming tool, was also used to remove non-informative amino acid regions from the alignment using the "strict method". The resulting alignment contained a suite of 240 amino acids.

Table: List of accession numbers (GenBank) and species used in phylogenetic analyses of PNP genes, Related to Figure 1.

PNP Accession	Species Name	Group
YP_079657	*Bacillus licheniformis ATCC 14580*	Bacteria
NP_465477	*Listeria monocytogenes EGD-e*	Bacteria
XP_001020972	*Tetrahymena thermophila*	Alveolata
EGF83310	*Batrachochytrium dendrobatidis JAM81*	Fungi
EEH43246	*Paracoccidioides brasiliensis Pb18*	Fungi
XP_506036	*Yarrowia lipolytica CLIB122*	Fungi
XP_002548301	*Candida tropicalis MYA-3404*	Fungi
NP_593927	*Schizosaccharomyces pombe 972h-*	Fungi
XP_001837849	*Coprinopsis cinerea okayama7#130*	Fungi
XP_001878763	*Laccaria bicolor S238N-H82*	Fungi
AMAG_14981.1**	*Allomyces macrogynus*	Fungi
RO3G_00999**	*Rhizopus oryzae*	Fungi
105288*	*Mucor circinelloides CBS277.49*	Fungi

EFX76980	*Daphnia pulex*	Crustacea
ACO14860	*Caligus clemensi*	Crustacea
XP_967070	*Tribolium castaneum*	Hexapoda
XP_001688760	*Anopheles gambiae str. PEST*	Hexapoda
XP_391850	*Apis mellifera*	Hexapoda
EFN71333	*Camponotus floridanus*	Hexapoda
XP_002427236	*Pediculus humanus corporis*	Hexapoda
EH641232		
EH634545***	*Laupala kohalensis*	Hexapoda
ACO14424	*Esox lucius*	Actinopterygii
NP_998476	*Danio rerio*	Actinopterygii
NP_001006720	*Xenopus (Silurana) tropicalis*	Amphibia
XP_002718082	*Oryctolagus cuniculus*	Mammalia
XP_001104622	*Macaca mulatta*	Mammalia
NP_000261	*Homo sapiens*	Mammalia
XP_001140576	*Pan troglodytes*	Mammalia
XP_001505186	*Equus caballus*	Mammalia
XP_002897613	*Phytophthora infestans T30-4*	Oomycetes

***** Protein ID, JGI (DOE Joint Genome Institute)
******Locus ID, Broad Institute
*** EST fragment containing PNP gene

Phylogenetic analyses were carried out using Maximum Likelihood and Bayesian methods. Maximum Likelihood phylogenies were performed with PhyML 3.0 (6) using the WAG substitution model, 100 bootstraps, 4 substitution rate categories and estimated gamma (Γ) parameters and proportion of invariant sites (I). Bayesian reconstructions were performed with Mr Bayes 3.1.2 (7) under the WAG+Γ4+I model of amino acid substitution. The Markov chain Monte Carlo searches were run for 10,000,000 generations, sampling the Markov chains every 10 generations; the first 25,000 trees were discarded as 'burn-in'. Finally, the CAT and LG models implemented in Phylobayes were also used to reconstruct the phylogeny of PNP (8). The two independent chains run for 10,000 (CAT) and 50,000 (LG) generations, even though both were found to rapidly converge (after 100 generations). The posterior distributions obtained under these independent runs were compared after a burn-in of 100 (CAT) and 1,000 (LG), resulting in maxdiff values much less than 0.1 (indicative of very good runs) for CAT and lower than 0.016 for LG (indicative of very good runs). The consensus trees were obtained by pooling all the trees from both chains.

The phylogeny of the genus *Encephalitozoon* (Figure S1.A) was reconstructed using the amino acid sequences of 20 conserved genes (8155 amino acids in total) from the complete genomes of *E. cuniculi* and *E. intestinalis*, the genome sequence survey data from *E. romaleae*, and sequences from an ongoing genome project from *E. hellem*. Homologues from complete or nearly complete genomes of *Nosema ceranae* and *Antonospora locustae* were used as outgroups .

Table: List of accession numbers (GenBank) and species used in phylogenetic analyses conserved Microsporidian genes.

Protein	*Antonospora locustae**	*Nosema ceranae*	*E. cuniculi*	*E. hellem*	*E. intestinalis*	*E. romaleae*
Actin	AAB86863	XP_002995436	XP_965880	AAB86862	XP_003072256	JN039386
DNA repair helicase RAD25	contig_340	XP_002995956	XP_965942	JN039409	XP_003072319	JN039394
Enolase	contig_489	XP_002995378	NP_586285	JN039405	XP_003073850	JN039388
Glucose-6-phosphate isomerase	contig_2954	XP_002996045	NP_597407	JN039408	XP_003072875	JN039392
Hsp70NP	AAC47660	XP_002995188	NP_586360	BAB69033	XP_003073899	JN039393
Isoleucyl tRNA synthetase	AAC41564	XP_002996071	CAD26020	BAD83624	XP_003073958	JN039397
Mannose-1-phosphate-guanylyltransferase	contig_1173	XP_002996275	NP_586375	JN039402	XP_003073916	JN039383
Methionine aminopeptidase 2	contig_868	XP_002996537	NP_586190	AAP51023	ADM12396	JN039385
MCM2	contig_119	XP_002995530	CAD25272	JN039410	ADM11370	JN039396
Pyruvate dehydrogenase E1 alpha subunit	contig_369	XP_002996136	XP_955659	JN039403	XP_003073591	JN039384
Phosphomannomutase	contig_1060	XP_002994973	NP_597365	JN039401	XP_003072832	JN039381
Pyruvate kinase	contig_493	XP_002996514	XP_955618	JN039404	XP_003073549	JN039387
RNA polymerase I largest subunit	AAT12325	XP_002996654	NP_584825	JN039411	XP_003073214	JN039398
Pyruvate dehydrogenase E1 beta subunit	contig_369	XP_002995305	NP_584800	JN039400	XP_003072763	JN039380
RNA polymerase II largest subunit	AAD12605	XP_002995402	CAD26175	JN039412	XP_003072524	JN039399
Trehalose-6-phosphate phosphatise	AAT12365	XP_002996623	XP_965922	JN039406	XP_003072299	JN039395
Translation elongation factor 1 alpha	contig_559	XP_002995330	NP_584794	JN039407	XP_003072757	JN039391
Tubulin alpha	AAC47419	XP_002995388	NP_586048	P92120	XP_003073238	JN039390
Tubulin beta	AAG48935	XP_002995929	NP_597591	Q24829	XP_003072575	JN039389
Transcription initiation factor TFIIB	contig_106	XP_002996560	NP_585866	JN052740	XP_003073070	JN039382

*Contigs from *A. locustae* can be found at http://forest.mbl.edu/cgi-bin/site/*Antonospora*01.

Extended acknowledgements

We would also like to thank Sara Selman, Douglas Whitman, Magne Osteras and Loïc Baerlocher for help with sequencing, genome assembly and handling of microsporidian material, and Stephane Aris-Brosou, Toni Gabaldon and two anonymous reviewers for important comments on a previous version of this manuscript. We acknowledge access to the *Allomyces macrogynus* and *Rhizopus oryzae* genome data generated by the Broad Institute and to the *Batrachochytrium dendrobatidis* and *Mucor circinelloides* genome data provided by the US Department of Energy Joint Genome Institute. We thank Hilary Morrison and acknowledge the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution for the use of data included in the *Antonospora locustae* Genome Project funded by NSF.

Supplemental References

1. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England) *16*, 944-945.
2. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.
3. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.
4. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol *17*, 540-552.
5. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxford, England) *25*, 1972-1973.
6. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.
7. Huelsenbeck, J.P., and Ronquist, F. (2001). MR BAYES: Bayesian inference of phylogenetic trees. Bioinformatics (Oxford, England) *17*, 754-755.
8. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics (Oxford, England) *25*, 2286-2288.

Appendix B : Supplemental information for Chapter 3

Table S1. Newly identified homologous introns inserted at cognate sites in *Encephalitozoon* genomes

	Intron	
Genomes	Locus	Coordinates*
E. intestinalis	EINT_100755	97465–97492c
	EINT_110355	47441–47469
E. hellem	EHEL_100810	100711–100735c
	EHEL_110370	46364–46390
E. romaleae	EROM_100680	88856–88883c
	EROM_110360	44929–44957
E. cuniculi	ECU10_0805	110242–110269c
	ECU11_0505	72842–72868

*Introns on the lagging strand are denoted with a c (complement).

Figure. S1. Venn diagram of shared gene content in *Encephalitozoon* species. The genes coding for the small and large ribosomal RNA subunits, present in multiple copies, were counted only once for this analysis.

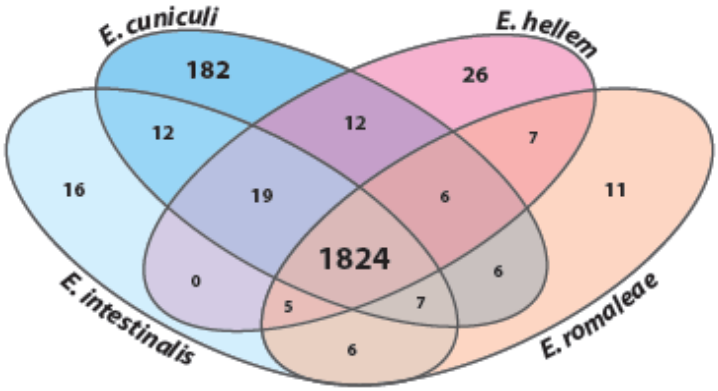
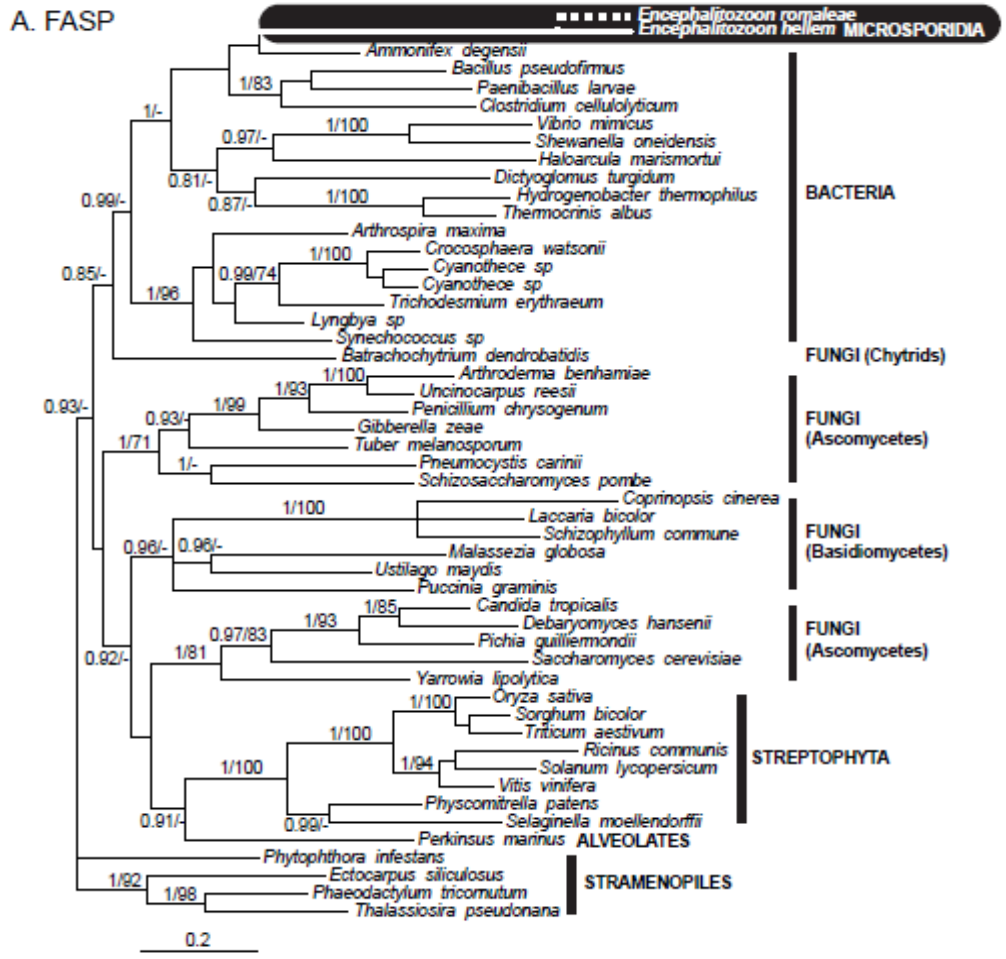


Figure. S2. Bayesian phylogenetic trees of folic acid synthase (FASP) (A) and dihydrofolate synthase (DHFS) (B) present in *Encephalitozoon hellem* and pseudogenized in *Encephalitozoon romaleae* (shown in white on black) but absent from other microsporidians. Numbers at nodes representing Bayesian posterior probabilities (Left) and bootstrap proportions (Right) are indicated when higher than 0.8 and 70%, respectively. The scale bar corresponds to the estimated number of amino acid substitutions per site. Trees are shown unrooted. The dashed lines leading to *E. romaleae* represent the very short pseudogenes found in this species that were not included in the phylogenetic reconstructions.



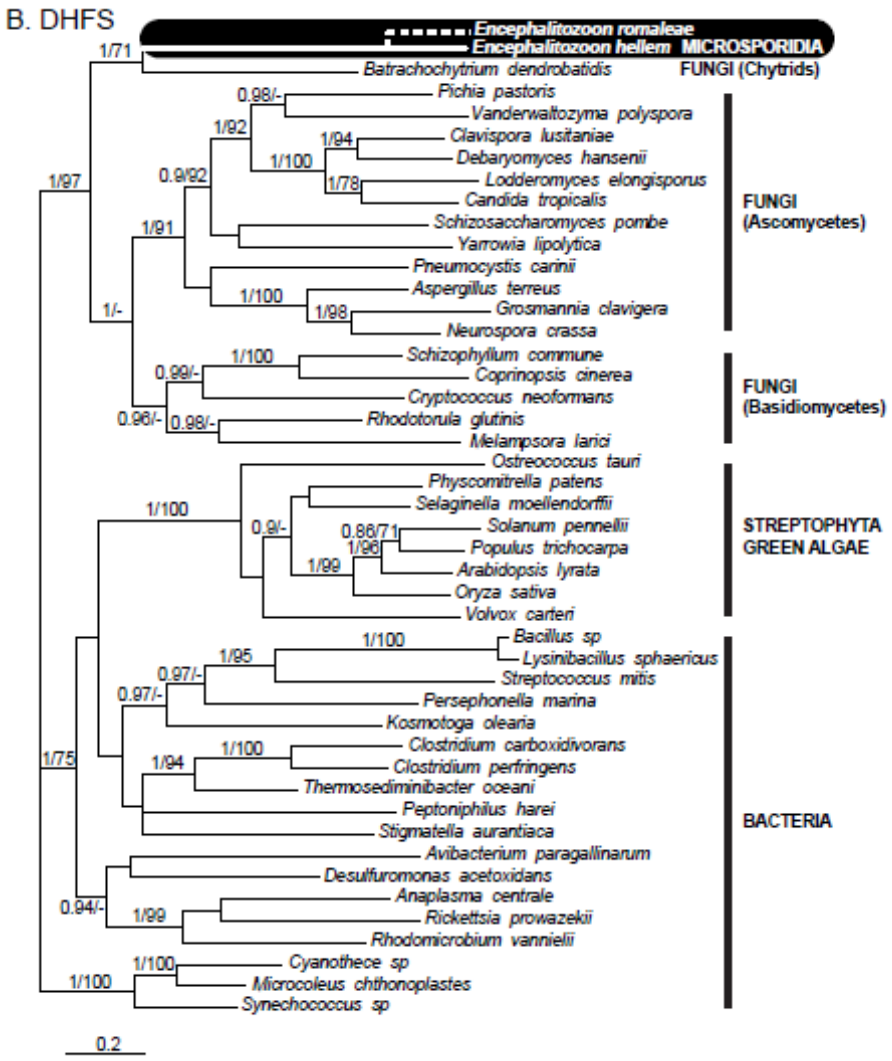
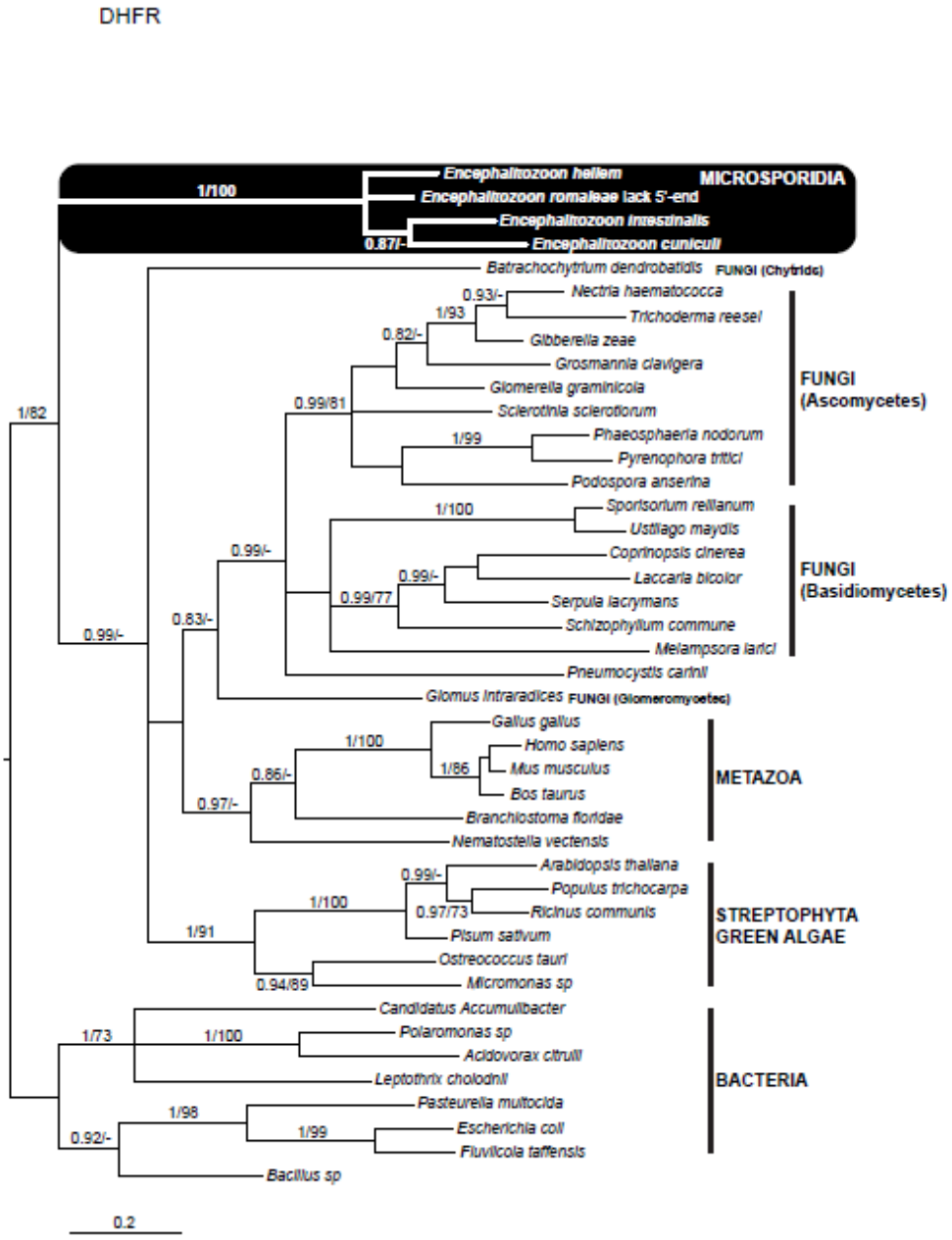
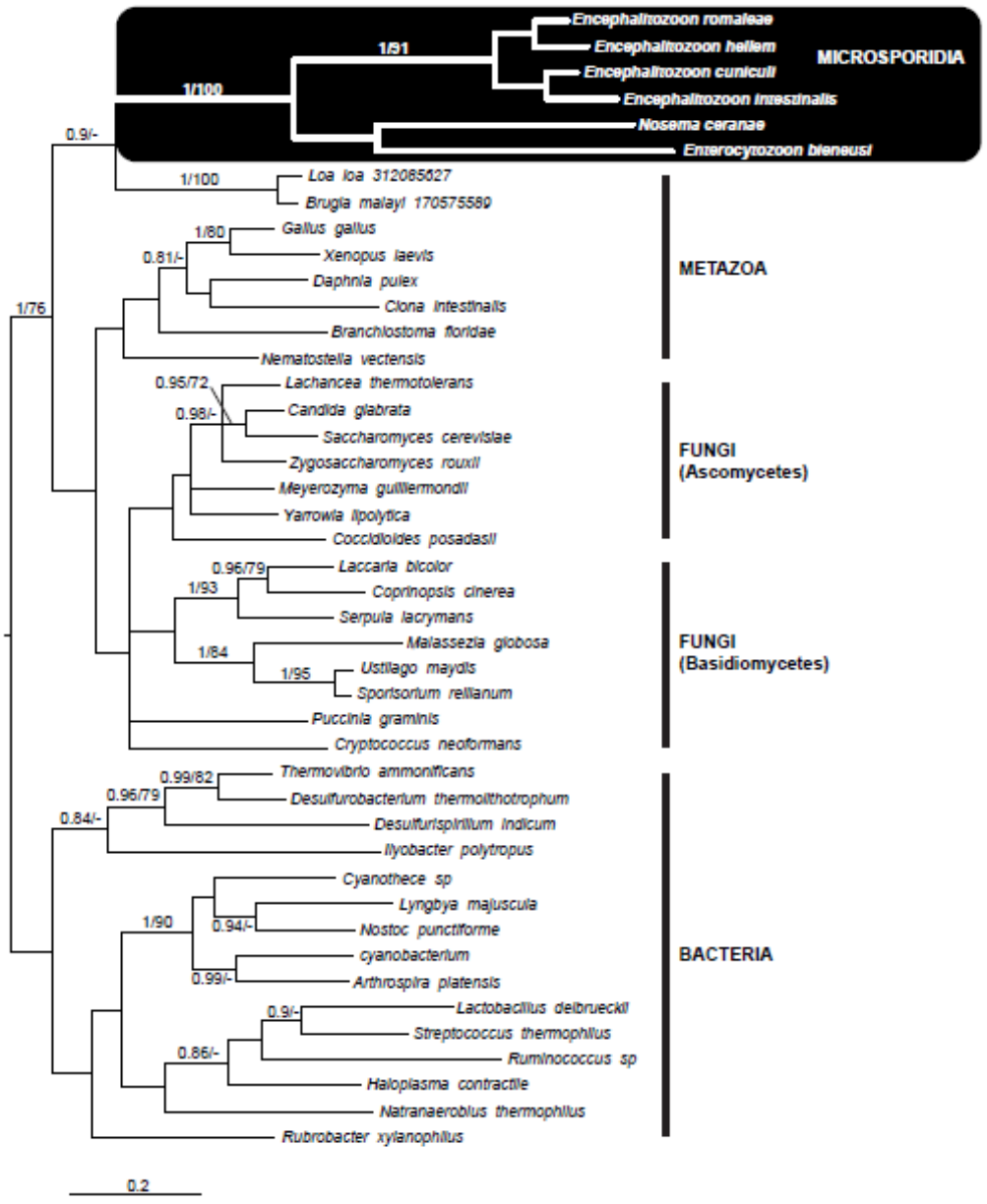

B. DHFS
Encephalitozoon romaleae
Encephalitozoon hellem
MICROSPORIDIA
Batrachochytrium dendrobatidis
FUNGI (Chytrids)
Pichia pastoris
Vanderwaltozyma polyspora
Clavispora lusitaniae
Debaryomyces hansenii
Lodderomyces elongisporus
Candida tropicalis
Schizosaccharomyces pombe
Yarrowia lipolytica
Pneumocystis carinii
Aspergillus terreus
Grosmannia clavigera
Neurospora crassa
FUNGI (Ascomycetes)
Schizophyllum commune
Coprinopsis cinerea
Cryptococcus neoformans
Rhodotorula glutinis
Melampsora larici
FUNGI (Basidiomycetes)
Ostreococcus tauri
Physcomitrella patens
Selaginella moellendorffii
Solanum pennellii
Populus trichocarpa
Arabidopsis lyrata
Oryza sativa
Volvox carteri
STREPTOPHYTA GREEN ALGAE
Bacillus sp
Lysinibacillus sphaericus
Streptococcus mitis
Persephonella marina
Kosmotoga olearia
Clostridium carboxidivorans
Clostridium perfringens
Thermosediminibacter oceani
Peptoniphilus harei
Stigmatella aurantiaca
Avibacterium paragallinarum
Desulfuromonas acetoxidans
Anaplasma centrale
Rickettsia prowazekii
Rhodomicrobium vannielii
Cyanothece sp
Microcoleus chthonoplastes
Synechococcus sp
BACTERIA
0.2

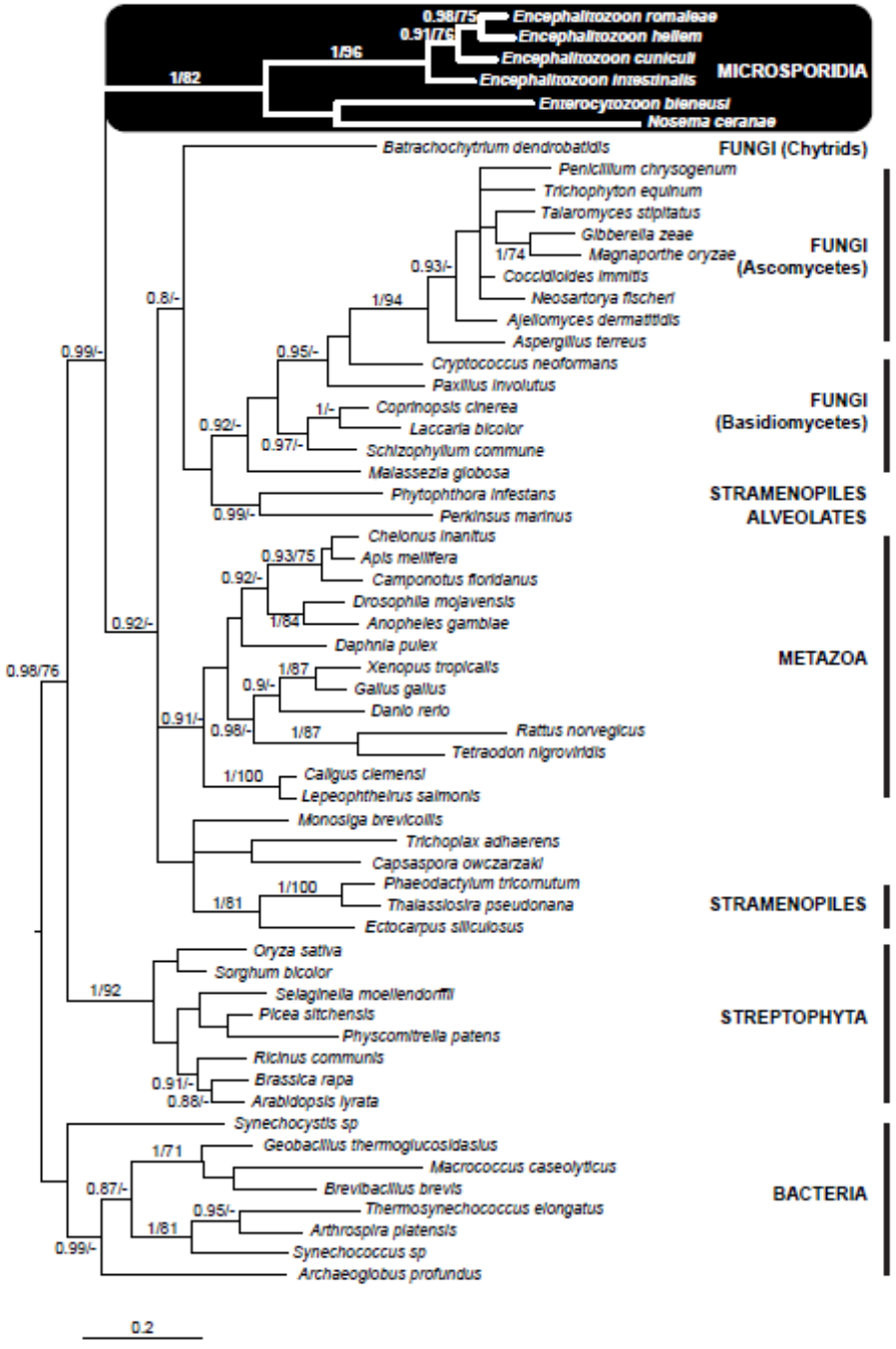
Figure. S3. Bayesian phylogenetic trees of folate- or purine-related genes inherited vertically in *Encephalitozoon* and other microsporidia (shown in white on black). Numbers at nodes represent Bayesian posterior probabilities (Left) and bootstrap proportions (Right) when higher than 0.8 and 70%, respectively. DHFR, dihydrofolate reductase; GK, guanylate kinase. The scale bar corresponds to the estimated number of amino acid substitutions per site. All trees are shown unrooted.



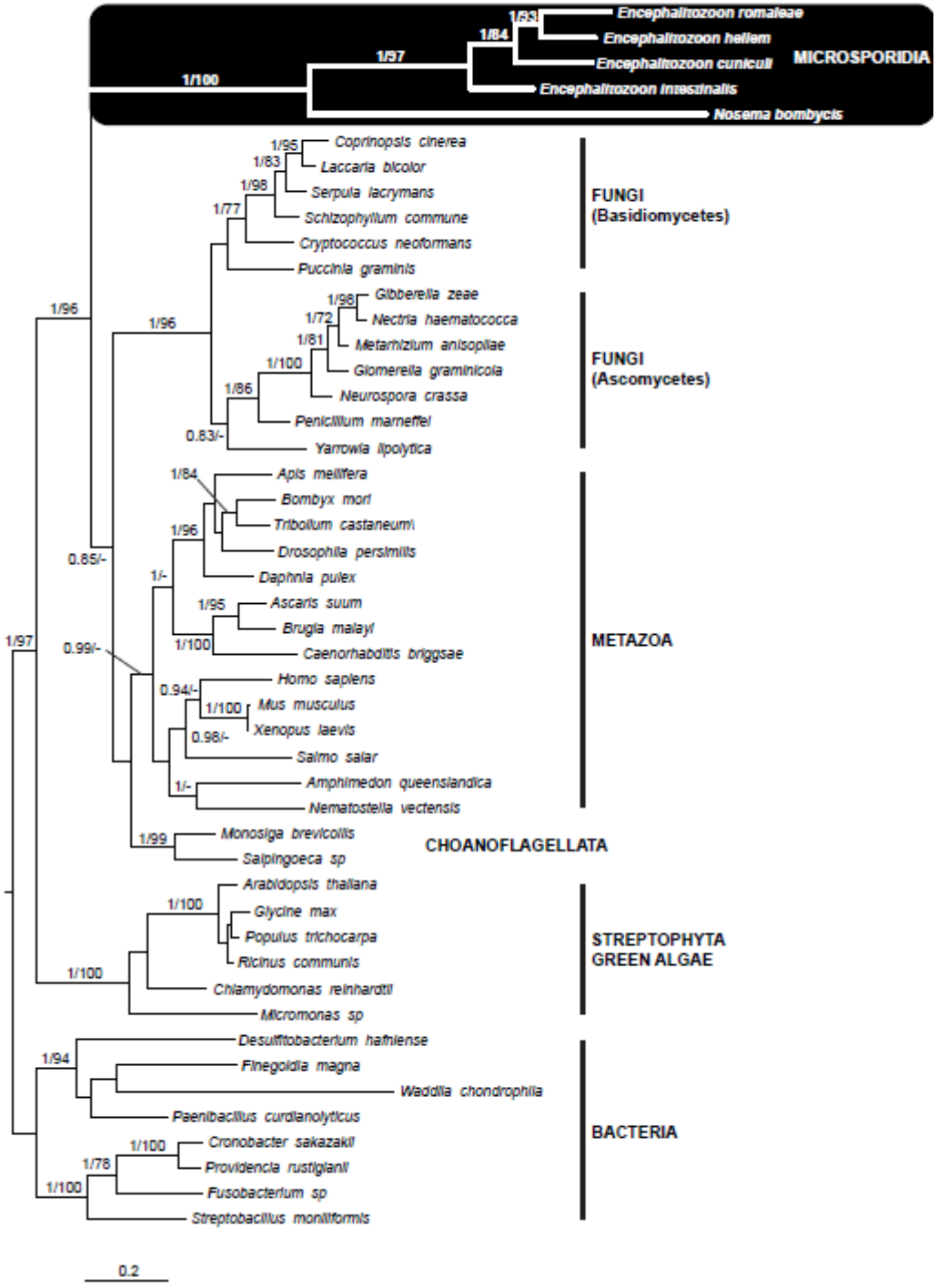
GK


Encephalitozoon romaleae
Encephalitozoon hellem
Encephalitozoon cuniculi
Encephalitozoon intestinalis
Nosema ceranae
Enterocytozoon bieneusi
MICROSPORIDIA
1/91
1/100
0.9/-
1/100
Loa loa 312065027
Brugia malayi 170575580
1/80
Gallus gallus
Xenopus laevis
0.81/-
Daphnia pulex
Ciona intestinalis
1/76
Branchiostoma floridae
Nematostella vectensis
METAZOA
0.95/72
Lachancea thermotolerans
0.98/-
Candida glabrata
Saccharomyces cerevisiae
Zygosaccharomyces rouxii
Meyerozyma guilliermondii
Yarrowia lipolytica
Coccidioides posadasii
FUNGI
(Ascomycetes)
0.96/79
Laccaria bicolor
1/93
Coprinopsis cinerea
Serpula lacrymans
Malassezia globosa
1/84
1/95
Ustilago maydis
Sporisorium reilianum
Puccinia graminis
Cryptococcus neoformans
FUNGI
(Basidiomycetes)
0.99/82
Thermovibrio ammonificans
0.95/79
Desulfurobacterium thermolithotrophum
0.84/-
Desulfurispirillum indicum
Ilyobacter polytropus
Cyanothece sp
Lyngbya majuscula
1/90
0.94/-
Nostoc punctiforme
cyanobacterium
0.99/-
Arthrospira platensis
0.9/-
Lactobacillus delbrueckii
Streptococcus thermophilus
0.86/-
Ruminococcus sp
Halopiasma contractile
Natranaerobius thermophilus
Rubrobacter xylanophilus
BACTERIA
0.2

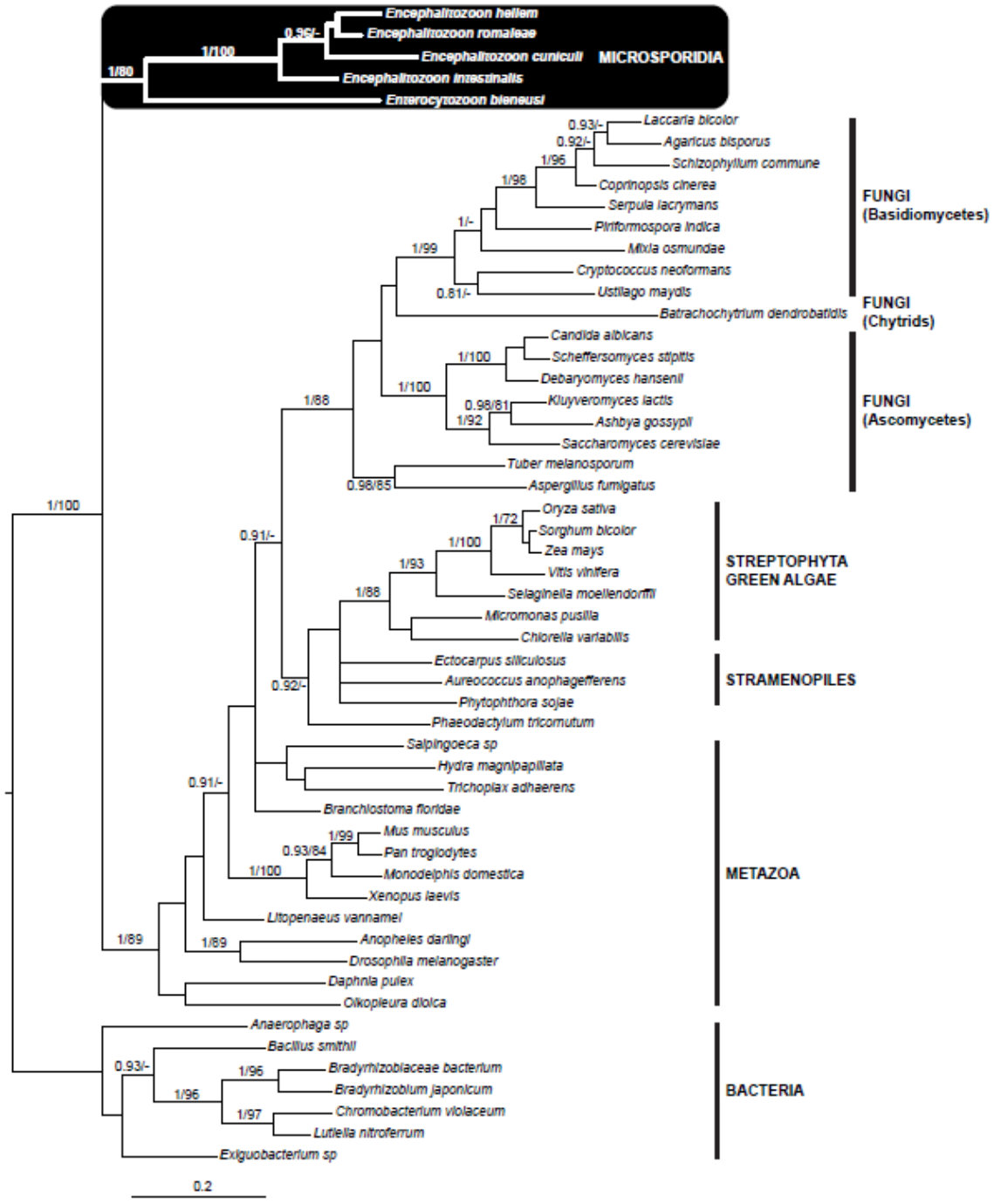
NDK


0.98/75
0.91/76
1/86
1/82
Encephalitozoon romaleae
Encephalitozoon hellem
Encephalitozoon cuniculi
Encephalitozoon intestinalis
Enterocytozoon bieneusi
Nosema ceranae
MICROSPORIDIA
Batrachochytrium dendrobatidis
FUNGI (Chytrids)
Penicillium chrysogenum
Trichophyton equinum
Talaromyces stipitatus
Gibberella zeae
Magnaporthe oryzae
Coccidioides immitis
Neosartorya fischeri
Ajellomyces dermatitidis
Aspergillus terreus
FUNGI (Ascomycetes)
Cryptococcus neoformans
Paxillus involutus
Coprinopsis cinerea
Laccaria bicolor
Schizophyllum commune
Malassezia globosa
FUNGI (Basidiomycetes)
Phytophthora infestans
Perkinsus marinus
STRAMENOPILES ALVEOLATES
Chelonus inanitus
Apis mellifera
Camponotus floridanus
Drosophila mojavensis
Anopheles gambiae
Daphnia pulex
Xenopus tropicalis
Gallus gallus
Danio rerio
Rattus norvegicus
Tetraodon nigroviridis
Callgus clemensi
Lepeophtheirus salmonis
METAZOA
Monosiga brevicollis
Trichoplax adhaerens
Capsaspora owczarzaki
Phaeodactylum tricornutum
Thalassiosira pseudonana
Ectocarpus siliculosus
STRAMENOPILES
Oryza sativa
Sorghum bicolor
Selaginella moellendorffi
Picea sitchensis
Physcomitrella patens
Ricinus communis
Brassica rapa
Arabidopsis lyrata
STREPTOPHYTA
Synechocystis sp
Geobacillus thermoglucosidasius
Macrococcus caseolyticus
Brevibacillus brevis
Thermosynechococcus elongatus
Arthrospira platensis
Synechococcus sp
Archaeoglobus profundus
BACTERIA
0.2

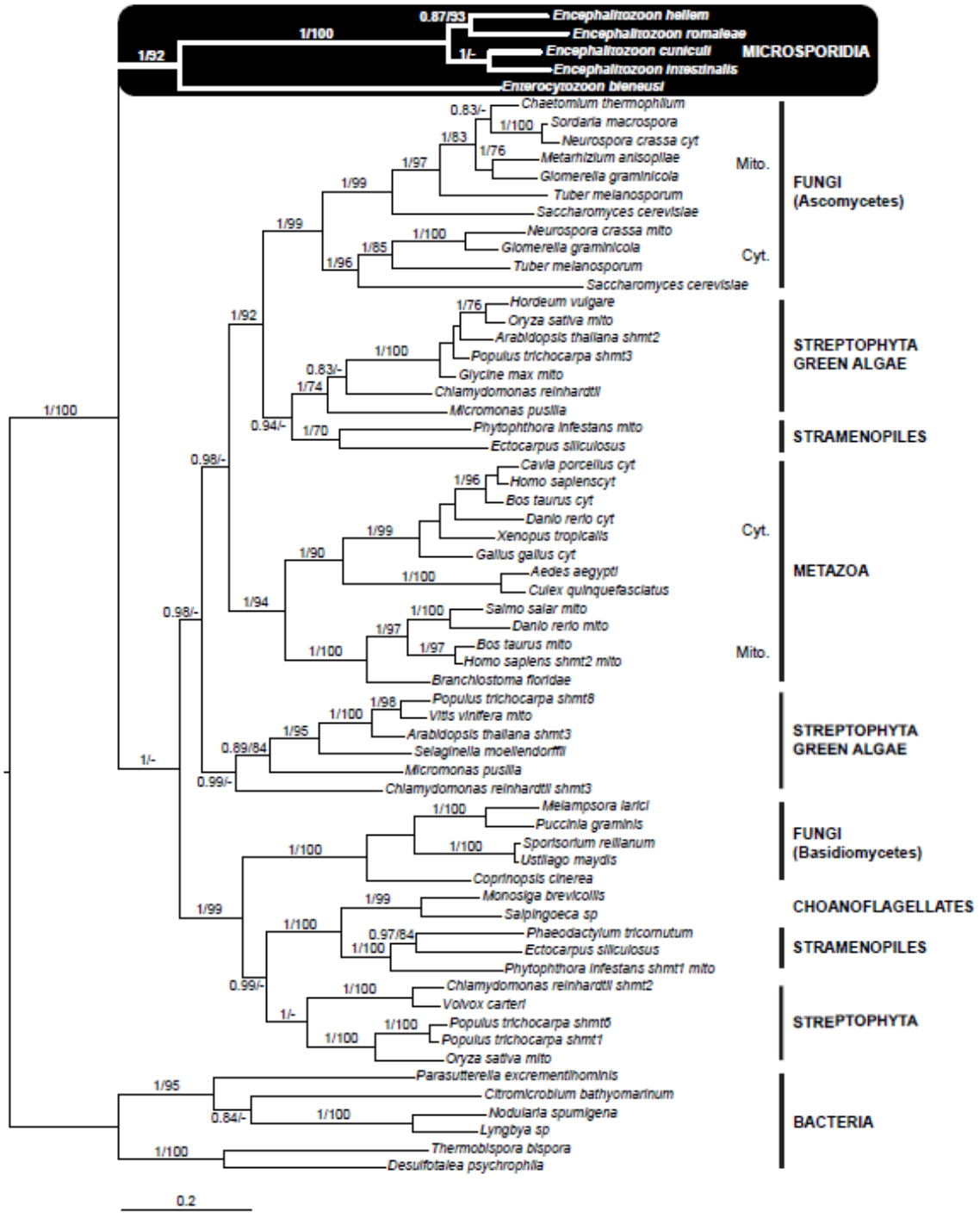
PK


Encephalitozoon romaleae
Encephalitozoon hellem
Encephalitozoon cuniculi
Encephalitozoon intestinalis
Nosema bombycis
MICROSPORIDIA
1/100
1/97
1/84
1/88
Coprinopsis cinerea
Laccaria bicolor
Serpula lacrymans
Schizophyllum commune
Cryptococcus neoformans
Puccinia graminis
FUNGI (Basidiomycetes)
Gibberella zeae
Nectria haematococca
Metarhizium anisopliae
Glomerella graminicola
Neurospora crassa
Penicillium marneffei
Yarrowia lipolytica
FUNGI (Ascomycetes)
Apis mellifera
Bombyx mori
Tribolium castaneum
Drosophila persimilis
Daphnia pulex
Ascaris suum
Brugia malayi
Caenorhabditis briggsae
Homo sapiens
Mus musculus
Xenopus laevis
Salmo salar
Amphimedon queenslandica
Nematostella vectensis
METAZOA
Monosiga brevicollis
Salpingoeca sp
CHOANOFLAGELLATA
Arabidopsis thaliana
Glycine max
Populus trichocarpa
Ricinus communis
Chlamydomonas reinhardtii
Micromonas sp
STREPTOPHYTA GREEN ALGAE
Desulfitobacterium hafniense
Finegoldia magna
Waddlia chondrophila
Paenibacillus curdlanolyticus
Cronobacter sakazakii
Providencia rustigianii
Fusobacterium sp
Streptobacillus moniliformis
BACTERIA
0.2

TS


Encephalitozoon hellem
Encephalitozoon romaleae
Encephalitozoon cuniculi
Encephalitozoon intestinalis
Enterocytozoon bieneusi
MICROSPORIDIA
Laccaria bicolor
Agaricus bisporus
Schizophyllum commune
Coprinopsis cinerea
Serpula lacrymans
Piriformospora indica
Mixia osmundae
Cryptococcus neoformans
Ustilago maydis
Batrachochytrium dendrobatidis
Candida albicans
Scheffersomyces stipitis
Debaryomyces hansenii
Kluyveromyces lactis
Ashbya gossypii
Saccharomyces cerevisiae
Tuber melanosporum
Aspergillus fumigatus
Oryza sativa
Sorghum bicolor
Zea mays
Vitis vinifera
Selaginella moellendorffii
Micromonas pusilla
Chlorella variabilis
Ectocarpus siliculosus
Aureococcus anophagefferens
Phytophthora sojae
Phaeodactylum tricornutum
Salpingoeca sp
Hydra magnipapillata
Trichoplax adhaerens
Branchiostoma floridae
Mus musculus
Pan troglodytes
Monodelphis domestica
Xenopus laevis
Litopenaeus vannamei
Anopheles darlingi
Drosophila melanogaster
Daphnia pulex
Oikopleura dioica
Anaerophaga sp
Bacillus smithii
Bradyrhizobiaceae bacterium
Bradyrhizobium japonicum
Chromobacterium violaceum
Lutiella nitroferrum
Exiguobacterium sp
FUNGI (Basidiomycetes)
FUNGI (Chytrids)
FUNGI (Ascomycetes)
STREPTOPHYTA GREEN ALGAE
STRAMENOPILES
METAZOA
BACTERIA
0.2

SHMT

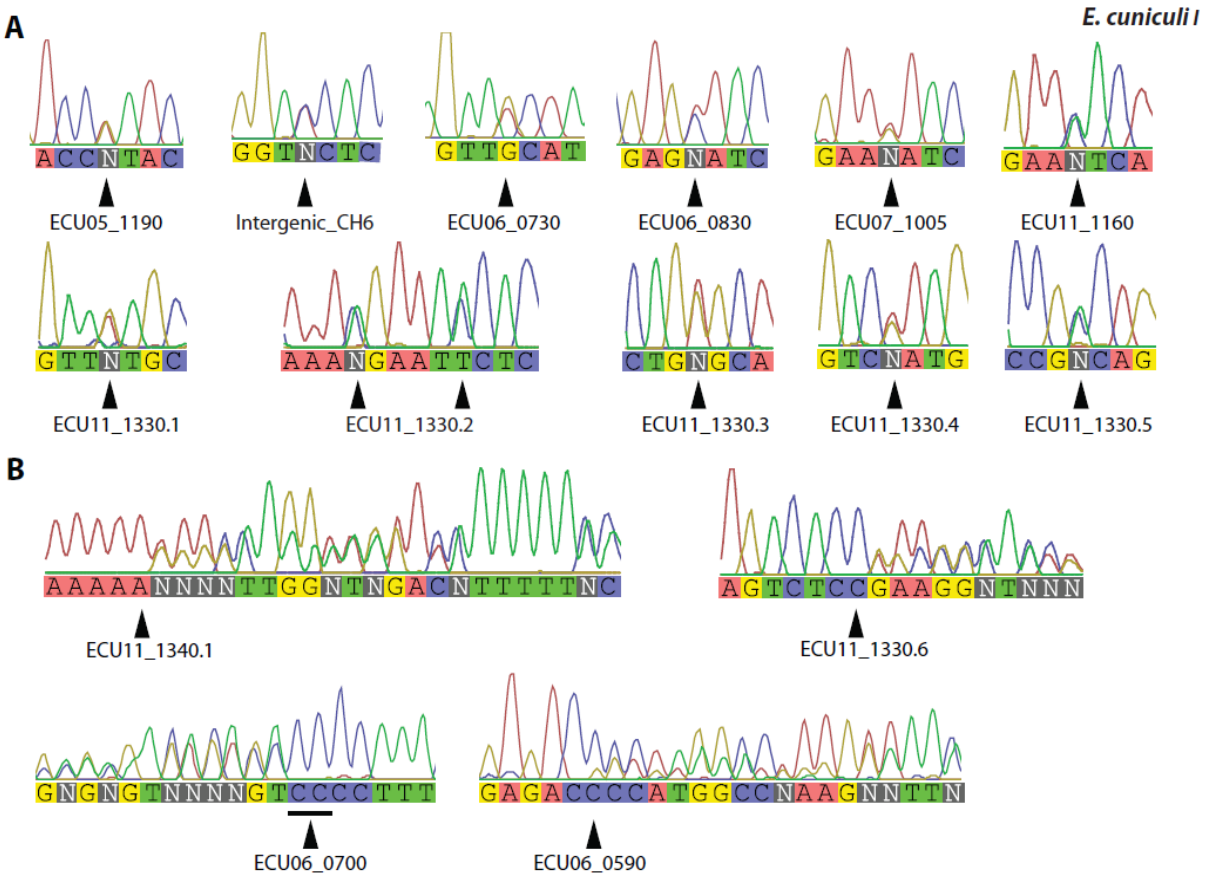

Encephalitozoon hellem
Encephalitozoon romaleae
Encephalitozoon cuniculi
Encephalitozoon intestinalis
Enterocytozoon bieneusi
MICROSPORIDIA
Chaetomium thermophilum
Sordaria macrospora
Neurospora crassa cyt
Metarhizium anisopliae
Glomerella graminicola
Tuber melanosporum
Saccharomyces cerevisiae
Neurospora crassa mito
Glomerella graminicola
Tuber melanosporum
Saccharomyces cerevisiae
Mito.
Cyt.
FUNGI (Ascomycetes)
Hordeum vulgare
Oryza sativa mito
Arabidopsis thaliana shmt2
Populus trichocarpa shmt3
Glycine max mito
Chlamydomonas reinhardtii
Micromonas pusilla
STREPTOPHYTA GREEN ALGAE
Phytophthora infestans mito
Ectocarpus siliculosus
STRAMENOPILES
Cavia porcellus cyt
Homo sapienscyt
Bos taurus cyt
Danio rerio cyt
Xenopus tropicalis
Gallus gallus cyt
Aedes aegypti
Culex quinquefasciatus
Salmo salar mito
Danio rerio mito
Bos taurus mito
Homo sapiens shmt2 mito
Branchiostoma floridae
Cyt.
Mito.
METAZOA
Populus trichocarpa shmt8
Vitis vinifera mito
Arabidopsis thaliana shmt3
Selaginella moellendorffii
Micromonas pusilla
Chlamydomonas reinhardtii shmt3
STREPTOPHYTA GREEN ALGAE
Melampsora larici
Puccinia graminis
Sporisorium reilianum
Ustilago maydis
Coprinopsis cinerea
FUNGI (Basidiomycetes)
Monosiga brevicollis
Salpingoeca sp
CHOANOFLAGELLATES
Phaeodactylum tricornutum
Ectocarpus siliculosus
Phytophthora infestans shmt1 mito
STRAMENOPILES
Chlamydomonas reinhardtii shmt2
Volvox carteri
Populus trichocarpa shmt6
Populus trichocarpa shmt1
Oryza sativa mito
STREPTOPHYTA
Parasutterella excrementihominis
Citromicrobium bathyomarinum
Nodularia spumigena
Lyngbya sp
Thermobispora bispora
Desulfotalea psychrophila
BACTERIA
0.2

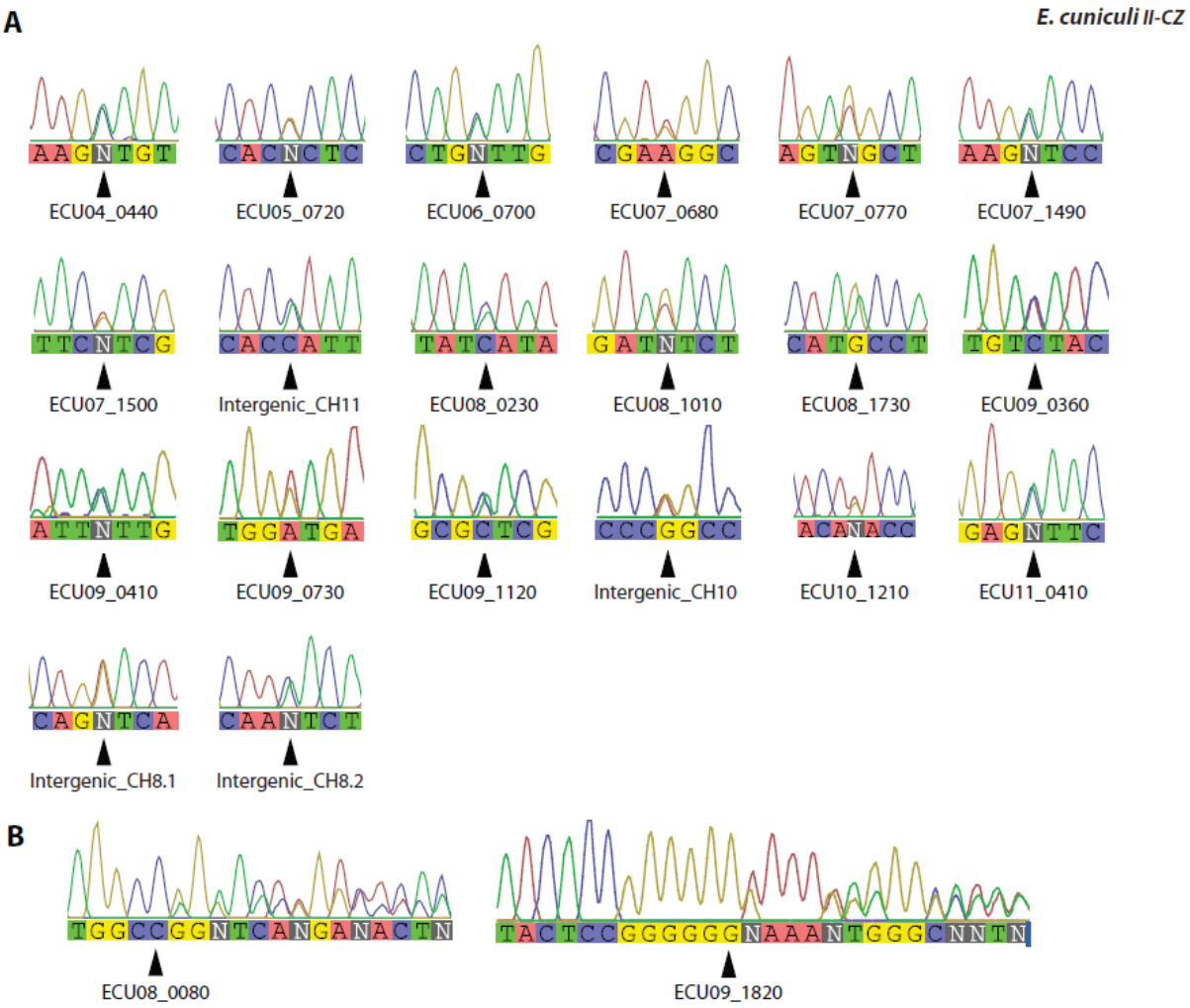
Dataset S1. Codon use in the *Encephalitozoon* species

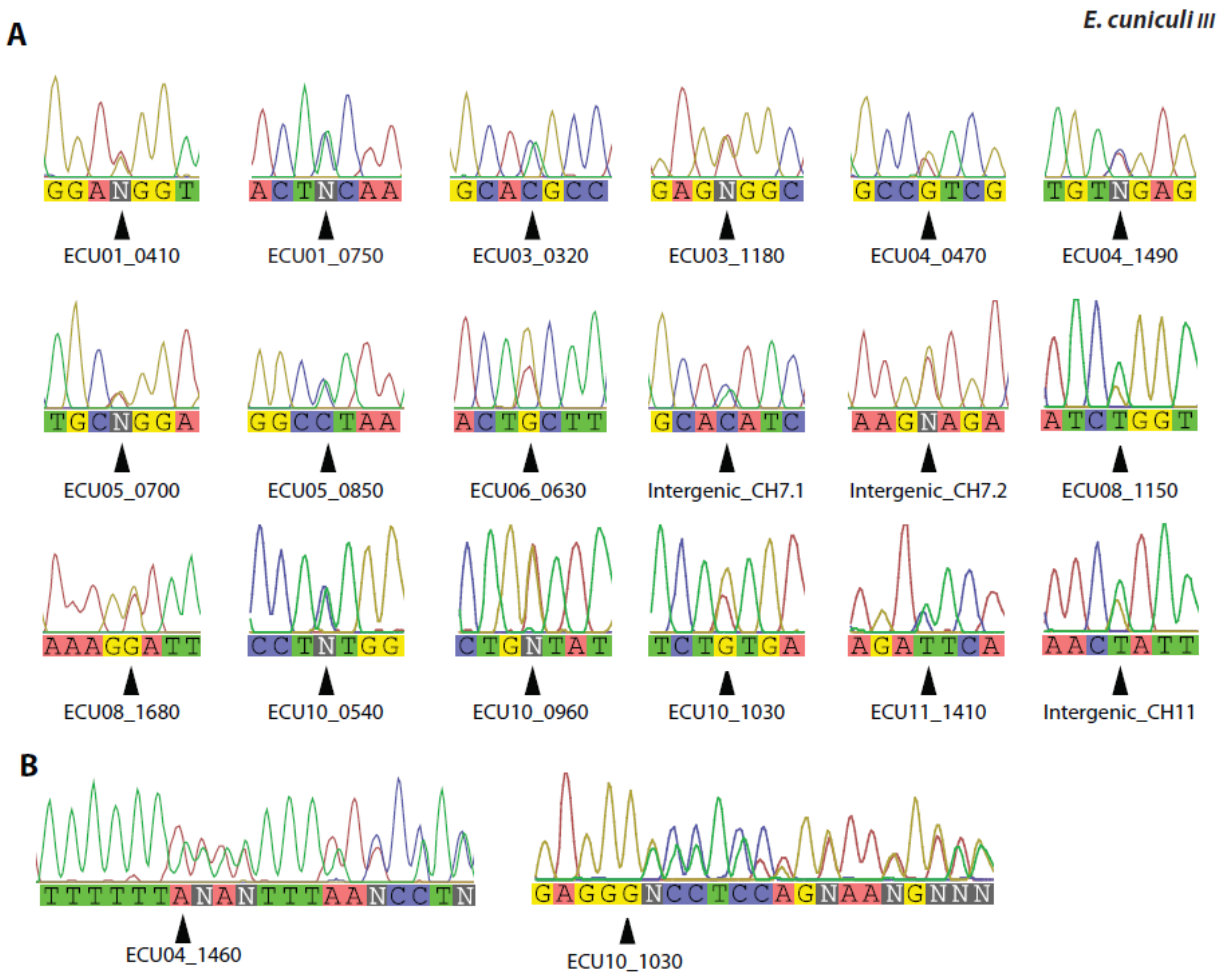
	T					C					A					G				
		E. cun	**E. int**	**E. hel**	**E. rom**		**E. cun**	**E. int**	**E. hel**	**E. rom**		**E. cun**	**E. int**	**E. hel**	**E. rom**		**E. cun**	**E. int**	**E. hel**	**E. rom**	
T	Phe	53%	61%	58%	61%	Ser	20%	23%	24%	26%	Tyr	46%	56%	52%	61%	Cys	34%	48%	41%	52%	**T**
	Phe	46%	38%	41%	39%	Ser	16%	15%	13%	14%	Tyr	53%	43%	47%	39%	Cys	65%	51%	58%	48%	**C**
	Leu	3%	8%	7%	11%	Ser	9%	12%	12%	16%	Ocr	100%	100%	100%	100%	Opl	100%	100%	100%	100%	**A**
	Leu	11%	16%	14%	17%	Ser	15%	13%	12%	10%	Amb	100%	100%	100%	100%	Trp	100%	100%	100%	100%	**G**
C	Leu	26%	29%	28%	30%	Pro	27%	33%	33%	38%	His	47%	63%	55%	67%	Arg	3%	3%	3%	4%	**T**
	Leu	17%	13%	14%	11%	Pro	22%	17%	19%	15%	His	52%	36%	44%	33%	Arg	4%	2%	2%	2%	**C**
	Leu	10%	12%	13%	14%	Pro	27%	32%	29%	34%	Gln	26%	44%	37%	48%	Arg	4%	7%	4%	6%	**A**
	Leu	31%	18%	22%	16%	Pro	22%	16%	18%	13%	Gln	73%	55%	62%	52%	Arg	8%	6%	5%	5%	**G**
A	Ile	31%	35%	34%	36%	Thr	17%	21%	21%	25%	Asn	39%	49%	47%	56%	Ser	12%	16%	14%	18%	**T**
	Ile	32%	24%	23%	22%	Thr	24%	20%	18%	19%	Asn	60%	50%	52%	44%	Ser	25%	17%	22%	17%	**C**
	Ile	36%	39%	41%	42%	Thr	35%	45%	38%	44%	Lys	24%	36%	30%	37%	Arg	36%	47%	42%	48%	**A**
	Met	100%	100%	100%	100%	Thr	23%	13%	20%	13%	Lys	75%	63%	69%	63%	Arg	43%	31%	41%	36%	**G**
G	Val	29%	34%	33%	35%	Ala	18%	22%	20%	25%	Asp	44%	55%	52%	63%	Gly	9%	9%	10%	12%	**T**
	Val	21%	16%	17%	15%	Ala	25%	22%	21%	20%	Asp	55%	44%	47%	37%	Gly	17%	10%	13%	11%	**C**
	Val	14%	20%	18%	23%	Ala	40%	47%	44%	46%	Glu	33%	48%	39%	49%	Gly	41%	54%	43%	48%	**A**
	Val	34%	28%	31%	27%	Ala	15%	8%	12%	8%	Glu	66%	51%	60%	51%	Gly	32%	26%	32%	30%	**G**

Appendix C : Supplemental information for Chapter 5

Supplemental Figure 1: *In vitro* validation of heterozygosity in *Encephalitozoon cuniculi*
Examples of validation of heterozygous loci using Sanger Sequencing. Chromatogram showing the heterozygous SNP and indel recurrently identified in *Encephalitozoon cuniculi* of genotype I, II, III. Polymorphism sites are identified by black arrow.




A
E. cuniculi II-CZ
AAGNTGT
ECU04_0440
CACNCTC
ECU05_0720
CTGNTTG
ECU06_0700
CGAAGGC
ECU07_0680
AGTNGCT
ECU07_0770
AAGNTCC
ECU07_1490
TTCNTCG
ECU07_1500
CACCATT
Intergenic_CH11
TATCATA
ECU08_0230
GATNTCT
ECU08_1010
CATGCCT
ECU08_1730
TGTCTAC
ECU09_0360
ATTNTTG
ECU09_0410
TGGATGA
ECU09_0730
GCGCTCG
ECU09_1120
CCCGGCC
Intergenic_CH10
ACANACC
ECU10_1210
GAGNTTC
ECU11_0410
CAGNTCA
Intergenic_CH8.1
CAANTCT
Intergenic_CH8.2
B
TGGCCGGNTCANGANACTN
ECU08_0080
TACTCCGGGGGGNAAANTGGGCNNTN
ECU09_1820


A
E. cuniculi III
GGANGGT
ECU01_0410
ACTNCAA
ECU01_0750
GCACGCC
ECU03_0320
GAGNGGC
ECU03_1180
GCCGTCG
ECU04_0470
TGTNGAG
ECU04_1490
TGCNGGA
ECU05_0700
GGCCTAA
ECU05_0850
ACTGCTT
ECU06_0630
GCACATC
Intergenic_CH7.1
AAGNAGA
Intergenic_CH7.2
ATCTGGT
ECU08_1150
AAAGGATT
ECU08_1680
CCTNTGG
ECU10_0540
CTGNTAT
ECU10_0960
TCTGTGA
ECU10_1030
AGATTCA
ECU11_1410
AACTATT
Intergenic_CH11
B
TTTTTTANANTTTAANCCTN
ECU04_1460
GAGGGNCCTCCAGNAANGNNN
ECU10_1030

Appendix D : Copyrights

PNAS

As a PNAS author, you and your employing institution or company retain extensive rights for use of your materials and intellectual property. You retain these rights and permissions without having to obtain explicit permission from PNAS, provided that you cite the original source:

The right to include your article in your thesis or dissertation

Current Biology

Permission asked from Rightslink service. See below.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 20, 2013

This is a License Agreement between Mohammed Selman ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Mohammed Selman
Customer address	30 montcalm st. Unit 15 Ottawa, ON K1S0A3
License number	3213300371461
License date	Aug 20, 2013
Licensed content publisher	Elsevier
Licensed content publication	Current Biology
Licensed content title	Acquisition of an animal gene by microsporidian intracellular parasites
Licensed content author	Mohammed Selman,Jean-François Pombert,Leellen Solter,Laurent Farinelli,Louis M. Weiss,Patrick Keeling,Nicolas Corradi
Licensed content date	9 August 2011
Licensed content volume number	21
Licensed content issue number	15
Number of pages	2
Start Page	R576
End Page	R577
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	print
Are you the author of this Elsevier article?	Yes
Will you be translating?	No

Eukaryotic cell

Authors may republish/adapt portions of their articles

ASM also grants the authors the right to republish discrete portions of his/her article in any other publication (including print, CD-ROM, and other electronic formats) of which he or she is author or editor, provided that proper credit is given to the original ASM publication. "Proper credit" means either the copyright lines shown on the top of the first page of the PDF version, or "Copyright © American Society for Microbiology, (insert journal name, volume number, year, page numbers and DOI)" of the HTML version.

Mobile Genetic Elements

Landes Bioscience grants it freely (via our License to Transfer or the CC-BY-NC license if your paper is published using the OA model). There is no need to secure rights from Landes Bioscience or The Copyright Clearance Center.

Journal of Eukaryotic Microbiology

Permission asked from Rightslink service. See below.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 20, 2013

This is a License Agreement between Mohammed Selman ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3213301306898
License date	Aug 20, 2013
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Eukaryotic Microbiology
Licensed content title	Latest Progress in Microsporidian Genome Research
Licensed copyright line	© 2013 The Author(s) Journal of Eukaryotic Microbiology © 2013 International Society of Protistologists
Licensed content author	Nicolas Corradi,Mohammed Selman
Licensed content date	Feb 27, 2013
Start page	309
End page	312
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Total	0.00 USD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or a society for whom a Wiley Company has exclusive publishing rights in relation to a particular journal (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).