Secreted proteins in microsporidian parasites: a functional and evolutionary perspective on host-parasite interactions.

Submitted by Scott Edward Campbell

to the University of Exeter as a thesis for the degree of
Doctor of Philosophy in Biological Science.

In September 2013

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature …….

Abstract

The Microsporidia form a phylum of obligate intracellular parasites known to cause disease in humans and a diverse range of economically important animal species. Once classified as 'primitive' eukaryotes, it is now recognised that the peculiarities of microsporidian genomics and cell biology are, in fact, the consequence of extreme reduction allowed by an intimate relationship with the host cell. Excluding survival as an extracellular spore, microsporidia are in direct contact with the host throughout their developmental lifecycle, from entry to egress. Host cell manipulations have been described in morphological terms, but despite this, characterisation of such processes at the molecular level remains challenging. The logistics of the microsporidian lifecycle suggest secreted proteins and membrane proteins with extracellular domains may be involved in virulence and implicated in host cell manipulation.

This study employs bioinformatic tools to predict secreted proteins in diverse microsporidia and comparative genomics to identify conserved proteins which may be required for host cell manipulation, pathogenicity and lifecycle progression. The protein complement secreted into the extracellular environment during microsporidian spore germination, a lifecycle stage required for host cell invasion, is identified experimentally. This analysis suggests that novel microsporidian specific hypothetical proteins, that is, proteins with no functional annotation or domain, play a significant role during parasite invasion of the host and provides the first identification of potential microsporidian effector proteins. Aiming to address microsporidian pathogenicity during intracellular stages, candidate virulence factor proteins, namely a hemolysin and a protein tyrosine phosphatase are also characterised and localised *in situ*.

Lastly, an animal-derived horizontal gene transfer event is used in conjunction with both the fossil record and molecular dating approaches to add timescale to the microsporidian diversification. This work suggests that microsporidia radiated recently, achieving extreme cellular diversity, acquiring a novel infection mechanism and undergoing vast speciation in a short evolutionary timescale, likely within the last 200 million years.

Table of contents

List of Figures

List of Tables

Acknowledgements

Firstly I would like to thank my project supervisor, Bryony Williams, for providing support, ideas and encouragement in perfect harmony with freedom and independence. This gave me the opportunity to explore my own ideas and make the project my own, for that I am indebted.

I would like to thank everybody who has contributed to this project in one form or another; however, I must notably thank Tom Richards, Darren Soanes, Tom Williams and Patrick Hamilton for their ongoing collaboration, constructive criticism and advice.

I would like to thank all my friends in Scotland and Exeter for providing humor, great nights out, some often well needed escapism from science and just generally being great friends.

Lastly, I would like to thank my Mum, Dad, Sister, Ritchie, Gran Campbell and Gran Doyle for all your love, support, encouragement and friendship throughout both my PhD and previous studies in Glasgow.

Chapter 1| General Introduction

1.1| An introduction to the microsporidia

Microsporidia comprise a diverse phylum of obligate intracellular parasites consisting of approximately 150 genera and more than 1300 identified species (Lee *et al*, 2010). Collectively, they are known to infect a wide range of hosts ranging from protists such as apicomplexans to humans (Vivier, 1975). Originally classified as pathogens of the immuno-compromised, microsporidian infection has since been identified and is increasingly prevalent in immuno-competent individuals (Cotte *et al*, 1999). The microsporidian species *Nosema bombycis* was first discovered over 150 years ago as the causative agent of pébrine disease, a condition responsible for destruction of the European silk industry (Nägeli, 1857). Subsequent scientific studies were conducted by Louis Pasteur in which he devised a successful method for *N. bombycis* control, thus rejuvenating the European silk industry (Pasteur, 1870). Nägeli initially described *N. bombycis* as a 'yeast like' pathogen, classifying it as a member of the Schizomycetes. However, microsporidia have since undergone numerous events of taxonomic re-classification based on morphological and cellular properties (Figure 1.1) (for review see Corradi and Keeling, 2009). As cell biology advanced, it became evident that microsporidia were 'unusual' in terms of their organellar composition. As microsporidia ostensibly lacked mitochondria, peroxisomes and centrioles but possessed prokaryote like 70S ribosomes, they were deemed to be early branching and primitive eukaryotes (Vossbrinck and Woese, 1986, Cavalier-Smith, 1987). These apparently primitive features led to formulation of the Archezoa hypothesis which proposed that the Microsporidia, along with parabasalids, Archamoebae and Metamonada, were

an early branching eukaryotic phylum which diverged prior to the endosymbiotic engulfing of the alpha-proteobacterium, which later became transformed into the mitochondrion via secondary reduction (Cavalier-Smith, 1983). The Archezoa hypothesis was supported by initial phylogenetic analysis of the SSU rRNA gene that placed microsporidia as the earliest branching eukaryotes (Vossbrinck *et al*, 1987). However, as phylogenetic methodology advanced it became apparent that the primitive placement of the Archezoan groups, including microsporidia, was a product of long-branch attraction in phylogeny caused by fast evolving divergent sequences (Vossbrinck *et al*, 1987, Bergsten, 2005). Phylogenetic studies using more advanced evolutionary models later revealed that microsporidia were in fact related to fungi (Edlind *et al*, 1996, Keeling and Doolittle, 1996, Hirt *et al*, 1999, Van de Peer *et al*, 2000). Numerous phylogenetic (for examples Keeling, 2003, James *et al*, 2006, Capella-Gutierrez *et al*, 2012) and gene synteny (Lee *et al*, 2008) studies have proposed several alternative microsporidian-fungal relationships, however the exact placement of the microsporidia within or as a sister group to the Fungi remains unresolved. It is clear that the microsporidian characteristics that originally defined them as primitive are a product of secondary reduction due to their obligate intracellular lifestyle.

Microsporidian infection of humans was first identified in 1959, however incidence remained low until the mid 1980's (Franzen and Müller, 2001). To date approximately 15 human-infective microsporidian species have been identified (Table. 1.1). Increased infection prevalence co-incided with the

AIDS epidemic thus indicating the potential for microsporidia to cause disease in immuno-compromised individuals (Desportes *et al*, 1985). Microsporidian infection of the immuno-compromised is not restricted to AIDS patients but has also been described in organ transplant patients receiving immuno-suppressive therapy (Rabodonirina *et al*, 1996, Sing *et al*, 2001). Microsporidia are known to infect all major human organ systems thus inflicting a diverse range of disease conditions and symptoms on human hosts including encephalitis, myositis, pulmonary infections, gastrointestinal infections, hepatitis and in most severe cases systemic disseminated infection and death (Orenstein *et al*, 1992). Microsporidian infection has also been recognised in immuno-competent individuals, normally characterised by self-limiting cases of gastrointestinal infection or keratoconjuntivitis (Sandfort *et al*, 1994, Bommala *et al*, 2011). Although usually mild, conditions including myositis and encephalitis have been observed in immuno-competent individuals and in rare cases can be fatal (Coyle *et al*, 2004, Ditrich *et al*, 2011). Identification in immuno-competent individuals has been recorded at high prevalence in healthy populations, for example *Enterocytozoon bieneusi* infection was identified in 67.5% of healthy individuals in a study in Cameroon (Nkinin *et al*, 2007). It remains unclear if the apparent infection increase is a consequence of the improvement in diagnosis sensitivity or is a true reflection of the state of current infection. Diagnosis of microsporidian infection has traditionally been achieved via microscopy based analysis of faeces and other body fluids such as duodenal aspirate (Weber *et al*, 1992). This requires morphological determination of spores based on their size and cellular properties and can be enhanced with the use of stains such as Giemsa stain

or fluorescence markers such as calcofluor white which stains spore wall chitin (Vavra and Hollisterz, 1993, Kotler *et al*, 1994). Molecular techniques including polymerase chain reaction (PCR) and serological methods such as enzyme linked immuno-sorbent assay (ELISA) have been established for microsporidian diagnosis. Despite limitations in terms of time and cost, these methods offer increased sensitivity and accurate species differentiation (da Silva *et al*, 1996, Enriquez *et al*, 1997). A range of drug treatments have been used to combat microsporidian infection with varying success depending on the site of infection (Costa and Weiss, 2000). Common human-infective microsporidian species including members of the genera *Encephalitozoon* and *Enterocytozoon bieneusi* are most commonly treated with albendazole and fumagillin or fumagillin respectively (Molina *et al*, 2002). It has been suggested that *E. bieneusi* resistance to albendazole is the result of a substitution at glutamine 198 of β-tubulin, one of six amino acid sites thought to be associated with drug sensitivity (Akiyoshi *et al*, 2007). These drugs are active against infection in a range of tissues and have been shown to actively combat infection in immuno-compromised (Dieterich *et al*, 1994, Molina *et al*, 2002).

Microsporidia are increasingly important veterinary pathogens causing infection, pathology and mortality in species of economic importance such as the honey bee *Apis mellifera*, Atlantic salmon species *Salmo salar* and the silkworm *Bombyx mori*. Colony collapse disorder (CCD) is a phenomenon in which honey bee colonies mysteriously lose their workers, thus decimating bee populations and the potential for crop pollination. It is prevalent in both

Europe and North America where it has been identified to cause up to 90% colony loss (Cox-Foster *et al*, 2007). The causes of CCD are thought to be multi-factorial with chemical crop treatments, environmental stresses and a range of pathogens linked to disease outbreaks. These pathogens include the microsporidian species *Nosema ceranae* and *Nosema apis* (Johnson *et al*, 2009). *Nosema* spp. have been shown to alter honey bee flight behaviour and pheromone production, inflict a nutritional stress and cause mortality in hives both experimentally and in apiaries (Higes *et al*, 2008, Mayack and Naug, 2009, Dussaubat *et al*, 2013). Due to the necessity of honey bee populations for crop pollination, CCD is a threat to global food security and understanding the contribution and importance of *N. apis* and *N. ceranae* to this phenomenon is a matter of worldwide concern. Microsporidial salmon gill disease (MSGD) is a condition seen in various species of salmon and trout, characterised by inflammation and cyst formation in infected gills thus leading to anoxia (Shaw *et al*, 1998). MSGD is predominantly caused by the microsporidian species *Loma salmonae* and can cause up to 30% mortality, inflicting a huge economic burden particularly in aquaculture-raised salmon in dense populations (Speare *et al*, 2007). As wild salmon populations continue to decrease, the need for producing salmon efficiently and in a cost effective manner is obvious. Understanding the biology, transmission and pathology of microsporidian pathogens of salmon is crucial to the development of this industry. *N. bombycis*, the causative agent of pébrine disease in the silkworm species *B. mori,* can cause complete collapse of local sericulture industry. Infections can persist chronically or in highly virulent outbreaks in which parasites are spread by both horizontal and vertical means (Bhat *et al*, 2009).

N. bombycis multiplies more quickly in immature larvae leading to high parasite infection load and forms cysts in the midgut, silk gland and gonads thus decreasing silk production and fertility (Jyothi and Patil, 2011).

In summary, microsporidia are a diverse phylum with the capacity to cause diverse pathology in a diverse range of host species including humans and economically important animals. Increased understanding of their biology, virulence and transmission will be required to develop new and more effective diagnosis and treatment methods to ensure the consequences of infection are kept to a minimum.

Species	Natural host(s)	Pathology
Brachiola algerae	**Human**, mosquito	Myositis
Brachiola connori	**Human**	Myositis
Brachiola vesicularum	**Human**	Myositis
Encephalitozoon cuniculi	**Human**, rabbit, rodents, fox, goat, horse, ticks	Encephalitis, myositis, hepatitis, sinusitis, keratoconjunctivitis
Encephalitozoon hellem	**Human**, chicken, parrot, ostrich, finch, budgerigar, ticks	Encephalitis, myositis, hepatitis, sinusitis, keratoconjunctivitis
Encephalitozoon intestinalis	**Human**, gorilla, donkey, dog, pig, cow, goat, tick	Encephalitis, myositis, hepatitis, sinusitis, keratoconjunctivitis
Enterocytozoon bieneusi	**Human**, cattle, rabbit, cat, chicken, dog, goat, pig, ticks	Diarrhoea, pulmonary infection
Loma salmonae	Salmon, trout	Salmon gill disease
Microsporidium africanum	**Human**	Ocular infection
Microsporidium ceylonensis	**Human**	Ocular infection
Nosema bombycis	Silkworm	Pebrine disease
Nosema ceranae	Honeybee	Colony collapse disorder
Nosema connori	**Human**	Disseminated infection
Nosema ocularum	**Humans**	Ocular infection
Paranucleospora theridon	Salmon, trout	Salmon gill disease
Pleistophora ronneafiei	**Human**, fish	Skeletal muscle infection
Trachipleistophora anthropopthera	**Humans**	Myositis, encephalitis
Trachipleistophora hominis	**Humans**	Hepatitis, ocular infection, myositis
Vittaforma corneae	**Humans**	Ocular infection, sinusitis, myositis

Table 1.1| Microsporidian species of medical and veterinary importance.

To date 15 human-infective species have been identified, with the ability to cause a diverse range of disease symptoms. Many human-infective species such as *Encephalitozoon cuniculi* and *Enterocytozoon bieneusi* have a diverse animal host range providing potential for zoonotic spread of infection. The major veterinary microsporidian infections that are of great economic importance are also described. This includes *N. ceranae* and *L. salmonae*, threats to food security, and *N. bombycis*, a threat to the sericulture, a major industry in many parts of Asia including China. Adapted from Didier, 2005.

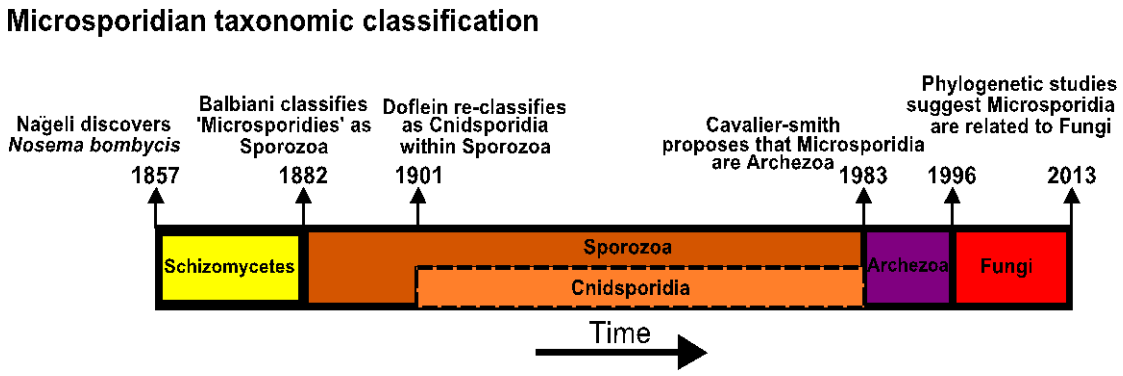


Figure 1.1| Taxonomic re-classification of the microsporidia across time. The microsporidian species *N. bombycis* was first discovered in 1852 and was classified as a member of the Schizomycetes. In 1882 the term 'Microsporidies' was proposed for the group and they were re-classified as members of the Sporozoa along with the Apicomplexa and Myxosporidia. Despite formulation of the sub-group Cnidsporidia, the microsporidia were classified as Sporozoa for approximately 100 years. In 1983, microsporidia were hypothesised to belong to the early-branching eukaryotic group, the Archezoa. From 1996 to present phylogenetic advances have shown a relationship between microsporidia and fungi. Adapted from Keeling, 2009.

1.2| The microsporidian lifecycle: how is life on the inside?

Microsporidia exist in the extracellular environment as spores ranging in size from 1-20 μm (Vemuganti *et al*, 2005). Microsporidian spores were once thought to be metabolically dormant; however studies in *Antonospora locustae* and *Trachipleistophora hominis* have since identified active glycolysis in extracellular lifecycle stages through immuno-localisation and proteomics (Dolgikh *et al*, 2011, Heinz *et al*, 2012). The spore wall consists of a proteinaceous exospore and an electron dense chitinaceous endospore which in combination provide extreme resistance to harsh environmental conditions. For example, spores have been shown to remain viable after 24 months in water under freezing conditions (Bigliardi *et al*, 1996, Koudela *et al*, 1999). Upon receiving an appropriate stimulus from a host cell, microsporidian spores undergo germination, the transitional step to their infectious lifecycle stage (Figure 1.2). Although the specific host stimuli required for germination are thought to be both complex and varied, germination can be induced successfully *in vitro* in some species by a variety of methods. For example, in *Spraguea lophii* and related species germination can be induced efficiently through addition of Ca^{2+} ions accompanied with an alkaline pH shift (Pleshinger and Weidner, 1985). This germination event triggers a huge transition in the microsporidian cell biology, characterised by an increased osmotic pressure and swelling of the posterior vacuole and polaroplast which in turn drives expulsion of the unique microsporidian infection apparatus, the polar tube, out of the cell (Xu and Weiss, 2005). The mechanism for this rapid increase in osmotic pressure remains contentious, however several possible explanations have been proposed. Firstly, it has been hypothesised that aquaporin proteins transport water directly into the spore or vacuole. *E. cuniculi* aquaporin (EcAQP)

has previously been shown to render *Xenopus laevis* oocysts permeable to water but has never been localised to the *E. cuniculi* posterior vacuole or characterised at the molecular level during the germination process (Ghosh *et al*, 2006). Secondly, it has been proposed that the increase in osmotic pressure is due to the trehalase-mediated breakdown of trehalose, a major fungal carbohydrate store, into glucose monomers, thus creating a suitable osmotic potential to induce water influx into the hypertonic spore (Dolgikh and Semenov, 2003). It has been estimated in the microsporidian species *Nosema algerae* (now *Anncaliia algerae*) that trehalose levels decrease by approximately 70% following germination, of which 70-78% is converted to glucose (Undeen and Meer, 1994). This trehalose decrease is not evident in other microsporidian species suggesting it is not the universal mechanism for turgor pressure increase. Other hypotheses include: 1) The release of calcium from the endomembrane system during early germination to increase the spore osmotic potential (Keohane and Weiss, 1998) or 2) An induction of spore swelling via the breakdown of hydrogen peroxide into oxygen and water by a catalase in the posterior vacuole, hypothesized to be a divergent peroxisome (Weidner and Findley, 2003). Genomic studies to date have provided no evidence for a peroxisomal catalase gene in microsporidia. It is possible that these mechanisms act in a multi-factorial manner to cause a rapid increase in cellular osmotic pressure or that they act individually depending on the specific host cell stimuli received. The stimuli required for germination are also likely to vary from species to species.

Once osmotic pressure reaches a suitable level, the polar tube is forced out of spore at the apical tip, the thinnest region of the spore wall. The polar

tube, which can be up to 100 μm in length and incredibly almost 100 times the length of the spore itself, transports infectious material directly into the host cell environment (Frixione *et al* 1992). Proteomic studies combined with immuno-fluorescence and immuno-electron microscopy have identified several components of the polar tube, termed polar tube proteins (PTP) (Dolgikh *et al*, 2005). PTP1, PTP2 and PTP3 of the microsporidian species *E. cuniculi* have been shown to intimately interact using yeast-2-hybrid analysis, however the mechanism of polar tube assembly remains unclear (Bouzahzah *et al*, 2010). Interestingly, PTP2 and PTP3 have also been show to interact with spore wall protein 5 (SWP5) in immuno-precipitation experiments in the microsporidian species *N. bombycis* (Li *et al*, 2012). The microsporidian polar tube when compared to other infection apparatus such as the appressorium of the ascomycete *Magnaporthe oryzae* is unique in that it is a pre-formed structure allowing rapid infection upon host cell contact (Williams, 2009). It was originally accepted that the polar tube acted as a 'hypodermic needle' directly puncturing the host cell membrane and thus injecting the infectious sporoplasm directly into the intracellular environment (Frixione *et al*, 1992). However, since histological evidence of polar tube penetration of the host cell membrane has remained elusive it has since been proposed that the polar tube, upon contact with the host cell membrane may induce an endocytic event causing invagination of the host cell membrane which then mediates phagocytic transport of infectious material into the cell (Bohne *et al*, 2011). Recent studies have identified the uptake of *N. bombycis* via phagocytosis, independently of germination in insect cells. However, post cell entry, these spores did not undergo intracellular development, demonstrating the importance of the germination process in initiating and establishing infection (Shun-feng *et al*, 2012). It has been

demonstrated that microsporidia adhere to host cell surface glycosaminoglycans via microsporidian endospore protein EnP1, which has been shown to modulate spore adherence to the host cell membrane *in vitro* in the genus *Encephalitozoon* via heparin binding motifs. Spore treatment with EnP1 antibody reduced both adherence and intracellular infection prevalence; however, the specific role of EnP1 in the overall infection cascade is currently unknown (Southern *et al*, 2007).

Once inside the host cell, microsporidia begin their second lifecycle stage, termed ‘merogony’, in which the polaroplast is thought to form a membrane for the infectious sporoplasm which has been transported via polar tube (Weidner *et al*, 1984). Intracellular development involves replication either directly in the host cell cytoplasm as seen in honey bee pathogen *N. ceranae* or in a parasitophorous vacuole (PV) as seen in human-infective pathogen *E. cuniculi*. Microsporidian meronts, which lack the robust spore wall and polar tube, replicate via either binary or multiple fission, this is followedby development through sporogony into a mature spore (Canning, 1988). This maturation process is characterised by a decrease in cell size and synthesis of the chitinaceous endospore, a process which involves significant morphological change (Keeling and Fast, 2002)**.** Within an infected host cell there is a catalogue of varying parasite developmental stages present, hence studying lifecycle transitions at the molecular level, remains an arduous task (Figure 1.2). Despite this, several genes including chitinases, chitin synthases and subtilisin proteases, have been shown to be developmentally regulated, demonstrating an increase in chitin synthase gene expression during meront to spore maturation (Ronnebaumer *et al*, 2006). Several studies have investigated the

origin and function of the PV in the microsporidian lifecycle and host interaction. The vacuole is not universally present across the microsporidia but is present in many human-infective species such as *Encephalitozoon intestinalis* and *Encephalitozoon hellem*. The PV can be identified in a primitive form at 1 minute post infection and has been shown to lack host cell endocytic pathway markers such as endosomal autoantigen 1 but possess host cell lipids (Fasshauer *et al*, 2005). To-date no parasitic proteins or lipids have been localised specifically to the PV membrane. The role of the PV in the microsporidian-host interaction remains unknown, however pores have been identified in the PV membrane which allow nutrient and metabolite influx into the intravacuolar space (Ronnebaumer *et al*, 2008). The pore size excludes passage of proteins bigger than 10 kDa, thus suggesting an alternative mechanism for vacuolar protein transport. Intravacuolar parasite development is an organised process in which meronts are thought to localise primarily to the vacuole periphery with spores and sporonts located internally, all of which are intermingled by a branching network of proteinaceous filaments with unknown function (Bohne *et al*, 2000, Ghosh *et al*, 2011).

Both the cues that prompt microsporidian egress and the mechanism by which they exit the host cell are currently unknown. It is possible that microsporidia secrete virulence factor proteins to aid egress, a function previously identified for putative proteases in the apicomplexan parasite *Plasmodium falciparum* (Yeoh *et al*, 2007). Recently it has been demonstrated that microsporidian species *Nematocida parisii* undergoes egress from intestinal cells of natural host *Caenorhabiditis elegans* via a non-lytic mechanism involving host actin and cytoskeletal re-arrangements, thus identifying a novel,

non-damaging mechanism for cell egress (Estes *et al*, 2011). Irrespective of the mechanism, upon host cell egress microsporidian spores are free to either persist in the extracellular environment or, if appropriately induced, can germinate again and re-infect a neighbouring host cell in a rapidly continuing infectious cycle, as is seen in *Vairimorpha necatrix* infection of the Beet Armyworm, *Spodoptera exigua* (Solter and Maddox, 1998).

An extreme intimacy between host and parasite is clear throughout the microsporidian developmental lifecycle, however the molecular mechanisms that underpin this interaction remain largelyunknown. Future molecular and cellular studies will provide new molecular understanding of how microsporidia sense, infect, replicate and egress from their host in what is certain to be an extremely complex catalogue of interactions.

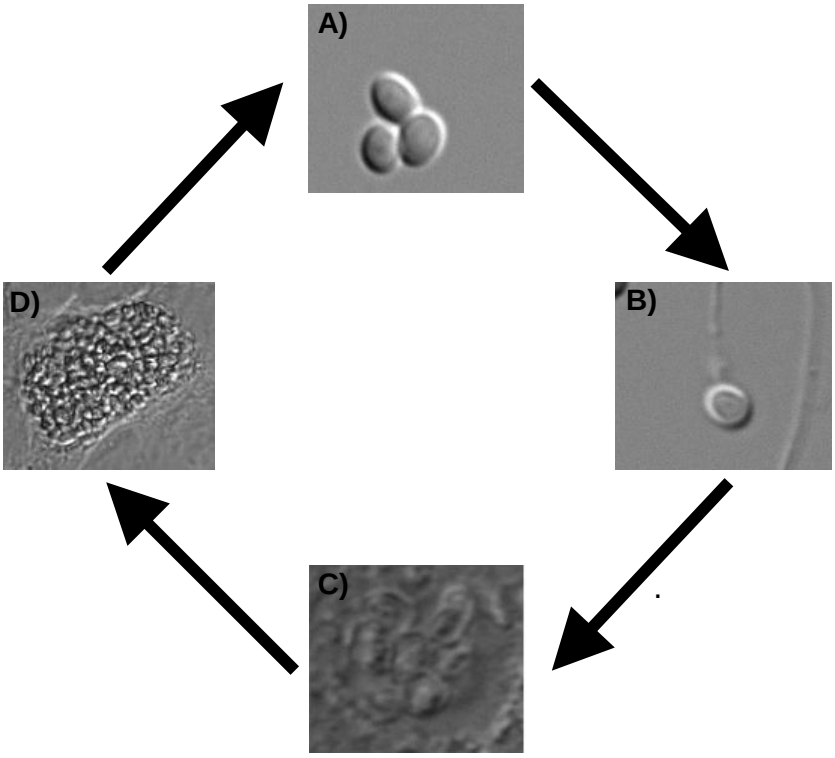


Figure 1.2| Overview of the microsporidian lifecycle.

A) Extracellular spores germinate upon receiving host cell stimuli and expel their polar tube. B) Infectious cellular material passes through the polar tube directly into the host cell environment. C) Meronts are formed and begin replication in immature vacuoles or cytoplasm. D) Mature infection in vacuole or cytoplasm characterised by an assortment of developmental stages prior to egress.

1.3| The microsporidian mitosome

Microsporidia were once hypothesised to be one of the earliest branching eukaryotic forms based on their unusual cell biology, which was originally interpreted as primitive (Cavalier-Smith, 1983). It appeared that they, along with various parasitic protists such as *Giardia lamblia*, *Trichomonas vaginalis* and *Entamoeba histolytica* and anaerobic fungi, lacked mitochondria, an otherwise universal eukaryotic organelle (Cavalier-Smith, 1987). The hypothesis that microsporidia diverged early in eukaryotic evolution initially received strong support based on both ribosomal characteristics and RNA sequence analysis (Vossbrinck and Woese, 1986, Vossbrinck *et al*, 1987)

However, following the identification of mitochondrial genes in diplomonads and trichomonads (Soltys and Gupta, 1994, Roger *et al*, 1996), an α-proteobacterial-derived mitochondrial Hsp70 gene was identified in the microsporidian species *V. necatrix*, *Antonospora locustae* and *E. cuniculi* (Germot *et al*, 1997, Hirt *et al*, 1997, Peyretaillade *et al*, 1998). This represents a hallmark of the past presence of mitochondria and re-established the theory that the ancestral eukaryote contained mitochondria. Identification of other classical mitochondrial markers such as the alpha and beta subunits of pyruvate dehydrogenase soon followed (Fast and Keeling, 2001). However these sequences lacked a recognisable mitochondrial targeting sequence, suggesting that they may be adapted to function cytosolically. The publication of the *E. cuniculi* genome provided more evidence of mitochondrial markers in microsporidia adding more weight to the idea that microsporidia once contained mitochondria (Katinka *et al*, 2001). Interestingly, the *E. cuniculi* genome data

demonstrated the absence of genes for oxidative phosphorylation and the Krebs cycle suggesting that microsporidia respire anaerobically.

Experimental validation of mitochondrial presence in microsporidia was lacking, despite identification of a reduced mitochondrial form, the mitosome, in the parasitic species *E. histolytica* (Tovar *et al*, 1999). Through localisation of Hsp70 in the microsporidian species *T. hominis* it was demonstrated that microsporidia also possessed mitosomes, tiny double membrane bound organelles which lack a mitochondrial genome (Williams *et al*, 2002). To determine the functional capacity of microsporidian mitosomes, the localisation of several traditional mitochondrial markers has been investigated, such as ferredoxin and glycerol-3-phosphate dehydrogenase in *E. cuniculi*, localising to the mitosome and cytoplasm respectively (Williams *et al*, 2008). The mitosomal localisation of ferredoxin suggests that iron-sulphur cluster assembly is carried out in the mitosome, a process demonstrated in *Saccharomyces cerevisiae* as the only truly essential mitochondrial function (Lill and Kispal, 2000). This was verified with localisation and characterisation of iron-sulphur cluster assembly components frataxin, cysteine desulphurase and scaffold protein IsuI in *E. cuniculi* and *T. hominis* (Figure 1.3). Several components were able to functionally complement orthologs in *S. cerevisiae*, with all three components localising to the *E. cuniculi* mitosome. In contrast IsuI and frataxin demonstrated cytosolic localisation in *T. hominis* creating a functional conundrum (Goldberg *et al*, 2008). As microsporidian mitosomes lack a genome, they must import all components destined for the organelle directly from the cytoplasm. It has been shown that some mitosomalbound proteins have retained the classic N-terminal leader sequences, which can be recognised and processed in *S. cerevisiae*.

However, the inner membrane peptidase complex required for N-terminal leader sequence cleavage is greatly simplified in *A. locustae* and is lost altogether in *E. cuniculi* suggesting pre-protein cleavage is not a requirement for mitochondrial protein import (Burri *et al*, 2006). The mitosomal protein import machinery in *E. cuniculi* has also undergone significant reduction. Bioinformatic analysis has indicated that *E. cuniculi* has lost otherwise ubiquitous import components including outer membrane translocases Tom20 and Tom22 and inner membrane translocases such as Tim17 and Tim44 (Figure 1.3). Experimental analysis demonstrated the *E. cuniculi* Tom70 could not rescue growth defects in *S. cerevisiae* knock-outs by complementing function despite localising to the mitochondria (Waller *et al*, 2009).

Like the mitosomes and hydrogenosomes of other parasitic protists such as *T*. *vaginalis*, *E. histolytica* and *G. lamblia*, microsporidian mitosomes have undergone severe reduction losing their genome and many 'core' mitochondrial pathways combined with simplification of the retained biochemical processes. Identification of these organelles implies that mitochondria were likely present in the ancestral eukaryote. Mitosomes appear to be a product of extreme specialisation and the host dependency demonstrated by these parasites. The major functional role of microsporidian mitosomes appears to be iron-sulphur cluster assembly, an essential function in yeast mitochondria. Taken together, these data suggest that iron-sulphur cluster assembly and not respiration is the signature function of mitochondria across eukaryotes.

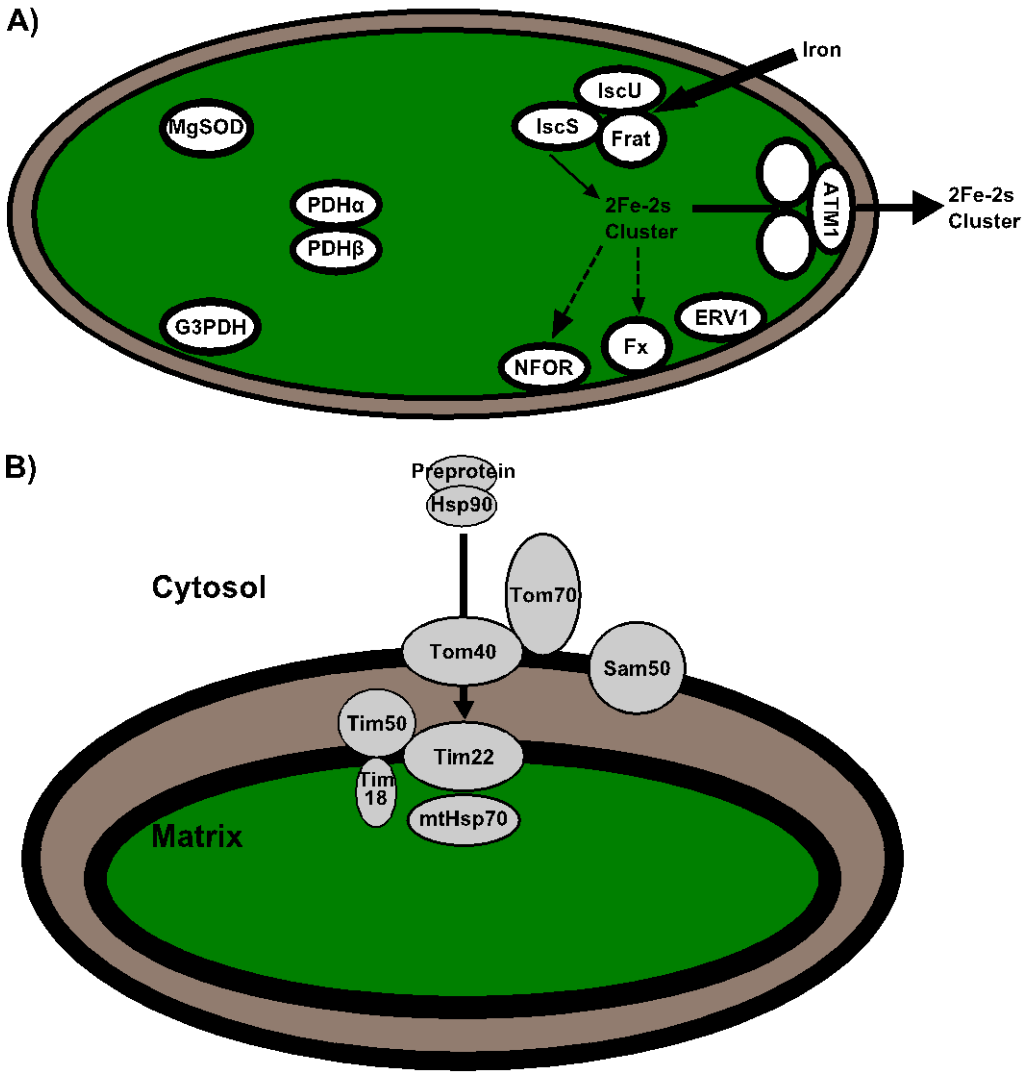


Figure 1.3| Metabolism and import in the microsporidian mitosome.
The microsporidian mitosome has undergone structural simplification, metabolic reduction and has most notably lost its genome. A) Model for metabolism in microsporidian mitosomes. Iron sulphur clusters are formed on scaffold protein IscU by assembly of components frataxin (Frat) and cysteine desulphurase (IscS) prior to transport to recipient protein via ATM1. B) Model of microsporidian mitosomal protein import machinery, which lacks several core components of the Tom and Tim protein complexes. Adapted from Burri *et al*, 2006 and Waller *et al*, 2009.

1.4| Microsporidian genomics: why less is sometimes more

The completion of the first microsporidian sequencing project of *E. cuniculi* revealed a tiny 2.9 megabase (Mb) genome distributed across 11 chromosomes with a coding capacity of approximately 2,000 proteins (Katinka *et al*, 2001). Comparison to other eukaryotes such as the yeast model *S. cerevisiae*, which is itself considered to possess a relatively small genome encoding ~ 6000 genes, demonstrates that *E. cuniculi* has undergone severe genomic reduction (Figure 1.4) (Goffeau *et al*, 1996). The genome of *E. cuniculi* revealed huge host dependence, a result of an obligate intracellular lifestyle, in which the parasite has lost many biochemical pathways or components for core cellular processes. This included TCA cycle, *de novo* protein and nucleotide biosynthesis machinery and the machinery required for RNA interference (RNAi) (RNA pol II, dicer and argonaute) causing speculation that gene knock-out would be the only amenable tool for genetic manipulation in microsporidia (Corradi and Slamovits, 2011, Heinz *et al*, 2012). A range of microsporidian sequencing projects have since been completed, including sampling of species from a diverse host range such as *N. parisii*, a natural pathogen of model nematode *C. elegans*, the honey bee pathogen *N. ceranae*, and *E. bieneusi*, a human-infective species sequenced directly from patient faeces (Akiyoshi *et al*, 2009, Cornman *et al*, 2009, Cuomo *et al*, 2012). This revealed a huge range in microsporidian genome size from 2.3 Mb in the human-infective *E. intestinalis*, one of the smallest known nuclear genomes, to the 24 Mb genome of *Octosporea bayeri* (now *Hamiltosporidium tvaerminnensis*) (Corradi *et al*, 2009, Corradi *et al*, 2010). Interestingly, this drastic variation in microsporidian genome size of up to ~ 10 fold is not translated to an increase in coding capacity and in fact

demonstrates the huge variation in the proportion of non-coding elements present and therefore reflects the level of genome compaction.

Microsporidia have adopted several strategies to compact their genomes. Firstly, they have extremely short intergenic distances which are on average 129 bp in *E. cuniculi*, far shorter than in *S. cerevisiae* which has a genome that is approximately 50% less gene dense and where intergenic spaces have an estimated average of 500 bp (Figure 1.4) (Dujon, 1996). This extremely short intergenic distance is caused by the lack of repetitive regions and the presence of very few selfish elements such as transposons (Keeling, 2009). Microsporidia have also managed a degree of genome compaction within genes themselves. They possess very few introns, in some cases none, and homologous proteins shared between *E. cuniculi* and *S. cerevisiae* are on average 15% shorter in the former (Figure 1.4) (Katinka *et al*, 2001). It has been shown that microsporidia contain overlapping genes in which promoters and termination signals have been relocated beyond initiating transcriptional codons on adjacent genes (Williams *et al*, 2005). Overlapping transcripts however do not provide evidence for a 'eukaryotic operon' as genes do not show a predisposition to be on the same strand. Microsporidian genome stability studies have also identified evolutionary idiosyncrasies. Comparative genomic analyses between species *E. cuniculi*, *A. locustae* and *E. bieneusi* revealed a strong degree of gene synteny even between distantly related species, for example 13% of genes have retained the same genomic location in *E. cuniculi* and *A. locustae* with reference to other genomic markers (Slamovits *et al*, 2004, Corradi *et al*, 2007). This is in stark contrast to the rate of sequence evolution, which is extremely rapid in eukaryotic terms and suggests that, peculiarly,

microsporidia possess rapidly evolving gene sequences within slowly evolving genomes.

Microsporidian genomes vary greatly in size, primarily due to differences in the proportion of non-coding genomic elements. They have lost many important biosynthetic genes and have adopted a life of host dependence in the intracellular environment. Microsporidia are universally reduced in terms of protein coding capacity, hence development of a genetic manipulation system such as gene knock-out or RNAi would provide a minimal model for studying host-parasite interactions in the intracellular environment (Texier *et al*, 2010).

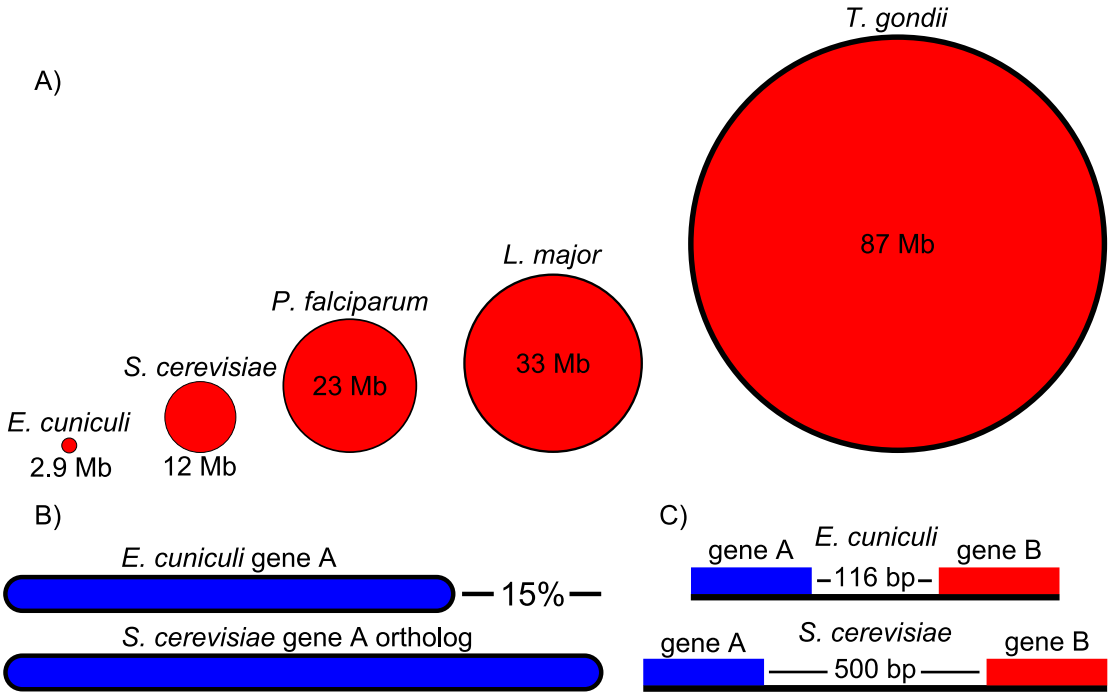


Figure 1.4| Comparison of microsporidian genome size to model and parasitic eukaryotes and examples of microsporidian genome reduction mechanisms.

A) Diagrammatic representation displaying the huge reduction of the *E. cuniculi* genome compared with model yeast *S. cerevisiae* and a range of intracellular parasite species. B) Comparison of *E. cuniculi* genes to *S. cerevisiae* orthologs identifies that *E. cuniculi* has compressed genes by an average of 15%. C) *E. cuniculi* has reduced intergenic distances when compared to *S. cerevisiae* due to a lack of repetitive and selfish elements. Data adapted from Katinka *el al*, 2001

1.5| The impact of horizontal gene transfer on microsporidian evolution

Horizontal gene transfer (HGT) has shaped the evolution of eukaryotic parasites where genes derived from prokaryotic or distantly-related eukaryotic organisms have been acquired, integrated and selected for, thus instantly providing the recipient with a new functional tool (Keeling and Palmer, 2008). Although the extent of HGT in eukaryotic evolution is currently unknown and probably underestimated, laterally acquired genes may have allowed eukaryotes to thrive in new environments and infect new hosts after the acquisition of important metabolic and pathogenicity genes (de Koning *et al*, 2000). In some cases, horizontal transfer of essential metabolic genes from distantly-related species has created a novel target for drug treatment in pathogenic species, such as the bacterial nucleotide biosynthesis genes in the human-infective intracellular apicomplexan parasite *Cryptosporidium parvum* (Striepen *et al*, 2004, Umejiego *et al*, 2008). The transfer of virulence genes or entire pathogenicity islands, among bacteria has been well characterised; more recently, cases of this same process have been identified in eukaryotic pathogens such as the fungal species *Pyrenophora tritici-repentis* where gene transfer of a virulence gene, ToxA, has been directly linked with the ability to cause pathology in the economically important wheat crop, *Triticum aestivum* (Friesen *et al*, 2006). Recent phylogenomic studies have identified vast numbers of eukaryotic HGT events; for example, many genes required for successful oomycete-plant infection may have been obtained from fungi, and of these, a surprising proportion are secreted effectors responsible for attacking plant host cells (Richards *et al*, 2011).

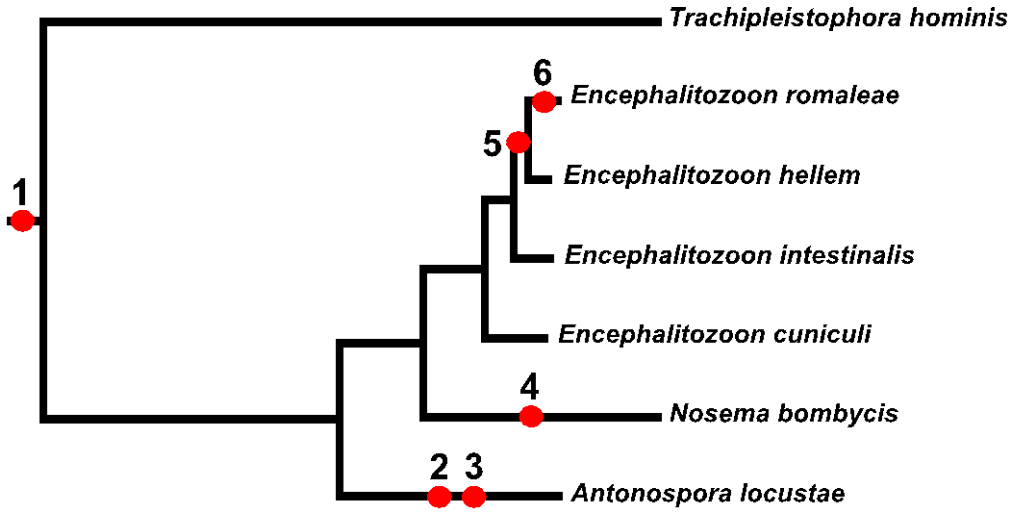
Microsporidia have hugely reduced genomes. Core genes and entire pathways have been lost and in many cases have been replaced by novel mechanisms to aid the acquisition of host metabolites and nutrients. Despite this genome reduction it has been shown that microsporidia have acquired a range of genes from other organisms, instantly providing the potential for a novel interaction with the host. The first example of HGT in microsporidia was apparent upon completion of the genome sequencing project for *E. cuniculi* (Katinka *et al*, 2001). This identified the presence of four 'foreign' adenylate transporters which appear to be of bacterial origin and have most likely been acquired from pathogenic species of *Rickettsia* or *Chlamydia* (Figure 1.5) (Richards *et al*, 2003). Adenylate transporters in these intracellular bacterial genera function in the exchange of bacterial ADP for host ATP as a means to acquire energy (Winkler and Neuhaus, 1999). This gene transfer provides *E. cuniculi*, an organism lacking genes required for the TCA cycle and respiratory electron transport chain, a means of scavenging ATP directly from the host thus ensuring the parasite is not dependent purely on glycolysis for energy production. Microsporidia are thought to lack peroxisomes based on the current absence of ultrastructural evidence and the lack of peroxisomal marker genes such as catalase in their genomes. However it has been identified that *A. locustae*has acquired a proteobacterial catalase into a genomic location with low gene density, thus providing the parasite with a new metabolic tool (Fast *et al*, 2003). This group II catalase, which is expressed in the spore, lacks an identifiable peroxisomal targeting sequence providing no evidence for a cryptic peroxisome. The gene may provide a means for *A. locustae* to cope with stress, a function previously identified for catalase Cta1p found in the fungal pathogen *Candida glabrata* (Cuellar-Cruz *et al*, 2008). The *A. locustae* catalase may

function in the extracellular spore as is seen by HPII in *Escherichia coli*, where it is expressed at a ten-fold increase during stationary phase (Loewen *et al*, 1985). Later it was shown that *A. locustae* had also acquired a photolyase enzyme by horizontal transfer making it the first described class II cyclobutane pyrimidine dimer photolyase in any fungal related organism (Slamovits and Keeling, 2004). Photolyases are a diverse group of repair enzymes which function by reversing UV induced DNA damage by photoreactivation (Britt, 1999). This transfer event therefore potentially provides the *A. locustae* spore with an increased environmental resistance.

Genomic studies of the silkworm pathogen *N. bombycis* identified that microsporidia have also acquired a manganese superoxide dismutase (MnSOD) from a bacterial source (Figure 1.5) (Xiang *et al*, 2010). Superoxide dismutase (SOD) enzymes are generally present in the mitochondria or cytosol and function in the breakdown of toxic superoxide into oxygen and hydrogen peroxide (Strain *et al*, 1998). Superoxide, the most abundant reactive oxygen species, can cause damage to a vast range of cellular macromolecules, hence SOD production is crucial to limit oxidative stress induced cell damage (Landis and Tower, 2005). *N. bombycis* possesses two MnSOD enzymes, suggesting a tandem duplication has occurred since the acquisition of the gene. These paralogs appear to possess different structural properties and targeting signals suggesting they may filldifferent functional roles in the *N. bombycis* stress response. Collectively this suggests that microsporidia have acquired a range of protective agents by HGT from a range of prokaryotic donors and that this may have significantly altered their ability to cope with both environmental and cellular stress.

More recently HGT from a eukaryotic donor to microsporidia has been identified with the acquisition of a purine nucleotide phosphorylase (PNP) and a folylpolyglutamate synthase (FPGS) to the microsporidian species *E. hellem* and *Encephalitozoon romaleae* and *E. romaleae* respectively (Figure 1.5) (Selman and Corradi, 2011, Selman *et al*, 2011). Both eukaryotic transfer events appear to have come from an arthropod species suggesting an ancestral arthropod host for both *Encephalitozoon* sister species and implying that the mammalian infectious *E. hellem* may have undergone a recent host switch or that it is capable of arthropod infection, which has yet to be identified. The genus *Encephalitozoon* lacks the *de novo* biosynthesis machinery for purine and pyrimidine nucleotides suggesting PNP may have a crucial role in salvage of these nucleotides from the host environment as is demonstrated in other parasitic species including apicomplexan *C. parvum*, which possesses a bacterial thymidine kinase (Striepen *et al*, 2004). FPGS functions in retention of folate and has been identified to function throughout the developmental lifecycle of many intracellular parasitic species including *Leishmania infantum* and *Leishmania tarentolae* (Fadili *et al*, 2002). This suggests that FPGS may have an important role in folate salvage during the intracellular development of *E. romaleae* where it is a limited resource.

It is clear that microsporidia have acquired genes to aid extracellular survival, cope with cellular stress and scavenge resources from the host. Future genomic and phylogenetic studies will reveal the true extent to which HGT has shaped the evolution of the microsporidia and the degree to which the host has inadvertently provided the parasite with new molecular tools.



	Gene	Donor	Study
1	ATP Transporters	Bacteria	Katinka *et al*
2	Catalase	Bacteria	Fast *et al*
3	Class II Photolyase	Undetermined	Slamovits and Keeling
4	MgSOD	Bacteria	Xiang *et al*
5	PNP	Arthropod	Selman *et al*
6	FPGS	Arthropod	Selman and Corradi

Figure 1.5| Summary of horizontal gene transfer in microsporidia.
Microsporidia have acquired genes from a range prokaryotic and eukaryotic sources by horizontal transfer. Transfer events are mapped onto an Hsp70 protein phylogeny demonstrating the timing of gene acquisition in relation to microsporidian evolution. Adapted from Selman and Corradi, 2011.

1.6| Microsporidian-host interactions and manipulations

Following germination, microsporidia are exposed to the host cell environment throughout their developmental lifecycle. They have the ability to live freely in the host cytoplasm, invade the host nucleus, replicate within a cytoplasmic PV or develop in a large lipid rich xenoma (Williams, 2009). Microsporidia have been shown to exert a range of complex manipulations on an infected host cell to aid their development, emphasising the host cell dependence suggested by microsporidian genome analysis. Firstly, upon successful colonisation of the intracellular environment, host organelles including the nucleus, mitochondria and the Golgi aggregate around sites of infection (Keeling and Fast, 2002). It has been shown that microsporidia have the ability to invade hostcell nuclei, for example, species of the genus *Enterospora* have been identified in the nuclei of various crustaceans such as the hermit crab *Eupagurus bernhardus* and European edible crab *Cancer pagurus* (Stentiford and Bateman, 2007). The diversity of intranuclear microsporidia and the extent of nuclear infection throughout crustaceans and other animals requires further investigation. The mechanism by which host cell nuclei are attracted to microsporidian infection sites is not known. It has however been identified that microsporidian infection can induce ultrastructural changes in host cell nuclei characterised by an increase in nuclear pore size and apparent absence of the central nuclear pore complex granule (Liu, 1972). The benefit of this change in host nuclear morphology to the parasite remains undetermined, however it has also been shown that microsporidia have the ability to disrupt the host cell cycle to perhaps maximise infective load and aid the spread of infection (Scanlon *et al*, 2000). Multinucleate host cells have been demonstrated as a consequence of *Vittaforma corneae* infection in green monkey kidney cell line E6 (Leitch *et al*,

2005). It has been proposed that this interaction involves manipulation of the host cytoskeleton, particularly of alpha and gamma tubulin.

Like host nuclei, mitochondria are also attracted to microsporidian infection sites within the host cell, as shown in the vacuolar pathogen *E. intestinalis* (Scanlon *et al*, 2004). Interestingly, this is a microtubule independent manipulation and results in no measurable increase in overall host ATP production. This phenomenon has previously been identified in the intracellular apicomplexan parasite *Toxoplasma gondii*. It was originally hypothesised that vacuolar membrane protein ROP2, which possessed a mammalian mitochondrial targeting signal, entered host mitochondria and anchored them to the infected vacuole membrane (Sinai and Joiner, 2001). However, this proposed function for ROP2 has since been discarded as *T. gondii* mutants deficient in the whole ROP2 locus were still capable of mitochondrial attraction, thus re-opening debate into the molecular mechanism responsible for this host cell manipulation (Pernas and Boothroyd, 2010). How microsporidia exert this effect on their host also remains unknown, however the benefit to the parasite is obvious. Microsporidia are only capable of ATP production via glycolysis, giving them limited energy without external means of acquisition (Katinka *et al*, 2001). However microsporidia possess four bacterial ATP transporters usually found in scavenging intracellular bacteria. All of these transporters have the capacity to transport ATP and three specifically localise to the parasite membrane (Tsaousis *et al*, 2008). This infers that microsporidia are capable of scavenging host ATP; hence being in close proximity with host mitochondria would be an obvious advantage.

A range of microsporidian species, generally those infecting fish are able to cause huge enlargement of their host cell, forming a hyper-parasitised spore factory termed a ‘xenoma’ (Lom and Dykova, 2005). These xenomas aggregate in adjacent cells and are capable of forming cysts of up to several centimetres in size, a phenomenon seen in *S. lophii* infection of the monkfish nervous ganglion. It has been proposed that the xenoma offers the parasite protection from the immune system and limits widespread dissemination throughout the host, causing debate as to whether the xenoma is in fact a parasite induced mechanism for immune evasion or a host formed structure to localise infection and prevent spread (Sitja-Bobadilla, 2008).

Microsporidia are transmitted via both horizontal and vertical means, with horizontal transmission usually associated with a high parasite burden and virulence and vertical transmission characterised by low virulence (Dunn and Smith, 2001). Little is known about the mechanism of vertical transmission in microsporidia or why this route, which is generally uncommon among eukaryotic pathogens, is favoured by certain species. *Dictyocoela duebenum* and *Nosema granulosis* infection have been identified in the host cell embryo, where they are associated with the nucleus to ensure equal partitioning of parasite burden during embryonic division (Weedall *et al*, 2006). Microsporidia are capable of optimising vertical transmission by altering the host-sex ratio in favour of females, a manipulation of the host at the population level. This is achieved by two methods. Firstly, microsporidia such as *N. granulosis* are capable of feminisation by altering androgenic gland hormone production in host species *Gammarus duebeni* to optimise for production of a female-heavy population of offspring (Rodgers-Gray *et al*, 2004). Secondly, microsporidia are capable of

male killing in which infection is highly virulent in males but asymptomatic in females, as has been identified in *Amblyospora connecticus* infection in the saltmarsh mosquito *Ochlerotatus cantator* (Andreadis, 2005).

Microsporidia have developed the means to manipulate their host at both the molecular and population level to meet their metabolic energy requirements and maximise their potential to disperse to new hosts. There are many unanswered questions at the molecular level, however, continued study in diverse host-parasite systems will be required to unveil the true intricacy and complexity of microsporidianinduced host cell manipulation.

1.7| The eukaryotic secretory pathway

In eukaryotes, protein synthesis occurs in the cytoplasmic dwelling ribosome. Peptides containing an NH_2 bearingN-terminal signal peptide are identified by the Signal Recognition Particle Complex (SRPC), directed to the Endoplasmic Reticulum (ER) membrane and subsequently into the ER lumen (Walter and Blobel, 1981). The signal peptide in eukaryotic cells acts as an N-terminal 'cellular postcode', directing proteins to the ER membrane. The signal peptide consists of 10 - 30 amino acids and 3 definedregions: 1) N- region – NH_2 terminal sequence containing polar positively charged amino acids 2) H- region consisting of hydrophobic amino acids 3) C- region – containing negatively charged amino acids at -1 and -3 position of the signal peptidase cleavage site which immediately proceeds the mature peptide (Figure 1.6) (Nothwehr and Gordon, 1990). The properties of the eukaryotic signal peptide have been investigated in great detail. Studies mutating the signal peptide sequence of human pre-apolipoprotein A have demonstrated that conservation of amino acid

properties particularly at the N-H region boundary, the H-C region boundary and C region at -1 position of signal peptidase cleavage site are crucial for co-translational cleavage (Nothwehr and Gordon, 1989, Nothwehr and Gordon, 1990). Altering the length of the H-region shows a significant effect on signal peptide cleavage efficiency with an H-region of 10-12 amino acids producing optimum signal peptide cleavage in humans (Nilsson *et al*, 1994). These conserved properties have allowed the development of bioinformatic tools such as SignalP (Petersen *et al*, 2011) and TargetP (Emanuelsson *et al*, 2000) to predict signal peptide presence based specifically on sequence or structural properties, thus allowing secretome prediction in the absence of experimental data. The accuracy of these secretion prediction programs has been assessed and estimated at up to 91% (Klee and Ellis, 2005). However, this analysis was conducted using model eukaryotes including human and mouse sequences and likely represents an over-estimate for most species (Klee and Ellis, 2005).

Peptide entry to the ER occurs via two mechanisms. Firstly, termed co-translational translocation; the signal peptide slows synthesis in the ribosome allowing recognition by SRPC. The ribosome-peptide-SRPC then binds SRP receptors at the ER membrane where the ribosome docks and the signal peptide is cleaved by signal peptidases (Rapoport, 2007). The mature peptide is then transported across the ER membrane to the lumen via the core Sec translocon (αβγ), a heterotrimeric transmembrane channel for protein translocation (Zhang and Miller, 2010). Secondly, post-translational translocation into the ER lumen involves the transport of complete peptides via the Sec62-Sec63 complex. Upon ER entry, translocated peptides bind luminal chaperone BiP which prevents retrograde movement back to the cytoplasm

(Pohlschröder *et al*, 1997). Proteins translocated into the ER are unfolded into polypeptide chains which upon entry bind BiP, which facilitates refolding, an event requiring protein disulfide enzymes in the ER lumen. Proteins arriving in the ER are *N*-glycosylated, via addition of *N*-linked oligosaccharides on asparagine (Asn) residues most commonly at Asn-xxx-serine/threonine sequence motif, thus increasing protein solubility (Zielinska *et al*, 2010). *N*-glycan modifications determine the fate of proteins within the ER whereby the *N*-glycan signature determines whether cargo is sent in vesicles to the Golgi, held intermediately to ensure correct folding or, in the case of misfolded protein, directed to the ER-associated degradation pathway (ERAD) (Wojczyk *et al*, 1998). ERAD is the quality control system of the ER that prevents aggregation of misfolded protein. Upon initiation of ERAD by ERManI, glycoproteins are ubiquinated by the OS-9 associated membrane ubiquitin-ligase complex (Hebert and Molinari, 2012). Misfolded cargo is then retro-translocated to the cytosol in a process involving AAA ATPase P97 and Derlin family proteins, where it is de-glycosylated by Png1 and degraded in the 26S proteasome (Hebert and Molinari, 2007).

Correctly folded *N*-glycoproteins are transported via BiP to integral membrane lectin proteins including vesicular integral membrane protein 36 (VIP36), ER-Golgi intermediate compartment (ERGIC53) and ERGL. VIP36 and ERGIC53 function as cargo receptors processing and facilitating glycoprotein export and interact directly with coat protein complex II (COPII), ERGL is thought to act as a regulator for ERGIC53 (Hauri *et al*, 2002, Nawa *et al*, 2007). Anterograde protein transport from the ER involves integration of cargo into COPII vesicles, a process requiring accurate interpretation of protein sorting

signals and retention of cargo until the coat is fully formed (Barlowe, 2003). COPII is a 5-protein complex consisting of a Sar1 GTPase, which is activated by ER localised Sec12, Sec23, cargo selection component Sec24 and structural proteins Sec13 and Sec31 (Stagg *et al*, 2006). COPII vesicles can vary in size from 50-90 nm depending on the cargo load (Fromme and Schekman, 2005). COPII vesicles are linked together by Golgi tethering protein p115, which is recruited by cis-Golgi membrane GTPase Rab1. This induces vesicle-Golgi fusion in a SNARE (Soluble N-ethylmaleimide sensitive fusion protein attachment protein receptor) dependent manner to transport cargo into the Golgi (Short *et al*, 2005). The Golgi complex represents the central processing site for protein secretion traditionally consisting of stacks of cisternae connected by tubular regions (Mogelsvang *et al*, 2004). Protein cargo and Golgi processing enzymes such as glycosyltransferases migrate through the Golgi in a cis-trans manner after which enzymes are 'recycled' back to the ER via retrograde transport in coat protein complex I (COPI) vesicles (Farquhar and Palade, 1981). COPI recruitment is mediated by the small GTPase Arf1 (Pepperkok *et al*, 2000). Two methods for cis-trans cargo transport through the Golgi have been proposed 1) The vesicular transport model and 2) The cisternal maturation model (Malhotra and Mayor, 2006). The vesicular transport model proposes that COPI vesicles mediate anterograde transport through the Golgi (Orci *et al*, 1986). In contrast the cisternae maturation model suggests that cisternae themselves are non-fixed structures that migrate in a cis-trans direction transporting cargo to the Trans-Golgi Network (TGN) (Pelham and Rothman, 2000). In *S. cerevisiae* three-dimensional confocal microscopy has revealed cisternae are extremely dynamic structures, providing evidence for the cisternal mediated cargo transport mechanism (Matsuura-Tokita *et al*, 2006). Irrespective

of the mechanism for cargo delivery to the TGN, cargo must be sorted and delivered to the extracellular membrane or out of the cell efficiently during this crucial stage of the late secretory pathway. Several trafficking routes have been identified for cargo transport from the TGN including 1) clathrin mediated vesicle transport to the endosomal system 2) protein incorporation and transport via secretory granules and the regulated secretory pathway and 3) vesicle mediated cargo transport via the constitutive secretory pathway (Traub and Kornfeld, 1997). Proteins destined for either regulated or constitutive secretion are bound to membrane receptors in the TGN and packaged into secretory granules. The regulatory secretory pathway involves concentration of cargo at a concentration of up to 200 times that present in Golgi-cisternae followed by cargo export upon receiving a specific cellular stimulus which in turn modifies concentration of intracellular signalling molecules (Burgess and Kelly, 1987). In contrast constitutively secreted cargo is not concentrated and is exported to the plasma membrane at a constant rate (Thiele *et al*, 1997). Secretory granule exocytosis is a highly orchestrated process involving Ca^{2+} import prompting disassembly of the actin cytoskeleton and SNARE reorganisation; followed by vesicle tethering, docking and cargo release; thus delivering proteins to their required functional location (Burgoyne and Morgan, 2003).

The secretory pathway has been studied in great detail, demonstrating a highly complex multi-organelle process. Molecular components of the secretory pathway must interact efficiently to ensure protein secretion to plasma membrane or extracellular environment.

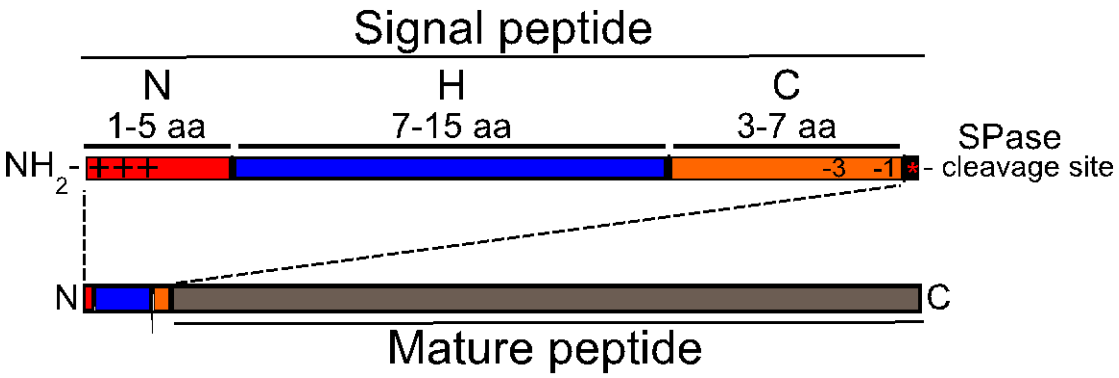


Figure 1.6| Structure and properties of the eukaryotic signal peptide.
The eukaryotic signal peptide has conserved properties that are recognised by the cellular secretory machinery for efficient protein export. The signal peptide contains three distinct subunits 1) N-region - comprising of an NH_2 cap followed by 1-5 positively charged amino acids. 2) H-region - normally 7-15 amino acids which are predominantly hydrophobic 3) C-region - composed of 3-7 amino acids followed by the signal peptidase cleavage site. Amino acids at the -3 and -1 position of the peptidase cleavage site are negatively charged. Upon signal peptide cleavage the mature peptide is directed through the ER-Golgi complex to the required destination. Adapted from Nothwehr and Gordon, 1990.

1.8| Non-classical protein secretion

Non-classical secretion describes proteins which are synthesised on free ribosomes and exported in an ER-Golgi independent manner (Nickel, 2003). Examples of such proteins include the biomedically significant cytokine Interleukin - 1β (IL-1β) and Fibroblast growth factor 1 (FGF1) (Prudovsky *et al*, 2003, Qu *et al*, 2007). Proteins exported by non-classical means lack an identifiable N-terminal targeting sequence and remain secreted upon brefeldin A inhibition of the ER-Golgi transport complex (Orci *et al*, 1991, Amzallag *et al*, 2004). There are 5 mechanisms for non-classical protein secretion currently described, these include: 1) Lysosomal secretion 2) Membrane transporters 3) Release of exosomes 4) Membrane blebbing (Nickel, 2005) and 5) Direct release following cell death (Dubyak, 2012). As with ER-Golgi dependent secretion there are bioinformatic tools for prediction of non-classical protein secretion such as SecretomeP, however this is currently limited to bacterial and mammalian proteins (Bendtsen *et al*, 2004).

Non-classical protein secretion plays a crucial role in delivering parasite virulence factors, both surface proteins and effectors. For example, a proteomic study of secreted proteins through the developmental lifecycle of pathogenic filiarial worm *Brugia malayi* identified approximately 40% of extracellular proteins were predicted to be secreted by non-classical means (Bennuru *et al*, 2009). Comparative proteomics in *Leishmania major* and *Leishmania donovani* has demonstrated that > 52% of the total secretome is exported via exosomes, which have been identified in cytoplasmic components of macrophage host cells (Silverman *et al*, 2010). It is clear that non-classical protein secretion is important for maintaining a range of cellular processes and in host-parasite

interactions, the extent of which will be determined by greater understanding of non-classical secretory mechanisms and the peptide properties at sequence or structural level that induce export in this fashion. Predicting or identifying non-classical secreted proteins in microsporidia holds a range of technical challenges. As bioinformatic tools are not currently applicable, identification of such proteins during intracellular stages would involve ER-Golgi transport disruption using secretion inhibitors such as brefeldin A, which would also severely disrupt host cell biology (Bendtsen *et al*, 2004). It may be possible to investigate non-classical protein secretion during germination of microsporidia using ER-Golgi transport inhibitors, however the role of these proteins in the interaction with the host or the extent to which they contribute to virulence cannot be determined with the genetic tools available currently.

1.9| Secreted proteins as virulence factors in parasitic protists

Colonisation of an intracellular host environment is a lifecycle trait common to a vast phylogenetic diversity of microorganisms and demonstrates convergent adoption of strategies to survive in this niche by many viruses, prokaryotic/eukaryotic pathogens and symbionts (Casadevall, 2008). Parasitic organisms have demonstrated the ability to survive and thrive in the intracellular environment of an enormous range of host cell types including host defence cells such as macrophages, where they develop either directly in the host cytoplasm or within a protective PV (Lingelbach, 2001). Despite drastic lifecycle variation between individual intracellular species, intracellular parasites share some basic common needs such as the ability to enter the host cell efficiently, to resist host defences and to acquire sufficient resources for multiplication and egress (Leiriao *et al*, 2004). In most cases, these complex host-parasite

interactions are poorly characterised particularly at the molecular level. However secretion signal prediction through bioinformatic analysis give a means of identifying a set of proteins that are putatively secreted into the host or sit in the parasite membrane at the interface of parasite and host (Cuervo *et al*, 2009). This 'secreted cellular artillery' is central to host cell manipulation and virulence in many eukaryotic protozoan parasites, where its significance is demonstrated by the evolution of specific secretory organelles such as rhoptries and micronemes in apicomplexan parasites *Plasmodium* and *Toxoplasma* (Sam-Yellowe *et al*, 1996, Kessler *et al*, 2008). Parasite secretome evolution is driven by the host cell environment, and a remarkable variety of host-cell manipulations and virulence mechanisms have been reported.

In *P. falciparum*, the most common causative agent of human malaria, bioinformatic analysis has identified an 11 amino acid signal required for the transport of proteins from the PV to the erythrocyte. This discovery led to the identification of more than 320 secreted proteins, including a range of heat shock proteins. This suggests an extremely complex host interaction despite the PV mediated isolation from the host environment (Hiller *et al*, 2004). This motif is termed the *P. falciparum* export element (PEXEL). Trafficking parasite proteins to the desired destination within the erythrocyte is complex. For example STEVOR proteins, a family of parasite transmembrane proteins require an additional 2 targeting domains including a host cell targeting signal to ensure transport to Maurer's clefts, structures of parasitic origin in the erythrocyte membrane important in antigenic variation (Przyborski *et al*, 2005). Several *P. falciparum* secreted proteins have been identified as virulence factors and in fact have been directly identified to function in inducing

erythrocyte rigidity and adhesion to host blood vessel lining, a phenotype strongly associated with mortality (Maier *et al*, 2008). *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) is an antigenically diverse protein family encoded by approximately 60 var genes and is responsible for mediating adhesion to host cell tissue lining (Baruch *et al*, 1995, Pasternak and Dzikowski, 2009). Knob-associated histidine rich protein (KAHRP) forms knobs on the infected erythrocyte membrane and is responsible for anchoring PfEMP1 to the erythrocyte surface (Crabb *et al*, 1997). PfEMP1 requires 8 genes for successful export to the membrane and in combination with KAHRP is responsible for mediating adhesion to the vascular lining (Maier *et al*, 2008). This cytoadherence together with rosetting, the clustering of non-infected erythrocytes around infected cells, is responsible for immune evasion, blocking the placenta and destroying circulation in organs and hence is a virulence phenotype responsible for up to 1 million deaths per annum (Miller *et al*, 2002).

Leishmania is a genus of intracellular parasites known to infect a wide range of vertebrates including humans, where they reside in macrophages. Several human-infective species of *Leishmania* have been identified as causative agents of human leishmaniasis, with the resultant pathology classified as cutaneous, mucocutaneous and in most severe cases visceral leishmanisis, caused by the species *L. donovani* and *L. infantum* (Berman, 1997). Several extracellularly secreted and membrane proteins have been identified as virulence factors in *Leishmania* infection such as surface metalloprotease Leishmanolysin (GP63). GP63 is encoded by a tandemly repeating homologous gene family consisting of 7 genes in *L. major* and up to 15 in *Leishmania mexicana* (Joshi *et al*, 2002). GP63 provides protection from macrophage

response via complement opsinisation thus providing resistance to complement-mediated killing (Brittingham *et al*, 1995). Interestingly, GP63 has also been shown to disrupt the actin cytoskeleton and tyrosine phosphorylation signalling in fibroblasts, a potential reservoir for the parasite, suggesting GP63 is responsible for exerting a diverse range of effects on host cell function (Halle *et al*, 2009). Secreted cysteine proteases (CPA and CPB) have also been linked to virulence by inhibiting the host Th1 response. In *L. mexicana* CPB deficient mutants could not induce lesion formation in Th1 competent mice but Th1 deficient mice developed high parasite load and lesions (Buxbaum *al*, 2003). Interestingly, *L. mexicana* CPA knockout mutants produced a full disease phenotype in BALB/c mice however CPB mutants displayed low infectivity and slow growth and double mutants were completely avirulent, suggesting an interaction between the two protease enzymes (Alexander *et al*, 1998). Together these studies demonstrate the importance of secreted and membrane proteins in *Leishmania* virulence and immune evasion both of which are crucial for survival in the macrophage and inflicting infection related pathology.

C. parvum is an intracellular apicomplexan parasite capable of producing self-limiting infection in immuno-competent and fatal chronic diarrhoea in immuno-compromised individuals (Barnes *et al*, 1998). A range of membrane bound glycoproteins have been implicated in *C. parvum* virulence by aiding motility, host cell adhesion and invasion including sporozoite ligand (CSL) and GP900 (Sturbaum *et al*, 2008). CSL and GP900 are capable of binding intestinal epithelial cell receptors to aid adherence and permit invasion of the host cell (Cevallos *et al*, 2000, Langer *et al*, 2001). In fellow apicomplexan parasite *T. gondii*, microneme proteins are secreted during invasion. For

example, MIC2 is secreted extracellularly to the host membrane interface and is then translocated to the posterior end of the parasite during invasion of the cell (Carruthers *et al*, 1999). *T. gondii* serine/threonine kinase ROP18 is secreted into the intracellular environment during parasite invasion which when tranfected into non-pathogenic type III *T. gondii* strain increased mortality by up to 5 orders of magnitude in mice (Taylor *et al*, 2006). Over expression of ROP18 also caused a dramatic increase in intravacuolar parasite proliferation in human foreskin fibroblasts emphasising the importance of effectors in *T. gondii* virulence (El Hajj *et al*, 2007). Chagas disease, a major parasitic concern to both public health and the economy in Latin America is caused by intracellular American trypanosome species *Trypanosoma cruzi* (Miles *et al*, 2003). Intracellular infection of *T. cruzi* involves manipulation of host cell cytoskeleton and signalling in which a GP63 homolog is thought to play a critical role (de Avalos *et al*, 2002, Cuevas *et al*, 2003). Several extracellular factors have also been implicated in *T. cruzi* virulence including Tc80, a proteinase which hydrolyses human type I and IV collagen and an acid-active hemolysin with cytotoxic activity *in vitro* (Andrews *et al*, 1990, Santana *et al*, 1997).

The importance of secreted and extracellular membrane proteins as virulence factors is not restricted to parasitic protists but is also of significant importance in intracellular fungi. *Histoplasma capsulatum* is an opportunistic facultative intracellular fungal species and the causative agent of histoplasmosis, a condition characterised by chronic lung infection (Wheat *et al*, 1990). Dwelling in phagolysosomes inside macrophages, a low calcium environment, *H. capsulatum* secretes a calcium binding protein (CBP1) into the intracellular environment suggesting a role in calcium acquisition (Batanghari *et*

al, 1998). CBP1 is not produced in the extracellular lifecycle stage where *H. capsulatum* exists as a saprophytic mould. A CBP1-deficient *H. capsulatum* knockout strain was unable to grow in calcium depleted conditions, destroy macrophages or establish infection in mice, demonstrating the importance of CBP1 in virulence (Sebghati *et al*, 2000). Similarly, *H. capsulatum* siderophore SID1 has been implicated in iron acquisition inside macrophages and is responsible for maintaining optimal growth during infection (Hwang *et al*, 2008).

It is clear that secreted proteins are crucial for establishing intracellular parasite infection, mediating both adherence and invasion in a diverse range of parasitic systems. Once in the intracellular environment, secreted and membrane proteins are required for proliferation, nutrient acquisition and combating the host immune system to maintain the optimum conditions for parasite growth. Identifying secreted proteins and understanding their function at the molecular level is crucial in gaining understanding of the complex host-parasite interactions and parasite induced cellular damage that can be responsible for huge mortality and economic burden worldwide.

1.10| Introduction to the current study

Since the initial discovery of microsporidia by Karl Wilhelm von Nägeli in 1857, over 1300 species have been identified and many aspects of the microsporidian lifecycle, cell biology and genomic properties have been elucidated successfully. Despite this, little is known about microsporidian infection at the molecular level. Infection-induced host cell manipulations have been identified morphologically; however, identifying the parasitic agents responsible has remained challenging. The current lack of a genetic manipulation system such a

gene knock-out or knockdown means molecular characterisation of microsporidian proteins requires heterologous expression in model organisms such as *E. coli* or *S. cerevisiae*.

The obligate nature of microsporidian lifecycle ensures that the secretome, comprising extracellularly secreted proteins and membrane proteins with extracellular domains are at the host-parasite interface throughout the developmental lifecycle, from entry to egress. The aim of this study is to conduct an evolutionary and functional analysis of secreted proteins in microsporidian parasites. This will involve bioinformatic analysis of the secretory pathway genes between diverse microsporidian species to determine the level of reduction in the microsporidian protein export pathway compared to model yeast *S. cerevisiae*. Bioinformatic tools will also be employed to compare the predicted secretome profile of microsporidian species to identify the 'core secretome' and determine patterns of secretome conservation and loss at the species and strain-specific level, where genomic data are available. Due to the retention of the 'core secretome' in highly reduced microsporidian genomes, this represents a list of potentially essential effectors which may be essential for virulence or lifecycle development in a wide range of host species. This bioinformatics analysis will be combined with a proteomic analysis of germination to identify proteins exported to the extracellular environment during this infectious lifecycle stage. This study uses *S. lophii* as a model system, due to the availability of large quantities of spores in combination with *in vitro* germination methods to provide the first experimental identification of potential effectors in microsporidia. In conjunction with this, variation in the whole cell proteome of germinated and non-germinated spores is also investigated. To

determine the function of predicted secreted proteins which may be implicated in virulence in the human-infective microsporidian species *E. cuniculi*; a hemolysin III protein, an endochitinase and a protein tyrosine phosphatase are characterised experimentally using immuno-fluorescence microscopy, heterologous expression systems and virulence assays. Taken together, these research chapters provide novel insights into microsporidian evolution, cell biology and host interaction and introduce new research questions which can be addressed in future work.

This study describes and adds functional understanding to the protein export machinery and secreted proteins in microsporidia, a set of proteins that together are undoubtedly involved in pathogenesis, lifecycle development and host cell manipulation in this obligate intracellular infection system. In this study we will use the term secretome to define the total set of proteins encoded in an organism's genome that are predicted to be exported to the extracellular environment and proteins secreted to the cell membrane with extracellular domains.

Chapter 2| General Methods

2.1| Microsporidian spore purification

Monkfish (*Lophius piscatorius*) xenomas containing *Spraguea lophii* spores (a generous gift from Brixham Trawler Agency) were manually dissected in 1 x PBS using forceps at room temperature. 10 mM EDTA/0.025% Trypsin (Sigma Aldrich) and Triton X-100 (Sigma Aldrich) at a final concentration of 0.05% were added. Xenomas were then incubated at 37^{o}C for 30 minutes. Subsequently, infected xenomas were disrupted manually using a glass homogenizer for 10 minutes. Lysate was transferred to a fresh falcon tube and washed 3 times in sterile 1 x PBS. The spore sample was then carefully decanted onto 5 ml of 100% Percoll (Sigma Aldrich) and centrifuged at 900 x g at 4^{o}C for 10 minutes. If a white spore pellet was not visible post-centrifugation this process was repeated at 1,200 - 1,500 x g. Supernatant containing percoll, monkfish cellular material and bacteria was carefully removed and the clean spore pellet was washed 3 times in sterile 1 x PBS. Spores were stored at 4^{o}C until required in 1 x PBS with addition of an antibiotic cocktail containing 10 µg/ml ampicillin, penicillin/streptomycin and kanamycin.

2.2| DNA extraction

A 200 µl pellet of purified *S. lophii* spores was resuspended in 400 µl Tris-EDTA buffer (TE) 10/1 pH 7.5 and ground in a pestle and mortar in liquid nitrogen for ~15 minutes. Powdered frozen material was transferred to 800 µl phenol (pH 7.9, Sigma Aldrich), mixed by inversion and then centrifuged at 10,000 x g for 10 minutes. 400 µl chloroform (Fisher Scientific) was added to the aqueous supernatant and this was centrifuged at 10,000 x g for 5 minutes. To precipitate DNA, 3 M sodium acetate (1/10 total volume) and 95% ethanol (3 x volume, Fisher Scientific) were added to aqueous supernatant which was stored on ice

for 30 minutes and centrifuged at 12,000 x g and 4^{o}C for 15 minutes. The DNA pellet was then washed in ice cold 70% ethanol. The resultant DNA pellet was air dried and resuspended in the desired volume of nuclease-free water (Fermentas) and quantified using a nanodrop spectrophotometer (Thermo scientific).

2.3| RNA extraction

Prior to beginning RNA extractions, bench space was cleaned thoroughly with ethanol followed by RNaseZAP (Invitrogen). Microsporidian spores were disrupted in liquid nitrogen manually using a pestle and mortar for 15-30 minutes, 1 ml of TRIzol (Invitrogen) was added to frozen cellular material post homogenisation. *E. coli* and *S. cerevisiae* cells were homogenised using the MP Fastprep homogenizer and sterile acid washed glass beads (Sigma Aldrich) in 1 ml of Trizol, 3 times for 20 seconds at 4.1 m/s. Following homogenization the samples were incubated in TRIzol at room temperature for 5 minutes. Subsequently, 0.2 ml of chloroform was added to samples, which were then incubated at room temperature for a further 5 minutes and centrifuged at 12,000 x g at 4^{o}C. The upper aqueous phase containing RNA was carefully transferred to a fresh tube containing 0.5 ml of isopropyl alcohol (Fisher Scientific), mixed and incubated at room temperature for a further 10 minutes. To pellet RNA the samples were then spun at 12,000 x g for 10 minutes at 4^{o}C. The supernatant was carefully discarded. The RNA pellet was then washed in 1 ml of 75% ice cold ethanol and centrifuged at 10,000 x g for 5 minutes at 4^{o}C. The ethanol was removed and the RNA pellet was allowed to air dry for approximately 10 minutes on ice. The RNA was then resuspended in nuclease-free water and quantified using a nanodrop spectrophotometer (Thermo scientific).

2.4| Protein extraction

Microsporidian spores were washed 3 times in sterile 1 x PBS and resuspended in 500 µl protein lysis buffer plus complete ULTRA protease inhibitors (Roche) (per 250 µl of *S. lophii* spores) following the last wash. Spores were then incubated on ice for 10-20 minutes with occasional vortexing. 200-300 µl acid washed glass beads (Sigma Aldrich) were added to the spores which were then homogenised using the MP Fastprep homogenizer, 3 times for 20 seconds at 4.1 m/s, with 5 minute incubations on ice between each Fastprep treatment. The homogenised samples were pelleted by centrifugation for 10-15 minutes at 12,000 x g and 4^{o}C. The resultant supernatant was stored on ice or at -20^{o}C for long-term storage. For protein extraction from mouse inner medullary collecting duct cells (IMCD-3), confluent cells were first trypsinised in TrypleXpress (Life technologies) at 37^{o}C and 5% CO_2 until cells had visibly detached from tissue culture flasks. Detached cells were then pelleted by centrifugation at 2,000 x g and room temperature. Cells were resuspended in RIPA buffer (Pierce Biotechnology) and incubated on ice for 30-45 minutes with gentle agitation by pipetting every 10 minutes. Lysate was then pelleted by centrifugation for 10 minutes at 12,000 x g and 4^{o}C.

2.5| Polymerase chain reaction (PCR) and plasmid ligation

PCR was carried out using Phusion High Fidelity Polymerase (New England Biolabs) for initial amplification of target genes from genomic DNA or PCR master mix (Promega) for verification of insert presence by colony PCR. Reactions were carried out following the manufacturer's guidelines unless otherwise stated with specific details explained in the methods section of subsequent results chapters. Specific template primer (MWG Eurofins)

annealing temperatures were initially trialed at the melting temperature (T_m) - 5^oC and optimised later if required. Amplified DNA was separated by DNA gel electrophoresis on 1% TAE agarose gels containing 0.5 mg/ml ethidium bromide (Sigma Aldrich) and visualised using the Image Master VDS gel imaging system. Resultant DNA amplicons were excised manually under UV light and gel purified using a Qiagen Gel Purification Kit (Qiagen) if required. Purified DNA was quantified using a nanodrop spectrophotometer (Thermo scientific) and ligated into the desired vector using T4 DNA ligase (Promega). Ligation reactions were incubated at room temperature for 3 hours for T-A cloning vectors (pGEM-T Easy, Promega), overnight at 14^oC for restriction digest cloning vectors (pYES2, Invitrogen and pET-14b, Novagen) and 30 minutes at room temperature for TOPO cloning reactions (pcDNA3.1/V5-6xHis, Invitrogen)

2.6| Bacterial transformation

2 µl of purified plasmid DNA was incubated with 50 µl of freshly thawed competent *Escherichia coli* cells. *E. coli* XL1-Blue strain (Stratagene) and *E. coli* C3010I strain (New England Biolabs) were used for plasmid propagation and recombinant protein expression respectively. Competent *E. coli* cells were then heat shocked at 42^oC for 20-45 seconds and immediately returned to ice for 2 minutes. 950 µl of pre-warmed Luria-Bertani (LB) media was added to cells which were then incubated at 37^oC for 45 minutes while shaking at 150-220 rpm. Cells were then plated on LB agar containing ampicillin (100 µg/ml) and incubated overnight (16-20 hours) at 37^oC.

2.7| Plasmid purification and restriction digest

Single bacterial colonies growing on LB-ampicillin were picked using sterile pipette tips and grown overnight (16-20 hours) in 4 ml LB broth containing ampicillin (100 µg/ml) in a 37^{o}C shaking incubator (180-220 rpm). The resultant bacterial culture was pelleted by centrifugation at 12,000 x g at room temperature. Plasmid DNA was extracted from a bacterial pellet using a Spin Column Plasmid DNA Miniprep Kit (NBS Biologicals). To verify the presence of a DNA insert or linearised plasmid for cloning, digests were incubated for 3 hours – overnight at 37^{o}C with the appropriate restriction enzymes (New England Biolabs). Insert dropout or plasmid linearisation was visualised by electrophoresis on 1% TAE agarose gels.

2.8| Fungal transformation

5 ml of Yeast Peptone Dextrose (YPD) was inoculated with a single colony of *Saccharomyces cerevisiae* wild type strain BY4741a (Euroscarf- Genotype MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0- unless otherwise stated) and incubated overnight in a 30^{o}C shaking incubator at 180 rpm. Details of the *S. cerevisiae* knock-out strains used are described in the methods section of the appropriate results chapter. The culture was then diluted to an optical density (OD) of 0.4 at 600 nm in ~50 ml YPD and incubated at 30^{o}C for an additional 2 hours. Cells were centrifuged at 2,000 x g for 2 minutes and resuspended in 40 ml 1 x TE. Cells were then pelleted at 2,000 x g for 2 minutes, resuspended in 1 x lithium acetate (LiAc, Sigma Aldrich)/0.5 x TE and incubated at room temperature for 10 minutes. 100 µl of *S. cerevisiae*/LiAc/TE suspension was aliquoted to a pre-mixed solution of 1 µg plasmid DNA and 100 µg of denatured herring sperm DNA (Sigma Aldrich). 700 µl of 1x LiAc/1x TE/40% polyethelene

glycol (PEG)-3350 (Sigma Aldrich) was added to the *S. cerevisiae*-DNA suspension which was then incubated at 30^{o}C for 30 minutes. 88 µl of dimethylsulfoxide (DMSO, Fisher Scientific) was added to the sample which was subsequently heat shocked at 42^{o}C for 7 minutes. The sample was then pelleted in a microcentrifuge at 12,000 x g for 10 seconds and resuspended in 1 ml of 1 x TE. Cells were re-pelleted, suspended in 100 µl TE and plated onto complete synthetic media (CSM) minus the appropriate amino acid (uracil/leucine) for selection.

2.9| Tissue culture

Mouse inner medullary collecting duct cell line (IMCD-3) and rabbit kidney epithelial cell line (RK-13) (generously provided by Dr Helen Dawe and Professor Martin Embley respectively) were grown in Dulbecco's modified Eagle's medium and Ham's F12 with glutamax (DMEM/F-12, Life technologies) and 10% fetal bovine serum (FBS, Life technologies) at 37^{o}C plus 5% CO_2. To passage cells once confluent, cells were rinsed with 1 x PBS and incubated at 37^{o}C/5% CO_2 in TrypleXpress (Life technologies) until the cell monolayer had visibly detached from the flask (20-30 minutes). To deactivate TrypleXpress, pre-warmed DMEM/F-12 was added to cells at a ratio of 4 ml DMEM/F-12: 1 ml trypsin. Cells were then pelleted by centrifugation at 1,000 x g for 2 minutes. Cells were resuspended in 3-5 ml DMEM/F-12/10% FBS and split at a ratio of between 1:4-1:8 into fresh tissue culture flasks (Thermo scientific) containing the desired volume of DMEM/F-12/10% FBS medium. If back up cell stocks were required, cells were flash frozen and stored in liquid nitrogen in cryopreservation media (DMEM/F-12/10% FBS + 10% DMSO).

2.10| Transient transfection of mammalian cells

IMCD-3 cells were grown in 6-well plates (Thermo scientific) in 2 ml OptiMEM (Life technologies) with 10% FBS medium until 70-90% confluent. For each well 2-4 µg of plasmid DNA and 8 µl of Lipofectamine 2000 (Invitrogen) were added to 200 µl of OptiMEM without serum respectively and incubated for 5 minutes at room temperature. DNA/OptiMEM was then added to Lipofectamine 2000/OptiMEM and incubated for a further 20 minutes at room temperature. The DNA-Lipofectamine complex (400 µl) was added to the IMCD-3 cell monolayer and mixed gently. Cells were then incubated at 37^{o}C with 5% CO_2 for 48 hours.

2.11| Recombinant protein expression and purification

Amplified DNA sequences encoding the desired protein were ligated into N-terminal His-Tag fusion expression vector pET-14b (Novagen). The construct was then sequence verified (MWG Eurofins). A 10 ml LB media starter culture was inoculated with *E. coli* C3010I containing pET14b construct and grown overnight at 37^{o}C in a 180 rpm shaking incubator. Subsequently, 1 L of LB media was inoculated with the starter culture and grown to an optical density at 600 nm of 0.4-0.6 at between 18-37^{o}C. Expression via the T7 promoter was induced with isopropylthio-β-galactoside (IPTG, Thermo Scientific) at a final concentration of 0.5 mM for 4 hours at 37^{o}C or overnight at 18^{o}C. Cell culture was then pelleted at 6,000 x g for 20 minutes at 4^{o}C. Cells were then resuspended in 50 ml 1 x PBS and lysed using a probe sonicator with 6 x 10 second treatments. Cells were cooled on ice for 20 seconds between each sonicator treatment. Lysed cells were pelleted at 6,000 x g for 20 minutes at 4^{o}C following the addition of complete ULTRA protease inhibitors (Roche). The

supernatant containing soluble protein was stored on ice. The cell pellet containing insoluble protein and inclusion bodies was resuspended in 50 ml 8 M urea and incubated at room temperature for 30 minutes with gentle agitation. The insoluble protein fraction was then centrifuged at 6,000 x g at 4^{o}C and the resultant supernatant was retained on ice. To purify His-tagged protein via affinity chromatography, soluble and insoluble crude protein extracts were passed over Ni^{2+} NTA agarose beads (5 Prime). The Ni^{2+} NTA agarose was then treated with an imidazole (Sigma Aldrich) gradient consisting of 5 x 5 ml washes with 10 mM, 30 mM, 50 mM, 200 mM and 500 mM imidazole in 1 x PBS and 8 M Urea for soluble and insoluble fractions respectively. Imidazole fractions were then checked for the presence of His-tagged target protein using both sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot.

2.12| SDS-PAGE and western blot

Protein samples of standardised concentrations were run by SDS-PAGE on 12% gels (see 2.16 for recipe) alongside a pre-stained protein marker (7-175 kDa, New England Biolabs). If protein was to be visualised directly, SDS-PAGE gels were stained overnight with Coomassie blue and subsequently destained using Destain I and Destain II as required. For western blot analysis, the gel was rinsed briefly in Towbin transfer buffer and sandwiched against a nitrocellulose transfer membrane (Bio-Rad) with filter paper (3 sheets per side). The gel-membrane sandwich was placed in a gel tank and immersed in ice cold Towbin transfer buffer. Protein transfer was run at 25 V for 1-2 hours. The gel-membrane sandwich was removed from the tank to verify transfer as indicated by the presence of pre-stained ladder on nitrocellulose membrane. The

membrane was subsequently washed briefly in Tris-buffered saline + 0.1% Tween 20 (Sigma Aldrich) (TBST) to remove remnants of SDS. The membrane was then blocked in 5% milk-TBST (Blotting grade non fat dry milk, Bio-Rad) with gentle agitation for 1 hour at room temperature. Following blocking, the membrane was rinsed in 0.5% milk-TBST for 30 minutes. The blot was then treated with primary antibody or serum at the pre-determined optimum or recommended concentration (1:100-1:2000) for 1 hour at room temperature or 4^{o}C overnight. The membrane was then washed 3 times in 1% milk-TBST for 10 minutes. The rinsed blot was incubated with secondary antibody labeled with horseradish peroxidase (1:1000-1:5000 concentration) at room temperature with agitation for 1-1.5 hours followed by 3 x 10 minute washes in TBST. The blot was then incubated with 5 ml ECL western blotting detection reagent (NBS biologicals) for 2 minutes in a dark room followed by incubation with x-ray film (Fujifilm) in a cassette. Blots were then developed using a Protec Optimax developer.

2.13| Immuno-localisation

Encephalitozoon cuniculi infected IMCD-3 and RK-13 cells were seeded onto sterile glass cover slips in 6-well plates and grown until confluent (or infection had reached the desired intensity) in DMEM/F-12/10% FBS at 37^{o}C and 5% CO_2. Once cells had become confluent, medium was removed and cells were washed twice in 2 ml PBS. Two methods for cell fixation were tested. Firstly, cells were incubated with 4% paraformaldehyde (Fisher Scientific) for 15 minutes at room temperature and subsequently washed with 3 x 5 minute 1 x PBS washes. The cells were then permeabilised with incubation in 1 x PBS/0.1% Triton X-100 (Sigma Aldrich) for 30 minutes. Secondly, cells were

incubated in acetone:methanol (1:1) (Fisher Scientific) for 2 hours at -20^{o}C. Following permeabilisation, cells were washed for 5 minutes, 3 times in 1 x PBS. Fixed cells were blocked for 15 minutes with 5% non-fat dry milk (Biorad) in PBS with gentle agitation. Cells were then rinsed in 0.5% milk/PBS solution for 30 minutes. Blocked cells were treated with 1% milk/PBS along with primary antibody or raised anti-sera at the optimum concentration for 1.5 hours (specific antibody/anti-sera details and concentrations are described in subsequent results chapters). Following three 1 x PBS washes, cells were incubated with fluorescently labeled secondary antibody in 1% PBS/milk at a 1:1000 dilution for 1-1.5 hours with gentle agitation at room temperature. If required 4', 6-diamidino-2-phenylindole (DAPI, Invitrogen) stain was applied to cells at a concentration of 1 µg/ml for the final 15 minutes of secondary antibody treatment. After three 1 x PBS washes, coverslips were inverted, mounted in Vectashield fluorescent mounting media (Vector Labs) and sealed with colourless nail varnish. Fluorescently labeled cells were then visualised with excitation and capture at the recommended wavelengths using Olympus IX-81 inverted fluorescent microscope.

2.14| *In vitro* spore germination

300 µl purified spores were washed 3 times in 1 x PBS. A non-germinated control sample was then treated with 100 µl of 0.25 mM EGTA per 100 µl of spores to inhibit germination. The sample to be germinated was treated with 100 µl of 0.5 M Gly-Gly buffer (Sigma Aldrich) (pH 7.0) per 100 µl of spores and incubated at room temperature for 30 minutes. 40 µl of calcium ionophore A23187 (0.001 µg/ml) (Sigma Aldrich) dissolved in DMSO was added to the sample followed immediately by 40 µl of 0.5 M Gly-Gly buffer (pH 9.0) per 100

µl of spores. Germination was verified instantly by light microscopy and efficiency was estimated to reach a maximum of 80%.

Chapter 3| Dissection of the microsporidian secretory pathway: a minimal and unconventional model for protein export.

3.1| Introduction

The eukaryotic secretory pathway involves the transport of proteins from the ribosome, through the Endoplasmic Reticulum (ER) and Golgi from which they are either released extracellularly or are incorporated into the plasma membrane (Rothman and Orci, 1992). This complex process is initiated by synthesis and recognition of an N-terminal signal peptide, which acts as a 'cellular postcode', directing cargo into the ER lumen (Heijne, 1990). Proteins which function in the ER and require retention in the organelle are characterised by the presence of a C-terminal HDEL/KDEL amino acid motif (HDEL in *Saccharomyces cerevisiae*, KDEL in animals) which is recognised by ER sorting receptors such as ERD2 in *S. cerevisiae* (Lewis and Pelham, 1990, Aoe *et al*, 1997). Successful secretion of protein cargo requires a complex multi-organelle gene network required to re-fold, glycosylate and package proteins correctly, to allow efficient transport across several defined biological membranes. Misfolded or incorrectly packaged cargo is otherwise exported for proteolysis in the cytoplasmic proteasome to prevent the aggregation of non-functional protein (Vashist and Ng, 2004). Secreted proteins are essential for maintaining the structural and functional integrity of the plasma membrane and for delivering effectors or signalling proteins to the extracellular environment, therefore efficient targeting and transport is a requirement for the functioning cell (Chaffin *et al*, 1998).

The Microsporidia, are highly reduced amongst eukaryotes in terms of genome size (Katinka *et* al, 2001, Keeling *et al*, 2005). Across the phylum they have lost various genes and complete biochemical pathways which have become redundant for their particular lifestyle in the intracellular environment.

This includes processes originally deemed ‘core’ or indispensible to the functioning of the cell, such as core carbon metabolism in the species *Enterocytozoon bieneusi*, thus demonstrating an extreme degree of specialisation (Keeling *et al*, 2010). Microsporidian peculiarities are not restricted to their genomic content or architecture; they are also evident in their cellular composition and organelles. Microsporidia harbour a unique infection apparatus, the polar tube; contain reduced mitochondria (mitosomes), lack peroxisomes and possess ‘prokaryote like’ 70S ribosomes (Vossbrinck and Woese, 1986, Cavalier-Smith, 1987, Vossbrinck *et al*, 1987).

The microsporidian Golgi, one of the key secretory pathway components, also displays unusual properties. It consists of a tubular network connected directly to the ER, plasma membrane and polar tube, with standard histochemical features, however, it apparently lacks vesicles generated by coat protein complexes (COPI and COPII) (Beznoussenko *et al*, 2007). These vesicles are generally required for anterograde cargo transport from the ER to the Golgi and the retrograde movement of ER or Golgi-localised proteins (Lee and Goldberg, 2010). This raises interesting questions as to how microsporidia transport proteins such as spore wall proteins to the plasma membrane, effectors into the intracellular host environment, and ‘recycle’ functional Golgi proteins in the absence of COPI and COPII vesicles. Microsporidia were initially hypothesised to lack one of the core components of the Sec61αβγ translocon responsible for the transport of protein cargo from the cytoplasm into the ER luminal space and from the ER back to the cytoplasm prior to proteolysis, Sec61β. Pore forming Sec components, Sec61α and Sec61γ, are essential for cell viability; however, Sec61β has been shown to be dispensable in *S.*

cerevisiae (Kelkar and Dobberstein, 2009). A divergent Sec61β ortholog has however since been identified in microsporidia which is heterologously targeted to the ER in *S. cerevisiae* and displays a strong degree of synteny between microsporidian species (Slamovits *et al*, 2006, Wu *et al*, 2007). This demonstrates the inability of methods such as Pfam and BLAST searching to identify all components of microsporidian pathways and leaves open the possibility that microsporidian pathways are not as simple as initially thought.

The objective of this work was to investigate the genetic composition of the microsporidian secretory pathway, from initial Sec translocon mediated peptide entry into the ER, through the Golgi, to the cell membrane or extracellular environment. The aim is to determine the level and signature of genetic reduction in the microsporidian secretory pathway and identify the genetic hallmark of the unusual microsporidian Golgi.

3.2| Methods

3.2.1| KEGG pathway analysis and database mining

To investigate patterns of retention and loss of secretory pathway genes in microsporidian genomes, *S. cerevisiae* proteins from the following molecular interaction pathways: 1) Protein export 2) Protein processing in the ER 3) SNARE interactions in vesicular transport and 4) Proteasome mediated proteolysis were downloaded from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) PATHWAY database (Kanehisa *et al*, 2004). *S. cerevisiae* reference proteins, including paralogs, were used to search for orthologs in the genomes of the following microsporidian species: *Encephalitozoon cuniculi*, *Encephalitozoon intestinalis*, *Vavraia culicus floridensis*, *Vittaforma corneae*, *Nematocida* sp. *Nematocida parisii* and *Edhazardia aedis* using the Broad Institute Microsporidia Comparative Database and the National Centre for Biotechnology Information (NCBI) protein database with a pre-defined e-value cut-off of $1e^{-10}$. NCBI searches were conducted against the non-redundant protein sequence database (nr database). This includes protein sequences from Non Redundant GenBank coding region translations (CDS), the Protein Data Bank (PBD), SwissProt and the Protein information Resource. Searches were conducted using standard scoring parameters including Blocks Substitution Matrix (BLOSUM) 62 and a conditional compositional score matrix adjustment. Broad Institute Microsporidia Comparative Database searches also included a BLOSUM 62 substitution matrix.

3.3| Results

3.3.1| Protein secretion machinery in the microsporidia

Following protein synthesis on the ribosome, the initial stage of the secretory pathway involves recognition of an NH_2 labelled N-terminal signal peptide by the Signal Recognition Particle Complex (SRPC) (Walter and Blobel, 1981). The SRPC in *S. cerevisiae* consists of 7 proteins, 5 cytoplasmic signal recognition particles (SRP72, SRP19, SRP14, SRP68 and SRP54) and 2 signal recognition particle receptors (SRPR and SRPRB) located in the ER membrane. These proteins must function coherently to transport proteins from the cytosol to the ER (Hann and Walter, 1991). The SRPC in microsporidia is extremely reduced in which SRP72, SRP14, SRP68 and receptor component SRPRB have been lost universally in the 7 species examined (Table 3.1). Of these components SRP72, SRP14 and SRP68 are non-essential in *S. cerevisiae*, however, yeast strains deficient in SRPRB, a gene required for anchoring SRPR to the ER membrane, are non-viable (Brown *et al*, 1994, Vandenbol and Fairhead, 2000). This suggests microsporidia have lost non-essential components of the SRPC and have perhaps adopted a novel means of SRPR docking to the ER membrane.

In co-translational translocation, protein cargo docked at the ER membrane is transported across the membrane into the ER lumen where the signal peptide is cleaved by signal peptidases (Johnson and van Waes, 1999). Microsporidia possess all components of the core Sec αβγ translocon, including a divergent Sec61β ortholog (Table 3.1) (Slamovits *et al*, 2006). In contrast, microsporidia possess an extremely reduced complement of signal peptidases where by enzymes SPCS1, SPCS2 and SPCS3 have been lost and only Sec11

has been retained (Table 3.1). Interestingly, SPCS3 appears to be essential in *S. cerevisiae* suggesting that microsporidia possess a novel signal peptidase cleavage mechanism or that they possess orthologs of SPCS3 and other signal peptidases but these genes are too divergent for identification by BLAST search analysis (Giaever *et al*, 2002). Sec11, the catalytic subunit of the signal peptidase complex is also essential in yeast (Giaever *et al*, 2002). Microsporidia possess the core components of the post-translational translocon, Sec62/63 and ER luminal chaperone BiP, suggesting this pathway is well conserved with 'higher eukaryotes'. Both mitochondrial inner membrane peptidases (IMP1 and IMP2) required for mitochondrial targeting signal cleavage have been lost in the microsporidia examined in this study, suggesting these proteins are functionally redundant in the microsporidian mitosome (Table 3.1). An IMP2 ortholog is present in the genome of the microsporidian species *Antonospora locustae* (GI: 63029087), however the functional implications of this remain unknown.

Upon arrival in the ER protein cargo must be glycosylated and correctly folded to ensure efficient transport to the Golgi. Microsporidia have lost many core components of the ER protein processing machinery such as *N*-glycan biosynthesis proteins GlcI and GlcII (Schallus *et al*, 2008), UGGT, which acts as an ER quality control service sorting incorrectly folded cargo (Dejgaard *et al*, 2004) and glycoprotein cargo receptor ERGIC53 (Table 3.1) (Hauri *et al*, 2000). This suggests that microsporidia have either simplified or highly adapted their cellular system for protein-processing and transport through the ER. Microsporidia possess terminal ER glycoprotein receptor VIP36 which facilitates protein export via direct interaction with COPII (Nawa *et al*, 2007). COPII components Sec13, Sec31, Sec23, Sec24 and GTPase Sar1 required for

anterograde cargo transport to the Golgi are predominantly present across microsporidia with some species level omissions (Table 3.1). Interestingly, structural COPII component Sec13 is absent from the genomes of both *Nematocida* species examined in this study. Future sequencing projects will determine if this core gene has been lost from the *Nematocida* genus (Table 3.1).

Similarly, the ER to Golgi transport machinery and genes involved in vesicle formation and cis-trans transport of cargo through the Golgi are also highly reduced in microsporidia (Table 3.2). Only Ykt6 and Stx1 are present ubiquitously in the microsporidian species studied, and Sec22 like Sec13 is absent from the genus *Nematocida* (Table 3.2). All other components of the cellular machinery required for protein transport through the Golgi are absent in microsporidia with the exception of Bos1, a SNARE protein required for vesicle transport from the ER to the Golgi in *S. cerevisiae*, which is present in *Encephalitozoon* spp. (Shim *et al*, 1991) and Stx16, a target SNARE (t-SNARE) required for vesicle docking which is present in *E. aedis* and *V. culicus floridensis* (Table 3.2) (Simonsen *et al*, 1998). Sec22 binds COPII at the Sec23/24 interface upon ER exit and in turn mediates vesicle transport to the Golgi (Mancias and Goldberg, 2007). Similarly, universal microsporidian secretory component Ykt6 and *Encephalitozoon* gene Bos1 also function in vesicle transport from the ER and in vesicle fusion with the Golgi membrane (Newman *et al*, 1990). All components required for vesicle transport through the Golgi are absent from microsporidian genomes in what may be a genetic reflection of the tubular Golgi (Beznoussenko *et al*, 2007). Stx1, a gene universally present in all microsporidia examined appears to be the only SNARE

protein capable of plasma membrane fusion and protein export present in microsporidian genomes (Han *et al*, 2004).

Table 3.1| Protein translocation into the ER in microsporidia.

	E. cuniculi	*E. intestinalis*	*V. corneae*	*E. aedis*	*V. culicus*	*Nematocida sp.*	*N. parisii*
Sec translocon							
Sec61α	●	●	●	●	●	●	●
Sec61β	○	○	○	○	○	○	○
Sec61γ	●	●	●	●	●	●	●
Sec62	●	●	●	●	●	●	●
Sec63	●	●	●	●	●	●	●
BiP	●	●	●	●	●	●	●
Signal recognition particles							
SRP72	○	○	○	○	○	○	○
SRP19	●	●	●	●	●	●	●
SRP14	○	○	○	○	○	○	○
SRP68	○	○	○	○	○	○	○
SRP54	●	●	●	●	●	●	●
SRPR	●	●	●	●	●	●	●
SRPRB	○	○	○	○	○	○	○
Signal peptidases							
SPCS1	○	○	○	○	○	○	○
SPCS2	○	○	○	○	○	○	○
SPCS3	○	○	○	○	○	○	○
SEC11	●	●	●	●	●	●	●
IMP1	○	○	○	○	○	○	○
IMP2	○	○	○	○	○	○	○
Protein processing in ER							
Sec13	●	●	●	○	●	○	○
Sec31	●	●	○	○	●	●	●
Sec23	●	●	●	●	●	●	●
Sec24	●	●	●	○	●	●	●
SAR1	●	●	●	●	●	●	●
ERMan1	○	○	○	○	○	○	○
CNX	○	○	○	○	○	○	○
GlcI	○	○	○	○	○	○	○
GlcII	○	○	○	○	○	○	○
Osts	○	○	○	○	○	○	○
UGGT	○	○	○	○	○	○	○
NEF	●	●	●	●	●	●	●
Hsp40	●	●	●	●	●	●	●

● = ortholog present ○ = ortholog absent

○ = ortholog identified in Slamovits *et al* 2006

Table 3.2| Protein cargo transport through the ER, Golgi and vesicle formation in microsporidia.

	E. cuniculi	*E. intestinalis*	*V. corneae*	*E. aedis*	*V. culicus*	*Nematocida sp.*	*N. parisii*
Vesicular transport							
Sec22	●	●	●	●	●	○	○
Sec20	○	○	○	○	○	○	○
Use1	○	○	○	○	○	○	○
Stx18	○	○	○	○	○	○	○
Bet1	○	○	○	○	○	○	○
Bos1	●	●	○	○	○	○	○
Stx5	○	○	○	○	○	○	○
Gos1	○	○	○	○	○	○	○
Ykt6	●	●	●	●	●	●	●
STX6	○	○	○	○	○	○	○
Stx16	○	○	○	●	●	○	○
Sft1	○	○	○	○	○	○	○
VAMP4	○	○	○	○	○	○	○
Stx1	●	●	●	●	●	●	●
SNAP23	○	○	○	○	○	○	○
VtiI	○	○	○	○	○	○	○

● = ortholog present ○ = ortholog absent

3.3.2| Retrograde protein transport and the fate of misfolded proteins

To limit the aggregation of non-functional protein in the ER, incorrectly folded cargo must be identified, ubiquinated, de-glycosylated and transported back to the cytoplasm, where it is degraded (Plemper and Wolf, 1999). The process involves the collaboration between two cellular pathways; 1) the ER-associated degradation pathway (ERAD) required for ubiquitination and retrograde transport of misfolded proteins, a process which involves the Sec61αβγ translocon and 2) the proteasome which is required for protein degradation (Kalies *et al*, 2005). Genome comparison with the *S. cerevisiae* ERAD pathway, which has been well described, reveals that microsporidia have retained the major components of what is a reasonably well-conserved ERAD pathway (Table 3.3). The microsporidia examined in this study have retained both the protein disulphide isomerases and derlin required for unfolding/transporting protein to the ER membrane and for cargo recognition at the ER membrane respectively (Kostova and Wolf, 2003, Schaheen *et al*, 2009). Microsporidia have lost ER-lectin OS-9 required for the recognition of asparagine-linked glycans on misfolded proteins (Su *et al*, 2012). Interestingly, export protein Bap31 has only been retained in the Encephalitozoonidae (Table 3.3). Following ubiquitination and export to the cytoplasm via the Sec61αβγ translocon, misfolded cargo is bound to cytoplasmic AAA-ATPase and chaperone p97 which inhibits bidirectional transport back into the ER via Sec61αβγ (Rabinovich *et al*, 2002). The p97 chaperone is present across the microsporidia. Subsequently, DOA1 interacts with p97 to bind cargo via the ubiquitin binding domain and interacts with DSK2 to present misfolded protein cargo to the 26S proteome (Mullally *et al*, 2006, Zhang *et al*, 2009). Both these ERAD components have been retained across the phylum (Table 3.3). Ufd2,

which was also identified in all microsporidia, functions in controlling AAA-ATPase p97 activity (Bohm *et al*, 2011).

Ubiquitin-tagged protein is then targeted to the 26S proteasome, consisting of a 20S core and a respective 19S lid and base, where, they are degraded into small peptides which are eventually recycled for subsequent translation (Glickman and Ciechanover, 2002). Comparison with *S. cerevisiae* reveals that microsporidia have retained a remarkably conserved 26S proteasome (Table 3.4). All microsporidia examined in this study have retained a complete 20S core proteasome suggesting these core subunits have an essential function across the phylum (Table 3.4). Microsporidia have retained many components of the 19S regulatory proteasome base and lid, though Rpn 8, 12, 13, 15 and nuclear proteasome activator PA200 are universally absent, the functional implications of which are not obvious (Table 3.4).

Table 3.3| Retrograde protein transport and ER to cytosol export of misfolded proteins.

	E. cuniculi	*E. intestinalis*	*V. corneae*	*E. aedis*	*V. culicus*	*Nematocida sp.*	*N. parisii*
ERAD pathway							
PDIs	●	●	●	●	●	●	●
OS-9	○	○	○	○	○	○	○
Bap31	●	●	○	○	○	○	○
Derlin	●	●	●	●	●	●	●
Ubx	○	○	○	○	○	○	○
p97 (cdc48)	●	●	●	●	●	●	●
NEF	●	●	●	●	●	●	●
sHSF	○	○	○	○	○	○	○
Np14	●	●	●	●	●	●	●
Ufd1	○	○	○	○	○	○	○
Otu1	○	○	○	○	○	●	●
DOA1	●	●	●	●	●	●	●
DSK2	●	●	●	●	●	●	●
Ufd2	●	●	●	●	●	●	●
RAD23	○	○	●	●	●	●	●
Png1	○	○	○	○	○	○	○

● = ortholog present ○ = ortholog absent

Table 3.4| The microsporidian proteasome.

	E. cuniculi	*E. intestinalis*	*V. corneae*	*E. aedis*	*V. culicus*	*Nematocida sp.*	*N. parisii*
19S - Lid							
Rpn3	●	●	●	●	○	●	●
Rpn5	●	●	●	●	●	●	●
Rpn6	●	●	●	●	●	●	●
Rpn7	●	●	●	●	●	●	●
Rpn8	○	○	○	○	○	○	○
Rpn9	●	●	○	●	●	○	○
Rpn11	●	●	●	●	●	●	●
Rpn12	○	○	○	○	○	○	○
Rpn15	○	○	○	○	○	○	○
Rpn10	●	●	○	●	●	○	●
19S - Base							
Rpn1	●	●	●	●	●	●	●
Rpn2	●	●	●	●	●	●	●
Rpn13	○	○	○	○	○	○	○
Rpt1	●	●	●	●	●	●	●
Rpt2	●	●	●	●	●	●	●
Rpt3	●	●	●	●	●	●	●
Rpt4	●	●	●	●	●	●	●
Rpt5	●	●	●	●	●	●	●
Rpt6	●	●	●	●	●	●	●
PA200	○	○	○	○	○	○	○
20S Core Proteasome							
α1	●	●	●	●	●	●	●
α2	●	●	●	●	●	●	●
α3	●	●	●	●	●	●	●
α4	●	●	●	●	●	●	●
α5	●	●	●	●	●	●	●
α6	●	●	●	●	●	●	●
α7	●	●	●	●	●	●	●
β1	●	●	●	●	●	●	●
β2	●	●	●	●	●	●	●
β3	●	●	●	●	●	●	●
β4	●	●	●	●	●	●	●
β5	●	●	●	●	●	●	●
β6	●	●	●	●	●	●	●
β7	●	●	●	●	●	●	●

● = ortholog present ○ = ortholog absent

3.4| Discussion

Investigation into the genetic composition of the microsporidian secretory pathway has revealed that microsporidia contain all components of the Sec61αβγ translocon, Sec62/63 complex and BiP, which are required for cargo transport into the ER. The Sec61αβγ translocon is required for co-translational translocation whereby peptides in the ribosome bind the SRPC during translation, migrate and bind SRP receptors (SRPR) on the ER surface. At this point the signal peptide is cleaved by signal peptidases and cargo is transported passively via the Sec61αβγ translocon pore (Halic *et al*, 2004). The Sec62/63 complex in contrast is responsible for post-translational translocation, a process in which complete peptides are unfolded and translocated into the ER lumen where they bind luminal chaperone BiP (Kar2 in *S. cerevisiae*) which begins the process of protein refolding (Matlack *et al*, 1997). This suggests microsporidia have the machinery to transport protein cargo into the ER via both co- and post-translational means. Comparative analysis identified that microsporidia have a reduced complement of SRP proteins compared to fungi. They appear to have lost SRP72, SRP14 and SRP68 as well as one of two SRPR protein subunits, SRPRB, a membrane-bound GTPase. Microsporidia however have retained SRP54, which is required for interpretation and binding of the secretion signal as well as the ‘docking’ of cargo with SRPR and Sec translocon proteins at the ER membrane (Bacher *et al*, 1999). Interestingly, a comparison of SRP proteins between *S. cerevisiae* and close ascomycete relative *Candida glabrata* demonstrates that SRP54 othologs between the two species share 82% amino acid identity, far more than SRP72, SRP14 and SRP68 orthologs which share 45%, 55% and 48% identity respectively (Altschul *et al*, 1997). It is therefore possible that orthologs of these more divergent SRP proteins are present in

microsporidian genomes but these orthologs are not identifiable by BLAST search analysis. Signal peptidases are required for signal peptide cleavage at the ER membrane. Interestingly, microsporidia appear to have lost 3 major components of the signal peptidase complex (SPCS1, SPCS2 AND SPCS3) and retained only core catalytic component Sec11. This suggests the microsporidian mechanism for signal peptide processing is both highly adapted and simplified. Sec11 is required for signal peptide cleavage, an essential process in the early secretory pathway, demonstrated by knockout studies in *S. cerevisiae* which accumulate protein cargo intended for export (Bohni *et al*, 1988). Microsporidia have also lost inner mitochondrial proteases (IMP1 and IMP2) from the mitosome, however an IMP2 ortholog is present in the genome of the microsporidian species *A. locustae* (GI: 63029087).

Microsporidian genomes have retained all components of the COPII protein coat, Sec13/31, Sec23/24 and SAR1, a GTPase required for initiating assembly and disassembly of the COPII complex. The COPII protein complex is required for transporting protein cargo from the ER to the Golgi in an anterograde fashion (Hughes and Stephens, 2008). It has been proposed by Beznoussenko *et al* 2007 that microsporidia transport secreted proteins via a tubular Golgi network, independently of COPII and retrograde transport vesicle complex COPI. This raises interesting questions as to why this protein complex has been retained in a highly reduced parasitic phylum. It is possible that COPII has adopted a new functional mechanism for cargo transport within the ‘tubular’ microsporidian secretory apparatus or that vesicle-mediated transport is a yet unidentified component of microsporidian secretion. Microsporidian genomes appear to lack ER ‘quality control’ components UGGT and GlcII which

recognise and attempt to re-fold incompletely or misfolded *N*-glycoproteins (Taylor *et al*, 2003, Dejgaard *et al*, 2004), despite the retention of protein disulphide isomerase orthologs required for protein folding. Interestingly, the machinery required for *N*-glycosylation is also absent, thus consolidating previous studies, which failed to conclusively identify *N*-linked glycans in microsporidia (Taupin *et al*, 2007). Continued molecular investigation will be required to determine if *N*-glycosylation is universally absent from microsporidia.

Transport of protein cargo through the Golgi, in a cis-trans manner, to the extracellular environment requires a complex interaction of SNARE proteins, proceeded by the formation of secretory or exocytic vesicles (Farmaki *et al*, 1999). This process involves continuous retrograde vesicular transport of SNARE proteins as a means of 'recycling' or returning components to their initial functional location (Lewis and Pelham, 1996). The microsporidian SNARE pathway for Golgi transport is extremely reduced. They have retained proteins Sec22 and Bos1 (the latter in the genus *Encephalitozoon* only) required for both anterograde and retrograde Golgi transport, as well as Ykt6 and Stx1 required for vesicle formation and cargo export. However, other core components such as Use1, Gos1 and Bet1 have been lost. It is possible that this severe reduction in transport proteins is a functional hallmark of the 'tubular' microsporidian Golgi or that microsporidia have adopted or acquired novel proteins for cargo export through the Golgi which function independently of vesicle formation. Future molecular and cell biology studies will be required to determine if anterograde cargo export and retrograde protein recycling occur within the tubular Golgi in a vesicle-independent fashion and to elicit the molecular mechanisms by which this process occurs. Understanding this process may reveal the mechanism of

polar tube formation in microsporidia, a unique infection organelle which has been shown to connect directly to the Golgi (Beznoussenko *et al*, 2007). To date little is known about the importance of ER-Golgi independent (non-classical) protein secretion in maintaining microsporidian cellular function and in delivering agents such as virulence factors to the host environment. Determining this will require molecular studies monitoring secretion in the presence of inhibitors such as brefeldin A which blocks ER to Golgi cargo transport (Misumi *et al*, 1986).

Terminally misfolded proteins are transported from the ER via the ERAD pathway to the cytosolic proteasome (Meusser *et al*, 2005). Interestingly, the microsporidian ERAD pathway lacks Png1, a protein that is required for de-glycolsylation of misfolded protein prior to export to the cytoplasm and proteolysis (Suzuki *et al*, 2000). Speculatively, this glycanase may have been lost due to functional redundancy as the upstream machinery required for *N*-glycosylation appears to be absent from sequenced microsporidian genomes. In contrast, *O*-glycoslylation has been described in microsporidian species such as *E. cuniculi*, *E. hellem*, *A. locustae* and *Paranosema grylii* and has been identified as a major post-translational modification of polar tube protein 1, a protein required for host cell invasion (Xu *et al*, 2004, Dolgikh *et al*, 2007, Bouzahzah and Weiss, 2010). The microsporidian proteasome for degradation of terminally misfolded proteins consists of a full 20S core, however non-ATPase 19S regulatory components Rpn12, Rpn13 and Rpn15 are absent as is an identifiable nuclear proteasomal activator PA200 (Ustrell *et al*, 2002). This suggests that microsporidia may have adapted the regulation of this key proteolytic machinery.

The microsporidian secretory pathway is peculiar both in terms of organelle structure and genetic composition. Microsporidia have previously been shown to possess a tubular Golgi and lack genes required for *N*-glycosylation. However, they have retained genes required for COPII formation, despite evidence suggesting that secretion occurs in a vesicle independent fashion. Other secretion-associated complexes and pathways such as the signal peptidase complex have undergone severe reduction suggesting either a simplified or alternative mechanism for signal peptide processing and cleavage. The tubular microsporidian Golgi lacks many SNARE proteins found in model yeast *S. cerevisiae* thus raising interesting questions regarding the molecular mechanism for anterograde and retrograde Golgi transport in the Microsporidia. This study demonstrates that the secretory pathway, consistent with the general biology of microsporidia, has undergone huge reduction, a process that is likely to be a hallmark of specialisation to life in the intracellular environment.

Chapter 4| An investigation of protein secretion during *Spraguea lophii* germination.

4.1| Introduction

Spraguea lophii is a microsporidian species known to infect anglerfish of the genus *Lophius*, a catch of commercial importance in East Asia and many other areas of the world (Thelohan, 1895). The species was first described in 1898 (Döflein, 1898) with the species and genus name assigned by Sprague and Vávra in 1976 (Weissenberg *et al*, 1976). *S. lophii* is thought to reside within a large clade of predominantly salt-water dwelling microsporidian species based on phylogenetic analysis of small subunit ribosomal DNA (Vossbrinck and Debrunner-Vossbrinck, 2005). *S. lophii* infection persists in nervous ganglion cells where the parasite forms large hyper-parasitised xenomas which aggregate to form cysts of several centimetres in diameter and the species has an infection prevalence of up to 83% (Freeman *et al*, 2011). To date little is known about the transmission of *S. lophii* or if any fitness effect is associated with infection, as previous studies have suggested that the xenoma may be the signature of a symbiotic co-existence between host and parasite (Lom and Dykova, 2005). Xenoma-formation is a lifecycle trait predominantly displayed by fish-infecting microsporidian species (Mansour *et al*, 2005), which differs drastically from the two more classical mammalian microsporidian intracellular lifecycle stages of either vacuolar or cytoplasmic merogony (Schottelius *et al*, 2000). It also remains unclear whether the xenoma wall is of host, parasite or mixed origin, a crucial question with regards to the basis of the host-parasite interaction. An original *S. lophii* genome sequencing survey described 120 kb of genome sequence; however a more recent deep sequencing survey has since been completed (Hinkle *et al*, 1997, Campbell *et al*, 2013). The *S. lophii* genome provides a platform to investigate the biology of the organism at the molecular level and to address questions regarding the contribution of parasitic

agents towards xenoma-formation through comparative genomics to other microsporidian species for which a complete genome exists, such as, *Nosema ceranae* and *Encephalitozoon intestinalis* (Cornman *et al*, 2009, Corradi *et al*, 2010). The *S. lophii* genome represents both the first xenoma-forming and fish-infecting microsporidian genome sequence.

The 'hyper-parasitic' nature of *S. lophii* infection provides simple purification of large quantities of spores, a major limitation for studying spore germination in other microsporidia, including species such as *Encephalitozoon hellem* and *Trachipleistophora hominis* with well described culture systems (Didier *et al*, 1991, Visvesvara *et al*, 1999). Spore germination although thought to be complex *in vivo* can be induced efficiently *in vitro,* for example, with the addition of H_2O_2, calcium ionophore A23187 or monovalent cations to spore containing medium (Weidner and Byrd, 1982, Frixione *et al*, 1994, Findley *et al*, 2005). As spore germination represents the microsporidian mechanism for infection, the need for understanding this process at the molecular level is apparent. Research to date has been directed primarily at identifying cellular and morphological changes during spore germination. Studies are in agreement that polar tube expulsion is caused by an increased osmotic pressure in the microsporidian posterior vacuole, however several methods have been proposed for inducing this phenonemon, including: 1) The import of water via aquaporin proteins to cause increased osmotic pressure in the posterior vacuole (Ghosh *et al*, 2006) and 2) Trehalase-mediated breakdown of the trehalose carbohydrate store into glucose monomers, thus creating a suitably high osmotic potential to induce water influx into the hypertonic spore (Frixione *et al*, 1992, Dolgikh and Semenov, 2003). Original hypotheses suggested that

polar tube expulsion during germination physically punctures the host cell providing a tunnel into the host for infectious material. However, recent studies have suggested that the polar tube tip may induce an endocytic event at the host cell membrane (Bohne *et al*, 2011). In contrast, the complex cascade of host-parasite interactions during germination-mediated infection has been largely overlooked.

To determine the degree of secretome conservation between diverse microsporidian species, microsporidian proteomes were downloaded from publically available genome databases and bioinformatic tools, such as OrthoMCL ortholog clustering software were employed to compare the predicted secretomes of microsporidian species including human-infective *E. cuniculi* and the honey-bee pathogen *N. ceranae.* This analysis was then used to identify unique secreted proteins at the species- and strain-specific (where available) level. To determine the extent to which HGT had contributed to microsporidian secretome evolution, phylogenetic methods were then employed to investigate the evolutionary origin of all species-specific secreted proteins identified by ortholog clustering analysis.

To identify potential *S. lophii* virulence factor proteins required for host cell invasion and early infection, the molecular basis of *S. lophii* spore germination was investigated using a combination of proteomics and transcriptomics. The transcriptome of the germinating cell was explored and used to identify a set of genes that were both highly expressed during germination and encoded a secretion signal. Similarly, whole cell proteomic changes during germination were profiled using a complex-mix proteomic

comparison of germinated and non-germinated spores. To identify a list of potential effector proteins secreted to the extracellular environment during *S. lophii* germination, a novel assay investigating the complement of proteins secreted into the extracellular medium was developed and employed in conjunction with existing *in vitro* germination assays and mass spectrometry. These results suggest that both *S. lophii* and microsporidian specific hypothetical proteins may play an important role in host cell invasion and early microsporidian infection. Due to the lack of a genetic manipulation system, protein characterisation in microsporidia currently relies on heterologous expression systems or indirect assays that are expensive and laborious. This study provides a novel method for identifying a small number of secreted hypothetical proteins which may play a significant role in microsporidian infection of the host cell, from the hundreds of hypothetical proteins encoded in microsporidian genomes.

4.2| Methods

4.2.1| Database mining and ortholog clustering

MicrosporidiaDB (http://microsporidiadb.org) was searched for predicted secreted proteins in the proteomes of microsporidian species (*Enecephalitozoon cuniculi* GB-M1, *Encephalitozoon hellem*, *Encephalitozoon intestinalis*, *Enterocytozoon bieneusi*, *Nosema ceranae*, *Nematocida parisii* ERTm1, *Nematocida* sp. ERTm2 and *N. parisii* ERTm3) using the following pre-defined SignalP score criteria: SignalP-NN conclusion score ≥ 3, SignalP-NN D-score ≥ 0.5 and a SignalP-HMM signal probability score ≥ 0.5 (Aurrecoechea *et al*, 2010, Petersen *et al*, 2011). Proteins classified as 'secreted' had to match all search criteria. Predicted secreted proteins in *S. lophii* were identified manually conducting a manual search of the proteome against the same SignalP criteria. The resultant predicted secretome files were clustered using OrthoMCL using a pre-defined e-value cut-off of $1e^{-20}$ and an inflation value of 1.5 to identify species-specific proteins with no ortholog in any other species. Species and strain-specific secreted proteins indentified in microsporidian species were searched against the National Centre for Biotechnology Information non-redundant protein database using BLASTp (protein-protein BLAST) with a pre-defined e-value cut-off of $1e^{-5}$ to identify orthologous protein sequences. Sequence hits against other microsporidian sequences were removed from the analysis (as false positives from OrthoMCL clustering). Protein sequences that retrieved hits from prokaryotic or non-fungal eukaryotic species were taken as potential cases of HGT and investigated by phylogenetic analysis. As the genome sequence of *E. bieneusi* was sequenced directly from spores isolated from faeces and many cases of bacterial contamination have been identified, *E.*

bieneusi proteins retrieving prokaryotic orthologs were deemed contaminants and removed from further analysis (Akiyoshi *et al*, 2009).

4.2.2| RNA extraction and transcriptomic analysis of germinated spores

S. lophii spores were purified from xenomas of North Atlantic monkfish species *L. piscatorius* as described in Chapter 2.1 (xenomas were kindly provided by Brixham Trawler Agency, Brixham UK). Purified *S. lophii* spores were germinated following the protocol described in Chapter 2.14. A 200 µl of pellet of germinated *S. lophii* spores was frozen with liquid nitrogen and disrupted manually using a pestle and mortar for approximately 15 minutes. RNA was then extracted following using TRIzol (Invitrogen) as described in Chapter 2.3. The RNA was resuspended in 50 µl Milli-Q water (Millipore) and quantified at 2260 ng/ul (absorbance: 1) 260/280= 2.12 2) 260/230 = 2.41). The RNA integrity was verified on a 1.5% TAE agarose gel by visualising the presence of 2 distinct ribosomal RNA bands prior to sequencing. Intact RNA was then sent to University of Exeter Sequencing Service for Illumina TruSeq RNA sequencing analysis. Raw Illumina reads obtained from RNA sequencing were filtered to remove adapter sequences and low quality reads, assembled and quantified (relatively) by Dr Tom Williams (University of Newcastle).

4.2.3| Whole cell protein analysis of germinated and non-germinated spores

S. lophii spores were purified from xenomas of North Atlantic monkfish species *L. piscatorius* as described in Chapter 2.1. Purified spores were adjusted to a volume of ~200 µl. Spores were then germinated as described in Chapter 2.14 alongside a non-germinated control sample. Spores were then washed three

times in sterile 500 µl 1 x PBS to remove germination medium and resuspended in 400 µl 1 x PBS following the final wash. Subsequently, samples were added to 200-300 µl pre-cooled sterile glass beads (Sigma Aldrich) and homogenised using the MP Fastprep homogenizer with 3 x 20 second bursts at 4.1 m/s. The spore samples were incubated on ice for 3-5 minutes between each Fastprep treatment to prevent overheating. Samples were then centrifuged at 12,000 x g for 15 minutes at 4°C. The resultant supernatant containing complex-mix protein for both germinated and non-germinated samples were then sent to the University of Exeter Mass Spectrometry Facility and analysed using the 6520 accurate mass quadrupole-time of flight (Q-TOF) mass spectrometer (Agilent Technologies). Peptide mass output was searched against the translated *S. lophii* genome database. Resultant protein hits from three biological replicates of both germinated and non-germinated samples were pooled and quality filtered, where any protein without two distinct peptide hits and a percent score peak intensity (% SPI) of ≥ 60% was removed from the results. The molecular function of proteins identified in germinated and non-germinated spores was predicted using InterPro protein sequence analysis and classification database (Hunter *et al*, 2012).

4.2.4| Identification of proteins secreted extracellularly during spore germination

S. lophii spores were purified as described in Chapter 2.1. Three biological spore replicates isolated from unique xenoma isolates were adjusted to a volume of 500 µl and germinated following the protocol described in Chapter 2.14. Germinated *S. lophii* spores along with a non-germinated control were then spun at 12,000 x g for 15 minutes at 4°C to remove spore material from

extracellular medium. The resultant supernatant was then collected and concentrated using Millipore Amicon 3 kDa centrifugal column to concentrate supernatant at 2,000 x g for 30 minutes at 4°C. The concentrated extracellular protein was then quantified and checked on 12% SDS gels. Complex mixtures of extracellular secreted proteins from germinated and non-germinated samples were sent to University of Exeter Mass Spectrometry Facility and analysed by mass spectrometry. The resultant peptide masses were searched against a translated *S. lophii* genome database. The experiment was conducted in triplicate and any protein without two distinct peptide hits and a percent score peak intensity (% SPI) of ≥ 60% was removed from the analysis.

4.3| Results

4.3.1| The microsporidian secretome

The term ‘secretome’ which is often simply described as the complete set of secreted proteins expressed by a cell, tissue or organism, in this study will be re-defined as the complete set of proteins secreted extracellularly and proteins secreted to the membrane which possess extracellular domains (Martinez *et al*, 2006). Secreted proteins play a range of functional roles in cellular and organismal processes such as signalling and cell defence, and in pathogensare well described virulence factors. For example, the secretome of *Plasmodium falciparum*, which consists of several hundred proteins, is known to remodel the erythrocyte host cell (van Ooij *et al*, 2008). Due to the obligate intracellular nature and intimate host-parasite interaction displayed throughout the microsporidian lifecycle, the secretome undoubtedly plays a role in invasion, host manipulation, lifecycle development and cell egress. The secretome for microsporidian species, as predicted by SignalP, ranges in size from 84 proteins in *E. romaleae* to 257 proteins in *Nematocida* sp. Ertm2 (Figure 4.1). Comparison of secretome size to that of the total proteome suggests that the microsporidian secretome comprises 5-10% of all proteins encoded in microsporidian genomes. Bioinformatic prediction tools such as SignalP, which has a reported accuracy of up to 91%, are modelled and tested on model organisms such as *Homo sapiens* and *Mus musculus* (Klee and Ellis, 2005). Experimental confirmation will be required to determine the true accuracy of such sequence based prediction programs on more divergent eukaryotes such as microsporidia (Campbell *et al*, 2013). The predicted secretome of microsporidian species predominantly consists of phylum, lineage or species-specific hypothetical proteins with no similarity to other sequences in NCBI or

identifiable functional domain in Pfam (<$1e^{-5}$) (Marchler-Bauer *et al*, 2011, Punta *et al*, 2012). Interestingly, there appears to be no obvious correlation between lifecycle and secretome conservation. Cytoplasmic dwelling species including *E. bieneusi* and *N. ceranae* show no greater level of secretome conservation to each other than they do with vacuolar dwelling species such as *E. cuniculi* or the xenoma-forming species *S. lophii*. Only 17 conserved secreted proteins, that is, secreted proteins retained in all microsporidia examined, were identified by ortholog clustering analysis. This included conserved hypothetical proteins, hydrolase enzymes and a range of transporters that may represent factors that were present in the 'ancestral microsporidian' and are essential for microsporidian development or lifecycle progression in a diverse range of host species. *S. lophii* is currently the only xenoma-forming microsporidian with an available genome sequence (Campbell *et al*, 2013). Future genome sequencing projects of xenoma-forming microsporidian species such as *Loma salmonae* and *Glugea stephani* will be required to employ comparative genomics and determine the contribution of the secretome to xenoma-formation. Phylogenetic analysis of secreted proteins could not confirm any cases of HGT, despite several proteins such as Eint_061610, EBI_21897 and NERG_00018 demonstrating a non-canonical taxonomic distribution.

Microsporidia also display variation in the number of species-specific secreted proteins translated from their genomes (Figure 4.2). For example, all secreted proteins encoded by closely-related species *E. romaleae* and *E. hellem* are present in another member of the *Encephalitozoon* genus. In contrast the secretomes of *S. lophii* and *N. ceranae* both contain over 40 members with no identifiable ortholog in any other microsporidian species for

which a genome sequence is available (Figure 4.2). This reflects the close relationship between sequenced *Encephalitozoon* species compared to *N. ceranae* and *S. lophii.* Few functional domains were identified in species-specific secreted proteins. However, an IncA domain protein was identified in the human-infective microsporidian *E. intestinalis* (GenBank Accession: Eint_020090). An IncA domain protein has previously been indentified as an essential component of the *Chlamydia trachomatis* infected vacuole membrane and is required for the fusion of multiple vacuoles within an infected host-cell (Fields *et al*, 2002). A similar function for IncA in *E. intestinalis,* a vacuolar dwelling microsporidian species, cannot be overlooked. Interestingly, *E. cuniculi* and *C. trachomatis* co-infection has been described previously with the two intracellular pathogens occupying independent vacuoles within a host cell (Lee *et al*, 2009). However, there are no identified IncA domain genes in the *E. cuniculi* genome and there is no evidence to suggest that this gene was acquired by *E. intestinalis* by HGT from *C. trachomatis*. An *Encephalitozoon* specific serine/theronine phosphatase (GenBank Accession: NP_586079) was also identified. It is possible this *Encephalitozoon*-specific adaptation is secreted to disrupt host cell signalling or protein-protein interactions by de-phosphorylating serine or threonine amino acids. Genome sequencing of closely related species or members of the same genera will identify if these secreted proteins are truly species-specific innovations. It is, of course, also possible that these secreted proteins are evolving sufficiently quickly due to a positive host-driven selective pressure that orthologs can not be identified by sequence-based BLAST analysis or that species-specific proteins have been incorrectly identified during genome annotation.

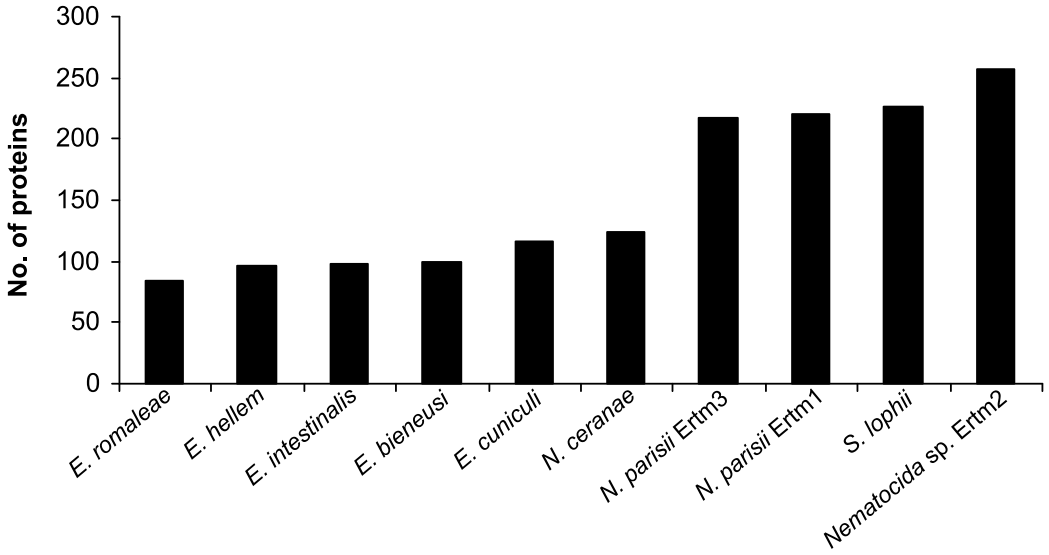


Figure 4.1| Predicted secretome size in diverse microsporidia.

Predicted secreted proteins matching SignalP criteria (NN conclusion score ≥ 3, NN D-score ≥ 0.5, HMM S-score ≥ 0.5) for diverse microsporidian species. Secretomes from microsporidian species *E. romaleae*, *E. hellem*, *E. intestinalis*, *E. cuniculi*, *E. bieneusi*, *N. ceranae*, *N. parisii* Ertm1, *N. parisii* Ertm3 and *Nematocida* sp. were downloaded from MicrosporidiaDB (Petersen, Brunak *et al.*2011) (Aurrecoechea, Brestelli *et al.*2010). The *S. lophii* proteome was searched against the same criteria manually. Secretome comparison shows > 2 fold variation in protein number ranging 84 predicted secreted proteins in *E. romaleae* to 257 in *Nematocida* sp. Ertm2. The *S. lophii* genome encodes a total of 227 predicted secreted proteins.

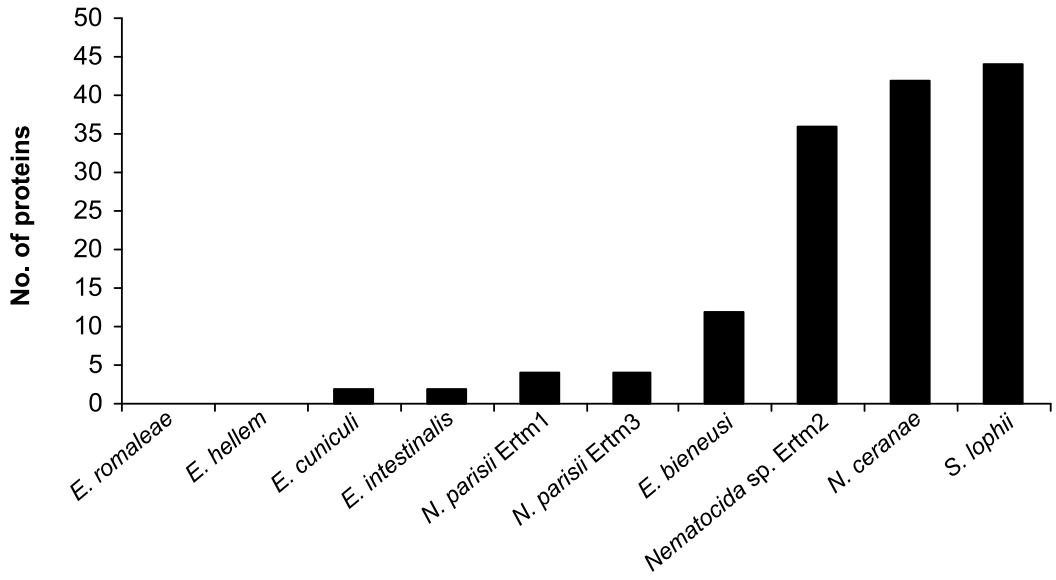


Figure 4.2| Number of species-specific secreted proteins.

Species and lineage specific-secreted proteins were identified using OrthoMCL protein clustering software. Species-specific protein output was then checked for false positives using BLASTp analysis. The number of species-specific proteins ranges from 0 in *E. hellem* and *E. romaleae* to 44 in *S. lophii*, thus reflecting the close relationship between members of the *Encephalitozoon* genus relative to the other species examined. These species-specific proteins may be crucial for lifecycle development and virulence in each specific host-parasite system.

4.3.2| The transcriptome of germinated *S. lophii* spores

To better understand microsporidian spore germination at the molecular level, *S. lophii* spores were purified from xenomas found in the ganglion cells of the spinal nerves of monkfish host *L. piscatorius* (Figure 4.3). RNA was extracted from *S. lophii* spores germinated *in vitro* (Figure 4.4). The corresponding cDNAs were sequenced by the non strand-specific TruSeq approach using Illumina sequencing by the University of Exeter Sequencing Service. The *de novo* transcriptome assembly contained 12,932 unique transcripts, of which 2,896 mapped to the *S. lophii* genome (data assembled by Dr Tom Williams, University of Newcastle). Although the most abundant transcripts corresponded to 18S ribosomal RNA, the most highly expressed gene in *S. lophii* spores corresponds to a hypothetical protein (SLOPH 2267, 3.27% of transcriptome) found in a range of microsporidian species including the human-infective *E. cuniculi* and the honey bee pathogen *N. ceranae* (Table 4.1). The gene, however, appears to be absent in other microsporidian species including *Caenorhabditis elegans* pathogen *N. parisii,* suggesting that the gene does not have a universal function in microsporidia during the germination process. Six other hypothetical proteins are found among the top 20 most highly expressed transcripts of which three are specific to *S. lophii* (SLOPH 857, 997, 655) and three are found in other microsporidian species (SLOPH 1855, 1788, 1203) (Table 4.1). This strongly suggests an important role for species, lineage-specific or very fast evolving proteins in *S. lophii* and microsporidian spore germination. Future sequencing projects of fish-infecting microsporidia will be required to determine if these *S. lophii* specific hypothetical proteins are adaptations required for infection of the fish host. Previous transcriptional studies have also demonstrated the importance of hypothetical proteins in the

early infection of *C. elegans* by *N. parisii* Ertm1. Cuomo *et al* demonstrated that 37/40 of the most highly expressed transcripts at 8 hours compared to 60 hours post infection were of hypothetical functional annotation (Aurrecoechea *et al*, 2010, Cuomo *et al*, 2012). Collectively these data suggests that these hypothetical proteins may represent good targets for future experimental work on the maintenance of the parasitic lifecycle both in *S. lophii* and other microsporidian species. Many expected candidates that function in core cellular pathways such as histone H3, protein kinases and zinc finger proteins were also found among the top 20 most highly expressed transcripts. Interestingly, actin, myosin and a rho1 GTPase involved in actin organisation are all highly expressed during germination; it is therefore possible that the germination process requires major re-organisation of the parasite cytoskeleton (Sit and Manser, 2011). Alternatively, these cytoskeletal genes may have an important function throughout the microsporidian lifecycle and may therefore also be highly expressed in non-germinated *S. lophii* spores.

SignalP4.1 was employed in conjunction with transcript abundance data to determine whether the 10 most highly expressed transcripts encoding secreted proteins (Table 4.2). Secreted proteins only comprise a small percentage of the overall transcriptome and are likely masked by more abundant housekeeping and core metabolic genes. Despite this, 6/10 most highly expressed secreted transcripts are hypothetical proteins all of which are specific to *S. lophii*. This suggests that protein secretion during germination is a specific adaptation to the *S. lophii – L. piscatorius* host-parasite system or that the proteins are evolving too rapidly for ortholog detection in other microsporidian species by conventional BLASTp analysis ($< 1e^{-5}$). Interestingly,

a leucine rich repeat protein (SLOPH 1278) is also present in the top 10 most highly expressed predicted secreted transcripts. This represents a member of a large expanded leucine-rich repeat gene family present in the *S. lophii* genome, of which many members are predicted to be secreted by both SignalP4.1 and TargetP1.1 but lack transmembrane domains as predicted by TMHMM2.0 (Emanuelsson *et al*, 2000, Krogh *et al*, 2001, Petersen *et al*, 2011). This suggests that these leucine rich repeat proteins may represent a large *S. lophii* effector family (Campbell *et al*, 2013). Expanded gene families have been described previously in microsporidia such as a non-related leucine enriched family of proteins in *T. hominis* and an InterB gene family in members of the genus *Encephalitozoon* and other human-infective species including *Vittaforma corneae* and *Anncalia algerae* (Dia *et al*, 2007, Heinz *et al*, 2012).

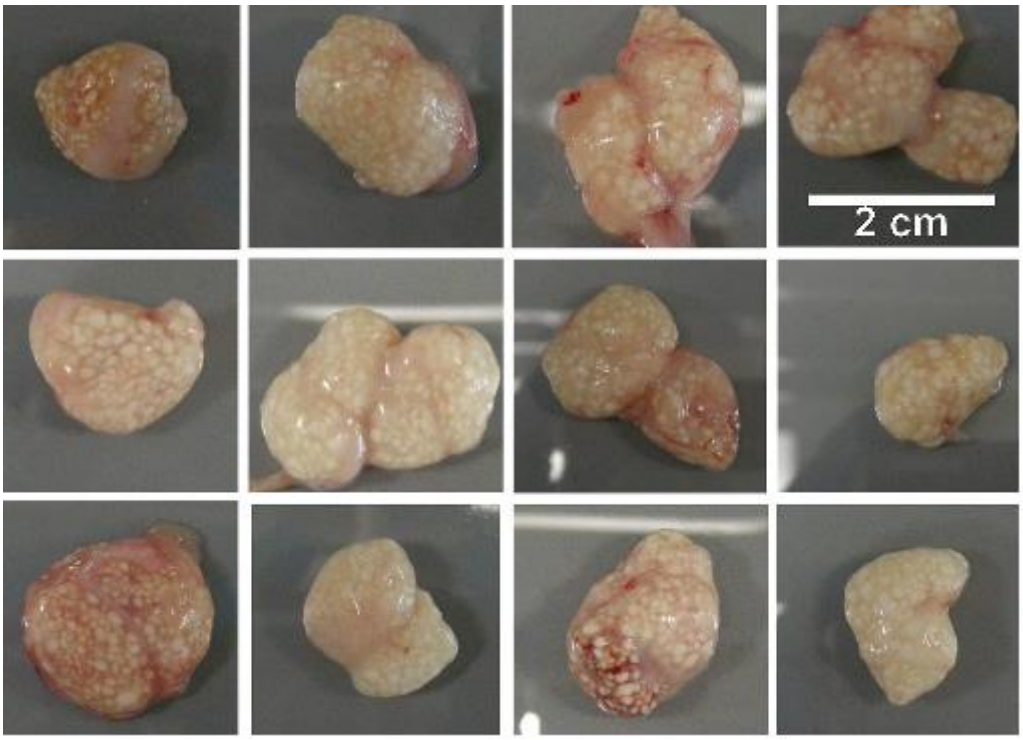


Figure 4.3| Aggregations of individual *S. lophii* filled xenomas.
S. lophii infects the nervous ganglion cells of local monkfish species *L. piscatorius* forming individual hyper-parasitised xenomas. These xenomas aggregate to form large cysts encased in a thin layer of fish tissue around the spinal cord. These cysts were a generous gift from Brixham Trawler Agency.

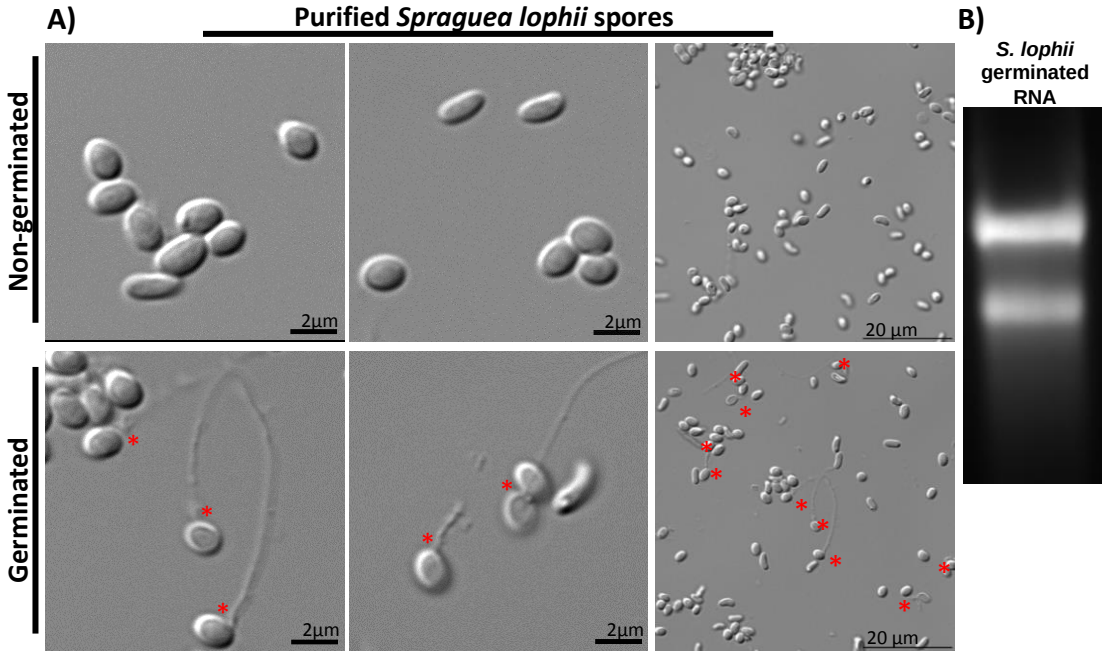


Figure 4.4| *In vitro* germination of *S. lophii* spores.

A) Examples of germinated *S. lophii* spores following *in vitro* germination using Gly-Gly buffer and calcium ionophore alongside a non-germinated control. The site of polar tube expulsion (indicated by *) is required for the transfer of infectious material into the host cell environment. B) RNA extracted from germinated *S. lophii* spores run on a 1.5% TAE agarose gel. Intact 28S and 18S ribosomal RNA bands are clearly visible thus demonstrating RNA integrity.

Table 4.1| Most abundant transcripts from germinated *S. lophii* spores.

Protein ID	Protein annotation	% of transcriptome
SLOPH 2267	Similar to hypothetical protein NCER 10017	3.27
SLOPH 2328	ADP-ribosylation factor	3.13
SLOPH 857	Hypothetical protein	2.78
*SLOPH 54	60S acidic ribosomal protein P2	2.13
SLOPH 997	Hypothetical protein	1.95
*SLOPH 2039	ubiquitin/40s ribosomal protein S27a fusion	1.76
+SLOPH 655	Hypothetical protein	1.57
SLOPH 2438	GTP-binding protein rho1	1.45
SLOPH 1855	Similar to hypothetical protein NCER 101353	1.25
*SLOPH 1540	RAS GTPase	1.03
*SLOPH 2369	14-3-3 protein	0.97
+SLOPH 431	Histone H3	0.95
SLOPH 2171	40S ribosomal protein S28 Intron	0.88
SLOPH 128	AN1-like Zinc finger protein	0.77
SLOPH 1788	Similar to hypothetical protein ECU09 1470	0.75
*SLOPH 101	Actin	0.75
SLOPH 2162	RNA binding domain	0.65
SLOPH 1405	Nucleoside diphosphate kinase	0.64
SLOPH 1378	Myosin regulatory light chain	0.61
SLOPH 1203	hypothetical protein ECU08 0980	0.61

+ = also found in secreted proteomics,
* = also found in germinated or non-germinated proteome

Table 4.2| Most abundant predicted secreted transcripts from germinated *S. lophii* spores.

Protein ID	Protein annotation	SignalP	% of transcriptome
SLOPH 891	Hypothetical protein	✔	0.47
SLOPH 2045	V-type ATP synthase	✔	0.05
SLOPH 556	Hypothetical protein	✔	0.04
SLOPH 168	Brr6 like protein	✔	0.03
SLOPH 774	Hypothetical protein	✔	0.02
SLOPH 1087	Hypothetical protein	✔	0.02
SLOPH 1211	ER membrane protein	✔	0.01
SLOPH 1278	Leucine rich repeat protein	✔	0.01
SLOPH 904	Hypothetical protein	✔	0.01
SLOPH 515	Hypothetical protein	✔	0.01

4.3.3| Proteomic analysis of germination and secretion.

Understanding spore germination at the molecular level is a requirement for dissecting the microsporidian interaction with the host during invasion and early infection (Takvorian *et al*, 2005). Despite the development of several artificial *in vitro* methods for microsporidian spore germination in a range of species, the host stimuli responsible for inducing this phenomenon *in vivo* are not currently known (Pleshinger and Weidner, 1985, Frixione *et al*, 1994, Frixione *et al*, 1997). To date, no microsporidian effector proteins with a role in invasion of the host cell have been identified or characterised experimentally. However, due to the vast quantity of *S. lophii* spores, well described spore purification and germination protocols, and an available genome sequence, *S. lophii* represents an attractive model system to study spore germination at the transcriptional and/or proteomic level.

To identify proteins expressed specifically in germinated *S. lophii* spores, complex-mixture whole protein extractions from both germinated and non-germinated spores were analysed by mass spectrometry. After pooling and filtering 3 biological replicates, the analysis showed little variation at the proteomic level between the two lifecycle stages, with unique peptide hits generally only present in 1/3 biological replicates (for summary see Figure 4.5, for full spore proteome see Appendix 2). Many components of core cellular pathways including histones, heat shock proteins and ribosomal proteins were identified in both lifecycle stages. Interestingly, glycolytic enzymes including a glyceraldehyde-3-phosphate dehydrogenase (SLOPH 399) and phosphofructokinase (SLOPH 1429) were identified in germinated and non-germinated spores, demonstrating that glycolytic pathways are active in

extracellular *S. lophii* spores. These data are consistent with studies in *Antonospora locustae* and *Trachipleistophora hominis* which identified active glycolytic enzymes in extracellular spores by immuno-localisation and proteomics (Dolgikh *et al*, 2011, Heinz *et al*, 2012). Coat protein complex II (COPII) components, Sec23 and Sec24, which function in anterograde protein transport through the ER and Golgi, were identified specifically in the germinated protein sample. This may indicate that protein secretion is activated or up-regulated specifically upon microsporidian spore germination. Overall a surprisingly conserved repertoire of proteins was identified in both lifecycle stages (Figure 4.5). The process of microsporidian spore germination can occur in less than 2 seconds, with a polar tube discharge speed greater than 100 µm/s (Frixione *et al*, 1992). It is likely that germination occurs too rapidly to allow for protein translation, with the parasite pre-packaging the proteins required for early infection prior to receipt of the germination stimulus. Therefore obvious changes are not evident between germinated and non-germinated spores at the proteomic level. Alternatively, it may be that the samples are dominated by highly expressed housekeeping proteins which function universally throughout the microsporidian lifecycle and that the proteins which are expressed specifically during germination are present in low abundance and hence were not detected by mass spectrometry. Gene ontology analysis of proteins identified in both germinated and non-germinated spores using Uniprot (Consortium, 2012) identify translation as the dominant function during both lifecycle stages. There are also numerous hypothetical proteins, proteins involved in ATP binding and proteins functioning in proteolysis (Figure 4.5).

To identify potential effector proteins involved in host cell invasion, the complement of proteins present in the extracellular medium following *in vitro* germination was explored. SDS-PAGE analysis identified several proteins that were consistently secreted to the extracellular medium during germination (Figure 4.6). Three hypothetical proteins were identified by band excision followed by LC/MS analysis, namely, SLOPH 477, a *Spraguea*-specific hypothetical protein and SLOPH 1709 and SLOPH 2344, which are widespread among microsporidian species (Figure 4.6). Importantly, no proteins were identified in the supernatant of non-germinated *S. lophii* spores, suggesting that these identified proteins are released specifically by *S. lophii* upon germination. Mass spectrometry analysis of the extracellular medium following germination identified a total 37 potential effector proteins across 3 biological replicates. This included 10 proteins which were predicted to be secreted by SignalP4.1 and 17 which were predicted to be directed to the secretory pathway by TargetP1.1; 10 proteins were predicted to be secreted by both prediction programs (Table 4.3). These secreted proteins display a vast range in size, ranging from 101 amino acids (SLOPH 433, histone H4) to 935 amino acids (SLOPH 1465, polar tube protein 3); suggesting *S. lophii* does not display a bias towards secreting small proteins during germination (Table 4.3). Of these 37 identified proteins, 19 are hypothetical proteins which are either specific to *S. lophii* or to the microsporidian phylum. A total of 5 hypothetical proteins were found in all 3 biological replicates, of which 3 are *S. lophii*-specific proteins (BLASTp e-value $<1x10^{-5}$) (SLOPH 477, 723, 762) and 2 are encoded in the genomes of other microsporidian species (SLOPH 1766, 1854) (for example peptide spectra see Figure 4.7). SLOPH 1854 shows significant sequence similarity to spore wall protein 7 (SWP 7) in *Nosema bombycis*, a spore wall component previously

identified by mass spectrometry analysis of extracellular membrane protein extracts, while SLOPH 1766 contains a Ricin-B-lectin domain which may mediate parasite binding to host cell surface carbohydrates (Wu *et al*, 2008, Campbell *et al* 2013). A further ten proteins were identified in 2/3 biological replicates. This included 6 proteins specific to *S. lophii* which may represent pathogenicity factors required specifically for invasion of the *L. piscatorius* host cell (Table 4.3). Interestingly, a second Ricin-B-lectin protein (SLOPH 691) with identifiable orthologs in other microsporidia was found in extracts of 2/3 replicates. This suggests that this protein family may contribute to microsporidian invasion across a diverse range of hosts. A chitin deacetylase which has been studied previously in *E. cuniculi* was also present in 2/3 replicates (SLOPH 204). This chitin deacetylase in *E. cuniculi* localised to the chitinaceous endospore, however the protein was unable to bind chitin (Brosson *et al*, 2005, Brosson *et al*, 2006, Urch *et al*, 2009). Urch *et al* therefore speculate that the protein has evolved to function as a lectin, binding carbohydrates within the *E. cuniculi* endospore (Urch *et al*, 2009). Hypothetical proteins SLOPH 2344 and SLOPH 1749, which were present in 2/3 biological replicates, have orthologs in all microsporidia examined and are predicted to be secreted in all species with the exception of *E. intestinalis* and *E. bieneusi* (SLOPH 2344) and *E. bieneusi* (SLOPH 1749) (Table 4.3, Table 4.4). A further 3 microsporidian-specific hypothetical proteins were identified in 2/3 extracellular protein replicates (SLOPH 1854, 1766, 691); orthologs of these proteins displayed a varied taxonomic distribution across microsporidian species in conjunction with a varied secretion prediction profile (Table 4.3, Table 4.4). Proteins including histones and a translation elongation factor which were identified in the analysis may represent abundant proteins within

microsporidian spores, which are not secreted but have been identified in the extracellular medium falsely and are in fact contaminants from the spore pellet. However, it is possible that other proteins identified in the extracellular medium which lack a secretion signal, particularly those found in at least 2 biological replicates, may be secreted by a non-classical/ER-Golgi independent mechanism.

The *S. lophii*-specific hypothetical proteins identified in this study may represent parasite factorsrequired for infection of the fish, or more specifically *L. piscatorius* host cell. Alternatively, these hypothetical proteins may be subject to a strong host-driven selective pressure causing them to evolve quickly, therefore, it is possible that orthologs exist in other microsporidian species but these are too divergent to be identified. This proteomic analysis consolidates the transcriptomic data in emphasising the importance of uncharacterised, hypothetical proteins in microsporidian germination. It is possible that these hypothetical proteins function specifically in host cell invasion. This proteomic dissection of microsporidian spore germination provides a novel approach to identify a small subset of hypothetical proteins which may have an important role in pathogenicity, representing strong candidates for further experimental characterisation from the 1000’s of hypothetical proteins encoded in microsporidian genomes.

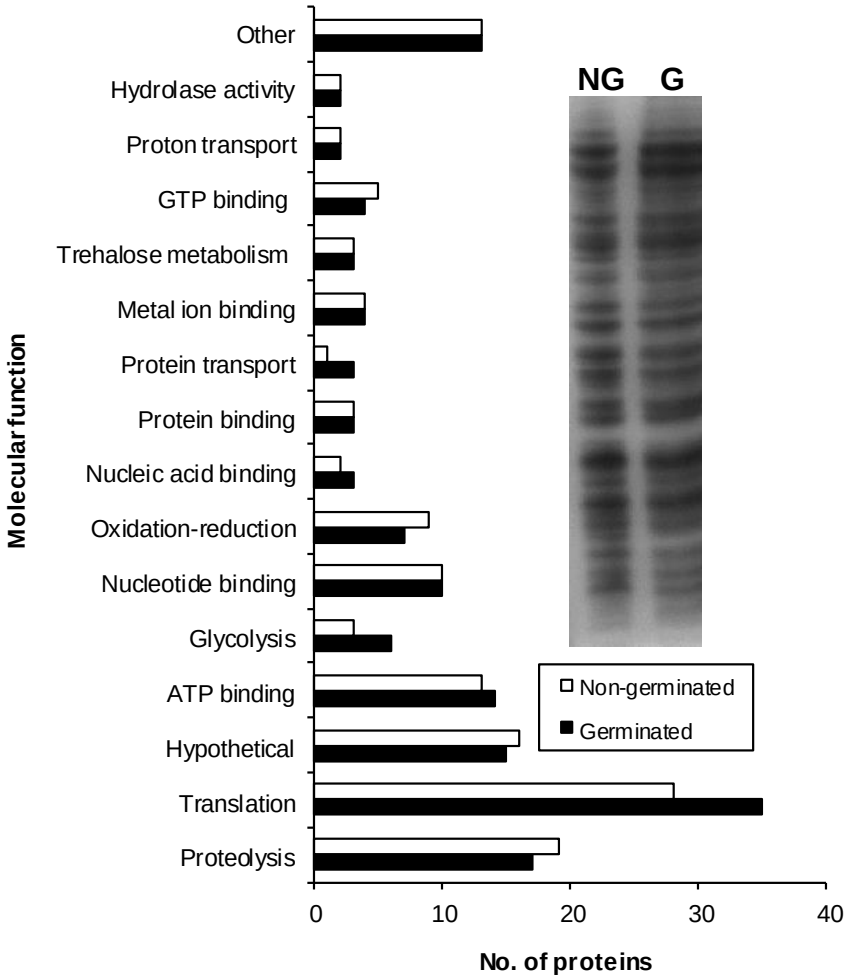


Figure 4.5| Gene ontology of proteins identified from germinated and non-germinated spores.

Mass spectrometry analysis identified a total of 143 and 142 proteins in the germinated and non-germinated whole cell protein extracts respectively. Proteomic analysis revealed no notable changes between the two lifecycle stages which could be linked to cellular function. The chart displays the predicted molecular function of proteins identified in both lifecycle stages thus identifying translation, proteolysis, ATP binding and hypothetical as the most prevalent predicted cellular functions during both extracellular *S. lophii* lifecycle stages.

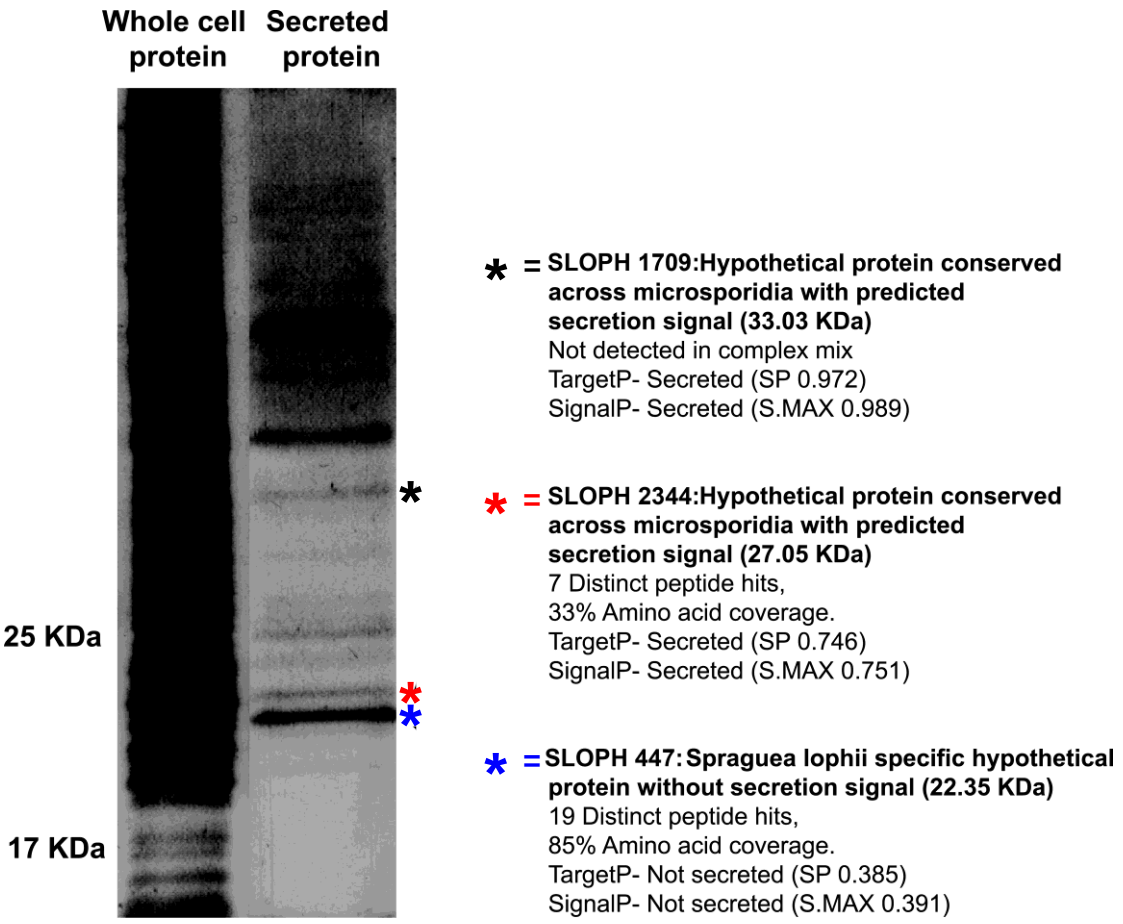


Figure 4.6| SDS-PAGE gel displaying total and secreted protein from germinated *S. lophii* spores.

Total and extracellular protein from germinated *S. lophii* spores was run on 12% SDS-PAGE gel and stained with Coomassie blue stain. Prominent and reoccurring low molecular weight extracellular protein bands were excised and analysed by mass spectrometry. All proteins identified are either *S. lophii*- or microsporidian-specific hypothetical proteins. SLOPH 2344 and 1709 are predicted to be secreted by both SignalP4.1 and TargetP1.1 with strong support.

Table 4.3| Identification of proteins secreted extracellularly during *S. lophii* spore germination.

S. lophii ID	Annotation	SignalP	TargetP	Replicate 1	Replicate 2	Replicate 3	Size (AAs)
SLOPH 477	Hypothetical protein			●	●	●	195
SLOPH 723	Hypothetical protein		S	●	●	●	235
SLOPH 1854	Hypothetical protein	S	S	●	●	●	270
SLOPH 762	Hypothetical protein	S	S	●	●	●	458
SLOPH 1766	Hypothetical protein		S	●	●	●	260
SLOPH 2241	Hypothetical protein		S	●	●		537
SLOPH 2344	Hypothetical protein	S	S	●	●		238
SLOPH 580	Hypothetical protein			●	●		438
SLOPH 607	Hypothetical protein			●	●		299
SLOPH 689	Hypothetical protein		S	●	●		425
SLOPH 204	Chitin deacetylase	S	S	●	●		257
SLOPH 1121	Hypothetical protein	S	S	●	●		399
SLOPH 854	Hypothetical protein			●	●		375
SLOPH 1749	Hypothetical protein			●	●		378
SLOPH 691	Hypothetical protein	S	S		●	●	211
SLOPH 1980	Translation elongation factor 2			●			847
SLOPH 1483	ATPase PP-loop			●			240
SLOPH 401	Glycerophosphoryl diester phosphodiesterase			●			355
SLOPH 373	Formin like protein		S	●			812
SLOPH 119	Alanyl-tRNA synthetase			●			852
SLOPH 1945	Thioredoxin				●		107
SLOPH 452	Hydroxyacylglutathione hydrolase		S		●		250
SLOPH 1662	Hypothetical protein		S		●		126
SLOPH 655	Hypothetical protein		S		●		215
SLOPH 692	Hypothetical protein	S			●		411
SLOPH 1083	Hypothetical protein	S	S		●		172
SLOPH 1419	Peptidyl-prolyl cis-trans isomerase				●		162
SLOPH 2233	Ribosomal protein L23					●	142
SLOPH 2557	Hypothetical protein	S	S			●	181
SLOPH 428	Histone H2A					●	193
SLOPH 430	Histone H3					●	144
SLOPH 433	Histone H4					●	101
SLOPH 439	HMG box protein					●	166
SLOPH 28	Ribosomal protein S12					●	130
SLOPH 431	Histone H3					●	139
SLOPH 1465	Polar tube protein 3	S	S			●	935
SLOPH 449	Hsp70					●	691

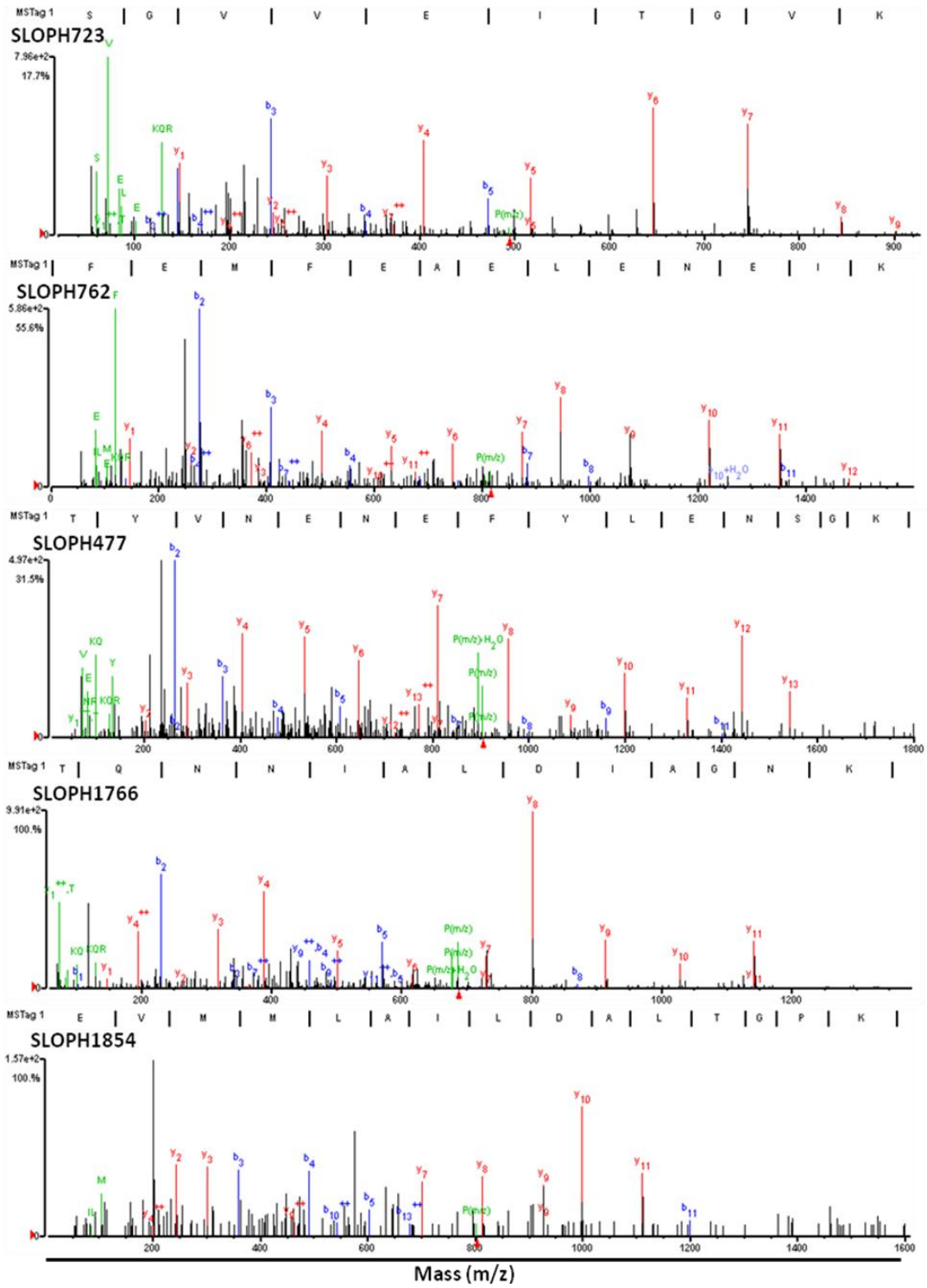


Figure 4.7| Mass spectra from 5 hypothetical proteins secreted during *S. lophii* germination.

An example of 5 spectra corresponding to a single peptide hit from each hypothetical protein present in all 3 replicates of *S. lophii* secretion proteomics (Table 4.3). Each histogram displays intensity (relative abundance) against mass to charge ratio (m/z) of b and y ions following peptide fractionation. The associated peptide sequences and *S. lophii* ID are listed above each spectrum.

Table 4.4| Secretion prediction properties of *S. lophii* secreted hypothetical proteins in other microsporidian species.

	SLOPH 1854		SlORF 1766		SlORF 2344		SlORF 1749		SlORF 691	
Species	SP	TP	SP	TP	SP	TP	SP	TP	SP	TP
E. cuniculi	●	●	●	●	●	●	●	●	●	●
E. intestinalis	●	●			○	○	●	●	●	●
E. hellem	●	●			●	●	●	●	●	
E. romaleae	●	●			●	●	●	●	○	●
N. ceranae	○	●			●	●	●	●	●	●
T. hominis	●	●	* ●	* ●	●	●	●	●	* ●	* ●
N. parisii					●	●	●	●		
V. corneae	●	○			●	●	●	○		
V. culicus	●	○	* ○	* ○	●	●	●	●	* ○	* ○
E. bieneusi	○	○			○	○	○	○	○	●

BlastP cutoff 1e-5 ● Secreted ○ Not Secreted * = paralogs present .SP = SignalP4.1, TP = TargetP1.1

4.4| Discussion

The role of secreted proteins in virulence in both extracellular and intracellular host-parasite systems is well characterised (Ravindran and Boothroyd, 2008, Bennuru *et al*, 2009, Lambertz *et al*, 2012). Post invasion, secreted and membrane proteins with extracellular domains of intracellular parasites have direct access to the host cell environment throughout the developmental lifecycle of the parasite. Microsporidia possess a unique infection mechanism catalysed by spore germination and expulsion of the polar tube and are known to induce several host cell changes to aid their intracellular development. For example, microsporidia are known to accumulate host mitochondria around the site of infection, presumably to aid the scavenging of host ATP, and disrupt the host cell cycle to optimise proliferation (Scanlon *et al*, 2000, Scanlon *et al*, 2004). Both microsporidian invasion and microsporidian-induced host cell manipulations have been characterised morphologically, however due to the lack of available tools, molecular understanding of these processes is currently lacking. Due to their ongoing access to the host cell environment, secreted proteins undoubtedly play a role in these processes. Bioinformatic prediction of microsporidian secreted proteins revealed a large range in predicted secretome size, from 84 proteins in *E. romaleae* to 257 in *Nematocida* sp. Ertm2, corresponding to approximately 5-10% of the overall proteome (Figure 4.1). Interestingly, no unique secretion motif, such as the 11 amino acid *Plasmodium* Export Element in *P. falciparum* or the RXLR effector motif in the genus *Phytophthora* was identified (Hiller *et al*, 2004, Schornack *et al*, 2009). Identification of one such motif would aid secretome identification in future sequenced species. In contrast to oomycete plant pathogens, such as *Phytophthora ramorum* which has acquired ~7.6% of its secretome by

horizontal gene transfer from fungi, only one case of HGT was identified in microsporidian secreted proteins (*E. cuniculi* GI: 449330242, see Chapter 6 for details) (Richards *et al*, 2011). The number of species-specific secreted proteins also varies considerably in microsporidian species, ranging from 0 in closely related species *E. romaleae* and *E. hellem* to 44 in *S. lophii*, thus reflecting the the close relationship of *Encephalitozoon* species (Figure 4.2). Future genome sequencing projects of closely related and xenoma-forming microsporidia will be required to identify the true extent of species-specific innovations in *S. lophii* and to determine the contribution of these secreted proteins to xenoma-formation. The role of the xenoma, a hypertrophic cell structure commonly formed by microsporidia infecting fish and crustaceans, is not known. Studies suggest the xenoma is either a parasite induced structure to evade host defences or a host structure deployed to isolate parasites and minimise the spread of infection (Lom and Dykova, 2005). Determining the proteomic composition of the xenoma wall will be required to identify its origin and the extent to which host and parasite factors contribute to its formation. Addressing this question is currently hindered by the absence of a sequenced *L. piscatorius* genome.

The predicted secretome of microsporidian parasites consists predominantly of proteins with no functional annotation. Due to the lack of a gene knockout or knock down system, microsporidian protein characterisation currently relies on heterologous expression systems in *Escherichia coli* and *Saccharomyces cerevisiae* which are time consuming, expensive, indirect and not applicable to large numbers of proteins (Campbell *et al*, 2013). In this study, using *S. lophii* as a model system to study germination, transcriptional and

proteomic analysis suggest, that there may be an important role for *S. lophii*- and microsporidian- specific secreted proteins during this invasive lifecycle stage. Despite the current lack of an *in vitro* culture system, *S. lophii* provides an excellent system for studying spore germination as large quantities of spores can be purified simply from cyst-like aggregations of xenomas. Interestingly, whole cell proteomic analysis identified no consistent variation between germinated and non-germinated spores. This suggests that spore germination occurs too rapidly for new protein translation and that microsporidia pre-package the proteins required for germination, host cell invasion and early development prior to this cellular event. To identify potential effector proteins secreted during germination, a complex mix proteomic analysis of the extracellular medium was employed in conjunction with *in vitro* germination of purified *S. lophii* spores. Mass spectrometry identified 37 proteins present in the extracellular medium across 3 biological replicates, of which 10 and 17 proteins were predicted to be secreted and targeted to the secretory pathway by SignalP4.1 and TargetP1.1 respectively (Table 4.3). A total of 15 proteins were identified in 2/3 or 3/3 biological replicates of which 5 were hypothetical proteins conserved in other microsporidia and 9 were hypothetical proteins encoded only in *S. lophii*. These effector proteins may perform *S. lophii* specific or conserved microsporidian functions required for host cell adhesion, invasion or the initial stages of intracellular development. No proteins were identified in the extracellular medium of non-germinated spores suggesting that release of these potential effector proteins is specific to the germination process. This study employs a novel method for identifying proteins secreted extracellularly during microsporidian spore germination, the lifecycle stage responsible for host cell infection. This analysis has identified a small subset of hypothetical proteins

from the thousands present in microsporidian genomes that can be characterised in future functional studies.

Chapter 5| Characterisation of *E. cuniculi* HlyIII: do pores mean prizes for an intracellular parasite?

5.1| Introduction

Hemolysins form a diverse group of prokaryotic and eukaryotic proteins defined by their ability to lyse erythrocytes *in vitro* (Nayak *et al*, 2013). Hemolysin proteins are well characterised in bacteria as important toxins and virulence factors deployed against the host by a range of pathogenic species (Goebel *et al*, 1988). Hemolysin proteins were originally classified as α, β, γ, δ, ε hemolysins based on studies in *Staphylococcus aureus* which identified variation in the mode of action and lytic phenotype produced by *S. aureus* pore-forming proteins (Wiseman, 1975, Ruzickova, 1994). *S. aureus* α -hemolysin (Hla), the first pore-forming bacterial exotoxin identified, has since been determined to play a crucial role in cytolytic induced virulence (Bhakdi and Tranum-Jensen, 1991). *S. aureus* Hla is required to induce necrotising pneumonia in mouse models and has been shown to induce inflammatory reactions responsible for many disease symptoms in the eye, skin and abdomen in other rodent models (Siqueira *et al*, 1997, Craven *et al*, 2009). Hla induced cytotoxicity is caused by both the induction of the host inflammatory response via activation of host caspase signaling leading to apoptosis and directly by heptameric pore formation in the host cell membrane upon binding an unidentified membrane receptor (Bantel *et al*, 2001). Similarly, *Escherichia coli* encodes a range of hemolysin proteins including HlyE, a gene which is present in all pathogenic *E. coli* strains identified to date but is absent from many non-pathogenic strains (Wallace *et al*, 2000). It is thought that HlyE, which is required for colonisation of the mammalian intestine, plays an important role in both bacterial iron acquisition and in promoting infection by inducing cytotoxic effects on immune cells, subsequently causing tissue damage (Green and Baldwin, 1997, del Castillo *et al*, 1997, Ludwig *et al*, 1999).

In contrast, eukaryotic hemolysins have been studied far less intensively. Hemolytic activity has been identified *in vitro* in many pathogens such as fungal species *Candida albicans*, *Candida glabrata* and *Candida dubliniensis*; however, a molecular dissection of this cytolytic mechanism or the resultant damage to the host is not known (Luo *et al*, 2001). Several fungal hemolysins have now been characterised at the molecular level. For example, the Asp-hemolysin of *Aspergillus fumigatus* is known to lyse erythrocytes and cause cytotoxicity in both human endothelial cells and murine macrophages (Ebina *et al*, 1985, Kumagai *et al*, 1999, Kumagai *et al*, 2001). Asp-hemolysin also induces increased production of proinflammatory cytokines such as interleukin-1 (IL-1) and tumor necrosis factor alpha (TNF- α) in both endothelial cells and macrophages. Hemolysins have also been implicated in fruiting intitiation in several fungal species such as the oyster mushroom *Pleurotus ostreatus* and the cocoa tree pathogen *Monoliophthora perniciosa* (Berne *et al*, 2007, Pires *et al*, 2009). For example, in *M. perniciosa* expression of MPRIA-1, a gene encoding an aegerolysin protein, is decreased during fungal stress but increased 4.3 fold in mycelial stages compared to primordia (Pires *et al*, 2009). This suggests hemolysin proteins in fungi are implicated in a range of diverse cellular functions and are also present in non-pathogenic species demonstrating the role of hemolysins does not universally lie in virulence. Similarly to fungi, many parasitic protists species have been shown to produce a hemolytic phenotype in *in vitro* lysis assays but the molecular mechanisms responsible for such a phenotype remain predominantly uncharacterised. For example, the human-infective parabasalid *Trichomonas vaginalis* is capable of erythrocyte lysis in all four human blood groups, whereas closely related cattle pathogen *Tritrichomonas foetus* could not lyse any human erythrocyte group,

demonstrating host specificity for hemolytic activity (De Carli *et al*, 1996). Pre-treatment of *T. vaginalis* with the lectin concanavalin A reduced hemolysis by up to 40%, suggesting that cytolytic activity is dependent on a cell surface-receptor interaction between the parasite and the erythrocyte (De Carli *et al*, 1996). The role of hemolysis in *T. vaginalis* infection or virulence *in vivo* is not currently known. A hemolysin protein has, however, been well characterised in the human-infective intracellular parasite *Trypanosoma cruzi*. The acid-active secreted hemolysin of *T. cruzi*, which functions optimally at pH 5.5, has been shown to lyse erythrocytes from humans and a range of animal species including dog, rat and sheep (Andrews and Whitlow, 1989). Studies have demonstrated that *T. cruzi* hemolysin is required for amastigote lysis of the acidic phagosome inside the host cell, a process which is inhibited at increased pH (Ley *et al*, 1990). A homologous acid-active hemolysin has been described in *Leishmania* (Noronha *et al*, 1994). This demonstrates that eukaryotic hemolysin proteins can have a direct role in pathogenesis and lifecycle progression.

The objective of this research chapter was to characterise a hemolysin III protein in the human-infective microsporidian species *Encephalitozoon cuniculi* (EcHlyIII, GenBank Accession: CAD25511), in both functional and evolutionary terms. EcHlyIII has an estimated size of 25 kDa and 7 predicted transmembrane domains (Bjellqvist *et al*, 1994, Krogh *et al*, 2001). EcHlyIII orthologs are widespread among microsporidian species and the gene represents the only hemolysin identified in microsporidian genomes to date. A phylogenetic analysis is conducted to determine the evolutionary relationship and taxonomic distribution of HlyIII orthologs and to identify any bias towards

the retention of orthologs in pathogenic or intracellular species. MEME motif elucidation software was also employed to identify any conserved sequence motifs in HlyIII orthologs which may have an important functional role in the hemolytic mechanism or in pore formation (Bailey *et al*, 2009). Fungal HlyIII orthologs are thought to function as zinc transporter proteins which play an important role in maintaining cellular zinc homeostasis and are induced through identified zinc response elements in the promoter region (Lyons *et al*, 2004). These fungal orthologs have also been implicated in fatty acid metabolism (Karpichev *et al*, 2002). To determine the functional relationship between microsporidian HlyIII proteins and fungal orthologs, termed Izh2 in *Saccharomyces cerevisiae*, a functional complementation was conducted under zinc stress conditions to determine if EcHlyIII can rescue Izh2 function in a *S. cerevisiae* Izh2 knock-out strain (Izh2Δ). To characterise EcHlyIII experimentally, the gene was cloned in a heterologous expression system, in *E. coli*, and tested for the ability to form pores in erythrocytes, using well described *in vitro* hemolysis assays (Baida and Kuzmin, 1996, Chen *et al*, 2004).

E. cuniculi proliferates inside a parasitophorous vacuole in the host cell cytoplasm in an organised manner, with spores located centrally and meronts present at the vacuole periphery (Scanlon *et al*, 2000). Previous studies have identified pores in the *E. cuniculi* vacuole membrane allowing the free transfer of metabolites and small proteins, of up to 10 kDa, between the intravacuolar space and host cell cytoplasm (Ronnebaumer *et al*, 2008). A possible role for EcHlyIII in vacuole pore formation, vacuole lysis or host cell membrane lysis to aid parasite egress, cannot be overlooked (Schottelius *et al*, 2000). The cellular localisation of EcHlyIII was also identified. Firstly, the protein was localised *in*

situ in *E. cuniculi* infected mouse kidney epithelial cell line, IMCD-3, using rabbit polyclonal antisera. Secondly, it was expressed and localised directly in IMCD-3 cells to determine any variation in targeting of the EcHlyIII between microsporidian and mammalian cells. In summary, this study investigated the evolutionary relationship of EcHlyIII with orthologs identified in other important pathogenic species and characterises the gene experimentally with particular reference to the potential role of EcHlyIII in *E. cuniculi* lifecycle development and virulence.

5.2| Methods

5.2.1| Phylogenetic analysis

E. cuniculi HlyIII (GenBank Accession: CAD25511) was used as a query sequence to examine the distribution of HlyIII orthologs across both prokaryotes and eukaryotes via protein BLAST searching with a pre-determined e-value of $1e^{-5}$. The following publically available genome databases were mined; National Centre for Biotechnology Information, The Joint Genome Institute and The Broad Institute Microsporidia Comparative Database. A total of 105 HlyIII sequences were retained after removal of redundant sequences and paralogs. The retained sequences were renamed into standard species-accession format using REFGEN/TREENAMER (Leonard *et al*, 2009). Sequences were aligned using MUSCLEv3.2 (Edgar, 2004) and masked manually in SEAVIEW to remove hypervariable sections of the sequence alignment. The resultant masked alignment contained 165 amino acids sites. Maximum likelihood analysis was conducted using PhyML3.0 (Guindon *et al*, 2010) under the Le and Gascuel (LG) substitution matrix (Le and Gascuel 2008) as deemed optimum by MODELGENERATOR (Keane, Creevey *et al.*2006), with an estimated gamma parameter and 1000 bootstrap replicates. To identify conserved sequence motifs in HlyIII orthologs, HlyIII sequences were submitted to Multiple Em for Motif Elucidation (MEME) in FASTA format. The following standard parameters were employed for protein motif elucidation; minimum motif length = 6 amino acids, maximum motif length = 50 amino acids and a maximum of 3 unique motifs per sequence. Output of identified motifs and graphics were downloaded and analysed with respect to the taxonomic and phylogenetic distribution of sequences.

5.2.2| Cloning of *E. cuniculi* HlyIII for expression in *E. coli*

The *E. cuniculi* HlyIII gene (654 base pairs (bp)) was amplified from *E. cuniculi* GB-M1 genomic DNA with Phusion high fidelity DNA polymerase (New England Biolabs) using gene specific primers EcHlyIIIF and EcHlyIIIR and an annealing temperature of 50^{o}C. *S. cerevisiae* Izh1 (951 bp) and Izh2 (954 bp) genes were amplified using primer sets ScIzh1F, ScIzh1R and ScIzh2F, ScIzh2R respectively from *S. cerevisiae* BY4741a genomic DNA (Euroscarf) under the same cyclic PCR conditions. PCR amplicons were analysed by electrophoresis, gel purified, digested at restriction sites XhoI and BamHI and ligated to pre-digested pET-14b vector (Novagen) as described in Chapter 2.5. HlyIII/pET-14b, Izh1/pET-14b, Izh2/pET-14b constructs were then transformed into *E. coli* XL1-Blue (Stratagene) as described in Chapter 2.6. Positive colonies were confirmed by PCR, restriction digest and sequencing (MWG Eurofins). Sequenced insert-bearing plasmids were purified and transformed to *E. coli* C3010I expression cells (Bioline) alongside an empty pET-14b plasmid to act as a negative control as described in Chapters 2.6 and 2.7 respectively.

5.2.3| Cloning of *E. cuniculi* HlyIII for expression in *S. cerevisiae* Izh2Δ

E. cuniculi HlyIII and *S. cerevisiae* Izh2 were amplified using primers EcHlyIIIfcF, EcHlyIIIfcR and ScIzh2fcF, ScIzh2fcR respectively from genomic DNA using Phusion high fidelity DNA polymerase (New England Biolabs) and a primer annealing temperature of 50^{o}C. PCR amplicons were analysed by electrophoresis, gel purified, digested at restriction sites HindIII and BamHI and ligated into pre-digested pYES2 vector (Invitrogen) as described in Chapter 2.5. *E. coli* XL-1 Blue cells were transformed with HlyIII/pYES2 and Izh2/pYES2 constructs as described in Chapter 2.6. Positive colonies were confirmed by

PCR, restriction digest and sequencing. Sequenced insert-bearing plasmids were then purified and transformed into *S. cerevisiae* Izh2Δ strain Y11693 (Euroscarf) as described in Chapters 2.6 and 2.8. Empty pYES2 was transformed into *S. cerevisiae* strains BY4741a and Y11693 to act as controls using the same protocol.

5.2.4| Cloning of *E. cuniculi* HlyIII for transient mammalian expression

E. cuniculi HlyIII was amplified using primers EcHlyIIIMaF and EcHlyIIIMaR from genomic DNA using Phusion high fidelity DNA polymerase (New England Biolabs) and a primer annealing temperature of 50^{o}C. The resultant PCR amplicon was analysed by electrophoresis and gel purified as described in Chapter 2.5. Purified PCR product was then ligated to pcDNA3.1D/V5-His-TOPO vector (Invitrogen) using 4 base TOPO overhang (cacc) following the manufacturer's instructions. HlyIII/pcDNA3.1D/V5-His-TOPO construct was then transformed into *E. coli* OneShot Top10 cells (Invitrogen) as described in Chapter 2.6. Positive colonies were confirmed by PCR, restriction digest (BamHI/XhoI) and sequencing. A sequence-confirmed insert-bearing plasmid was then transformed into mammalian cell line IMCD-3 as described in Chapter 2.10 alongside an empty pcDNA3.1D/V5-His-TOPO control.

5.2.5| Hemolysis assays

E. coli C3010I cells bearing plasmids HlyIII/pET-14b, Izh1/pET-14b, Izh2/pET-14b and pET-14b were grown overnight at 37^{o}C in 4 ml LB broth. Cell cultures were adjusted to OD 0.6 at 595 nm using sterile LB broth. 2 ml of normalised *E. coli* C3010I inoculum was then added to a blood broth medium solution containing 95 ml LB broth, 5 ml defibrinated horse blood (Fisher Scientific) and

0.5 mM IPTG. IPTG was required for induction of recombinant expression via the bacteriophage T7 promoter. Cultures were then incubated at 18^{o}C with shaking at 150 rpm. At regular time intervals (0, 2, 4, 6, 8, 12, 24 and 48 hours post induction of gene expression), 100 µl samples were taken from cultures and spun at 10,000 x g for 15 seconds. The supernatant was then added to a 96-well plate and the OD was calculated at 595 nm for 3 biological replicates and three technical replicates at each timepoint. Percentage hemolysis was calculated using the following equation in which A595 describes the absorbance determined at 595 nm for each sample; % *Hemolysis = (A595 SDS – A595 Sample)/ (A595 PBS – A595 Sample) x 100*. SDS was used as a positive control as the detergent produces 100% erythrocyte lysis. 1 x PBS was used as a negative control.

5.2.6| Reverse Transcription PCR (RT-PCR)

Semi-quantitative RT-PCR was used to determine consistent expression of recombinant proteins *E. cuniculi* HlyIII, *S. cerevisiae* Izh1 and *S. cerevisiae* Izh2 in *E. coli* C3010I cells. Expression was conducted as described in Chapter 2.11 with cells harvested 4 hours post IPTG induction. Harvested cells were resuspended in 2 ml PBS and flash frozen in liquid nitrogen. Cells were disrupted manually using a pestle and mortar. RNA was extracted using TRIzol (Invitrogen) as described in Chapter 2.3, treated with DNAse I (Invitrogen), quantified using a nanodrop spectrophotometer and normalised to 3 µg per sample. RNA was then reverse transcribed using M-MLV reverse transcriptase (Promega) in the following reaction; 3 µg RNA, 5 µl x 5 reverse transcriptase buffer, 5 µl DNTPS, 2 µl MMLV reverse transcriptase and 1.5 µl random primers (Promega) at 37^{o}C for 1.5 hours. Following reverse transcription 1.5 µl template

cDNA was used to amplify 150 bp internal fragments using the following gene specific primer sets: 1) EcHlyIIIIntF/EcHlyIIIIntR 2) ScIzh1IntF/ ScIzh1IntR 3) ScIzh2IntF/ScIzh2R. The same PCR was conducted for the *E. coli* housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GapA) for an expression control for each recombinant expression clone using primer set EcoGapAF/EcoGapAR.

5.2.7| Functional complementation of EcHlyIII in *S. cerevisiae* Izh2Δ

S. cerevisiae BY4741a containing pYES2 (+ve wild type control) and *S. cerevisiae* Izh2Δ Y11693 (Euroscarf) containing HlyIII/pYES2, Izh2/pYES2 and pYES2 (-ve control) constructs were grown for 24 hours in 5 ml complete synthetic media (CSM) broth lacking uracil (Ura) at 30^{o}C and 180 rpm. Cultures were then pelleted by centrifugation at 1,200 x g and washed twice in sterile PBS. Cells concentration was determined using a hemocytometer and all cultures were adjusted to a concentration of 1x10^{8} cells/ml. The standardised cultures were diluted serially and spotted onto CSM/-Ura/galactose control medium and CSM/-Ura/galactose/12 mM ZnCl stress medium manually using a replicator. Galactose was required for recombinant expression by induction of the pYES2 Gal1 promoter. Plates were incubated at 30^{o}C for 72 hours.

5.2.8| Recombinant expression and purification of *E. cuniculi* HlyIII

Four independent 4 ml starter cultures were inoculated with *E. coli* C3010I previously transformed with HlyIII/pET-14b construct and grown overnight in LB broth at 37^{o}C with shaking at 180 rpm. 2 ml of each starter culture was then inoculated into 100 ml LB medium and grown at 150 rpm at 18^{o}C, 25^{o}C, 30^{o}C and 37^{o}C to determine the optimum temperature for recombinant *E. cuniculi*

HlyIII expression. Cells were induced using IPTG at an OD600 nm of 0.4 as described in Chapter 2.11. Soluble and insoluble protein was harvested 4 hours post induction for samples cultures grown at 30^{o}C and 37^{o}C and 16 hours post induction for cultures grown at 18^{o}C and 25^{o}C, to compensate for slower bacterial growth as described in Chapter 2.11. Soluble and insoluble protein was quantified using a nanodrop spectrophotometer, normalised using 1 x PBS/ 8 M Urea (Soluble - 2.02 mg/ml, Insoluble – 4.6 mg/ml) and size fractionated using SDS-PAGE (20 µl per sample). Following electrophoresis, protein was transferred to a nitro-cellulose membrane by western blot as described in Chapter 2.12. Following blocking, the blot was treated with mouse anti-his (Invitrogen, Isotype – IgG_1 kappa) at a 1:1000 dilution, followed by goat anti-mouse IgG horse radish peroxidase conjugate secondary antibody a concentration of 1:2000, as described in Chapter 2.12. Following blot development using ECL, band intensities were compared, and the 18^{o}C (soluble fraction) was identified as the optimum expression temperature. Optimised recombinant expression and Ni^{2+} NTA affinity chromatography purification was then conducted at 18^{o}C in 1 L cultures as described in Chapter 2.11, purifying only proteins in the soluble protein fraction. The purity of each imidazole fraction was visualised by SDS-PAGE and Coomassie blue staining. Purified protein was subsequently sent to Eurogentec for production of anti-sera in two SPF rabbits (hosts #27 and #28) following the 28 day speedy polyclonal antibody production program. During the Eurogentec 28 day polyclonal antibody production package EcHlyIII immunizations (0.152 mg of purified EcHlyIII protein per injection) were administered in non-Freund adjuvant on days 0, 7, 10 and 18. Bleeds were taken from rabbit hosts on days 0 (pre-immune sera), 21 (production bleed) and 28 (final bleed) of the immunization program.

5.2.9| Determining anti-HlyIII serum specificity

Anti-serum was received from Eurogentec from two SPF rabbits (hosts #27 and #28). The specificity of the pre-immune serum and final bleed post immunisation serum from both rabbits was tested via western blot against protein from *E. cuniculi* spores, *E. cuniculi* infected IMCD-3 cells and from IMCD-3 cells, which served as a negative control. 40 µg of both *E. cuniculi* infected IMCD-3 protein and IMCD-3 protein was used in western blot analysis, however, due to limitations in purifying *E. cuniculi* spores only 12 µg of protein could be run. Western blots were conducted as described in Chapter 2.12 both at 1:100 dilutions for primary antibody treatment followed by a 1:2000 dilution of goat anti-rabbit IgG-HRP conjugated secondary antibody (Invitrogen).

5.2.10| Immuno-localisation of *E. cuniculi* HlyIII *in situ*

IMCD-3 cells grown in tissue culture flasks as described in Chapter 2.8 were infected with *E. cuniculi* GB-M1 spores (a generous gift from Martin Embley, University of Newcastle). Infection was monitored daily via microscopy until intracellular parasitophorous vacuoles were visible in IMCD-3 cells, signifying that *E. cuniculi* had developed from spores into the merogony stage of its lifecycle. Cells were then passaged as described in Chapter 2.9 and seeded on to sterile coverslips in 6-well tissue culture plates. Infected cells were grown on coverslips for 48-72 hours and subsequently fixed and permeabilised in 1:1 acetone:methanol as described in Chapter 2.13. Fixed slides were blocked and treated with pre-immune and final bleed serum from rabbit 27 at 1:100 dilutions. Cells were subsequently treated with Goat anti-rabbit IgG (H+L) HiLyte Fluor-488 antibody (Anaspec) at a 1:1000 dilution in the dark at room temperature for 1.5 hours. DAPI was added to cells at a concentration of 1µg/ml to stain nuclei

during the final 15 minutes of the secondary antibody incubation. Cells were mounted in Vectashield mounting medium and visualised using the Olympus IX-81 inverted fluorescence microscope. Captured images were converted to 8-bit image files and overlaid using Metamorph Image Software (Metamorph).

5.2.11| Transient expression and immuno-localisation of *E. cuniculi* HlyIII in IMCD-3 cells

IMCD-3 cells were passaged and seeded onto sterile coverslips in 6-well plates and grown in 2 ml Opti-MEM medium until they had reached between 70-90% confluence. Cells were transfected with 2 µg plasmid DNA of pcDNA3.1D/V5-His-TOPO and *E. cuniculi* HlyIII/pcDNA3.1D/V5-His-TOPO using Lipofectamine 2000. Cells were grown at 37^{o}C and 5% CO_2 for 48 hours to ensure transient expression under the constitutive cytomegalovirus (CMV) promoter and subsequently fixed/permeabilised with parafomaldehyde and Triton X-100 respectively, as described in Chapter 2.13. Cells were blocked and treated with mouse anti-his primary antibody at 1:1000 dilution for 1 hour at room temperature followed by secondary antibody treatment with Goat anti-mouse IgG_1 (γ_1) Alexa Fluor-568 (Invitrogen) for 1.5 hours. Cells were then mounted and visualised using Olympus IX-81 microscope.

5.2.12| List of oligonucleotides

Table 5.1| Names and sequences of oligonucleotides used in this study.

Oligo name	Sequence (5'- 3')	Restriction site
EcHlyIIIF	gcat**ctcgag**atggcaaaacagaaacctatgctt	XhoI
EcHlyIIIR	cgta**ggatcc**ttatatttccagatagggtagagat	BamHI
ScIzh1F	cgat**ctcgag**atgagtatcaccactactaga	XhoI
ScIzh1R	cgat**ggatcc**tcatcaagggtggatatgaga	BamHI
ScIzh2F	cgat**ctcgag**atgtcaactttattagaaagg	XhoI
ScIzh2R	cgat**ggatcc**atcctaggagacaatcccgtt	BamHI
EcHlyIIIfcF	cgat**aagctt**atggcaaaacagaaacctatgctt	HindIII
EcHlyIIIfcR	ggat**ggatcc**ttattatttccagatagggtagagat	BamHI
ScIzh2fcF	cgat**aagctt**atgtcaactttattagaaagg	HindIII
ScIzh2fcR	ggat**ggatcc**atcctaggagacaatcccgtt	BamHI
EcHlyIIIMaF	**cacc**atggcaaaacagaaacctatgctt	TOPO cloning overhang
EcHlyIIIMaR	ttatatttccagatagggtagagatata	Blunt end TOPO cloning
EcHlyIIIIntF	ctggaccacatctccatatttcttctc	Internal
EcHlyIIIIntR	ttcatcaggatgatctttagaatacca	Internal
ScIzh1IntF	gcaacactctgaaaagcaaagcaattt	Internal
ScIzh1IntR	actaaagtaacaatcgtgaaaagactg	Internal
ScIzh2IntF	taagaattgctaccttaggaaataagt	Internal
ScIzh2IntR	cgatcccaaagctaacggtaataagcg	Internal
EcoGapAF	ccaggacatcgtttccaacgcttcctg	Internal
EcoGapAR	gagacgggccatcaacggttttctgag	Internal

Restriction site sequences for plasmid integration are displayed in bold.

5.3| Results

5.3.1| Evolutionary diversity of EcHlyIII orthologs.

Database mining with BLASTp and tBLASTn sequence homology searches retrieved 105 EcHlyIII orthologs (<$1e^{-5}$) after removal of long-branches, redundant and paralogous sequences. This included sequences belonging to members of evolutionarily diverse groups such as Microsporidia, Apicomplexa and Bacteria, overall demonstrating a sporadic taxonomic distribution for EcHlyIII orthologs (Figure 5.1). Maximum likelihood analysis of HlyIII orthologs produced a phylogenetic tree consisting of two distinct clades with strong bootstrap support (Figure 5.1). Firstly, a clade containing the Bacteria, Microsporidia, Nematoda, Chromalveolata and Apicomplexa and, secondly, a fungal clade in which members have been implicated in zinc transport and homeostasis (Lyons *et al*, 2004). This contradicts current understanding and recent phylogenetic studies which place microsporidia either as an early branching fungal phylum or as a sister group to the fungi (Capella-Gutierrez *et al*, 2012, Cuomo *et al*, 2012). Phylogenetic analysis provides no evidence that the gene has spread between taxonomic groups or been acquired by microsporidia via horizontal gene transfer (HGT). Several bacterial hemolysins have been acquired by HGT such as the thermostable direct hemolysin in various strains of species *Vibrio parahemolyticus*, *Vibrio cholerae* and *Vibrio mimicus*, suggesting hemolysin family proteins are amenable to HGT in prokaryotes (Nishibuchi *et al*, 1996). No such examples have been described in eukaryotes. It is possible, due to the highly divergent nature of HlyIII orthologs that the gene is present in other eukaryotic groups, but is too divergent to be identified. The relationship of HlyIII orthologs included in phylogenetic analysis is likely driven by convergent evolution and, in the case of intracellular

microsporidia and apicomplexa, may be caused by a common selective pressure from the host cell environment.

MEME motif elucidation analysis of the 105 HlyIII protein sequences identified 3 conserved sequence motifs (Figure 5.2A), which display a varied distribution and relative location on the protein across different taxonomic groups (Figure 5.2B) (Bailey *et al*, 2009). These sequence motifs range in size from 16-50 amino acids and contain several amino acid sites conserved across all taxonomic groups such as; Motif 1: Histidine (H-site 4 and 6), Phenylalanine (F-5) Motif 2: Aspartic acid (D-2), Isoleucine (I-6,10) and Motif 3: Tryptophan (W-5,18), Alanine (A-8), Glycine (G-11,34), Proline (P-22) (Figure 5.2B). These conserved sites maybe required to maintain the functional integrity of the protein. Active functional protein motifs have been described previously in hemolysin proteins, for example, the N-terminal region (amino acids 1-31) is required for erythrocyte binding in *V. parahemolyticus* thermostable direct hemolysin. This essentiality was demonstrated through the use of neutralising monoclonal antibodies (Tang *et al*, 1997). The motif distribution identified in microsporidian species (Motif order from N-terminus 2-3-1), with the exception of *Trachipleistophora hominis* and *Vavraia culicus floridensis*, is shared only with bacteria and nematode species *Caenorhabditis remanei*. Fungi, chromalveolates, apicomplexans and microsporidian species *T. hominis* and *V. culicus floridensis*, in contrast, only contain 2/3 conserved domains (Motif order from N-terminus 2-1), lacking the central 50 amino acid Motif 3. Such variation in the presence and distribution of sequence motifs in HlyIII orthologs will likely translate into significant functional variation. It is possible that the distribution Motif 2-Motif 3-Motif 1, which is found in bacterial species *Vibrio vulnificus* and

Bacillus cereus for which HlyIII has been characterised as a pore forming protein, is required for hemolytic activity (Baida and Kuzmin, 1996, Chen *et al*, 2004).

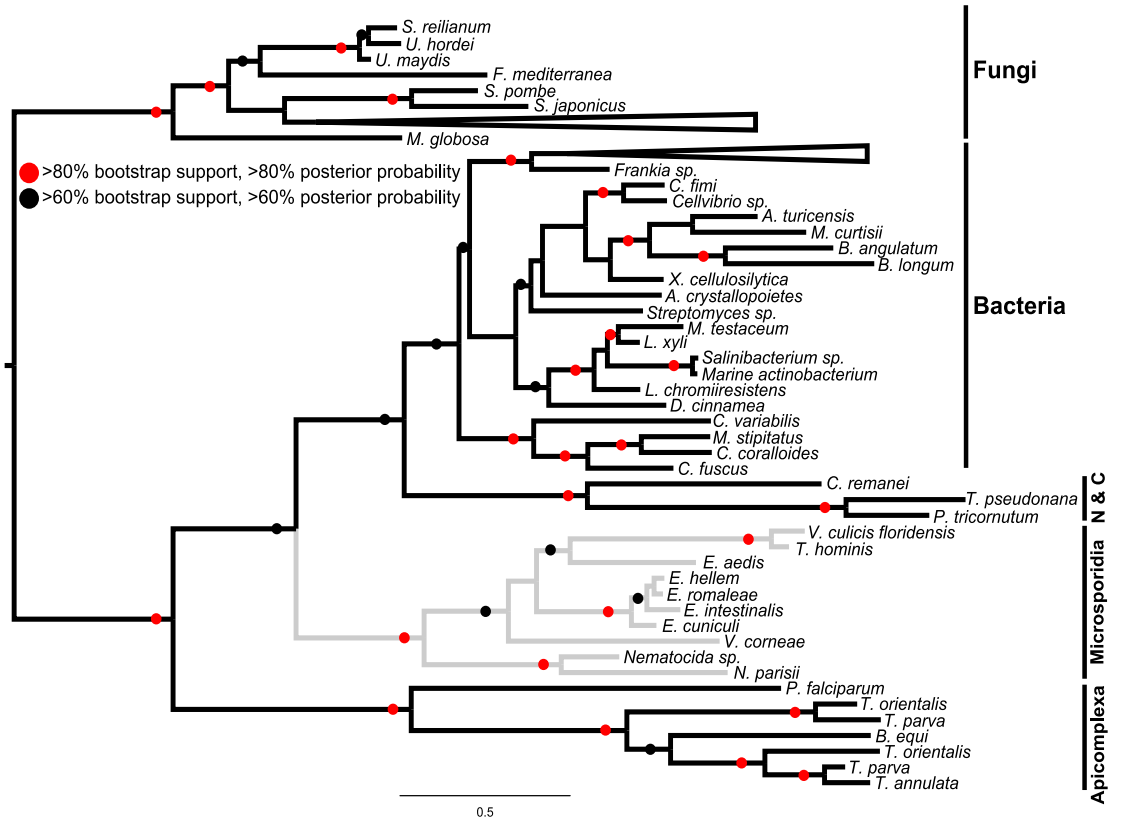


Figure 5.1| Phylogenetic analysis of EcHlyIII.

Maximum likelihood phylogeny of EcHlyIII and 104 orthologs identified by BLAST analysis. The tree demonstrates the evolutionary relationship of EcHlyIII to orthologs in both microsporidia and more divergent taxonomic groups such as the Apicomplexa and Bacteria. The tree was constructed in PhyML3.0 (Guindon *et al*, 2010) from 165 amino acid sites and 1000 bootstrap replicates using the LG amino acid substitution model (Le and Gascuel, 2008). Microsporidian species are highlighted by grey branches, N & C represent the Nematoda and Chromalveolata.

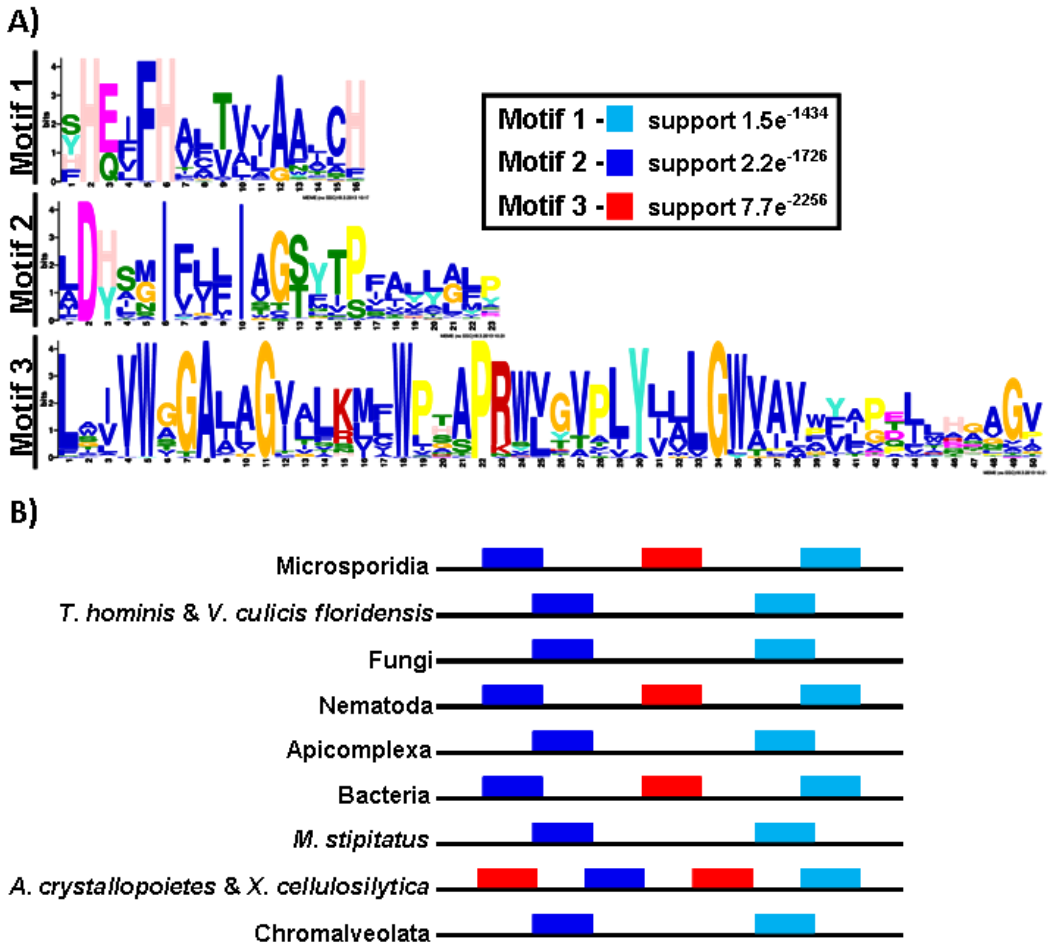


Figure 5.2| Sequence motif identification and distribution in 105 hemolysin III orthologs.

A) Sequences and e-value support for 3 motifs identified by MEME motif elucidation analysis of the 105 hemolysin III orthologs analysed (Bailey *et al*, 2009). B) Relative distribution of the 3 conserved sequence motifs in the taxonomic groups included in phylogenetic analysis. The motif distribution identified in microsporidia is present only in the phylum Nematoda and in Bacteria.

5.3.2| Generation of EcHlyIII, *S. cerevisiae* Izh1 and *S. cerevisiae* Izh2 expression constructs

EcHlyIII, *S. cerevisiae* Izh1 and Izh2 genes were amplified by PCR from genomic DNA as described in Chapter 5.2.3. The sizes of PCR amplicons were verified on TAE agarose gels (Figure 5.3). PCR bands for all genes, namely EcHlyIII 657 bp, *S. cerevisiae* Izh1 951 bp and *S. cerevisiae* Izh2 954 bp, matched the expected gene size, demonstrating the specificity of the oligonucleotides used for PCR amplification (Figure 5.3). PCR products were purified and cloned as described in Chapter 5.2.3. Clones used for functional analyses were checked for insert presence through size verification of PCR amplicons (colony PCR) and drop out fragments (restriction digest) (Figure 5.4). All PCR and restriction digest sizes for EcHlyII, *S. cerevisiae* Izh1 and Izh2 corresponded to the expected size of the gene (Figure 5.4).

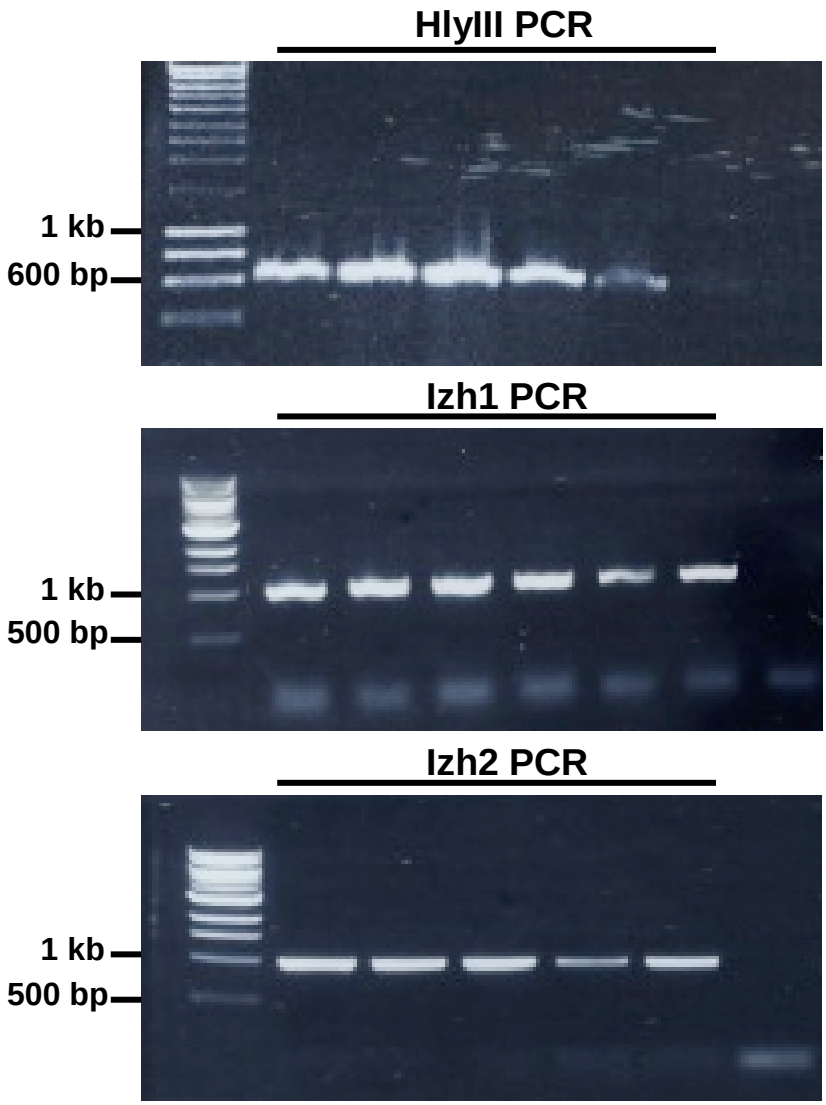


Figure 5.3| PCR amplification of EcHlyIII and *S. cerevisiae* genes Izh1 and Izh2.

PCR amplification of EcHlyIII (657 bp), *S. cerevisiae* Izh1 (951 bp) and *S. cerevisiae* Izh2 (954 bp) with gene specific primer sets EcHlyIIIF/R, ScIzh1F/R and ScIzh2F/R for digestion and ligation into pET14b at restriction sites XhoI and BamHI. Size and specificity of amplicons was verified on 1% TAE agarose gels.

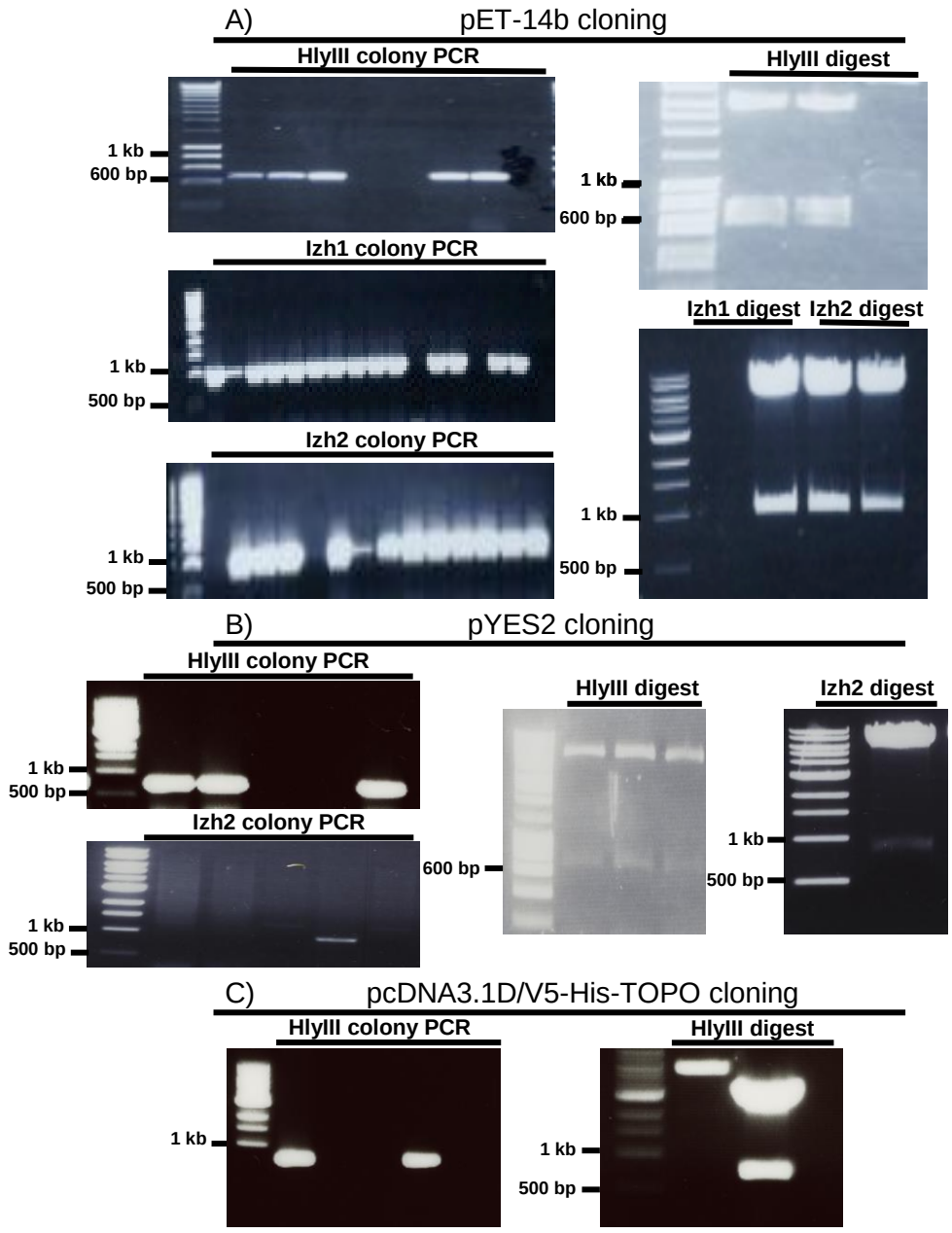


Figure 5.4| Colony PCR and restriction digest verification of EcHlyIII, *S. cerevisiae* Izh1 and *S. cerevisiae* Izh2 constructs.

A) Colony PCR and restriction digest of constructs EcHlyIII/pET14b, ScIzh1/pET14b and ScIzh2/pET14b with enzymes XhoI and BamHI. Sizes of gene products digested from cloning vector were verified on 1% TAE agarose gels. B) Colony PCR and digest of constructs EcHlyIII/pYES2 and ScIzh2/pYES2 with restriction enzymes HindIII and BamHI. C) Colony PCR and digest of construct EcHlyIII/pcDNA3.1D/V5-His-TOPO with restriction enzymes BamHI and XhoI.

5.3.3| Is EcHlyIII a divergent zinc transporter?

Database mining and phylogenetic analysis identified that EcHlyIII shared sequence similarity with a fungal zinc transporter protein, termed Izh2 in *S. cerevisiae*. Surprisingly, microsporidian HlyIII sequences do not form a sister group to the fungal Izh2 clade, as would have been expected, given the well supported evolutionary relationship between microsporidia and fungi. To determine the ability of EcHlyIII to act as a functional zinc transporter or in zinc homeostasis, a complementation assay was conducted in *S. cerevisiae* Izh2Δ using 12 mM ZnCl as an inducer of phenotypic zinc stress (Figure 5.5) (Lyons *et al*, 2004). Interestingly, gene expression studies suggest Izh2 may also function in the regulation of lipid and phosphate metabolism, however the exact role of the gene in these processes is not known (Karpichev *et al*, 2002).

The growth based assay included wild type *S. cerevisiae* BY4741a transformed with pYES2, *S. cerevisiae* Izh2Δ complemented with the native Izh2 gene, *S. cerevisiae* Izh2Δ transformed with EcHlyIII and *S. cerevisiae* Izh2Δ harboring an empty pYES2 vector, which served as a negative control. All *S. cerevisiae* strains grow equally well on Complete Synthetic Media (CSM) lacking uracil and containing galactose as carbon source to induce heterologous gene expression via the Gal1 promoter (West *et al*, 1984). *S. cerevisiae* BY4741a and Izh2Δ control strains grow as expected under zinc stress showing no growth defect and significantly reduced growth respectively (Figure 5.5). The reduced growth phenotype of *S. cerevisiae* Izh2Δ strain is caused by activation of zinc associated cell death pathways leading to activation of cellular endonucleases and apoptosis (Sunderman, 1995). Expression of native Izh2 in *S. cerevisiae* Izh2Δ background strain rescues the

zinc induced reduced growth phenotype present in *S. cerevisiae* Izh2Δ, and in fact, grows more efficiently under zinc stress than wild type strain BY4741a. It is probable that this improved growth is because the expression level of Izh2Δ achieved by induction of over expression promoter Gal1 is greater than achieved in wild type *S. cerevisiae* BY4741a strain in response to zinc (Figure 5.5). In contrast, expression of EcHlyIII does not rescue the zinc induced growth defect in *S. cerevisiae* Izh2Δ, suggesting EcHlyIII is not a functional zinc transporter and cannot functionally complement this role in *S. cerevisiae* or that the protein has not been folded in correctly and is not functional in this heterologous expression system. No zinc response elements such as ACCTTTAGGGT and TCCTCTAGGGT which are present upstream of *S. cerevisiae* Izh2 gene were identified within the 1000 bases upstream of the EcHlyIII start codon, providing no evidence that zinc drives expression of the gene (Lyons *et al*, 2004).

Taken together, these data suggest that despite a reasonable degree of sequence similarity, EcHlyIII may have evolved a novel functional role since the microsporidian-fungal divergence, perhaps as the zinc transport activity present in the ancestral gene is redundant in the intracellular environment. Analysis of EcHlyIII expression levels and sub-cellular localisation of the protein in *S. cerevisiae* will be required to determine how the protein is targeted and to fully understand the functional relationship between EcHlyIII and *S. cerevisiae* Izh2. It is possible that this functional variation was driven by the intimate *E. cuniculi* interaction with the host cell or by the change in nutritional composition upon invasion of the intracellular environment.

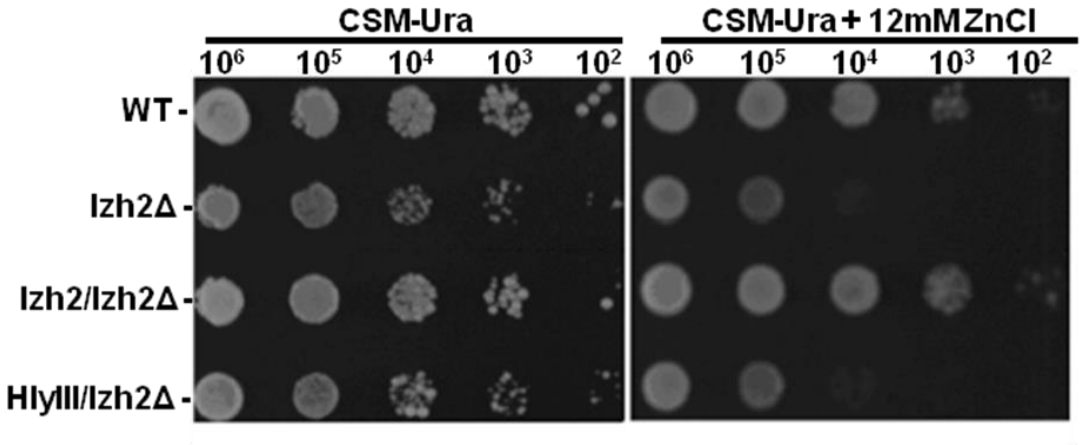


Figure 5.5| Functional complementation of EcHlyIII in *S. cerevisiae* Izh2Δ. *S. cerevisiae* wild type strain BY4741a (WT), Izh2Δ strain Y11693, Izh2Δ complemented with the native Izh2 gene and Izh2Δ complemented with EcHlyIII were grown for 72 hours on both standard Complete Sythetic Media (CSM) without uracil (Ura) containing galactose and under zinc stress (Lyons *et al*, 2004). All strains grow consistently on standard CSM-Ura. Under zinc stress Izh2Δ shows a clear growth defect, EchlyIII cannot complement Izh2 function.

5.3.4| EcHlyIII is capable of erythrocyte lysis *in vitro*.

Recombinant expression of EcHlyIII, *S. cerevisiae* Izh1 and *S. cerevisiae* Izh2 genes in *E. coli* C3010I cells was conducted to determine their hemolytic potential *in vitro*, using pET14b as negative control. EcHlyIII shares both sequence similarity and a conserved secondary structure (7 transmembrane domains) with *S. cerevisiae* Izh2 (identity = 31% BLASTp score = $1.6e^{-12}$). EcHlyIII shares less sequence similarity with Izh2 paralog Izh1 (identity = 25%, BLASTp score = $1.9e^{-4}$). Paralogous zinc transporter proteins Izh1 and Izh2 were included in hemolysis assays as controls to determine if cell lysis/pore formation is a non-specific event caused by contact of the 7 transmembrane domains with a foreign membrane or receptor, rather than a reflection of true protein function.

Hemolysis assays demonstrate a significant difference in hemolytic activity between EcHlyIII and *S. cerevisiae* Izh transporters at all time points except 2 hours post induction (HlyIII/Izh2 p-values: 2h = 0.52, 4h = 1.32^{-3}, 6h = 6.1^{-7}, 8h = 3.8^{-4}, 12h = 7.31^{-7}, 24h = 7.53^{-6}, 48h =1.81^{-4}), with lysis quantified at 75.8% (EcHlyIII), 20.6% (ScIzh1), 24.3% (ScIzh2) and 13.4% (pET14b) after 48 hours expression (Figure 5.6). This consolidates the complementation study (Figure 5.5) and suggests EcHlyIII and *S.cerevisiae* Izh2 do not occupy the same functional role and that erythrocyte lysis *in vitro* is not a coincidental result of contact with a zinc transporter with 7 transmembrane domains. No erythrocyte lysis was produced by purified EcHlyIII or with crude protein extracts from *E. coli* expressing EcHlyIII. Similarly, culture supernatants from *Bartonella bacilliformis*, a bacterial species known to cause severe anaemia through hemolysis, were unable to lyse erythrocytes despite clear hemolytic activity

when bacterial cells were applied directly onto red blood cells. It is suggested that *B*. *bacilliformis* hemolysis is a contact dependent process, with structural integrity and cellular localisation required for hemolytic function (De Carli *et al*, 1996, Hendrix, 2000). *In vitro* hemolytic activity has been described in two EcHlyIII orthologs from human-infective bacterial species *B. cereus* and *V. vulnificus* (Baida and Kuzmin, 1996, Chen *et al,* 2004). These studies demonstrate that HlyIII is a pore forming protein, producing a pore diameter of ~3 nm and that pore formation is a temperature dependent process, with an optimum of 37^{o}C. Interestingly, a *V. vulnificus* HlyIII mutant showed attenuated virulence in mice with a median lethal dose (LD50) of 8 x 10^{6} cells compared to 5 x 10^{5} cells in the wild-type (Chen *et al*, 2004). This temperature dependent lysis is not evident in EcHlyIII assays, in which the protein is capable of efficient erythrocyte lysis at 18^{o}C. Future studies such as site directed mutagenesis will be required to determine the functional significance of the three sequence motifs identified in HlyIII orthologs with relation to the pore forming activity identified (Figure 5.2). Hemolytic orthologs, EcHlyIII, *B. cereus* HlyIII and *V. vulnificus* HlyIII share a conserved presence and distribution of the three sequence motifs. In contrast *S. cerevisiae* Izh2 and other fungal orthologs lack centrally located sequence Motif 3 (Figure 5.2).

Semi-quantitative RT-PCR was conducted to identify any variation in heterologous gene expression levels of EcHlyIII, ScIzh1 and ScIzh2. RNA was extracted from *E. coli* four hours post expression induction under identical experimental conditions as where employed in hemolysis assays. There was no identifiable difference in recombinant expression levels of EcHlyIII, Izh1 and

Izh2 suggesting that the > 3-fold difference in lytic phenotype was not a consequence of varying gene expression levels (Figure 5.7).

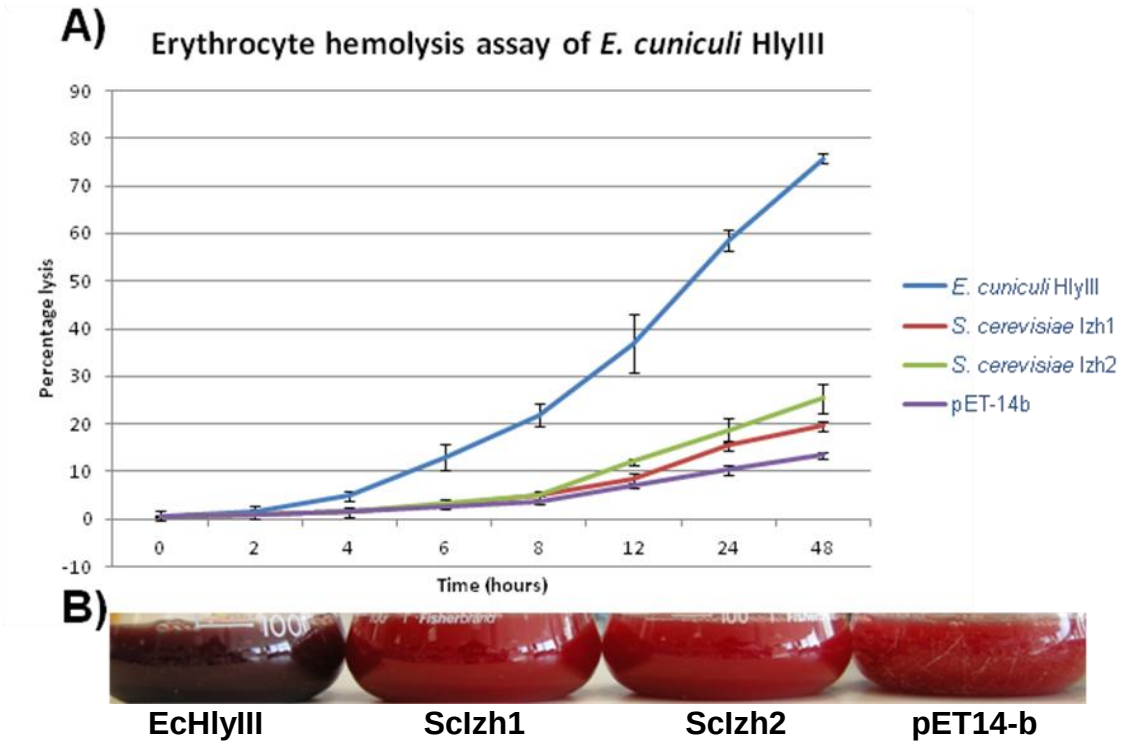


Figure 5.6| Hemolytic activity of EcHlyIII *in vitro*.

A) Hemolysis assay determining the percentage of erythrocyte lysis as a consequence of expression of EcHlyIII and *S. cerevisiae* genes Izh1 and Izh2 expressed in *E. coli* over 48 hours in LB medium containing 5% defibrinated horse blood. IPTG was added to medium at T0 to a final concentration of 0.5 mM. *E. coli* transformed with empty expression vector pET-14-b served as a negative control. B) Example of cell lysis induced by EcHlyIII expression in *E.coli* C3010I at 12 hours post expression induction. There was no identifiable lysis in *E.coli* C3010I expressing *S.cerevisiae* Izh1 (ScIzh1), Izh2 (ScIzh2) or pET14b negative control.

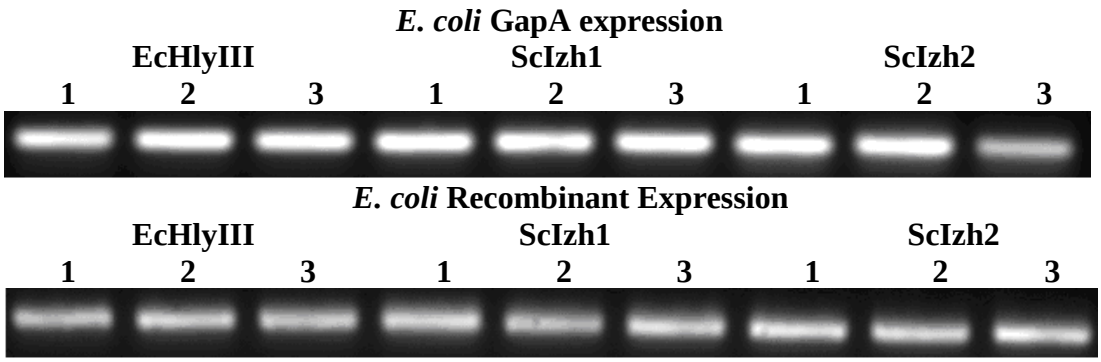


Figure 5.7| Recombinant expression of genes expressed in hemolysis assay.

Recombinant expression levels of 150 base pair internal fragments of *E. coli* housekeeping gene (GapA) and plasmid expressed genes EcHlyIII and *S. cerevisiae* Izh1 (ScIzh1) and Izh2 (ScIzh2) as determined by RT-PCR. PCRs were carried out on normalised cDNA using gene specific primers and run on 1% TAE agarose. No bands were evident in negative controls. RT-PCR demonstrates no identifiable difference in the expression levels of EcHlyIII, ScIzh1 or ScIzh2 suggesting variation in expression level is not responsible for variation in lytic phenotype.

5.3.5| EcHlyIII displays a nuclear localisation both *in situ* and in mammalian cells.

To identify the optimum expression temperature of EcHlyIII in *E. coli*, soluble and insoluble protein was extracted from *E. coli* expressing HlyIII at temperatures ranging from 18-37^{o}C. Proteins extracts were screened by western blot against the N-terminal his-tag, demonstrating optimum expression of soluble EcHlyIII at 18^{o}C (Figure 5.8A). Recombinant expression of eukaryotic proteins in *E. coli* at temperatures lower than the standard optimal growth temperature of 37^{o}C has been shown to increase expression levels and solubility, reduce inclusion body formation and improve functionality (Sahdev *et al*, 2008, Bill *et al*, 2011).

EcHlyIII was purified via nickel affinity chromatography from 1 L *E. coli* cultures grown at 18^{o}C. Crude soluble protein extract was applied to Ni^{2+} charged agarose resin, followed by an imidazole gradient wash to remove non-specific protein via competitive binding of Ni^{2+} (Figure 5.8B) (Bornhorst and Falke, 2000). Purified EcHlyIII was identified in the 250 mM imidazole fraction by SDS-PAGE and verified by western blot. Purified EcHlyIII antigen was subsequently sent to Eurogentec for polyclonal antibody production in rabbit. The pre-immune serum and resultant HlyIII anti-serum was screened by western blot against protein extracted from *E. cuniculi* infected IMCD-3 cells, *E. cuniculi* spores and IMCD-3 cells (Figure 5.9). Some non-specific binding was evident in pre-immune serum western blot, particularly against protein extracted from IMCD-3 cells. Pre-immune serum showed little activity against *E. cuniculi* protein antigens either from intracellular parasitic stages or from spores, suggesting that the individual had never built an antibody response to an *E. cuniculi* antigen prior to challenge with EcHlyIII. This is of particular significance

as *E. cuniculi* is a natural pathogen of rabbits and infection in laboratory rabbit populations is common (Cox *et al*, 1980). Final-bleed serum containing polyclonal antibodies against a protein of ~25 kDa, the expected size of the HlyIII antigen, was identified in *E. cuniculi* infected IMCD-3 cells. As expected no antibody response was identified against IMCD-3 proteins. Interestingly, no response was evident against protein extracted from *E. cuniculi* spores, suggesting that EcHlyIII is not expressed in extracellular stages or that quantity of HlyIII present in 12 µg total protein is below the limit of detection. It may be possible that EcHlyIII is required functionally only during intracellular lifecycle stages. Final bleed serum also produced an unexpected band of ~40 kDa against a protein from *E. cuniculi* infected IMCD-3 cells and *E. cuniculi* spores demonstrating that this polyclonal anti-serum is not completely specific to EcHlyIII. Despite the presence of EcHlyIII as the prominent band in western blot analysis of anti-serum against *E. cuniculi* infected IMCD-3 protein, non-specific serum activity was identified in all western blot analyses (Figure 5.9). This included testing serum in both BSA and milk based blocking solutions and testing at a reduced 1:500 serum concentration. Figure 5.9 represents the optimum western blot conditions identified. It is possible this 40 kDa band identified in Figure 5.9 represents dimers formed by EcHlyIII and incomplete denaturation during SDS-PAGE analysis.

To determine the cellular localisation of EcHlyIII *in situ*, *E. cuniculi* infected IMCD-3 cell monolayers were treated with anti-EcHlyIII polyclonal antibody from final bleed serum followed by Goat anti-rabbit IgG HiLyte Fluor-488 conjugate secondary antibody. Immuno-fluorescence microscopy revealed a specific green fluorescent signal located within *E. cuniculi* meronts (Figure 5.10). Various co-localisation studies were conducted using previously

characterised microsporidian polyclonal antibodies including a vacuolar branching network protein and a spore wall protein (red) (a generous gift from Dr Kaya Ghosh, Albert Einstein College of Medicine) (Figure 5.10, Ghosh *et al*, 2011). EcHlyIII did not co-localise with either marker. *E. cuniculi* HlyIII did however co-localise with the nuclear cell stain DAPI, suggesting that EcHlyIII localises to the parasite nucleus, or based on the predicted transmembrane structure of the protein, the nuclear membrane. As a negative control *E. cuniculi* infected IMCD-3 cells were treated with pre-immune serum and visualised with immuno-fluorescence microscopy. No specific fluorescent signal was identified within either *E. cuniculi* meronts or IMCD-3 cells, demonstrating the absence of antibodies against *E. cuniculi* (Figure 5.11). To investigate the localisation of EcHlyIII in a heterologous expression system, the gene was expressed directly in mouse kidney epithelial cell line IMCD-3. EcHlyIII was expressed transiently over 48 hours under the constitutive CMV promoter and geneticin selection. Immuno-localisation was conducted treating cells with mouse anti-his monoclonal antibody to directly target the C-terminal polyhistidine tag expressed with EcHlyIII, followed by Goat anti-mouse IgG Alexa Fluor-568 secondary antibody. In agreement with *in situ* localisation, EcHlyIII localised to the nucleus and more specifically appears to associate with the nuclear membrane in IMCD-3 cells (red), as confirmed by DAPI co-localisation (blue) (Figure 5.12).

To determine any non-specific binding of primary or secondary antibody to mouse kidney epithelial cells, IMCD-3 cells were transfected with empty vector pcDNA3.1D/V5-His-TOPO and subjected to antibody treatment. Immuno-fluorescence identified no consistent signal in empty vector control IMCD-3 cells (red), suggesting nuclear signal is specific to the C-terminal his-tag of EcHlyIII

(Figure 5.13). This confirms EcHlyIII localises to the nucleus both *in situ* and in a mammalian expression system using two distinct antibody treatments. This nuclear localisation is novel for studied hemolysin proteins, which are predominantly secreted extracellularly or localise to the cell membrane. Interestingly, no nuclear retention signal, such as the monopartite nuclear localisation signal consensus sequence: K-K/R-X-K/R or the non-classical nuclear localisation signal motif: PY-NLS, required for nuclear transport could be identified in EcHlyIII (Dingwall *et al*, 1988, Lee *et al*, 2006). The nuclear localisation of EcHlyIII in both *E. cuniculi* and *Mus musculus* demonstrates conservation of EcHlyIII protein targeting between microsporidia and mammals. This localisation suggests that despite displaying hemolytic activity *in vitro* when expressed in *E. coli*, EcHlyIII is not involved in vacuolar pore formation, virulence or cell egress but instead has a non-canonical or novel function. Because of the presence of an unidentified ~40 kDa protein band in EcHlyIII antibody western blot, a nuclear localisation cannot be confirmed conclusively. However, nuclear localisation *in situ* coupled with nuclear localisation when expressed in a heterologous system, suggests that the nuclear localisation identified for EcHlyIII is accurate. Due to the lack of complete specificity of anti-serum against EcHlyIII, a nuclear localisation cannot, however, be confirmed. The lack of the contaminating 40 kDa band in pre-immune screen suggests it may be linked directly to challenge with the EcHlyIII protein antigen.

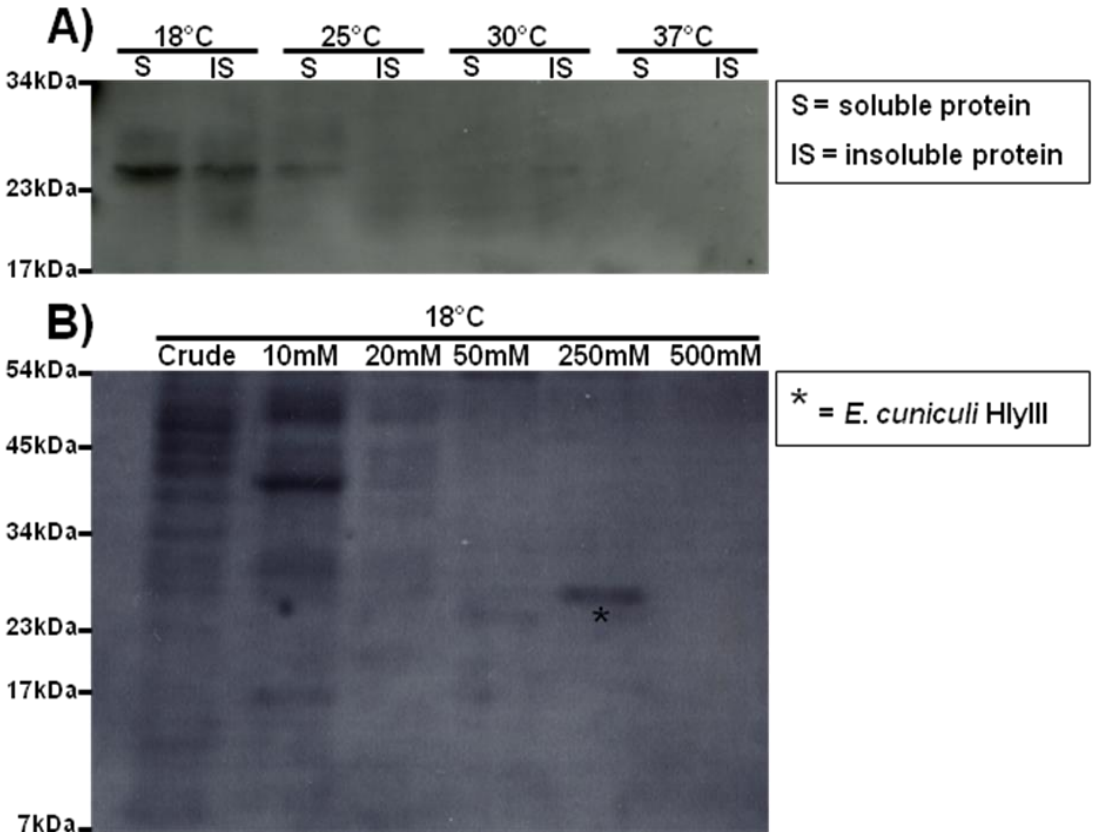


Figure 5.8| Recombinant expression and purification of EcHlyIII.

A) Anti-his western blot of soluble and insoluble protein extracts from *E. coli* expressing EcHlyIII at a temperature range of 18-37^{o}C. EcHlyIII expression in *E. coli* occurs optimallyat 18^{o}C. B) Imidazole elution of purified EcHlyIII following nickel affinity chromatography. Imidazole fractions and crude protein extracts from *E.coli* C3010I expressing EcHlyIII were run on a 12% SDS gel and stained with Coomassie blue.

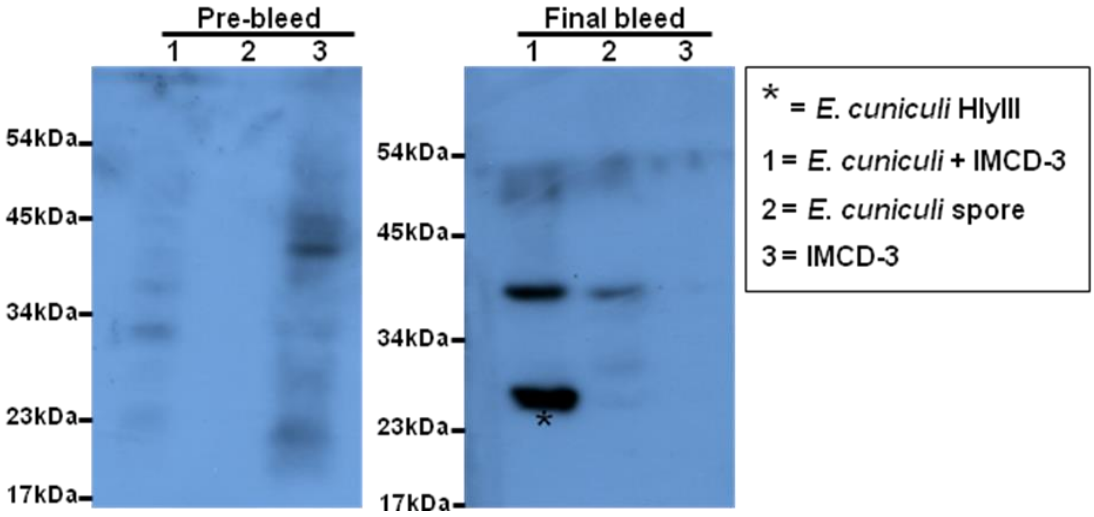


Figure 5.9| Anti-serum specificity following anti-EcHlyIII polyclonal antibody production.

Western blots of whole protein extracts from *E. cuniculi* infected IMCD-3 cells, purified *E. cuniculi* spores and from IMCD-3 cells using pre-immune serum and final production bleed serum as primary treatment. A strong expression band corresponding to ~25kDa, the predicted size of EcHlyIII, was identified in *E. cuniculi* infected IMCD-3 cells. This may indicate that EcHlyIII is not expressed in extracellular spores. Non-specific binding was also evident in both pre- and final bleed serum.

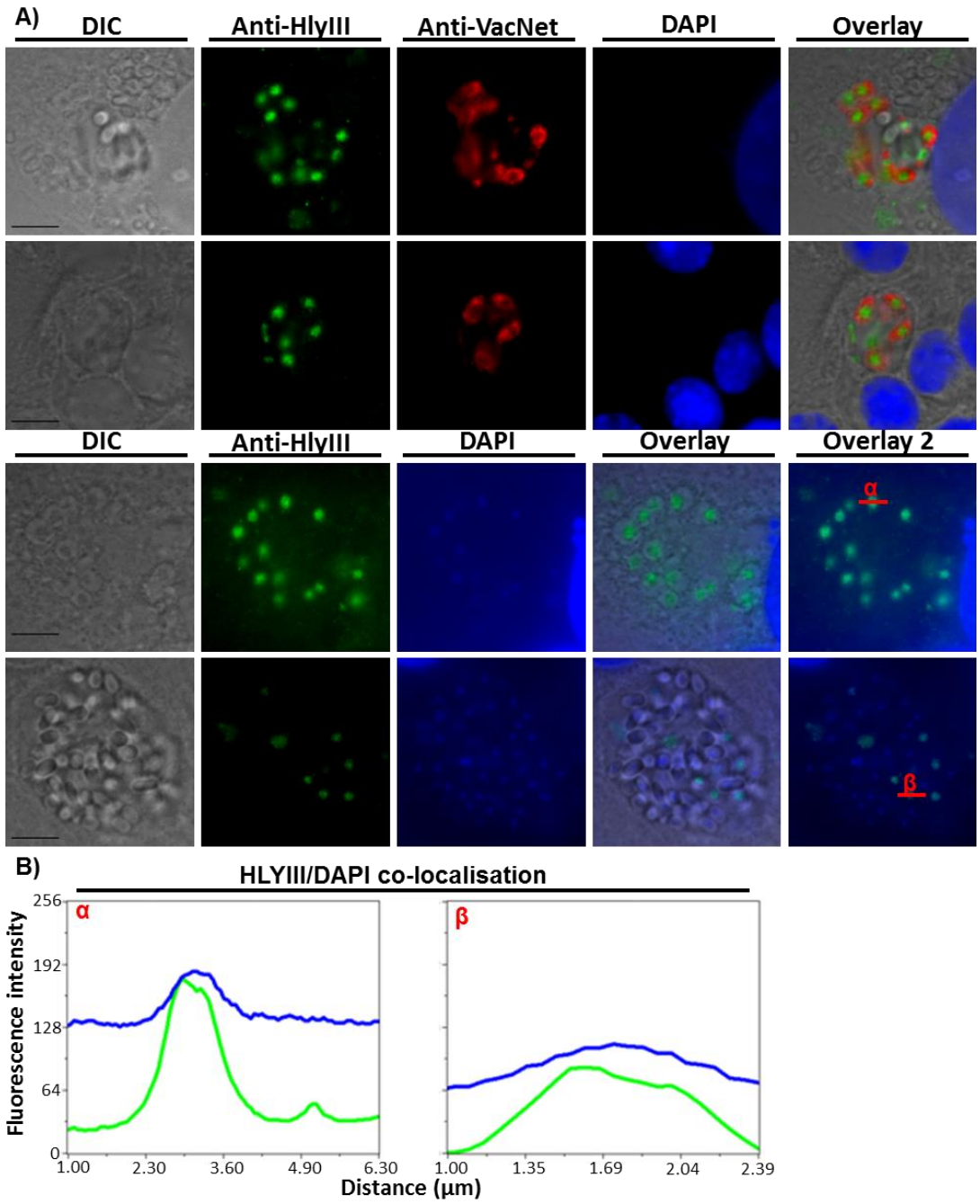


Figure 5.10| Immuno-localisation of EcHlyIII in *E. cuniculi* infected IMCD-3 cells.

A) *E. cuniculi* meronts growing in IMCD-3 cells react specifically with rabbit anti-serum containing polyclonal antibodies raised against EcHlyIII (green). DAPI was used to stain host and parasite nuclei and confirm nuclear localisation of EcHlyIII (scale bar = 5 µm). A vacuolar network filament protein (Anti-VacNet, Red) is used as another point of reference within *E. cuniculi* infected vacuoles (Ghosh *et* al, 2011) B) Fluorescent intensity charts showing HlyIII/DAPI co-localisation across sections of individual meronts.

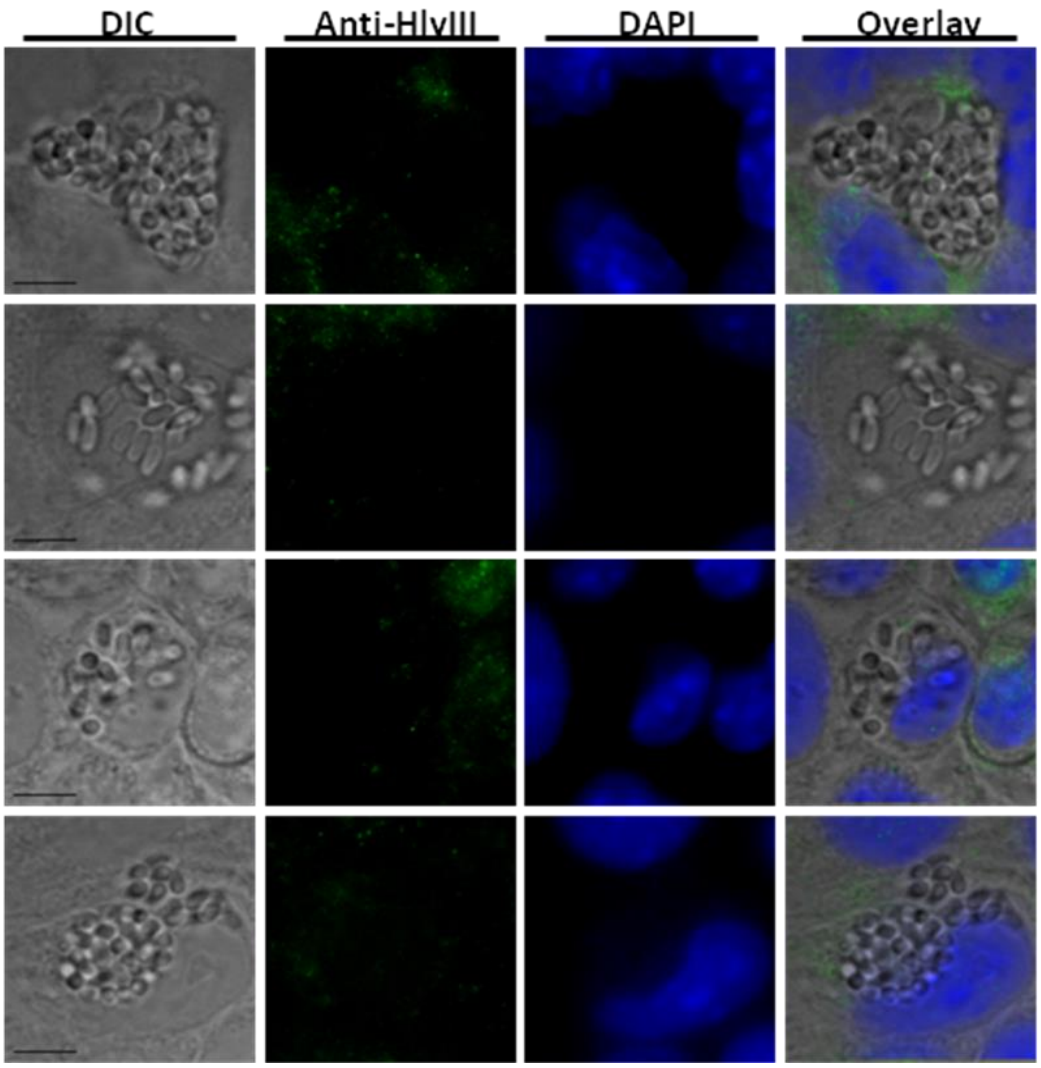


Figure 5.11| Screening pre-immune serum for reactivity against *E. cuniculi* in infected IMCD-3 cells.

No reactivity is evident in *E. cuniculi* meronts or spores in IMCD-3 cells after treatment with rabbit pre-immune serum (green). DAPI (blue) was used to stain host and parasite nuclei. This suggests that final bleed anti-serum reaction with *E. cuniculi* meronts is specific to EcHlyIII challenge and not a previous antigen the individual was inadvertently exposed to (scale bar = 5 µm).

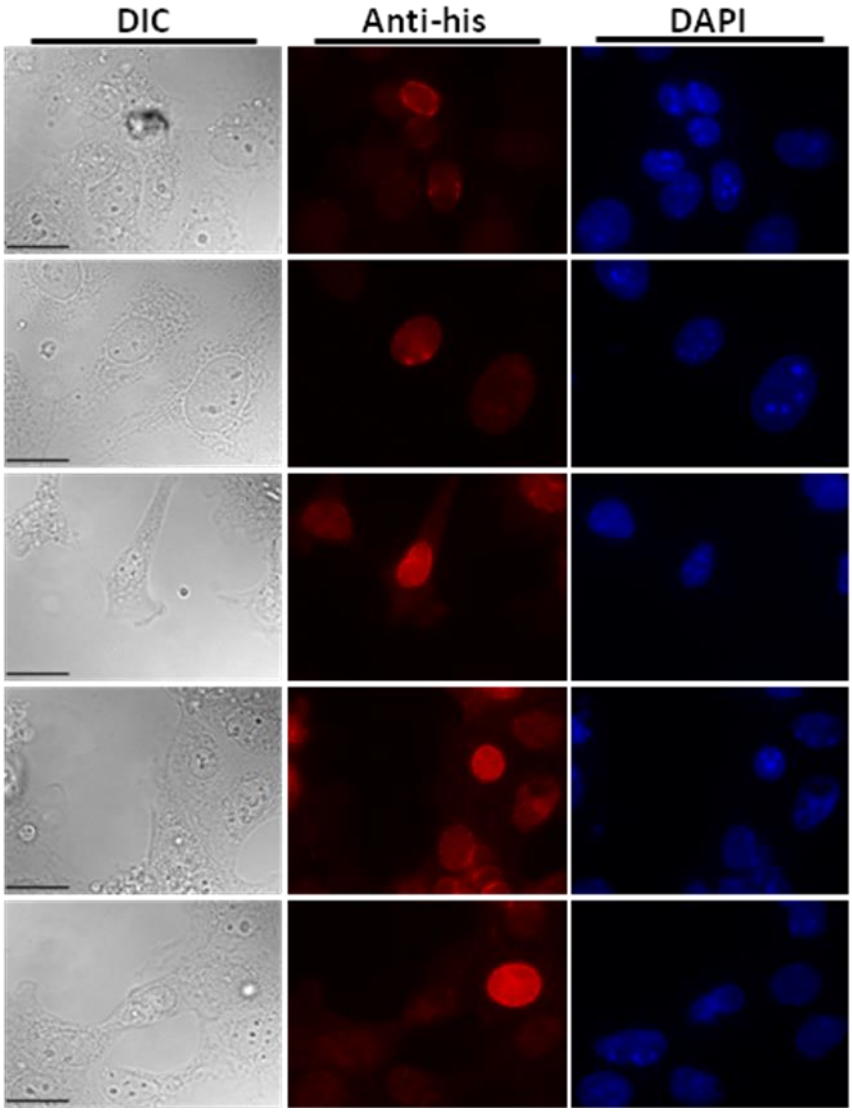


Figure 5.12| Immuno-localisation of EcHlyIII in mammalian cells.
EcHlyIII was cloned in mammalian expression vector pcDNA3.1D/V5-His-TOPO, transfected into IMCD-3 cells and expressed with a C-terminal His-tag. Anti-his antibodies (red) reacted specifically with the nucleus of IMCD-3 cells and appear to associate the nuclear membrane. DAPI was used to stain nuclei and confirm co-localisation (blue) (scale bar = 20 µm).

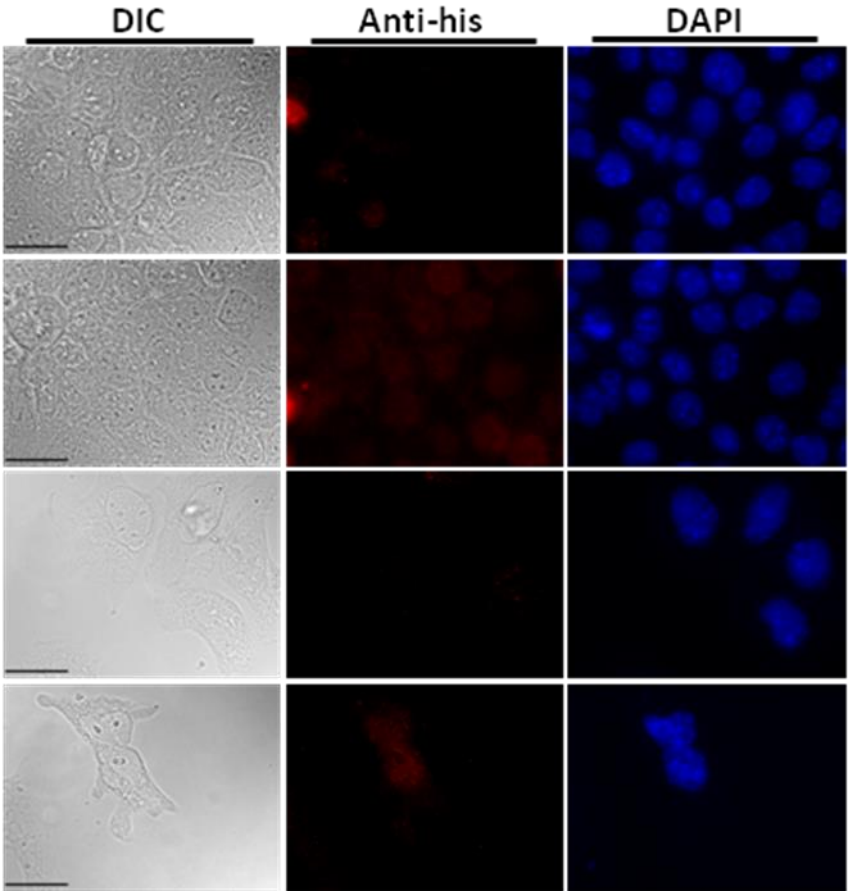


Figure 5.13| Screening anti-his antibody against IMCD-3 cells transfected with pcDNA3.1D/V5-His-TOPO.

As a negative control, IMCD-3 cells were transfected with empty expression vector pcDNA3.1D/V5-His-TOPO and probed with anti-his antibodies (red). No specific staining of IMCD-3 cells was identified. DAPI was used to stain cell nuclei (blue) (scale bar = 20 µm).

5.4| Discussion

Hemolysins are a diverse and ubiquitous group of proteins capable of erythrocyte lysis *in vitro* via formation of pores in the red blood cell membrane. Hemolysin proteins occupy a range of functional roles in both prokaryotes and eukaryotes, and have been adopted by many pathogenic species as a virulence mechanism, involved in direct host cell damage or lifecycle development. The presence of hemolysin proteins in non-pathogenic species including saprotrophicbacterial species *Leptospira biflexa* and avirulent strains of *Listeria monocytogenes* demonstrates that the group is not universally linked to virulence (Farber *et al*, 1991, Picardeau *et al*, 2008).

Database mining using EcHlyIII as a query sequence demonstrates a discontinuous taxonomic distribution of HlyIII orthologs present in a diverse range of kingdoms and phyla, including the Microsporidia, Apicomplexa and Bacteria (Figure 5.1). Due to the highly divergent nature of the gene which, for example, displays only 53% identity between the microsporidian species *E. cuniculi* and *Edharzardia aedis*, it is of course possible that orthologous sequences exist in non-represented groups and that these have not been identified by homology based search analysis (Altschul *et al*, 2005). In maximum likelihood analysis the Microsporidia form a monophyletic group (85% bootstrap support) nested among non-related species, forming a separate clade to their fungal relatives (Figure 5.1). To fully verify the phylogenetic placement of microsporidian HlyIII orthologs, future analyses could be repeated using a site-heterogeneous mixture model (CAT) which has been shown to significantly reduce long-branch attraction artifacts (Lartillot *et al*, 2007).

Three conserved protein sequence motifs were identified in the 105 HlyIII orthologs, which displayed both varying length (from 16 to 50 amino acids) and distribution in terms of both presence/absence and relative position on the protein (Figure 5.2). Microsporidia (excluding *T. hominis* and *V. culicus floridensis*) displayed the same motif distribution (Motif order 2-3-1) as bacteria and *C. remanei*. Fungi and microsporidian species *T. hominis* and *V. culicus floridensis* lacked centre 50 amino acid Motif 3. *E. cuniculi* and previously studied pore forming HlyIII orthologs in *B. cereus and V. vulnificus* have all retained Motif 3, which is lacking in non-hemolytic Izh2 of *S. cerevisiae*. It is therefore possible that Motif 3 is required for pore formation. Site directed mutagenesis replacing single amino acids tryptophan at site 65 and leucine at site 66 reduced the lytic activity of the thermostable hemolysin in *V. parahemolyticus* greater than 150-fold (Baba *et al*, 1992). This demonstrates the impact of even single amino acid substitutions on hemolytic activity. Speculatively, the conserved amino acid sites leucine (Motif 3, site 1) and tryptophan (Motif 3, sites 5 and 18) may be crucial for hemolytic activity in HlyIII orthologs. The significance of motif presence and distribution in terms of hemolytic activity or function is not obvious and requires future experimental investigation.

Fungal HlyIII orthologs such as *S. cerevisiae* Izh2 have previously been characterised in zinc transport and metabolism (Lyons *et al*, 2004). EcHlyIII could not complement this function when expressed in *S. cerevisiae* under zinc stress (Figure 5.5). Speculatively, HlyIII orthologs in microsporidia may have adopted a new functional role, driven by the intimacy of the host-parasite interaction. Alternatively, microsporidian HlyIII may have a role in zinc transport

but the high level of sequence diversity, despite a reasonably well conserved secondary structure, may render EcHlyIII non-functional in yeast as the protein is not recognised by Izh2 interaction partners or targeted in the same manner as the native Izh2 gene. The role of Izh2 orthologs in pathogenic or facultatively intracellular fungi such as *Crytococcus neoformans* is not known.

EcHlyIII displays hemolytic activity of up to 75.8% after 48 hours when expressed in *E. coli* in erythrocyte lysis assays (Figure 5.6). This hemolytic activity was not identified in *E. coli* expressing *S. cerevisiae* Izh2 or paralogous protein Izh1, despite consistent heterologous expression levels (Figure 5.7). These data, taken in conjunction with the lack of functional complementation suggests that EcHlyIII is specifically capable of pore formation and occupies a novel functional role compared to fungal orthologs. EcHlyIII orthologs are widespread in microsporidia. Further investigation will determine if the hemolytic activity displayed by EcHlyIII is universal among microsporidian orthologs. Interestingly, microsporidian species *T. hominis* and *V. culicus floridensis* possessed HlyIII orthologs with identical motif presence/distribution to Izh2, lacking central Motif 3. Determining the hemolytic potential of these proteins could identify the significance of the 50 amino acid Motif 3 in pore formation.

Surprisingly, EcHlyIII displayed a nuclear localisation both *in situ* and in mammalian cells (Figures 5.10 and 5.12). Through localisation of recombinantly expressed EcHlyIII in the larger IMCD-3 mammalian cell, it appears that EcHlyIII is localising to the mouse nuclear membrane. This, taken together with the fact that the protein has 7 predicted transmembrane domains, suggests EcHlyIII most likely sits in the nuclear membrane (Figure 5.12) (Krogh *et al*,

2001). This is a novel localisation for any identified hemolysin, the majority of which are either secreted extracellularly or are pore forming proteins located in the extracellular membrane. This suggests a non-canonical function for HlyIII in *E. cuniculi* and likely a functional role independent of lifecycle progression, vacuole/cell lysis or virulence, as was originally hypothesised. Anti-EcHlyIII serum suggests that the protein is not expressed in the spore/extracellular lifecycle stage, it is possible that EcHlyIII is only required during the intracellular development or proliferation of *E. cuniculi*. Serum and pre-immune serum showed some non-specific binding in western blot analysis. However, a consistent nuclear signal was present only upon treatment with EcHlyIII anti-serum, with no consistent fluorescent signal identified from pre-immune serum treatment (Figure 5.9). Despite consistent localisation of EcHlyIII both *in situ* and in a mammalian expression system, the lack of serum specificity identified in western blot analysis means an alternative cellular localisation for EcHlyIII can not be excluded. It is possible that EcHlyIII anti-serum is binding to another antigen such as the 40 kDa band identified in western blot analysis and that over expression of the foreign EcHlyIII gene in heterologous mammalian expression system has led to non-native packaging and targeting of the protein (Figure 5.9). Future studies identifying the 40 kDa contaminating protein directly by mass spectrometry based N-terminal sequencing or purifying IgG from immune serum using methods such as Protein A or Protein G affinity purification may improve specificity and determine the localisation of EcHlyIII definitively. No nuclear localisation signals were identified in EcHlyIII or any other microsporidian HlyIII sequence. It is possible that microsporidia have developed a novel mechanism for nuclear protein targeting or that nuclear proteins of microsporidia possess a non-canonical retention signal which does

not confirm to the sequence motifs identified in model organisms such as *Homo sapiens* and *S. cerevisiae* (Timmers *et al*, 1999, Lott *et al*, 2011). Analysis of *E. cuniculi* nuclear histone proteins H3 (GI: 449330163) and H4 (GI: 449329984) also failed to identify the presence of a definitive nuclear retention signal using NucPred prediction software suggesting the presence of such a signal needs investigation at the genomic level (Brameier *et al*, 2007). Whether HlyIII localises to the nuclear membrane in all microsporidia is not known. It is difficult to hypothesise a functional role for EcHlyIII and in particular why a novel pore forming protein would be required in the nuclear membrane. Hypothetically, EcHlyIII may be a divergent pore-forming transporter responsible for the delivery of metabolites to and from the nucleus but the lack of molecular tools available in *E. cuniculi* impedes further protein characterisation.

Currently, there is no transformation or gene knock-out system for any microsporidian. The genome of *E. cuniculi* lacks all components of the RNAi machinery that could be use for gene knockdown. However, dicers, argonaute and RNA-dependent RNA-polymerase genes have been found in the genomes of *Nosema ceranae*, *T. hominis* and *Spraguea lophii* (Cornman *et al*, 2009, Heinz *et al*, 2012, Campbell *et al*, 2013). This demonstrates that gene knockdown is a possible genetic manipulation system which could be used widely across the microsporidian phylum but that the machinery required has been lost from highly reduced species such as members of the *Encephalitozoon* genus, many of which are important pathogens of humans. The development of molecular tools and model microsporidian-host systems will be required to determine the function of proteins such as HlyIII in microsporidia, and the nature in which these proteins influence virulence and parasite development.

Chapter 6| Chitin and the origin of the microsporidia.

6.1| Introduction

Chitinase enzymes function in the hydrolysis of glycosidic bonds in chitin, a 1,4ß-linked polymer of *N*-acetyl-ß-D-glucosamine (GlcNAc) and one of the most abundant polysaccharides on earth (Merzendorfer and Zimoch, 2003). Chitinases are a group of enzymes belonging to the glycosyl hydrolase (GH) families 18, 19 and 20, and are found to be broadly distributed in both prokaryotic and eukaryotic species. These chitinase families differ significantly in terms of both sequence similarity and catalytic mechanism, with GH 18 chitinases producing β-anomeric products and GH19 enzymes, by contrast, producing α-anomers (Funkhouser and Aronson, 2007, Seidl 2008). Within these glycosyl hydrolase families, chitinase enzymes are further classified based on cleavage pattern, into groups that include exochitinases, endochitinases or *N*-acetyl-ß-D-glucosaminidases. Endochitinases cleave chitin at any point within the polymer while exochitinases and *N*-acetyl-ß-D-glucosaminidases specifically cleave chitin at the terminal end chain, thus releasing non-reducing *N*-acetyl-ß-D-glucosamine residues (Seidl, 2008, Romão-Dumaresq *et al* 2012).

Chitin contributes the major structural component of the arthropod and crustacean exoskeleton and the fungal cell wall. These chitin-containing organisms require chitinase enzymes for exoskeletal and cell wall remodelling, thus contributing to lifecycle and morphological development (Gooday, 1999). Chitinases have been implicated in both pathogenesis and host defence in a plethora of biological systems. Chitinolytic organisms are also being explored as potential biocontrol agents, where the chitinase activity of symbiotic fungi, bacteria or transgenic plants has been harnessed to protect crop species

against attack by pathogenic fungi, nematodes or insects (Herrera-Estrella and Chet, 1999). Bacteria commonly secrete chitinase enzymes into the extracellular environment in order to break down chitin into GlcNAc monomers that can be scavenged and used as an external energy source (Bhattacharya *et al*, 2007). Chitinase enzymes have also been linked to virulence in a range of insect-infecting fungi. In *Metarhizium anisopliae*, a pathogen of cotton stainer bugs, chitinase knockout mutants show significantly reduced virulence (Boldo *et al*, 2009), and secreted chitinases have been directly implicated in host cuticle penetration in the entomopathogenic fungus *Beauveria bassiana* (Fan *et al*, 2007). Plants also possess a range of chitinase enzymes that are generally up-regulated during stress conditions and function in a range of different processes; most notably in defence against invading fungal pathogens and arthropod grazers (Kasprzewska, 2003). Chitinase accumulation around invading fungal hyphae has been demonstrated and has led to the classification of chitinases as crucial agents of the plant defence system (Collinge *et al*, 1993).

Microsporidia possess a thick chitinaceous endospore in their extracellular spore stage. In contrast chitin is thought to be absent in microsporidian meronts, the initial intracellular lifecycle stage following germination, and is subsequently deposited at the parasite membrane during meront to sporont differentiation (Keeling and Fast, 2002). Despite this complex morphological rearrangement of chitin throughout the microsporidian lifecycle, microsporidia possess a reduced set of genes for chitin metabolism. For example, the genome of *Encephalitozoon cuniculi* encodes only one class I GH 19 chitinase gene, one chitin synthase gene and one chitin deacetylase gene.

This is significantly fewer than the number of chitinase genes present in other fungal phyla such as the Ascomycota in which the human-infective *Aspergillus fumigatus*, for example, possesses a total of 14 GH 18 family chitinases and at least 7 chitin synthase genes (Hurtado-Guerrero and van Aalten, 2007, Abad *et al*, 2010). Of these 14 chitinases present in the *A. fumigatus* genome, 4 are classified as GH 18 family class II genes and the remaining 10 belong to GH 18 class III (Taib *et al*, 2005). Despite the identification of very few genes with a predicted function in chitin remodelling, little is known about microsporidian chitin metabolism. Previous studies have demonstrated that both chitin synthase and endochitinase genes show an up-regulation in expression during meront to sporont development in *E. cuniculi* (Ronnebaumer *et al*, 2006). Interestingly, the chitin synthase inhibitors nikkomycin Z and polyoxin D have been shown to reduce parasite infection and cause cell damage to both spores and meronts in *Encephalitozoon hellem* (nikkomycin Z only) and *E. cuniculi* (both nikkomycin Z and polyoxin D) (Bigliardi *et al*, 2000, Sobottka *et al*, 2002). *E. cuniculi* chitin deacetylase has been shown to localise specifically to the plasma membrane but surprisingly the protein could not deacetylase chitooligosaccharides or beta-chitin (Brosson *et al*, 2005, Urch *et al*, 2009).

Several examples of horizontal gene transfer have been identified in microsporidian parasites to date, giving various species in the phylum a range of new tools for survival in both the extracellular environment and harsh/nutrient-limited host cell environment. From bacteria, microsporidia have obtained protective agents such as catalase (Fast *et al*, 2003), class II photolyase (Slamovits and Keeling, 2004), manganese superoxide dismutase (Xiang *et al*, 2010), ATP transporters (Katinka *et al*, 2001), a GTP

cyclohydrolase and a folic acid synthase (Pombert *et al* 2012). These HGT events have come from a range of prokaryotic donor groups and, with the exception of the ATP transporter genes which are present in all studied microsporidia, are lineage- or species-specific, indicating the approximate timing of the transfer in relation to the diversification of identified microsporidian species (see Figure 1.5 for summary). More recently, a number of horizontally transferred genes derived from a eukaryotic donor have been identified in microsporidia. These include a purine nucleoside phosphorylase for purine salvage in *E. hellem* and *Encephalitozoon romaleae*, and an additional folate salvage enzyme, folylpolyglutamate synthase in *E. romaleae* (Selman *et al*, 2011, Selman and Corradi, 2011). It seems likely that these eukaryotic HGTs are from an ancestral arthropod host of *E. hellem* and *E. romaleae* and that *E. hellem* has since undergone a recent host-switching event to infect vertebrates, including humans (Selman *et al*, 2011). The microsporidian pathogen *Nosema bombycis* has acquired a range of transposable elements (TEs) from host species *Bombyx mori* by HGT and these host-derived TEs have since undergone proliferation, thus contributing to the increased genome size of *N. bombycis* when compared to close relative *Nosema antheraeae* (Pan *et al*, 2013). This proliferation has been a major mechanism of *N. bombycis* genome expansion in which transposable elements contribute over 38% of the overall genome content (Pan *et al*, 2013). This is in stark contrast to other microsporidia such as *Encephalitozoon intestinalis* which is thought to lack transposable elements altogether, despite this, the advantage of transposable element expansion to the parasite is not obvious (Corradi *et al*, 2010).

In this study, the changes in chitin morphology during both germination and intracellular stages of *E. cuniculi* infection are investigated, with the aim to identify changes in chitin structure in the microsporidian spore during polar tube expulsion, an event requiring vast morphological rearrangement. An HGT of a family-19 endochitinase from an insect donor to the microsporidia is also described. This represents the first family-19 chitinase found in any member of the Fungi or their relatives (Hartl *et al*, 2012). This family-19 endochitinase is present in all microsporidia for which there is a complete genome sequence, indicating an ancient HGT event, potentially from an insect host to the 'ancestral microsporidian'. Due to the absence of fungal orthologs it is clear that this HGT event occurred after the microsporidian-fungal divergence, particularly as fungal chitinases identified to date are exclusively family-18 glycosyl hydrolases (Seidl, 2008). Several fossil-verified animal divergence dates are then used to calibrate a molecular clock for the microsporidia. This is the first example where an HGT has been used as a means to add a timescale to the evolution of a parasitic group. Most notably, this study demonstrates that the phylum Microsporidia radiated recently and that their huge diversification in terms of host range, genetic and cellular biology has occurred in a short evolutionary timescale, a direct contradiction to their previous classification as an ancient eukaryotic phylum (Cavalier-Smith, 1987). The presence of this endochitinase in distantly related and highly reduced microsporidian genomes suggests an essential function and due to the ability of these microsporidian species to infect and cause pathology in diverse hosts, a potential role in cellular development or virulence cannot be overlooked. It is possible that the acquisition of this endochitinase by the ancestral microsporidian has contributed, in conjunction

with other genetic anomalies, to the success of the microsporidian phylum which has resulted in vast speciation and host differentiation.

Methods

6.2.1| Cloning of *E. cuniculi* endochitinase Cht1 in and expression *S. cerevisiae*

E. cuniculi endochitinase (Cht1) (GenBank Accession: CAD27103) was amplified using primers EcEndochitPYF and EcEndochitPYR from genomic DNA using Phusion high fidelity DNA polymerase (New England Biolabs) at a primer annealing temperature of 58^{o}C. PCR amplicons were analysed by electrophoresis, gel purified, digested at restriction sites HindIII and EcoRI and ligated into pre-digested pYES2 vector (Invitrogen) as described in Chapter 2.5. *E. coli* XL-1 Blue cells were transformed with Cht1/pYES2 construct as described in Chapter 2.6. Positive colonies were confirmed by PCR, restriction digest and sequencing. Sequenced insert-bearing plasmid was then purified and transformed into *Saccharomyces cerevisiae* wild type strain BY4741a (Euroscarf) and plated onto CSM-Ura media for selection, alongside a pYES2 negative control as described in Chapters 2.6 and 2.8.

6.2.2| Over expression screen of *E. cuniculi* Cht1 in *S. cerevisiae*

S. cerevisiae BY4741a cells containing either pYES2 or *E. cuniculi* Cht1/pYES2 were grown for 16 hours in 5 ml YPD and 24 hours in 5 ml CSM-Ura at 30^{o}C at 180 rpm. Cultures were then pelleted by centrifugation at 1,200 x g and washed twice in sterile 1 x PBS. Cell concentrations for each culture were determined using a hemocytometer and adjusted to a concentration of 1 x 10^{8} cells/ml. Standardised cultures were then diluted serially and spotted (10 µl) onto solid YPD, YPG, CSM-URA and CSM-URA+galactose media and subsequently grown for 48 hours (YPD/YPG) and 72 hours (CSM) at 30^{o}C. Galactose was included in YPG and CSM-URA+galactose media to induce expression of Cht1

via the Gal1 promoter and identify the effects of *E. cuniculi* Cht1 on *S. cerevisiae* growth and/or chitin. For liquid growth assays, *S. cerevisiae* BY4741 clones containing pYES2 and *E. cuniculi* Cht1/pYES2 were grown overnight in both YPD and CSM-Ura broth as described above. Cells were then pelleted and adjusted to O.D. 0.1 at 660 nm in their respective growth media (YPD, YPG, CSM-Ura and CSM-URA+galactose). Standardised cell cultures were then grown in 150 µl cultures in a 96-well plate format at 30^{o}C in a Softmax Versa plate reader (Molecular Devices) until cultures had reached stationary growth phase (approximately 24 hours for YPD/YPG cultures and 30 hours for CSM-URA/CSM-URA+galactose cultures).

6.2.3| Chitin staining of *E. cuniculi* spores, germinated spores and meronts

E. cuniculi-infected IMCD-3 cells were trypsinised from tissue culture flasks using TrypleXpress (Life technologies) as described in Chapter 2.9. The infected cell suspension was then washed twice in H_2O by centrifugation at 1,500 x g at room temperature. Following the last wash cells were resuspended in H_2O (1 ml/200 µl cells) and incubated overnight at 4^{o}C to lyse host cell material. Following overnight incubation cells were vortexed at full speed for approximately 1 minute and carefully pipetted onto 50% percoll (Sigma Aldrich) (500 µl cells/500 µl percoll) in 1.5 ml eppendorf tubes. Cells were subsequently centrifuged through percoll at 4,000 x g for 5 minutes at 4^{o}C to pellet spores, thus purifying them from host cell lysate. The supernatant was carefully removed. Purified spores were then germinated *in vitro* with 1 minute incubation in 0.1% H_2O_2 diluted in PBS and stained with 0.01% Fluorescent Brightener 28 (FB-28, Sigma Aldrich), a calcofluor white derivative, in PBS for 15 minutes in

the dark at room temperature, alongside a non-germinated control. Germinated and non-germinated *E. cuniculi* spores were then washed twice in 1 x PBS to remove excess FB-28 and reduce non-specific background staining. Washed cells were mounted in Vectashield fluorescent mounting media and visualised using an Olympus IX-81 microscope using DIC and fluorescence at excitation 436 nm: emission 488 nm. To visualise chitin morphology during intracellular *E. cuniculi* development, infected IMCD-3 cells were grown in 6-well plates on sterile coverslips until confluent in DMEM/F-12/10% FBS at 37^{o}C/5% $CO_{2.}$ Cell monolayers were washed twice in 1 x PBS with gentle agitation at room temperature and subsequently stained in 0.01% FB-28 for 15 minutes. Cells were mounted and visualised as described above.

6.2.4| Database searching

E. cuniculi family-19 endochitinase was used to search publicly available genome databases (National Centre for Biotechnology Information, The Joint Genome Institute, The Broad Institute, Baylor College of Medicine Genome Sequencing Centre, the Norwegian Institute for Marine Research and the National Science Foundation Plant Genome Database) with a pre-defined e-value cut-off of $1e^{-5}$. After removal of redundant or paralogous protein sequences including the product of a recent duplication in *Culex quinquefasciatus* (GenBank Accession: XP_001870031), 93 eukaryotic and prokaryotic taxa were included in the phylogenetic analyses. An endochitinase ortholog was also found in whole genome shotgun data from ectoparasitic copepod *Lepeophtheirus salmonis* (GenBank Accession: ADND02099729) and unexpectedly associated within the microsporidia in phylogenetic analysis. Due to the presence of orthologs of many microsporidian specific genes including

spore wall and polar tube proteins in the *L. salmonis* genome shotgun dataset, the high prevalence of microsporidian infection in this species (Freeman *et al*, 2003) and the absence of the gene from a second sequenced *L. salmonis* isolate (Norwegian Institute of Marine Research) this sequence has been assumed to be a contaminant, hence it has been excluded from further analysis. To assess the degree of gene synteny, 7 kb genome fragments neighbouring endochitinase orthologs in microsporidian genomes were analysed using the Broad Institute microsporidian comparative database genome map tool.

6.2.5| Phylogenetic analysis

Endochitinase orthologs were renamed in standard species/accession number format using REFGEN/TREENAMER (Leonard *et al*, 2009). Sequences were aligned using MUSCLEv3.2 (Edgar, 2004) and masked manually in SEAVIEW to remove hypervariable sequence regions. Trees from the masked alignment file of 148 amino acid sites were then constructed using both Bayesian and maximum likelihood methodologies. Bayesian analysis was conducted using PhyloBayes version 3.3f (Lartillot *et al.*2009) CAT20 model for 72768 and 72703 cycles on each respective chain, with the first quarter discarded as ‘burn in’. Trees were sampled every tenth cycle, until chains had converged with a maximum discrepancy between bipartition frequencies in the chains of <0.1. A consensus tree was created from the retained samples. Maximum likelihood trees were constructed using RAxML version 7.2.6 under the Whelan and Goldman (WAG) substitution model (Whelan and Goldman, 2001) as recommended as optimum by the MODELGENERATOR substitution model prediction program (Keane *et al*, 2006) with bootstrap support values calculated

from 1000 pseudo-replicates. To further test for differential rate problems and long-branch attraction artefacts the data was recoded into Dayhoff groups, a reduced alphabet approach (Susko and Roger, 2007) and previous analysis in RAxML and PhyloBayes was repeated. Because compositional bias in sequence alignments can have a profound effect on phylogenetic reconstruction, phylogenetic analyses was re-run using a protein LogDet distance approach utilising an LDDist-like strategy to compensate for invariant sites (Thollesson, 2004), TK02 correction, inversion of negative determinants and with 1000 bootstrap pseudo-replicates using a Bio-Neighbour Joining tree search in the P4 package. LogDet analysis was conducted by Finlay Maguire, Natural History Museum, London.

6.2.6| Molecular dating

Divergence dates were estimated from BEAST v. 1.6.1 using a Bayesian Markov Chain Monte Carlo (MCMC) approach (Drummond and Rambaut, 2007). Posterior probability distributions of node ages were obtained using WAG with rate variation (four gamma categories) and an uncorrelated relaxed molecular clock branch rate model under log-normal distribution to compensate for varying substitution rates across amino acid sites and heterogeneous rates of evolution between branches respectively. A Yule tree prior (Yule 1925) was employed, as recommended for species tree analysis. Two independent MCMC chains were run for a minimum of 7,013,000 generations (burn-in 10%), with parameters sampled every 1000 generations until chains had converged. In order to assess that the MCMC chains had stabilised, lnL plots were examined using Tracer v. 1.4 (Rambaut and Drummond, 2003). Molecular dating was conducted by Dr Patrick Hamilton, University of Exeter.

6.2.7| Prediction of protein localisation and domain searching

Based on the final tree topology *E. cuniculi* Cht1p orthologs were analysed by secretion/localisation prediction programs SignalP4.1 (Petersen *et al*, 2011), TargetP1.1 (Emanuelsson *et al*, 2007) and Pfam (Punta *et al*, 2012) to identify the presence/absence of predicted functional domains. Cut-off for significant prediction values were set using server recommendations in SignalP and TargetP and at $1e^{-5}$ in Pfam.

6.2.8| List of oligonucleotides

Table 6.1| Names and sequences of oligonucleotides used in this study.

Oligo name	Sequence (5'-3')	Restriction site
EcEndochitPYF	cgat**aagctt**atgctggtcttcttcttgggcctg	HindIII
EcEndochitPYR	gcat**gaattc**ctaactgtagcacccactttcatc	EcoRI

The oligonucleotides listed above were required for cloning the *E. cuniculi* endochitinase into *S. cerevisiae* expression vector pYES2 using restriction site overhangs HindIII and EcoRI. Restriction site sequences within primers are highlighted in bold.

6.3| Results

6.3.1| Chitin and morphological development of microsporidia

Microsporidia survive in the extracellular environment as a spore containing a thick chitinaceous endospore and a proteinaceous exospore wall (Vavra *et al*, 1993). Upon receipt of an appropriate stimulus the spore germinates and expels its polar tube, a unique infection organelle required for the transport of infectious material into the host cell (Takvorian *et al*, 2005). Microsporidia then undergo binary or multiple fission during a lifecycle stage termed 'merogony' before development into a mature spore (Canning, 1988). This complex lifecycle involves huge changes in cellular structure and morphology. To investigate structural changes in microsporidian chitin during polar tube expulsion, dormant and germinated *E. cuniculi* spores were stained with calcofluor white derivative, FB-28 (Figure 6.1). Dormant spores displayed a uniform staining pattern, demonstrated both visually and quantitatively; indicative of an intact and equally distributed chitinaceous endospore (Figure 6.1). In contrast, germinated spores show a clear region of chitin degradation at the site of polar tube expulsion, as demonstrated by a reduction in FB-28 staining intensity (Figure 6.1). Polar tube expulsion occurs at the apical tip of the spore, the region with the thinnest spore wall (Frixione *et al*, 1992). FB-28 studies clearly demonstrate chitin degradation around the site of polar tube expulsion. It is possible that the N-terminal secretion signal directs Cht1 to the spore apical tip to aid this process or that the protein functions at alternative point in the morphological development of microsporidia (see Figure 6.2 for functional model). However, without definitive molecular evidence this function cannot be assumed. Alternatively, it is possible that *E. cuniculi* and other microsporidia secrete Cht1 into the host cell environment to degrade host substrates and utilise the breakdown products for

metabolism. This process has been described in bacteria as a means of acquiring both carbon and nitrogen (Larsen *et al*, 2010). Although this process has never been characterised, microsporidia have a reduced set of carbon metabolism genes and in the case of *Enterocytozoon bieneusi*, have lost carbon metabolism genes altogether, hence Cht1 secretion into the host may provide elegant means to acquire carbon (Keeling *et al*, 2010) (Figure 6.2).

Once inside the host cell *E. cuniculi* develops within a vacuole containing parasites in various developmental stages. FB-28 staining of mature *E. cuniculi* vacuoles inside IMCD-3 cells demonstrates an absence of chitin in *E. cuniculi* merogony (Figure 6.3). However, chitin can be identified in spores which are also present within the vacuole (Figure 6.3). Taken together, these studies demonstrate that chitin metabolism is a ubiquitous process ongoing both in extracellular and intracellular stages of the microsporidian lifecycle.

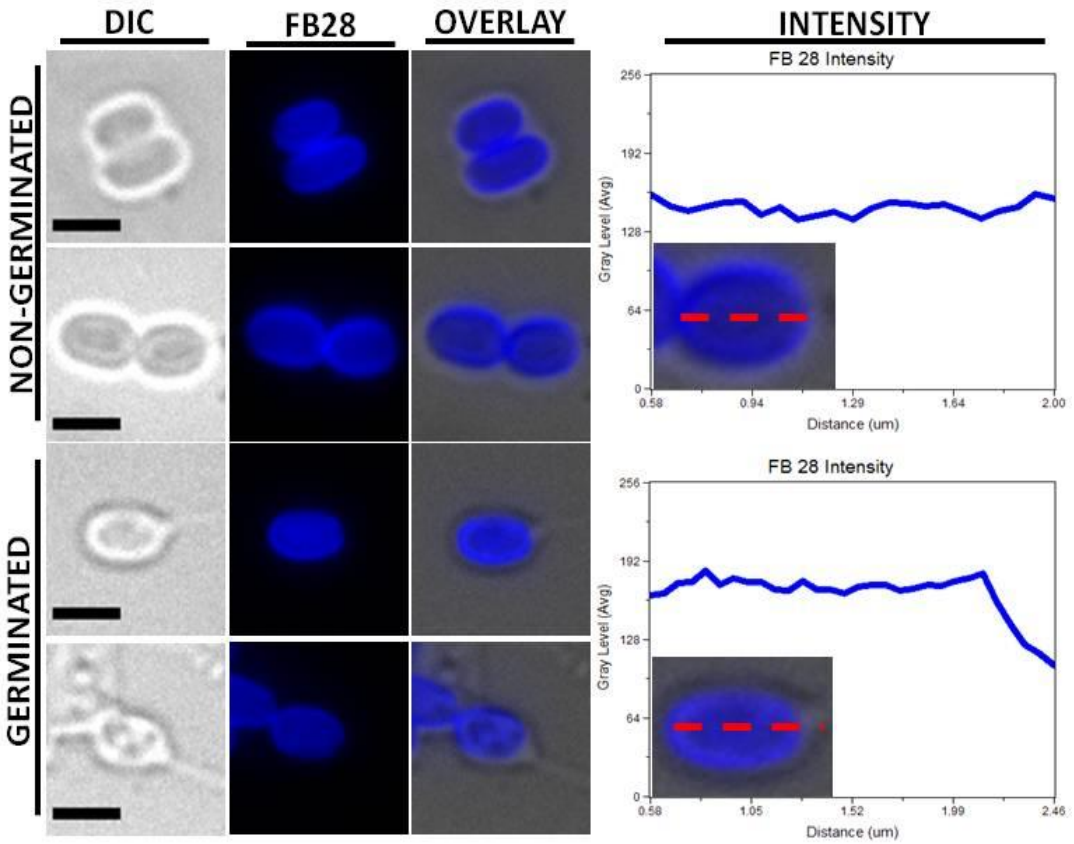


Figure 6.1| Chitin degradation at the site of polar tube expulsion during *E. cuniculi* germination.

Dormant and germinated *E. cuniculi* spores were stained directly with FB-28 to determine the presence and morphology of chitin in the endospore. Germination was easily determined by presence of the extruding polar tube. Fluorescence intensity across spore diameter and at the site of polar tube expulsion was measured using Metamorph Image Analysis software. It is evident that dormant spores possess an intact chitinaceous endospore, however at the polar tube expulsion site there is evidence for chitin degradation, demonstrated clearly in the overlaid image and fluorescence intensity graph of germinated samples. It cannot be determined in the absence of specific molecular evidence whether this chitin degradation/loss at polar tube expulsion site in an enzymatic or a physical process (scale bar = 2 µm).

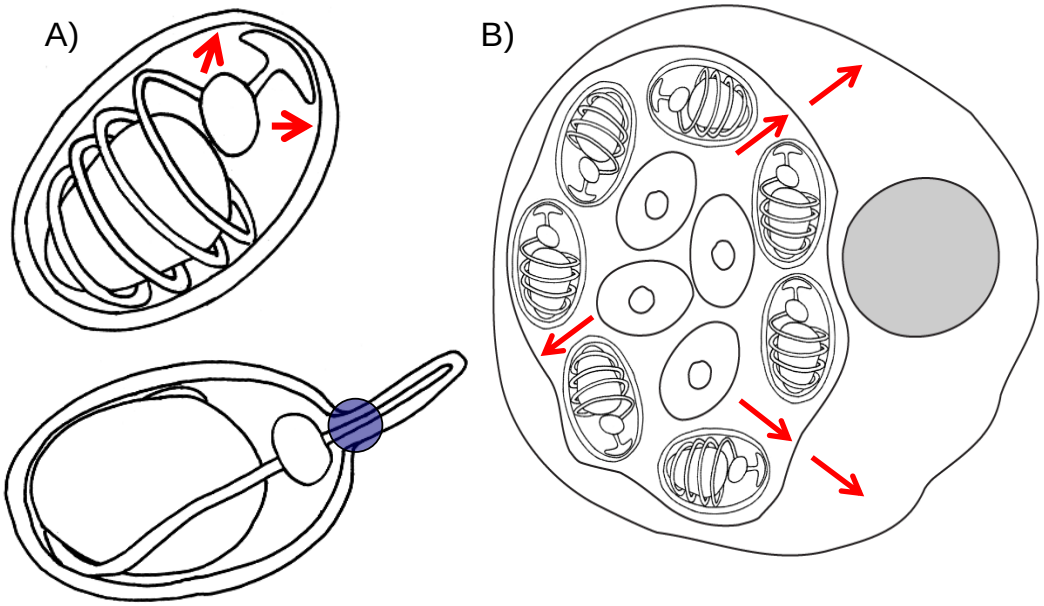


Figure 6.2| Hypothetical functional model for the microsporidian endochitinase Cht1.

A) – Cht1 is secreted to the chitinaceous microsporidian endospore. Hydrolysis of the endospore weakens the spore wall and allows efficient polar tube expulsion upon receipt of host cell stimuli. The polar tube then punctures the host cell and injects infectious sporoplasm. The blue circle represents a possible site for Cht1 mediated chitin degradation at the spore apical tip.

B) - Cht1 is secreted through the parasitophorous vacuole or directly into the host cell cytoplasm. The specific host target substrate is then hydrolysed and the resultant breakdown products are scavenged for parasite metabolism.

Adapted from Williams and Keeling, 2011

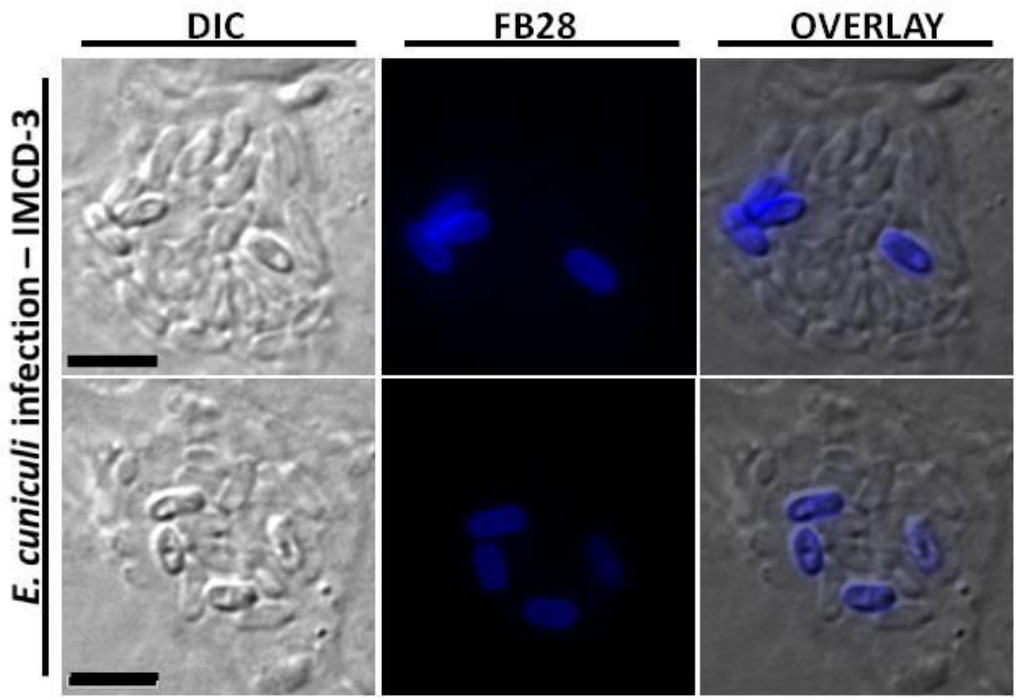


Figure 6.3| Chitin morphology during intracellular development of *E. cuniculi.*

E. cuniculi infected IMCD-3 cells were treated with FB-28 to determine variation in chitin morphology through the intracellular lifecycle stages of *E. cuniculi* infection. Parasites within vacuoles represent a range of developmental stages, namely, spores, sporonts and meronts. Chitin is only present in the wall of mature *E. cuniculi* spores (scale bar = 5 µm).

6.3.2| *E. cuniculi* Cht1 displays chitin substrate specificity

E. cuniculi Cht1 gene was amplified using the species-specific primer set EcEndochitPYF/EcEndochitPYR (Table 6.4A) and PCR. The resultant 1791 bp amplicon was size verified by electrophoresis on a 1% TAE agarose gel (Figure 6.4A). Size verified PCR products were purified, digested, ligated to pYES2 and transformed into *E. coli* as described in Chapter 6.2.1. *E. coli* colonies growing under selection were screened by colony PCR (Figure 6.4B). Insert-bearing plasmids were purified from *E. coli* clones and further verified for Cht1 presence by restriction digest and sequencing (Figure 6.4C).

The optimal conditions for chitinase activity vary significantly between members of what is a diverse enzyme family, particularly in terms of pH and temperature (Rabeeth *et al*, 2011). For example, the family-18 chitinase CBFOS-03 from *Bacillus licheniformis* displays optimum activity at pH 9 and 55^{o}C whereas, in contrast, a chitinase enzyme purified from an acidophilic *Bacillus* strain functions optimally at pH 5 and 37^{o}C (Natsir *et al*, 2002, Islam *et al*, 2010). Similar functional diversity is evident in chitin substrate specificity. Individual chitinases generally exhibit preferential activity against a specific chitin substrate such as GlcNAc, chitosan, which is a deacetylated chitin form, or to a chitin oligomer of a specific length (Brunner *et al*, 1998). Chitin provides strength to the fungal cell wall forming anti-parallel chains linked internally by hydrogen bonds and externally linked covalently to β(1,3)-glucan (Lenardon *et al*, 2010). To determine the activity of *E. cuniculi* Cht1 against the chitin cell wall of *S. cerevisiae*, the gene was overexpressed in yeast alongside a pYES2 negative control. Cells were grown on solid nutrient rich YPD media and minimal media lacking uracil, both of which contained glucose for repression of

the pYES2 Gal1 promoter and would act as experimental growth controls (Figure 6.5, YPD and –Ura+Glu). To test the consequence of Cht1 expression on the growth of *S. cerevisiae*, these conditions were replicated with the addition of galactose for induction of the Gal1 promoter and expression of Cht1 (Figure 6.5, YPG and –Ura+Gal). *S. cerevisiae* Cht1/pYES2 and pYES2, as expected, displayed similar growth on glucose-containing repression media when grown for 48 hours on YPD and 72 hours on minimal media (Figure 6.5). Expression of Cht1 on both YPG and minimal media containing galactose also showed no identifiable growth defect in *S. cerevisiae* when compared to the pYES2 bearing control strain (Figure 6.5). To consolidate this result, the experiment was repeated in liquid media under the same experimental conditions (Figure 6.6). Cultures were grown for approximately 23 hours (YPD/YPG) and 28 hours (-URA/-URA+Gal) until growth curves had reached stationary growth phase. In agreement with growth on solid medium, no growth defect was observed for either Cht1 expressing or pYES2 control strains and no difference between their growth rate was evident. Taken together, these data suggest that *E. cuniculi* Cht1 is not capable of *S. cerevisiae* cell wall degradation under these growth conditions, thus suggesting that the protein is either not functional at this specific pH and/or temperature or that Cht1 is not active against fungal cell wall chitin substrate. However, it is also possible that despite the presence of a predicted secretion signal by both SignalP and TargetP that *E. cuniculi* Cht1 is not targeted in a manner that allows it to access to the cell wall when recombinantly expressed in *S. cerevisiae* (Emanuelsson *et al*, 2000, Petersen *et al*, 2011).

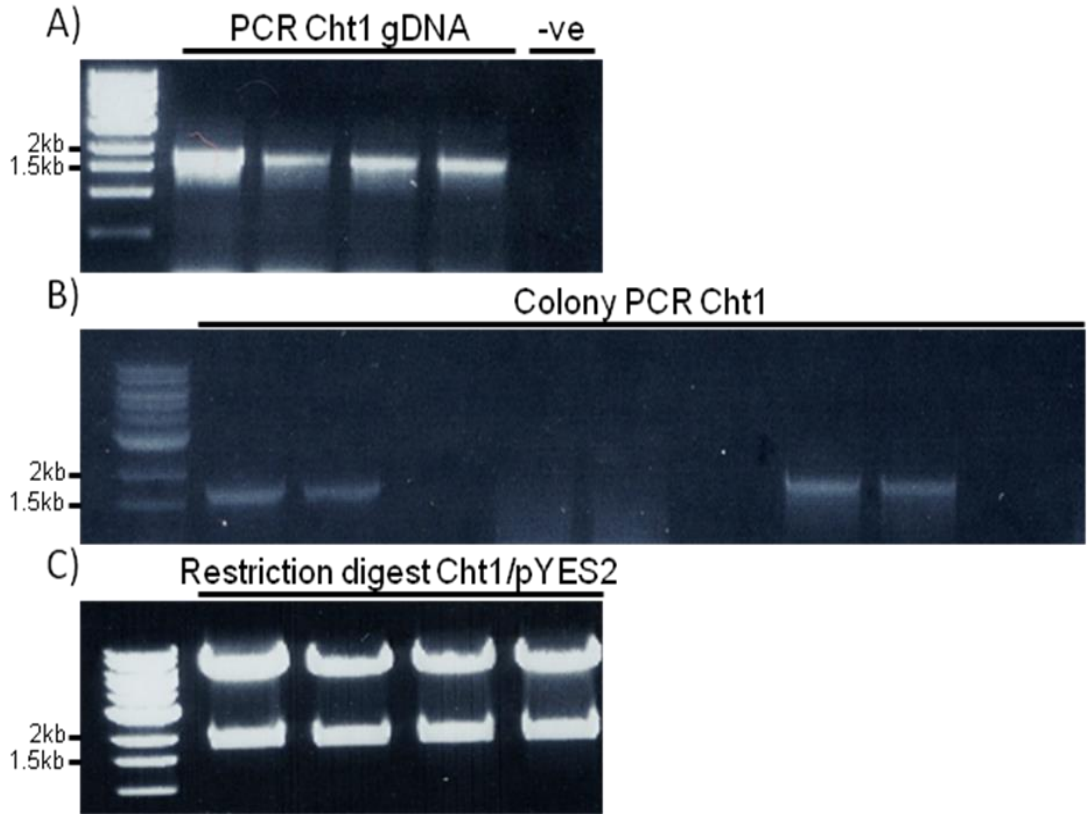


Figure 6.4| Amplification and cloning of *E. cuniculi* Cht1 into *S. cerevisiae* expression plasmid pYES2.

A) Amplification of Cht1 from *E. cuniculi* GB-M1 genomic DNA using species-specific gene primers EcEndochitPYF and EcEndochitPYR (Table 6.1). B) Colony PCR of Cht1 from *E. coli* XL-1 Blue colonies grown under ampicillin selection following transformation with Cht1/pYES2. PCR was conducted using a T7 promoter primer and EcEndochitPYR to reduce false positives. C) Restriction digest confirmation of Cht1 insert in four colony PCR positive clones.

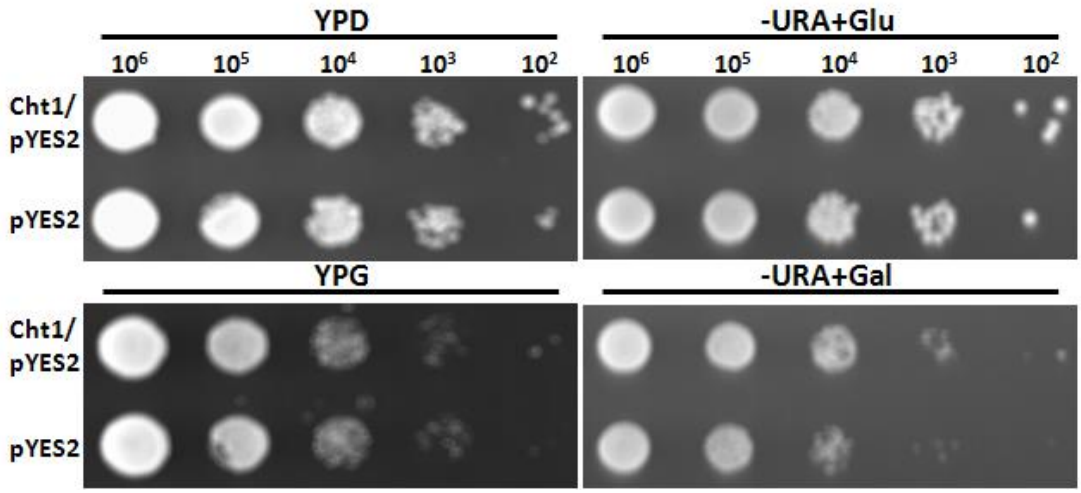


Figure 6.5| Over expression screen of *E. cuniculi* Cht1 in *S. cerevisiae* on solid growth medium.

S. cerevisiae BY4741a expressing *E. cuniculi* Cht1 and empty plasmid control pYES2 were diluted serially then spotted on control/promoter repression media: Yeast Peptone Dextrose and Complete Synthetic Media –URA and expression media: Yeast Peptone Galactose and Complete Synthetic Media + Galactose –URA. Cells were grown at 30^{o}C for 48 hours (YPD/YPG) and 72 hours (-URA). Cht1 expression in *S. cerevisiae* showed no growth or morphology defect suggesting either that *E. cuniculi* Cht1 is not active against the yeast chitin substrate or it is not targeted/secreted specifically to the chitin cell wall.

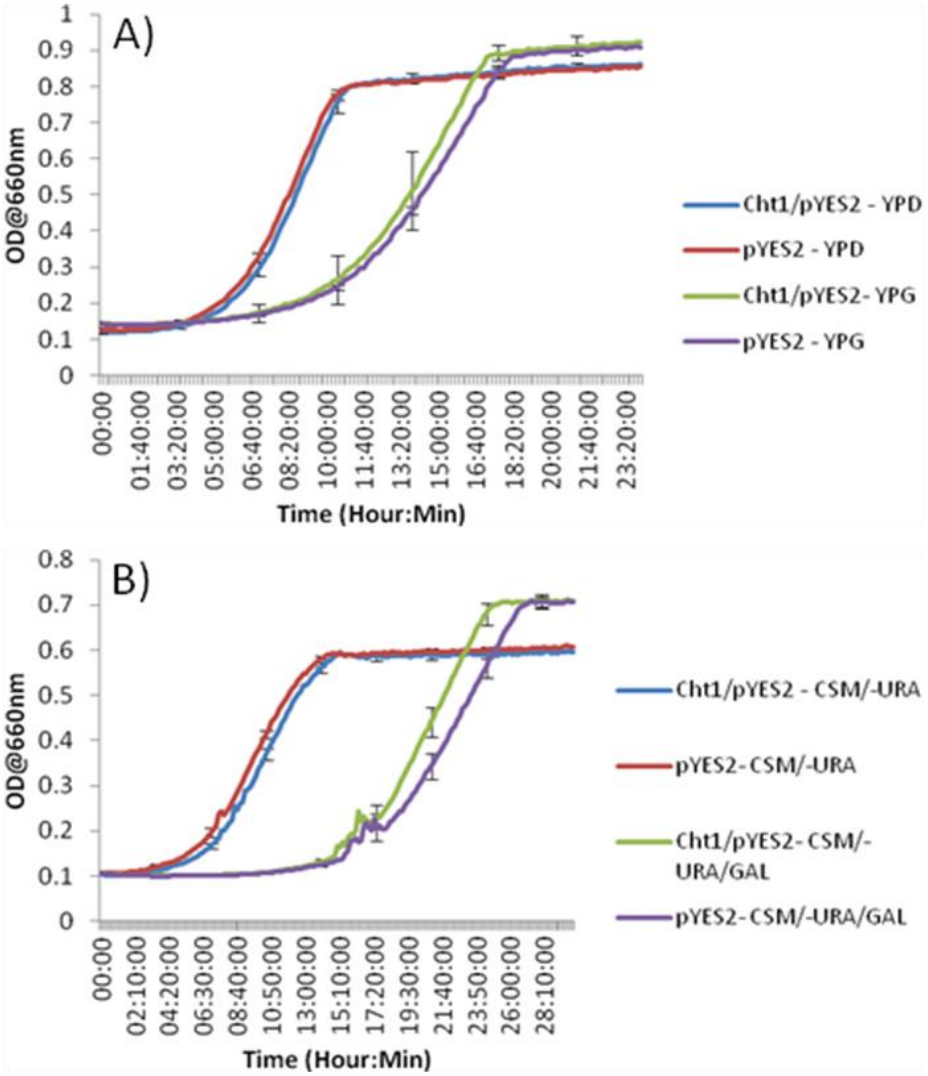


Figure 6.6| Over expression screen of *E. cuniculi* Cht1 in *S. cerevisiae* in liquid growth medium.

S. cerevisiae BY4741a expressing *E. cuniculi* Cht1 and pYES2 control were grown in liquid media. A) Growth of *S. cerevisiae* strains in YPD and YPG and B) Growth in CSM/–URA and CSM/-URA/GAL. Growth curves were calculated by measuring OD at 660 nm with time points taken every 10 minutes in an automated plate reader at 30^{o}C. Curves were plotted untill all strains reached stationary phase, 24 hours in A) and 30 hours in B). Error bars were calculated from 3 biological replicates, with two technical replicates of each and were plotted every 20 data points to aid readability. Similar to expression on solid media, Cht1 expression in liquid culture results in no significant growth defect in *S. cerevisiae*.

6.3.3| Horizontal gene transfer of Cht1 and the microsporidian radiation

Database searching using *E. cuniculi* Cht1 as a query sequence identified putative endochitinase orthologs in eubacteria and several eukaryotic lineages including plants, animals and chromalveolates (Figure 6.7). The analysis identified numerous inferred cases of gene loss within the metazoa and other eukaryote groups, but found only one verifiable case of gene duplication in the animals. As expected no Cht1 orthologs were found in sampled fungal genomes (databases checked 10/2012) as family-19 endochitinase genes have never been identified in the fungal kingdom (Hartl *et al*, 2012). However, Cht1 orthologs were identified in all microsporidian species for which there is a complete genome sequence, a taxon distribution which represents a broad selection of currently known microsporidian lineages. Phylogenetic analysis of this endochitinase gene family revealed a moderately supported monophyletic clade of microsporidia (Posterior Probability (PP): 0.82, Maximum Likelihood Bootstrap Support (ML BS): 69%). This suggests that the gene was present in the last common ancestor of all known extant microsporidia. Interestingly, phylogenetic analyses demonstrated that the microsporidia fall within the animal (eumetazoa) clade with significant topology support at multiple nodes. Specifically, phylogenetic trees demonstrated that the microsporidian Cht1 family form a clade nested within the class Insecta with strong topology support (Node B, PP: 1, ML BS: 99%), and branch as a sister group to the wasp species *Nasonia vitripennis* (Node C, PP: 0.98, ML BS: 94%, Figure 6.7). This suggests a gene transfer event has occurred from within the insects to the microsporidia (Figure 6.9). To minimise the effect of long-branch attraction problems, phylogenetic analysis was repeated using a reduced alphabet approach in which the amino acid dataset was converted into Dayhoff classes (Dayhoff and

Schwartz, 1978). Trees constructed using the reduced alphabet approach demonstrated the same topological relationship in key nodes supporting HGT with moderate to strong support (Figure 1, Node B, PP: 0.95, ML BS: 64% and Node C, PP: 0.99, ML BS: 91%). Secondly, to control for the possibility that compositional bias in the dataset may be a factor influencing the phylogenetic placement of the microsporidia, phylogenetic analysis was repeated using a protein LogDet distance approach (Foster and Hickey, 1999). This analysis again recovered the same tree topology relationship with regards to the placement of the microsporidia within the Insecta (Figure 6.7, Node B, BS: 62% and Node C, BS: 38%), albeit showing reduced bootstrap support values (LogDet analysis conducted by Finlay Maguire, Natural History Museum, London). Importantly, Cht1 orthologs demonstrated a moderate degree of gene synteny between diverse microsporidian species strongly suggesting that the gene had been acquired ancestrally and had not spread between microsporidian species by HGT (Figure 6.8)

The distribution of this HGT among the microsporidia provides a unique opportunity to use an animal-derived HGT combined with animal fossils and molecular clock approaches to estimate the age of divergence for the major microsporidian parasite group. Therefore, Bayesian molecular dating analysis was conducted using BEAST (Drummond and Rambaut, 2007), assuming no prior topology constraints, using three fossil-verified calibration points within the Metazoa (Molecular dating analysis conducted by Dr Patrick Hamilton, University of Exeter). The dates included were 1. subkingdom Eumetazoa - 587 MYA (Peterson and Butterfield, 2005) 2. Hymenoptera and Diptera within the Insecta - 300 MYA (Labandeira and Phillips, 1996) and 3. mosquitoes species

Culex quinquefasciatus and *Aedes aegypti* - 38.5 MYA (Foley *et al*, 1998). Dates included standard deviations of +/- 29 million years for Eumetazoa and +/- 5 million years the latter two calibration points. The analysis identified the same tree topology shown in Figure 6.7 and indicated that the radiation of the microsporidian group sampled occurred 199 MYA (95% CI: 139-264 MYA) (Figure 6.10). As BEAST, under certain circumstances may become stuck on local optima the analysis was repeated with 4 monophyletic groups pre-defined (A-D in Figure 6.7) to verify the dates only and not the topological relationship. This supported initial dating analysis and predicts the microsporidian radiation to have occurred 218 MYA (95% CI: 151-308 MYA).

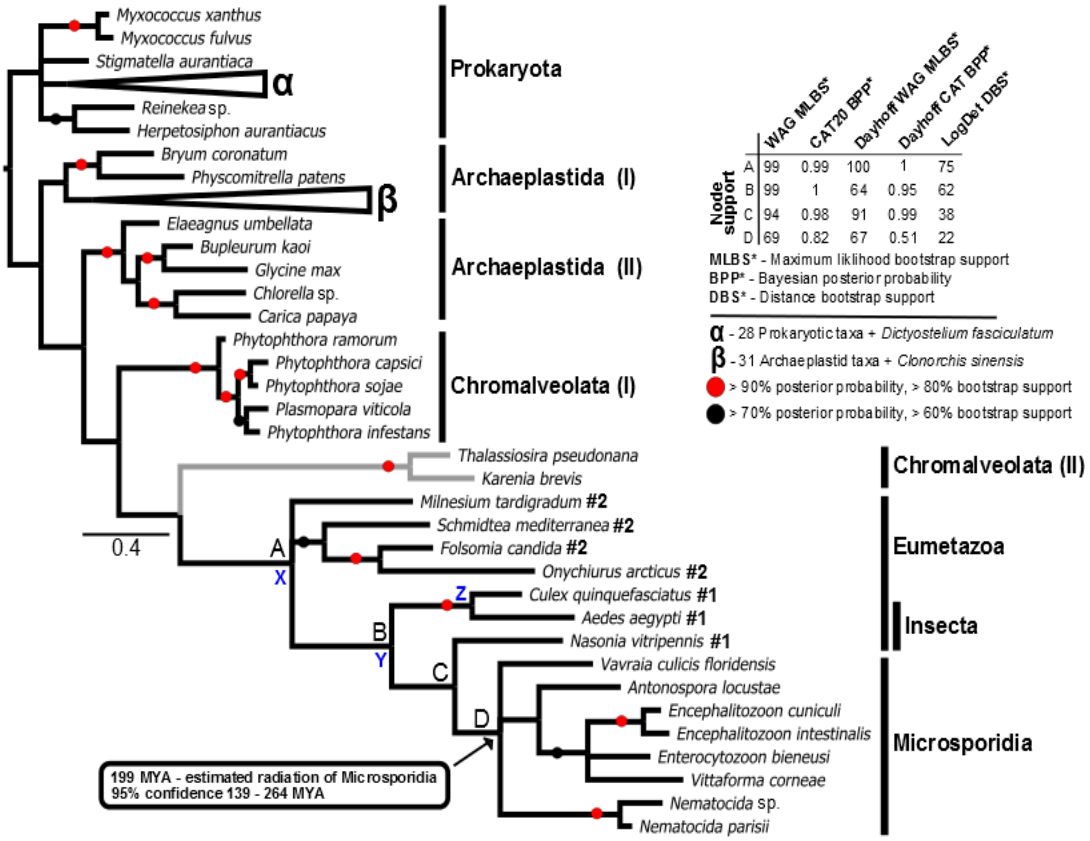


Figure 6.7| Phylogenetic distribution of endochitinase orthologs across 93 prokaryotic and eukaryotic species.

Key nodes supporting HGT (A, B, C) and monophyly of the microsporidia (D) are labelled and have corresponding support values for a range of different phylogenetic analysis methods tabulated. Nodes B and C represent key nodes for inferring HGT, thus demonstrating that the microsporidian endochitinases branch with and within the eumetazoan clade. Fossil verified animal divergence dates are labelled on appropriate branches (X- 587 MYA, Y- 300 MYA and Z- 38.5 MYA). *Clonorchis sinensis* and *Dictyostelium fasciculatum* are labelled as they group within collapsed branches of Archaeplastida and prokaryotic taxa but do not belong to these respective taxonomic groups, suggesting additional cases of gene transfer of this gene family into eukaryotic genomes.

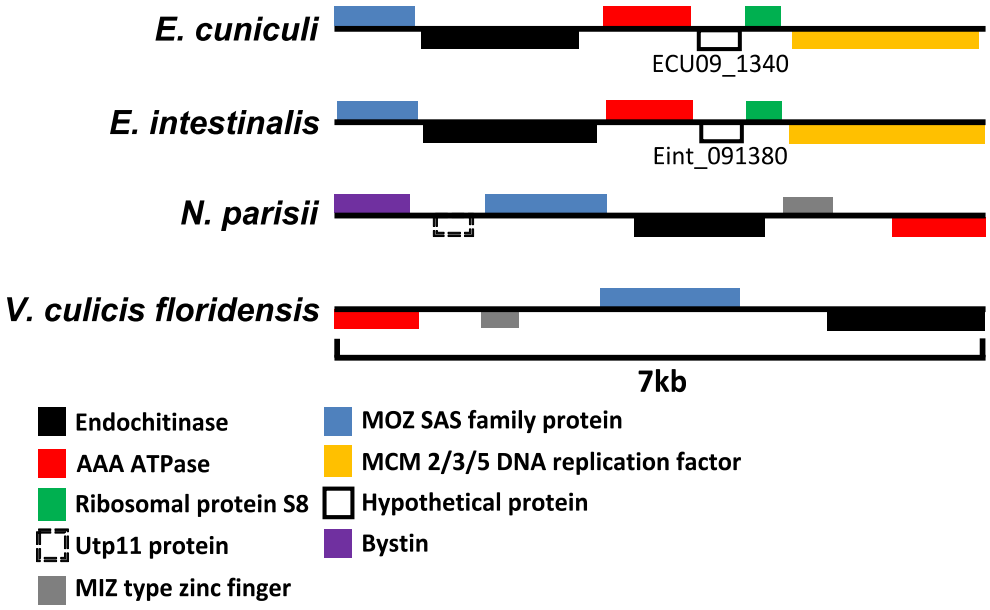


Figure 6.8| Genomic location of endochitinase orthologs and relative synteny in microsporidian genomes.

The 7 kb genomic regions containing the endochitinase gene of the microsporidian species *E. cuniculi*, *E. intestinalis*, *Nematocida parisii* and *Vavraia culicus floridensis* were mapped using the Broad Institute Microsporidia comparative database genome map tool to demonstrate the conserved location of endochitinase orthologs in relation to other conserved microsporidian genomic markers such as the MOZ SAS family and AAA-ATPase protein encoding genes.

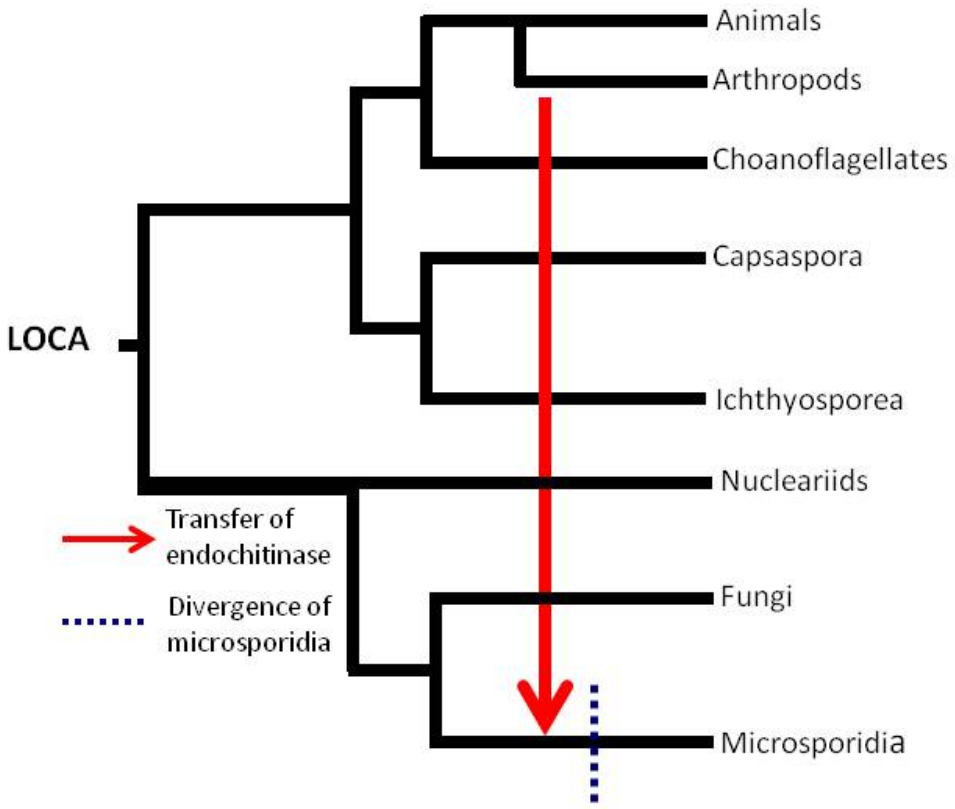


Figure 6.9| HGT model of Cht1 from the arthopod donor to the ancestral microsporidian.

Cht1 was aqcuired by microsporidia from an arthropod donor prior to the radiation of sequenced members of the phylum but after the microsporidian-fungal divergence. The event has been mapped onto an interpretation of the Opisthokont tree. (LOCA = Last Opisthokont Common Ancestor).

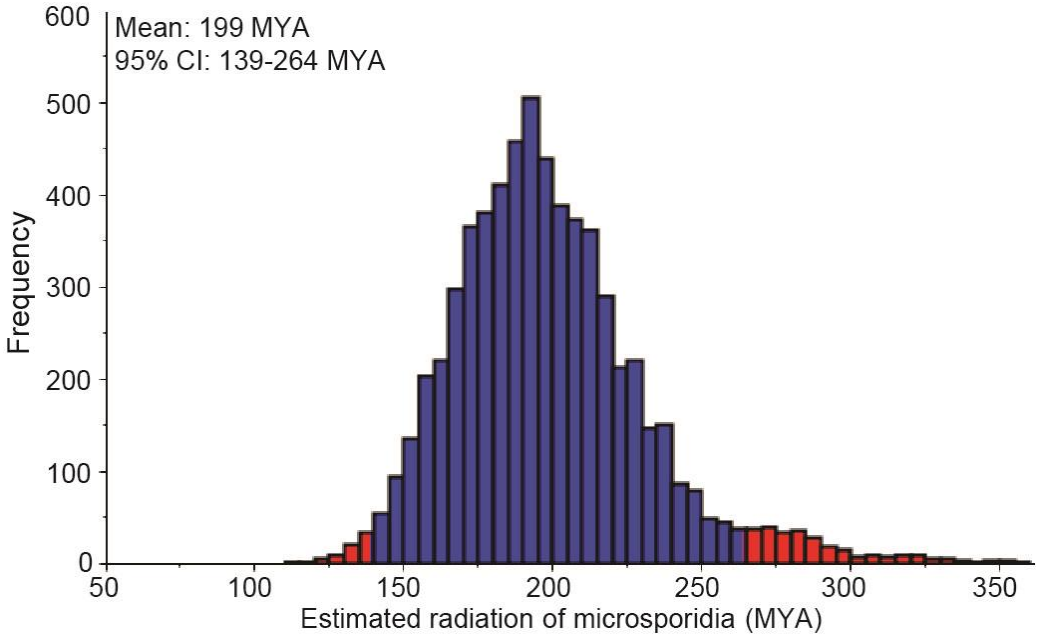


Figure 6.10| Graphical representation for the estimated date of microsporidian radiation.

Frequency output of BEAST v. 1.6.1 MCMC chains assuming no prior topology constraints were plotted using Tracer v. 1.4. The mean for the microsporidian radiation was estimated at 199 MYA with 95% confidence this event occurred between 139-264 MYA (blue graph bars).

6.3.4| Prediction of functional domains and secretion of Cht1 orthologs

Divergent Cht1 orthologs from both prokaryotes and eukaryotes all possess a glycoside hydrolase family-19 domain (GH19, Pfam). Family-19 chitinases show no sequence or structural similarity to family-18 chitinases and even have a different catalytic mechanism producing alpha anomeric products compared to beta anomers produced by family-19 enzymes (Brameld *et al*, 1998, Brameld and Goddard, 1998). In microsporidia the functional GH19 domain is present towards the C-terminus of the protein suggesting the N-terminus may function in protein targeting.

Despite the ubiquitous presence of the GH19 domain in Cht1 orthologs, predicted secretion patterns are more inconsistent (Table 6.2). Secretion prediction data from SignalP4.1 and TargetP1.1 for Cht1 orthologs from representative taxa of main clades from phylogenetic analysis were tabulated. The results of the prediction algorithms vary even between relatively closely related species such as *Aedes aegypti* and *N. vitripennis*, potentially suggesting that Cht1 orthologs may have specific and diverse functional roles in each species. Among microsporidian Cht1 orthologs, the chitinase is predicted to be secreted in species *E. cuniculi*, *E. intestinalis* and *Vittaforma corneae* and not secreted in *E. bieneusi*, *N. parisii*, *Nematocida* sp. or *V. culicis floridensis* (*A. locustae* was not included in this analysis as the EST data did not include the N-terminus). There appears to be no obvious link between Cht1 secretion prediction and lifecycle or host range, although *E. cuniculi*, *E. intestinalis* and *V. corneae*, species for which Cht1 is predicted to be secreted extracellularly, all have the ability to infect and cause disease in humans.

Table 6.2| Secretion and Pfam domain prediction profile from taxa representing all clades of Cht1 phylogenetic analysis.

Species	Clade	SignalP 4.0	Smax value	TargetP 1.1	SP Value
E. cuniculi	Microsporidia	Secreted	0.881	Secreted	0.677
E. intestinalis	Microsporidia	Secreted	0.89	Secreted	0.679
E. bieneusi	Microsporidia	Not Secreted	0.113	Not Secreted	0.054
N. vitripennis	Insecta	Secreted	0.966	Secreted	0.956
A. aegypti	Insecta	Not Secreted	0.212	Not Secreted	0.16
M. tardigradum	Eumetazoa	Not Secreted	0.128	Not Secreted	0.043
P. capsici	Chromalveolata (I)	Not Secreted	0.301	Not Secreted	0.035
K. brevis	Chromalveolata (II)	Not Secreted	0.131	Not Secreted	0.05
O. sativa	Archaeplastida (I)	Secreted	0.991	Secreted	0.991
A. thaliana	Archaeplastida (I)	Secreted	0.953	Secreted	0.893
G. max	Archaeplastida (II)	Secreted	0.93	Secreted	0.936
S. lividens	Prokaryota	Secreted	0.926	Secreted	0.823
S. roseum	Prokaryota	Secreted	0.986	Secreted	0.964
Reinekea.sp.	Prokaryota	Secreted	0.928	Secreted	0.795

6.4| Discussion

Cht1 represents the only chitinase identified in sequenced microsporidian genomes to date, however, its role in microsporidian development remains unclear. The conservation of Cht1 in distantly related microsporidian species with a wide host range suggests an essential function, particularly given the widespread loss of metabolic enzymes during the reductive evolution of these parasites (Sobottka *et al*, 2002). Chitinases have been characterised in a range of organisms including plants, fungi and bacteria with a variety of functional roles such as cell wall remodelling, virulence and defence from invading pathogens. Based on the general biology of microsporidia and the properties of the protein, two plausible functional models for Cht1 are proposed (Figure 6.2). Firstly, Cht1 is secreted to the thick chitinaceous endospore during germination to allow efficient polar tube expulsion or at another stage of morphological development such as meront to sporont differentiation, where expression of Cht1 has been shown to be up-regulated (Ronnebaumer *et al*, 2006). Figure 6.1 demonstrates chitin degradation/rearrangement at site of polar tube expulsion in germinated *E. cuniculi* spores. However, without direct molecular evidence in the form of immuno-localisation of the protein to the site of polar tube expulsion this function cannot be attributed conclusively to Cht1. Secondly, it is possible that Cht1 may be secreted into the extracellular environment, either directly into the host cell cytoplasm or via the parasitophorous vacuole in vacuole-dwelling species. Once exposed to the host cell environment Cht1 may act on a specific host substrate where by resultant by-products may be scavenged by the parasite for metabolism. Interestingly, both of these functions will require secretion through the ER and tubular microsporidian Golgi, which has been show to connect directly to the spore wall (Beznoussenko *et al*, 2007).

However, Cht1 orthologs in microsporidian species *E. bieneusi*, *N. parisii*, *Nematocida* sp. and *V. culicus floridensis* are not predicted to be secreted. These hypotheses can be constrained by experiments demonstrating the narrow substrate range of Cht1, as over expression studies of the gene in *S. cerevisiae* BY4741a show no defects in growth or changes in cell wall morphology on solid or liquid growth medium (Figures 6.5 and 6.6). It seems unlikely that the role or significance of Cht1 in microsporidian development will be completely determined with the lack of genetic tools such as gene knockout or knockdown available at present.

The developmental lifecycle of the Microsporidia has been studied in detail in a variety of different species and hosts (Wittner and Weiss, 1999). Despite this morphological characterisation of the distinct microsporidian lifecycle stages (germination, sporogony, merogony);, little is known as to how the cell controls this development at the molecular level or the stimuli that induce it. Chitin is a major component of the microsporidian endospore which is encapsulated by an exterior proteinaceous exospore (Schottelius *et al*, 2000). In contrast chitin appears to be lacking in the second intracellular lifecycle stage, the merogony (Figure 6.3). To date the mechanism of chitin synthesis, remodelling and degradation in microsporidia is uncharacterised, however understanding this mechanism will be crucial in understanding microsporidian lifecycle development.

HGT has had a significant impact on the evolution of the Microsporidia, providing species with novel tools which can be deployed in the host-parasite arms race. Microsporidia have acquired a range of genes from bacterial donors,

thus providing novel means for microsporidian species to scavenge metabolites and resist host and environmental stresses. More recently genes have been identified in *E. hellem*, *E. romaleae* and *N. bombycis* which have been acquired by HGT from insect donors (Selman *et al*, 2011, Selman and Corradi, 2011, Pan *et al*, 2013). However, Cht1 is the first putatively animal-derived transfer present in all studied microsporidian genomes. The syntenic location of Cht1 relevant to other genomic markers and a microsporidian tree topology which is congruent with recent studies such as the concatenated protein phylogeny produced by Capella-Gutierrez *et al* provide no evidence that Cht1 has been spread between microsporidian species by HGT (Capella-Gutierrez *et al*, 2012). Robust phylogenetic analysis including reduced alphabet analysis to minimise the effect of fast evolving sites (Susko and Roger, 2007) and LogDet distance analysis to minimise compositional bias in the dataset all infer HGT from an insect donor. However, it remains unclear whether this transfer event may indicate an ancient insect host for the 'ancestral microsporidian'. This hypothesis can only be tested with improved sampling of both host and parasite lineages. This study describes a novel method of dating parasitic speciesusing HGT from a donor organism with a fossil record to provide fossil verified divergence dates at local tree nodes to calibrate a molecular clock for the recipient group of interest, in this case, the Microsporidia. BEAST molecular dating analysis using 3 fossil verified animal divergence dates suggest microsporidia radiated approximately 200 million years ago. This estimate is considerably more recent than the radiation of the fungal crown which is estimated to be at 760 MYA-1 BYA (Lücking *et al*, 2009). It is, of course, possible that there are as yet unsequenced or undiscovered microsporidian species that branch before the radiation of the lineages examined in this study.

Microsporidia therefore could have potentially diversified before the acquisition of this family-19 endochitinase gene. Such data would demonstrate that this HGT is not a marker for the wider microsporidian group; however, the distribution of genes already identified and the syntenic genomic location of Cht1 orthologs (Figure 6.8) suggest that this HGT must have occurred prior to a substantial diversification of this phylum.

Microsporidia have undergone rapid change since divergence from their opisthokont relatives (e.g. animals, fungi and choanoflagellates). This has included the acquisition of a set of microsporidian–specific innovations such as the polar tube allowing invasion of the animal cell, the jettison of around 2000 unnecessary genes, and compaction of their remaining genomic material (Krylov *et al*, 2003).This study uses a novel approach, hijacking a HGT event to date the origin and timescale to the evolution of a major parasite group, suggesting that microsporidia have undergone radical evolutionary change in a relatively short timescale. This hypothesis therefore infers that the major groups of microsporidia, representing considerable diversity in cell structure, genome composition and host range, may have radiated within the last 200 million years.

Chapter 7| Characterisation of *E. cuniculi* tyrosine phosphatase Ppt1 and the impact of infection on the host phosphoproteome.

7.1| Introduction

Protein phosphorylation is a reversible post-translational modification required for the delicate control of core cellular signalling processes (Bononi *et al*, 2011). Phosphorylation, which is maintained through a complex balance of phosphorylation and dephosphorylation events, induced respectively by kinases and phosphatases, plays a crucial role in the regulation of signalling pathways required for cell division, proliferation and apoptosis (Bauman and Scott, 2002). Sequence motif analysis suggests the human proteome contains more than 100,000 phosphorylation sites, yet due to the rapid and tightly regulated nature of phosphorylation/dephosphorylation events, only ~2,000 have been verified and characterised experimentally (Zhang *et al*, 2002). Eukaryotic protein phosphatases are functionally diverse and are incorporated into three distinct protein families, namely, serine phosphatases, threonine phosphatases and tyrosine phosphatases (Barford, 1996).

The human genome contains 107 protein tyrosine phosphatases and ~ 30 serine/threonine phosphatase genes, a minimal dephosphorylation repertoire in comparison to the 518 identified protein kinases (Venter *et al*, 2001, Alonso *et al*, 2004). This demonstrates the ability of relatively few phosphatase enzymes to control the specific dephosphorylation of 1,000s of diverse phosphoproteome substrates which are involved in a range of essential cellular functions (Shi, 2009). More specifically, protein tyrosine phosphatases (PTP) are characterised by the presence of a conserved active site motif HCXXGXXRS(T), known as the PTP signature motif, where cysteine (C) is required for catalysis (Denu and Dixon, 1998). PTPs have a diverse range of substrates from proteins to mRNA, and based on structure and specificity are

divided into four subfamilies; 1) phosphotyrosine specific PTPs 2) Dual specificity phosphatases 3) Cdc25 phosphatases and 4) Low molecular weight phosphatases (Wang *et al*, 2003). Furthermore, PTPs are also classified by their cellular localisation, either as receptor PTPs (rPTPs) containing at least one transmembrane domain or non-receptor PTPs which are predominantly cytoplasmic or secreted extracellularly (Paul and Lombroso, 2003).

rPTPs have been identified and partially characterised in a range of parasitic protists including *Trypanosoma cruzi*, *Leishmania major* and *Entamoeba histolytica* and pathogenic fungi such as *Candida albicans*, *Candida parapsilosis* and *Crytococcus neoformans* (Gomes *et al*, 2011). In the intracellular human-infective trypanosomatid species *Leishmania donovani*, rPTPs function optimally at a pH of 5-5.5, likely demonstrating adaptation to life in the acidic phagolysosome found inside macrophages (Shakarian *et al*, 2002). rPTPs have been well characterised in *Leishmania* and are implicated in several unique virulence mechanisms. Gene knock-out studies of *Leishmania* PTP (LPTP1) demonstrates that the gene is essential for amastigote survival in mice but is non-essential for promastigotes in culture, suggesting LPTP1 may represent a potential drug target against the human-infective parasite lifecycle stage (Nascimento *et al*, 2006). The receptor phosphatase of *Leishmania donovani* has been shown to inhibit production of host oxidative metabolites O_2 and H_2O_2 in human neutrophils to aid intracellular survival and provide resistance to the host immune response (Remaley *et al*, 1984, Remaley *et al*, 1985). rPTPs have also been implicated in epithelial cell adhesion in the pathogenic fungus *C. parapsilosis*, in which irreversible inhibition of the enzyme by sodium orthovanadate significantly decreased fungal adhesion (Kiffer-

Moreira *et al*, 2007, Gomes *et al*, 2011). Taken together these studies clearly demonstrate a role for phosphatases in pathogenicity and host-interaction in divergent protistan and fungal species.

As with many core protein families, the complement of microsporidian PTPs is severely reduced. Ontology-based searching of the NCBI protein database identified only 2 non-related PTPs in *Encephalitozoon cuniculi* and *Nematocida parisii* Ertm1 and 3 PTPs in *Trachipleistophora hominis*, far fewer than are present in higher fungi such as *Saccharomyces cerevisiae* which contains ~10 PTP enzymes (Altschul *et al*, 1997, Cherry *et al*, 1998). PTPs retained in microsporidia may represent essential genes that may function in lifecycle progression, cellular development or host interactions. This study partially characterises a PTP in the human-infective microsporidian parasite *E. cuniculi* (EcPpt1, GenBank Accession: CAD26663) and investigates the effect of *E. cuniculi* infection on the host cell phosphoproteome and nitric oxide production. EcPpt1 is a predicted secreted protein with an estimated size of 22.6 kDa, which possesses 4 transmembrane domains and has a mitochondrial targeting signal located internally of the signal peptide cleavage site, in what is a predicted extracellular protein domain (Bjellqvist *et al*, 1994, Claros and Vincens, 1996, Krogh *et al*, 2001, Petersen *et al*, 2011). Speculatively, this protein may be involved in anchoring host mitochondria around the site of intracellular parasite development, a phenomenon which has previously been described by a morphological analysis of *Encephalitozoon* infection in the host (Scanlon *et al*, 2004). Alternatively, EcPpt1 may be involved in *E. cuniculi* adhesion or in disruption of host cell signalling pathways as a means to aid development and proliferation. EcPpt1 orthologs (BLASTp $<1e^{-5}$) were found in

all microsporidia for which there is a complete genome sequence and a diverse range of eukaryotes including Mammals, Fungi, Algae and Diplomonads (Altschul *et al*, 2005). Interestingly, orthologs were only identified in two closely related bacterial species; *Marinobacter manganoxydans* and *Marinobacter adhaerens*. The ubiquitous presence of Ppt1 in microsporidia, including extremely reduced species such as *Encephalitozoon hellem* and *Encephalitozoon intestinalis* suggests the gene may have an essential function. *S. cerevisiae* ortholog Phs1 ($6e^{-11}$, 30% amino acid identity) is an essential ER protein involved in long chain fatty acid biosynthesis (Yu *et al*, 2006, Kihara *et al*, 2008). The ability of EcPpt1 to complement Phs1 in *S. cerevisiae* is tested in this study by conducting a functional complementation assay in tetracycline-regulated Phs1 knock-out strain (Phs1Δ). The localisation of EcPpt1 is identified in three different cellular systems. Firstly, EcPpt1 is localised *in situ* in *E. cuniculi* infected rabbit kidney cells (RK-13) using a specific polyclonal antibody raised against two immunogenic EcPpt1 peptides. Secondly, EcPpt1 is tagged with C-terminal green fluorescent protein (GFP), expressed under the Gal1 promoter and localised in *S. cerevisiae* alongside a Phs1 control to determine if EcPpt1 and *S. cerevisiae* Phs1 share the same cellular localisation. *S. cerevisiae* Phs1 has previously been shown to localise to the ER (Huh *et al*, 2003). Lastly, EcPpt1 is localised in the mammalian cell line IMCD-3 using monoclonal anti-his antibodies against the C-terminal his-tag expressed in pcDNA3.1D/V5-His-TOPO.

Recently it was shown that microsporidian hexokinase proteins encode a secretion signal. This secretion signal is predicted to direct the protein extracellularly where it is hypothesised to catalyze host glycolysis to increase

host ATP production, which, can subsequently by scavenged by the parasite (Cuomo *et al*, 2012). To determine the effect of *E. cuniculi* on the host cell phosphoproteome, quantitative tandem mass tag mass spectrometry is conducted in conjunction with phosphoprotein purification to identify the phosphoproteome of infected and non-infected IMCD-3 cells. It is possible that host proteins displaying increased or decreased phosphorylation may be targeted and disrupted by the respective action of parasite kinases and phophatases or that these changes in the phosphorylation levels represent key host-cell signaling events induced in response to parasite infection. Overall, the main aims of this chapter are to 1) Localise EcPpt1 and determine if the protein is likely to function in host mitochondrial aggregation and 2) Identify the impact of *E. cuniculi* on the host phosphoproteome.

7.2| Methods

7.2.1| Cloning of EcPpt1 and *S. cerevisiae* Phs1 for expression in *S. cerevisiae*

EcPpt1 and *S. cerevisiae* Phs1 were amplified from genomic DNA extracted from *E. cuniculi* GB-M1 spores and *S. cerevisiae* BY4741a cells with respective primer sets EcPpt1GateF/EcPpt1GateR and ScPhs1GateF/ScPhs1GateR using Phusion high fidelity DNA polymerase (New England Biolabs). Thirty cycles of the following PCR conditions were employed following an initial 30 second denaturation step at 98^{o}C; 98^{o}C for 30 seconds, 52^{o}C for 30 seconds, 72^{o}C for 1 minute, completed by a single final extension step of 72^{o}C for 10 minutes. The resultant PCR products were run on 1% TAE agarose gels and gel purified as described in Chapter 2.5. Purified Ppt1 and Phs1 PCR products were ligated into yeast gateway expression vectors pAG426Gal (for functional complementation) and pAG426Gal-EGFP (for GFP localisation) by LR clonase (Invitrogen) mediated att-overhang homologous recombination in the following reaction: 50 ng PCR product, 150 ng (1 µl), 4 µl TE buffer and 2 µl of LR clonase for 1 hour at 25^{o}C (Alberti *et al*, 2007). Following a 1 hour incubation, 1 µl of Proteinase K was added and the reaction was incubated at 37^{o}C 10 minutes. 2 µl of each reaction was then transformed into 50 µl of *E. coli* OneShot Top10 cells (Invitrogen) under both ampicillin and bacterial death gene (ccdB) selection. Insert presence and integrity in resultant colonies was verified by colony PCR. Verified constructs EcPpt1/pAG426Gal, *S. cerevisiae* Phs1/pAG426Gal, EcPpt1/pAG426Gal-EGFP and *S. cerevisiae* Phs1/pAG426Gal-EGFP were then transformed into *S. cerevisiae* Phs1Δ and plated on CSM-ura agar as described in Chapter 2.8.

7.2.2| Cloning of EcPpt1 for expression in IMCD-3 cells

EcPpt1 was amplified using primer set EcPhs1MaF and EcPhs1MaR from genomic DNA using Phusion high fidelity DNA polymerase (New England Biolabs) following the manufacturer's instructions and a primer annealing temperature of 52^{o}C. The resultant PCR amplicon was analysed by gel electrophoresis on a 1% TAE agarose gel and purified as described in Chapter 2.5. Purified product was then ligated to pcDNA3.1D/V5-His-TOPO (Invitrogen) using topoisomerase mediated ligation of CACC in the following reaction for 30 minutes at room temperature; 5 ng EcPpt1, 5 ng (1 µl) pcDNA3.1D/V5-His-TOPO, 1 µl TOPO salt solution and 2 µl $_{d}H_{2}0$. 2 µl of ligation mix was then transformed into 50 µl *E. coli* OneShot Top10 cells (Invitrogen) as described in Chapter 2.6. Positive colonies were confirmed by PCR, restriction digest (BamHI/XhoI) and sequencing. Verified construct and empty vector negative control vector was then transfected into IMCD-3 cells as described in Chapter 2.10.

7.2.3| Functional complementation and GFP localisation of EcPpt1 in *S. cerevisiae*

S. cerevisiae BY4741a containing pYES2 (+ve wild type control) and *S. cerevisiae* tetracycline-regulated Phs1 knock-out strain (Phs1Δ) containing constructs *S. cerevisiae* Phs1/pAG426Gal, EcPpt1/pAG426Gal and pAG426Gal (-ve control) were grown for overnight in YPD broth at 30^{o}C at 180 rpm. Cultures were then pelleted by centrifugation at 1,200 x g and washed twice in sterile 1 x PBS. Cell concentration was determined by manual counts using a hemocytometer and all cultures were adjusted to a concentration of 1 x 10^{8} cells/ml. The standardised cultures were serially diluted and spotted onto YPD agar control medium and YPG medium containing doxycycline manually using a

replicator. Galactose was required for recombinant expression in order to induce the pYES2 Gal1 promoter. Doxycycline, a tetracycline derivative, was required in the media to switch off native *S. cerevisiae* Phs1 function, which is an essential gene, via the tetracycline-regulated $TetO_7$ promoter (Mnaimneh *et al*, 2004). Plates were incubated at 30^{o}C for 48 hours.

S. cerevisiae BY4741a containing constructs Phs1/pAG426Gal-EGFP and EcPpt1/pAG426Gal-EGFP were grown overnight in 5 ml YPD broth in a shaking incubator at 30^{o}C and 180 rpm. Subsequently, cultures were pelleted by centrifugation at 1,200 x g and washed twice in sterile 1 x PBS. Cells were then resuspended in CSM-Ura+galactose, to switch on gene expression via induction of the Gal1 promoter and grown at 30^{o}C and 180 rpm for 4-6 hours. Cells were then pelleted by centrifugation, washed twice and resuspended in 1 x PBS. Cells were mounted and visualised using the Olympus IX-81 inverted fluorescence microscope under DIC and fluorescence at an excitation/emission wavelength of 488/510 nm.

7.2.4| Anti-EcPpt1 polyclonal antibody production and verification of serum specificity

Anti-EcPpt1 polyclonal antibodies were raised in New Zealand White Rabbits (hosts #6083 and #6084) by Pacific Immunology against the immunogenic peptide sequences EAANISAKKSNSRYL and SRKVKAYKSNRKDL coupled to BSA carrier protein. Polyclonal antibody production was conducted over a 13 week program with immunization boosters administered on weeks 1, 3, 6 and 10. The initial peptide immunization was administered in Complete Freunds Adjuvant with all subsequent immunizations given in Incomplete Freunds Ajuvant. Production bleeds occurred throughout the program on weeks 7, 9, 11

and 12 with final bleeds occurring on week 13. The specificity of the anti-EcPpt1 serum from the final bleed from both rabbits was tested by western blot against protein from *E. cuniculi* spores, germinated *E. cuniculi* spores*, S. lophii* spores and from IMCD-3 cells. A total of 40 µg of protein was loaded per sample. Western blots were conducted as described in Chapter 2.12 at 1:500 dilutions for primary antibody treatment followed by a 1:2000 dilution of goat anti-rabbit IgG-HRP conjugated secondary antibody (Invitrogen).

7.2.5| Localisation of EcPpt1 *in situ* in *E. cuniculi* infected RK-13 cells

RK-13 cells heavily infected with *E. cuniculi* (gift from Martin Embley, University of Newcastle) were maintained continuously in DMEM/F-12 + 10% FBS as described in Chapter 2.9. Prior to immuno-localisation infected cells were passaged and seeded on to sterile coverslips in 6-well tissue culture plates. Infected cells were grown on coverslips for 48 hours and subsequently fixed and permeabilised in 1:1 acetone:methanol as described in Chapter 2.13. Fixed slides were blocked and treated with pre-immune and final bleed anti-serum at 1:500 dilutions in 1% milk/PBS for 1.5 hours. Cells were subsequently treated with Goat anti-rabbit IgG (H+L) HiLyte Fluor-488 antibody (Anaspec) at 1:1000 dilution in 1% milk/1 x PBS in the dark at room temperature for 1.5 hours. DAPI was added to cells at a final concentration of 1 µg/ml to stain nuclei during the final 15 minutes of secondary antibody incubation. Cells were mounted in Vectashield mounting medium and visualised using the Olympus IX-81 inverted fluorescence microscope. Images captured were converted to 8-bit image files and overlaid using Metamorph Image Software (Metamorph). The specificity of serum was verified by western blot as detailed in Chapter 2.12.

7.2.6| Localisation of EcPpt1 in IMCD-3 cells

IMCD-3 cells were passaged and seeded onto sterile coverslips and grown in 2 ml Opti-MEM medium until between 70-90% confluent. Cells were transfected with 2 µg of both pcDNA3.1D/V5-His-TOPO and EcPpt1/pcDNA3.1D/V5-His-TOPO using Lipofectamine 2000 (Invitrogen). Cells were grown at 37^{o}C and 5% CO_2 for 48 hours to ensure transient expression under the constitutive CMV promoter and subsequently fixed/permeabilised with parafomaldehyde and Triton X-100 respectively, as described in Chapter 2.13. Cells were blocked and treated with mouse anti-his primary antibody at 1:1000 dilution for 1 hour at room temperature followed by secondary antibody treatment with Goat anti-mouse IgG_1 (γ_1) Alexa Fluor-568 (Invitrogen) for 1.5 hours and DAPI stain. Cells were then mounted with Vectashield mounting medium and visualised using Olympus IX-81 microscope to verify a consistent localisation of EcPpt1 in comparison to empty vector negative control.

7.2.7| Phosphoproteomic analysis of *E. cuniculi* infected IMCD-3 cells

Protein was extracted from 3 replicates of both IMCD-3 cells and *E. cuniculi* infected IMCD-3 cells (~30% infection prevalence) as described in Chapter 2.4 using RIPA lysis buffer (4 ml/ml of cells) and Halt protease and phosphatase inhibitors (1/100 dilution) (Thermo scientific). Protein was quantified using a nanodrop spectrophotometer (Thermo scientific) and adjusted to a concentration of 4 mg/ml using RIPA buffer and 3 kDa exclusion protein concentration columns (Millipore). The integrity and relative concentration of the 6 protein samples was verified by SDS-PAGE as described in Chapter 2.12. Samples were then frozen at -20^{o}C and sent to Dr Kate Heesom, University of Bristol Proteomics Facility for tandem mass tagging, phosphoprotein

enrichment, nano LC-MS/MS identification and Mascot database searching against a pooled *E. cuniculi* and mouse proteome (Thompson *et al*, 2003). For TMT labeling each protein sample was labeled with a MS/MS isobaric reporter tag of unique mass (126-131 Da) using the Thermo Scientific TMT^6 mass tagging kit. Subsequently samples were pooled and enriched using TiO_2 phosphoprotein enrichment and clean up kit (Thermo Scientific). Phosphoproteins in the three biological replicates of both IMCD-3 and *E. cuniculi* infected IMCD-3 cells were then identified quantitatively using the specific reporter tag identified by LC-MS/MS analysis.

7.2.8| List of oligonucleotides

Table 7.1| Names and sequences of oligonucleotides used in this study.

Oligo name	Sequence (5'- 3')	Cloning method
EcPpt1GateF	**CCAACTTTGTACAAAAAAGCAGG CT**ATGGCGTGTCCAAGTTATATC	Gateway
EcPpt1GateR	**CCAACTTTGTACAAGAAAGCTGG GT**TAAATCTTTCCTGTTGCT	Gateway
ScPhs1GateF	**CCAACTTTGTACAAAAAAGCAGG CT**ATGTCAAAAAAACTTGCGTCA	Gateway
ScPhs1GateR	**CCAACTTTGTACAAGAAAGCTGG GT**AATTAGTTTCTTCCCGAA	Gateway
EcPhs1MaF	**CACC**ATGGCGTGTCCAAGTTATATC	TOPO cloning
EcPhs1MaR	TAAATCTTTCCTGTTGCTCTT	TOPO cloning

Gateway recombination sequences and TOPO ligation overhang displayed in bold.

7.3| Results

7.3.1| Generation of EcPpt1 and *S. cerevisiae* Phs1 expression constructs

Prior to experimental characterisation EcPpt1 and *S. cerevisiae* Phs1 were amplified by PCR using gene specific sets EcPpt1GateF/EcPpt1GateR and ScPhs1GateF/ ScPhs1GateR respectively. The resultant PCR products were run on 1% TAE agarose gels and size verified (EcPpt1 = 579 bp, *S. cerevisiae* Phs1 = 654 bp, Figure 7.1). PCR amplicons were gel purified, ligated into gateway cloning vectors pAG426Gal and pAG426Gal-EGFP by LR clonase mediated homologous recombination and transformed to *E. coli* cells under selection from bacterial death gene (ccdB) and ampicillin (Alberti *et al*, 2007). *E. coli* colonies were checked for insert presence by colony PCR using gene specific primers (Figure 7.2). Sequence verified constructs were transformed into *S. cerevisiae* under uracil selection.

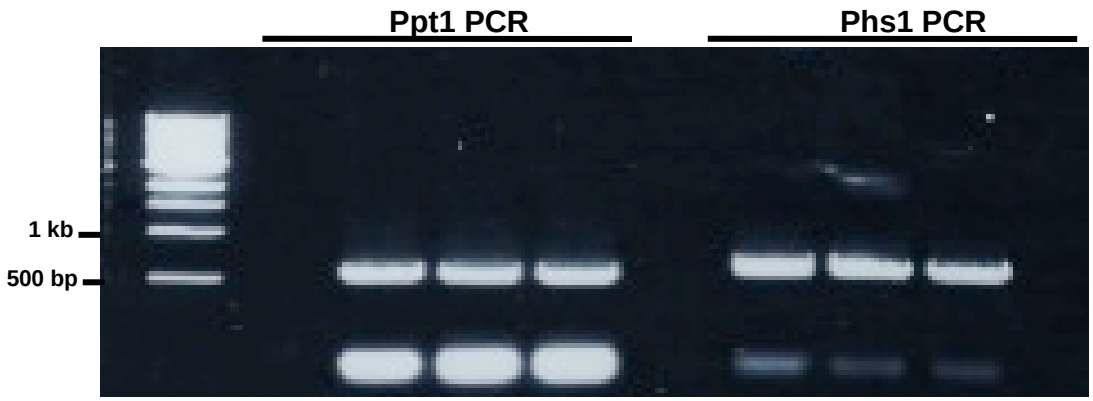


Figure 7.1| PCR amplification of EcPpt1 and *S. cerevisiae* Phs1 genes.
PCR amplification of EcPpt1 (579 bp) and *S. cerevisiae* Phs1 (654 bp) with gene specific primer sets EcPpt1GateF/R and ScPhs1GateF/R for integration into cloning vectors using gateway mediated *in vitro* recombination. The size and specificity of gene amplicons was verified on 1% TAE agarose gels.

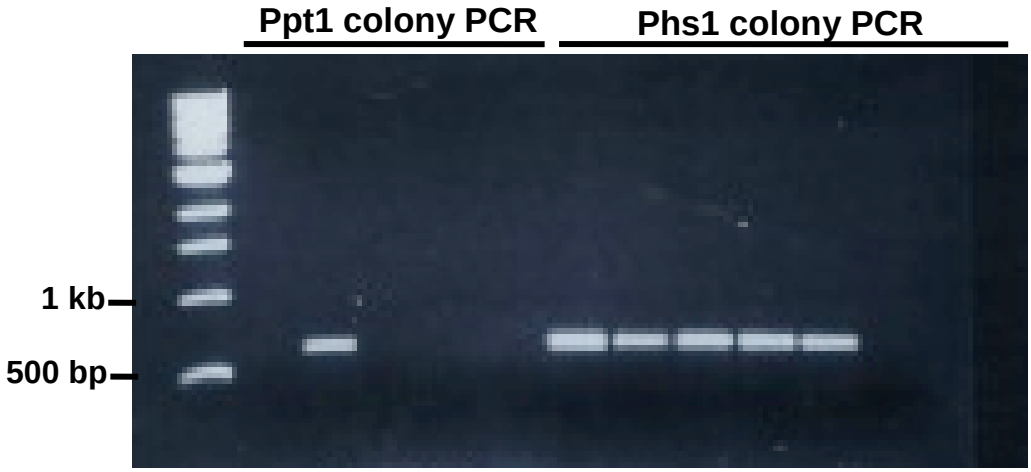


Figure 7.2| Confirmation of EcPpt1 and *S. cerevisiae* Phs1 clones.
The presence of EcPpt1 and *S. cerevisiae* Phs1 inserts in vector pAG426Gal determined by colony PCR using gene specific primers. PCR products were verified by sequencing (MWG Eurofins).

7.3.2| Functional complementation and GFP localisation of EcPpt1 in *S. cerevisiae*

BLAST search analysis of EcPpt1 identified a potential ortholog, Phs1, in *S. cerevisiae* ($6e^{-11}$, 30% identity) (Altschul *et al*, 2005). Phs1 has previously been described as an essential gene in *S. cerevisiae* with an ER and vacuolar localisation (Yu *et al*, 2006, Kihara *et al*, 2008). *S. cerevisiae* Phs1 is thought to function in sphingolipid and phosphate metabolism (Dickson *et al*, 2006). Sphingolipids are among the most prevalent eukaryotic lipids and are required for maintaining membrane structure and regulating diverse signaling pathways (Futerman and Riezman, 2005). Phs1 has been shown to provide resistance to high levels of phytosphingosine and dihydrosphingosine, however the specific function of the protein in sphingolipid metabolism is not known (Schuldiner *et al*, 2005).

To determine if EcPpt1 has the same functional role as *S. cerevisiae* Phs1 in sphingolipid metabolism, a functional complementation was carried out in *S. cerevisiae* tetracycline-regulated Phs1 knock-out strain (Phs1Δ). Due to the essential nature of *S. cerevisiae* Phs1, stress conditions were not required to induce a growth phenotype and functional complementation assay was carried out on YPD and YPG media for repression/expression of cloned constructs. Functional complementation reveals that EcPpt1 is not capable of rescuing the function of *S. cerevisiae* Phs1 once native Phs1 expression has been switched off with the addition of doxycycline to YPG media, displaying growth comparable with Phs1Δ (Figure 7.3A). Recombinantly expressed *S. cerevisiae* Phs1 does partially complement native gene function but still shows a slightly reduced growth phenotype compared to that of wild type *S. cerevisiae*

strain BY4741a, which displays optimal growth under both conditions (Figure 7.3A). Phs1Δ shows no growth in the presence of doxycycline demonstrating the requirement of Phs1 for viability. These data suggest that EcPpt1 and *S. cerevisiae* Phs1 may have evolved different functional roles and that EcPpt1 is not capable of complementing the function of Phs1 in sphingolipid metabolism. Alternatively, it is possible EcPpt1 it is not correctly folded or targeted in the heterologous expression system. To determine if EcPpt1 and Phs1 share the same cellular localisation, the two genes were expressed with a C-terminal GFP tag in plasmid pAG426Gal-EGFP and localised by fluorescence microscopy in *S. cerevisiae* (Figure 7.3B). The GFP signal shows that EcPpt1 and *S. cerevisiae* Phs1 display different sub-cellular localisations. Phs1 shows an internal localisation consistent with the ER localisation described previously (Kihara *et al*, 2008). EcPpt1 shows a more peripheral localisation suggesting the protein may be targeted to the cell membrane when expressed in *S. cerevisiae*.

This difference in localisation supports functional complementation data and suggests that EcPpt1 and *S. cerevisiae* Phs1 perform different cellular functions and, perhaps, that the sequence similarity between the two proteins is a result of conserved amino acids within the shared tyrosine phosphatase domain. Interestingly, *E. cuniculi* lacks phosphosphingosine phosphatase and phosphosphingosine lyase genes, which are known to interact with Phs1 (Cherry *et al*, 1998, Dickson *et al*, 2006). *E. cuniculi* appears to have undergone either severe reduction in sphingolipid metabolism or adopted a non-canonical sphingolipid metabolism pathway. It is therefore possible that EcPpt1 has evolved a new cellular function, unrelated to sphingolipid biosynthesis or

metabolism. Idiosyncrasies in *E. cuniculi* lipid metabolism have previously been described, for example acetyl-CoA-carboxylase, biotin-[acetyl-CoA-carboxylase]-ligase and fatty acid synthase are all absent from the *E. cuniculi* genome (Katinka *et al*, 2001, Akiyoshi *et al*, 2009, Campbell *et al*, 2013). Acetyl-CoA-carboxylase and biotin-[acetyl-CoA-carboxylase]-ligase are present in the genomes of *T. hominis*, *N. parisii* and *Spraguea lophii*, however *S. lophii* is known to lack a fatty acid synthase gene (Cuomo *et al*, 2012, Heinz *et al*, 2012, Campbell *el al*, 2013) . This suggests lipid metabolism pathways are amenable to adaptation in microsporidia, a process perhaps driven by the unique host cell environment occupied by each parasitic species.

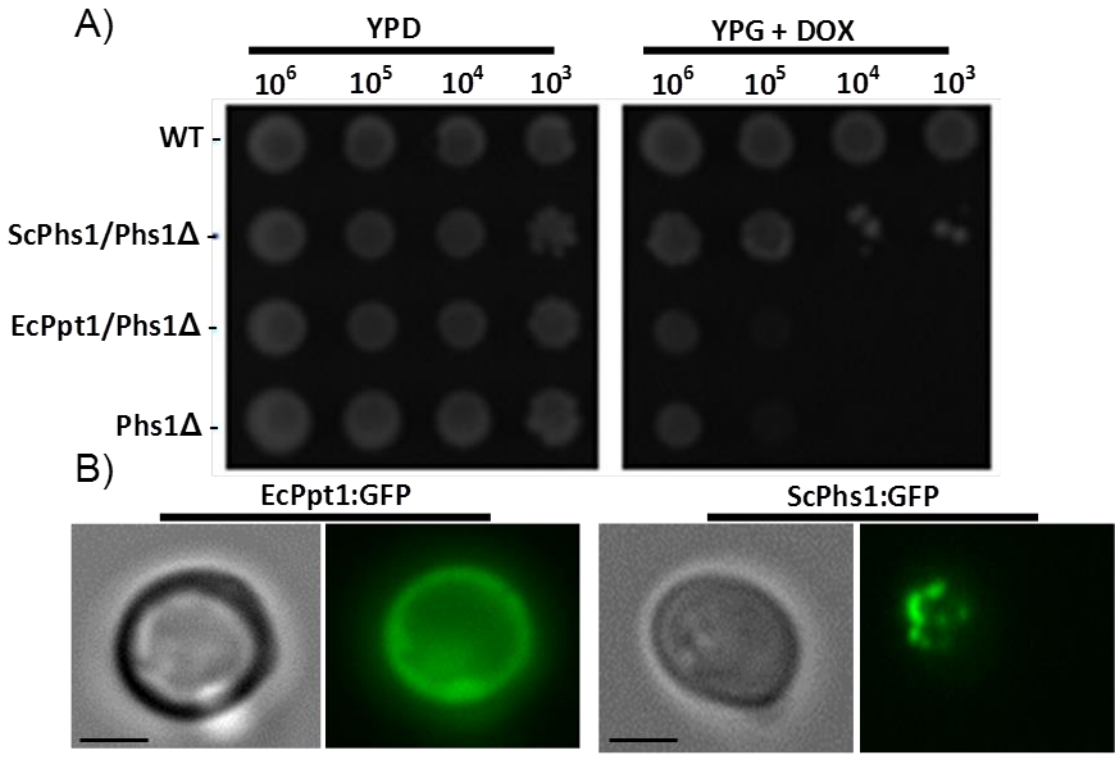


Figure 7.3| Functional complementation of EcPpt1 in tetracycline-regulated *S. cerevisiae* Phs1Δ.

A) *S. cerevisiae* wild type strain BY4741a (WT), Phs1Δ strain YJL097w, Phs1Δ complemented with native Phs1 gene and Phs1Δ complemented with EcPpt1 were grown for 48 hours on both standard yeast peptone dextrose media (YPD) and yeast peptone galactose (YPG) media containing doxycycline. All strains grow consistently on standard YPD. Under doxycycline and hence Phs1 knock-out conditions, there is a clear reduced growth phenotype in the *S. cerevisiae* strain expressing EcPpt1, suggesting EcPpt1 cannot complement *S. cerevisiae* Phs1 function. B) GFP-localisation of EcPpt1 and *S. cerevisiae* Phs1 in *S. cerevisiae*. EcPpt1 displays a peripheral localisation compared to the ER localisation of Phs1, as described in (Huh *et al*, 2003). This suggests that EcPpt1 and *S. cerevisiae* Phs1 occupy different cellular localisations in their native cells (scale bar = 2µm).

7.3.3| EcPpt1 localises to the cell membrane *in situ* and in mammalian cells

Polyclonal antibodies raised in rabbit against two immunogenic 15 amino acid peptides coupled to BSA (Peptides 1) SRKVKAYKSNRKDL and 2) EAANISAKKSNSRYL - synthesized by Pacific Immunology) were used to localise EcPpt1 in *E. cuniculi* infected rabbit kidney cell line RK-13. Anti-serum from final production bleed (host #6083), which was taken on week 13 of the immunization program, was used as primary antibody treatment. This was followed by goat anti-rabbit IgG HiLyte Fluor-488 secondary antibody treatment to determine the cellular localisation of EcPpt1, producing a green fluorescence signal upon excitation with light at a wavelength of 488 nm. Immuno-fluorescence microscopy reveals a specific signal for EcPpt1 at the cell membrane of both *E. cuniculi* spores and meronts (Figure 7.4A). DAPI stain was used to stain host and parasite nuclei to provide a point of reference within the cell. This experimental localisation is in agreement with sequence-based cellular prediction programs such as TargetP1.1 which suggest that EcPpt1 is targeted to the secretory pathway (probability = 0.951) and TMHMM2.0 which predicts the presence of 4 transmembrane domains are responsible for docking EcPpt1 to the parasite extracellular membrane (Krogh *et al*, 2001, Emanuelsson *et al*, 2007). Cumulatively, the data suggests that EcPpt1 is a receptor PTP (rPTP), a protein family which act in a highly regulated antagonistic manner against tyrosine kinases to balance global cellular tyrosine phosphorylation (van der Wijk *et al*, 2005). This membrane localisation provides no evidence that EcPpt1 is involved in host mitochondrial accumulation despite identification of a mitochondrial targeting signal located internally of the predicted signal peptide (Claros and Vincens, 1996, Scanlon *et al*, 2004). The specificity of anti-EcPpt1

was determined by western blot against *E. cuniculi* proteins (Figure 7.4B). The blot clearly shows a single protein band matching the predicted 22.6 kDa size of EcPpt1 (Bjellqvist *et al*, 1994). EcPpt1 is expressed in both germinated spores and non-germinated spores with no cross reactivity with protein from the microsporidian species *S. lophii* or mouse kidney epithelial cells.

Screening of pre-immune serum by immuno-fluorescence microscopy in *E. cuniculi* infected RK-13 cells revealed no specific or consistent signal (Figure 7.5). This suggests the response identified in final bleed serum is specific to EcPpt1 peptide antigens and not a prior antigen that the individual had inadvertently or previously been exposed to. To verify the localisation of EcPpt1 the protein was expressed and localised in mouse kidney epithelial cell line IMCD-3 (Figure 7.6). In agreement with *in situ* studies, localisation of EcPpt1 in the mammalian expression system also identified a membrane/peripheral localisation (Figure 7.6, Red). This membrane localisation of EcPpt1 in IMCD-3 cells demonstrates that EcPpt1 targeting is conserved between microsporidia, yeast and mammals and provides accumulating evidence that the membrane localisation identified *in situ* is accurate and that EcPpt1 is likely a rPTP. As a negative control, IMCD-3 cells were transformed with empty vector pcDNA3.1D/V5-His-TOPO and treated with mouse anti-his monoclonal to determine if there is any unspecific binding of the antibody to native IMCD-3 proteins (Figure 7.7). Immuno-fluorescence microscopy revealed no specific antibody signal against IMCD-3 cells (Figure 7.7, Red) suggesting the membrane localisation signal is specific to the c-terminal his-tag expressed with EcPpt1. This study describes the first phosphatase localised in any microsporidian species to date. The reduced genome of *E. cuniculi* encodes

only 1 additional PTP (GI: 19068749) which is predicted to localise to the nucleus and cytoplasm, hence it seems likely, that *E. cuniculi* possesses one rPTP and one non-receptor PTP which together are responsible for global tyrosine dephosphorylation (Horton *et al*, 2007). Future analysis such as identification of EcPpt1 protein interaction partners will be required to determine both the functional role and signaling pathways in which EcPpt1 functions and whether the protein has any role is disrupting host cell signaling.

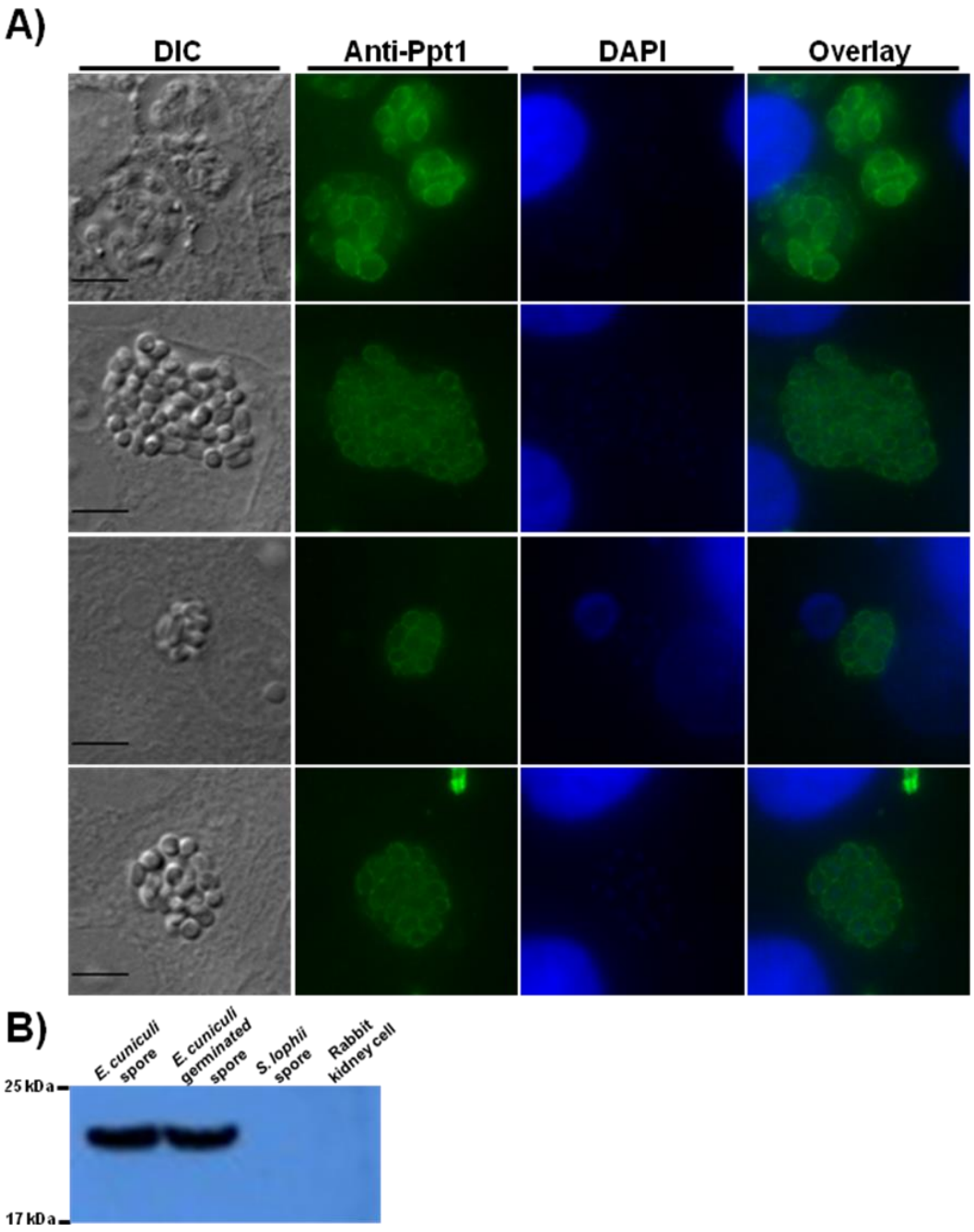


Figure 7.4| Immuno-localisation of EcPpt1 in *E. cuniculi* infected RK-13 cells.

A) *E. cuniculi* meronts and spores growing in RK-13 cells show specific staining with rabbit anti-serum containing polyclonal antibodies raised against immunogenic peptides from EcPpt1 (green). B) Western blot of whole protein extracts from purified *E. cuniculi* spores, germinated *E. cuniculi* spores, *S. lophii* spores and from IMCD-3 cells. A strong protein band corresponding to ~22.5 kDa and the predicted size of EcPpt1 was identified in both dormant and germinated *E. cuniculi* spores. (scale bar = 5 µm).

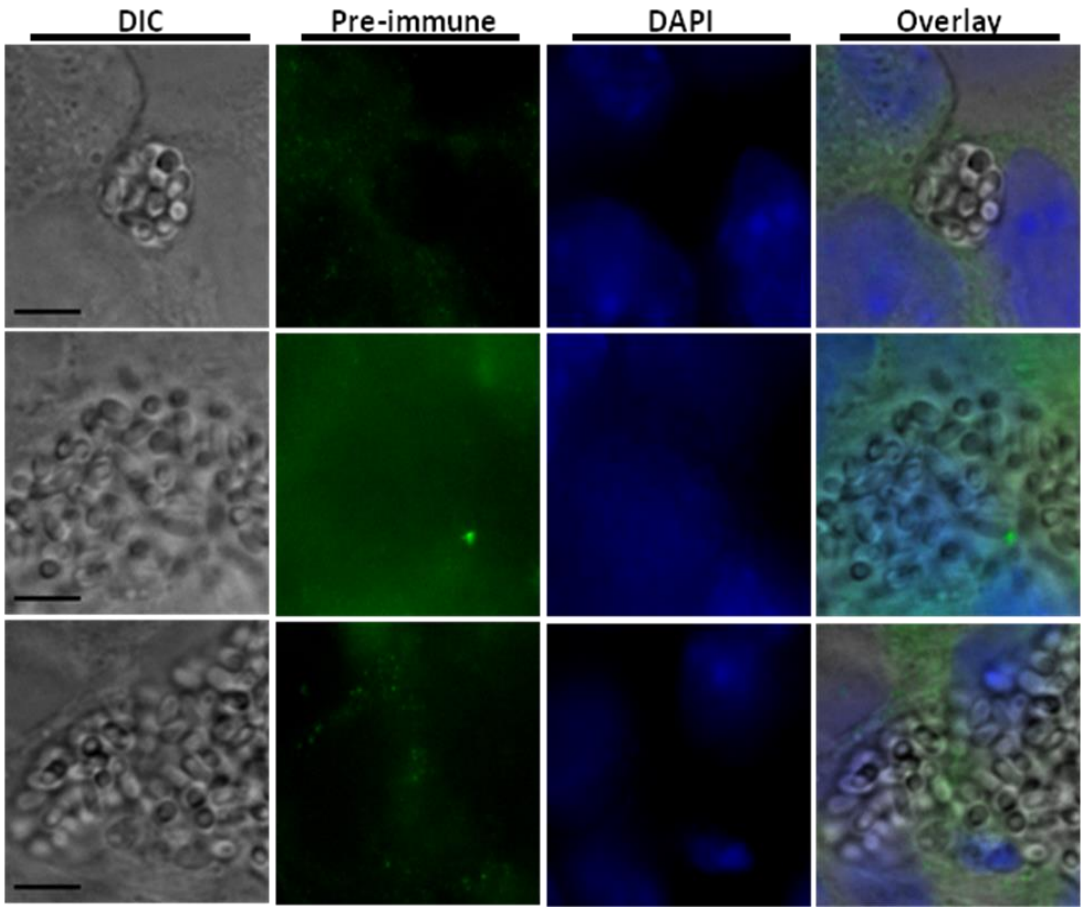


Figure 7.5| Screening of pre-immune serum against *E. cuniculi* infected RK-13 cells.

No consistent or specific reactivity is evident in *E. cuniculi* spores or meronts in RK-13 cells after treatment with rabbit pre-immune serum (green). DAPI (blue) was used to stain host and parasite nuclei. This suggests that final bleed anti-serum reaction with *E. cuniculi* is specific to EcPpt1 peptide challenge (scale bar = 5 µm).

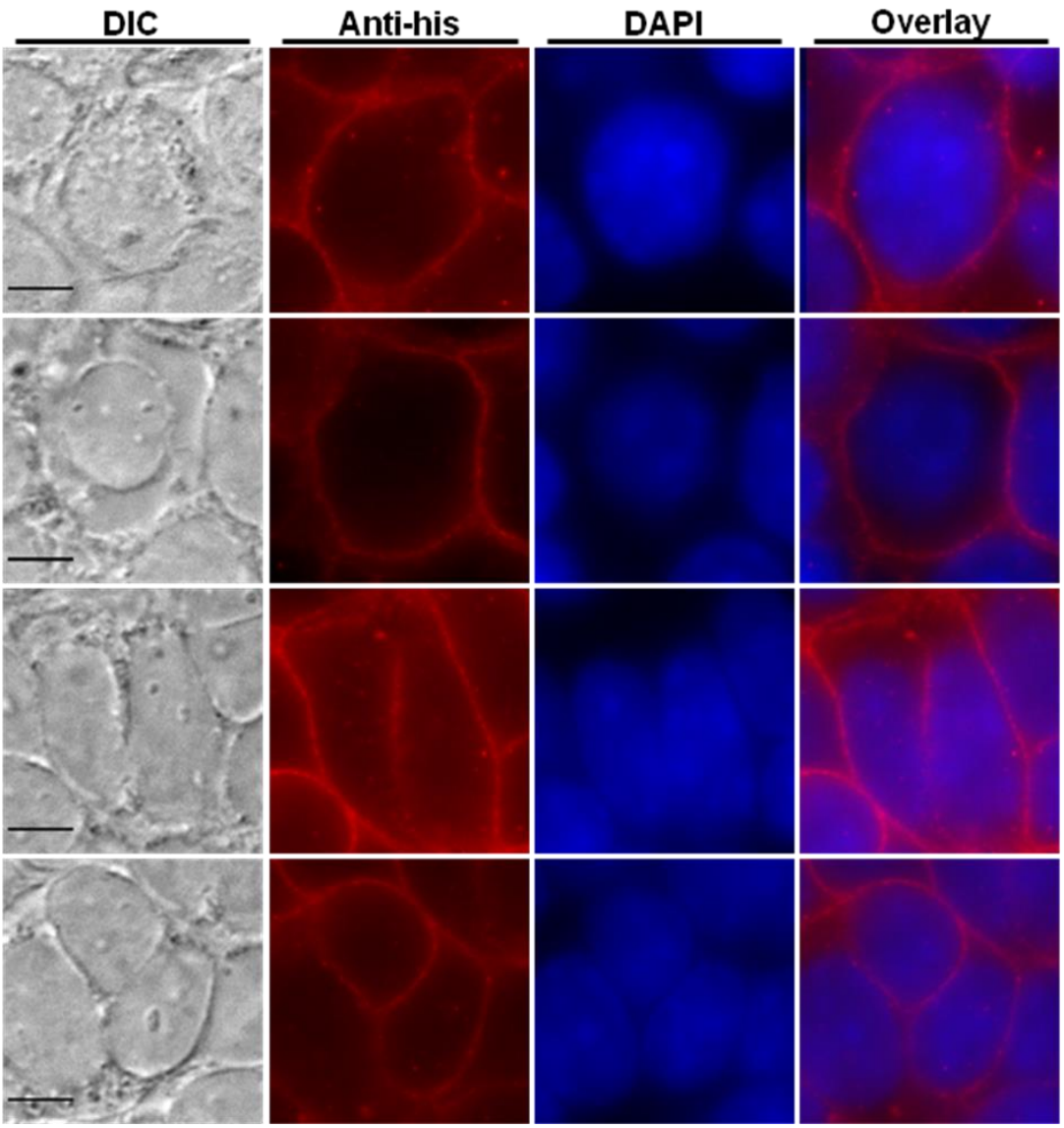


Figure 7.6| Immuno-localisation of EcPpt1 in mammalian cell line IMCD-3. EcPpt1 was cloned in mammalian expression vector pcDNA3.1D/V5-His-TOPO, transfected into IMCD-3 cells and expressed with a C-terminal His-tag. Monoclonal anti-his antibodies (red) were used as primary treatment, displaying a peripheral or membrane localisation for EcPpt1. DAPI stain was used to stain nuclei as a reference (blue) (scale bar = 5 µm).

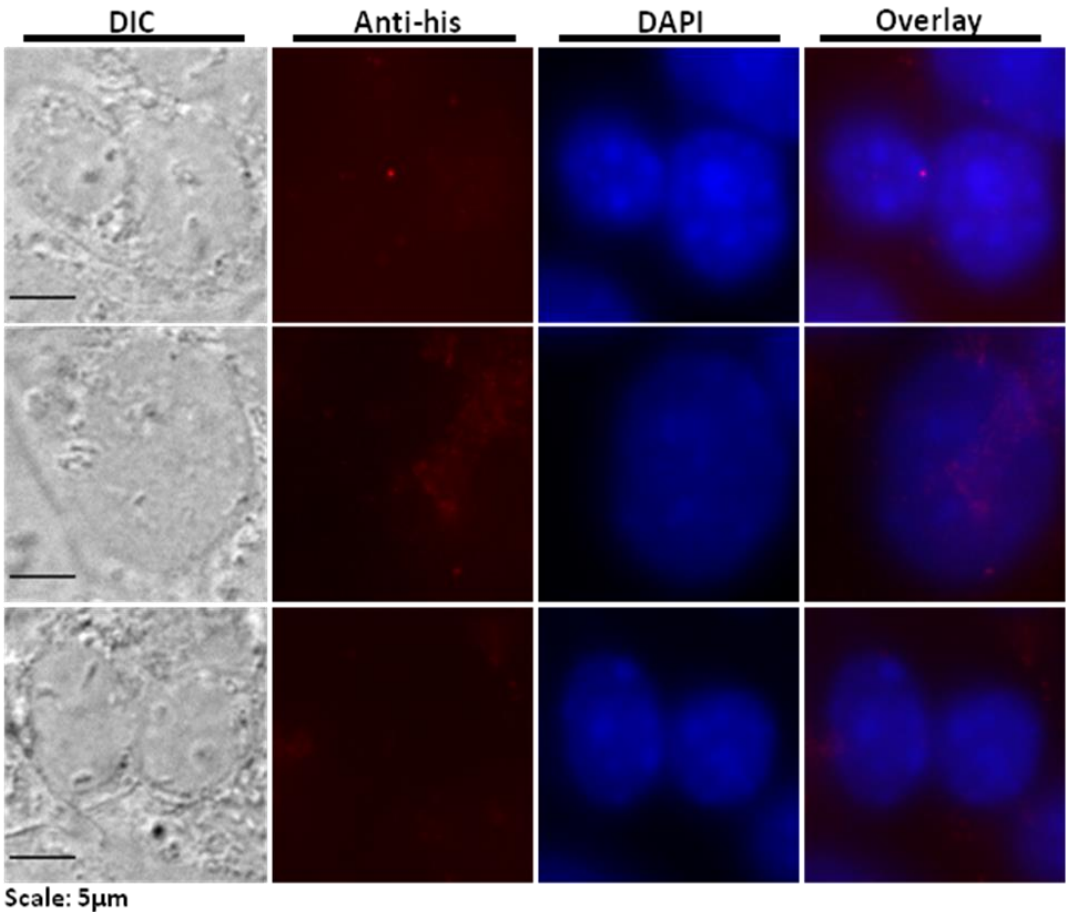


Figure 7.7| Screening anti-his antibody against mammalian cell line IMCD-3 transfected with empty expression vector pcDNA3.1D/V5-His-TOPO.
IMCD-3 cells were transfected with empty expression vector pcDNA3.1D/V5-His-TOPO and probed with anti-his antibodies, as a negative control (red). No specific staining of IMCD-3 cells was identified (red). DAPI was used to stain cell nuclei (blue).

7.3.4| The impact of *E. cuniculi* infection on the host cell phosphoproteome

To determine the impact of *E. cuniculi* infection on the host cell phosphoproteome total cell protein was extracted from 3 biological replicates of both non-infected and *E. cuniculi* infected IMCD-3 cells. Protein concentration was standardised and checked on a 12% SDS-PAGE gel. To quantify changes in phosphoproteome abundance, standardised protein samples were sent to the University of Bristol Proteomics Facility for tandem mass tag mass spectrometry analysis and phosphoprotein enrichment (Thompson *et al*, 2003). Mass spectrometry analysis and database searching of peptides against a pooled *E. cuniculi* and *M. musculus* proteome identified a total of 846 proteins conserved across all replicates (3 infected and 3 non-infected) of which 844 were mouse proteins and 2 were hypothetical proteins belonging to *E. cuniculi* (GI: 19074080 and 449329807, Appendix 3). The presence of so few *E. cuniculi* proteins in phosphoproteome analysis is likely explained by the > 35 times larger *M. musculus* proteome swamping the M/S analysis. Based on mean abundance across the 3 biological replicates for each lifecycle stage, a total of 38 mouse proteins showed increased phosphorylation during *E. cuniculi* infection and 806 proteins displayed decreased phosphorylation (Appendix 3).

The list of 38 proteins displaying increased phosphorylation during *E. cuniculi* infection includes many core protein families including histones, transcription factors and zinc finger proteins (Table 7.2.). Nucleolar protein Esf1 displays the greatest increase in phosphorylation during *E. cuniculi* infection, showing a 2.4 fold increase in abundance following phosphoprotein enrichment. Esf1 has been characterised previously in *S. cerevisiae* where it has been

implicated in 18S rRNA biogenesis, pre-rRNA processing and may represent a co-factor for RNA helicase mediated release of small nucleolar RNA (Peng *et al*, 2004, Bleichert and Baserga, 2007). Gene ontology analysis identified mouse proteins with predicted functions in zinc ion binding and RNA binding displayed the greatest increase in phosphorylation during *E. cuniculi* infection (Figure 7.8). RNA binding proteins, such as those containing zinc finger domains, are known to contribute to a range of core cellular process including transcriptional regulation, mRNA stability, differentiation and apoptosis (Ravasi *et al*, 2003). Parasitic manipulation of host processes such as apoptosis may be crucial for the optimisation of proliferation, egress and the spread of infection to adjacent cells. Approximately 35% of the 30 proteins showing the greatest increase in phosphorylation during *E. cuniculi* infection are predicted to localise specifically to the nucleolus (Figure 7.9). The nucleolus is an organelle that regulates and coordinates the assembly of ribosomal subunits and plays important roles in cell growth, proliferation and cell cycle regulation (Visintin and Amon, 2000, Andersen *et al*, 2005). Microsporidia of the genus *Encephalitozoon* have been shown previously to arrest the host cell cycle (Scanlon *et al*, 2000). It is possible this cell cycle disruption is due to the secretion of parasite proteins which manipulate the nucleolar phosphoprotein balance.

A total of 806 proteins show decreased phosphorylation during *E. cuniculi* infection (for top 30 see Table 7.3, for full list see Appendix 3). Speculatively, these proteins may represent the targets of secreted or cell membrane located parasite phosphatases as a means of interfering with host signaling pathways. Among the proteins showing the greatest decrease in

phosphorylation include; 1) a cadherin-6 protein known to function in cell-cell adhesion 2) a component of the kinesin light chain which functions in cellular cargo transport and is required for a range of core processes including meiosis and mitosis and 3) an SH_3 –ankyrin repeat family protein (Table 7.3) (Paul *et al*, 1997, Vale *et al*, 2010). Gene ontology analysis of the predicted molecular function of the 30 host proteins showing the greatest decrease in phosphorylation during *E. cuniculi* infection suggests down-regulated phosphoproteins predominantly function in DNA binding (Figure 7.8) (Conesa *et al*, 2005). However, several proteins among those showing the greatest decrease in phosphorylation still lack functional annotation and localisation confirmation. Down-regulated phosphoproteins are predicted to predominantly localise to the nucleus and plasma membrane in IMCD-3 cells (Figure 7.9). It is, therefore possible, that microsporidian agents or pathogenicity factors disrupt host nuclear function to optimise proliferation or plasma membrane components to aid egress and entry to adjacent cells. Taken together, these data demonstrate that *E. cuniculi* infections severely alter the host cell phosphoproteome and that disruption of the host cell phosphorylation-dephosphorylation balance by parasite kinases and phosphatases may represent a mechanism for optimisation of intracellular conditions for lifecycle development.

Table 7.2| Mouse proteins displaying increased phosphorylation during *E. cuniculi* infection.

Uniprot Accession	Annotation	Ratio
Q3V1V3	ESF1 homolog	2.434
A2AWT6	Nucleolar transcription factor 1	1.927
Q8CI11	Guanine nucleotide-binding protein-like 3	1.917
Q5U4C1	G-protein coupled receptor-associated sorting protein 1	1.838
D3YZJ1	Sequestosome-1	1.815
Q9DAY9	Nucleophosmin	1.739
G3UYA6	Yorkie homolog	1.602
Q03145	Ephrin type-A receptor 2	1.499
Q9JIK5	Nucleolar RNA helicase 2	1.452
Q91VE6	MKI67 FHA domain-interacting nucleolar phosphoprotein	1.445
P09405	Nucleolin	1.431
P53986	Monocarboxylate transporter 1	1.423
Q7TPV4	Myb-binding protein 1A	1.374
P14873	Microtubule-associated protein 1B	1.322
Q80XU3	Nuclear ubiquitous casein	1.295
O08784	Treacle protein	1.278
F6XMS4	Protein Pus7	1.242
D3Z2T9	Cyclin-dependent kinase 1	1.211
P39689	Cyclin-dependent kinase inhibitor 1	1.208
P62806	Histone H4	1.204
Q9DCE5	p21-activated protein kinase-interacting protein 1	1.201
F8WI35	Histone H3	1.146
E9Q616	Protein Ahnak	1.129
Q8C0S5	CAP-Gly domain-containing linker protein 1	1.116
Q0VEE6	Zinc finger protein 800	1.109
P10711	Transcription elongation factor A protein 1	1.097
A2BE92	Protein SET	1.074
Q9CYA6	Zinc finger CCHC domain-containing protein 8	1.063
P15919	V(D)J recombination-activating protein 1	1.062
Q6DFW4	Nucleolar protein 58	1.049
F8WIX8	Histone H2A	1.039
Q91VY9	Zinc finger protein 622	1.032
P47856	Glutamine--fructose-6-phosphate aminotransferase 1	1.028
Q99LL5	Periodic tryptophan protein 1 homolog	1.021
E9QPW8	Microtubule-associated protein	1.017
B1AU75	Nuclear autoantigenic sperm protein	1.002
Q7TPE5	Probable RNA polymerase II nuclear localization protein	1.002

Table 7.3| Top 30 mouse proteins displaying the greatest decrease phosphorylation during *E. cuniculi* infection.

Uniprot Accession	Annotation	Ratio
P97326	Cadherin-6	0.18
D3Z710	Kinesin light chain 3	0.2
D3Z5K9	SH3 and multiple ankyrin repeat domains protein 2	0.22
Q5SXA9	Protein KIBRA	0.23
E9QAX1	Brain-specific angiogenesis inhibitor 1-associated protein 2	0.24
P98203	Armadillo repeat protein	0.26
Q9QYC0	Alpha-adducin	0.26
G3V011	Unconventional myosin-If	0.26
Q80X35	Microtubule-associated protein	0.27
F7AR27	Protein Fam98c	0.27
Q8C5R2	Uncharacterised protein C10orf47 homolog	0.28
Q8VBT6	Apolipoprotein B receptor	0.28
E9Q4K7	Protein Kif13b	0.28
P05627	Transcription factor AP-1	0.29
P46062	Signal-induced proliferation-associated protein 1	0.29
Q8R0W0	Epiplakin	0.29
Q91VW5	Golgin subfamily A member 4	0.3
Q99JP6	Homer protein homolog 3	0.3
A2AB59	Rho GTPase-activating protein 27	0.31
E9Q3E2	Synaptopodin	0.31
Q6PDN3	Myosin light chain kinase, smooth muscle	0.31
G3UZU0	DNA (cytosine-5)-methyltransferase 3A	0.33
Q7TT50	Serine/threonine-protein kinase MRCK beta	0.33
E9Q0S6	Protein Tns1	0.34
F7AL76	Chromodomain-helicase-DNA-binding protein 8	0.34
B7ZC23	Nuclear receptor coactivator 5	0.34
Q99J21	Mucolipin-1	0.36
H3BK79	Nucleobindin-1	0.36
P27889	Hepatocyte nuclear factor 1-beta	0.37
E9Q8P5	PDZ and LIM domain protein 5	0.4
P62996	Transformer-2 protein homolog beta	0.46

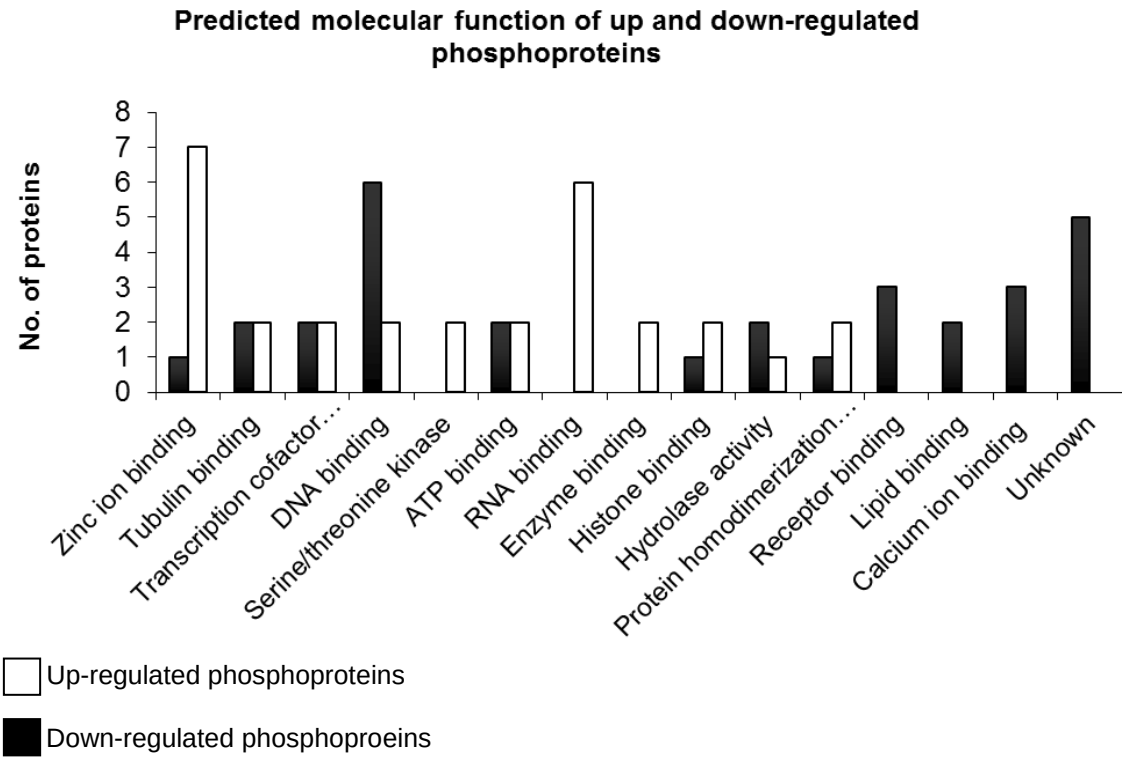


Figure 7.8| Molecular function summary of top 30 proteins displaying the greatest decrease and increase in phosphorylation during *E. cuniculi* infection.

The molecular function of mouse proteins showing increased and decreased phosphorylation during *E. cuniculi* infection was predicted using BLAST2GO. Proteins involved in zinc ion binding and DNA binding displayed the greatest respective phosphorylation increase and decrease during *E. cuniculi* infection.

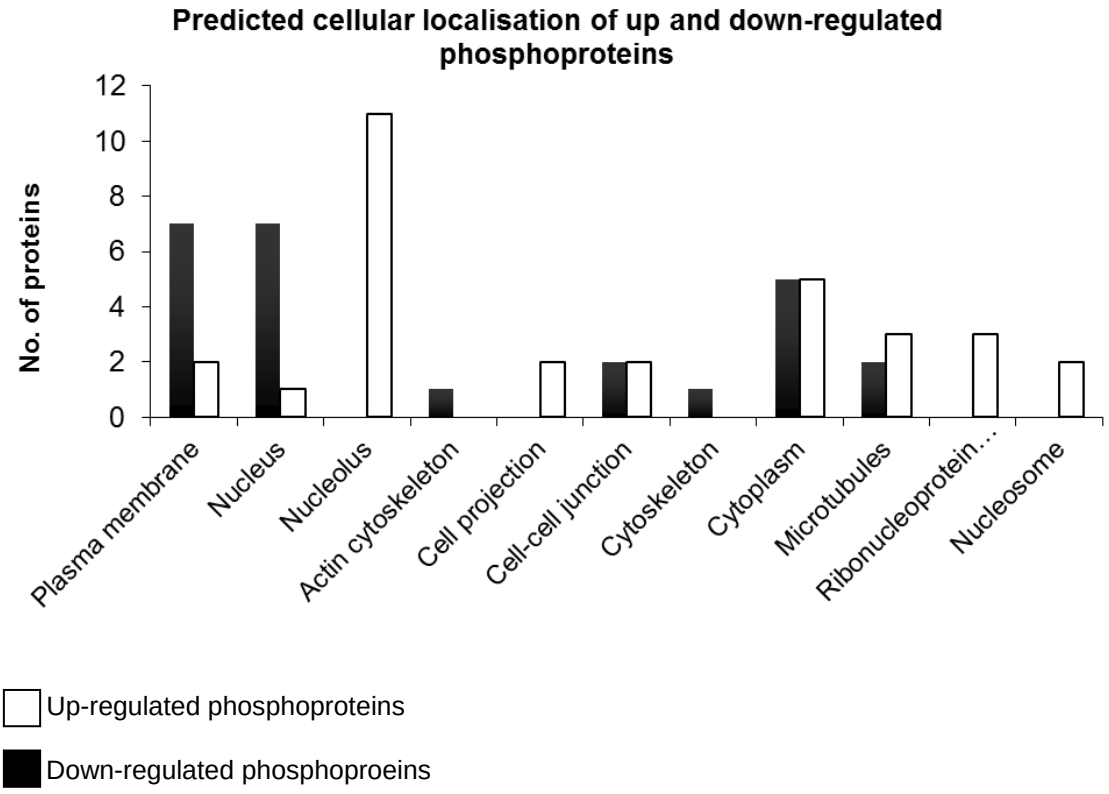


Figure 7.9| Predicted cellular localisation of top 30 proteins displaying the greatest decrease and increase in phosphorylation during *E. cuniculi* infection.

The cellular localisation of mouse proteins showing increased and decreased phosphorylation during *E. cuniculi* infection was predicted using BLAST2GO. Proteins localising to the nucleolus and cytoplasm showed the largest increase in phosphorylation during *E. cuniculi* infection. In contrast, plasma membrane and nucleus showed the most significant decrease in phosphorylation.

7.4| Discussion

Phosphorylation is a key reversible post-translational modification required for the regulation of protein-protein interactions, protein subcellular localisation and protein degradation. In eukaryotes the event predominantly occurs at serine, threonine and tyrosine residues (Delom and Chevet, 2006). Phosphorylation is regulated by the intimate balance between the ATP dependent action of protein kinases and the antagonistic action of protein phosphatases. These protein families act together as a switch mechanism activating and deactivating a range of complex cellular processes (Hunter, 1995). It is estimated that at least 30% of all proteins contain covalently bound phosphate, demonstrating the importance of protein phosphorylation in the homeostatic control of core cellular processes including the cell cycle and energy metabolism (Ficarro *et al*, 2002).

The intricate control of phosphoproteins required for cellular signalling can be exploited by pathogens, which in some cases manipulate the host with non-native kinases or phosphatases. A recent study by Cuomo *et al* 2012 identified a microsporidian hexokinase family protein that has acquired an N-terminal signal peptide capable of directing the protein to the extracellular environment *in vitro*. The authors hypothesise that this microsporidian protein could mimic host hexokinase function and boost the production of metabolites which could subsequently be scavenged by the parasite (Cuomo *et al*, 2012). Parasite phosphatases as agents capable of host cell manipulation are best described in intracellular parasites *Leishmania, T. cruzi* and *T. gondii*. These parasites use secreted or membrane phosphatases to aid adherence and internalisation, disrupt host MAPK signalling and down-regulate host iNOS production to optimise intracellular survival (Martiny *et al*, 1996, Martiny *et al*,

1999, Nandan *et al*, 1999). For example, protein phosphatase 2C from intracellular apicomplexan parasite *T. gondii* is secreted from rhoptries during invasion and is targeted directly to the host cell nucleus where it can disrupt the host cell command centre (Gilbert *et al*, 2007).

Bioinformatic analysis identified a total of 32 genes encoding protein kinases in the genome of *E. cuniculi*, of which 30% where predicted to regulate the cell cycle and 28% had no identifiable homolog in any non-microsporidian eukaryote (Miranda-Saavedra *et al*, 2007). In contrast, little is known about the complement or function of protein phosphatases present in the *E. cuniculi* genome. In this study an *E. cuniculi* protein tyrosine phosphatase, termed EcPpt1, is localised and partially characterised. The protein localises to the cell membrane both *in situ* and in mammalian cells, and shows a peripheral localisation in yeast. This strongly suggests that EcPpt1 is a receptor tyrosine phosphatase, which is targeted to the membrane consistently in three unique and diverse opisthokont systems. Future studies will be required to elucidate the complete cellular function of EcPpt1, however due to the transmembrane nature of the protein and expression in both extracellular and intracellular lifecycle stages, a possible role in adherence, host manipulation or intravacuolar signalling can not be overlooked. The retention of predicted EcPpt1 orthologs in diverse and reduced microsporidian species suggests the protein has an essential cellular function. Sequence similarity searching using BLASTp suggested *S. cerevisiae* Phs1, a protein involved in sphingolipid metabolism was a potential ortholog of EcPpt1. However, EcPpt1 could not complement Phs1 function suggesting either that the homology between the two proteins was due to conserved sequence properties of their tyrosine phosphatase

domains or that EcPpt1 has acquired a new functional role since the divergence of microsporidia from their fungal relatives. It is also possible that EcPpt1 does not fold and function properly in *S. cerevisiae* cells. Interestingly, *E. cuniculi* also lacks the Phs1 interaction partners; phosphosphingosine phosphatase and phosphosphingosine lyase, suggesting *E. cuniculi* maintains a non-canonical sphingolipid biosynthesis pathway (Cherry *et al*, 1998, Dickson *et al*, 2006).

E. cuniculi infection causes severe change in the host phosphoproteome. Mass spectrometry analysis identified a total of 38 proteins displayed increased phosphorylation and 844 proteins showed decreased phosphorylation during *E. cuniculi* infection. These data suggest that *E. cuniculi* infection has a huge impact on host cell signalling, however the extent to which this is a host response to infection or a parasite-induced host manipulation cannot be determined from this assay. A significant proportion of up-regulated host phosphoproteins (~35%) were predicted to function in the nucleolus. It is possible that *E. cuniculi* secretes effectors to specifically target the host nucleolus as a means of disrupting the host cell cycle, a phenotype that has previously been described in *E. cuniculi* infection (Scanlon *et al*, 2000). No host nuclear targeting or retention signals have been identified in *E. cuniculi* predicted secreted proteins to date. However effectors targeting the host nuclei have been described in other pathogens including oomycetes and bacteria (Bozkurt *et al*, 2012). Host proteins displaying increased and decreased phosphorylation have a diverse range of predicted functional roles including DNA, zinc-ion and ATP binding. This demonstrates that *E. cuniculi* infection of the host causes a global disruption to signalling and does not appear to be directed to a small subset of specific pathways or processes. Subsequent

experimental analysis treating *E. cuniculi* with kinase or phosphatase inhibitors, such as PTP inhibitor sodium orthovanadate, may help decipher the extent to which these parasite enzymes disrupt the host phosphoproteome balance (Hecht and Zick, 1992).

In summary, EcPpt1 localises to the cell membrane in both *in situ* and in divergent expression systems. Despite a reasonable degree of sequence similarity, EcPpt1 cannot complement the functionality of *S. cerevisiae* Phs1 and this suggests that the apparent homology of the two proteins is misleading and based on conserved properties of the tyrosine phosphatase domain, which is common to both proteins. *E. cuniculi* infection severely alters the host cell phosphoproteome, a crucial signalling component in many core cellular pathways. Subsequent research will be required to determine the extent to which this altered phosphoproteome is a result of host manipulation by parasite kinases or phosphatases or a host-mediated response to infection. Microsporidia possess a reduced complement of kinases and phosphatases thus providing a minimal platform to investigate the role of these important enzymes in the interaction with the host cell.

Chapter 8| General Discussion

8.1| General Discussion

Microsporidia are a diverse and increasingly important group of obligate intracellular parasites known to cause disease in immuno-competent and immuno-compromised humans as well as economically important animal species, including trout and salmon (McConnachie *et al*, 2013). Microsporidian research has been ongoing for over 150 years since the discovery of *Nosema bombycis*, the causative agent of pébrine disease, a condition responsible for the collapse of the European silk industry (Nägeli, 1857). Since that time, the morphology, phylogeny, cell biology and physiology of microsporidia has been studied in some detail. Early hypotheses suggested that the Microsporidia were an early branching 'primitive' eukaryotic phylum which lacked mitochondria (Cavalier-Smith, 1983). However, genomic advances led to the sequencing of the *Encephalitozoon cuniculi* genome which demonstrated the presence of a range of conserved mitochondrial markers (Katinka *et al*, 2001). Improved phylogenetic models and subsequent studies suggested that the Microsporidia were most probably related to Fungi (Hirt *et al*, 1999, Keeling *et al*, 2000, Van de Peer *et al*, 2000). Mitochondrial Hsp70 was then localised to tiny double membrane bound reduced mitochondria, termed mitosomes, thus demonstrating definitively that microsporidia were not primitive 'amitochondriate' eukaryotes but in fact, highly reduced parasites primed for life in the host intracellular environment (Williams *et al*, 2002).

Genomic and cellular investigations have more recently indentified a huge range in microsporidian genome size, from 2.3-24 Mb, but the presence of a surprisingly conserved proteome size (Keohane *et al*, 1996, Delbac *et al*, 2001, Corradi *et al*, 2009, Corradi *et al*, 2010). Several microsporidian-induced

host cell manipulations have been described morphologically including the accumulation of host mitochondria around the site of infection and the disruption of the host cell cycle (Scanlon *et al*, 2000, Scanlon *et al*, 2004). Despite this, little is known about the virulence factors required for host cell invasion or the determinants of microsporidian pathogenicity in the intracellular environment, particularly at the molecular level (Campbell *et al*, 2013). Following spore germination and the rapid expulsion of the polar tube, microsporidia have constant access to the host cell environment throughout their developmental lifecycle, until egress, demonstrating an extremely intimate host-parasite relationship (Cali and Takvorian, 1999). Proteins secreted extracellularly and membrane proteins with extracellular domains, in this study collectively termed the secretome, represent the interface between microsporidia and their host. It therefore seems inevitable that the microsporidian secretome has a crucial function in lifecycle progression, host cell manipulation and virulence. Despite this, these proteins are poorly characterised, particularly during intracellular stages, a process that is undoubtedly hindered by the current lack of a microsporidian genetic manipulation system.

This study aims to begin addressing this omission in current understanding by identifying potentially significant secreted proteins involved in lifecycle progression or virulence by comparative bioinformatics, evolutionary analysis and experimental protein characterisation, by employing a combination of heterologous expression assays and immuno-localisation studies. The project initially employs comparative genomics and secretion prediction software to predict the ‘core’ microsporidian secretome and identify species-specific secreted proteins that may be significant in individual host-parasite

systems. It seems plausible that many predicted ‘core’ secreted or membrane proteins will perform an essential function at a particular stage or stages of microsporidian development (Katinka *et al*, 2001, Thomarat *et al*, 2004). However, it must be considered that secretion prediction programs including SignalP and TargetP are trained on model organisms. Because of this and the accelerated rate of microsporidia gene evolution, the degree of prediction accuracy in microsporidia cannot be determined and will require future experimental validation (Klee and Ellis, 2005, Emanuelsson *et al*, 2007, Campbell *et al*, 2013).

Secondly, the study focuses on identifying proteins secreted extracellularly during spore germination, the crucial lifecycle stage for host cell infection. Using *Spraguea lophii* as a ‘model’ system, due to the availability of large quantities of purifiable spores from the monkfish host, in conjunction with *in vitro* germination methods, a set of potential effector proteins were identified experimentally. Of these 37 potential effectors, 5 were conserved in three biological replicates and a further 10 were conserved in 2/3 biological replicates (Campbell *et al*, 2013). This list was enriched in *S. lophii*-specific and microsporidian-specific hypothetical proteins with no functional annotation, suggesting an important role for these proteins in host cell infection and demonstrating that species-specific secreted proteins may have evolved for infection of new host species. It is, of course possible that these *S. lophii* specific hypothetical proteins are present in other microsporidian species but that these orthologs are evolving quickly and hence, are not identifiable by conventional BLAST search analysis (Altschul *et al*, 1997). This study provides the first identification of potential effector proteins in any microsporidian species,

providing a small list of hypothetical proteins from the 1000's present in microsporidian genomes. These potential effectors proteins are strong candidates for future experimental characterisation.

Lastly, this study partially characterises three predicted secreted proteins in the human-infective microsporidian *E. cuniculi*, namely, a hemolysin, a protein tyrosine phosphatase and an endochitinase. *E. cuniculi* HlyIII displays clear pore-forming activity in heterologous expression assays suggesting a possible role in virulence or lifecycle development, where HlyIII may be involved in vacuole or host cell lysis prior to pathogen egress. Hemolysin proteins have been implicated in pathogenicity in a range of prokaryotic and eukaryotic species including *Staphlococcus aureus* and *Trypanosoma cruzi* (Wiseman, 1975, Andrews and Whitlow, 1989, Ley *et al*, 1990). However, this work shows localisation of *E. cuniculi* HlyIII both *in situ* using an anti-protein antibody and in a mammalian expression system and suggests a nuclear or nuclear membrane localisation. This novel hemolysin localisation suggests a non-canonical function that could not be fully determined in the absence of a genetic manipulation system. *E. cuniculi* protein tyrosine phosphatase was localised to the extracellular membrane both *in situ* with immuno-fluorescence microscopy using a specific anti-peptide antibody and in a mammalian expression system, suggesting the protein is a receptor protein tyrosine phosphatase. The effect of *E. cuniculi* infection on the host cell phosphoproteome was also investigated. This quantitative proteomic analysis identified that *E. cuniculi* infection caused a significant increase/decrease in the phosphorylation levels of many host proteins. Future experimental analysis will determine if these phosphoproteomic changes are the respective action of parasite kinases and phosphatases, such

as the receptor tyrosine phosphatase localised in this study, altering host signalling or are the result of the host response to infection.

The genome of *E. cuniculi*, like other microsporidia, is predicted to encode just one chitinase gene (Katinka *et al*, 2001). Previous studies investigating gene expression levels using RT-PCR have described an up-regulation on this endochitinase during meront to sporont development of *E. cuniculi* (Ronnebaumer *et al*, 2006). Despite this, the molecular function, subcellular localisation and extent to which this endochitinase contributes to microsporidian morphological development, remains unknown. The microsporidian spore consists of a proteinaceous exospore and a chitinaceous endospore, yet little is known regarding the extensive morphological rearrangements of the spore wall during germination, a process which occurs extremely rapidly (Peuvel-Fanget *et al*, 2006). Addressing this question using *in vitro* germination methods and FB-28 chitin staining identifies clear chitin degradation at the spore apical tip during germination. It remains unclear whether this degradation is due to the action of an *E. cuniculi* chitinase or is a consequence of the physical damage caused by expulsion of the polar tube. Phylogenetic analysis of this *E. cuniculi* gene, which is present in all microsporidia for which there is an available genome sequence, but is absent from genomes of ‘higher’ fungi, revealed it had been acquired by horizontal transfer from an insect donor. Two other cases of HGT from insects to members of the genus *Encephalitozoon* have been described but this case is unique as it appears to have been acquired by the ‘ancestral microsporidian’ prior to the radiation of the phylum (Selman and Corradi, 2011, Selman *et al*, 2011). This HGT provides a unique opportunity to combine phylogenetics with molecular

dating approaches and fossil-verified insect divergence dates to estimate the date of the microsporidian radiation using both molecular and geological data. The results of this study suggest that microsporidia radiated recently, likely within the last 200 million years. This demonstrates that microsporidia have acquired a unique infection mechanism, have undergone vast speciation and have hugely diversified their host range, over a short evolutionary timescale, a conclusion in direct contradiction to original theories suggesting the phylum represented ancient eukaryotes (Cavalier-Smith, 1983).

This study provides valuable evolutionary and functional insight into secreted proteins in microsporidia. Future studies targeting secreted proteins which are common to diverse microsporidian species will be required to identify proteins which function in virulence or lifecycle progression across the phylum, in species which collectively infect a diverse range of hosts. To determine the contribution of secreted proteins to microsporidian pathogenicity, either across the phylum or at the species-specific level will require characterisation of 'hypothetical' proteins, which, for example, comprise nearly 50% of the *E. cuniculi* proteome (Katinka *et al*, 2001). To add further complication, microsporidian species have undergone various gene-family expansions including the InterB family proteins in the Encephalitozoonidae, *Vittaforma corneae* and *Anncalia algerae* and leucine rich repeat proteins in *S. lophii* and *V. corneae*, of which many members are predicted to be secreted extracellularly (Dia *et al*, 2007, Campbell *et al*, 2013). It is possible that these gene family expansions contribute significantly to the microsporidian-host interaction. Little is currently known regarding the specific microsporidian agents which function in virulence or interaction with the host, despite the identification and description

of many host cell manipulations morphologically (for review see Williams, 2009). Despite the difficulties caused by the obligate nature of the microsporidian lifecycle, microsporidian research must prioritise the development of an efficient and stable genetic modification system. The advent of such a system will be required for the characterisation of secreted proteins on the scale necessary to understand the degree to which these proteins influence the microsporidian-host interaction and microsporidian virulence.

References

Abad, A., J. V. Fernandez-Molina, J. Bikandi, A. Ramirez, J. Margareto, J. Sendino, F. L. Hernando, J. Ponton, J. Garaizar and A. Rementeria (2010). What makes *Aspergillus fumigatus* a successful pathogen? Rev Iberoam Micol 27(4): 155-182.

Akiyoshi, D. E., H. G. Morrison, S. Lei, X. Feng, Q. Zhang, N. Corradi, H. Mayanja, J. K. Tumwine, P. J. Keeling, L. M. Weiss and S. Tzipori (2009). Genomic Survey of the Non-Cultivatable Opportunistic Human Pathogen, *Enterocytozoon bieneusi*. PLoS Pathog **5**(1).

Akiyoshi, D. E., L. M. Weiss, X. Feng, B. A. P. Williams, P. J. Keeling, Q. Zhang and S. Tzipori (2007). Analysis of the β-Tubulin Genes from *Enterocytozoon bieneusi* Isolates from a Human and Rhesus Macaque. Journal of Eukaryotic Microbiology **54**(1): 38-41.

Alberti, S., A. D. Gitler and S. Lindquist (2007). A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. Yeast **24**(10): 913-919.

Alexander, J., G. H. Coombs and J. C. Mottram (1998). *Leishmania mexicana* Cysteine Proteinase-Deficient Mutants Have Attenuated Virulence for Mice and Potentiate a Th1 Response. The Journal of Immunology **161**(12): 6794-6801.

Alonso, A., J. Sasin, N. Bottini, I. Friedberg, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon and T. Mustelin (2004). Protein Tyrosine Phosphatases in the Human Genome. Cell **117**(6): 699-711.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**(17): 3389-3402.

Altschul, S. F., J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schaffer and Y. K. Yu (2005). Protein database searches using compositionally adjusted substitution matrices. FEBS J **272**(20): 5101-5109.

Amzallag, N., B. J. Passer, D. Allanic, E. Segura, C. Théry, B. Goud, R. Amson and A. Telerman (2004). TSAP6 Facilitates the Secretion of Translationally Controlled Tumor Protein/Histamine-releasing Factor via a Nonclassical Pathway. Journal of Biological Chemistry **279**(44): 46104-46112.

Andersen, J. S., Y. W. Lam, A. K. L. Leung, S.-E. Ong, C. E. Lyon, A. I. Lamond and M. Mann (2005). Nucleolar proteome dynamics. Nature **433**(7021): 77-83.

Andreadis, T. G. (2005). Evolutionary strategies and adaptations for survival between mosquito-parasitic microsporidia and their intermediate copepod hosts: a comparative examination of *Amblyospora connecticus* and *Hyalinocysta chapmani* (Microsporidia: Amblyosporidae). Folia Parasitol (Praha) **52**(1-2): 23-35.

Andrews, N. W., C. K. Abrams, S. L. Slatin and G. Griffiths (1990). A *T. cruzi*-secreted protein immunologically related to the complement component C9: Evidence for membrane pore-forming activity at low pH. Cell **61**(7): 1277-1287.

Andrews, N. W. and M. B. Whitlow (1989). Secretion by *Trypanosoma cruzi* of a hemolysin active at low pH. Molecular and Biochemical Parasitology **33**(3): 249-256.

Aoe, T., E. Cukierman, A. Lee, D. Cassel, P. J. Peters and V. W. Hsu (1997). The KDEL receptor, ERD2, regulates intracellular traffic by recruiting a GTPase-activating protein for ARF1. EMBO J **16**(24): 7305-7316.

Aurrecoechea, C., J. Brestelli, B. P. Brunk, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. T. Kraemer, W. Li, J. A. Miller, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, G. Srinivasamoorthy, C. J. Stoeckert, R. Thibodeau, C. Treatman and H. Wang (2010). EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Research **38**: D415-D419.

Baba, K., S. Yamasaki, M. Nishibuchi and Y. Takeda (1992). Examination by site-directed mutagenesis of the amino acid residues of the thermostable direct hemolysin of *Vibrio parahaemolyticus* required for its hemolytic activity. Microb Pathog **12**(4): 279-287.

Bacher, G., M. Pool and B. Dobberstein (1999). The ribosome regulates the GTPase of the beta-subunit of the signal recognition particle receptor. J Cell Biol **146**(4): 723-730.

Baida, G. E. and N. P. Kuzmin (1996). Mechanism of action of hemolysin III from *Bacillus cereus*. Biochim Biophys Acta **1284**(2): 122-124.

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res **37**: W202-208.

Bantel, H., B. Sinha, W. Domschke, G. Peters, K. Schulze-Osthoff and R. U. Janicke (2001). alpha-Toxin is a mediator of *Staphylococcus aureus*-induced cell death and activates caspases via the intrinsic death pathway independently of death receptor signaling. J Cell Biol **155**(4): 637-648.

Barford, D. (1996). Molecular mechanisms of theprotein serine/threonine phosphatases. Trends in Biochemical Sciences **21**(11): 407-412.

Barlowe, C. (2003). Signals for COPII-dependent export from the ER: what's the ticket out? Trends Cell Biol **13**(6): 295-300.

Barnes, D. A., A. Bonnin, J. X. Huang, L. Gousset, J. Wu, J. Gut, P. Doyle, J. F. Dubremetz, H. Ward and C. Petersen (1998). A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion. Mol Biochem Parasitol **96**(1-2): 93-110.

Baruch, D. I., B. L. Pasloske, H. B. Singh, X. Bi, X. C. Ma, M. Feldman, T. F. Taraschi and R. J. Howard (1995). Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell **82**(1): 77-87.

Batanghari, J. W., G. S. Deepe, Jr., E. Di Cera and W. E. Goldman (1998). *Histoplasma* acquisition of calcium and expression of CBP1 during intracellular parasitism. Mol Microbiol **27**(3): 531-539.

Bauman, A. L. and J. D. Scott (2002). Kinase- and phosphatase-anchoring proteins: harnessing the dynamic duo. Nat Cell Biol **4**(8): E203-206.

Bendtsen, J. D., L. J. Jensen, N. Blom, G. Von Heijne and S. Brunak (2004). Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel **17**(4): 349-356.

Bennuru, S., R. Semnani, Z. Meng, J. M. C. Ribeiro, T. D. Veenstra and T. B. Nutman (2009). *Brugia malayi* Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling. PLoS Negl Trop Dis **3**(4).

Bergsten, J. (2005). A review of long - branch attraction. Cladistics **21**(2): 163-193.

Berman, J. D. (1997). Human Leishmaniasis: Clinical, Diagnostic, and Chemotherapeutic Developments in the Last 10 Years. Clinical Infectious Diseases **24**(4): 684-703.

Berne, S., J. Pohleven, I. Vidic, K. Rebolj, F. Pohleven, T. Turk, P. Macek, A. Sonnenberg and K. Sepcic (2007). Ostreolysin enhances fruiting initiation in the oyster mushroom (*Pleurotus ostreatus*). Mycol Res **111**: 1431-1436.

Beznoussenko, G. V., V. V. Dolgikh, E. V. Seliverstova, P. B. Semenov, Y. S. Tokarev, A. Trucco, M. Micaroni, D. Di Giandomenico, P. Auinger, I. V. Senderskiy, S. O. Skarlato, E. S. Snigirevskaya, Y. Y. Komissarchik, M. Pavelka, M. A. De Matteis, A. Luini, Y. Y. Sokolova and A. A. Mironov (2007). Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. J Cell Sci **120**: 1288-1298.

Bhakdi, S. and J. Tranum-Jensen (1991). Alpha-toxin of *Staphylococcus aureus*. Microbiol Rev **55**(4): 733-751.

Bhat, S. A., I. Bashir and A. S. Kamili (2009). Microsporidiosis of silkworm, *Bombyx mori* (*Lepidoptera bombycidae*): A review. Afr. J. Agric. Res **4**: 1519-1523.

Bhattacharya, D., A. Nagpure and R. K. Gupta (2007). Bacterial Chitinases: Properties and Potential. Critical Reviews in Biotechnology **27**(1): 21-28.

Bigliardi, E., A. M. Bernuzzi, S. Corona, S. Gatti, M. Scaglia and L. Sacchi (2000). In Vitro Efficacy of Nikkomycin Z against the Human Isolate of the

Microsporidian Species *Encephalitozoon hellem*. Antimicrobial Agents and Chemotherapy **44**(11): 3012-3016.

Bigliardi, E., M. G. Selmi, P. Lupetti, S. Corona, S. Gatti, M. Scaglia and L. Sacchi (1996). Microsporidian spore wall: ultrastructural findings on *Encephalitozoon hellem* exospore. J Eukaryot Microbiol **43**(3): 181-186.

Bill, R. M., P. J. F. Henderson, S. Iwata, E. R. S. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate and H. Vogel (2011). Overcoming barriers to membrane protein structure determination. Nat Biotech **29**(4): 335-340.

Binns, D., E. Dimmer, R. Huntley, D. Barrell, C. O'Donovan and R. Apweiler (2009). QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics **25**(22): 3045-3046.

Bjellqvist, B., B. Basse, E. Olsen and J. E. Celis (1994). Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis **15**(3-4): 529-539.

Bleichert, F. and S. J. Baserga (2007). The Long Unwinding Road of RNA Helicases. Molecular Cell **27**(3): 339-352.

Bogdan, C. (2001). Nitric oxide and the immune response. Nat Immunol **2**(10): 907-916.

Bohm, S., G. Lamberti, V. Fernandez-Saiz, C. Stapf and A. Buchberger (2011). Cellular functions of Ufd2 and Ufd3 in proteasomal protein degradation depend on Cdc48 binding. Mol Cell Biol **31**(7): 1528-1539.

Bohne, W., K. Bottcher and U. Gross (2011). The parasitophorous vacuole of *Encephalitozoon cuniculi*: biogenesis and characteristics of the host cell-pathogen interface. Int J Med Microbiol **301**(5): 395-399.

Bohne, W., D. J. Ferguson, K. Kohler and U. Gross (2000). Developmental expression of a tandemly repeated, glycine- and serine-rich spore wall protein in the microsporidian pathogen *Encephalitozoon cuniculi*. Infect Immun **68**(4): 2268-2275.

Bohni, P. C., R. J. Deshaies and R. W. Schekman (1988). SEC11 is required for signal peptide processing and yeast cell growth. J Cell Biol **106**(4): 1035-1042.

Boldo, J., A. Junges, K. do Amaral, C. Staats, M. Vainstein and A. Schrank (2009). Endochitinase CHI2 of the biocontrol fungus *Metarhizium anisopliae* affects its virulence toward the cotton stainer bug *Dysdercus peruvianus*. Current Genetics **55**(5): 551-560.

Bommala, M. L., S. Nalamada, S. Sharma and P. Garg (2011). Microsporidial keratoconjunctivitis in an immunocompetent patient with a past history of laser in situ keratomilieusis surgery. Indian J Med Microbiol **29**(4): 428-430.

Bononi, A., C. Agnoletto, E. De Marchi, S. Marchi, S. Patergnani, M. Bonora, C. Giorgi, S. Missiroli, F. Poletti, A. Rimessi and P. Pinton (2011). Protein Kinases and Phosphatases in the Control of Cell Fate. Enzyme Research **2011**: 26.

Bornhorst, J. A. and J. J. Falke (2000). Purification of proteins using polyhistidine affinity tags. Methods Enzymol **326**: 245-254.

Bouzahzah, B., F. Nagajyothi, K. Ghosh, P. M. Takvorian, A. Cali, H. B. Tanowitz and L. M. Weiss (2010). Interactions of *Encephalitozoon cuniculi* polar tube proteins. Infect Immun **78**(6): 2745-2753.

Bouzahzah, B. and L. Weiss (2010). Glycosylation of the major polar tube protein of *Encephalitozoon cuniculi*. Parasitology Research **107**(3): 761-764.

Bozkurt, T. O., S. Schornack, M. J. Banfield and S. Kamoun (2012). Oomycetes, effectors, and all that jazz. Current Opinion in Plant Biology **15**(4): 483-492.

Brameier, M., A. Krings and R. M. MacCallum (2007). NucPred-predicting nuclear localisation of proteins. Bioinformatics **23**(9): 1159-1160.

Brameld, K. A. and W. A. Goddard (1998). The role of enzyme distortion in the single displacement mechanism of family-19 chitinases. Proc Natl Acad Sci U S A **95**(8): 4276-4281.

Brameld, K. A., W. D. Shrader, B. Imperiali and W. A. Goddard Iii (1998). Substrate assistance in the mechanism of family-18 chitinases: theoretical studies of potential intermediates and inhibitors. Journal of Molecular Biology **280**(5): 913-923.

Britt, A. B. (1999). Molecular genetics of DNA repair in higher plants. Trends in Plant Science **4**(1): 20-25.

Brittingham, A., C. J. Morrison, W. R. McMaster, B. S. McGwire, K. P. Chang and D. M. Mosser (1995). Role of the *Leishmania* surface protease gp63 in complement fixation, cell adhesion, and resistance to complement-mediated lysis. The Journal of Immunology **155**(6): 3102-3111.

Brosson, D., L. Kuhn, F. Delbac, J. Garin, C. P. Vivarès and C. Texier (2006). Proteomic analysis of the eukaryotic parasite *Encephalitozoon cuniculi* (microsporidia): a reference map for proteins expressed in late sporogonial stages. Proteomics **6**(12): 3625-3635.
Brosson, D., L. Kuhn, G. Prensier, C. P. Vivares and C. Texier (2005). The putative chitin deacetylase of *Encephalitozoon cuniculi*: a surface protein implicated in microsporidian spore-wall formation. FEMS Microbiol Lett **247**(1): 81-90.

Brown, J. D., B. C. Hann, K. F. Medzihradszky, M. Niwa, A. L. Burlingame and P. Walter (1994). Subunits of the *Saccharomyces cerevisiae* signal recognition particle required for its functional expression. EMBO J **13**(18): 4390-4400.

Brunner, F., A. Stintzi, B. Fritig and M. Legrand (1998). Substrate specificities of tobacco chitinases. Plant J **14**(2): 225-234.

Burri, L., B. A. P. Williams, D. Bursac, T. Lithgow and P. J. Keeling (2006). Microsporidian mitosomes retain elements of the general mitochondrial targeting system. Proc Natl Acad Sci U S A **103**(43): 15916-15920.

Buxbaum, L. U., H. Denise, G. H. Coombs, J. Alexander, J. C. Mottram and P. Scott (2003). Cysteine Protease B of *Leishmania mexicana* Inhibits Host Th1 Responses and Protective Immunity. The Journal of Immunology **171**(7): 3711-3717.

Cai Shun-feng, Qiu Hai-hong, Li Ming-qian, Feng Zhen-zhen (2012). Phagocytic Uptake of *Nosema bombycis* (Microsporidia) Spores by Insect Cell Lines. Journal of Integrative Agriculture **11**(8): 1321-1326.

Cali, A. and P. M. Takvorian (1999). Developmental morphology and lifecycles of the microsporidia. The microsporidia and microsporidiosis. American Society for Microbiology, Washington, DC: 85-128.

Campbell, S. E., T. A. Williams, A. Yousuf, D. M. Soanes, K. H. Paszkiewicz, B. A. P. Williams (2013). The genome of *Spraguea lophii* and the basis of host-microsporidian interactions. PLoS Genet **9**(8)

Canning, E. U. (1988). Nuclear division and chromosome cycle in microsporidia. Biosystems **21**(3–4): 333-340.

Capella-Gutierrez, S., M. Marcet-Houben and T. Gabaldon (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC biology **10**(47).

Carruthers, V. B., O. K. Giddings and L. D. Sibley (1999). Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. Cellular Microbiology **1**(3): 225-235.

Casadevall, A. (2008). Evolution of intracellular pathogens. Annu Rev Microbiol **62**: 19-33.

Cavalier-Smith, T. (1983). A 6-kingdom classification and a unified phylogeny. Endocytobiology II: 1027-1034.

Cavalier-Smith, T. (1987). Eukaryotes with no mitochondria. Nature **326**(6111): 332-333.

Cevallos, A. M., N. Bhat, R. Verdon, D. H. Hamer, B. Stein, S. Tzipori, M. E. Pereira, G. T. Keusch and H. D. Ward (2000). Mediation of *Cryptosporidium parvum* infection in vitro by mucin-like glycoproteins defined by a neutralizing monoclonal antibody. Infect Immun **68**(9): 5167-5175.

Chaffin, W. L., J. L. Lopez-Ribot, M. Casanova, D. Gozalbo and J. P. Martinez (1998). Cell wall and secreted proteins of *Candida albicans*: identification, function, and expression. Microbiol Mol Biol Rev **62**(1): 130-180.

Chen, Y. C., M. C. Chang, Y. C. Chuang and C. L. Jeang (2004). Characterization and virulence of hemolysin III from *Vibrio vulnificus*. Curr Microbiol **49**(3): 175-179.

Cherry, J. M., C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng and D. Botstein (1998). SGD: *Saccharomyces* Genome Database. Nucleic Acids Research **26**(1): 73-79.

Claros, M. G. and P. Vincens (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem **241**(3): 779-786.

Collinge, D. B., K. M. Kragh, J. D. Mikkelsen, K. K. Nielsen, U. Rasmussen and K. Vad (1993). Plant chitinases. The Plant Journal **3**(1): 31-40.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**(18): 3674-3676.

Consortium, T. U. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research **40**(D1): 71-75.

Cornman, R. S., Y. P. Chen, M. C. Schatz, C. Street, Y. Zhao, B. Desany, M. Egholm, S. Hutchison, J. S. Pettis, W. I. Lipkin and J. D. Evans (2009). Genomic Analyses of the Microsporidian *Nosema ceranae*, an Emergent Pathogen of Honey Bees. PLoS Pathog **5**(6).

Corradi, N., D. E. Akiyoshi, H. G. Morrison, X. Feng, L. M. Weiss, S. Tzipori and P. J. Keeling (2007). Patterns of Genome Evolution among the Microsporidian Parasites *Encephalitozoon cuniculi*, *Antonospora locustae* and *Enterocytozoon bieneusi*. PLoS ONE **2**(12).

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

Corradi, N. and P. J. Keeling (2009). Microsporidia: a journey through radical taxonomical revisions. Fungal Biology Reviews **23**(1): 1-8.

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

Corradi, N. and C. H. Slamovits (2011). The intriguing nature of microsporidian genomes. Brief Funct Genomics **10**(3): 115-124.

Costa, S. F. and L. M. Weiss (2000). Drug treatment of microsporidiosis. Drug Resistance Updates **3**(6): 384-399.

Cotte, L., M. Rabodonirina, F. Chapuis, F. Bailly, F. Bissuel, C. Raynal, P. Gelas, F. Persat, M. A. Piens and C. Trepo (1999). Waterborne outbreak of

intestinal microsporidiosis in persons with and without human immunodeficiency virus infection. Journal of Infectious Diseases **180**(6): 2003-2008.

Cox-Foster, D. L., S. Conlan, E. C. Holmes, G. Palacios, J. D. Evans, N. A. Moran, P.-L. Quan, T. Briese, M. Hornig, D. M. Geiser, V. Martinson, D. vanEngelsdorp, A. L. Kalkstein, A. Drysdale, J. Hui, J. Zhai, L. Cui, S. K. Hutchison, J. F. Simons, M. Egholm, J. S. Pettis and W. I. Lipkin (2007). A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. Science **318**(5848): 283-287.

Cox, J. C., D. Pye, J. W. Edmonds and R. Shepard (1980). An investigation of Encephalitozoon cuniculi in the wild rabbit *Oryctolagus cuniculus* in Victoria, Australia. J Hyg (Lond) **80**(2): 295-300.

Coyle, C. M., L. M. Weiss, L. V. Rhodes III, A. Cali, P. M. Takvorian, D. F. Brown, G. S. Visvesvara, L. Xiao, J. Naktin and E. Young (2004). Fatal myositis due to the microsporidian *Brachiola algerae*, a mosquito pathogen. New England Journal of Medicine **351**(1): 42-47.

Crabb, B. S., B. M. Cooke, J. C. Reeder, R. F. Waller, S. R. Caruana, K. M. Davern, M. E. Wickham, G. V. Brown, R. L. Coppel and A. F. Cowman (1997). Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell **89**(2): 287-296.

Craven, R. R., X. Gao, I. C. Allen, D. Gris, J. B. Wardenburg, E. McElvania-TeKippe, J. P. Ting and J. A. Duncan (2009). *Staphylococcus aureus* α-Hemolysin Activates the NLRP3-Inflammasome in Human and Mouse Monocytic Cells. PLoS ONE **4**(10).

Cuellar-Cruz, M., M. Briones-Martin-del-Campo, I. Canas-Villamar, J. Montalvo-Arredondo, L. Riego-Ruiz, I. Castano and A. De Las Penas (2008). High resistance to oxidative stress in the fungal pathogen *Candida glabrata* is mediated by a single catalase, Cta1p, and is controlled by the transcription factors Yap1p, Skn7p, Msn2p, and Msn4p. Eukaryot Cell **7**(5): 814-825.

Cuervo, P., J. B. De Jesus, L. Saboia-Vahia, L. Mendonça-Lima, G. B. Domont and E. Cupolillo (2009). Proteomic characterization of the released/secreted proteins of *Leishmania braziliensis* promastigotes. Journal of Proteomics **73**(1): 79-92.

Cuevas, I. C., J. J. Cazzulo and D. O. Sánchez (2003). gp63 Homologues in *Trypanosoma cruzi*: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection. Infection and Immunity **71**(10): 5739-5749.

Cuomo, C. A., C. A. Desjardins, M. A. Bakowski, J. Goldberg, A. T. Ma, J. J. Becnel, E. S. Didier, L. Fan, D. I. Heiman, J. Z. Levin, S. Young, Q. Zeng and E. R. Troemel (2012). Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res **22**(12): 2478-2488.

da Silva, A. J., D. A. Schwartz, G. S. Visvesvara, H. De Moura, S. B. Slemenda and N. J. Pieniazek (1996). Sensitive PCR diagnosis of Infections by

Enterocytozoon bieneusi (microsporidia) using primers based on the region coding for small-subunit rRNA. Journal of clinical microbiology **34**(4): 986-987.

Dayhoff, M. O. and R. M. Schwartz (1978). A model of evolutionary change in proteins. In Atlas of protein sequence and structure

de Avalos, S. V., I. J. Blader, M. Fisher, J. C. Boothroyd and B. A. Burleigh (2002). Immediate/Early Response to *Trypanosoma cruzi* Infection Involves Minimal Modulation of Host Cell Transcription. Journal of Biological Chemistry **277**(1): 639-644.

De Carli, G. A., P. Brasseur, A. C. d. Silva, A. Wendorff and M. Rott (1996). Hemolytic activity of *Trichomonas vaginalis* and *Tritrichomonas foetus*. Memórias do Instituto Oswaldo Cruz **91**: 107-110.

de Koning, A. P., F. S. L. Brinkman, S. J. M. Jones and P. J. Keeling (2000). Lateral Gene Transfer and Metabolic Adaptation in the Human Parasite *Trichomonas vaginalis*. Molecular Biology and Evolution **17**(11): 1769-1773.

Dejgaard, S., J. Nicolay, M. Taheri, D. Y. Thomas and J. J. Bergeron (2004). The ER glycoprotein quality control system. Curr Issues Mol Biol **6**(1): 29-42.

del Castillo, F. J., S. C. Leal, F. Moreno and I. del Castillo (1997). The *Escherichia coli* K-12 sheA gene encodes a 34-kDa secreted haemolysin. Mol Microbiol **25**(1): 107-115.

Delbac, F., I. Peuvel, G. Metenier, E. Peyretaillade and C. P. Vivares (2001). Microsporidian Invasion Apparatus: Identification of a Novel Polar Tube Protein and Evidence for Clustering of ptp1 and ptp2 Genes in Three *Encephalitozoon* Species. Infection and Immunity **69**(2): 1016-1024.

Delom, F. and E. Chevet (2006). Phosphoprotein analysis: from proteins to proteomes. Proteome Sci **4:** 8-15.

Denu, J. M. and J. E. Dixon (1998). Protein tyrosine phosphatases: mechanisms of catalysis and regulation. Current Opinion in Chemical Biology **2**(5): 633-641.

Desportes, I., Y. L. Charpentier, A. Galian, F. Bernard, B. Cochand - Priollet, A. Lavergne, P. Ravisse and R. Modigliani (1985). Occurrence of a New Microsporidan: *Enterocytozoon bieneusi* sp. in the Enterocytes of a Human Patient with AIDS1. Journal of Eukaryotic Microbiology **32**(2): 250-254.

Dia, N., L. Lavie, G. Metenier, B. S. Toguebaye, C. P. Vivares and E. Cornillot (2007). InterB multigenic family, a gene repertoire associated with subterminal chromosome regions of *Encephalitozoon cuniculi* and conserved in several human-infecting microsporidian species. Curr Genet **51**(3): 171-186.

Dickson, R. C., C. Sumanasekera and R. L. Lester (2006). Functions and metabolism of sphingolipids in *Saccharomyces cerevisiae*. Progress in Lipid Research **45**(6): 447-465.

Didier, P. J., E. S. Didier, J. M. Orenstein and J. A. Shadduck (1991). Fine Structure of a New Human Microsporidian, *Encephalitozoon hellem*, in Culture. Journal of Eukaryotic Microbiology **38**(5): 502-507.

Didier, E. S. (2005) Microsporidiosis: An emerging and opportunistic infection in humans and animals. Acta Tropica **94**(1): 61-76.

Dieterich, D. T., E. A. Lew, D. P. Kotler, M. A. Poles and J. M. Orenstein (1994). Treatment with albendazole for intestinal disease due to *Enterocytozoon bieneusi* in patients with AIDS. Journal of Infectious Diseases **169**(1): 178-183.

Dingwall, C., J. Robbins, S. M. Dilworth, B. Roberts and W. D. Richardson (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. J Cell Biol **107**(3): 841-849.

Ditrich, O., A. Chrdle, B. Sak, V. Chmelík, J. Kubále, I. Dyková and M. Kváč (2011). *Encephalitozoon cuniculi* genotype I as a causative agent of brain abscess in an immunocompetent patient. Journal of clinical microbiology **49**(7): 2769-2771.

Doflein, F. (1898). Studien zur Naturgeschichte der Protozoen: Über Polytomella agilis Aragao, nebst Bemerkungen über die Kernteilung bei den Protozoen und den Stoffwechsel der Zuckerflagellaten.

Dolgikh, V. V. and P. B. Semenov (2003). Trehalose catabolism in microsporidia *Nosema grylli* spores. Parazitologiia **37**(4): 333-342.

Dolgikh, V. V., P. B. Semenov and G. V. Beznusenko (2007). Protein glycosylation in the spores of the microsporidia Paranosema (Antonospora) grylli. Tsitologiia **49**(7): 607-613.

Dolgikh, V. V., P. B. Semenov, A. A. Mironov and G. V. Beznoussenko (2005). Immunocytochemical identification of the major exospore protein and three polar-tube proteins of the microsporidia *Paranosema grylli*. Protist **156**(1): 77-87.

Dolgikh, V. V., I. V. Senderskiy, O. A. Pavlova, A. M. Naumov and G. V. Beznoussenko (2011). Immunolocalization of an Alternative Respiratory Chain in *Antonospora* (Paranosema) *locustae* Spores: Mitosomes Retain Their Role in Microsporidial Energy Metabolism. Eukaryotic Cell **10**(4): 588-593.

Drummond, A. J. and A. Rambaut (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **7**: 214.

Dubyak, G. R. (2012). P2X7 receptor regulation of non-classical secretion from immune effector cells. Cellular Microbiology **14**(11): 1697-1706.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends in Genetics **12**(7): 263-270.

Dunn, A. M. and J. E. Smith (2001). Microsporidian lifecycles and diversity: the relationship between virulence and transmission. Microbes Infect **3**(5): 381-388.

Dussaubat, C., A. Maisonnasse, D. Crauser, D. Beslay, G. Costagliola, S. Soubeyrand, A. Kretzchmar and Y. Le Conte (2013). Flight behavior and pheromone changes associated to *Nosema ceranae* infection of honey bee workers *Apis mellifera* in field conditions. Journal of Invertebrate Pathology.

Ebina, K., S. Ichinowatari and K. Yokota (1985). Studies on toxin of *Aspergillus fumigatus*. Fashion of binding of Asp-hemolysin to human erythrocytes and Asp-hemolysin-binding proteins of erythrocyte membranes. Microbiol Immunol **29**(2): 91-101.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**(5): 1792-1797.

Edlind, T. D., J. Li, G. S. Visvesvara, M. H. Vodkin, G. L. McLaughlin and S. K. Katiyar (1996). Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. Molecular phylogenetics and evolution **5**(2): 359.

El Hajj, H., M. Lebrun, S. T. Arold, H. Vial, G. Labesse and J. F. Dubremetz (2007). ROP18 Is a Rhoptry Kinase Controlling the Intracellular Proliferation of *Toxoplasma gondii*. PLoS Pathog **3**(2): e14.

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc **2**(4): 953-971.

Emanuelsson, O., H. Nielsen, S. Brunak and G. von Heijne (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol **300**(4): 1005-1016.

Enriquez, F. J., O. Ditrich, J. D. Palting and K. Smith (1997). Simple diagnosis of *Encephalitozoon* sp. microsporidial infections by using a panspecific antiexospore monoclonal antibody. Journal of clinical microbiology **35**(3): 724-729.

Estes, K. A., S. C. Szumowski and E. R. Troemel (2011). Non-lytic, actin-based exit of intracellular parasites from *C. elegans* intestinal cells. PLoS Pathog **7**(9).

Fadili, A. E., C. Kündig and M. Ouellette (2002). Characterization of the folylpolyglutamate synthetase gene and polyglutamylation of folates in the protozoan parasite *Leishmania*. Molecular and Biochemical Parasitology **124**(1–2): 63-71.

Fan, Y., W. Fang, S. Guo, X. Pei, Y. Zhang, Y. Xiao, D. Li, K. Jin, M. J. Bidochka and Y. Pei (2007). Increased Insect Virulence in *Beauveria bassiana* Strains Overexpressing an Engineered Chitinase. Applied and Environmental Microbiology **73**(1): 295-302.

Farber, J. M., J. I. Speirs, R. Pontefract and D. E. Conner (1991). Characteristics of nonpathogenic strains of *Listeria monocytogenes*. Can J Microbiol **37**(8): 647-650.

Farmaki, T., S. Ponnambalam, A. R. Prescott, H. Clausen, B. L. Tang, W. Hong and J. M. Lucocq (1999). Forward and retrograde trafficking in mitotic animal cells. ER-Golgi transport arrest restricts protein export from the ER into COPII-coated structures. Journal of Cell Science **112**(5): 589-600.

Farquhar, M. G. and G. E. Palade (1981). The Golgi apparatus (complex) (1954-1981) from artifact to center stage. J Cell Biol **91**(3): 77-103.

Fasshauer, V., U. Gross and W. Bohne (2005). The parasitophorous vacuole membrane of *Encephalitozoon cuniculi* lacks host cell membrane proteins immediately after invasion. Eukaryot Cell **4**(1): 221-224.

Fast, N. M. and P. J. Keeling (2001). Alpha and beta subunits of pyruvate dehydrogenase E1 from the microsporidian *Nosema locustae*: mitochondrion-derived carbon metabolism in microsporidia. Molecular and Biochemical Parasitology **117**(2): 201-209.

Fast, N. M., J. S. Law, B. A. Williams and P. J. Keeling (2003). Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution. Eukaryot Cell **2**(5): 1069-1075.

Ficarro, S. B., M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt and F. M. White (2002). Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. Nat Biotechnol **20**(3): 301-305.

Fields, K. A., E. Fischer and T. Hackstadt (2002). Inhibition of fusion of *Chlamydia trachomatis* inclusions at 32 degrees C correlates with restricted export of IncA. Infect Immun **70**(7): 3816-3823.

Findley, A. M., E. H. Weidner, K. R. Carman, Z. Xu and J. S. Godbar (2005). Role of the posterior vacuole in *Spraguea lophii* (Microsporidia) spore hatching. Folia Parasitol (Praha) **52**(1-2): 111-117.

Foley, D. H., J. H. Bryan, D. Yeates and A. Saul (1998). Evolution and Systematics of *Anopheles*: Insights from a Molecular Phylogeny of Australasian Mosquitoes. Molecular phylogenetics and evolution **9**(2): 262-275.

Foster, P. G. and D. A. Hickey (1999). Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. Journal of Molecular Evolution **48**(3): 284-290.

Franzen, C. and A. Müller (2001). Microsporidiosis: human diseases and diagnosis. Microbes and infection **3**(5): 389-400.

Freeman, M., A. Bell and C. Sommerville (2003). A hyperparasitic microsporidian infecting the salmon louse, *Lepeophtheirus salmonis*: an rDNA - based molecular phylogenetic study. Journal of fish diseases **26**(11): 667-676.

Freeman, M. A., H. Yokoyama, A. Osada, T. Yoshida, A. Yamanobe and K. Ogawa (2011). *Spraguea* (Microsporida: Spraguidae) infections in the nervous

system of the Japanese anglerfish, *Lophius litulon* (Jordan), with comments on transmission routes and host pathology. Journal of Fish Diseases **34**(6): 445-452.

Friesen, T. L., E. H. Stukenbrock, Z. Liu, S. Meinhardt, H. Ling, J. D. Faris, J. B. Rasmussen, P. S. Solomon, B. A. McDonald and R. P. Oliver (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. Nat Genet **38**(8): 953-956.

Frixione, E., L. Ruiz, J. Cerbon and A. H. Undeen (1997). Germination of *Nosema algerae* (Microspora) spores: conditional inhibition by D2O, ethanol and Hg^{2+} suggests dependence of water influx upon membrane hydration and specific transmembrane pathways. J Eukaryot Microbiol **44**(2): 109-116.

Frixione, E., L. Ruiz, M. Santillán, L. V. de Vargas, J. M. Tejero and A. H. Undeen (1992). Dynamics of polar filament discharge and sporoplasm expulsion by microsporidian spores. Cell Motility and the Cytoskeleton **22**(1): 38-50.

Frixione, E., L. Ruiz and A. H. Undeen (1994). Monovalent Cations Induce Microsporidian Spore Germination In Vitro. Journal of Eukaryotic Microbiology **41**(5): 464-468.

Fromme, J. C. and R. Schekman (2005). COPII-coated vesicles: flexible enough for large cargo? Curr Opin Cell Biol **17**(4): 345-352.

Funkhouser, J. D. and N. N. Aronson, Jr. (2007). Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol Biol **7**: 96.

Futerman, A. H. and H. Riezman (2005). The ins and outs of sphingolipid synthesis. Trends in Cell Biology **15**(6): 312-318.

Germot, A., H. Philippe and H. Le Guyader (1997). Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Mol Biochem Parasitol **87**(2): 159-168.

Ghosh, K., C. D. Cappiello, S. M. McBride, J. L. Occi, A. Cali, P. M. Takvorian, T. V. McDonald and L. M. Weiss (2006). Functional characterization of a putative aquaporin from *Encephalitozoon cuniculi*, a microsporidia pathogenic to humans. Int J Parasitol **36**(1): 57-62.

Ghosh, K., E. Nieves, P. Keeling, J.-F. Pombert, P. P. Henrich, A. Cali and L. M. Weiss (2011). Branching Network of Proteinaceous Filaments within the Parasitophorous Vacuole of *Encephalitozoon cuniculi* and *Encephalitozoon hellem*. Infection and Immunity **79**(3): 1374-1385.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo,

D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis and M. Johnston (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418**(6896): 387-391.

Gilbert, L. A., S. Ravindran, J. M. Turetzky, J. C. Boothroyd and P. J. Bradley (2007). *Toxoplasma gondii* Targets a Protein Phosphatase 2C to the Nuclei of Infected Host Cells. Eukaryotic Cell **6**(1): 73-83.

Glickman, M. H. and A. Ciechanover (2002). The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. Physiological Reviews **82**(2): 373-428.
Goebel, W., T. Chakraborty and J. Kreft (1988). Bacterial hemolysins as virulence factors. Antonie van Leeuwenhoek **54**(5): 453-463.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver (1996). Life with 6000 genes. Science **274**(5287): 546, 563-547.

Goldberg, A. V., S. Molik, A. D. Tsaousis, K. Neumann, G. Kuhnke, F. Delbac, C. P. Vivares, R. P. Hirt, R. Lill and T. M. Embley (2008). Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. Nature **452**(7187): 624-628.

Gomes, M. T., A. H. Lopes, J. Meyer-Fernandes, and Roberto (2011). Possible Roles of Ectophosphatases in Host-Parasite Interactions. Journal of Parasitology Research **2011**.

Gooday, G. W. (1999). Aggressive and defensive roles for chitinases. EXS **87**: 157-169.

Green, J. and M. L. Baldwin (1997). HlyX, the FNR homologue of *Actinobacillus pleuropneumoniae*, is a [4Fe-4S]-containing oxygen-responsive transcription regulator that anaerobically activates FNR-dependent class I promoters via an enhanced AR1 contact. Mol Microbiol **24**(3): 593-605.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology **59**(3): 307-321.

Halic, M., T. Becker, M. R. Pool, C. M. Spahn, R. A. Grassucci, J. Frank and R. Beckmann (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature **427**(6977): 808-814.

Halle, M., M. A. Gomez, M. Stuible, H. Shimizu, W. R. McMaster, M. Olivier and M. L. Tremblay (2009). The *Leishmania* surface protease GP63 cleaves

multiple intracellular proteins and actively participates in p38 mitogen-activated protein kinase inactivation. Journal of Biological Chemistry **284**(11): 6893-6908.

Han, X., C.-T. Wang, J. Bai, E. R. Chapman and M. B. Jackson (2004). Transmembrane Segments of Syntaxin Line the Fusion Pore of Ca2+-Triggered Exocytosis. Science **304**(5668): 289-292.

Hann, B. C. and P. Walter (1991). The signal recognition particle in *S. cerevisiae*. Cell **67**(1): 131-144.
Hartl, L., S. Zach and V. Seidl-Seiboth (2012). Fungal chitinases: diversity, mechanistic properties and biotechnological potential. Appl Microbiol Biotechnol **93**(2): 533-543.

Hauri, H. P., F. Kappeler, H. Andersson and C. Appenzeller (2000). ERGIC-53 and traffic in the secretory pathway. J Cell Sci **113**(4): 587-596.

Hauri, H. P., O. Nufer, L. Breuza, H. B. Tekaya and L. Liang (2002). Lectins and protein traffic early in the secretory pathway. Biochemical Society symposium (69): 73-82.

Hebert, D. N. and M. Molinari (2007). In and Out of the ER: Protein Folding, Quality Control, Degradation, and Related Human Diseases. Physiological Reviews **87**(4): 1377-1408.

Hebert, D. N. and M. Molinari (2012). Flagging and docking: dual roles for *N*-glycans in protein quality control and cellular proteostasis. Trends in Biochemical Sciences **37**(10): 404-410.

Hecht, D. and Y. Zick (1992). Selective inhibition of protein tyrosine phosphatase activities by H_2O_2 and vanadate In vitro. Biochemical and Biophysical Research Communications **188**(2): 773-779.

Heijne, G. (1990). The signal peptide. The Journal of Membrane Biology **115**(3): 195-201.

Heinz, E., T. A. Williams, S. Nakjang, C. J. Noël, D. C. Swan, A. V. Goldberg, S. R. Harris, T. Weinmaier, S. Markert, D. Becher, J. Bernhardt, T. Dagan, C. Hacker, J. M. Lucocq, T. Schweder, T. Rattei, N. Hall, R. P. Hirt and T. M. Embley (2012). The Genome of the Obligate Intracellular Parasite *Trachipleistophora hominis*: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. PLoS Pathog **8**(10).

Hendrix, L. R. (2000). Contact-dependent hemolytic activity distinct from deforming activity of *Bartonella bacilliformis*. FEMS Microbiol Lett **182**(1):119-124.

Hermoso, T., Z. Fishelson, S. I. Becker, K. Hirschberg and C. L. Jaffe (1991). Leishmanial protein kinases phosphorylate components of the complement system. EMBO J **10**(13): 4061-4067.

Herrera-Estrella, A. and I. Chet (1999). Chitinases in biological control. EXS **87**: 171-184.

Higes, M., R. Martín - Hernández, C. Botías, E. G. Bailón, A. V. González - Porto, L. Barrios, M. J. del Nozal, J. L. Bernal, J. J. Jiménez and P. G. Palencia (2008). How natural infection by *Nosema ceranae* causes honeybee colony collapse. Environmental Microbiology **10**(10): 2659-2669.

Hiller, N. L., S. Bhattacharjee, C. van Ooij, K. Liolios, T. Harrison, C. Lopez-Estraño and K. Haldar (2004). A Host-Targeting Signal in Virulence Proteins Reveals a Secretome in Malarial Infection. Science **306**(5703): 1934-1937.
Hinkle, G., H. G. Morrison and M. L. Sogin (1997). Genes coding for reverse transcriptase, DNA-directed RNA polymerase, and chitin synthase from the microsporidian *Spraguea lophii*. Biol Bull **193**(2): 250-251.

Hirt, R. P., B. Healy, C. R. Vossbrinck, E. U. Canning and T. M. Embley (1997). A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Current biology **7**(12): 995-998.

Hirt, R. P., J. M. Logsdon, B. Healy, M. W. Dorey, W. F. Doolittle and T. M. Embley (1999). Microsporidia are related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci U S A **96**(2): 580-585.

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier and K. Nakai (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res **35**: 585-587.

Hughes, H. and D. J. Stephens (2008). Assembly, organization, and function of the COPII coat. Histochem Cell Biol **129**(2): 129-151.

Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman and E. K. O'Shea (2003). Global analysis of protein localization in budding yeast. Nature **425**(6959): 686-691.

Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. A. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P. D. Thomas, C. H. Wu, C. Yeats and S.-Y. Yong (2012). InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Research **40**: 306-312.

Hunter, T. (1995). Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. Cell **80**(2): 225-236.

Hurtado-Guerrero, R. and D. M. van Aalten (2007). Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. Chem Biol **14**(5): 589-599.

Hwang, L. H., J. A. Mayfield, J. Rine and A. Sil (2008). *Histoplasma* requires SID1, a member of an iron-regulated siderophore gene cluster, for host colonization. PLoS Pathog **4**(4): e1000044.

Islam, S. A., K. M. Cho, S. J. Hong, R. K. Math, J. M. Kim, M. G. Yun, J. J. Cho, J. Y. Heo, Y. H. Lee, H. Kim and H. D. Yun (2010). Chitinase of *Bacillus licheniformis* from oyster shell as a probe to detect chitin in marine shells. Appl Microbiol Biotechnol **86**(1): 119-129.

James, T. Y., F. Kauff, C. L. Schoch, P. B. Matheny, V. Hofstetter, C. J. Cox, G. Celio, C. Gueidan, E. Fraker and J. Miadlikowska (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443**(7113): 818-822.

Johnson, A. E. and M. A. van Waes (1999). The translocon: a dynamic gateway at the ER membrane. Annu Rev Cell Dev Biol **15**: 799-842.

Johnson, R. M., J. D. Evans, G. E. Robinson and M. R. Berenbaum (2009). Changes in transcript abundance relating to colony collapse disorder in honey bees (*Apis mellifera*). Proc Natl Acad Sci U S A **106**(35): 14790-14795.

Joshi, P. B., B. L. Kelly, S. Kamhawi, D. L. Sacks and W. R. McMaster (2002). Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. Molecular and Biochemical Parasitology **120**(1): 33-40.

Jyothi, N. and C. Patil (2011). Light and Electron Microscopic Study on the Development of *Nosema Bombycis* in the Midgut of Silkworm *Bombyx Mori*. Int. J. Indust. Entomol **23**(1): 107-113.

Kalies, K. U., S. Allan, T. Sergeyenko, H. Kroger and K. Romisch (2005). The protein translocation channel binds proteasomes to the endoplasmic reticulum membrane. EMBO J **24**(13): 2284-2293.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno and M. Hattori (2004). The KEGG resource for deciphering the genome. Nucleic Acids Research **32**: 277-280.

Karpichev, I. V., L. Cornivelli and G. M. Small (2002). Multiple regulatory roles of a novel *Saccharomyces cerevisiae* protein, encoded by YOL002c, in lipid and phosphate metabolism. Journal of Biological Chemistry **277**(22): 19609-19617.

Kasprzewska, A. (2003). Plant chitinases--regulation and function. Cell Mol Biol Lett **8**(3): 809-824.

Katinka, M. D., S. Duprat, E. Cornillot, G. Metenier, F. Thomarat, G. Prensier, V. Barbe, E. Peyretaillade, P. Brottier, P. Wincker, F. Delbac, H. El Alaoui, P. Peyret, W. Saurin, M. Gouy, J. Weissenbach and C. P. Vivares (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature **414**(6862): 450-453.

Keane, T. M., C. J. Creevey, M. M. Pentony, T. J. Naughton and J. O. McLnerney (2006). Assessment of methods for amino acid matrix selection and

their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol **6**: 29.

Keeling, P. (2009). Five Questions about Microsporidia. PLoS Pathog **5**(9).

Keeling, P. J. (2003). Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genetics and Biology **38**(3): 298-309.

Keeling, P. J., N. Corradi, H. G. Morrison, K. L. Haag, D. Ebert, L. M. Weiss, D. E. Akiyoshi and S. Tzipori (2010). The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. Genome Biol Evol **2**: 304-309.

Keeling, P. J. and W. F. Doolittle (1996). Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Molecular Biology and Evolution **13**(10): 1297-1305.

Keeling, P. J. and N. M. Fast (2002). Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol **56**: 93-116.

Keeling, P. J., N. M. Fast, J. S. Law, B. A. Williams and C. H. Slamovits (2005). Comparative genomics of microsporidia. Folia Parasitol (Praha) **52**(1-2): 8-14.

Keeling, P. J., M. A. Luker and J. D. Palmer (2000). Evidence from Beta-Tubulin Phylogeny that Microsporidia Evolved from Within the Fungi. Molecular Biology and Evolution **17**(1): 23-31.

Keeling, P. J. and J. D. Palmer (2008). Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet **9**(8): 605-618.

Kelkar, A. and B. Dobberstein (2009). Sec61beta, a subunit of the Sec61 protein translocation channel at the Endoplasmic Reticulum, is involved in the transport of Gurken to the plasma membrane. BMC Cell Biology **10**(1): 11.

Keohane, E. M., P. M. Takvorian, A. Cali, H. B. Tanowitz, M. Wittner and L. M. Weiss (1996). Identification of a microsporidian polar tube protein reactive monoclonal antibody. J Eukaryot Microbiol **43**(1): 26-31.

Keohane, E. M. and L. M. Weiss (1998). Characterization and function of the microsporidian polar tube: a review. Folia Parasitol (Praha) **45**(2): 117-127.

Kessler, H., A. Herm-Götz, S. Hegge, M. Rauch, D. Soldati-Favre, F. Frischknecht and M. Meissner (2008). Microneme protein 8 – a new essential invasion factor in *Toxoplasma gondii*. Journal of Cell Science **121**(7): 947-956.

Kiffer-Moreira, T., A. A. de Sa Pinheiro, W. S. Alviano, F. M. Barbosa, T. Souto-Padron, L. Nimrichter, M. L. Rodrigues, C. S. Alviano and J. R. Meyer-Fernandes (2007). An ectophosphatase activity in *Candida parapsilosis* influences the interaction of fungi with epithelial cells. FEMS Yeast Res **7**(4): 621-628.

Kihara, A., H. Sakuraba, M. Ikeda, A. Denpoh and Y. Igarashi (2008). Membrane topology and essential amino acid residues of Phs1, a 3-hydroxyacyl-CoA dehydratase involved in very long-chain fatty acid elongation. Journal of Biological Chemistry **283**(17): 11199-11209.

Klee, E. W. and L. B. Ellis (2005). Evaluating eukaryotic secreted protein prediction. BMC Bioinformatics **6**: 256.

Kostova, Z. and D. H. Wolf (2003). For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin-proteasome connection. EMBO J **22**(10): 2309-2317.

Kotler, D., T. Giang, M. Garro and J. Orenstein (1994). Light microscopic diagnosis of microsporidiosis in patients with AIDS. The American journal of gastroenterology **89**(4): 540.

Koudela, B., S. Kucerova and T. Hudcovic (1999). Effect of low and high temperatures on infectivity of *Encephalitozoon cuniculi* spores suspended in water. Folia Parasitol (Praha) **46**(3): 171-174.

Krogh, A., B. Larsson, G. von Heijne and E. L. Sonnhammer (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305**(3): 567-580.

Krylov, D. M., Y. I. Wolf, I. B. Rogozin and E. V. Koonin (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res **13**(10): 2229-2235.

Kumagai, T., T. Nagata, Y. Kudo, Y. Fukuchi, K. Ebina and K. Yokota (1999). Cytotoxic activity and cytokine gene induction of Asp-hemolysin to murine macrophages. Nihon Ishinkin Gakkai Zasshi **40**(4): 217-222.

Kumagai, T., T. Nagata, Y. Kudo, Y. Fukuchi, K. Ebina and K. Yokota (2001). Cytotoxic activity and cytokine gene induction of Asp-hemolysin to vascular endothelial cells. Yakugaku Zasshi **121**(4): 271-275.

Labandeira, C. and T. Phillips (1996). A Carboniferous insect gall: insight into early ecologic history of the Holometabola. Proc Natl Acad Sci U S A **93**(16): 8470-8474.

Lambertz, U., J. M. Silverman, D. Nandan, W. R. McMaster, J. Clos, L. J. Foster and N. E. Reiner (2012). Secreted virulence factors and immune evasion in visceral leishmaniasis. J Leukoc Biol **91**(6): 887-899.

Landis, G. N. and J. Tower (2005). Superoxide dismutase evolution and life span regulation. Mechanisms of Ageing and Development **126**(3): 365-379.

Langer, R. C., D. A. Schaefer and M. W. Riggs (2001). Characterization of an intestinal epithelial cell receptor recognized by the *Cryptosporidium parvum* sporozoite ligand CSL. Infect Immun **69**(3): 1661-1670.

Larsen, M. H., J. J. Leisner and H. Ingmer (2010). The chitinolytic activity of *Listeria monocytogenes* EGD is regulated by carbohydrates but also by the virulence regulator PrfA. Appl Environ Microbiol **76**(19): 6470-6476.

Lartillot, N., H. Brinkmann and H. Philippe (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol.

Lartillot, N., T. Lepage and S. Blanquart (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics **25**(17): 2286-2288.

Le, S. Q. and O. Gascuel (2008). An improved general amino acid replacement matrix. Mol Biol Evol **25**(7): 1307-1320.

Lee, B. J., A. E. Cansizoglu, K. E. Suel, T. H. Louis, Z. Zhang and Y. M. Chook (2006). Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell **126**(3): 543-558.

Lee, C. and J. Goldberg (2010). Structure of Coatomer Cage Proteins and the Relationship among COPI, COPII, and Clathrin Vesicle Coats. Cell **142**(1): 123-132.

Lee, S. C., L. M. Weiss and J. Heitman (2009). Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. Commun Integr Biol **5**(2): 414-417.

Lee, S. C., N. Corradi, E. J. Byrnes III, S. Torres-Martinez, F. S. Dietrich, P. J. Keeling and J. Heitman (2008). Microsporidia evolved from ancestral sexual fungi. Current Biology **18**(21): 1675-1679.

Lee, S. C., N. Corradi, S. Doan, F. S. Dietrich, P. J. Keeling and J. Heitman (2010). Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLoS One **5**(5): 10539.

Leiriao, P., C. D. Rodrigues, S. S. Albuquerque and M. M. Mota (2004). Survival of protozoan intracellular parasites in host cells. EMBO Rep **5**(12): 1142-1147.

Leitch, G. J., A. P. Shaw, M. Colden-Stanfield, M. Scanlon and G. S. Visvesvara (2005). Multinucleate host cells induced by *Vittaforma corneae* (Microsporidia). Folia Parasitol (Praha) **52**(1-2): 103-110.

Lenardon, M. D., C. A. Munro and N. A. R. Gow (2010). Chitin synthesis and fungal pathogenesis. Current Opinion in Microbiology **13**(4): 416-423.

Leonard, G., J. R. Stevens and T. A. Richards (2009). REFGEN and TREENAMER: automated sequence data handling for phylogenetic analysis in the genomic era. Evol Bioinform Online **5**: 1-4.

Lewis, M. J. and H. R. Pelham (1990). A human homologue of the yeast HDEL receptor. Nature **348**(6297): 162-163.

Lewis, M. J. and H. R. B. Pelham (1996). SNARE-Mediated Retrograde Traffic from the Golgi Complex to the Endoplasmic Reticulum. Cell **85**(2): 205-215.

Ley, V., E. S. Robbins, V. Nussenzweig and N. W. Andrews (1990). The exit of *Trypanosoma cruzi* from the phagosome is inhibited by raising the pH of acidic compartments. J Exp Med **171**(2): 401-413.

Li, Z., G. Pan, T. Li, W. Huang, J. Chen, L. Geng, D. Yang, L. Wang and Z. Zhou (2012). SWP5, a Spore Wall Protein, Interacts with Polar Tube Proteins in the Parasitic Microsporidian *Nosema bombycis*. Eukaryotic Cell **11**(2): 229-237.

Lill, R. and G. Kispal (2000). Maturation of cellular Fe-S proteins: an essential function of mitochondria. Trends Biochem Sci **25**(8): 352-356.

Lingelbach, K. (2001). Life in vacuoles – a strategy for parasite survival. International Journal for Parasitology **31**(12).

Liu, T. P. (1972). Ultrastructural changes in the nuclear envelope of larval fat body cells of *Simulium vittatum* (Diptera) induced by microsporidian infection of *Thelohania bracteata*. Tissue and Cell **4**(3): 493-501.

Loewen, P. C., J. Switala and B. L. Triggs-Raine (1985). Catalases HPI and HPII in Escherichia coli are induced independently. Arch Biochem Biophys **243**(1): 144-149.

Lom, J. and I. Dykova (2005). Microsporidian xenomas in fish seen in wider perspective. Folia Parasitol (Praha) **52**(1-2): 69-81.

Lott, K., A. Bhardwaj, P. J. Sims and G. Cingolani (2011). A minimal nuclear localization signal (NLS) in human phospholipid scramblase 4 that binds only the minor NLS-binding site of importin alpha1. Journal of Biological Chemistry **286**(32): 28160-28169.

Lücking, R., S. Huhndorf, D. H. Pfister, E. R. Plata and H. T. Lumbsch (2009). Fungi evolved right on track. Mycologia **101**(6): 810-822.

Ludwig, A., S. Bauer, R. Benz, B. Bergmann and W. Goebel (1999). Analysis of the SlyA-controlled expression, subcellular localization and pore-forming activity of a 34 kDa haemolysin (ClyA) from *Escherichia coli* K-12. Mol Microbiol **31**(2): 557-567.

Luo, G., L. P. Samaranayake and J. Y. Yau (2001). *Candida* species exhibit differential in vitro hemolytic activities. J Clin Microbiol **39**(8): 2971-2974.

Lyons, T. J., N. Y. Villa, L. M. Regalla, B. R. Kupchak, A. Vagstad and D. J. Eide (2004). Metalloregulation of yeast membrane steroid receptor homologs. Proc Natl Acad Sci U S A **101**(15): 5506-5511.

Maier, A. G., M. Rug, M. T. O'Neill, M. Brown, S. Chakravorty, T. Szestak, J. Chesson, Y. Wu, K. Hughes, R. L. Coppel, C. Newbold, J. G. Beeson, A. Craig, B. S. Crabb and A. F. Cowman (2008). Exported Proteins Required for

Virulence and Rigidity of *Plasmodium falciparum*-Infected Human Erythrocytes. Cell **134**(1): 48-61.

Malhotra, V. and S. Mayor (2006). Cell biology: the Golgi grows up. Nature **441**(7096): 939-940.

Malin, D. H., J. R. Lake, M. Shenoi, T. P. Upchurch, S. C. Johnson, W. E. Schweinle and C. D. Cadle (1998). The nitric oxide synthesis inhibitor nitro-L-arginine (L-NNA) attenuates nicotine abstinence syndrome in the rat. Psychopharmacology (Berl) **140**(3): 371-377.

Mancias, J. D. and J. Goldberg (2007). The transport signal on Sec22 for packaging into COPII-coated vesicles is a conformational epitope. Mol Cell **26**(3): 403-414.

Mansour, L., G. Prensier, S. B. Jemaa, O. K. Hassine, G. Metenier, C. P. Vivares and E. Cornillot (2005). Description of a xenoma-inducing microsporidian, *Microgemma tincae* n. sp., parasite of the teleost fish *Symphodus tinca* from Tunisian coasts. Dis Aquat Organ **65**(3): 217-226.

Marchler-Bauer, A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng and S. H. Bryant (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res **39**: D225-229.

Martinez, G., K. Georgas, G. A. Challen, B. Rumballe, M. J. Davis, D. Taylor, R. D. Teasdale, S. M. Grimmond and M. H. Little (2006). Definition and spatial annotation of the dynamic secretome during early kidney development. Developmental Dynamics **235**(6): 1709-1719.

Martiny, A., J. R. Meyer-Fernandes, W. de Souza and M. A. Vannier-Santos (1999). Altered tyrosine phosphorylation of ERK1 MAP kinase and other macrophage molecules caused by *Leishmania* amastigotes. Molecular and Biochemical Parasitology **102**(1): 1-12.

Martiny, A., M. A. Vannier-Santos, V. M. Borges, J. R. Meyer-Fernandes, J. Assreuy, N. L. Cunha e Silva and W. de Souza (1996). *Leishmania* induced tyrosine phosphorylation in the host macrophage and its implication to infection. Eur J Cell Biol **71**(2): 206-215.

Matlack, K. E. S., K. Plath, B. Misselwitz and T. A. Rapoport (1997). Protein Transport by Purified Yeast Sec Complex and Kar2p Without Membranes. Science **277**(5328): 938-941.

Matsuura-Tokita, K., M. Takeuchi, A. Ichihara, K. Mikuriya and A. Nakano (2006). Live imaging of yeast Golgi cisternal maturation. Nature **441**(7096): 1007-1010.

Mayack, C. and D. Naug (2009). Energetic stress in the honeybee *Apis mellifera* from *Nosema ceranae* infection. Journal of Invertebrate Pathology **100**(3): 185-188.

McConnachie, S. H., N. J. Guselle and D. J. Speare (2013). Retention of viable microsporidial (*Loma salmonae*) spores within the blue mussel (*Mytilus edulis*): Use of an experimental laboratory model probing pathogen transfer within a multi-trophic aquaculture setting. Aquaculture **5**(0): 376–379.

Merzendorfer, H. and L. Zimoch (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. Journal of Experimental Biology **206**(24): 4393-4412.

Meusser, B., C. Hirsch, E. Jarosch and T. Sommer (2005). ERAD: the long road to destruction. Nat Cell Biol **7**(8): 766-772.

Miles, M. A., M. D. Feliciangeli and A. R. d. Arias (2003). American trypanosomiasis (Chagas' disease) and the role of molecular epidemiology in guiding control strategies. BMJ **326**(7404): 1444-1448.

Miller, L. H., D. I. Baruch, K. Marsh and O. K. Doumbo (2002). The pathogenic basis of malaria. Nature **415**(6872): 673-679.

Miranda-Saavedra, D., M. J. Stark, J. C. Packer, C. P. Vivares, C. Doerig and G. J. Barton (2007). The complement of protein kinases of the microsporidium *Encephalitozoon cuniculi* in relation to those of *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. BMC Genomics **8**: 309.

Misumi, Y., Y. Misumi, K. Miki, A. Takatsuki, G. Tamura and Y. Ikehara (1986). Novel blockade by brefeldin A of intracellular transport of secretory proteins in cultured rat hepatocytes. Journal of Biological Chemistry **261**(24): 11398-11403.

Mnaimneh, S., A. P. Davierwala, J. Haynes, J. Moffat, W. T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez, M. Trochesset, D. Morse, N. J. Krogan, S. L. Hiley, Z. Li, Q. Morris, J. Grigull, N. Mitsakakis, C. J. Roberts, J. F. Greenblatt, C. Boone, C. A. Kaiser, B. J. Andrews and T. R. Hughes (2004). Exploration of essential gene functions via titratable promoter alleles. Cell **118**(1): 31-44.

Mogelsvang, S., B. J. Marsh, M. S. Ladinsky and K. E. Howell (2004). Predicting function from structure: 3D structure studies of the mammalian Golgi complex. Traffic **5**(5): 338-345.

Mohebiany, A. N., R. M. Nikolaienko, S. Bouyain and S. Harroch (2013). Receptor-type tyrosine phosphatase ligands: looking for the needle in the haystack. FEBS Journal **280**(2): 388-400.

Molina, J. M., M. Tourneur, C. Sarfati, S. Chevret, A. de Gouvello, J. G. Gobert, S. Balkan and F. Derouin (2002). Fumagillin treatment of intestinal microsporidiosis. N Engl J Med **346**(25): 1963-1969.

Mullally, J. E., T. Chernova and K. D. Wilkinson (2006). Doa1 is a Cdc48 adapter that possesses a novel ubiquitin binding domain. Mol Cell Biol **26**(3): 822-830.

Nägeli, K. (1857). Uber die neue Krankheit der Seidenraupe und verwandte Organismen. Bot. Z **15**: 760-761.

Nascimento, M., W. W. Zhang, A. Ghosh, D. R. Houston, A. M. Berghuis, M. Olivier and G. Matlashewski (2006). Identification and characterization of a protein-tyrosine phosphatase in *Leishmania*: Involvement in virulence. Journal of Biological Chemistry **281**(47): 36257-36268.

Natsir, H., D. Chandra, Y. Rukayadi, M. T. Suhartono, J. K. Hwang and Y. R. Pyun (2002). Biochemical characteristics of chitinase enzyme from *Bacillus* sp. of Kamojang Crater, Indonesia. J Biochem Mol Biol Biophys **6**(4): 279-282.

Nawa, D., O. Shimada, N. Kawasaki, N. Matsumoto and K. Yamamoto (2007). Stable interaction of the cargo receptor VIP36 with molecular chaperone BiP. Glycobiology **17**(9): 913-921.

Nayak, A. P., B. J. Green and D. H. Beezhold (2013). Fungal hemolysins. Med Mycol **51**(1): 1-16.

Newman, A. P., J. Shim and S. Ferro-Novick (1990). BET1, BOS1, and SEC22 are members of a group of interacting yeast genes required for transport from the endoplasmic reticulum to the Golgi complex. Mol Cell Biol **10**(7): 3405-3414.

Nickel, W. (2003). The mystery of nonclassical protein secretion. European Journal of Biochemistry **270**(10): 2109-2119.

Nickel, W. (2005). Unconventional secretory routes: direct protein export across the plasma membrane of mammalian cells. Traffic **6**(8): 607-614.

Nilsson, I., P. Whitley and G. von Heijne (1994). The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase. The Journal of Cell Biology **126**(5): 1127-1132.

Nishibuchi, M., J. M. Janda and T. Ezaki (1996). The thermostable direct hemolysin gene (tdh) of *Vibrio hollisae* is dissimilar in prevalence to and phylogenetically distant from the tdh genes of other vibrios: implications in the horizontal transfer of the tdh gene. Microbiol Immunol **40**(1): 59-65.

Nkinin, S. W., T. Asonganyi, E. S. Didier and E. S. Kaneshiro (2007). Microsporidian infection is prevalent in healthy people in Cameroon. Journal of clinical microbiology **45**(9): 2841-2846.

Noronha, F. S., F. J. Ramalho-Pinto and M. F. Horta (1994). Identification of a putative pore-forming hemolysin active at acid pH in *Leishmania amazonensis*. Braz J Med Biol Res **27**(2): 477-482.

Nothwehr, S. F. and J. I. Gordon (1989). Eukaryotic signal peptide structure/function relationships. Identification of conformational features which

influence the site and efficiency of co-translational proteolytic processing by site-directed mutagenesis of human pre(delta pro)apolipoprotein A-II. Journal of Biological Chemistry **264**(7): 3979-3987.

Nothwehr, S. F. and J. I. Gordon (1990). Structural features in the NH2-terminal region of a model eukaryotic signal peptide influence the site of its cleavage by signal peptidase. Journal of Biological Chemistry **265**(28): 17202-17208.

Nothwehr, S. F. and J. I. Gordon (1990). Targeting of proteins into the eukaryotic secretory pathway: Signal peptide structure/function relationships. BioEssays **12**(10): 479-484

Olivier, M., D. J. Gregory and G. Forget (2005). Subversion Mechanisms by Which *Leishmania* Parasites Can Escape the Host Immune Response: a Signaling Point of View. Clinical Microbiology Reviews **18**(2): 293-305.

Orci, L., B. S. Glick and J. E. Rothman (1986). A new type of coated vesicular carrier that appears not to contain clathrin: its possible role in protein transport within the Golgi stack. Cell **46**(2): 171-184.

Orci, L., M. Tagaya, M. Amherdt, A. Perrelet, J. G. Donaldson, J. Lippincott-Schwartz, R. D. Klausner and J. E. Rothman (1991). Brefeldin A, a drug that blocks secretion, prevents the assembly of non-clathrin-coated buds on Golgi cisternae. Cell **64**(6): 1183-1195.

Orenstein, J. M., D. T. Dieterich and D. P. Kotler (1992). Systemic dissemination by a newly recognized intestinal microsporidia species in AIDS. AIDS **6**(10): 1143.

Pan, G., J. Xu, T. Li, Q. Xia, S. L. Liu, G. Zhang, S. Li, C. Li, H. Liu, L. Yang, T. Liu, X. Zhang, Z. Wu, W. Fan, X. Dang, H. Xiang, M. Tao, Y. Li, J. Hu, Z. Li, L. Lin, J. Luo, L. Geng, L. Wang, M. Long, Y. Wan, N. He, Z. Zhang, C. Lu, P. J. Keeling, J. Wang, Z. Xiang and Z. Zhou (2013). Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. BMC Genomics **14**: 186.

Pasternak, N. D., and R. Dzikowski (2009). PfEMP1: An antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*. The International Journal of Biochemistry & Cell Biology **41**(7):1463-1466.

Pasteur, L. (1870). Études sur la maladie des vers à soie: Notes et documents, Gauthier-Villars, successeur de Mallet-Bachelier.

Paul, R., C. M. Ewing, J. C. Robinson, F. F. Marshall, K. R. Johnson, M. J. Wheelock and W. B. Isaacs (1997). Cadherin-6, a Cell Adhesion Molecule Specifically Expressed in the Proximal Renal Tubule and Renal Cell Carcinoma. Cancer Res **57**: 2741-2748.

Paul, S. and P. J. Lombroso (2003). Receptor and nonreceptor protein tyrosine phosphatases in the nervous system. Cell Mol Life Sci **60**(11): 2465-2482.

Pelham, H. R. B. and J. E. Rothman (2000). The Debate about Transport in the Golgi—Two Sides of the Same Coin? Cell **102**(6): 713-719.

Peng, W. T., N. J. Krogan, D. P. Richards, J. F. Greenblatt and T. R. Hughes (2004). ESF1 is required for 18S rRNA synthesis in *Saccharomyces cerevisiae*. Nucleic Acids Research **32**(6): 1993-1999.

Pepperkok, R., J. A. Whitney, M. Gomez and T. E. Kreis (2000). COPI vesicles accumulating in the presence of a GTP restricted arf1 mutant are depleted of anterograde and retrograde cargo. J Cell Sci **113**: 135-144.

Pernas, L. and J. C. Boothroyd (2010). Association of host mitochondria with the parasitophorous vacuole during *Toxoplasma* infection is not dependent on rhoptry proteins ROP2/8. Int J Parasitol **40**(12): 1367-1371.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth **8**(10): 785-786.

Peterson, K. J. and N. J. Butterfield (2005). Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record. Proc Natl Acad Sci U S A **102**(27): 9547-9552.

Peuvel-Fanget, I., V. Polonais, D. Brosson, C. Texier, L. Kuhn, P. Peyret, C. Vivares and F. Delbac (2006). EnP1 and EnP2, two proteins associated with the *Encephalitozoon cuniculi* endospore, the chitin-rich inner layer of the microsporidian spore wall. Int J Parasitol **36**(3): 309-318.

Peyretaillade, E., V. Broussolle, P. Peyret, G. Méténier, M. Gouy and C. P. Vivarès (1998). Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. Molecular biology and evolution **15**(6): 683-689.

Picardeau, M., D. M. Bulach, C. Bouchier, R. L. Zuerner, N. Zidane, P. J. Wilson, S. Creno, E. S. Kuczek, S. Bommezzadri, J. C. Davis, A. McGrath, M. J. Johnson, C. Boursaux-Eude, T. Seemann, Z. Rouy, R. L. Coppel, J. I. Rood, A. Lajus, J. K. Davies, C. Medigue and B. Adler (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. PLoS One **3**(2).

Pires, A. B., K. P. Gramacho, D. C. Silva, A. Goes-Neto, M. M. Silva, J. S. Muniz-Sobrinho, R. F. Porto, C. Villela-Dias, M. Brendel, J. C. Cascardo and G. A. Pereira (2009). Early development of *Moniliophthora perniciosa* basidiomata and developmentally regulated genes. BMC Microbiol **9**: 158.

Plemper, R. K. and D. H. Wolf (1999). Endoplasmic reticulum degradation. Reverse protein transport and its end in the proteasome. Mol Biol Rep **26**(1-2): 125-130.

Pleshinger, J. and E. Weidner (1985). The microsporidian spore invasion tube. IV. Discharge activation begins with pH-triggered Ca^{2+} influx. J Cell Biol **100**(6): 1834-1838.

Pohlschröder, M., W. A. Prinz, E. Hartmann and J. Beckwith (1997). Protein Translocation in the Three Domains of Life: Variations on a Theme. Cell **91**(5): 563-566.

Pombert, J. F., M. Selman, F. Burki, F. T. Bardell, L. Farinelli, L. F. Solter, D. W. Whitman, L. M. Weiss, N. Corradi and P. J. Keeling (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**(31): 12638-12643.

Prudovsky, I., A. Mandinova, R. Soldi, C. Bagala, I. Graziani, M. Landriscina, F. Tarantini, M. Duarte, S. Bellum, H. Doherty and T. Maciag (2003). The non-classical export routes: FGF1 and IL-1α point the way. Journal of Cell Science **116**(24): 4871-4881.

Przyborski, J. M., S. K. Miller, J. M. Pfahler, P. P. Henrich, P. Rohrbach, B. S. Crabb and M. Lanzer (2005). Trafficking of STEVOR to the Maurer's clefts in *Plasmodium falciparum* infected erythrocytes. EMBO J **24**(13): 2306-2317.

Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman and R. D. Finn (2012). The Pfam protein families database. Nucleic Acids Research **40**(D1): D290-D301.

Qu, Y., L. Franchi, G. Nunez and G. R. Dubyak (2007). Nonclassical IL-1β Secretion Stimulated by P2X7 Receptors Is Dependent on Inflammasome Activation and Correlated with Exosome Release in Murine Macrophages. The Journal of Immunology **179**(3): 1913-1925.

Rabeeth, M., A. Anitha and G. Srikanth (2011). Purification of an antifungal endochitinase from a potential biocontrol Agent *Streptomyces griseus*. Pak J Biol Sci **14**(16): 788-797.

Rabinovich, E., A. Kerem, K.-U. Fröhlich, N. Diamant and S. Bar-Nun (2002). AAA-ATPase p97/Cdc48p, a Cytosolic Chaperone Required for Endoplasmic Reticulum-Associated Protein Degradation. Molecular and Cellular Biology **22**(2): 626-634.

Rabodonirina, M., M. Bertocchi, I. Desportes-Livage, L. Cotte, H. Levrey, M. Piens, G. Monneret, M. Celard, J. Mornex and M. Mojon (1996). *Enterocytozoon bieneusi* as a cause of chronic diarrhea in a heart-lung transplant recipient who was seronegative for human immunodeficiency virus. Clinical infectious diseases **23**(1): 114-117.

Rambaut, A. and A. Drummond (2003). Tracer: MCMC trace analysis tool. University of Oxford, Oxford.

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature **450**(7170): 663-669.

Ravasi, T., T. Huber, M. Zavolan, A. Forrest, T. Gaasterland, S. Grimmond and D. A. Hume (2003). Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Res **13**(6B): 1430-1442.

Ravindran, S. and J. C. Boothroyd (2008). Secretion of proteins into host cells by Apicomplexan parasites. Traffic **9**(5): 647-656.

Remaley, A. T., R. H. Glew, D. B. Kuhns, R. E. Basford, A. S. Waggoner, L. A. Ernst and M. Pope (1985). *Leishmania donovani*: surface membrane acid phosphatase blocks neutrophil oxidative metabolite production. Exp Parasitol **60**(3): 331-341.

Remaley, A. T., D. B. Kuhns, R. E. Basford, R. H. Glew and S. S. Kaplan (1984). Leishmanial phosphatase blocks neutrophil O_2 production. Journal of Biological Chemistry **259**(18): 11173-11175.

Richards, T. A., R. P. Hirt, B. A. P. Williams and T. M. Embley (2003). Horizontal Gene Transfer and the Evolution of Parasitic Protozoa. Protist **154**(1): 17-32.

Richards, T. A., D. M. Soanes, M. D. M. Jones, O. Vasieva, G. Leonard, K. Paszkiewicz, P. G. Foster, N. Hall and N. J. Talbot (2011). Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc Natl Acad Sci U S A **108**(37): 15258-15263.

Rodgers-Gray, T. P., J. E. Smith, A. E. Ashcroft, R. E. Isaac and A. M. Dunn (2004). Mechanisms of parasite-induced sex reversal in *Gammarus duebeni*. Int J Parasitol **34**(6): 747-753.

Roger, A. J., C. G. Clark and W. F. Doolittle (1996). A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. Proc Natl Acad Sci U S A **93**(25): 14618-14622.

Romão-Dumaresq, A. S., W. L. de Araújo, N.J. Talbot and C. R. Thorton (2012) RNA Interference of Endochitinases in the Sugarcane Endophyte *Trichoderma virens* 223 Reduces Its Fitness as a Biocontrol Agent of Pineapple Disease. PLoS ONE **7**(10)

Ronnebaumer, K., U. Gross and W. Bohne (2008). The nascent parasitophorous vacuole membrane of *Encephalitozoon cuniculi* is formed by host cell lipids and contains pores which allow nutrient uptake. Eukaryot Cell **7**(6): 1001-1008.

Ronnebaumer, K., J. Wagener, U. Gross and W. Bohne (2006). Identification of novel developmentally regulated genes in *Encephalitozoon cuniculi*: an endochitinase, a chitin-synthase, and two subtilisin-like proteases are induced during meront-to-sporont differentiation. J Eukaryot Microbiol **53**: 74-76.

Rothman, J. E. and L. Orci (1992). Molecular dissection of the secretory pathway. Nature **355**(6359): 409-415.

Ruzickova, V. (1994). A rapid method for the differentiation of *Staphylococcus aureus* hemolysins. Folia Microbiol (Praha) **39**(2): 112-114.

Sahdev, S., S. K. Khattar and K. S. Saini (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem **307**(1-2): 249-264.

Sam-Yellowe, T. Y. (1996). Rhoptry organelles of the apicomplexa: Their role in host cell invasion and intracellular survival. Parasitology Today **12**(8): 308-316.

Sandfort, J., A. Hannemann, H. Gelderblom, K. Stark, R. L. Owen and B. Ruf (1994). *Enterocytozoon bieneusi* infection in an immunocompetent patient who had acute diarrhea and who was not infected with the human immunodeficiency virus. Clinical infectious diseases **19**(3): 514-516.

Santana, J. M., P. Grellier, J. Schrevel and A. R. Teixeira (1997). A *Trypanosoma cruzi*-secreted 80 kDa proteinase with specificity for human collagen types I and IV. Biochem J **325**(1): 129-137.

Scanlon, M., G. J. Leitch, G. S. Visvesvara and A. P. Shaw (2004). Relationship between the Host Cell Mitochondria and the Parasitophorous Vacuole in Cells Infected with *Encephalitozoon* Microsporidia. Journal of Eukaryotic Microbiology **51**(1): 81-87.

Scanlon, M., A. P. Shaw, C. J. Zhou, G. S. Visvesvara and G. J. Leitch (2000). Infection by microsporidia disrupts the host cell cycle. J Eukaryot Microbiol **47**(6): 525-531.

Schaheen, B., H. Dang and H. Fares (2009). Derlin-dependent accumulation of integral membrane proteins at cell surfaces. Journal of Cell Science **122**(13): 2228-2239.

Schallus, T., C. Jaeckh, K. Feher, A. S. Palma, Y. Liu, J. C. Simpson, M. Mackeen, G. Stier, T. J. Gibson, T. Feizi, T. Pieler and C. Muhle-Goll (2008). Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein *N*-glycosylation. Mol Biol Cell **19**(8): 3404-3414.

Schornack, S., E. Huitema, L. M. Cano, T. O. Bozkurt, R. Oliva, M. Van Damme, S. Schwizer, S. Raffaele, A. Chaparro-Garcia, R. Farrer, M. E. Segretin, J. Bos, B. J. Haas, M. C. Zody, C. Nusbaum, J. Win, M. Thines and S. Kamoun (2009). Ten things to know about oomycete effectors. Mol Plant Pathol **10**(6): 795-803.

Schottelius, J., C. Schmetz, N. P. Kock, T. Schuler, I. Sobottka and B. Fleischer (2000). Presentation by scanning electron microscopy of the lifecycle of microsporidia of the genus *Encephalitozoon*. Microbes Infect **2**(12): 1401-1406.

Schuldiner, M., S. R. Collins, N. J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J. F. Greenblatt, J. S. Weissman and N. J. Krogan (2005). Exploration of the function and organization of the yeast

early secretory pathway through an epistatic miniarray profile. Cell **123**(3): 507-519.

Sebghati, T. S., J. T. Engle and W. E. Goldman (2000). Intracellular Parasitism by *Histoplasma capsulatum*: Fungal Virulence and Calcium Dependence. Science **290**(5495): 1368-1372.

Seidl, V. (2008). Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. Fungal Biology Reviews **22**(1): 36-42.

Selman, M. and N. Corradi (2011). Microsporidia: Horizontal gene transfers in vicious parasites. Mobile genetic elements **1**(4): 251-255.

Selman, M., J.-F. Pombert, L. Solter, L. Farinelli, L. M. Weiss, P. Keeling and N. Corradi (2011). Acquisition of an animal gene by microsporidian intracellular parasites. Current Biology **21**(15): R576-R577.

Shakarian, A. M., M. B. Joshi, E. Ghedin and D. M. Dwyer (2002). Molecular dissection of the functional domains of a unique, tartrate-resistant, surface membrane acid phosphatase in the primitive human pathogen *Leishmania donovani*. Journal of Biological Chemistry **277**(20): 17994-18001.

Shaw, R. W., M. L. Kent and M. L. Adamson (1998). Modes of transmission of *Loma salmonae* (Microsporidia). Diseases of Aquatic Organisms **33**(2): 151-156.

Shi, Y. (2009). Serine/Threonine Phosphatases: Mechanism through Structure. Cell **139**(3): 468-484.

Shim, J., A. P. Newman and S. Ferro-Novick (1991). The BOS1 gene encodes an essential 27-kD putative membrane protein that is required for vesicular transport from the ER to the Golgi complex in yeast. J Cell Biol **113**(1): 55-64.

Short, B., A. Haas and F. A. Barr (2005). Golgins and GTPases, giving identity and structure to the Golgi apparatus. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1744**(3): 383-395.

Silverman, J. M., J. Clos, C. C. de'Oliveira, O. Shirvani, Y. Fang, C. Wang, L. J. Foster and N. E. Reiner (2010). An exosome-based secretion pathway is responsible for protein export from *Leishmania* and communication with macrophages. J Cell Sci **123**(6): 842-852.

Simonsen, A., B. Bremnes, E. Ronning, R. Aasland and H. Stenmark (1998). Syntaxin-16, a putative Golgi t-SNARE. Eur J Cell Biol **75**(3): 223-231.
Sinai, A. P. and K. A. Joiner (2001). The *Toxoplasma gondii* protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane. J Cell Biol **154**(1): 95-108.

Sing, A., K. Tybus, J. Heesemann and A. Mathis (2001). Molecular Diagnosis of an *Enterocytozoon bieneusi* Human Genotype C Infection in a Moderately Immunosuppressed Human Immunodeficiency Virus-Seronegative Liver-

Transplant Recipient with Severe Chronic Diarrhea. Journal of clinical microbiology **39**(6): 2371-2372.

Siqueira, J. A., C. Speeg-Schatz, F. I. Freitas, J. Sahel, H. Monteil and G. Prevost (1997). Channel-forming leucotoxins from *Staphylococcus aureus* cause severe inflammatory reactions in a rabbit eye model. J Med Microbiol **46**(6): 486-494.

Sit, S.-T. and E. Manser (2011). Rho GTPases and their role in organizing the actin cytoskeleton. Journal of Cell Science **124**(5): 679-683.

Sitja-Bobadilla, A. (2008). Living off a fish: a trade-off between parasites and the immune system. Fish Shellfish Immunol **25**(4): 358-372.

Slamovits, C. H., L. Burri and P. J. Keeling (2006). Characterization of a divergent Sec61beta gene in microsporidia. J Mol Biol **359**(5): 1196-1202.

Slamovits, C. H., N. M. Fast, J. S. Law and P. J. Keeling (2004). Genome compaction and stability in microsporidian intracellular parasites. Curr Biol **14**(10): 891-896.

Slamovits, C. H. and P. J. Keeling (2004). Class II Photolyase in a Microsporidian Intracellular Parasite. Journal of Molecular Biology **341**(3): 713-721.

Sobottka, I., K. Bartscht, P. Schafer, T. Weitzel, J. Schottelius, N. Kock and R. Laufs (2002). In vitro activity of polyoxin D and nikkomycin Z against *Encephalitozoon cuniculi*. Parasitol Res **88**(5): 451-453.

Solter, L. F. and J. V. Maddox (1998). Timing of an Early Sporulation Sequence of Microsporidia in the Genus *Vairimorpha* (Microsporidia: Burenellidae). Journal of Invertebrate Pathology **72**(3): 323-329.

Soltys, B. J. and R. S. Gupta (1994). Presence and cellular distribution of a 60-kDa protein related to mitochondrial hsp60 in *Giardia lamblia*. J Parasitol **80**(4): 580-590.

Southern, T. R., C. E. Jolly, M. E. Lester and J. R. Hayman (2007). EnP1, a microsporidian spore wall protein that enables spores to adhere to and infect host cells *in vitro*. Eukaryot Cell **6**(8): 1354-1362.

Speare, D., R. Markham and N. Guselle (2007). Development of an effective whole-spore vaccine to protect against microsporidial gill disease in rainbow trout (*Oncorhynchus mykiss*) by using a low-virulence strain of *Loma salmonae*. Clinical and Vaccine Immunology **14**(12): 1652-1654.

Stagg, S. M., C. Gurkan, D. M. Fowler, P. LaPointe, T. R. Foss, C. S. Potter, B. Carragher and W. E. Balch (2006). Structure of the Sec13/31 COPII coat cage. Nature **439**(7073): 234-238.

Stentiford, G. D. and K. S. Bateman (2007). *Enterospora* sp., an intranuclear microsporidian infection of hermit crab *Eupagurus bernhardus*. Diseases of aquatic organisms **75**(1): 73-78.

Strain, J., C. R. Lorenz, J. Bode, S. Garland, G. A. Smolen, D. T. Ta, L. E. Vickery and V. C. Culotta (1998). Suppressors of Superoxide Dismutase (SOD1) Deficiency in *Saccharomyces cerevisiae*: Identification of proteins predicted to mediate iron-sulphur cluster assembly. Journal of Biological Chemistry **273**(47): 31138-31144.

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

Sturbaum, G. D., D. A. Schaefer, B. H. Jost, C. R. Sterling and M. W. Riggs (2008). Antigenic differences within the *Cryptosporidium hominis* and *Cryptosporidium parvum* surface proteins P23 and GP900 defined by monoclonal antibody reactivity. Molecular and Biochemical Parasitology **159**(2): 138-141.

Su, W., Y. Liu, Y. Xia, Z. Hong and J. Li (2012). The *Arabidopsis* homolog of the mammalian OS-9 protein plays a key role in the endoplasmic reticulum-associated degradation of misfolded receptor-like kinases. Mol Plant **5**(4): 929-940.

Sunderman, F. W., Jr. (1995). The influence of zinc on apoptosis. Ann Clin Lab Sci **25**(2): 134-142.

Susko, E. and A. J. Roger (2007). On Reduced Amino Acid Alphabets for Phylogenetic Inference. Molecular Biology and Evolution **24**(9): 2139-2150.

Suzuki, T., H. Park, N. M. Hollingsworth, R. Sternglanz and W. J. Lennarz (2000). PNG1, a yeast gene encoding a highly conserved peptide: *N*-glycanase. J Cell Biol **149**(5): 1039-1052.

Taib, M., J. W. Pinney, D. R. Westhead, K. J. McDowell and D. J Adams (2005). Differential expression and extent of fungal/plant and fungal/bacterial chitinases of *Aspergillus fumigatus*. Archives of Microbiology **184**(1): 78-81.

Takvorian, P. M., L. M. Weiss and A. Cali (2005). The early events of *Brachiola algerae* (Microsporidia) infection: spore germination, sporoplasm structure, and development within host cells. Folia Parasitol (Praha) **52**(1-2): 118-129.

Tang, G., T. Iida, K. Yamamoto and T. Honda (1997). Analysis of functional domains of *Vibrio parahaemolyticus* thermostable direct hemolysin using monoclonal antibodies. FEMS Microbiol Lett **150**(2): 289-296.

Taupin, V., E. Garenaux, M. Mazet, E. Maes, H. Denise, G. Prensier, C. P. Vivares, Y. Guerardel and G. Metenier (2007). Major *O*-glycans in the spores of two microsporidian parasites are represented by unbranched manno-oligosaccharides containing alpha-1,2 linkages. Glycobiology **17**(1): 56-67.

Taylor, S., A. Barragan, C. Su, B. Fux, S. J. Fentress, K. Tang, W. L. Beatty, H. E. Hajj, M. Jerome, M. S. Behnke, M. White, J. C. Wootton and L. D. Sibley (2006). A Secreted Serine-Threonine Kinase Determines Virulence in the Eukaryotic Pathogen *Toxoplasma gondii*. Science **314**(5806): 1776-1780.

Taylor, S. C., P. Thibault, D. C. Tessier, J. J. Bergeron and D. Y. Thomas (2003). Glycopeptide specificity of the secretory protein folding sensor UDP-glucose glycoprotein:glucosyltransferase. EMBO Rep **4**(4): 405-411.

Texier, C., C. Vidau, B. Vigues, H. El Alaoui and F. Delbac (2010). Microsporidia: a model for minimal parasite-host interactions. Curr Opin Microbiol **13**(4): 443-449.

Thelohan, P. (1895). Recherches sur les Myxosporidies, G. Carré & P. Klincksieck.

Thollesson, M. (2004). LDDist: a Perl module for calculating LogDet pair-wise distances for protein and nucleotide sequences. Bioinformatics **20**(3): 416-418.

Thomarat, F., C. P. Vivares and M. Gouy (2004). Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. J Mol Evol **59**(6): 780-791.

Thompson, A., J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed and C. Hamon (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem **75**(8): 1895-1904.

Timmers, A. C. J., R. Stuger, P. J. Schaap, J. van 't Riet and H. A. Raué (1999). Nuclear and nucleolar localization of *Saccharomyces cerevisiae* ribosomal proteins S22 and S25. FEBS Letters **452**(3): 335-340.

Tovar, J., A. Fischer and C. G. Clark (1999). The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. Molecular Microbiology **32**(5): 1013-1021.

Tsaousis, A. D., E. R. Kunji, A. V. Goldberg, J. M. Lucocq, R. P. Hirt and T. M. Embley (2008). A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature **453**(7194): 553-556.

Umejiego, N. N., D. Gollapalli, L. Sharling, A. Volftsun, J. Lu, N. N. Benjamin, A. H. Stroupe, T. V. Riera, B. Striepen and L. Hedstrom (2008). Targeting a Prokaryotic Protein in a Eukaryotic Pathogen: Identification of Lead Compounds against Cryptosporidiosis. Chemistry & Biology **15**(1): 70-77.

Undeen, A. H. and R. K. V. Meer (1994). Conversion of Intrasporal Trehalose into Reducing Sugars During Germination of *Nosema algerae* (Protista: Microspora) Spores: A Quantitative Study. Journal of Eukaryotic Microbiology **41**(2): 129-132.

Urch, J. E., R. Hurtado-Guerrero, D. Brosson, Z. Liu, V. G. Eijsink, C. Texier and D. M. van Aalten (2009). Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite *Encephalitozoon cuniculi*. Protein Sci **18**(6): 1197-1209.

Ustrell, V., L. Hoffman, G. Pratt and M. Rechsteiner (2002). PA200, a nuclear proteasome activator involved in DNA repair. EMBO J **21**(13): 3516-3525.

Vale, R. D., T. S. Reese and M. P. Sheetz (1985). Identification of a Novel Force-Generating Protein, Kinesin, Involved in Microtubule-Based Motility. Cell **42**(1): 39-50

Van de Peer, Y., A. Ben Ali and A. Meyer (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. Gene **246**(1-2): 1-8.

van der Wijk, T., C. Blanchetot and J. den Hertog (2005). Regulation of receptor protein-tyrosine phosphatase dimerization. Methods **35**(1): 73-79.

van Ooij, C., P. Tamez, S. Bhattacharjee, N. L. Hiller, T. Harrison, K. Liolios, T. Kooij, J. Ramesar, B. Balu, J. Adams, A. Waters, C. Janse and K. Haldar (2008). The Malaria Secretome: From Algorithms to Essential Function in Blood Stage Infection. PLoS Pathog **4**(6): e1000084.

Vandenbol, M. and C. Fairhead (2000). Mass-murder deletion of 19 ORFs from *Saccharomyces cerevisiae* chromosome XI. Gene **247**(1-2): 45-52.

Vashist, S. and D. T. W. Ng (2004). Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. The Journal of Cell Biology **165**(1): 41-52.

Vavra, J., R. Dahbiova, W. S. Hollister and E. U. Canning (1993). Staining of microsporidian spores by optical brighteners with remarks on the use of brighteners for the diagnosis of AIDS associated human microsporidioses. Folia Parasitol (Praha) **40**(4): 267-272.

Vemuganti, G., J. Joseph and S. Sharma (2005). Microsporidia: Emerging Ocular Pathogens.

Venter, J. C., *et al* (2001). The Sequence of the Human Genome. Science **291**(5507): 1304-1351.

Visintin, R. and A. Amon (2000). The nucleolus: the magician's hat for cell cycle tricks. Current Opinion in Cell Biology **12**(6): 752.

Visvesvara, G. S., H. Moura, G. J. Leitch and D. A. Schwartz (1999). Culture and propagation of microsporidia. The Microsporidia and Microsporidiosis. ASM Press, Washington, DC: 363-392.

Vivier, E. (1975). The microsporidia of the protozoa. Protistologica **11**: 345-361.

Vossbrinck, C. R. and B. A. Debrunner-Vossbrinck (2005). Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. Folia Parasitol (Praha) **52**(1-2): 131-142.

Vossbrinck, C. R., J. V. Maddox, S. Friedman, B. A. Debrunner-Vossbrinck and C. R. Woese (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature **326**(6111): 411-414.

Vossbrinck, C. R. and C. R. Woese (1986). Eukaryotic ribosomes that lack a 5.8S RNA. Nature **320**(6059): 287-288.

Wallace, A. J., T. J. Stillman, A. Atkins, S. J. Jamieson, P. A. Bullough, J. Green and P. J. Artymiuk (2000). *E. coli* Hemolysin E (HlyE, ClyA, SheA): X-Ray Crystal Structure of the Toxin and Observation of Membrane Pores by Electron Microscopy. Cell **100**(2): 265-276.

Waller, R. F., C. Jabbour, N. C. Chan, N. Celik, V. A. Likic, T. D. Mulhern and T. Lithgow (2009). Evidence of a reduced and modified mitochondrial protein import apparatus in microsporidian mitosomes. Eukaryot Cell **8**(1): 19-26.

Walter, P. and G. Blobel (1981). Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in-vitro-assembled polysomes synthesizing secretory protein. J Cell Biol **91**(2 Pt 1): 551-556.

Wang, W. Q., J. P. Sun and Z. Y. Zhang (2003). An overview of the protein tyrosine phosphatase superfamily. Curr Top Med Chem **3**(7): 739-748.

Weber, R., R. T. Bryan, R. L. Owen, C. M. Wilcox, L. Gorelkin and G. S. Visvesvara (1992). Improved light-microscopical detection of microsporidia spores in stool and duodenal aspirates. New England Journal of Medicine **326**(3): 161-166.

Weedall, R. T., M. Robinson, J. E. Smith and A. M. Dunn (2006). Targeting of host cell lineages by vertically transmitted, feminising microsporidia. International Journal for Parasitology **36**(7): 749-756.

Weidner, E. and W. Byrd (1982). The microsporidian spore invasion tube. II. Role of calcium in the activation of invasion tube discharge. J Cell Biol **93**(3): 970-975.

Weidner, E., W. Byrd, A. N. N. Scarborough, J. Pleshinger and D. Sibley (1984). Microsporidian Spore Discharge and the Transfer of Polaroplast Organelle Membrane into Plasma Membrane. Journal of Eukaryotic Microbiology **31**(2): 195-198.

Weidner, E. and A. Findley (2003). Catalase in microsporidian spores before and during discharge. Biol Bull **205**(2): 236-237.

Weissenberg, R., J. LA Bulla and T. Cheng (1976). Microsporidian interactions with host cells. Comparative pathobiology. Volume 1. Biology of the Microsporidia 203-237.

West, R. W., Jr., R. R. Yocum and M. Ptashne (1984). *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. Mol Cell Biol **4**(11): 2467-2478.

Wheat, L. J., P. A. Connolly-Stringfield, R. L. Baker, M. F. Curfman, M. E. Eads, K. S. Israel, S. A. Norris, D. H. Webb and M. L. Zeckel (1990). Disseminated histoplasmosis in the acquired immune deficiency syndrome: clinical findings, diagnosis and treatment, and review of the literature. Medicine **69**(6): 361-374.

Whelan, S. and N. Goldman (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Molecular Biology and Evolution **18**(5): 691-699.

Wiles, T. J. and M. A. Mulvey (2013). The RTX pore-forming toxin alpha-hemolysin of uropathogenic *Escherichia coli*: progress and perspectives. Future Microbiol **8**(1): 73-84.

Williams, B. A. (2009). Unique physiology of host-parasite interactions in microsporidia infections. Cell Microbiol **11**(11): 1551-1560.

Williams, B. A., A. Cali, P. M. Takvorian and P. J. Keeling (2008). Distinct localization patterns of two putative mitochondrial proteins in the microsporidian *Encephalitozoon cuniculi*. J Eukaryot Microbiol **55**(2): 131-133.

Williams, B. A. P., R. P. Hirt, J. M. Lucocq and T. M. Embley (2002). A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature **418**(6900): 865-869.

Williams, B. A. P. and P. J. Keeling (2011). Microsporidia – Highly Reduced and Derived Relatives of Fungi. Springer Verlag.

Williams, B. A. P., C. H. Slamovits, N. J. Patron, N. M. Fast and P. J. Keeling (2005). A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc Natl Acad Sci U S A **102**(31): 10936-10941.

Winkler, H. H. and H. E. Neuhaus (1999). Non-mitochondrial ATP transport. Trends in Biochemical Sciences **24**(2): 64-68.

Wiseman, G. M. (1975). The hemolysins of *Staphylococcus aureus*. Bacteriological Reviews **39**(4): 317-344.

Wittner, M. and L. M. Weiss (1999). The microsporidia and microsporidiosis. Washington, D.C., ASM Press.

Wojczyk, B. S., M. Stwora-Wojczyk, S. Shakin-Eshleman, W. H. Wunner and S. L. Spitalnik (1998). The role of site-specific *N*-glycosylation in secretion of soluble forms of rabies virus glycoprotein. Glycobiology **8**(2): 121-130.

Wu, Z., Y. Li, G. Pan, G. Li, J. Hu, H. Liu, Z. Zhou and Z. Xiang (2007). A Complete Sec61 Complex in *Nosema Bombycis* and Its Comparative Genomics Analyses1. Journal of Eukaryotic Microbiology **54**(4): 379-380.

Wu, Z., Y. Li, G. Pan, G. Li, J. Hu, Z. Zhou and Z. Xiang (2008). Proteomic analysis of spore wall proteins and identification of two spore wall proteins from *Nosema bombycis* (Microsporidia). Proteomics **8**(12): 2447-2461.

Xiang, H., G. Pan, C. Vossbrinck, R. Zhang, J. Xu, T. Li, Z. Zhou, C. Lu and Z. Xiang (2010). A Tandem Duplication of Manganese Superoxide Dismutase in *Nosema bombycis* and Its Evolutionary Origins. Journal of Molecular Evolution **71**(5): 401-414.

Xiang, H., G. Pan, C. R. Vossbrinck, R. Zhang, J. Xu, T. Li, Z. Zhou, C. Lu and Z. Xiang (2010). A tandem duplication of manganese superoxide dismutase in *Nosema bombycis* and its evolutionary origins. J Mol Evol **71**(5-6): 401-414.

Xu, Y., P. M. Takvorian, A. Cali, G. Orr and L. M. Weiss (2004). Glycosylation of the major polar tube protein of *Encephalitozoon hellem*, a microsporidian parasite that infects humans. Infect Immun **72**(11): 6341-6350.

Xu, Y. and L. M. Weiss (2005). The microsporidian polar tube: a highly specialised invasion organelle. Int J Parasitol **35**(9): 941-953.

Yeoh, S., R. A. O'Donnell, K. Koussis, A. R. Dluzewski, K. H. Ansell, S. A. Osborne, F. Hackett, C. Withers-Martinez, G. H. Mitchell, L. H. Bannister, J. S. Bryans, C. A. Kettleborough and M. J. Blackman (2007). Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. Cell **131**(6): 1072-1083.

Yu, L., L. Pena Castillo, S. Mnaimneh, T. R. Hughes and G. W. Brown (2006). A survey of essential gene function in the yeast cell division cycle. Mol Biol Cell **17**(11): 4736-4747.

Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character **213**: 21-87.

Zhang, B. and T. F. Miller (2010). Hydrophobically stabilized open state for the lateral gate of the Sec translocon. Proc Natl Acad Sci U S A **107**(12): 5399-5404.

Zhang, D., T. Chen, I. Ziv, R. Rosenzweig, Y. Matiuhin, V. Bronner, M. H. Glickman and D. Fushman (2009). Together, Rpn10 and Dsk2 can serve as a polyubiquitin chain-length sensor. Mol Cell **36**(6): 1018-1033.

Zhang, H., X. Zha, Y. Tan, P. V. Hornbeck, A. J. Mastrangelo, D. R. Alessi, R. D. Polakiewicz and M. J. Comb (2002). Phosphoprotein Analysis Using Antibodies Broadly Reactive against Phosphorylated Motifs. Journal of Biological Chemistry **277**(42): 39379-39387.

Zielinska, D. F., F. Gnad, J. R. Wiśniewski and M. Mann (2010). Precision Mapping of an *In Vivo N*-Glycoproteome Reveals Rigid Topological and Sequence Constraints. Cell **141**(5): 897-907.

Appendix 1| Recipes for standard reagents and media

Luria-Bertani broth

10 g Bacto-tryptone (Difco)

5 g Yeast extract (Oxoid)

10 g NaCl (Fisher Scientific)

dH_20 to 1 Litre and autoclave

Luria-Bertani agar

10 g Bacto-tryptone

5 g Yeast extract

10 g NaCl

20 g Bacteriological agar No.2 (Lab M)

$dH_20$to 1 Litre and autoclave

Yeast Peptone Dextrose broth

10 g Yeast extract

20 g Bacto-peptone (Oxoid)

20 g Dextrose (Fisher Scientific)

$dH_20$to 1 Litre and autoclave

Yeast Peptone Dextrose agar

10 g Yeast extract

20 g Bacto-peptone

20 g Dextrose

20 g Bacteriological agar No.2

$dH_20$to 1 Litre and autoclave

Complete synthetic media (broth)

6.9 g Yeast Nitrogen Base (Formedium)

690 mg Synthetic amino acid drop out medium (Formedium) minus Leucine (770 mg minus Uracil)

10 g Glucose

$dH_20$to 1 Litre and autoclave

Complete synthetic media (agar)

6.9 g Yeast Nitrogen Base

690 mg Synthetic amino acid drop out medium minus Leucine (770 mg minus Uracil)

10 g Glucose

$dH_20$to 1 litre and autoclave

20 g Bacteriological agar No.2

| Tris-acetate EDTA electrophoresis buffer

4.8 g Tris base (Melford)

0.37 g EDTA (Fisher Scientific)

1.4 ml Glacial acetic acid (Fisher Scientific)

$dH_20$to 1 litre

Protein lysis buffer

300 mg Tris base

pH to 8.8 in 45 ml $_dH_20$

440 mg Sodium chloride

500 µl Triton X-100

$dH_20$to 50 ml

SDS running buffer (per 12% gel)

2.9 ml $_dH_20$

4 ml 1M Tris, pH 8.8

3 ml 38% Acrylamide/2% Bisacrylamide (Amresco)

108 µl 10% Sodium dodecyl sulphate (Fisher Scientific)

54 µl 10% Ammonium persulpate (Sigma Aldrich)

10 µl Tetramethylethylenediamine (TEMED, Amresco)

SDS stacking buffer (per gel)

2.85 ml $_dH_20$

570 µl 1M Tris, pH 6.8

350 µl 38% Acrylamide/2% Bisacrylamide

40 µl 10% Sodium dodecyl sulphate

38 µl 10% Ammonium persulpate

5 µl Tetramethylethylenediamine

SDS tank buffer

3.1 g Tris base

14.4 g Glycine

1 g Sodium dodecyl sulphate

$dH_20$to 1 Litre

Destain I

75 ml Glacial acetic acid

100 ml Methanol

$dH_20$to 1 Litre

Destain II

70 ml Glacial acetic acid

50 ml Methanol

$dH_20$to 1 Litre

Coomassie blue

2 g Coomassie Brilliant Blue R-250

500 ml Methanol

100 ml Acetic acid

$dH_20$to 1 litre

Phosphate-buffered saline

8 g NaCl

0.2 g KCl

1.44 g Na_2HPO_4

0.24 g KH_2PO_4

$dH_20$to 1 Litre and autoclave if required

Tris-buffered saline + Tween (TBST)

3 g Tris

8.8 g NaCl

0.2 g KCl

$dH_20$to 800 ml

500 µl Tween-20

pH to 7.4

$dH_20$to 1 litre and filter if required

Tris-EDTA buffer

1.2 g Tris

0.37 g EDTA

$dH_20$to 1 litre

pH 7.5 using HCl

Towbin Transfer Buffer

3.1 g Tris base

14.4 g Glycine

200 ml Methanol

$dH_20$to 1 litre

Appendix 2| A full list of proteins identified in whole cell protein extracts from germinated and non-germinated spores.

Germinated		Non-germinated	
S. lophii ID	**Annotation**	**S. lophii ID**	**Annotation**
SLOPH 10	20S proteasome subunit beta type-	SLOPH 10	20S proteasome subunit beta type-1
SLOPH 101	Actin	SLOPH 101	Actin
SLOPH 108	Adenylate kinase	SLOPH 108	Adenylate kinase
SLOPH 1081	Hypothetical	SLOPH 1081	Hypothetical
SLOPH 11	26S protease regulatory subunit 6	SLOPH 11	26S protease regulatory subunit 6
SLOPH 1179	Hypothetical	SLOPH 1179	Hypothetical
SLOPH 12	26S protease regulatory subunit 6a	SLOPH 12	26S protease regulatory subunit 6a
SLOPH 1230	Inorganic pyrophosphatase	SLOPH 1230	Inorganic pyrophosphatase
SLOPH 1233	Isoleucine tRNA synthetase	SLOPH 1233	Isoleucine tRNA synthetase
SLOPH 13	26S protease regulatory subunit 8	SLOPH 13	26S protease regulatory subunit 8
SLOPH 1334	Lysyl-tRNA synthetase	SLOPH 1334	Lysyl-tRNA synthetase
SLOPH 136	Argonaute protein	SLOPH 136	Argonaute protein
SLOPH 1414	Peptidase M1	SLOPH 1414	Peptidase M1
SLOPH 1416	Peptidase M18	SLOPH 1416	Peptidase M18
SLOPH 1419	Peptidyl-prolyl cis-trans isomerase	SLOPH 1419	Peptidyl-prolyl cis-trans isomerase
SLOPH 143	Asparaginyl-tRNA synthetase	SLOPH 143	Asparaginyl-tRNA synthetase
SLOPH 1448	Phosphomannomutase	SLOPH 1448	Phosphomannomutase
SLOPH 1466	Polar tube protein PTP2	SLOPH 1466	polar tube protein PTP2
SLOPH 147	Aspartyl-tRNA synthetase	SLOPH 147	Aspartyl-tRNA synthetase
SLOPH 1488	Prolyl-tRNA synthetase	SLOPH 1488	Prolyl-tRNA synthetase
SLOPH 1540	RAS GTPase	SLOPH 1540	RAS GTPase
SLOPH 1572	Ribosomal protein S26	SLOPH 1572	Ribosomal protein S26
SLOPH 16	26S Proteasome core subunit-alpha-6	SLOPH 16	26S Proteasome core subunit-alpha-6
SLOPH 1662	Hypothetical	SLOPH 1662	Hypothetical
SLOPH 1788	Hypothetical	SLOPH 1788	Hypothetical
SLOPH 1854	Hypothetical	SLOPH 1854	Hypothetical
SLOPH 1923	T complex protein 1 subunit beta	SLOPH 1923	T complex protein 1 subunit beta
SLOPH 1925	T-complex protein 1 subunit alpha	SLOPH 1925	T-complex protein 1 subunit alpha
SLOPH 1945	Thioredoxin	SLOPH 1945	Thioredoxin
SLOPH 1976	Transketolase	SLOPH 1976	Transketolase
SLOPH 1979	Translation elongation factor 1 alpha	SLOPH 1979	Translation elongation factor 1 alpha
SLOPH 1980	Translation elongation factor 2	SLOPH 1980	Translation elongation factor 2
SLOPH 1995	Trehalose-6-phosphate phosphatase	SLOPH 1995	Trehalose-6-phosphate phosphatase
SLOPH 1996	Triose phosphate isomerase	SLOPH 1996	Triose phosphate isomerase
SLOPH 2039	Ubiquitin/40s ribosomal protein S27a	SLOPH 2039	Ubiquitin/40s ribosomal protein S27a fusion
SLOPH 204	Chitin/polysaccharide deacetylase	SLOPH 204	Chitin/polysaccharide deacetylase
SLOPH 2044	UTP-glucose-1-phosphate uridylyltransferase	SLOPH 2044	UTP-glucose-1-phosphate uridylyltransferase
SLOPH 2046	V-type ATPase subunit A	SLOPH 2046	V-type ATPase subunit A

SLOPH 2084	Zinc finger protein	SLOPH 2084	Zinc finger protein
SLOPH 2170	Aldose reductase	SLOPH 2170	Aldose reductase
SLOPH 2185	Ubiquitin-activating enzyme E1	SLOPH 2185	Ubiquitin-activating enzyme E1
SLOPH 2189	40S ribosomal protein S2	SLOPH 2189	40S ribosomal protein S2
SLOPH 2210	60S ribosomal protein L6	SLOPH 2210	60S ribosomal protein L6
SLOPH 2213	Actin-depolymerizing factor	SLOPH 2213	Actin-depolymerizing factor
SLOPH 2232	Guanine nucleotide binding protein beta subunit	SLOPH 2232	Guanine nucleotide binding protein beta subunit
SLOPH 2233	Ribosomal protein L23	SLOPH 2233	Ribosomal protein L23
SLOPH 2310	Enolase	SLOPH 2310	Enolase
SLOPH 2338	Alpha trehalose-phosphate synthase	SLOPH 2338	Alpha trehalose-phosphate synthase
SLOPH 2363	Glucose-6-phosphate isomerase	SLOPH 2363	Glucose-6-phosphate isomerase
SLOPH 2369	14-3-3 protein	SLOPH 2369	14-3-3 protein
SLOPH 24	26S proteosome subunit alpha-4	SLOPH 24	26S proteosome subunit alpha-4
SLOPH 25	40S ribosomal protein S0	SLOPH 25	40S ribosomal protein S0
SLOPH 2449	Alpha-alpha-trehalase	SLOPH 2449	Alpha-alpha-trehalase
SLOPH 255	Cytosol aminopeptidase	SLOPH 255	Cytosol aminopeptidase
SLOPH 2557	Hypothetical	SLOPH 2557	Hypothetical
SLOPH 263	DEAD/DEAH box helicase	SLOPH 263	DEAD/DEAH box helicase
SLOPH 2653	metallopeptidase M24	SLOPH 2653	Metallopeptidase M24
SLOPH 28	40S ribosomal protein S12	SLOPH 28	40S ribosomal protein S12
SLOPH 33	40S ribosomal protein S18	SLOPH 33	40S ribosomal protein S18
SLOPH 346	Eukaryotic translation initiation factor 4A	SLOPH 346	Eukaryotic translation initiation factor 4A
SLOPH 348	Eukaryotic translation initiation factor 6	SLOPH 348	Eukaryotic translation initiation factor 6
SLOPH 377	Fructose biphosphate aldolase B	SLOPH 377	Fructose biphosphate aldolase B
SLOPH 38	40S ribosomal protein S27	SLOPH 38	40S ribosomal protein S27
SLOPH 399	Glyceraldehyde-3-phosphate dehydrogenase	SLOPH 399	Glyceraldehyde-3-phosphate dehydrogenase
SLOPH 40	40S ribosomal protein S3	SLOPH 40	40S ribosomal protein S3
SLOPH 403	Glycosyl transferase	SLOPH 403	Glycosyl transferase
SLOPH 409	GTP binding protein	SLOPH 409	GTP binding protein
SLOPH 416	Heat shock protein 70	SLOPH 416	Heat shock protein 70
SLOPH 417	Heat shock protein 90	SLOPH 417	Heat shock protein 90
SLOPH 418	Heat shock protein 90	SLOPH 418	Heat shock protein 90
SLOPH 43	40S ribosomal protein S4	SLOPH 43	40S ribosomal protein S4
SLOPH 433	Histone H4	SLOPH 433	Histone H4
SLOPH 449	Hsp70 protein	SLOPH 449	Hsp70 protein
SLOPH 45	40S ribosomal protein S6	SLOPH 45	40S ribosomal protein S6
SLOPH 452	Hydroxyacylglutathione hydrolase	SLOPH 452	Hydroxyacylglutathione hydrolase
SLOPH 477	Hypothetical	SLOPH 477	Hypothetical
SLOPH 50	6-phosphogluconate dehydrogenase	SLOPH 50	6-phosphogluconate dehydrogenase
SLOPH 53	60S acidic ribosomal protein P0	SLOPH 53	60S acidic ribosomal protein
SLOPH 54	60S acidic ribosomal protein P2	SLOPH 54	60S acidic ribosomal protein P2
SLOPH 55	60S ribosomal protein L10a	SLOPH 55	60S ribosomal protein L10a
SLOPH 57	60S ribosomal protein L12	SLOPH 57	60S ribosomal protein L12
SLOPH 577	Hypothetical	SLOPH 577	Hypothetical
SLOPH 59	60S ribosomal protein L18	SLOPH 59	60S ribosomal protein L18
SLOPH 596	Hypothetical	SLOPH 596	Hypothetical
SLOPH 6	20S proteasome component	SLOPH 6	20S proteasome component
SLOPH 607	Hypothetical	SLOPH 607	Hypothetical
SLOPH 61	60S ribosomal protein L2	SLOPH 61	60S ribosomal protein L2
SLOPH 62	60s ribosomal protein L21	SLOPH 62	60s ribosomal protein L21
SLOPH 63	60S ribosomal protein L22	SLOPH 63	60S ribosomal protein L22
SLOPH 655	Hypothetical	SLOPH 655	Hypothetical
SLOPH 66	60S ribosomal protein L27	SLOPH 66	60S ribosomal protein L27
SLOPH 68	60S ribosomal protein L3	SLOPH 68	60S ribosomal protein L3
SLOPH 69	60S ribosomal protein L3	SLOPH 69	60S ribosomal protein L3
SLOPH 691	Hypothetical	SLOPH 691	Hypothetical
SLOPH 7	20S proteasome component beta 3	SLOPH 7	20S proteasome component beta 3
SLOPH 723	Hypothetical	SLOPH 723	Hypothetical
SLOPH 73	60S ribosomal protein L4	SLOPH 73	60S ribosomal protein L4

SLOPH 749	Hypothetical	SLOPH 749	Hypothetical
SLOPH 75	60S ribosomal protein L5	SLOPH 75	60S ribosomal protein L5
SLOPH 78	60S ribosomal protein L8	SLOPH 78	60S ribosomal protein L8
SLOPH 82	AAA ATPase	SLOPH 82	AAA ATPase
SLOPH 888	Hypothetical	SLOPH 888	Hypothetical
SLOPH 99	Acetyl-coenzyme A synthetase	SLOPH 99	Acetyl-coenzyme A synthetase
SLOPH 1386	NAD-dependent glycerol-3-phosphate dehydrogenase	SLOPH 1386	NAD-dependent glycerol-3-phosphate dehydrogenase
SLOPH 139	Arsenite-transporting ATPase	SLOPH 139	Arsenite-transporting ATPase
SLOPH 1395	Nonsense-mediated mRNA decay protein	SLOPH 1465	Polar tube protein 3
SLOPH 1429	Phosfructokinase	SLOPH 1490	Proteasome subunit beta
SLOPH 1437	Phosphoglycerate kinase	SLOPH 1539	RAN-specific GTPase activating protein
SLOPH 1438	Phosphoglyceromutase	SLOPH 1566	Ribosomal protein L15
SLOPH 1491	Proteasome subunit beta type 2	SLOPH 1628	Serine/threonine protein kinase
SLOPH 1517	Protein transport protein SEC23	SLOPH 1766	Hypothetical
SLOPH 1567	Ribosomal protein L34	SLOPH 18	26s proteasome regulatory subunit
SLOPH 1591	RNA recognition motif domain cotaining protein	SLOPH 1884	Spore wall protein 12
SLOPH 1608	SEC24-related protein	SLOPH 1949	Threonyl-tRNA synthetase
SLOPH 1639	Seryl-tRNA synthetase	SLOPH 2164	Glucosamine 6-phosphate N-acetyltransferase
SLOPH 1717	Hypothetical	SLOPH 2241	Hypothetical
SLOPH 1926	T-complex protein 1 subunit delta	SLOPH 2251	Peroxiredoxin
SLOPH 1928	T-complex protein 10	SLOPH 2268	NADH-cytochrome b5 reductase
SLOPH 1957	Transcription elongation complex subunit (Cdc68)	SLOPH 230	Hypothetical
SLOPH 1977	Transketolase	SLOPH 2344	Hypothetical
SLOPH 2050	Vacuolar ATP synthase subunit B	SLOPH 2373	Hypothetical
SLOPH 2178	Hypothetical	SLOPH 2431	20S proteasome alpha type-1 subunit
SLOPH 2217	WD-40 repeat-containing protein	SLOPH 2448	Hypothetical
SLOPH 2267	Hypothetical	SLOPH 34	40S Ribosomal protein S19
SLOPH 2334	T-complex protein 1 epsilon subunit	SLOPH 35	40S ribosomal protein S20
SLOPH 2418	T-complex protein 1 zeta subunit	SLOPH 392	Glucose-6-phosphate isomerase
SLOPH 2665	Arginine/alanine aminopeptidase	SLOPH 479	Hypothetical
SLOPH 2673	Proteasome B-type subunit	SLOPH 48	40S ribosomal protein S9
SLOPH 2686	Serine hydroxymethyltransferase	SLOPH 49	5'- 3'exoribonuclease
SLOPH 27	40S ribosomal protein S11	SLOPH 535	Hypothetical
SLOPH 29	40S ribosomal protein S13	SLOPH 548	Hypothetical
SLOPH 30	40S ribosomal protein S14	SLOPH 689	Hypothetical
SLOPH 31	40S ribosomal protein S15A	SLOPH 76	60S ribosomal protein L7
SLOPH 37	40s ribosomal protein s24	SLOPH 762	Hypothetical
SLOPH 393	Glutaminyl-tRNA synthetase	SLOPH 835	Hypothetical
SLOPH 394	Glutaminyl-tRNA synthetase	SLOPH 854	Hypothetical
SLOPH 415	Heat shock protein 101	SLOPH 887	Hypothetical
SLOPH 42	40S ribosomal protein S3A	SLOPH 962	Hypothetical
SLOPH 424	Histidyl-tRNA synthetase	SLOPH 1073	Hypothetical
SLOPH 44	40S ribosomal protein S5	SLOPH 1175	Hypothetical
SLOPH 473	Hypothetical	SLOPH 1347	Methionine aminopeptidase
SLOPH 580	Hypothetical		
SLOPH 79	60S ribosomal protein L9		

Appendix 3| *E. cuniculi* infection and the host phosphoproteome: a quantitative comparison of protein phosphorylation between non-infected IMCD-3 cells and *E. cuniculi* infected IMCD-3 cells.

Accession	Coverage	Description	Ratio
Q3V1V3	0.95	ESF1 homolog	2.78
Q8CI11	3.9	Guanine nucleotide-binding protein-like 3	2.13
Q9DAY9	28.02	Nucleophosmin	2.02
A2AWT6	1.1	Nucleolar transcription factor 1	2.13
D3YZJ1	4.71	Sequestosome-1	2.10
Q5U4C1	0.82	G-protein coupled receptor-associated sorting protein 1	2.07
G3UYA6	5.1	Yorkie homolog	1.77
P09405	0.85	Nucleolin	1.52
Q91VE6	3.15	MKI67 FHA domain-interacting nucleolar phosphoprotein	1.62
Q9JIK5	6.35	Nucleolar RNA helicase 2	1.59
Q03145	2.25	Ephrin type-A receptor 2	1.72
P53986	9.33	Monocarboxylate transporter 1	1.54
P14873	1.75	Microtubule-associated protein 1B	1.45
Q7TPV4	1.04	Myb-binding protein 1A	1.54
O08784	4.47	Treacle protein	1.41
Q9DCE5	2.09	p21-activated protein kinase-interacting protein 1	1.32
Q80XU3	17.52	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	1.30
P39689	5.03	Cyclin-dependent kinase inhibitor 1	1.33
F6XMS4	11.36	Protein Pus7 (Fragment)	1.33
D3Z2T9	5.5	Cyclin-dependent kinase 1 (Fragment)	1.37
F8WI35	4.44	Histone H3	1.26
P62806	31.07	Histone H4	1.33
E9Q616	2.33	Protein Ahnak	1.20
P10711	4.65	Transcription elongation factor A protein 1	1.16
Q99LL5	3.99	Periodic tryptophan protein 1 homolog	1.13
Q6DFW4	4.48	Nucleolar protein 58	1.17
Q8C0S5	1.9	CAP-Gly domain-containing linker protein 1 (Fragment)	1.10
Q0VEE6	1.66	Zinc finger protein 800	1.21
A2BE92	5.3	Protein SET (Fragment)	1.16
Q9CYA6	1.55	Zinc finger CCHC domain-containing protein 8	1.15
E9QPW8	3.96	Microtubule-associated protein	1.11
B1AU75	2.85	Nuclear autoantigenic sperm protein	1.10
P47856	1.87	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 1	1.05
F8WIX8	5.6	Histone H2A	1.12
F6S0D5	18.75	Protein 0610010K14Rik (Fragment)	1.07
Q91VY9	1.68	Zinc finger protein 622	1.14
P15919	0.58	V(D)J recombination-activating protein 1	1.15
P70698	2.37	CTP synthase 1	1.10
Q3UFB2	3.04	Box C/D snoRNA protein 1	1.05
E9Q8N5	1.09	CLIP-associating protein 2	0.98
Q7TPE5	4.25	Probable RNA polymerase II nuclear localization protein SLC7A6OS	1.06
E9Q4T0	1.18	Holliday junction recognition protein	1.03
P48725	0.93	Pericentrin	0.98
Q8R3F0	5.67	Larp2 protein	1.03
Q8R5F7	1.66	Interferon-induced helicase C domain-containing protein 1	1.02
F6R782	2.53	IQ domain-containing protein E (Fragment)	0.99
Q91XC0	2.01	LIM domain-containing protein ajuba	1.06
Q80UZ2	2.62	Protein SDA1 homolog	0.93
Q45VK5	1.99	Interleukin enhancer-binding factor 3	1.05
P60330	0.85	Separin	1.07

P10853	12.7	Histone H2B type 1-F/J/L	1.06
P20152	3.43	Vimentin	1.07
Q6ZWU9	13.1	40S ribosomal protein S27	0.98
F6WNT4	2.12	C-Jun-amino-terminal kinase-interacting protein 4 (Fragment)	0.91
E9QQ48	3.3	Disks large-associated protein 5 (Fragment)	0.93
E9PUA2	0.88	Terminal uridylyltransferase 7	1.05
Q3UHX0	1.31	Nucleolar protein 8	0.94
F7CBP1	0.81	Eukaryotic translation initiation factor 4 gamma 2	0.98
Q8VCE2	5.38	GPN-loop GTPase 1	0.97
Q80Y44	2.86	Probable ATP-dependent RNA helicase DDX10	0.95
Q5F2E7	2.46	Nuclear fragile X mental retardation-interacting protein 2	0.94
O88792	5	Junctional adhesion molecule A	1.07
Q9Z1Z0	3.23	General vesicular transport factor p115	0.96
H7BX26	1.01	Centrosomal protein of 170 kDa	0.91
E9PZT1	8.85	Nischarin (Fragment)	0.96
F2Z431	22.67	Importin subunit alpha-2	0.99
Q91VL0	4.14	BCL2/adenovirus E1B 19 kDa protein-interacting protein 2	0.99
Q6NS46	0.75	Protein RRP5 homolog	0.92
Q3UPF5	0.95	Zinc finger CCCH-type antiviral protein 1	0.97
Q9CZP3	2.4	Protein Zfp655	0.97
Q99J36	6.57	THUMP domain-containing protein 1	0.88
A2ATT5	5.59	ADP-sugar pyrophosphatase (Fragment)	0.94
J3KMM5	1.1	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0.92
Q9CZX5	4.22	PIN2/TERF1-interacting telomerase inhibitor 1	0.96
F6YBV1	3.53	CCR4-NOT transcription complex subunit 3 (Fragment)	0.96
G3UYV7	19.64	40S ribosomal protein S28 (Fragment)	0.93
H3BJQ2	18.84	SUMO-activating enzyme subunit 2 (Fragment)	0.87
Q7TMY8	0.32	E3 ubiquitin-protein ligase HUWE1	0.96
F6SPK0	5.73	Ubiquitin-conjugating enzyme E2 J1 (Fragment)	0.89
P51660	1.36	Peroxisomal multifunctional enzyme type 2	0.94
O35841	2.58	Apoptosis inhibitor 5	0.92
Q6P1B9	8.18	Bin1 protein	0.93
H3BKN0	5.54	tRNA (cytosine(34)-C(5))-methyltransferase	0.91
Q8K4L0	1.72	ATP-dependent RNA helicase DDX54	0.91
E9PX48	1.67	Dedicator of cytokinesis protein 7	0.91
Q3UW53	1.3	Protein Niban	0.91
P49452	2.1	Centromere protein C 1	0.96
Q6P9R1	2.19	ATP-dependent RNA helicase DDX51	0.94
F8WHU8	11.76	Protein 1810035L17Rik (Fragment)	0.93
Q9DCF9	4.32	Translocon-associated protein subunit gamma	0.91
H3BKV3	8.62	tRNA (guanine(26)-N(2))-dimethyltransferase (Fragment)	0.89
Q80W47	2.25	WD repeat domain phosphoinositide-interacting protein 2	0.91
Q8BK35	4.13	MCG2065, isoform CRAc	0.86
P54731	2.62	FAS-associated factor 1	0.88
E9PVC6	1.8	Eukaryotic translation initiation factor 4 gamma 1	0.89
Q8C0T5	1.18	Signal-induced proliferation-associated 1-like protein 1	0.93
Q8R0F5	3.07	RNA-binding motif protein, X-linked 2	0.88
E9Q0U7	0.98	Heat shock protein 105 kDa	0.90
Q923A2	3.78	Protein Spindly	0.87
Q3V2L1	4.4	Histone RNA hairpin-binding protein	0.95
D3Z736	3.38	L-lactate dehydrogenase	0.88
F6UND7	1.91	Tyrosine-protein kinase HCK	0.77
Q9DBJ3	1.95	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	0.78
Q80YS6	2.87	Actin filament-associated protein 1	0.86
Q3TXS7	1.78	26S proteasome non-ATPase regulatory subunit 1	0.89
E9Q5C9	3.7	Protein Nolc1	0.86
J3QP81	2.88	CLIP-associating protein 1	0.85

Q8BG09	1.97	Transmembrane protein 184B	0.85
E9Q1Q0	3.56	Protein Slc4a1ap	0.84
D3YXV8	1.1	Leukocyte receptor cluster member 8 homolog (Fragment)	0.89
O54692	1.8	Centromere/kinetochore protein zw10 homolog	0.84
D3Z0Q8	8.26	Golgin subfamily A member 3 (Fragment)	0.82
A2AI19	0.89	Glutamate [NMDA] receptor subunit zeta-1	0.88
D3Z4C3	3.13	Zinc finger and BTB domain-containing protein 7A	0.87
Q9ESV0	2.33	ATP-dependent RNA helicase DDX24	0.83
P59672	0.96	Ankyrin repeat and SAM domain-containing protein 1A	0.80
Q9DBY8	1.75	Nuclear valosin-containing protein-like	0.86
F7AVU1	2.31	Malcavernin	0.87
Q9ERU9	0.23	E3 SUMO-protein ligase RanBP2	0.85
O88477	2.08	Insulin-like growth factor 2 mRNA-binding protein 1	0.85
Q8BI72	3.37	CDKN2A-interacting protein	0.84
A2APD7	12.84	Nucleolar protein 56 (Fragment)	0.84
D3Z2R5	2.69	Protein Sepn1	0.83
Q5SVQ0	4.4	Histone acetyltransferase KAT7	0.85
P56183	3.44	Ribosomal RNA processing protein 1 homolog A	0.85
Q3TC46	1.56	Protein PAT1 homolog 1	0.79
Q8K1J5	4.12	Protein SDE2 homolog	0.82
G3UWX1	2.57	Replication factor C subunit 1	0.82
Q6A099	1.44	MKIAA0248 protein (Fragment)	0.82
F8VPM0	0.69	Bromodomain adjacent to zinc finger domain protein 2A	0.79
D3Z7B1	1.14	Protein 1700021K19Rik	0.79
Q99020	4.56	Heterogeneous nuclear ribonucleoprotein A/B	0.78
J3QNW0	1.27	DNA (cytosine-5)-methyltransferase 1	0.90
Q9D0B6	7.07	UPF0368 protein Cxorf26 homolog	0.80
Q8CH25	1.45	SAFB-like transcription modulator	0.81
D3Z7N2	10.08	Elongation factor 1-delta (Fragment)	0.78
G3UWZ0	1.16	Bromodomain adjacent to zinc finger domain protein 1A	0.76
F6UMA3	1.92	Putative methyltransferase NSUN7 (Fragment)	0.77
A2AVJ7	1.23	Ribosome binding protein 1	0.79
Q91W92	3.42	Cdc42 effector protein 1	0.78
Q6PFD9	2.27	Nucleoporin 98	0.78
A2AT02	2.65	NSFL1 (P97) cofactor (P47)	0.77
P42128	5.84	Forkhead box protein K1	0.86
Q8VDG3	1.28	Poly(A)-specific ribonuclease PARN	0.81
Q3TGG2	6.74	Ataxin-2-like protein	0.79
Q8BHJ9	1.54	Pre-mRNA-splicing factor SLU7	0.79
Q6ZWY8	15.91	Thymosin beta-10	0.82
Q8K2K6	1.96	Arf-GAP domain and FG repeat-containing protein 1	0.79
Q7TPD0	1.73	Integrator complex subunit 3	0.82
B1AQG3	5.61	Novel protein (5730455P16Rik) (Fragment)	0.85
Q9JI99	3.02	Sphingosine-1-phosphate phosphatase 1	0.79
E9PVP1	0.83	Protein Aim1	0.78
Q8R361	2.64	Rab11 family-interacting protein 5	0.79
Q7M739	1.87	Nuclear pore complex-associated intranuclear coiled-coil protein TPR	0.79
D3YUG5	1.84	MKL/myocardin-like protein 1 (Fragment)	0.78
D3YV48	6.43	RNA-binding protein 33	0.77
O55098	0.83	Serine/threonine-protein kinase 10	0.76
Q3U9G9	3.51	Lamin-B receptor	0.76
Q8BJF9	4.69	Charged multivesicular body protein 2b	0.77
Q6P5B0	1.54	RRP12-like protein	0.83
E9Q3V6	6.23	Septin-2	0.77
Q62376	2.9	U1 small nuclear ribonucleoprotein 70 kDa	0.82
Q61846	1.56	Maternal embryonic leucine zipper kinase	0.76
E9Q663	20.9	RNA (guanine-9-)-methyltransferase domain-containing protein 2	0.83

Q8BP27	19.12	Swi5-dependent recombination DNA repair protein 1 homolog	0.77
Q05D44	5.59	Eukaryotic translation initiation factor 5B	0.76
P62071	7.84	Ras-related protein R-Ras2	0.78
D3Z3A0	9.74	MCG126099, isoform CRAb	0.78
B1AZA5	3.42	Transmembrane protein 245	0.75
Q9WTV7	2.17	E3 ubiquitin-protein ligase RLIM	0.72
Q504P4	2.55	Heat shock cognate 71 kDa protein	0.76
Q5BL07	1.25	Peroxisome biogenesis factor 1	0.73
Q9ESX5	3.73	H/ACA ribonucleoprotein complex subunit 4	0.79
D6RJ29	27.66	Protein Stx4a	0.76
Q8VDJ3	1.89	Vigilin	0.74
H7BXC3	7.19	Triosephosphate isomerase	0.73
Q8CCJ3	2.02	E3 UFM1-protein ligase 1	0.77
Q3UZ39	9.05	Leucine-rich repeat flightless-interacting protein 1	0.86
P20029	1.37	78 kDa glucose-regulated protein	0.79
P07901	5.73	Heat shock protein HSP 90-alpha	0.71
E9Q6J5	2.11	Protein Bod1l	0.78
Q8VDZ4	2.1	Palmitoyltransferase ZDHHC5	0.73
Q8R1B4	1.54	Eukaryotic translation initiation factor 3 subunit C	0.73
E9PUF4	1.22	RNA-binding protein 26	0.80
Q8C1Y8	2.92	Vacuolar fusion protein CCZ1 homolog	0.75
F6SMY7	0.28	Probable E3 ubiquitin-protein ligase MYCBP2 (Fragment)	0.73
Q3TQV3	21.95	Adapter molecule crk	0.73
E9Q066	1.39	La-related protein 4	0.74
Q3UHX2	9.94	28 kDa heat- and acid-stable phosphoprotein	0.81
F6R6U0	1.29	AT-rich interactive domain-containing protein 4B (Fragment)	0.78
O70435	5.49	Proteasome subunit alpha type-3	0.74
Q9R0L6	1.43	Pericentriolar material 1 protein	0.73
A2AMI6	2.36	Eukaryotic translation initiation factor 4 gamma 3	0.73
P63094	2.28	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	0.77
Q8R3C0	1.71	Mini-chromosome maintenance complex-binding protein	0.79
P62908	5.35	40S ribosomal protein S3	0.79
Q8BR65	6.71	Sin3 histone deacetylase corepressor complex component SDS3	0.71
D3YTR5	3.07	Methyl-CpG-binding domain protein 3	0.86
E9PVH6	7.28	Zinc finger protein 106 (Fragment)	0.72
Q3UGS4	18.56	Protein FAM195B	0.79
H3BL90	38.1	Disks large-associated protein 4 (Fragment)	0.75
H9KUX8	0.39	Baculoviral IAP repeat-containing protein 6	0.70
Q8BFY7	13.85	Protein FAM64A	0.77
P21619	2.18	Lamin-B2	0.78
Q0P678	1.69	Zinc finger CCCH domain-containing protein 18	0.73
Q9R1C7	1.26	Pre-mRNA-processing factor 40 homolog A	0.76
B2RRE7	3.16	OTU domain-containing protein 4	0.75
Q9JLV1	3.64	BAG family molecular chaperone regulator 3	0.72
Q3B7Z2	6.13	Oxysterol-binding protein 1	0.74
Q6NV99	1.82	ortholog of H. sapiens family with sequence similarity 29, member A	0.69
Q9CQV4	1.72	Protein FAM134C	0.76
Q3UIR3	1.34	E3 ubiquitin-protein ligase DTX3L	0.72
F6WR45	2.48	Chromodomain-helicase-DNA-binding protein 4 (Fragment)	0.76
Q8CJ53	2.82	Cdc42-interacting protein 4	0.71
P97820	0.81	Mitogen-activated protein kinase kinase kinase kinase 4	0.74
D6RJ71	30.77	Centromere protein A, isoform CRAa	0.74
D3YZC7	1.02	Bromodomain-containing protein 8	0.74
Q8BYR2	1.86	Serine/threonine-protein kinase LATS1	0.73
Q8K310	2.6	Matrin-3	0.70
E9PUU4	0.86	Gem-associated protein 5	0.78
D6RJK7	1.79	TATA element modulatory factor	0.71

O35245	1.86	Polycystin-2	0.73
Q922U1	2.05	U4/U6 small nuclear ribonucleoprotein Prp3	0.69
G5E8P1	1.23	MCG7283	0.69
Q8BP78	6.67	Protein FRA10AC1 homolog	0.74
D3YWA4	21.62	PEST proteolytic signal-containing nuclear protein	0.68
O70310	2.42	Glycylpeptide N-tetradecanoyltransferase 1	0.74
Q8BIZ6	2.35	Smad nuclear-interacting protein 1	0.72
G3UX48	3.24	Protein PRRC2A	0.73
P48754	0.55	Breast cancer type 1 susceptibility protein homolog	0.75
P26645	11	Myristoylated alanine-rich C-kinase substrate	0.72
Q99MR6	1.03	Serrate RNA effector molecule homolog	0.72
A2AIV2	0.61	Protein virilizer homolog	0.69
E9Q6M7	1.27	Pumilio homolog 1	0.79
Q9WTU0	2.37	Lysine-specific demethylase PHF2	0.73
Q6P4T2	0.75	Activating signal cointegrator 1 complex subunit 3-like 1	0.66
Q3ULB0	0.91	Protein Rbm6	0.72
Q8BGT8	4.27	Phytanoyl-CoA hydroxylase-interacting protein-like	0.74
D3Z313	10.91	Chromobox protein homolog 3 (Fragment)	0.70
Q9DC47	3.02	Protein Zfp869	0.69
A2ALU4	0.81	Protein Shroom2	0.65
Q8R0A0	3.61	General transcription factor IIF subunit 2	0.71
Q80ZA1	3.41	Cellular tumor antigen p53	0.78
Q9JKB3	3.32	DNA-binding protein A	0.77
Q9JJY4	1.09	Probable ATP-dependent RNA helicase DDX20	0.71
P62259	11.37	14-3-3 protein epsilon	0.69
J3QNB1	4.29	La-related protein 1	0.72
Q9WV92	1.61	Band 4.1-like protein 3	0.71
Q9D281	2.46	Protein Noxp20	0.71
Q80X50	2.44	Ubiquitin-associated protein 2-like	0.72
Q8BRF7	2.97	Sec1 family domain-containing protein 1	0.71
Q99KX1	4.86	Myeloid leukemia factor 2	0.72
Q0VGB7	4.56	Serine/threonine-protein phosphatase 4 regulatory subunit 2	0.75
Q8JZQ9	7.47	Eukaryotic translation initiation factor 3 subunit B	0.73
F8WH95	4.17	Lysosomal-associated transmembrane protein 5 (Fragment)	0.71
F7D1J2	1.84	Receptor-interacting serine/threonine-protein kinase 1 (Fragment)	0.72
Q3TWW8	5.9	Protein Srsf6	0.69
P58871	6.69	182 kDa tankyrase-1-binding protein	0.70
G5E870	1.43	E3 ubiquitin-protein ligase TRIP12	0.71
P97855	3.66	Ras GTPase-activating protein-binding protein 1	0.70
P81122	0.91	Insulin receptor substrate 2	0.70
F7CPL2	4.73	Drebrin (Fragment)	0.71
F8WHC5	4.61	Arginine/serine-rich coiled-coil protein 2	0.72
Q62419	2.17	Endophilin-A2	0.71
D3YVH4	9.74	Splicing factor 1 (Fragment)	0.73
E9QNQ4	0.6	Rap guanine nucleotide exchange factor 2	0.72
Q8CJF7	1.2	Protein ELYS	0.70
J3QJV7	2.19	Rab GTPase-binding effector protein 1	0.68
Q69ZK6	1.19	Probable JmjC domain-containing histone demethylation protein 2C	0.70
P28667	14	MARCKS-related protein	0.71
F6R6F1	1.06	RalBP1-associated Eps domain-containing protein 1 (Fragment)	0.74
A2BI12	17.74	PC4 and SFRS1 interacting protein 1	0.70
Q5SSI6	3.08	U3 small nucleolar RNA-associated protein 18 homolog	0.68
D3YUW4	5.75	DNA ligase 1 (Fragment)	0.73
Q3UQN2	1.11	FCH domain only protein 2	0.66
E9QKZ7	0.71	Protein Zfml	0.69
Q6P5D3	0.58	Putative ATP-dependent RNA helicase DHX57	0.69
A2AS44	1.99	Plakophilin 4 (Fragment)	0.68

Q91WT4	4.62	DnaJ homolog subfamily C member 17	0.71
Q91V89	3.87	Protein Ppp2r5d	0.67
Q69ZX6	3.4	MORC family CW-type zinc finger protein 2A	0.71
Q9D8Y8	3.75	Inhibitor of growth protein 5	0.66
P97868	1.9	E3 ubiquitin-protein ligase RBBP6	0.71
Q8VBX6	0.49	Multiple PDZ domain protein	0.73
A2A864	0.5	Integrin beta	0.69
Q8R0L9	2.78	Transcriptional adapter 3	0.69
Q91Z96	2.2	BMP-2-inducible protein kinase	0.69
Q8C1D8	15.54	Protein IWS1 homolog	0.69
B0V2N8	5.68	Annexin (Fragment)	0.70
Q3U4S0	3.61	Pantothenate kinase 2 (Hallervorden-Spatz syndrome)	0.74
Q6QI06	0.82	Rapamycin-insensitive companion of mTOR	0.72
F6S5I0	2.36	Myosin phosphatase Rho-interacting protein (Fragment)	0.70
Q640N2	3.75	ADP-ribosylation factor-like protein 13B	0.67
P70255	6.61	Nuclear factor 1 C-type	0.69
F6QK97	5.7	DNA-directed RNA polymerase III subunit RPC3 (Fragment)	0.66
Q921F4	3.05	Heterogeneous nuclear ribonucleoprotein L-like	0.67
E9PVX6	1.38	Protein Mki67	0.67
Q8BJU0	5.4	Small glutamine-rich tetratricopeptide repeat-containing protein alpha	0.70
F6RU76	2.01	Protein 4933403O08Rik (Fragment)	0.73
D3Z6I8	4.86	Tropomyosin alpha-3 chain	0.67
E9Q309	0.52	Protein Cep350	0.71
B2RUP2	1.01	Protein unc-13 homolog D	0.61
F7ADT6	3.28	GTPase-activating protein and VPS9 domain-containing protein 1	0.65
A2BGG7	17.76	Nuclease-sensitive element-binding protein 1 (Fragment)	0.67
Q3UZD0	5.36	SWI/SNF complex subunit SMARCC1	0.69
Q9ERL0	4.94	Btk-PH-domain binding protein	0.68
E9QKG6	0.98	Ankyrin repeat domain-containing protein 17	0.66
A6X8Z3	3.24	Insulin-like growth factor 2 mRNA binding protein 2	0.66
P15920	2.45	V-type proton ATPase 116 kDa subunit a isoform 2	0.70
Q8BU14	6.03	Translocation protein SEC62	0.65
E9Q3K3	1.37	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 1	0.67
J3QPI4	17.57	Uncharacterised protein	0.65
A1A535	0.72	Ventricular zone-expressed PH domain-containing protein 1	0.67
Q9JK91	1.45	DNA mismatch repair protein Mlh1	0.66
Q6NZR5	1.21	Protein Skiv2l	0.64
Q9Z204	4.47	Heterogeneous nuclear ribonucleoproteins C1/C2	0.67
E9Q9Q7	14.75	Actin-binding LIM protein 1	0.64
Q8C079	2.03	Protein FAM40A	0.68
H9KV01	1.12	Protein SON	0.68
Q05CL8	5.79	La-related protein 7	0.68
Q5SXY1	1.03	Cytospin-B	0.63
F7B6X4	5.76	Ataxin-2 (Fragment)	0.63
A2AP93	3.67	Mitogen activated protein kinase kinase kinase 7	0.70
F7ANA1	1.68	PHD and RING finger domain-containing protein 1 (Fragment)	0.67
E9QKG5	1.69	Protein PRRC2C	0.69
Q3UHJ0	2.5	AP2-associated protein kinase 1	0.66
F8WH48	6.01	WD repeat-containing protein 26 (Fragment)	0.68
Q9CPN8	4.84	Insulin-like growth factor 2 mRNA-binding protein 3	0.68
Q62172	1.7	RalA-binding protein 1	0.68
Q9CXL3	13.85	Uncharacterised protein C7orf50 homolog	0.69
P08775	1.42	DNA-directed RNA polymerase II subunit RPB1	0.71
P58802	2.4	TBC1 domain family member 10A	0.63
P51612	3.12	DNA repair protein complementing XP-C cells homolog	0.60
Q80U49	1.78	Protein KIAA0284	0.64
J3QK23	2.84	Uncharacterised protein	0.69

A6PWC3	0.98	Nardilysin	0.67
A2A7T3	3.02	Arginine glutamic acid dipeptide (RE) repeats	0.66
Q8K298	1.34	Actin-binding protein anillin	0.67
F7A3Q3	9.8	Serine/threonine-protein kinase MARK2 (Fragment)	0.63
H3BKM2	11.76	Nuclear-interacting partner of ALK	0.64
A2A9P6	6.27	Cell division cycle 2-like 1	0.62
Q5PSV9	4.8	Mediator of DNA damage checkpoint protein 1	0.63
Q8R149	9.26	BUD13 homolog	0.66
G3X920	3.27	Armadillo repeat containing 8, isoform CRAb	0.64
G3X8R2	4.48	Steroid receptor RNA activator 1 (Fragment)	0.64
P47930	3.68	Fos-related antigen 2	0.68
F8WHR6	0.88	Probable ATP-dependent RNA helicase DDX46	0.63
Q9ESL4	2.62	Mitogen-activated protein kinase kinase kinase MLT	0.63
F6RTE1	10.76	Serine/arginine-rich-splicing factor 9 (Fragment)	0.69
Q8R4U7	1.97	Leucine zipper protein 1	0.65
Q8C850	2.15	Scavenger receptor class A member 3	0.68
Q80WQ6	0.97	Inactive rhomboid protein 2	0.69
Q62470	0.76	Integrin alpha-3	0.64
Q9JKP8	14.73	Chromatin accessibility complex protein 1	0.74
Q61136	3.77	Serine/threonine-protein kinase PRP4 homolog	0.65
P26369	2.32	Splicing factor U2AF 65 kDa subunit	0.64
A2AHC4	1.21	Calmodulin regulated spectrin-associated protein 1 (Fragment)	0.62
Q9CRA8	9.79	Exosome complex component RRP46	0.63
Q8K2D3	2.95	Enhancer of mRNA-decapping protein 3	0.68
F8VPM7	2.86	ELKS/Rab6-interacting/CAST family member 1	0.66
B1ATP7	2.43	ElaC homolog 2 (E. coli)	0.67
A2A6A1	1.59	G patch domain-containing protein 8	0.66
Q9DBR1	1.68	5'-3' exoribonuclease 2	0.67
Q99J77	1.95	N-acetylneuraminic acid synthase (Sialic acid synthase)	0.65
Q99LG1	10.44	Transmembrane protein 51	0.62
E9Q784	2.6	Protein Zc3h13	0.63
E9Q7D5	1.52	Protein Arhgef5	0.69
D3Z4K1	28.23	mRNA cap guanine-N7 methyltransferase (Fragment)	0.65
A2AQ43	1.99	Formin binding protein 1	0.66
Q6P542	6.33	ATP-binding cassette sub-family F member 1	0.63
Q99MR1	0.86	PERQ amino acid-rich with GYF domain-containing protein 1	0.64
Q9CWU4	10	UPF0690 protein C1orf52 homolog	0.64
Q61235	4.81	Beta-2-syntrophin	0.63
Q8R1Q8	3.63	Cytoplasmic dynein 1 light intermediate chain 1	0.63
Q9CW03	1.73	Structural maintenance of chromosomes protein 3	0.65
F8VQK7	4.32	Protein Gm16519	0.64
Q80V24	4.18	Transcription cofactor vestigial-like protein 4	0.62
Q9JMK2	2.4	Casein kinase I isoform epsilon	0.66
P36916	2.97	Guanine nucleotide-binding protein-like 1	0.69
F8WI22	12.76	Bcl-2-associated transcription factor 1	0.64
Q501J7	5.33	Phosphatase and actin regulator 4	0.65
Q505F5	2.93	Leucine-rich repeat-containing protein 47	0.58
Q7TNF9	2	Protein FAM117A	0.68
P11499	5.25	Heat shock protein HSP 90-beta	0.62
E9QNG1	1.27	Intersectin-2	0.66
Q8SWL7	4.81	Putative uncharacterised protein ECU010760	0.66
E9Q8D5	2.86	CCR4-NOT transcription complex subunit 2	0.63
Q61033	8.23	Lamina-associated polypeptide 2, isoforms alpha/zeta	0.66
Q99P88	0.72	Nuclear pore complex protein Nup155	0.64
Q8R010	7.81	Aminoacyl tRNA synthase complex-interacting multifunctional protein 2	0.64
E9Q5F4	7.92	Actin, cytoplasmic 1, N-terminally processed (Fragment)	0.67
Q8VBT0	9.35	Thioredoxin-related transmembrane protein 1	0.62

Q04692	2.25	SWI/SNF-related regulator of chromatin subfamily A containing DEAD/H box 1	0.62
D3Z7R0	6.54	[3-methyl-2-oxobutanoate dehydrogenase kinase, mitochondrial	0.68
E9PW51	8.4	Pumilio homolog 2 (Fragment)	0.64
Q99LI7	1.67	Cleavage stimulation factor subunit 3	0.61
E9PYJ2	3.51	Partitioning defective 3 homolog	0.63
A2AN08	0.33	E3 ubiquitin-protein ligase UBR4	0.59
E9PYX7	1.68	Afadin	0.62
A2A791	1.23	Zinc finger MYM-type protein 4	0.59
Q6P4S8	0.96	Integrator complex subunit 1	0.61
D6RE33	2.2	Enhancer of mRNA-decapping protein 4	0.64
Q3TB04	1.22	FERM domain-containing protein 4A	0.62
Q8BZR9	4.72	Uncharacterised protein C17orf85 homolog	0.64
Q9Z188	1.27	Dual specificity tyrosine-phosphorylation-regulated kinase 1B	0.61
O35126	2.13	Atrophin-1	0.63
E9Q026	1.8	Protein 5730419I09Rik	0.60
Q9DAS9	15.28	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	0.63
Q01853	1.61	Transitional endoplasmic reticulum ATPase	0.65
F8VQK5	2.03	SAM and SH3 domain-containing protein 1	0.61
J3KMH5	1.79	Melanoma inhibitory activity protein 3	0.63
Q91YE7	2.33	RNA-binding protein 5	0.55
B7ZD61	3.3	RNA binding motif protein 39 (Fragment)	0.62
D3YVN1	3.92	Actin-like protein 6A (Fragment)	0.63
D6RG95	2.02	Protein Ythdc1	0.61
O09000	0.5	Nuclear receptor coactivator 3	0.62
E9Q0A7	1.8	Oxidation resistance protein 1	0.59
E9QL08	2.99	Transforming acidic coiled-coil-containing protein 2	0.61
O70589	1.62	Peripheral plasma membrane protein CASK	0.63
Q8BZX4	1.82	Splicing regulatory glutamine/lysine-rich protein 1	0.63
O08582	3.74	GTP-binding protein 1	0.67
Q60596	2.54	DNA repair protein XRCC1	0.64
Q8K327	2.49	Chromosome alignment-maintaining phosphoprotein 1	0.64
P47911	4.15	60S ribosomal protein L6	0.64
F6TGJ2	12.84	Protein phosphatase 1 regulatory subunit 7 (Fragment)	0.60
Q60668	5.07	Heterogeneous nuclear ribonucleoprotein D0	0.64
F6YRW4	3.45	F-box and leucine-rich repeat protein 11	0.63
Q5XJE5	1.95	RNA polymerase-associated protein LEO1	0.61
P35564	5.92	Calnexin	0.63
A2ADS6	2.74	InaD-like (Drosophila)	0.65
Q8C129	1.27	Leucyl-cystinyl aminopeptidase	0.65
A2AR02	4.52	Peptidyl-prolyl cis-trans isomerase G	0.61
P50636	2.14	E3 ubiquitin-protein ligase RNF19A	0.60
F6XR65	6.7	Protein Srsf11 (Fragment)	0.63
O88532	1.3	Zinc finger RNA-binding protein	0.62
E9Q740	4.59	Protein Srp72	0.61
Q9D168	1.95	Integrator complex subunit 12	0.63
B2RXW8	1.13	Ppfia1 protein	0.60
E9PZM7	2.68	Protein Scaf11	0.62
Q80VW7	0.57	AT-hook-containing transcription factor	0.61
Q8CHY6	3.18	Transcriptional repressor p66 alpha	0.58
Q3TS02	4.48	ATP citrate lyase	0.63
Q8JZX4	3.46	Splicing factor 45	0.59
J3QPH6	9.36	Uncharacterised protein	0.61
Q3THK7	1.59	GMP synthase [glutamine-hydrolyzing]	0.62
P48678	11.28	Prelamin-A/C	0.63
Q9EQC8	4.17	Papillary renal cell carcinoma (Translocation-associated)	0.62
Q3UPH7	0.79	Rho guanine nucleotide exchange factor 40	0.58
O55022	10.26	Membrane-associated progesterone receptor component 1	0.64

A2BH40	0.74	AT-rich interactive domain-containing protein 1A	0.59
Q3U3C9	0.99	Genetic suppressor element 1	0.60
D3Z5B2	4.46	FERM domain-containing protein 8 (Fragment)	0.60
Q6PAM1	2.89	Alpha-taxilin	0.61
P35486	3.59	Pyruvate dehydrogenase E1 component subunit alpha, somatic form,	0.62
O88587	4.91	Catechol O-methyltransferase	0.61
Q8CHW4	1.39	Translation initiation factor eIF-2B subunit epsilon	0.59
Q5EG47	4.65	5'-AMP-activated protein kinase catalytic subunit alpha-1	0.58
Q80XQ2	3.31	TBC1 domain family member 5	0.62
Q99LJ0	3.13	CTTNBP2 N-terminal-like protein	0.62
G3UWD4	5.8	Periphilin 1, isoform CRAb	0.59
Q8BTI8	22.57	Serine/arginine repetitive matrix protein 2	0.61
Q99K01	2.41	Pyridoxal-dependent decarboxylase domain-containing protein 1	0.62
Q6A0A2	4.18	La-related protein 4B	0.60
Q60864	1.47	Stress-induced-phosphoprotein 1	0.61
F6RDB7	3.03	Transcriptional regulator ATRX (Fragment)	0.61
Q9WTQ5	3.56	A-kinase anchor protein 12	0.60
Q922Y1	7.74	UBX domain-containing protein 1	0.59
Q8K3X4	4	Interferon regulatory factor 2-binding protein-like	0.59
E0CX59	7.26	Protein Pkp2	0.59
Q9Z0U1	4.8	Tight junction protein ZO-2	0.57
F6Y765	3	Probable histone-lysine N-methyltransferase NSD2 (Fragment)	0.61
Q9CSN1	2.99	SNW domain-containing protein 1	0.63
Q6ZQ03	2.52	Formin-binding protein 4	0.59
Q3TN85	9.32	UV excision repair protein RAD23 homolog A (Fragment)	0.64
Q9D706	1.67	RNA polymerase II-associated protein 3	0.59
F7DCW4	8.41	Polypyrimidine tract-binding protein 1 (Fragment)	0.61
Q8CDK1	1.29	Protein Gm5591	0.61
O54724	3.83	Polymerase I and transcript release factor	0.60
O88874	1.81	Cyclin-K	0.60
F7BPW6	1.21	Protein Sec16a (Fragment)	0.57
A2AU90	13.85	Protein Trp53bp1 (Fragment)	0.59
Q9WV60	2.62	Glycogen synthase kinase-3 beta	0.58
Q9WVR4	2.97	Fragile X mental retardation syndrome-related protein 2	0.60
P54276	0.88	DNA mismatch repair protein Msh6	0.60
Q8VHE0	2.5	Translocation protein SEC63 homolog	0.58
P70302	1.61	Stromal interaction molecule 1	0.63
F6SZ47	2.67	Serine/threonine-protein kinase B-raf (Fragment)	0.60
Q8K3W3	4.3	Protein CASC3	0.60
Q3TIX9	2.13	U4/U6.U5 tri-snRNP-associated protein 2	0.56
Q923B1	4.73	Lariat debranching enzyme	0.60
Q99N57	2.62	RAF proto-oncogene serine/threonine-protein kinase	0.62
Q8BGU5	3.23	Cyclin-Y	0.60
Q9Z1T1	1.27	AP-3 complex subunit beta-1	0.56
Q3UMU9	7.32	Hepatoma-derived growth factor-related protein 2	0.59
P14069	7.87	Protein S100-A6	0.60
Q569Z6	8.2	Thyroid hormone receptor-associated protein 3	0.60
E9Q4Y9	3.64	WW domain-containing adapter protein with coiled-coil	0.59
Q61879	0.96	Myosin-10	0.54
P25206	5.17	DNA replication licensing factor MCM3	0.58
Q3U0V1	3.07	Far upstream element-binding protein 2	0.59
Q6PFR5	6.76	Transformer-2 protein homolog alpha	0.60
Q9CWK3	7.89	CD2 antigen cytoplasmic tail-binding protein 2	0.58
Q62261	1.14	Spectrin beta chain, non-erythrocytic 1	0.58
B1AYU7	4.43	Novel protein	0.59
B7ZN27	0.7	Cad protein	0.62
B1AY13	0.69	Ubiquitin carboxyl-terminal hydrolase 24	0.60

G5E8B1	0.73	Protein tyrosine phosphatase, non-receptor type 13	0.54
Q64012	19.87	RNA-binding protein Raly	0.61
Q8BJ71	1.47	Nuclear pore complex protein Nup93	0.60
Q673H1	5.85	Tumor suppressor candidate gene 1 protein homolog	0.58
Q5F2E8	1.9	Serine/threonine-protein kinase TAO1	0.64
Q9CRA5	5.37	Golgi phosphoprotein 3	0.58
Q3UMJ8	2.18	Transcriptional repressor GLI3R (Fragment)	0.57
O70279	1.88	Protein DGCR14	0.60
E9QPI5	1.05	Sister chromatid cohesion protein PDS5 homolog A	0.57
Q149C2	1.92	TRAF3-interacting protein 1	0.59
E0CXA0	17.33	Hepatoma-derived growth factor (Fragment)	0.56
P97792	3.01	Coxsackievirus and adenovirus receptor homolog	0.59
Q80U93	0.86	Nuclear pore complex protein Nup214	0.60
Q6DID3	1.58	Protein SCAF8	0.65
Q5FWK3	2.51	Rho GTPase-activating protein 1	0.55
Q4VA53	5.26	Sister chromatid cohesion protein PDS5 homolog B	0.55
Q8R0S1	1.94	Cyclic AMP-dependent transcription factor ATF-7	0.57
D3YYI8	10.79	Histone deacetylase	0.58
G5E8W1	3.33	TRIO and F-actin-binding protein	0.58
D3Z619	2.83	Pre-mRNA 3'-end-processing factor FIP1	0.57
G3UY87	5.65	PCTP-like protein (Fragment)	0.61
Q9DBR0	2.77	A-kinase anchor protein 8	0.58
Q9QWF0	1.1	Chromatin assembly factor 1 subunit A	0.58
Q9CT10	8.96	Ran-binding protein 3	0.57
Q148V8	7.69	Protein FAM83H	0.58
Q8BRN2	5.57	Euchromatic histone methyltransferase 1 (Fragment)	0.60
F6TFN2	3.62	Protein Lmo7 (Fragment)	0.55
Q9R0E1	3.1	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	0.59
P46467	5.86	Vacuolar protein sorting-associated protein 4B	0.55
D3YWB4	3.62	Rab5 GDP/GTP exchange factor (Fragment)	0.60
Q8CHI8	0.26	E1A-binding protein p400	0.64
D3YXK2	6.19	Scaffold attachment factor B1	0.56
Q3UPL0	1.95	Protein transport protein Sec31A	0.55
Q8BH76	3.25	DNA polymerase delta subunit 3	0.55
Q5SU71	1.44	E3 ubiquitin/ISG15 ligase TRIM25	0.55
A2A8V8	25.41	Serine/arginine repetitive matrix 1	0.57
Q4FK66	2.24	Pre-mRNA-splicing factor 38A	0.59
Q9EST5	3.68	Acidic leucine-rich nuclear phosphoprotein 32 family member B	0.52
Q3URN1	3.92	Protein Gm16401	0.57
Q6PDM2	4.44	Serine/arginine-rich splicing factor 1	0.59
A2A7S8	3.46	Uncharacterised protein KIAA1522	0.56
D6RII9	18.87	Protein PML	0.56
F7CPS1	4.37	3-hydroxyacyl-CoA dehydratase 4 (Fragment)	0.55
Q62093	3.17	Serine/arginine-rich splicing factor 2	0.56
Q6PBC0	9.3	CDP-diacylglycerol synthase (Phosphatidate cytidylyltransferase) 2	0.50
A2AI69	10.53	Protein kinase, cAMP dependent regulatory, type I, alpha	0.60
Q0VBL3	6.24	Protein Rbm15	0.58
E9QN52	4.82	Leucine-rich repeat flightless-interacting protein 2	0.55
E9PUK3	2	Formin-binding protein 1-like	0.55
Q5NCM6	2.6	Epsin 2	0.61
Q8C0J6	2.34	Ankyrin repeat domain-containing protein SOWAHC	0.56
Q6NXJ0	0.84	Protein WWC2	0.55
A2ARV4	0.47	Low-density lipoprotein receptor-related protein 2	0.56
P49717	0.93	DNA replication licensing factor MCM4	0.53
P20444	1.19	Protein kinase C alpha type	0.59
F6RJ39	9.51	Apoptotic chromatin condensation inducer in the nucleus (Fragment)	0.58
P05784	4.73	Keratin, type I cytoskeletal 18	0.59

Q9Z2R6	11.25	Protein unc-119 homolog A	0.58
E9PYT0	2.2	Rho GTPase-activating protein 5	0.53
B9EHJ3	4.19	Tight junction protein ZO-1	0.56
Q9D8Z1	3.37	Activating signal cointegrator 1 complex subunit 1	0.55
E9Q6R7	0.29	Protein Utrn	0.56
Q5SV54	1.18	chromosome region, candidate 19 (ALS2CR19)	0.58
B2RUJ8	3.57	Rho GTPase activating protein 12	0.56
D3Z6P4	4.18	Intersectin-1 (Fragment)	0.53
Q8BGC0	2.25	HIV Tat-specific factor 1 homolog	0.54
E9PUD0	21.5	Zinc finger Ran-binding domain-containing protein 2	0.53
Q62018	1.45	RNA polymerase-associated protein CTR9 homolog	0.57
B0R030	10.49	Synaptosomal-associated protein (Fragment)	0.54
E9PZG4	1.26	Histone deacetylase 7	0.53
F8VQE9	1.32	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3	0.60
J3QK72	3.44	Uncharacterised protein	0.56
Q99PG2	2.84	Opioid growth factor receptor	0.50
B2RUR8	2.86	OTU domain containing 7B	0.51
F8WJ93	1.62	Echinoderm microtubule-associated protein-like 4	0.55
Q80SY5	2.4	Pre-mRNA-splicing factor 38B	0.56
Q3UN10	0.86	Wolframin	0.54
Q3URZ6	1.95	AHNAK nucleoprotein 2	0.55
Q2KN98	1.16	Cytospin-A	0.59
Q9Z277	1.69	Tyrosine-protein kinase BAZ1B	0.55
Q3UJB0	3.53	Protein Sf3b2	0.54
Q3TQI7	5.54	Uncharacterised protein C9orf78 homolog	0.57
E9Q1P8	5.26	Interferon regulatory factor 2-binding protein 2	0.52
Q3TSG4	4.56	Probable alpha-ketoglutarate-dependent dioxygenase ABH5	0.57
D3Z2H7	1.64	Catenin delta-1	0.52
D3YZP9	2.56	Coiled-coil domain-containing protein 6	0.52
F7BTP0	8.48	Non-homologous end-joining factor 1 (Fragment)	0.57
Q7TMI3	1.62	E3 ubiquitin-protein ligase UHRF2	0.54
Q8BL97	6.74	Serine/arginine-rich splicing factor 7	0.63
E9QKV6	0.82	Unconventional myosin-IXb	0.53
A2AG58	1.46	Uncharacterised protein CXorf23 homolog	0.54
Q8BH50	8.16	Uncharacterised protein C18orf25 homolog	0.55
Q80X82	1.87	Symplekin	0.53
Q80VC9	2.32	Calmodulin-regulated spectrin-associated protein 3	0.54
O54774	1.42	AP-3 complex subunit delta-1	0.52
E9Q6Q2	1.31	Serine/threonine-protein kinase WNK2	0.51
B1AXN9	1.54	Ribosomal protein S6 kinase alpha-3	0.58
O35691	4.69	Pinin	0.54
Q8C0D7	4.12	Inhibitor of growth protein 4	0.55
F8VQL9	2.59	Nuclear receptor corepressor 2	0.51
Q80U87	1.48	Ubiquitin carboxyl-terminal hydrolase 8	0.54
Q64707	3.74	U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit-related	0.52
Q3UIL6	0.98	Pleckstrin homology domain-containing family A member 7	0.53
G3X8S2	2.22	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB	0.55
Q3TLD5	4.9	Unconventional prefoldin RPB5 interactor	0.51
Q8K2F8	1.95	Protein LSM14 homolog A	0.54
B1AZ15	6.85	Cobl-like 1	0.53
F6ZIE1	12.31	Calcium/calmodulin-dependent protein kinase type II subunit delta (Fragment)	0.53
Q99N13	1.36	Histone deacetylase 9	0.52
O35954	0.97	Membrane-associated phosphatidylinositol transfer protein 1	0.55
G5E8C3	3.93	G protein-coupled receptor, family C, group 5, member A	0.52
Q8C170	0.35	Unconventional myosin-IXa	0.55
P97310	4.31	DNA replication licensing factor MCM2	0.54
O55111	0.8	Desmoglein-2	0.51

A8Y5K6	2.1	Eukaryotic translation initiation factor 4E nuclear import factor 1	0.53
D3YZ06	12.24	Heat shock protein beta-1	0.53
Q66JV4	0.96	RNA-binding protein 12B-B	0.55
F7AYW2	0.68	Protein PRRC2B (Fragment)	0.55
Q6PHZ5	1.47	Putative RNA-binding protein 15B	0.57
Q91WB4	3.61	Pleckstrin homology domain-containing family F member 2	0.50
E9QAF9	1.24	Protein TANC1	0.55
G3UWY3	2.18	Protein cordon-bleu	0.56
Q8BJS4	1.37	SUN domain-containing protein 2	0.53
F6PYU5	0.83	Protein Fam38a (Fragment)	0.53
Q80TP3	0.54	E3 ubiquitin-protein ligase UBR5	0.54
B1AUA8	7.61	Histone deacetylase 6 (Fragment)	0.56
D3YUW7	2.53	Cingulin	0.49
D3Z6R9	32.65	DNA-(apurinic or apyrimidinic site) lyase (Fragment)	0.52
Q3UL36	5.17	Arginine and glutamate-rich protein 1	0.52
Q8SVY7	5.92	Putative uncharacterised protein ECU040030	0.52
B7ZP47	1.01	Wapal protein	0.55
P43346	5	Deoxycytidine kinase	0.54
F6RUK9	2.03	Pleckstrin homology-like domain family B member 1 (Fragment)	0.49
D3Z5T2	8.82	Armadillo repeat-containing protein 10	0.53
Q99LD4	4.13	COP9 signalosome complex subunit 1	0.58
G3X9H8	2.57	E3 ubiquitin-protein ligase	0.55
E9Q855	2.54	Secretory carrier-associated membrane protein 3	0.57
P26231	4.18	Catenin alpha-1	0.54
F6XCT0	0.65	Microtubule-actin cross-linking factor 1 (Fragment)	0.52
Q62394	5.68	Zinc finger protein 185	0.51
A2AM69	12.79	Heterochromatin protein 1, binding protein 3 (Fragment)	0.53
Q8C052	1.64	Microtubule-associated protein 1S	0.51
Q5U4C3	0.88	Splicing factor, arginine/serine-rich 19	0.51
Q9ERG0	7.04	LIM domain and actin-binding protein 1	0.52
G5E8L6	2.84	Killer cell lectin-like receptor subfamily G member 2	0.50
H3BJL7	7.02	Homeobox protein cut-like 1 (Fragment)	0.51
Q6PGG2	1.44	GEM-interacting protein	0.50
Q8BGV8	2.16	Mitochondrial dynamic protein MID51	0.51
A2AFI3	2.33	RNA binding motif protein, X chromosome	0.50
P97433	0.53	Rho guanine nucleotide exchange factor 28	0.48
E9PX94	2.01	Coiled-coil and C2 domain-containing protein 1A	0.47
Q8BGD9	5.73	Eukaryotic translation initiation factor 4B	0.51
A6PWE1	2.56	Coilin (Fragment)	0.48
E9QMZ5	0.84	Plectin	0.48
Q5U3K5	2.48	Rab-like protein 6	0.53
Q8BQ30	2.86	Phostensin	0.51
H3BJU7	5.54	Rho guanine nucleotide exchange factor 2	0.49
E9PZD2	2.16	Protein Micall2	0.48
Q9QXV1	3.59	Chromobox protein homolog 8	0.52
E9Q3G8	1.09	Protein Nup153	0.47
A2ACM0	0.67	Novel protein containing six WD40 domains at C-terminus	0.48
Q9DBT5	2.26	AMP deaminase 2	0.44
O35295	5.86	Transcriptional activator protein Pur-beta	0.50
F6W2Q5	1.05	Epidermal growth factor receptor substrate 15	0.47
A2A485	2.19	Protein Zmynd8	0.50
Q9WVA4	5.53	Transgelin-2	0.41
Q99KG3	1.18	RNA-binding protein 10	0.49
P70441	9.86	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	0.50
P83741	0.46	Serine/threonine-protein kinase WNK1	0.50
Q99M08	12.31	Uncharacterised protein C4orf3 homolog	0.50
F7AI27	3.65	Leucine-rich repeat-containing protein 16A (Fragment)	0.48

P46935	1.47	E3 ubiquitin-protein ligase NEDD4	0.46
D3Z0V2	1.59	Rho guanine nucleotide exchange factor 7	0.47
H7BX82	1.79	TBC1 domain family member 4 (Fragment)	0.47
Q3UID0	1.24	SWI/SNF complex subunit SMARCC2	0.48
Q8BVL3	5.53	Sorting nexin-17	0.48
Q3TDQ1	1.82	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	0.49
P26350	12.61	Prothymosin alpha	0.49
E9PXE6	0.68	Formin-like protein 2	0.49
D6RI80	8.57	Phosphatidylinositol 4-kinase beta (Fragment)	0.46
Q3U9H3	5.56	Bcl-associated death promoter, isoform CRAa	0.47
E9Q8K9	3.62	Sorbin and SH3 domain-containing protein 2	0.50
D3YU01	6.59	Pleckstrin homology domain-containing family A member 1	0.49
Q56A10	0.73	Zinc finger protein 608	0.47
Q6IQX8	1.24	Protein Zfp219	0.48
P11679	5.51	Keratin, type II cytoskeletal 8	0.47
Q80WJ7	3.11	Protein LYRIC	0.46
E9QKY7	1.63	Transcription factor 20	0.48
E9QMJ1	1.44	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1	0.45
Q4VAC9	5.44	Pleckstrin homology domain-containing family G member 3	0.46
P49586	8.17	Choline-phosphate cytidylyltransferase A	0.47
Q64727	2.44	Vinculin	0.48
Q6P5G6	3.85	UBX domain-containing protein 7	0.48
F7C521	9.62	Protein Rod1	0.46
Q8BTW9	1.52	Serine/threonine-protein kinase PAK 4	0.47
Q62048	10	Astrocytic phosphoprotein PEA-15	0.47
D3Z564	2.86	TGF-beta-activated kinase 1 and MAP3K7-binding protein 2 (Fragment)	0.46
A2AI08	1.2	Taperin	0.42
Q8BGS1	2.46	Band 4.1-like protein 5	0.46
Q3UVG3	1.08	Protein FAM91A1	0.45
Q3UYI5	1.97	Ral guanine nucleotide dissociation stimulator-like 3	0.45
Q9ER72	1.81	Cysteine--tRNA ligase, cytoplasmic	0.44
H3BJV7	6.19	Cell division cycle-associated 7-like protein (Fragment)	0.45
Q8BLH7	5.82	HIRA-interacting protein 3	0.46
A2AKD7	2	Alpha-1-syntrophin	0.53
F8VPR8	0.83	Putative peptidyl-prolyl cis-trans isomerase	0.47
Q8VDP4	1.19	DBIRD complex subunit KIAA1967 homolog	0.47
Q9D219	0.77	B-cell CLL/lymphoma 9 protein	0.45
Q01279	1.57	Epidermal growth factor receptor	0.49
P50543	16.33	Protein S100-A11	0.50
P38532	2.48	Heat shock factor protein 1	0.46
Q9R0U0	5.73	Serine/arginine-rich splicing factor 10	0.58
Q8VDD5	1.07	Myosin-9	0.44
Q8C6B2	5.67	Rhotekin	0.45
Q91XA2	7.38	Golgi membrane protein 1	0.39
Q9JI10	4.23	Serine/threonine-protein kinase 3	0.44
B9EKS2	0.68	Jumonji domain containing 1B	0.52
Q91YN9	3.33	BAG family molecular chaperone regulator 2	0.45
D3YX85	1.68	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2	0.46
O54781	1.76	SRSF protein kinase 2	0.44
Q9DBX5	2.43	Phospholipase A2	0.41
Q99K30	1.23	Epidermal growth factor receptor kinase substrate 8-like protein 2	0.42
F8WI30	3.37	Sorting nexin-7	0.43
A2AIR0	9.77	E3 ubiquitin-protein ligase BRE1A (Fragment)	0.46
P97825	19.48	Hematological and neurological expressed 1 protein	0.46
E9QPT4	2.66	Polyhomeotic-like protein 3	0.46
B2FDH7	48.78	Ubiquitin-conjugating enzyme E2 E3 (Fragment)	0.48
F7CGP0	0.64	Ral GTPase-activating protein subunit alpha-2 (Fragment)	0.44

E9Q5B0	2.39	Calmodulin-regulated spectrin-associated protein 3	0.45
Q8C0E3	3.9	Tripartite motif-containing protein 47	0.45
Q9ET54	0.78	Palladin	0.42
F6TH70	7.43	Dynamin-1 (Fragment)	0.44
E9QP01	2.06	Protein capicua homolog	0.44
Q8C2T6	1.94	DNA polymerase alpha subunit B	0.42
P19426	12.53	Negative elongation factor E	0.41
Q8C142	2.6	Low density lipoprotein receptor adapter protein 1	0.42
Q8N7N5	2.37	DDB1- and CUL4-associated factor 8	0.42
G3X971	3.34	Ankyrin 3, epithelial, isoform CRAi	0.43
E9QP46	0.2	Nesprin-2	0.41
G5E8M1	2.1	Protein tyrosine phosphatase, non-receptor type 14, isoform CRAa	0.41
Q6PCM2	1.13	Integrator complex subunit 6	0.44
Q9WTX2	3.19	Interferon-inducible double stranded RNA-dependent protein kinase activator A	0.42
Q91Z58	1.82	Uncharacterised protein C6orf132 homolog	0.42
Q8BH07	5.31	ADP-ribosylation factor-like protein 6-interacting protein 6	0.44
Q921G9	4.54	Tight junction protein 3	0.44
D3YUC9	8.77	Methionine-R-sulfoxide reductase B3, mitochondrial	0.43
Q8K4Z3	4.26	NAD(P)H-hydrate epimerase	0.43
Q6PAJ1	1.18	Breakpoint cluster region protein	0.40
P57016	5.68	Ladinin-1	0.44
Q9D952	0.39	Envoplakin	0.42
F6RIN9	4.23	Protein scribble homolog (Fragment)	0.44
E9Q3Y4	1.12	Lipopolysaccharide-responsive and beige-like anchor protein	0.42
P59017	2.3	Bcl-2-like protein 13	0.42
Q5SSZ5	0.9	Tensin-3	0.41
Q9QYB5	2.41	Gamma-adducin	0.42
E9QQ30	1.08	Eukaryotic translation initiation factor 2-alpha kinase 3	0.42
D3Z1Z8	20.59	Stathmin (Fragment)	0.40
E0CXN3	10	Protein Ddx23 (Fragment)	0.43
D3Z607	10.47	Non-histone chromosomal protein HMG-14	0.41
Q9CQT2	3.77	RNA-binding protein 7	0.43
F6T941	2.04	Transcription factor E2-alpha (Fragment)	0.42
Q9JLQ0	3.92	CD2-associated protein	0.40
H3BKL6	3.14	Cutaneous T-cell lymphoma-associated antigen 5 homolog	0.38
E9Q449	0.5	Protein Dennd4c	0.42
P97863	1.58	Nuclear factor 1 B-type	0.40
A2RSJ4	0.62	UHRF1-binding protein 1-like	0.41
Q6P9Q4	1.67	FH1/FH2 domain-containing protein 1	0.41
D6RHM9	15.15	Alpha-parvin	0.40
Q8R3Y8	2.23	Interferon regulatory factor 2-binding protein 1	0.40
D3Z197	1.65	Histone-lysine N-methyltransferase NSD3	0.38
Q9WTK7	2.52	Serine/threonine-protein kinase STK11	0.40
Q3UDW8	1.68	Heparan-alpha-glucosaminide N-acetyltransferase	0.38
Q5PPR2	1.14	Exocyst complex component 1	0.36
G3UXH4	2.99	Serine/threonine-protein kinase N2	0.41
A8Y5C3	23.97	PDZ and LIM domain 4 (Fragment)	0.40
Q61686	5.76	Chromobox protein homolog 5	0.37
Q8K3A9	1.5	7SK snRNA methylphosphate capping enzyme	0.41
Q6PR54	0.54	Telomere-associated protein RIF1	0.42
J3QNE8	1.01	Putative sodium-coupled neutral amino acid transporter 10	0.41
E9Q7G0	1.34	Protein Numa1	0.40
Q8C4S8	1.2	DENN domain-containing protein 2A	0.41
D3Z1N4	0.73	Protein polybromo-1	0.41
A2ALU3	2.16	DENN domain-containing protein 1A (Fragment)	0.39
O70439	5.75	Syntaxin-7	0.40
F8WH20	1.59	Oxysterol-binding protein	0.39

D3YUY7	11.57	Tripartite motif-containing protein 3 (Fragment)	0.39
Q62318	5.76	Transcription intermediary factor 1-beta	0.38
B1ASH8	5.17	Rho guanine nucleotide exchange factor (GEF) 16 (Fragment)	0.38
Q9R078	2.96	5'-AMP-activated protein kinase subunit beta-1	0.40
A2AIX0	1.16	72 kDa inositol polyphosphate 5-phosphatase	0.34
F8VQ28	7.11	Paxillin	0.39
G5E898	0.63	Periplakin	0.39
Q61165	2.32	Sodium/hydrogen exchanger 1	0.36
Q99LH9	2.04	SH3 domain-binding protein 5-like	0.38
Q9ER00	3.65	Syntaxin-12	0.41
Q8VC03	1.56	Echinoderm microtubule-associated protein-like 3	0.39
Q00PI9	5.1	Heterogeneous nuclear ribonucleoprotein U-like protein 2	0.38
Q9CQN9	4.33	Sentrin-specific protease 7	0.38
Q62433	4.16	Protein NDRG1	0.39
Q69ZR9	0.87	Protein FAM208A	0.42
Q80XC3	1.1	USP6 N-terminal-like protein	0.38
Q61823	4.15	Programmed cell death protein 4	0.34
P62754	4.42	40S ribosomal protein S6	0.39
Q8C0C3	2.07	Protein Ttc15	0.35
Q8BVQ5	3.11	Protein phosphatase methylesterase 1	0.32
F7C134	0.65	Nuclear receptor corepressor 1 (Fragment)	0.37
A2ALS7	2.22	Protein Rap1gap (Fragment)	0.40
G3X8S6	1.26	Bicaudal C homolog 1 (Drosophila), isoform CRAd	0.36
P62996	7.64	Transformer-2 protein homolog beta	0.46
Q99J21	3.1	Mucolipin-1	0.36
H3BK79	4.8	Nucleobindin-1 (Fragment)	0.36
E9Q0S6	3.18	Protein Tns1	0.34
Q91VW5	0.67	Golgin subfamily A member 4	0.30
F7AL76	0.39	Chromodomain-helicase-DNA-binding protein 8 (Fragment)	0.34
P27889	3.94	Hepatocyte nuclear factor 1-beta	0.37
E9Q8P5	3.35	PDZ and LIM domain protein 5	0.40
G3UZU0	39.02	DNA (cytosine-5)-methyltransferase 3A (Fragment)	0.33
B7ZC23	4.11	Nuclear receptor coactivator 5	0.34
A2AB59	2.07	Rho GTPase-activating protein 27	0.31
E9Q3E2	2.73	Synaptopodin	0.31
Q6PDN3	1.39	Myosin light chain kinase, smooth muscle	0.31
P05627	2.4	Transcription factor AP-1	0.29
Q7TT50	0.76	Serine/threonine-protein kinase MRCK beta	0.33
Q8C5R2	2.97	Uncharacterised protein C10orf47 homolog	0.28
P46062	1.54	Signal-induced proliferation-associated protein 1	0.29
Q8R0W0	1.83	Epiplakin	0.29
Q99JP6	3.37	Homer protein homolog 3	0.30
Q8VBT6	0.74	Apolipoprotein B receptor	0.28
P98203	2.18	Armadillo repeat protein deleted in velo-cardio-facial syndrome homolog	0.26
Q9QYC0	6.12	Alpha-adducin	0.26
Q80X35	3	Microtubule-associated protein	0.27
F7AR27	5.19	Protein Fam98c (Fragment)	0.27
E9Q4K7	0.71	Protein Kif13b	0.28
G3V011	1.4	Unconventional myosin-If	0.26
Q5SXA9	0.91	Protein KIBRA	0.23
E9QAX1	3.34	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 2	0.24
D3Z5K9	0.95	SH3 and multiple ankyrin repeat domains protein 2	0.22
D3Z710	2.38	Kinesin light chain 3	0.20
P97326	1.77	Cadherin-6	0.18

Appendix 4| Campbell, S. E., T. A. Williams, A. Yousuf, D. M. Soanes, K. H. Paszkiewicz, B. A. P. Williams (2013). The genome of *Spraguea lophii* and the basis of host-microsporidian interactions. PLoS Genet 9(8)

OPEN ACCESS Freely available online

PLOS | GENETICS

The Genome of *Spraguea lophii* and the Basis of Host-Microsporidian Interactions

Scott E. Campbell[1], Tom A. Williams[2], Asim Yousuf[1], Darren M. Soanes[1], Konrad H. Paszkiewicz[1], Bryony A. P. Williams[1]*

1 Biosciences, College of Life and Environmental Sciences, University of Exeter, Devon, United Kingdom, **2** Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne, Tyne and Wear, United Kingdom

Abstract

Microsporidia are obligate intracellular parasites with the smallest known eukaryotic genomes. Although they are increasingly recognized as economically and medically important parasites, the molecular basis of microsporidian pathogenicity is almost completely unknown and no genetic manipulation system is currently available. The fish-infecting microsporidian *Spraguea lophii* shows one of the most striking host cell manipulations known for these parasites, converting host nervous tissue into swollen spore factories known as xenomas. In order to investigate the basis of these interactions between microsporidian and host, we sequenced and analyzed the *S. lophii* genome. Although, like other microsporidia, *S. lophii* has lost many of the protein families typical of model eukaryotes, we identified a number of gene family expansions including a family of leucine-rich repeat proteins that may represent pathogenicity factors. Building on our comparative genomic analyses, we exploited the large numbers of spores that can be obtained from xenomas to identify potential effector proteins experimentally. We used complex-mix proteomics to identify proteins released by the parasite upon germination, resulting in the first experimental isolation of putative secreted effector proteins in a microsporidian. Many of these proteins are not related to characterized pathogenicity factors or indeed any other sequences from outside the Microsporidia. However, two of the secreted proteins are members of a family of RICIN B-lectin-like proteins broadly conserved across the phylum. These proteins form syntenic clusters arising from tandem duplications in several microsporidian genomes and may represent a novel family of conserved effector proteins. These computational and experimental analyses establish *S. lophii* as an attractive model system for understanding the evolution of host-parasite interactions in microsporidia and suggest an important role for lineage-specific innovations and fast evolving proteins in the evolution of the parasitic microsporidian lifecycle.

Citation: Campbell SE, Williams TA, Yousuf A, Soanes DM, Paszkiewicz KH, et al. (2013) The Genome of *Spraguea lophii* and the Basis of Host-Microsporidian Interactions. PLoS Genet 9(8): e1003676. doi:10.1371/journal.pgen.1003676

Editor: Joseph Heitman, Duke University Medical Center, United States of America

Received March 1, 2013; **Accepted** June 12, 2013; **Published** August 22, 2013

Funding: This work was supported by a BBSRC studentship to SEC (http://www.bbsrc.ac.uk), a Marie Curie postdoctoral fellowship to TAW (http://cordis.europa.eu/fp7/home_en.html) and a Royal Society University Research Fellowship to BAPW (http://royalsociety.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: b.a.p.williams@exeter.ac.uk

Introduction

Microsporidia are a diverse phylum of obligate intracellular parasites related to fungi. Over 1300 species have been described in approximately 160 genera and, as is the case for other microbial eukaryotes, a vast undescribed diversity is thought to exist in the environment [1]. Microsporidia are important pathogens of a broad range of animal groups: they can infect immunocompromized humans, such as those with HIV/AIDS, and are major pathogens of fish and invertebrates, representing a significant threat to sericulture [2] and fisheries [3]. Their unusual lifecycle has also attracted attention, particularly the unique mechanism by which microsporidia gain entrance to host cells. Outside the host cell, microsporidia exist as a resistant spore containing a coiled polar tube. Upon coming into contact with a host cell, or appropriate stimulus, the spore rapidly everts this tube, penetrating the host cell membrane and delivering the spore contents to the host cytoplasm, where proliferation and the next round of spore production occurs [4].

In addition to their importance as parasites of animals, Microsporidia have attracted much attention as eukaryotic model systems for reductive genome evolution. The 2.9 Mb genome of the microsporidian *Encephalitozoon cuniculi* was one of the first eukaryotic genomes to be sequenced [5]. Analysis of the *E. cuniculi* genome revealed a highly reduced and streamlined genome which had lost or simplified many biochemical pathways, had truncated genes, shortened intergenic spaces and had almost entirely lost introns and repetitive DNA [5]; its close relative *Encephalitozoon intestinalis* has an even smaller genome, at 2.3 Mb [6]. Interestingly, while microsporidian genomes are consistently smaller than those of their opisthokont relatives, there is a ten-fold difference in genome size within the phylum, with some genomes as large as 24 Mb [7]. To date, it has proven difficult to relate this genomic variation to differences in parasite biology or host preference. This is because all microsporidia sequenced so far share a broadly conserved core proteome, with differences in genome size due largely to changes in gene density, transposon content, and expansions of uncharacterized, lineage-specific or fast evolving protein families [8–10]. As the main differences in coding capacity among sequenced microsporidian lineages, it seems reasonable to hypothesize that these lineage-specific or fast evolving proteins

Author Summary

Microsporidia are unusual intracellular parasites that infect a broad range of animal cells. In comparison to their fungal relatives, microsporidian genomes have shrunk during evolution, encoding as few as 2000 proteins. This minimal molecular repertoire makes them a reduced model system for understanding host-parasite interactions. A number of microsporidian genomes have now been sequenced, but the lack of a system for genetic manipulation makes it difficult to translate these data into a better understanding of microsporidian biology. Here we present a deep sequencing project of *Spraguea lophii*, a fish-infecting microsporidian that is abundantly available from environmental samples. We use our sequence data combined with germination protocols and complex-mix proteomics to identify proteins released by the cell at the earliest stage of germination, representing potential pathogenicity factors. We profile the RNA expression pattern of germinating cells and identify a set of highly transcribed hypothetical genes. Our study provides new insight into the importance of uncharacterized, lineage-specific and/or fast evolving proteins in microsporidia and provides new leads for the investigation of virulence factors in these enigmatic parasites.

play a role in mediating host-parasite interactions. However, because they lack any detectable similarity to genes from model eukaryotes, the functions of these proteins are difficult to predict using bioinformatics. Combined with the current lack of a system for genetic manipulation in these parasites, this makes understanding the basis of host-microsporidian interactions extremely challenging. Beyond identification of proteins of the spore wall and polar tube, very little known about molecular basis of spore germination and eversion of the polar tube, or the exact details of how the microsporidian sporoplasm is transferred into the host cell. Even though microsporidia can have drastic effects on the organization of the host cell, we know little of the virulence factors and effector proteins that bring about these changes or how they are delivered into the host cell environment either directly or via the parasitophorous vacuole in human infective species such as *E. cuniculi* and *E. intestinalis*.

Spraguea lophii is a microsporidian that infects the monkfish *Lophius piscatorius* and *Lophius budegassa*, inhabiting both the North Atlantic and Mediterranean regions [11]. Compared to those microsporidia that have been sequenced so far, *S. lophii* is an attractive model for identifying microsporidian effector proteins and investigating host-parasite interactions, despite the current lack of an *in vitro* culture system [12]. Infection with *S. lophii* results in the formation of xenomas, large clusters of spore-filled cells in the vagal nerves of the fish that can be several centimeters in diameter and contain spores in various stages of development (Figure 1). In the monkfish, no fitness effects are known to be associated with *Spraguea* infection, and the prevalence rate can be as high as 83% [13,14]. However, xenoma formation in salmon infected with related microsporidia commonly localizes to the gills and is a considerable threat in aquaculture [15]. *S. lophii* spores are easily purified from xenomas in large quantities, providing an opportunity to study germination and to perform experiments that are difficult or impossible in other microsporidia due to the difficulty of obtaining sufficient amounts of parasite material. In addition, *S. lophii* is the first microsporidian parasite of fish to be sequenced, with the potential to provide new insights into host-parasite interactions in these economically important vertebrates.

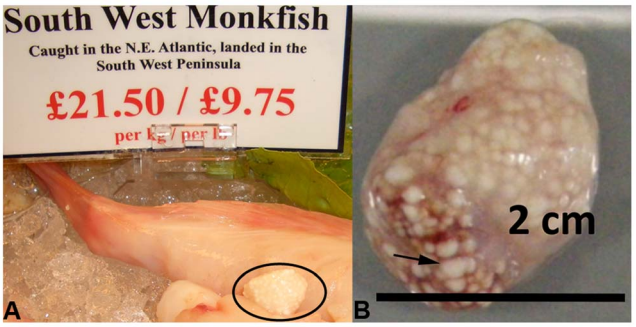


Figure 1. Microsporidian xenomas *in situ*. *S. lophii* grows and replicates inside large cysts visible to the naked eye; these cysts are comprised of swollen, spore-filled fish cells called xenomas. **A**. Cyst of *S. lophii* xenomas *in situ* in commercially available monkfish (circled). **B**. Close up of a cyst of xenomas. Arrow points to a single enlarged fish cell (xenoma).
doi:10.1371/journal.pgen.1003676.g001

Our aim here is to provide a genomic resource that will facilitate future work on this promising model microsporidian.

S. lophii was the first microsporidian to be explored with genome-scale sequencing, and 120 Kb of the genome has previously been published [16]. Here we present 4.98 Mb of unique sequence from the *S. lophii* genome as determined by Illumina sequencing, representing 70–80% of the complete genome (estimated at 6.2–7.3 Mb [17–19]) and, based on our analyses, the great majority of the coding DNA. We investigated the evolution of the *S. lophii* proteome, using OrthoMCL [20] to identify proteins that are unique to *S. lophii* and may therefore be associated with the unique xenoma formation seen in microsporidian infections of fish. To explore the utility of *S. lophii* as a model for understanding core features of microsporidian biology, we combined complex mix proteomics with TruSeq transcriptomics to characterize the proteins expressed and secreted during spore germination, a key time point in the lifecycle of this intracellular parasite. Our results highlighted the importance of microsporidia-specific and fast evolving proteins in germination and host interaction.

Results

Genome architecture of *Spraguea lophii*

The sequence data are summarized in Table 1. Our sequencing resulted in 1392 contigs over 500 bp in length to give a total of 4982 Kb (This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ATCN00000000. The version described in this paper is version ATCN01000000). These contigs had a maximum length of 46788 bp and an N50 length of 5923. The average coverage

Table 1. Summary of sequence data.

Total number of analysed contigs (over 500 bp)	1392
Average length	3578
Total amount of assembled sequence data in contigs >500 bp	4982267
Total number with open reading frames	1116
Total number of annotated proteins or truncated proteins	2543
Total number of tRNAs	53

doi:10.1371/journal.pgen.1003676.t001

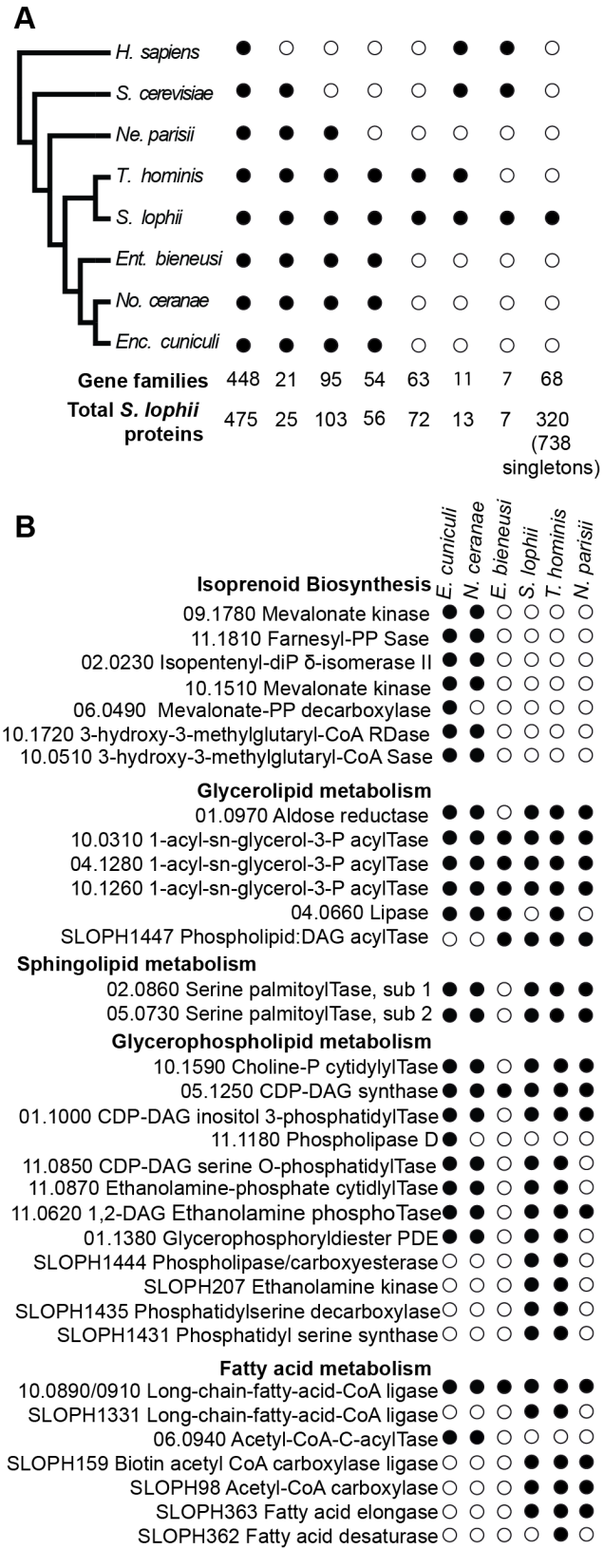


Figure 2. Comparison of the gene content of *S. lophii* to published microsporidian genomes. A. Phylogenetic distribution of annotated *S. lophii* proteins. Gene families (OrthoMCL clusters) were mapped on to a multigene phylogeny of microsporidia and their opisthokont relatives (see also Figure 7A.). **B**. Graphical illustration of the variation in fatty acid synthesis enzymes encoded by six different microsporidian genomes. Although the *S. lophii* genome is larger than that of *E. cuniculi*, it has lost some components of the pathway that are conserved in the smaller genome, such as isoprenoid biosynthesis. Reference numbers are given for *E. cuniculi* or *S. lophii* locus ID's.
doi:10.1371/journal.pgen.1003676.g002

across these contigs is 70× (Mode = 29×, Median = 59×). The overall GC content is low at 23.4%, rising slightly to 25.7% in protein coding regions. The karyotype has been investigated both by pulse field electrophoresis and by using an ultrathin multiwire proportional chamber-based detector [17–19]. This predicts a variable karyotype between isolates from different geographic regions, but the consensus is that there are 15 chromosomes, of which 10–13 are unique, and the overall genome size is estimated at 6.2 to 7.3 Mb. In our assembly we identified 2,573 predicted open reading frames or fragments of, which made up 52% of our assembly. This gives it an intermediate coding density amongst microsporidia, which is consistent with an emerging pattern for microsporidian genomes where coding density decreases with increasing genome size. For example, *Trachipleistophora hominis* has 34% coding DNA and an estimated genome size of 8.5–11.6 Mb, while 86% of the 2.9 Mb *E. cuniculi* genome is made up of coding DNA [5,8]. Our assembly covers 70–80% of the *S. lophii* genome based on size estimates, but is likely enriched for coding regions. To evaluate the completeness of our assembly, we searched the *S. lophii* genome for the presence of genes involved in a range of core metabolic pathways as described previously [8,21] (Table S1). Our assembly encodes most major metabolic pathways in full, including glycolysis and trehalose metabolism as well as a full complement of transfer RNAs (tRNAs) and protein kinases (Table S1). We identified only a few absences, many of which are also absent in related microsporidia such as *T. hominis*, lending support to our coverage of the protein-coding component of the genome. As an independent check on the completeness of our assembly, we compared the transcripts from our *de novo* transcriptome assembly (see below) to the genes predicted on the genome. Based on BLAST similarity to genes from other microsporidia, there were only 20 additional *S. lophii* genes in the transcriptome that did not map to the genome assembly, 5 of which represent transcripts from LTR retrotransposons. Taken together, these analyses suggest that our assembly represents a largely complete sampling of the coding component of the *S. lophii* genome.

Loss and retention of metabolic pathways in *Spraguea lophii*

To investigate the evolution of gene content in *Spraguea*, we mapped the taxonomic distribution of microsporidian gene families onto a cladogram (Figure 2A) (derived from a multiprotein phylogeny - see below). We built gene families using OrthoMCL [20] on a broad sampling of microsporidian genomes, with *Homo sapiens* and *Saccharomyces cerevisiae* as opisthokont outgroups. Our analysis included genomes covering a broad taxonomic spectrum of sequenced microsporidia, including *Nematocida parisii* Ertm1 [22], *T. hominis* [8], *Nosema ceranae* [23], *E. cuniculi* [5], and *Enterocytozoon bieneusi* [24]. The results indicated that 19% of the predicted proteins are shared with all sampled opisthokonts, 1% are specific to sampled fungi, 4% are specific to microsporidia and conserved across the group, 3% are found in clusters of proteins only present in *T. hominis* and *S. lophii*. 42% the 2499 analysed *S.*

lophii proteins, do not cluster with proteins from any of the other organisms in our analysis and of these 30% cluster with other *S. lophii* proteins, indicating that they are part of multiprotein families within the *S. lophii* genome. Table S2 gives a complete classification of the *S. lophii* proteome by OrthoMCL analysis.

E. cuniculi and other microsporidia have a reduced lipid metabolism repertoire [5,24]. A key missing step is the initial reaction of fatty acid synthesis, the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase. Homologues of both acetyl-CoA carboxylase and the biotin-[acetyl-CoA-carboxylase] ligase that it depends upon are present in the *S. lophii* genome as well as in the *T. hominis* and *N. parisii* genomes, suggesting that they are capable of performing this reaction. However, as in *E. cuniculi*, no fatty acid synthase is evident, though both a fatty acid elongase and desaturase are present (Figure 2B). Interestingly, the presence of these additional components in *S. lophii* was predicted on the basis of comparative liquid chromatography of the lipid composition of *E. cuniculi* and *S. lophii* spores which showed a higher level of docosahexaenoic acid, an unsaturated fatty acid, in *S. lophii* than *E. cuniculi* [25]. In comparison to *E. cuniculi*, the *S. lophii* genome also encodes more enzymes for glycerophospholipid synthesis, allowing for a greater variety of interconversions between different types of phospholipid for membrane integration. In contrast, and despite more complexity in some aspects of fatty acid metabolism, no components of the isoprenoid biosynthesis pathway were present in the *S. lophii* genome, though these are encoded in the *E. cuniculi* and *N. ceranae* genomes [5,23] (Figure 2B). These are also absent from our transcriptome data, and from the genomes of *T. hominis*, *N. parisii* and *E. bieneusi*. Taken together, these results suggest that several lineages of microsporidia have independently lost the ability to biosynthesize isoprenoids, a capability that is otherwise conserved across the tree of life [24]. This pathway has been shown to be essential in prokaryotes, and whilst some parasitic Apicomplexa have replaced the classical mevalonate biosynthesis pathway with the alternative MEP pathway [26], it too is absent from these microsporidian genomes. These data suggest that some microsporidia scavenge sterols from the environment. This pathway has been lost in microsporidia with diverse hosts: *E. bieneusi* (mammals, insects), *T. hominis* (mammals, insects), *N. parisii* (worms) and *S. lophii* (fish), suggesting that there is nothing specific about the host biochemical environment that is driving the loss of this pathway.

Key components of the RNAi system are encoded by the *S. lophii* genome, including a dicer protein, an argonaute protein and fragments of an RNA dependent RNA polymerase. This is consistent with emerging genomic data for microsporidia with larger genomes such as *N. ceranae* and *T. hominis* that possess transposable elements [8,23]. This suggests that RNAi was present in the common ancestor of microsporidia but has been secondarily lost in highly reduced genomes such as *E. cuniculi* and *E. bieneusi* [5,8,24]. A comparable situation exists in ascomycete fungi, where RNAi has been lost from the compact *S. cerevisiae* genome but is conserved in other budding yeasts [27].

Relatively few proteins with homology to characterized domains are found in *S. lophii* but no other microsporidia. There are proteins annotated on the basis of similarity to PFAM domains such as kinases, phosphatases and acetyltransferases, which are difficult to relate to specific functions within the cell. One notable exception is a glutamate-ammonia ligase domain containing protein, which can catalyze the generation of glutamine from glutamate and ammonia [28]. This protein is nested between genes with homologs in other microsporidian genomes (a DNA binding protein and acetyl-CoA carboxylase) (Figure S1) and in phylogenetic analysis does not fall into a clade with other fungi, but rather with prokaryotes meaning that it is not clear whether this gene was acquired by lateral transfer or by vertical inheritance. Fish excrete their nitrogenous waste products as ammonia across gills, but glutamate-ammonia ligases are expressed in the brains of fish and other vertebrates to protect from fluctuations in ammonia levels [29]. This protein may have a similar role in protection against ammonia stress in the microsporidian: Whilst the fish is alive the microsporidia may be protected by host ammonia defense mechanisms, however, once the fish dies, microbial degradation of the fish can increase ammonia levels [30]. As the spores of *S. lophii* are embedded deeply within the nervous tissue of the monkfish, they may have to be liberated after the death of the fish, and this glutamate-ammonia ligase may allow the spores to survive fluctuations in ammonia levels in the decaying fish tissue.

Despite the relatively close relationship of *T. hominis* to *S. lophii* (Figure 2A), these two species share few genes that are not found in other microsporidian genomes in our comparison, and most of these are uncharacterized, lineage-specific or fast evolving genes with no similarity to genes in other lineages. A handful of genes shared between *T. hominis* and *S. lophii* have a function that can be predicted on the basis of homology to characterized proteins from model organisms. These include an arsenite transporting ATPase, which might act as an efflux pump in the cell membrane and a cold shock domain protein, which could allow the cell to survive at lower than optimal temperatures [31].

Leucine rich repeat (LRR) proteins

The largest protein family expansion present in *S. lophii* is a family of proteins containing LRRs (Figure 3). Fungal genomes generally encode fewer LRR proteins than their animal relatives [32] though expansion of LRR protein families as pathogenicity factors is known in pathogenic fungi [33]. The genome of *E. cuniculi* encodes just 9 LRR genes in total [5], yet in stark contrast, we have found 97 ORFs encoding fragments of LRR proteins in the *S. lophii* genome, 52 of which appear complete (that is, they have a predicted start and stop codon), and 35 of which appear in our transcriptome data. We used PFAM and MEME [34,35] to identify common conserved motifs, SignalP 4.1 and TargetP 1.1 to look for presence or absence of a secretion signal, and TMHMM 2.0 [36,37] to look for evidence of transmembrane domains that could anchor the proteins in the membrane of the parasite as a potential LRR receptor protein. MEME analysis shows the presence of three different leucine-enriched motifs in the proteins (Figure 3). We also found that the majority of the proteins (37/52) have a predicted signal peptide but no transmembrane domain, meaning that they are potentially a family of secreted parasite effector proteins (Figure 3). Leucine rich repeat proteins often mediate protein-protein interactions, particularly through the formation of dimers [38]. One possibility is that, if these microsporidian LRR proteins are secreted into the host, they could potentially interfere with the formation of dimers of host proteins, disturbing the functional dimer-monomer cycle by sequestering them into inactive dimers, a mechanism seen in mammalian cells [39]. Perhaps surprisingly, similar sequences are also found in the distantly related microsporidian parasite of humans, *Vittaforma corneae*, but in no other microsporidian species for which there is an available genome sequence (Table S3). A large family of leucine-rich proteins was recently reported in another microsporidian, *T. hominis*, although the two families are not related and no predicted secretion signals were reported for the *T. hominis* family [8].

Although the *S. lophii* genome encodes 97 LRR-containing ORFs, we have evidence of expression for only 35. Therefore, we

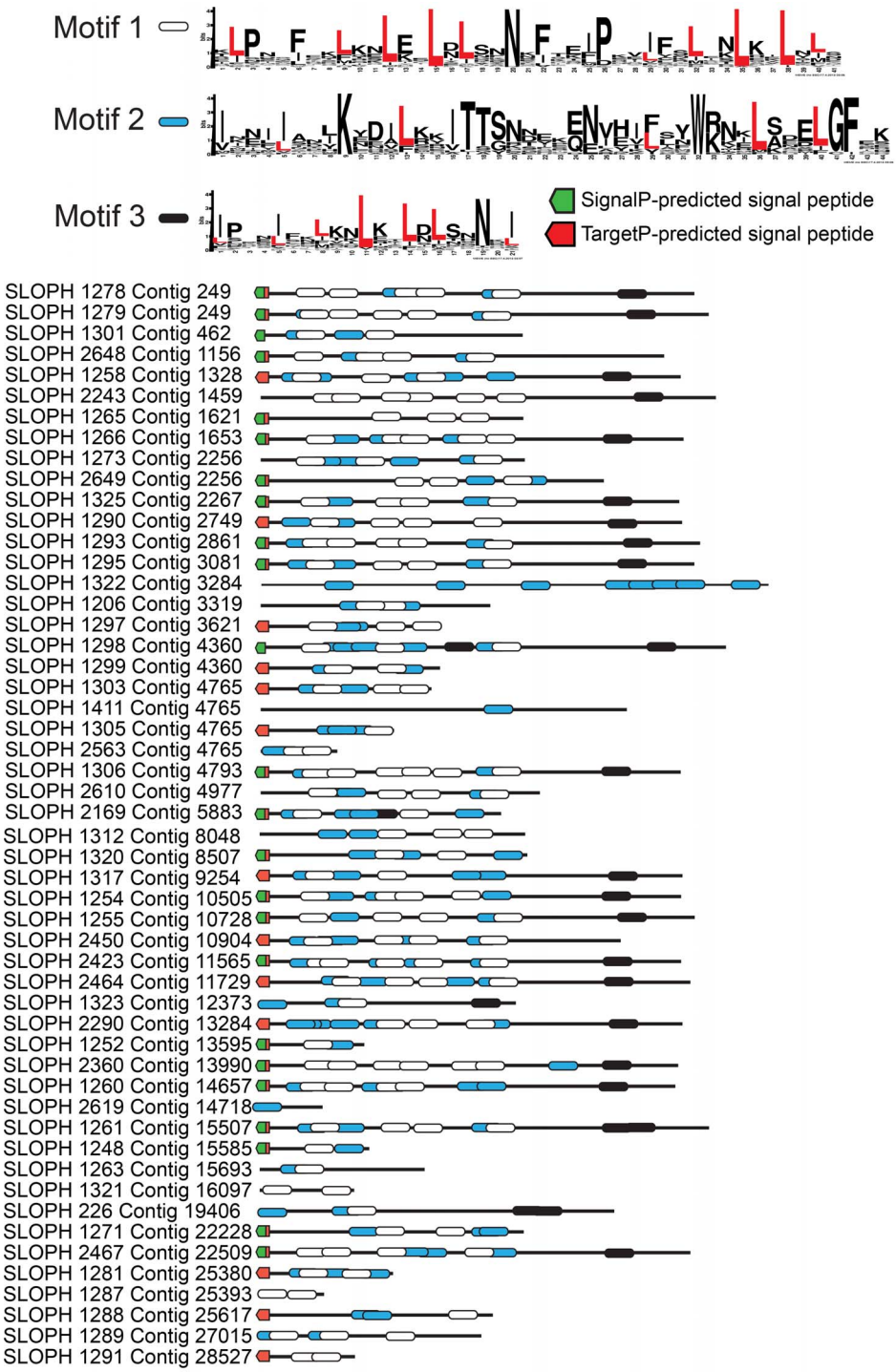


Figure 3. An expanded family of 52 complete leucine-rich repeat proteins in the *S. lophii* genome. This family represents the largest *S. lophii*-specific protein family expansion, and interestingly, many of its members have N-terminal signal peptides, raising the possibility that some may be secreted or targeted to the parasite cell surface for host interactions. Leucine-rich motifs at non-overlapping sites with a p-value lower than 0.0001 were identified using MEME [34] and signal peptides identified using SignalP 4.1 (green) and TargetP (red) [37].
doi:10.1371/journal.pgen.1003676.g003

cannot exclude the possibility that some family members are never expressed and are pseudogenes. In other eukaryotic parasites such as *Trypanosoma* and *Giardia*, large protein families contain many pseudogenes which provide the genetic variation for the ongoing process of antigenic switching [40,41]. If microsporidian multigene families interact with the host, switching of expression from gene to gene may allow escape from the fish adaptive immune response over the course of infection [42].

The transcriptome of germinated *S. lophii* spores

The cDNAs of artificially germinated spores were sequenced by the non strand-specific TruSeq approach using Illumina

sequencing. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GALE00000000. The version described in this paper is the first version, GALE01000000. We used Trinity RNA-Seq [43] to assemble the spore transcriptome, and RSEM [44] to quantify the relative abundance of transcripts. Our *de novo* transcriptome assembly contained 12,932 unique transcripts, of which 2,896 mapped to the *S. lophii* genome. Although relatively small in number, these transcripts made up 67.7% of the transcriptome by abundance, and likely represent the majority of the *S. lophii* genes in the dataset. 2,514 of these transcripts mapped to existing *S. lophii* genes, so that 1,598 of the predicted 2,539 open reading frames had at least one matching transcript. A small proportion of transcripts (398) mapped to regions of the genome assembly without gene predictions; manual inspection of these cases provided evidence for 30 additional genes that had been missed by the initial annotation process.

The remaining third of transcripts that did not map to the *S. lophii* genome assembly might represent contaminants, unmapped *S. lophii* genes, or artifacts of the assembly process. To distinguish between these possibilities, we searched these transcripts against the NCBI nr database using BLASTX [45]. Many showed high sequence identity to *Pseudomonas* and *Flavobacterium* genes and likely represent contaminants. We did, however, identify 20 additional genes with best BLAST hits among the Microsporidia, particularly *T. hominis* and *Vavraia culicis*; these are most likely *S. lophii* genes from the unassembled portion of the genome (Table S4). Interestingly, this set of 20 genes included one LTR and four non-LTR retrotransposons with similarity to those found on the *T. hominis* genome [8]. While the *T. hominis* retrotransposons are largely fragmented and pseudogenized, these results demonstrate that at least some of their homologues in *S. lophii* remain active.

Overall, our analyses of the *S. lophii* transcriptome provided support for the completeness of our genome assembly and for our gene calling approach, and also provided some new insights into microsporidian biology. Of the 1,986 genes in our genomic data that have complete open reading frames, that is, they have a start and a stop codon, 265 have complete coverage and for these, the RNA transcript was on average 132 base pairs longer than the gene. Five of these show more than one gene in the transcripts. However, given the short read-length of Illumina sequences and the possibility that transcripts of adjacent divergent genes could erroneously assemble, it is not possible to say whether these transcripts are overrunning into downstream genes as seen in other microsporidian species [46,47].

Interestingly, although the most abundant transcripts corresponded to 18S ribosomal RNA (Table 2), the most highly expressed protein in *S. lophii* spores is an uncharacterized ORF, with homologues found only in a limited number of other microsporidia. Indeed, while the list of highly transcribed proteins contained many of the expected candidates (ribosomal proteins, ATP-binding proteins, transcription factors, and proteins involved in energy metabolism), these were intermingled with a number of uncharacterized proteins annotated as hypotheticals (see Table 2), strongly suggesting an important role for novel, lineage-specific or very fast evolving proteins in *S. lophii* and microsporidian biology. An interesting observation is that microsporidia-specific proteins conserved in multiple species make up a larger proportion of the transcriptome than *Spraguea*-specific proteins (Figure 4), and a larger proportion of the transcriptomic data than they do of the genomic data. These potentially novel proteins originating in the common ancestor of microsporidia may be good targets for future experimental work on the maintenance of the parasitic lifecycle both in *S. lophii* and other microsporidian species.

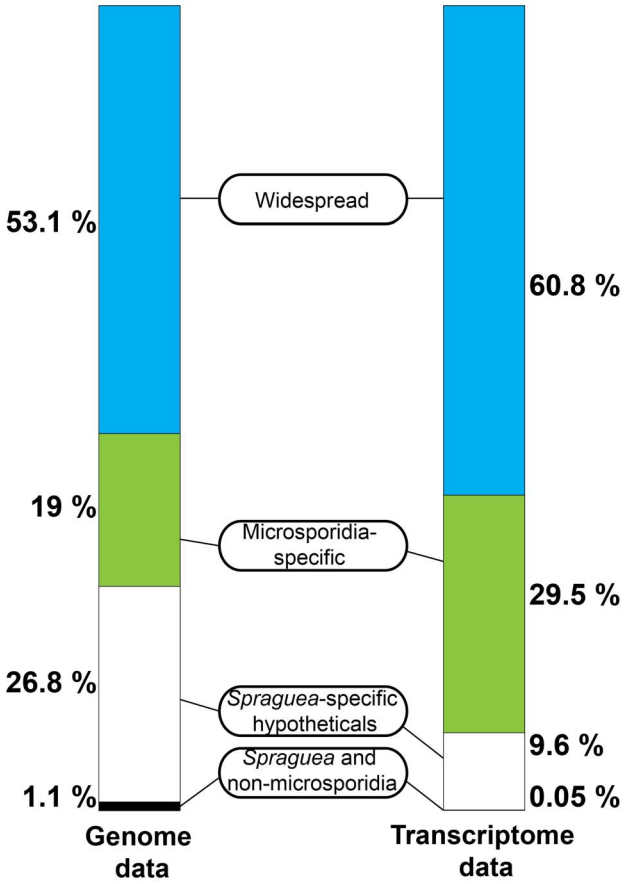


Figure 4. Taxonomic distribution of predicted ORFs in the genome and transcriptome of *Spraguea*. For taxonomic profiling, BLASTP was used to search against the NCBI nr database in January 2013, with a cutoff value of $e<1\times10^{-5}$. ORFs were classified into the four illustrated categories according to the results. Microsporidia- and *S. lophii*-specific ORFs and fast-evolving proteins with no similarity to proteins from other lineages make up a substantial proportion of the transcriptome, suggesting an important role for lineage-specific (or fast evolving) genes in the evolution of microsporidia.
doi:10.1371/journal.pgen.1003676.g004

Splicing and introns in *S. lophii*

Our *de novo* transcriptome assembly also enabled us to investigate the splicing of putative introns in *S. lophii* protein-coding genes. The number of introns in microsporidian genomes is greatly reduced compared to their opisthokont relatives [48]. *E. cuniculi*, *N. ceranae*, and *T. hominis* encode a relatively small (6–78) number of spliceosomal introns, which are largely confined to ribosomal proteins, while the *Nematocida* genus appears to have lost both introns and the splicing machinery entirely [22]. *S. lophii* does encode conserved components of the splicing machinery, so we searched its coding sequences for introns using a two-step approach. First, we scanned the genome with a consensus microsporidian intron motif built from comparisons of the introns in *E. cuniculi* and *T. hominis* [5,8,49]. This search returned hits to 8 genes, 6 of which encode ribosomal proteins (Figure 5); the two non-ribosomal proteins included genes encoding the DNA replication licensing factor Mcm1 and a poly(A) binding protein. We then searched the *S. lophii* transcriptome for transcripts containing deletions relative to the genome assembly, which might also indicate the presence of introns. Surprisingly, this analysis identified only two transcripts from which the predicted intron

Table 2. The twenty most abundant protein-coding transcripts in the *S. lophii* transcriptome.

Protein ID	Protein annotation	Abundance as percentage of transcriptome
SLOPH 2267	hypothetical protein SLOPH 2267	3.2715733
SLOPH 2328	ADP-ribosylation factor	3.133362
SLOPH 857	hypothetical protein SLOPH 857	2.7781082
*SLOPH 54	60S acidic ribosomal protein P2	2.127622
SLOPH 997	hypothetical protein SLOPH 997	1.9514969
*SLOPH 2039	ubiquitin/40s ribosomal protein S27a fusion	1.7599442
+SLOPH 655	hypothetical protein SLOPH 655	1.5733076
SLOPH 2438	GTP-binding protein rho1	1.4452069
SLOPH 1855	hypothetical protein SLOPH 1855	1.245524
*SLOPH 1540	RAS GTPase	1.0318709
*SLOPH 2369	14-3-3 protein	0.969516
+SLOPH 431	Histone H3	0.9530647
SLOPH 2171	40S ribosomal protein S28	0.8755324
SLOPH 128	AN1-like Zinc finger protein	0.7700206
SLOPH 1788	hypothetical protein SLOPH 1788	0.7537534
*SLOPH 101	actin	0.7490758
SLOPH 2162	rRNA binding protein	0.644547
SLOPH 1405	nucleoside diphosphate kinase	0.6361997
SLOPH 1378	myosin regulatory light chain	0.6137582
SLOPH 1203	hypothetical protein SLOPH 1203	0.6130135

[+]Also found in secreted proteomics data.
*Also found in germinated or non-germinated proteome data.
doi:10.1371/journal.pgen.1003676.t002

sequence had been spliced, corresponding to two of the eight genes identified by our motif scan (ribosomal protein S23 and poly(A) binding protein, Figure 5), transcripts for the other six genes still contained the putative intron motif. A comparison of the sequences of the two actively spliced introns with those of the other intron-like sequences revealed three striking differences (Figure 5): the spliced introns are much longer, are out of frame with the coding sequence, and are located further downstream from the 5′ end of the gene (89 and 152 nucleotides 3′ of the start codon, as opposed to directly adjacent to that codon in all other cases). Of the eight putative intron-containing genes we identified in *Spraguea*, five have orthologues in *E. cuniculi* that also contain an intron (S17, L27a, S24, L5 and poly(A) binding protein), and the efficiency with which those introns are spliced parallels our results with the *S. lophii* transcriptome. The introns in *E. cuniculi* S17, L27a, S24 and L5, for which we did not detect splicing in *Spraguea*, are also short [49] and are among the least efficiently spliced genes in *E. cuniculi*, with less than 15% of transcripts experiencing splicing (a figure which drops to 5% for L5) [50]. In contrast, the *E. cuniculi* orthologue of the actively-spliced poly(A) binding protein contains the longest and most frequently spliced intron in *E. cuniculi*, with over 80% of transcripts spliced. Thus, it appears that the properties determining intron splicing efficiency are conserved between these two distantly related microsporidia; it will be interesting to see if they hold more generally for other intron-containing microsporidian genomes.

These observations raise the question of how genes containing intron-like sequences that are rarely, if ever, spliced can be

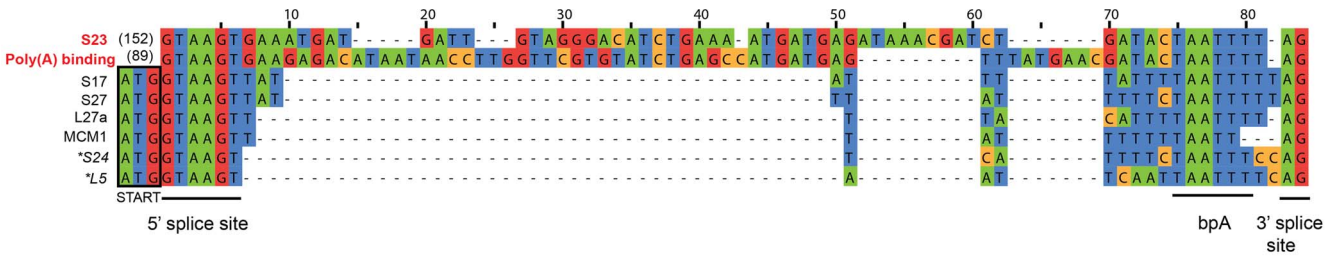


Figure 5. Putative introns in *S. lophii*. The regions corresponding to the conserved microsporidian motifs that interact with the splicing machinery (5′ and 3′ splice sites, branch point bpA) are indicated [49]. The splicing of introns highlighted in red (S23 and poly(A) binding protein) was confirmed at the transcriptome level; transcripts from the other genes retained the intron sequence, suggesting that they are rarely, if ever, spliced in *S. lophii*. The actively spliced introns are located 152 and 89 nucleotides downstream of the translation start site, while all others are immediately adjacent to the start codon (outlined in black). Genes marked with an asterisk could, in principle, be read through to produce a short insertion in the protein sequence, but the others cannot be read through because they contain an in-frame stop codon.
doi:10.1371/journal.pgen.1003676.g005

adequately expressed in *Spraguea*. The genes containing these motifs encode some of the most widely conserved and functionally important proteins in cellular life forms, including components of the ribosome and a DNA replication factor. Several of these intron-containing proteins were identified in our whole cell protein analysis of germinated and non-germinated spores. Ribosomal proteins S27, S24 and L5 were present in our germinated sample and S27 and L5 were also present in the proteome of dormant spores (Table S5). For two of the six genes (S24 and L5), inspection of the intron-like sequence suggests a simple explanation: indels in these sequences have caused a frameshift such that the intron can be read through from the upstream ATG without encountering an in-frame stop codon. Thus, a full-length protein containing a short N-terminal insertion could be expressed from these transcripts in the absence of splicing. The other four introns contain in-frame stop codons such that translation from the upstream ATG is not possible unless the intron is spliced. We considered the possibility that translation could instead begin at an alternative start codon downstream of the intron. However, initiation at the next available, in-frame ATG would result in substantial N-terminal deletions (covering 25–37% of the coding sequence) for three of these genes, and in the case of ribosomal protein S17 no suitable alternative start codon is available. Thus, it remains unclear whether these genes can be expressed without splicing, or whether translation depends on a rate of splicing too low to be detected in our assay; it was recently suggested that low rates of transcript degradation might partially ameliorate this problem in *E. cuniculi* [50].

Proteomics of germination and secretion

Dissecting the interactions between microsporidia and their host requires an understanding of the process of spore germination, in which the spore leaves dormancy and rapidly expels a long polar tube, through which the spore's cellular contents exit the spore and enter the host cell. The specific host cell stimuli inducing microsporidian spore germination are unidentified and likely complex, however several successful methods for artificial spore germination *in-vitro* have been described in a range of species [51–53]. At present, both the changes within the spore that trigger germination, and the identity of the secreted effector proteins used by the parasite gain entry into the host cell and control its biology are unknown. Genetic manipulation is a powerful tool for identifying these factors in many pathogens, but is not yet available for any microsporidian. However, *S. lophii* xenomas are densely packed with spores, providing an abundant source of parasite material for proteomic comparisons between the dormant and germinated spore stages.

To identify any proteins present in artificially germinated but not dormant spores, we analyzed whole protein extractions of both lifecycle stages with mass spectrometry of complex protein mixtures. After pooling and filtering 3 biological replicates our analysis showed no consistent variation at the proteomic level between the two lifecycle stages (Table S5). We did find components of many core pathways in germinated and non-germinated spores, such as histones, heat shock protein and ribosomal proteins. We also find glycolytic enzymes in both germinated and non-germinated spores, which is consistent with recent work that glycolytic pathways are specifically active in the spore stage [8,54]. Two components of the secretory pathway (Sec23 and Sec24) were identified only in germinated spores, which may indicate its activation specifically upon germination, however these were not consistently found in all three replicates and overall we found a surprisingly conserved repertoire of proteins between the two samples. It may be that germination happens too rapidly to allow for translation, with microsporidia pre-packaging the proteins needed for immediate use upon recognition of the germination stimulus and therefore that obvious changes in protein complement may come later in development in meront and sporogonial stages. Alternatively it may be that the samples are dominated by highly expressed housekeeping proteins, and the proteins that vary between the two samples may be present at low levels not easily detectable by complex mix proteomics.

Next, we investigated the complement of proteins present in the extracellular medium after *in-vitro* germination. Here, we consistently retrieved a small subset of proteins that were visualized by SDS-PAGE (Figure S2). Importantly, no proteins were identified from the supernatant of non-germinated *S. lophii* spores, suggesting that the identified proteins are released specifically by the parasite upon germination. These could be proteins released by the sporoplasms on early infection or by the spore during germination. We retrieved 37 proteins from three quality-filtered replicates (Figure 6). Of these, 11/37 proteins are predicted by SignalP 4.1 to have a secretion signal and 17 are predicted by TargetP 1.1 to be directed to the secretory pathway; in some cases these predictions overlap (Figure 6). Proteins secreted to the extracellular medium range in size from 101 amino acids to 935 amino acids however there is no obvious correlation between size and the presence or absence of a predicted secretion signal (Figure 6). Five proteins are consistently present in all three replicates. Three of these have no similarity to any other proteins in the NCBI nr database ($e<1\times10^{-5}$) (SLOPH 477, 723, 762), while two of the proteins are found in other microsporidia (SLOPH 1766, 1854). One of these has features which make it particularly interesting in the context of potential effector proteins: it is part of a multigene family whose members have predicted Ricin-B-lectin domains and is found in several copies in *S. lophii*, as well as in several other microsporidian genomes (SLOPH 1766-see below). The other secreted protein shared with other microsporidia has significant sequence similarity to a spore wall protein (SWP7) in *Nosema bombycis* (SLOPH 1854).

Ten proteins are found in two of three replicates. Six of these proteins have no similarity to any proteins in the NCBI database ($e<1\times10^{-5}$) and are potentially unique to *S. lophii*, demonstrating that species-specific innovations may play a crucial role in *S. lophii* invasion of the host cell. Four proteins are found in *S. lophii* and other microsporidia, including SLOPH 1749, which contains an exonuclease/endonuclease/phosphatase domain, SLOPH 2344, a spore wall protein and another RICIN-B lectin domain protein (SLOPH 691 - see below). A chitin deacetylase, which is conserved across the Microsporidia, was also identified in 2/3 biological replicates. The orthologue of this protein has been studied in *E. cuniculi*. In this species it is expressed at high levels during sporogonial lifecycle stages and localizes to the endospore, where it accumulates in paramural bodies [55,56]. However, later functional analysis demonstrated that the protein was unable deacetylate chitooligosaccharides or bind chitin or any glycans, suggesting a non-canonical function [57]. Urch *et al* speculate that this chitin deacetylase may have evolved from a carbohydrate active enzyme to a lectin and may bind carbohydrates in the *E. cuniculi* cell wall [57]. A total of five microsporidia-specific uncharacterized proteins were observed in at least 2/3 biological replicates (SLOPH 1854, 1766, 2344, 1749, 691) (Figure 6). Orthologues of these proteins display a varied taxonomic distribution and secretion prediction profile across microsporidia, although two of them - SLOPH 2344 and SLOPH 1749 - are present in all microsporidia examined (with the exception of *N. parisii*) and are predicted to be secreted in all species expect *E. intestinalis* and *E. bieneusi* (SLOPH 2344), or just *E. bieneusi* (SLOPH 1749) (Table S6). Some proteins lacking predicted secretion signals are potentially highly expressed proteins that are released during

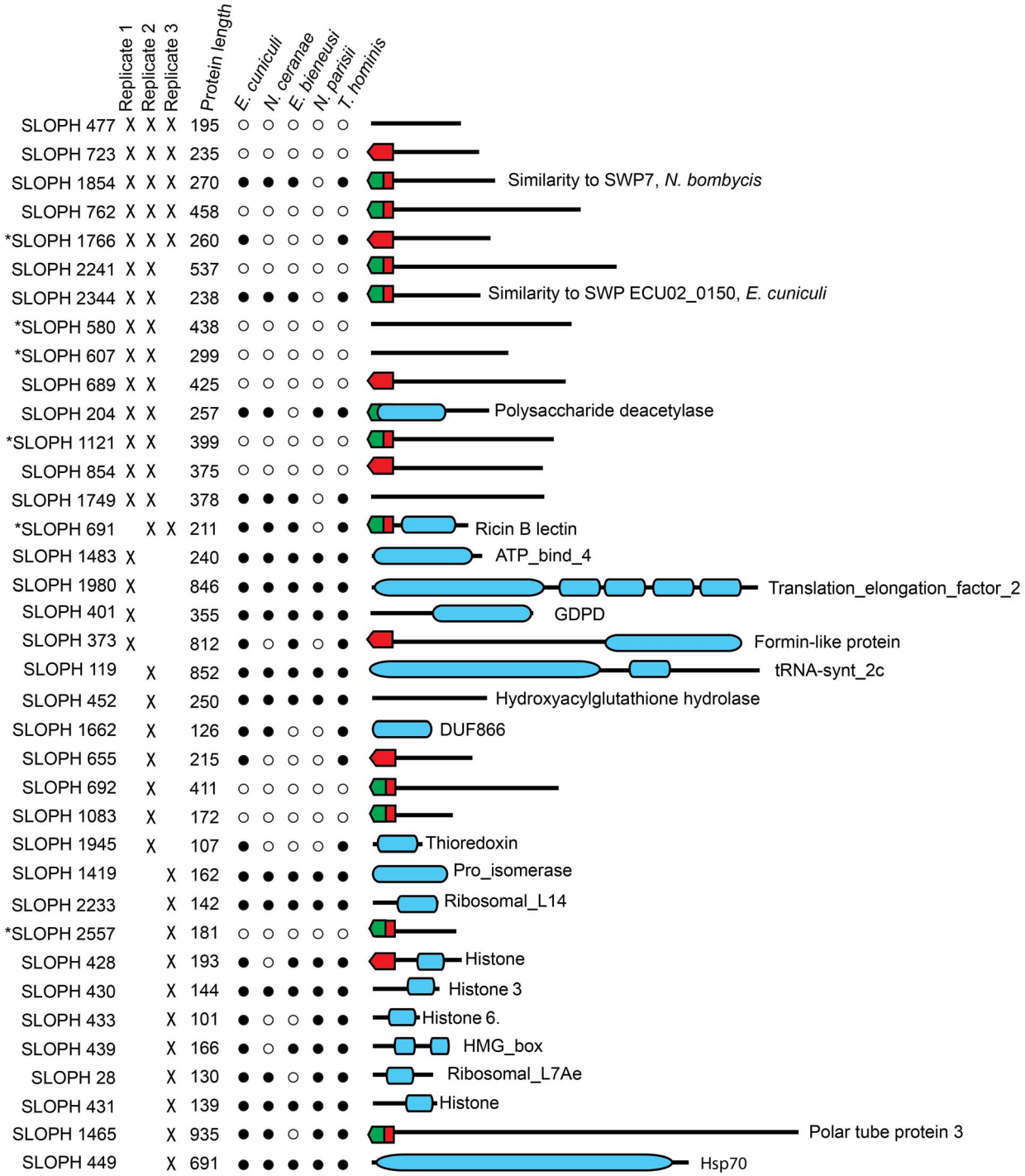


Figure 6. Proteins found in mass spectrometry analysis of medium from germinated *S. lophii* spores. Targeting signals predicted by SignalP (green) or TargetP (red) are indicated. Presence in each replicate of the experiment is indicated with a cross. Stars next to proteins indicate the presence of multiple orthologues of the protein in the *S. lophii* genome. Presence or absence of the protein in other microsporidian genomes is indicated with a full or empty circle respectively. Recognized PFAM domains are indicated by blue bars.
doi:10.1371/journal.pgen.1003676.g006

the germination protocol and are therefore potentially false positives, for example histone proteins and translation elongation factor which were found in single replicate. Others, that are present in 2 or 3 replicates, may be secreted via a non-canonical secretion signal not detected by bioinformatics prediction programs.

The *S. lophii*-specific proteins may be host-driven innovations that mediate interactions specifically with the fish host, representing cases of lineage-specific adaptation or microsporidian-specific proteins that are fast evolving and not easily recognized between species. Some of the identified proteins are part of multigene families with other homologs in the *S. lophii* genome (Starred in Figure 6,

Table S7) though proteomic data correspond to one member of the family. In agreement with our transcriptome analysis, these results demonstrate the importance of uncharacterized, hypothetical proteins in microsporidian germination, host cell invasion and early infection. This approach also represents a powerful method for identifying the presumably small proportion of novel parasite effectors or virulence factors from among the hundreds, or thousands, of hypothetical proteins found in sequenced microsporidian genomes. The need for such streamlining approaches is obvious, as a genetic manipulation protocol for microsporidia is currently lacking and hence protein characterization relies on heterologous expression systems or indirect assays which are time consuming and not easily applied to a large number of proteins.

A family of lectin-like proteins conserved in the Microsporidia

Two of the identified secreted proteins (SLOPH 1766 and SLOPH 691) are conserved in other microsporidia and show similarity to RICIN B-lectin proteins, with weakly conserved motifs involved in carbohydrate binding. Lectins have diverse roles in parasites, and can mediate adhesion of the parasite to the host cell during infection, but they can also play roles in immune evasion and their binding to host proteins can trigger different developmental pathways in the parasite [58,59]. The members of this family form clusters in the genomes of *E. cuniculi* and *N. ceranae* (Figure 7A), reminiscent of clusters of effector proteins in fungal pathogens such as *Ustilago maydis* [60] and the RXLR and Crinkler effector families in oomycetes of the genus *Phytophthora* [61]. In *E. cuniculi*, four proteins are found in a single syntenic block, whereas in *N. ceranae*, a block of six proteins sit together in the genome with two others elsewhere. Interrogating the *E. cuniculi* genome, which has well-assembled chromosomes, these genes are not found at the end of chromosomes in the subtelomeric regions as is the case for effector proteins in other parasite species [62]. It is difficult to draw conclusions about the relative genomic location of the corresponding genes from our *S. lophii* data as several are found on short contigs, but we did identify two clusters of two genes (Figure 7A). A phylogeny of this protein family suggests that most of the paralogs within each species arose from species-specific duplications (Figure S3), again mirroring the pattern seen for effector families in other eukaryotic pathogens such as *Ustilago maydis* [60], *Phytophthora* [61] and *Trichomonas* [63]. The distribution of lectin-like proteins and expanded gene families were plotted onto a multigene phylogeny of microsporidia to give an overview of their distribution across the phylum (Figure 7B); the patchy distribution of these families suggests a process of differential loss and expansion during the radiation of microsporidia. In particular, we did not detect homologues of the lectin-like proteins in the *N. parisii* genome, suggesting that this family may have evolved after the divergence of *N. parisii* from the other sequenced microsporidia.

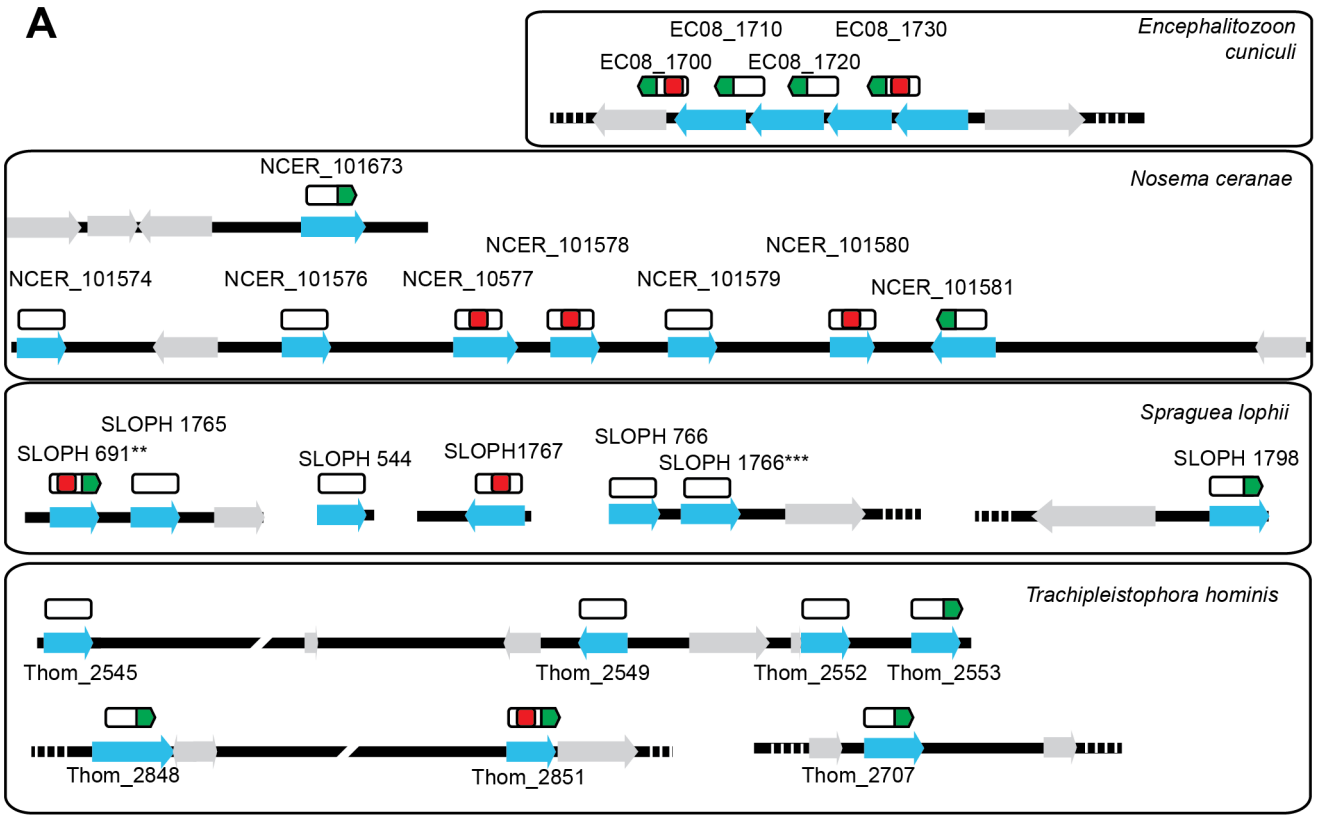
Discussion

Although the *S. lophii* genome is larger than that of *E. cuniculi*, there is no clear trend towards a greater retention of broader ancestral metabolism. Like *T. hominis*, *S. lophii* has apparently lost the ability to biosynthesize isoprenoids, although it has retained more enzymes involved in other aspects of lipid metabolism. These results demonstrate that the loss of major biosynthetic pathways did not occur only in the common ancestor of all microsporidia, but has been an ongoing process throughout the evolution of the group, giving rise to important lineage-specific differences in metabolism.

Beyond reduction in genome size, a relative scarcity of introns is a conserved and striking feature of microsporidian genomes. The remaining introns tend to be short and, in many cases, located at the 5′ ends of genes encoding ribosomal proteins [5]. Our comparisons of splicing in *S. lophii* and *E. cuniculi* suggest that the molecular determinants of splicing efficiency are conserved in these distantly-related species, with longer, out-of-frame introns experiencing the highest levels of splicing. The observation that some inefficiently spliced, intron-like sequences can be read through without producing a truncated or frameshifted protein suggests a plausible, though speculative, mechanism by which introns could be lost from microsporidian genomes: once read-through is possible, subsequent deletions could reduce or remove the newly expressed insertion, resulting in complete loss of the former intronic sequence. We note, however, that the introns we compared in our assay are conserved between *S. lophii* and *E. cuniculi*, and so have not been lost in the period of time since the divergence of these species.

One of the most striking features of the *S. lophii* genome is the presence of a leucine rich repeat protein family, which may represent an expanded gene family of pathogenicity factors. Such expanded gene families are characteristic of other fungal pathogens such as *Batrachochytrium dendrobatidis* and *Blumeria graminis*, have also been reported in the microsporidia *T. hominis*, *E. cuniculi*, *Anncaliia algerae* and *Vittaforma corneae* [8,64–66], and are also considered of importance in host-parasite interactions more generally [67]. Whilst the *S. lophii* members of this family have predicted N-terminal secretion signals, there is no evidence of an obvious conserved peptide motif involved in directing protein secretion into the host such as the oomycete crinkler motifs or the conserved tripeptide motif found in *B. graminis* which could help to define the microsporidian secretome [65,68].

In addition to the conserved microsporidian proteome and an expanded complement of LRR proteins, *S. lophii* encodes a large number of predicted hypothetical proteins. Some of these were only expressed at low levels in our transcriptome analysis, potentially representing false positive ORF calls, but there were also a significant number of highly expressed microsporidia-specific transcripts and together with *S. lophii*-specific transcripts, these made up 39.1% of the total expression in *S. lophii* spores (Figure 4). Thus, lineage-specific or fast evolving proteins likely play an important role in *S. lophii* biology, and the same may well be true for all microsporidia. Although we now have genome sequences for a number of microsporidia, our understanding of their basic biology, pathogenicity and host interactions is limited by the current lack of a genetic manipulation system for these intracellular parasites. Recent work has shown that identification of proteins bearing secretion signals through bioinformatics can reveal interesting strategies by which microsporidia may alter the host environment in their favor. Cuomo *et al.* have identified functional secretion signals in a family of microsporidian hexokinases; when secreted, these effectors may stimulate host metabolism, providing more metabolites for the parasite [22]. These results demonstrate the utility of bioinformatics in predicting signal peptides and identifying microsporidian proteins that may be targeted to the cell surface or secreted. However, signal peptide prediction tools are trained on model eukaryotes, and their accuracy in predicting the full secretome from the highly divergent sequences of microsporidia is unclear. Even when these signals are accurately identified, heterologous characterization is laborious and expensive. One way forward may be through alternatives to genetic transformation, such as those we have explored in this article. Here we have used secretion proteomics to detect candidate virulence factors that are likely to be secreted from the parasite as it germinates, leaves the polar tube and enters


A
Encephalitozoon cuniculi
EC08_1710
EC08_1730
EC08_1700
EC08_1720
Nosema ceranae
NCER_101673
NCER_101578
NCER_101580
NCER_101574
NCER_101576
NCER_10577
NCER_101579
NCER_101581
Spraguea lophii
SLOPH 1765
SLOPH 691**
SLOPH 544
SLOPH1767
SLOPH 766
SLOPH 1766***
SLOPH 1798
Trachipleistophora hominis
Thom_2545
Thom_2549
Thom_2552
Thom_2553
Thom_2848
Thom_2851
Thom_2707
Ricin B lectin domain as detected by PFAM at gathering threshold
SignalP4 predicted secretion signal
*** Detected in all three secreted proteome replicates
** Detected in all two of three secreted proteome replicates
Assembly gap
Contig/Scaffold continues

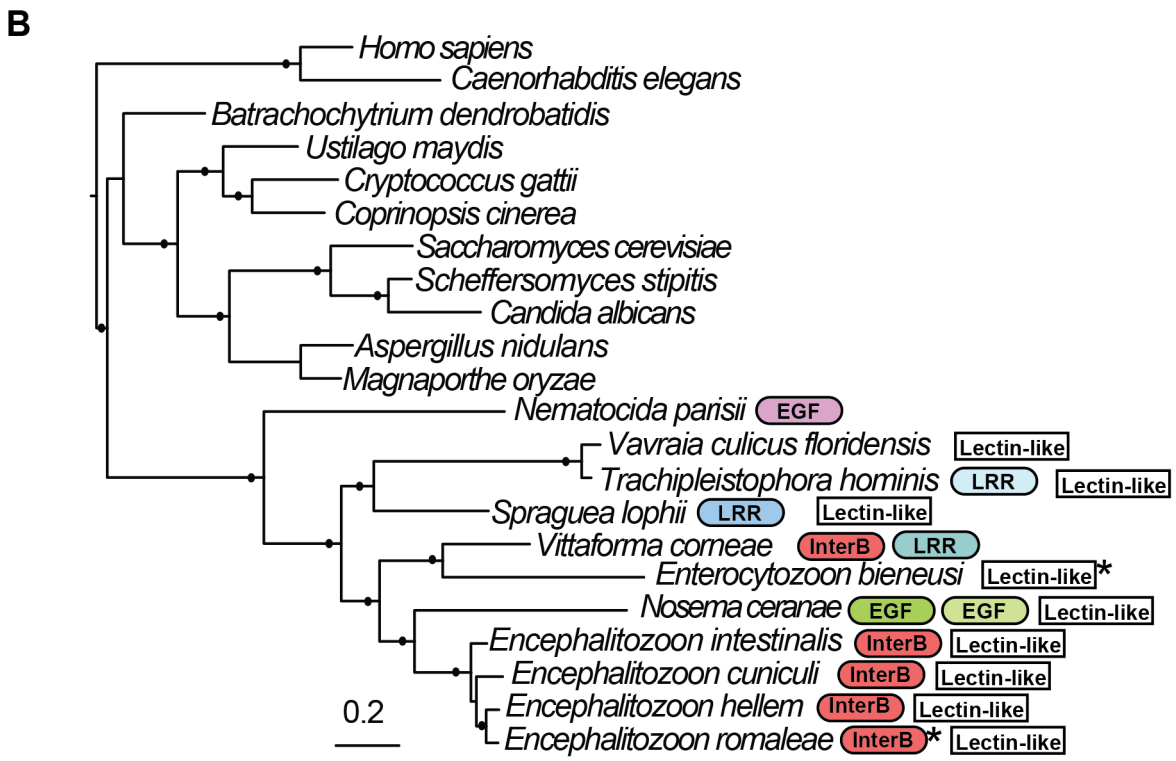

B
Homo sapiens
Caenorhabditis elegans
Batrachochytrium dendrobatidis
Ustilago maydis
Cryptococcus gattii
Coprinopsis cinerea
Saccharomyces cerevisiae
Scheffersomyces stipitis
Candida albicans
Aspergillus nidulans
Magnaporthe oryzae
Nematocida parisii EGF
Vavraia culicus floridensis Lectin-like
Trachipleistophora hominis LRR Lectin-like
Spraguea lophii LRR Lectin-like
Vittaforma corneae InterB LRR
Enterocytozoon bieneusi Lectin-like*
Nosema ceranae EGF EGF Lectin-like
Encephalitozoon intestinalis InterB Lectin-like
Encephalitozoon cuniculi InterB Lectin-like
Encephalitozoon hellem InterB Lectin-like
Encephalitozoon romaleae InterB* Lectin-like
0.2

Figure 7. Genomic context of lectin-like proteins and phylogenetic distribution of expanded gene families in microsporidia. A. Putative lectin-like proteins are shown in their genomic context as blue arrows and unrelated flanking genes are shown as grey arrows. Above these (white rectangles) are shown hypothetical translated proteins (not to scale) to illustrate the positions of predicted motifs. The proteins SLOPH 691 and 1766 that were identified in our secretion proteomics were BLASTed against the *Spraguea lophii* genome at a cutoff of $e<1\times10^{-5}$ to identify other members of the lectin-like family. These proteins also show BLASTP similarity to lectin-like proteins in *E. cuniculi* and *N. ceranae*, as indicated. **B**. RAxML 8 protein phylogeny of microsporidia showing the phylogenetic context of *S. lophii*. Published expanded gene families were mapped onto this phylogeny, including InterB proteins [66], leucine rich-repeat proteins (LRR) [8], and other expanded gene families (EGF) as published [22,23]. Stars indicate the detection of a single homolog of the protein in the genome. Shading is used to denote different families of apparently unrelated LRR proteins. Solid circles on branches indicate a posterior probability of 1 in our Bayesian phylogenetic analysis.
doi:10.1371/journal.pgen.1003676.g007

the host cell. Two of these identified secreted proteins are a part of a family that is broadly conserved across the Microsporidia, with local duplications giving rise to syntenic blocks of family members in several microsporidian genomes. Based on their sequence similarity to lectin domains, they may be involved in binding to carbohydrates on host proteins, and their conservation suggests they play an important general role in microsporidian parasitism. We combined our secretion proteomics with an analysis of the *S. lophii* transcriptome, which identified a further set of highly expressed, microsporidia-specific hypotheticals. These proteins are ideal candidates for further characterization to better understand the molecular basis of parasitism in the Microsporidia, both within the parasite cell and in its host interactions.

Materials and Methods

Spore purification

DNA was extracted from two separate clusters of cysts both collected from local Monkfish (*Lophius piscatorius*) landed at either Plymouth or Brixham caught in the North Atlantic. Spores were separated from fish material in a three-step process. Xenomas filled with microsporidia were removed from the fish tissue manually. Samples were then soaked overnight at 4°C in EDTA (10 mM) containing Triton X 100 at 0.05% and trypsin 0.025% (w/v). The samples were then homogenized in a glass homogenizer until they formed a fine suspension, and washed three times in sterile PBS. The spore suspension was then cleaned by centrifugation through 100% Percoll (Sigma) at 4°C at 1600×g for 15 minutes. Samples were further washed three times in sterile 1× PBS before resuspension and storage at 4°C in 1 ml of sterile 1× PBS with addition of an antibiotic cocktail of 10 µg/ml Ampicillin, Penicillin/Streptomycin and Kanamycin.

DNA extraction and library preparation

Aliquots of 200 µl of purified spores were resuspended in 400 µl TE 10/1 pH 7.5 and ground in a pestle and mortar in liquid nitrogen for 15 minutes. Powdered frozen material was transferred to 800 µl phenol (pH 7.9) and mixed by inversion then centrifuged at 10,000×g for 10 minutes. 400 µl chloroform was added to the aqueous supernatant and this was centrifuged at 10,000×g for 5 minutes. DNA was precipitated from the aqueous layer using a standard ethanol precipitation protocol [69].

Preparation of DNA library for Illumina sequencing

For sample one, the DNA was sheared using a Covalis Nebuliser. Sheared fragments of approximately 400 bp in size were selected using a polyacrylamide gel and used to create a library using the Illumina paired-end sample preparation protocol (revision A June 2008). A single lane of paired-end 36 bp sequence reads was sequenced. For sample two, DNA was sheared by Biorupter sonication and fragments of approximately 600 bp were selected by polyacrylamide electrophoresis for library preparation as above. A single lane of paired-end reads of 76 bp was sequenced giving a total of 325 Mb of data.

Data assembly and annotation

The data from the two runs were pooled and assembled together using Velvet 0.7.50 using a kmer length of 31 and coverage cutoff of 5. Estimated expected kmer coverage using automatic calibration was 6. Contigs of 500 bp or larger were retained and analysed further. This resulted in a set of 3080 contigs, which showed a bimodal distribution of GC content with peaks at approximately 23% GC and 58% GC, indicative of DNA from a second organism (based on sequence identity, a close relative or strain of *Pseudomonas fluorescens*). A histogram of GC content revealed two distinct distributions with little overlap (Figure S4); we therefore discarded all contigs with GC content greater than 40% as potential *Pseudomonas* contaminants. All retained contigs were used to BlastN search gi229587578 *Pseudomonas fluorescens* SBW25 chromosome and gi146343893 *Pseudomonas fluorescens* SBW25 plasmid pQBR103, to verify that there was no *P. fluorescens* DNA in the remaining contigs. Removed high GC contigs were also used to BLAST search the same genomes and any contigs with a BLASTN hit value higher than 1×10^{-30} were used to BLASTX search the same microsporidia used in our comparative analysis to verify that there was no obvious microsporidian content. Remaining contigs were parsed to find unknown stretches of DNA left by paired-end assembly of 5 base pairs or more and split at these points. This left a set of 1392 contigs that were further analyzed. These contigs were annotated using Artemis 13.2.0. All ORFs of 100 amino acids or more were analyzed by BLASTP and PFAM search and an annotation given on the basis of these searches [35]. Each contig was searched for the presence of tRNAs using tRNAscan-SE v.1.23 [70], which were annotated onto the contigs using Artemis.

OrthoMCL and BLAST analysis

56274 proteins sequences from eight species (predicted proteomes of *H. sapiens*, *S. cerevisiae*, *N. ceranae*, *E. bieneusi*, *E. cuniculi*, *N. parisii*, *T. hominis* and those 2499 *S. lophii* proteins without predicted frameshifts) were clustered using OrthoMCL: BLAST e-value cutoff = 1e-5, Inflation value = 1.5. Output files were parsed and sorted into categories shown in Figure 2A.

In vitro germination

300 µl of purified spores were washed 3 times in 1× PBS. A non-germinated control sample was then treated with 100 µl of 0.25 mM EGTA per 100 µg of spores to inhibit germination. The sample to be germinated was treated with 100 µl of 0.5 M Gly-Gly buffer (pH 7.0) per 100 µg of spores and incubated at room temperature for 30 minutes. 40 µl of Calcium Ionophore A23187 (1 µg/µl) (Sigma-Aldrich) dissolved in DMSO was added to the sample followed immediately by 40 µl of 0.5 M Gly-Gly buffer (pH 9.0) per 100 µg of spores. Germination was verified instantly by light microscopy and efficiency was estimated to reach a maximum of 80%.

RNA preparation for sequencing

200 µl of germinated *S. lophii* spores were frozen with liquid nitrogen and disrupted manually using a pestle and mortar. RNA was then extracted following the standard TRIzol (Invitrogen)

protocol. RNA was resuspended in 50 µl Milli-Q water (Millipore) and quantified as 2260 ng/ul (absorbance 260/280 = 2.12; 260/230 = 2.41). RNA integrity was verified on 1.5% TAE agarose gel prior to sequencing. PolyA RNA was isolated from 4 µg total RNA and the TruSeq library was prepared according to the Illumina TruSeq RNA sample preparation guide (Part # 15008136 Rev. A November 2010). 15 cycles of PCR were used to amplify the library and 5 µl of the 6 nM library was loaded onto a flow cell with two other TruSeq libraries and run on the HiSeq 2000 with a paired-end 100 bp run.

Transcriptome assembly and analysis

The raw Illumina reads obtained from RNA sequencing were filtered with fastq-mcf [71] to remove adapter sequences and low quality reads (with a quality score <28). The filtered reads were assembled using the Trinity package [43]. Reads were mapped back onto the transcriptome assembly using Bowtie [72], and the abundance of each transcript was estimated using RSEM [44].

Identification of proteins secreted into the extracellular medium during spore germination

500 µl of purified *S. lophii* spores were germinated using the protocol described previously. Germinated *S. lophii* spores along with non-germinated control were then spun at 12,000×g for 15 minutes at 4°C. The resultant supernatant was then collected and concentrated using Millipore Amicon 3 kDa centrifugal concentration column at 2,000×g for 30 minutes at 4°C. The concentrated extracellular protein was then quantified and checked on 12% SDS gels. Complex mixtures of extracellular secreted proteins from germinated and non-germinated samples were sent for analysis on an 6520 accurate mass quadrupole time of flight (Q-TOF) mass spectrometer (Agilent Technologies) and resultant masses were searched against our translated *S. lophii* genome database. The experiment was conducted in triplicate and any protein without two distinct peptide hits and a percent score peak intensity (% SPI) of ≥60% was removed from the analysis.

Supporting Information

Figure S1 A glutamate-ammonia ligase *in S. lophii*. **A**. Phylogeny of glutamate-ammonia ligase. PhyloBayes phylogenetic tree of eukaryotic glutamate-ammonia ligase proteins using the C60 empirical mixture model. Black circles show nodes with posterior probability equal to 0.99 and squares show posterior probability support of 1. **B**. Genomic context of the glutamate ammonia ligase gene. The gene is located between genes found in other microsporidian genomes. Bars and agarose gel images above show the sizes of PCR products amplified spanning the glutamate ammonia ligase gene and adjacent genes in the genome. These observations suggest that this is a *bona fide S. lophii* gene potentially acquired by horizontal transfer.
(TIF)

Figure S2 SDS-PAGE of *S. lophii* proteins secreted into the extracellular medium during germination. Protein from extracellular medium following germination was loaded onto a 12% SDS gel along with an *S. lophii* whole cell protein control. *S. lophii* hypothetical proteins (SLOPH 2344, SLOPH 477 and SLOPH 1709) were identified in the extracellular medium following band excision and mass spectrometry. SLOPH 2344 is predicted to be secreted to the extracellular environment by both SignalP and TargetP. No peptide hits were identified in the N terminal of the protein suggesting signal peptide cleavage may be responsible for the lower than predicted molecular weight visualized on SDS-PAGE.
(TIF)

Figure S3 Phylogeny of microsporidian lectin-like proteins. PhyloBayes (C20 model) tree of lectin-like proteins from representative microsporidia. Support values are provided as Bayesian posterior probabilities.
(TIF)

Figure S4 Histogram of GC content of sequenced contigs. The chosen GC cutoff point for identifying contaminant contigs is indicated by an arrow.
(TIF)

Table S1 Presence and absence of enzymes in selected metabolic pathways in our *S. lophii* genome survey and across Microsporidia. Table expanded from Keeling *et al* [21].
(XLSX)

Table S2 Classification of *S. lophii* predicted open reading frames according to OrthoMCL analysis.
(DOCX)

Table S3 List of conserved leucine rich proteins in *S. lophii* and other microsporidian genomes.
(XLSX)

Table S4 List of genes identified from RNA transcripts alone that are not included in our genomic data.
(DOCX)

Table S5 A list of proteins identified from germinated and non-germinated *S. lophii* spores.
(DOCX)

Table S6 Properties of putative *S. lophii* secreted hypothetical proteins in other microsporidian species.
(XLSX)

Table S7 Expanded gene families that have proteins identified in our germination secretome.
(DOCX)

Acknowledgments

We thank Matt Bailey at Brixham Trawler Agents for providing material and Robert Hirt for critical comments on the manuscript.

Author Contributions

Conceived and designed the experiments: SEC TAW BAPW. Performed the experiments: SEC AY BAPW. Analyzed the data: SEC TAW DMS KHP BAPW. Wrote the paper: SEC TAW BAPW.

References

1. Keeling P (2009) Five Questions about Microsporidia. PLoS Pathog 5: e1000489.
2. Singh T, Bhat MM, Khan MA (2012) Microsporidiosis in the Silkworm, *Bombyx mori* L. (Lepidoptera: Bombycidae). Pertanika Journal of Tropical Agricultural Science 35: 387–406.
3. Speare DJ, Lovy J (2011) *Loma salmonae* related species. In: Woo PTK, Buchmann K, editors. Fish Parasites: Pathobiology and Protection. UK: CABI Publishing.
4. Schottelius J, Schmetz C, Kock NP, Schuler T, Sobottka I, et al. (2000) Presentation by scanning electron microscopy of the life cycle of microsporidia of the genus *Encephalitozoon*. Microbes Infect 2: 1401–1406.
5. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414: 450–453.
6. Corradi N, Pombert JF, 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.
7. Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ (2009) Draft genome sequence of the Daphnia pathogen *Octosporea bayeri*: insights into the gene content

of a large microsporidian genome and a model for host-parasite interactions. Genome Biol 10: R106.
8. Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, et al. (2012) The Genome of the Obligate Intracellular Parasite *Trachipleistophora hominis*: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. PLoS Pathog 8: e1002979.
9. Williams BA, Lee RC, Becnel JJ, Weiss LM, Fast NM, et al. (2008) Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. BMC Genomics 9: 200.
10. Krylov DM, Wolf YI, Rogozin IB, Koonin EV (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res 13: 2229–2235.
11. Döflein F (1898) Studien zur Naturgeschichte der Protozoen. III. Über die Myxosporidien. Zool Jahrb Abt Anat 11: 281–350.
12. Weidner E, Byrd W (1982) The microsporidian spore invasion tube. II. Role of calcium in the activation of invasion tube discharge. J Cell Biol 93: 970–975.
13. Freeman MA, Yokoyama H, Osada A, Yoshida T, Yamanobe A, et al. (2011) *Spraguea* (Microsporida: Spraguidae) infections in the nervous system of the Japanese anglerfish, *Lophius litulon* (Jordan), with comments on transmission routes and host pathology. J Fish Dis 34: 445–452.
14. Cañás L, Sampedro MP, Fariña AC. (2010) Influence of host biological features on macroparasites of the two European anglerfish species, *Lophius piscatorius* and *Lophius budegassa*, off north and northwest Spain. J Parasitol 96:191–193.
15. Putz RE, Hoffman GL, Dunbar CE (1965) 2 New Species of Plistophora (Microsporides) from North American Fish with a Synopsis of Microsporidea of Freshwater and Euryhaline Fishes. Journal of Protozoology 12: 228–236.
16. Hinkle G, Morrison HG, Sogin ML (1997) Genes coding for reverse transcriptase, DNA-directed RNA polymerase, and chitin synthase from the microsporidian *Spraguea lophii*. Biol Bull 193: 250–251.
17. Mansour L, Cheikali C, Desaunais P, Coulon JP, Daubin J, et al. (2004) Description of an ultrathin multiwire proportional chamber-based detector and application to the characterization of the *Spraguea lophii* (Microspora) two-dimensional genome fingerprint. Electrophoresis 25: 3365–3377.
18. Biderre C, Pages M, Metenier G, David D, Bata J, et al. (1994) On small genomes in eukaryotic organisms: molecular karyotypes of two microsporidian species (Protozoa) parasites of vertebrates. C R Acad Sci III 317: 399–404.
19. Mansour L, Ben Hassine OK, Vivares CP, Cornillot E (2013) *Spraguea lophii* (Microsporidia) parasite of the teleost fish, *Lophius piscatorius* from Tunisian coasts: Evidence for an extensive chromosome length polymorphism. Parasitol Int 62: 66–74.
20. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178–2189.
21. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, et al. (2010) The Reduced Genome of the Parasitic Microsporidian *Enterocytozoon bieneusi* Lacks Genes for Core Carbon Metabolism. Genome Biology and Evolution 2: 304–309.
22. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, et al. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res 12: 2478–2488.
23. Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, et al. (2009) Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. PLoS Pathog 5: e1000466.
24. Akiyoshi DE, Morrison HG, Lei S, Feng X, Zhang Q, et al. (2009) Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. PLoS Pathog 5: e1000261.
25. El Alaoui H, Bata J, Bauchart D, Dore JC, Vivares CP (2001) Lipids of three microsporidian species and multivariate analysis of the host-parasite relationship. J Parasitol 87: 554–559.
26. Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc Natl Acad Sci U S A 97: 13172–13177.
27. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, et al. (2009) RNAi in budding yeast. Science 326: 544–550.
28. Eisenberg D, Gill HS, Pfluegl GM, Rotstein SH (2000) Structure-function relationships of glutamine synthetases. Biochim Biophys Acta 1477: 122–145.
29. Suarez I, Bodega G, Fernandez B (2002) Glutamine synthetase in brain: effect of ammonia. Neurochem Int 41: 123–142.
30. Huss HH (1995) Quality and quality changes in fresh fish. Food and agriculture organization of the united nations.
31. Phadtare S, Alsina J, Inouye M (1999) Cold-shock response and cold-shock proteins. Curr Opin Microbiol 2: 175–180.
32. Soanes DM, Talbot NJ (2010) Comparative genome analysis reveals an absence of leucine-rich repeat pattern-recognition receptor proteins in the kingdom Fungi. PLoS One 5: e12725.
33. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, et al. (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459: 657–662.
34. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28–36.
35. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290–301.
36. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305: 567–580.
37. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953–71.
38. Kobe B, Deisenhofer J (1995) A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374: 183–186.
39. Berthold J, Schenkova K, Ramos S, Miura Y, Furukawa M, et al. (2008) Characterization of RhoBTB-dependent Cul3 ubiquitin ligase complexes–evidence for an autoregulatory mechanism. Exp Cell Res 314: 3453–3465.
40. Deitsch KW, Lukehart SA, Stringer JR (2009) Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol 7: 493–503.
41. Marcello L, Barry JD (2007) Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res 17: 1344–1352.
42. Rauta PR, Nayak B, Das S (2012) Immune system and immune responses in fish and their role in comparative immunity study: a model for higher organisms. Immunol Lett 148: 23–33.
43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652.
44. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
46. 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: 388–93.
47. Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ (2005) A high frequency of overlapping gene expression in compacted eukaryotic genomes. Proc Natl Acad Sci U S A 102: 10936–10941.
48. Keeling PJ, Slamovits CH (2004) Simplicity and complexity of microsporidian genomes. Eukaryot Cell 3: 1363–1369.
49. Lee RCH, Gill EE, Roy SW, Fast NM (2010) Constrained Intron Structures in a Microsporidian. Molecular Biology and Evolution 27: 1979–1982.
50. Grisdale C, Bowers L, Didier E, Fast N (2013) Transcriptome analysis of the parasite *Encephalitozoon cuniculi*: an in-depth examination of pre-mRNA splicing in a reduced eukaryote. BMC Genomics 14: 207.
51. Pleshinger J, Weidner E (1985) The microsporidian spore invasion tube. IV. Discharge activation begins with pH-triggered Ca2+ influx. J Cell Biol 100: 1834–1838.
52. Frixione E, Ruiz L, Undeen AH (1994) Monovalent cations induce microsporidian spore germination in vitro. J Eukaryot Microbiol 41: 464–468.
53. Frixione E, Ruiz L, Cerbon J, Undeen AH (1997) Germination of *Nosema algerae* (Microspora) spores: conditional inhibition by D_2O, ethanol and Hg^{2+} suggests dependence of water influx upon membrane hydration and specific transmembrane pathways. J Eukaryot Microbiol 44: 109–116.
54. Dolgikh VV, Senderskiy IV, Pavlova OA, Naumov AM, Beznoussenko GV (2011) Immunolocalization of an alternative respiratory chain in *Antonospora* (*Paranosema*) *locustae* spores: mitosomes retain their role in microsporidial energy metabolism. Eukaryot Cell 10: 588–593.
55. Brosson D, Kuhn L, Prensier G, Vivares CP, Texier C (2005) The putative chitin deacetylase of *Encephalitozoon cuniculi*: a surface protein implicated in microsporidian spore-wall formation. FEMS Microbiol Lett 247: 81–90.
56. Brosson D, Kuhn L, Delbac F, Garin J, P. Vivarès C, et al. (2006) Proteomic analysis of the eukaryotic parasite *Encephalitozoon cuniculi* (microsporidia): a reference map for proteins expressed in late sporogonial stages. PROTEOMICS 6: 3625–3635.
57. Urch JE, Hurtado-Guerrero R, Brosson D, Liu Z, Eijsink VG, et al. (2009) Structural and functional characterization of a putative polysaccharide deacetylase of the human parasite *Encephalitozoon cuniculi*. Protein Sci 18: 1197–1209.
58. Loukas A, Maizels RM (2000) Helminth C-type lectins and host-parasite interactions. Parasitol Today 16: 333–339.
59. Petri WA, Jr., Haque R, Mann BJ (2002) The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. Annu Rev Microbiol 56: 39–64.
60. Kamper J, Kahmann R, Bolker M, Ma LJ, Brefort T, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444: 97–101.
61. Schornack S, Huitema E, Cano LM, Bozkurt TO, Oliva R, et al. (2009) Ten things to know about oomycete effectors. Mol Plant Pathol 10: 795–803.
62. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, et al. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome Biol 7: R12.
63. Noel CJ, Diaz N, Sicheritz-Ponten T, Safarikova L, Tachezy J, et al. (2010) *Trichomonas vaginalis* vast BspA-like gene family: evidence for functional diversity from structural organisation and transcriptomics. BMC Genomics 11: 99.
64. Abramyan J, Stajich JE (2012) Species-specific chitin-binding module 18 expansion in the amphibian pathogen *Batrachochytrium dendrobatidis*. MBio 3: e00150–00112.
65. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, et al. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330: 1543–1546.

66. Dia N, Lavie L, Metenier G, Toguebaye BS, Vivares CP, et al. (2007) InterB multigenic family, a gene repertoire associated with subterminal chromosome regions of *Encephalitozoon cuniculi* and conserved in several human-infecting microsporidian species. Curr Genet 51: 171–186.
67. Fankhauser N, Nguyen-Ha TM, Adler J, Maser P (2007) Surface antigens and potential virulence factors from parasites detected by comparative genomics of perfect amino acid repeats. Proteome Sci 5: 20.
68. Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, et al. (2010) Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A 107: 17421–17426.
69. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.
70. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33: W686–689.
71. Aronesty E (2011) Command-line tools for processing biological sequencing data. ea-utils http://code.google.com/p/ea-utils: Expression Analysis, Durham, NC
72. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.