Wiredu Boakye, D., Jaroenlak, P., Prachumwat, A., Williams, T., Bateman, K. S., Itsathitphaisarn, O., ... Williams, B. A. P. (2017). Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae microsporidia. *Environmental Microbiology*, *19*(5), 2077–2089. https://doi.org/10.1111/1462-2920.13734

Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae microsporidia

Dominic Wiredu Boakye[1], Pattana Jaroenlak[2,3], Anuphap Prachumwat[4], Tom A. Williams[5], Kelly S. Bateman[6], Ornchuma Itsathitphaisarn[2,3], Kallaya Sritunyalucksana[4,], Konrad H. Paszkiewicz[1], Karen A. Moore[1], Grant D. Stentiford[6*] , Bryony A. P. Williams[1*]

[1] Biosciences, College of Life and Environmental Sciences, University of Exeter, EX4 4QD, UK.

[2] Department of Biochemistry, Faculty of Science, Mahidol University, Rama VI Rd., Bangkok 10400, Thailand

[3] Center of Excellence for Shrimp Molecular Biology and Biotechnology, Faculty of Science, Mahidol University, Rama VI Rd., Bangkok 10400, Thailand

[4] Shrimp–Virus Interaction Laboratory (ASVI), National Center for Genetic Engineering and Biotechnology (BIOTEC), Rama VI Rd., Bangkok 10400, Thailand

[5] School of Earth Sciences, University of Bristol, BS8 1TH, UK

[6] European Union Reference Laboratory for Crustacean Diseases, Centre for Environment Fisheries and Aquaculture Science, Weymouth Laboratory, Weymouth, Dorset DT4 8UB, UK

[*]Correspondence should be addressed to b.a.p.williams@exeter.ac.uk or grant.stentiford@cefas.co.uk

Originality-Significance Statement: This study uses genomics to explore metabolic reduction of microsporidia in the highly economically important Enterocytozoonidae family, showing that uniquely amongst eukaryotes, this whole clade has lost any clear intrinsic means of generating ATP. It also presents the first genome sequence of an obligate intranuclear eukaryotic parasite and the first sequence for the shrimp pathogen EHP, which is crucial to underpinning the development of new tools to diagnose, monitor and potentially mitigate the negative impacts of this major emerging aquacultural pathogen.

Summary:

Glycolysis and oxidative phosphorylation are the fundamental pathways of ATP generation in eukaryotes. Yet in microsporidia, endoparasitic fungi living at the limits of cellular streamlining, oxidative phosphorylation has been lost: energy is obtained directly from the host or, during the dispersive spore stage, via glycolysis. It was therefore surprising when the first genome from the Enterocytozoonidae – a major family of human and animal-infecting microsporidians - appeared to have lost glycolysis. Here we sequence and analyse genomes from four additional members of this family, shedding new light on their unusual biology. Our survey includes the genome of *Enterocytozoon hepatopenaei*, a major aquacultural parasite currently causing substantial economic losses in shrimp farming, and *Enterospora canceri*, a pathogen that lives exclusively inside the nuclei of its crab host. Our analysis of gene content across the clade suggests that *Ent. canceri*'s adaptation to intranuclear life is underpinned by the expansion of transporter families. We demonstrate that this entire lineage of pathogens has lost glycolysis and, uniquely amongst eukaryotes, lacks any obvious intrinsic means of generating energy. Our study provides an important resource for the investigation of host-pathogen interactions and reductive evolution in one of the most medically and economically important microsporidian lineages.

Introduction:

The microsporidia are important emergent pathogens of almost all known animal phyla in all major biomes (Stentiford et al., 2016). The Enterocytozoonidae family is home to two of the most economically important microsporidian species. *Enterocytozoon bieneusi* is the most prevalent human infecting microsporidian with some studies reporting prevalence of 58% in immunocompetent individuals. *Enterocytozoon hepatopenaei* is a recently emerged pathogen of the farmed shrimp species *Penaeus monodon* and *Penaeus vannamei* and severely retards the growth of the shrimp and has rapidly spread across south-east Asia (Newman, 2015). Microsporidia are also pre-eminent model systems for understanding the processes of metabolic, cellular and genomic reduction in eukaryotes, with genomes as small as 2.3 Mb and as few as 1,990 encoded genes (Corradi et al., 2010). A large set of core eukaryotic genes were jettisoned early in the evolutionary history of microsporidia (Nakjang et al., 2013). This left them without a mitochondrial genome and without the ability to generate ATP via oxidative phosphorylation (Williams et al., 2002; Nakjang et al., 2013), whilst their close relative *Mitosporidium daphniae* retains a mitochondrial genome and components of the electron transport chain (Haag et al., 2014). With the absence of oxidative phosphorylation in microsporidia, glucose is only partially metabolised via glycolysis to release 7 % of its full ATP potential (Berg et al., 2007). However, this energy supply is complemented by the import of ATP molecules from the host with the help of horizontally acquired ATP/ATP translocases (Richards et al., 2003; Tsaousis et al., 2008; Nakjang et al., 2013; Hacker et al., 2014). Thus, microsporidians are considered to be entirely reliant on glycolysis and ATP import from their hosts for energy (Heinz et al., 2012; Haag et al., 2014). One emerging idea is that glycolysis occurs primarily in the spore stage but shuts down in intracellular life stages, where the parasite can tap into the host's energy pools (Heinz et al., 2012). This is based on data that demonstrate that the enzyme responsible for the first ATP-forming reaction in glycolysis is enriched in the cytosol of the extracellular spores in comparison to the cytosol of the intracellular meront stages of *Trachipleistophora hominis* (Heinz et al., 2012). Moreover, the two enzymes responsible for the reoxidation of NADH for glycolysis (Alternative oxidase

and glycerol-3-phosphate dehydrogenase) were found to accumulate in the cytosol of *Antonospora locustae* spores in a study by Dolgikh *et al.*, 2011.

When the genome of the human infecting microsporidian *E. bieneusi* was deeply surveyed using whole-genome shotgun sequencing, there was a surprising revelation that most genes involved in glycolysis were absent, in addition to the loss of those involved in the oxidative phosphorylation pathway (Akiyoshi et al., 2009; Keeling et al., 2010). This implies that this parasite relies entirely on ATP import from its host with no obvious means of generating ATP, even in the extracellular spore stage (Akiyoshi et al., 2009; Keeling et al., 2010).

As organisms with highly reduced metabolisms, microsporidians are highly dependent on their hosts and have an expanded repertoire of transport proteins that allow them to exploit the rich environment in the host cell cytoplasm. Interestingly, meronts of *E. bieneusi* form tight physical associations with the host mitochondria, and this, in other microsporidian species, has been shown to maximise the host ATP-harvesting efficiency of the parasite (Desportes et al., 1985; Hacker et al., 2014). However, this association of microsporidian life stages with host mitochondria is not universal across the group (Hacker et al., 2014). In fact, some of the aquatic host-infecting microsporidian species belonging to the Enterocytozoonidae family inhabit the host nucleus rather than the cytoplasm. These species include *Enterospora canceri*, *Enterospora nucleophila*, several members of the genus *Nucelospora*, and, *Desmozoon lepeophtherii (=Paranucleospora theridion)* (Chilmonczyk et al., 1991; Hedrick et al., 1991; Lom and Dykova, 2002; Stentiford and Bateman, 2007; Stentiford et al., 2007; Freeman and Sommerville, 2009; Nylund et al., 2010; Freeman et al., 2013). These species have no access to the host mitochondria so cannot use this strategy to maximise ATP uptake.

Despite the economic and evolutionary importance of the Enterocytozoonidae, at present, *E. bieneusi* is the only member of the group with a sequenced genome. This leaves several intriguing unanswered questions about this group. Firstly, what drove these taxa to inhabit the

host nucleus and, what are the metabolic benefits and parasite adaptations to life in this unusual environment? Secondly, when was glycolysis was lost in the microsporidia and does the absence of this pathway in *E. bieneusi* represent an anomaly within the phylum?

To better understand this intriguing parasite family we sequenced, annotated and analysed the genomes of several members of the Enterocytozoonidae. These include the intranuclear crab parasite *Enterospora canceri*, the cytoplasmic crab parasites *Hepatospora eriocheir* and *Hepatospora eriocheir canceri,* and, the cytoplasmic shrimp parasite *Enterocytozoon hepatopenaei.* The latter is of particular interest given its recent emergence and significant economic impact on the global shrimp farming industry (Rajendran et al., 2016; Stentiford et al., 2016). This pathogen is reported to have reached epidemic levels in shrimp farms across South East Asia, where the infection results in slow shrimp growth (Thitamadee et al., 2016). For this reason, this pathogen is currently a major threat to sustainable shrimp farming in Asia (Thitamadee et al., 2016). Thus, our data provide a resource for understanding the metabolic strategy for intranuclear existence, as well as a foundation for investigating host-parasite interactions (and potential intervention strategies) for economically important microsporidians such as *E. hepatopenaei*. Additionally, our genome data reveals that the genes that encode glycolysis show a pattern of parallel decay across the Enterocytozoonidae clade: All of these species have incomplete glycolytic pathways, but different genes have been lost in each species, implying a common loss of the selective pressure to retain glycolysis in these diverse species of microsporidia.

Results and Discussion:

Establishing the interrelationships of the Enterocytozoonidae using a multi protein phylogeny: In order to investigate the phylogenetic relationships between our studied species within the Enterocytozoonidae, we generated a 21-protein concatenated phylogeny. This demonstrated the close relationships of the crustacean parasites *Ent. canceri*, and *E.*

hepatopenaei to the human pathogen *E. bieneusi*, with *Hepatospora* forming an outgroup to these (Figure 1 + 2 and Supplementary figure 1). This phylogeny implies that there has been a rapid change in lifestyle within the group with drastic switches in both cellular localisation (*E. hepatopenaei* and *E. bieneusi* are cytoplasmic, while *Ent. canceri*, is nuclear) and, host type (*E. hepatopenaei* and *Ent. canceri* in aquatic crustaceans, and *E. bieneusi* in humans, mammals and birds) over a relatively short evolutionary timescale. The finding that *E. bieneusi* is phylogenetically nested amongst microsporidians with aquatic hosts suggests that this species has emerged to infect human and terrestrial animal hosts via zoonotic transfer from an aquatic environment.

Parallel decay of the glycolysis and other metabolic pathways across the Enterocytozoonidae: The discovery that *E. bieneusi* lacked a functional glycolytic pathway opened up a number of questions about how these organisms fuel their complex lifecycle (Akiyoshi et al., 2009; Keeling et al., 2010). One of these was whether the loss was unique to this species or had occurred more broadly within the Microsporidia. Firstly we took a conserved protein set identified by Keeling et al. (Keeling et al., 2010). We used OrthoMCL to cluster all publically available microsporidia predicted open reading frames and clusters representing these conserved gene sets were used to populate supplementary table 1. In a second approach, the same conserved protein set was used to interrogate our predicted ORFs and our raw contigs for the Enterocytozoonidae using BLASTP and TBLASTN. This revealed that that all surveyed representatives of the Enterocytozoonidae lacked multiple components of the glycolytic pathway. We find 4 glycolytic enzymes retained in *Ent. canceri*, 2 in *E hepatopenaei* and 4 in each of the *H. eriocheir* genomes. Although none of these genomes were sequenced to completion, coverage of our genomes is high (Table 1) and it is unlikely that our observations can be explained by a failure to sequence the missing components of glycolysis. Using the presence or absence of a set of 44 broadly conserved genes (transcriptional control proteins) to gauge genome completeness, we estimate that our assemblies are 84-100% complete [*H. eriocheir canceri* (40/44), *H. eriocheir* (40/44), *Ent.*

canceri (42/44)*, E. hepatopaenei* (43/44)] (Supplementary table 1). Assuming a genome completeness of between 84% and 100%, we calculate the probabilities of observing at least this many glycolytic losses simply as the result of incomplete sequencing (rather than true absences) to be 2.869×10^{-08} for *Ent. canceri,* 6.975×10^{-10} for *E. hepatopenaei,* 8.629×10^{-05} for *H. eriocheir canceri* and 8.629×10^{-05} for *H. eriocheir*. In addition, the genomes from our two independently sequenced and assembled *Hepatospora eriocheir* sub-species share the same pattern of glycolytic enzyme loss corroborating our assumptions that these gene absences are true losses and not a consequence of incomplete sequencing. Overall, the data suggest a common absence of a functional glycolytic pathway in the Enterocytozoonidae. Our data also show that this loss did not occur as a single ancestral event but highlights a pattern of parallel decay of the pathway with different genes lost in each lineage (Figure 2). This suggests a relaxation of the selective pressure to keep glycolysis at the base of the Enterocytozoonidae followed by pathway decay via loss of different enzymes. What is common to each species is that one or more genes encoding a hexokinase-like protein are retained, along with genes coding for a handful of other components. However and crucially, phophoglycerate kinase and pyruvate kinase , enzymes that catalyze the ATP-generating steps of glycolysis, are consistently missing.

The conservation of ATP-expensive pyrimidine and purine synthesis pathways (Figure 3) together with the absence of glycolytic enzymes involved in ATP-generating steps and the retention of seemingly functional hexokinases (Figure 4) suggests that the cytosol of these Enterocytozoonidae lineages represent a potential ATP sink. One consequence of this is an increase in ADP/ATP ratio within the cytosol, which has been shown in *Chlamydiae* to stimulate ATP uptake via ATP/ADP translocases (Trentmann et al., 2007). Thus, a high ADP/ATP ratio in the microsporidian meront potentially generated by the broken glycolytic pathway and ATP-intensive pathways such as pyrimidine and purine synthesis pathways may be responsible for stimulating ATP import into the microsporidia via ATP/ADP translocases.

Other core metabolic enzymes were identified and mapped onto the multi-protein phylogeny. Whilst individual microsporidian lineages have lost enzymes of different metabolic pathways independently, the Enterocytozoonidae have undergone the most dramatic metabolic reduction described to date. In addition to the loss of most glycolytic enzymes, they have also lost the pentose phosphate pathway and most genes involved in fatty acid and trehalose metabolism (Figures 1 & 2).

The nature of microsporidian hexokinases in non-glycolytic lineages: Within the microsporidian phylum hexokinase is found either in single or multiple copies. Based on sequence similarity, we identified hexokinase-like genes in the Enterocytozoonidae that were divergent in comparison to other microsporidian hexokinases. This may reflect the fact that they may not be functioning in the same pathway as those present in glycolytic microsporidia. As seen previously in other microsporidia (Nakjang et al., 2013), our data shows a pattern of independent gene duplications of these hexokinases across the Enterocytozoonidae, including multiple genes in *E. bieneusi,* single gene copies in *Ent. canceri* and *E. hepatopenaei* and a duplication event at the base of the *Hepatospora* lineage (Figure 4). An attractive explanation for the retention and expansion of the hexokinase gene family in these seemingly non-glycolytic organisms is that the redundant copy is secreted to modify host metabolism and increase metabolic rewards for the pathogen as has been previously hypothesized to occur in other glycolytic microsporidia (Cuomo et al., 2012). However, no signal sequences that would typically direct these proteins to the secretory pathway were detected in any of the Enterocytozoonidae hexokinase proteins using either SignalP or WOLFPSORT (Horton et al., 2007; Petersen et al., 2011).

Interestingly, a protein tyrosine phosphatase A (PTPA) domain was found to be fused to the N-terminus of the single copy of the *Ent. canceri* divergent hexokinase, the significance of which is unclear. In other organisms, these tyrosine phosphatases perform regulatory roles by

dephosphorylating specific tyrosine residues (Tonks, 2006). One possibility is that the *Ent. canceri* hexokinase has acquired a similar regulatory role.

The divergent nature of the gene sequences of these hexokinases suggested that they might no longer be functioning as typical hexokinases. To investigate this, we compared amino acid sequence conservation across known active sites between the new members of the Enterocytozoonidae sequenced in this study, hexokinases from published microsporidian genomes and the hexokinase 1 (HXK1) of *Saccharomyces cerevisiae*. In cases where a genome of a species encodes multiple hexokinase homologues, the additional hexokinase copies had lost some of the active sites identified in *S. cerevisiae* HXK1. Similarly, amino acid residues known to be important for the function of the *S. cerevisiae* HXK1 (Kuser et al., 2008) were frequently substituted or even deleted (Figure 4). As such, it is possible that, for these genomes, duplicate hexokinases may be non-functional. Taken together, these observations suggest that at least one hexokinase copy in each microsporidian genome, including members of the Enterocytozoonidae retained its active sites indicating that they may be capable of performing a conserved hexokinase function.

.

Inferring the role of the conserved hexokinase gene in the metabolically reduced genomes of the Enterocytozoonidae: Hexokinase catalyzes phosphorylation of glucose to glucose-6-phosphate (G6P). G6P is not recognized by glucose transporters , but in model organisms is channeled into one of at least four metabolic pathways: glycolysis, dolichyl-P-mannose synthesis, trehalose synthesis and the pentose phosphate pathway (Figure 3) (Mueckler, 1994).

Our gene content analysis demonstrates that out of these four metabolic pathways, only the dolichyl-P-mannose (DPM) synthesis pathway has a complete repertoire of enzymes (See figure 3) across representatives of the Enterocytozoonidae (See figure 3). The retention of the

dolichyl-P-mannose synthesis pathway in these metabolically reduced lineages could be due its important role in the production of building blocks for biosynthesis and protein glycosylation (Wujek et al., 2004; Shental-Bechor and Levy, 2008; Vagin et al., 2009; Moharir et al., 2013).

In the pentose phosphate pathway, only glucose-6-phosphate dehydrogenase (G6PD) and hexokinase were retained across all members of the Enterocytozoonidae, suggesting an alternative role for the coupled reaction catalyzed by these two enzymes. In this context, it is interesting to note that the reducing equivalent, NADPH, released during the conversion of glucose-6-phosphate to 6-phosphogluconolate by G6PD is key for lipid synthesis and scavenging harmful reactive oxygen species (ROS). As intracellular parasites, microsporidia are subject to challenge by host defense mechanisms including ROS (Leiro et al., 2001). We hypothesize that hexokinase and G6PD may be retained, in part, due to their role in protecting the parasite against damage from host-derived ROS. Retention of the ROS pathway in all analyzed species in this study is consistent with this hypothesis (Figure 3).

Absence of an intrinsic ATP generating mechanism in the Enterocytozoonidae: The extreme metabolic reduction observed in these *Enterocytozoonidae* lineages suggest they are completely reliant on horizontally acquired ATP/ADP translocases that are universal in the Microsporidia for ATP acquisition from the host cell (Haag et al., 2014). These would provide ATP for the Enterocytozoonidae merogonial stages, but it does not explain how these lineages acquire or generate ATP during their extracellular spore stage for the presumably energetic process of spore germination (Vavra and Lukes, 2013).

To explore the possibility that microsporidia have retained alternative ATP generation pathways, we used BLAST searches to look for enzymes involved in alternative exergonic metabolic pathways such as mixed acid fermentation, the Entner-Doudoroff pathway, and

lipid and protein oxidation. However, none of the microsporidian genomes surveyed had the hallmark enzymes of these pathways either. In addition, each of these pathways feeds into the citric acid cycle, which is absent in the Microsporidia.

Phagocytosis has been observed as an alternative form of host cell invasion by some microsporidia (Franzen et al., 2005). There is also evidence to suggest that host cell types parasitized by the presently studied Enterocytozoonidae lineages (hepatopancreatic epithelial cells and mammalian enterocytes), have phagocytic properties (Alday-Sanz et al., 2002; Neal et al., 2006; Bu et al., 2010; Liu et al., 2013). It is therefore possible that the spores of the Enerocytozoonidae lineages could invade their respective host cells by phagocytosis. This would bring the microsporidian into contact with the host ATP supply and if ATP/ADP translocases were expressed at this time, they could provide the spore with the ATP needed to germinate and to escape from the phagocytic vesicle. As mentioned, above, a truncated glycolytic pathway could potentially stimulate ATP uptake in the microsporidian by generating a high ADP/ATP ratio. In the case of *Ent. canceri*, entry into the host cell by phagocytosis would need to be followed by a second step in which the parasite gains access to the nucleoplasm, where it completes its lifecycle. One possibility is that *Ent. canceri* could enter the nucleus by means of polar tube extrusion following germination in the cytoplasm. It is worth noting that members of the Cryptomycota (Syn. Rozellomycota), a related group of parasitic fungi, are able to gain access to the nucleoplasm without a polar tube (Michel et al., 2000; Michel et al., 2009; Corsaro et al., 2014) although the mechanism by which they do so remains unknown. Phagocytosis is typically mediated by pathogen ligand binding to phagocytic receptor. If this is the main means of host cell invasion by these Enterocytozoonidae lineages, then identifying and blocking the appropriate ligands may inhibit pathogen entry into the host cell and provide a means of controlling infection.

Comparing the plasma membrane transporter repertoire of nuclear and cytoplasm-infecting microsporidians: To investigate the adaptation of *Ent. canceri* to its unique

intranuclear niche, we compared the predicted complement of membrane transporters across the Enterocytozoonidae to determine whether the intranuclear *Ent. canceri* genome encoded a different transporter repertoire to that found in sister genera inhabiting the cytoplasm. Interestingly, the Enterocytozoonidae as a whole encode fewer homologues of characterised membrane transport proteins than other sequenced microsporidia (median 23, compared to 46 and 48 in the outgroups *Encephalitozoon cuniculi* and *Trachipleistophora hominis*, respectively), suggesting a reduction in transport repertoire in their common ancestor (Supplementary table 2). However, the branch leading to *Ent. canceri* experienced a substantial gain in characterised membrane transporters, encoding a number (45 transporters) similar to that of other microsporidians. These lineage-specific gains included additional representatives of the ABC, lipid (P-type ATPase), choline (CTL), amino acid, nucleotide sugar (UAA) and cation transporter families, raising the possibility that, against a background of extreme genome reduction, this increased transporter repertoire is associated with adaptation to the intranuclear environment.

The predicted secretomes of the Enterocytozoonidae: A key question related to members of the Enterocytozoonidae is how these closely-related organisms, with vastly different parasitic habits and host preferences, interact with their hosts. To screen for proteins secreted by these parasites into their host cells, we screened each of the predicted proteomes with SignalP 4.0 and identified those proteins with appropriate signal sequences to allow for translocation to the host cell cytoplasm (or nucleoplasm) (Petersen et al., 2011). We filtered out those proteins with multiple transmembrane domains that may be anchored into parasite membranes. Although this list is likely to include some proteins more generally associated with the secretory pathway or cell wall biosynthesis, we also expected to find potential effectors. Whilst dominated by hypothetical proteins, the list also included proteins with conserved domains allowing prediction of their function (Supplementary table 3). Predicted secreted proteins present in more than one Enterocytozoonidae genus, and thus representing candidate effectors include putative peptidyl-prolyl cis-trans isomerases and

glycosyltransferases, both protein families that have been identified as effectors in other microbes (Unal and Steinert, 2014; Lu et al., 2015).

This study has provided a major new resource for the investigation of host-pathogen interactions, phylogeny, genetic reduction and pathogen evolution in arguably the most economically and evolutionarily important families within the microsporidian phylum (Stentiford et al., 2016). We provide the first genome sequence for the major yield-limiting shrimp pathogen *E. hepatopenaei*. This genomic data in conjunction with previously-published information on the parasite's transmission cycle (Tangprasittipap et al., 2012; Tangprasittipap et al., 2013), a growing body of genomic data for host shrimp species (Baranski et al., 2014, You et al., 2010; Castillo-Juarez et al., 2015; Yu et al., 2015) and the capacity to study this pathogen in the shrimp host (Salachan et al., 2017) will undoubtedly underpin the development of new tools to diagnose, monitor and potentially mitigate the negative impacts of *E. hepatopenaei*. As a whole our data provide resounding evidence of the consistent loss of glycolysis into the Enterocytozoonidae. We demonstrate that glycolytic enzymes were not lost in a single event in the ancestor of the group but rather there has been a common loss of the selective pressure to retain glycolysis followed by a parallel erosion of the pathway by differential loss of the enzymes across the members of the genus. Microsporidia are already primary models of metabolic reduction in eukaryotes and the Enterocytozoonidae take this reduction even further making them the only eukaryotic group to have eliminated all canonical ATP-generating pathways. Whilst there are precedents for loss of glycolysis amongst prokaryotes that form very close host associations, for example *Mycoplasma hominis* and *Nanoarchaeum equitans,* (Waters et al., 2003; Pereyre et al., 2009), these organisms do not have an extracellular spore stage. Loss of glycolysis makes the Enterocytozoonidae unique amongst eukaryotes and leaves the enigma of how they complete their life cycle, surviving out of association with the host and activating their spores to invade host cells in the absence of their own energy generation system.

Materials and Methods

Purification of *Enterospora canceri* and *Hepatospora eriocheir canceri* spores: European edible crab adults (*Cancer pagurus*) were purchased from local fishermen in Weymouth, UK (50°34'N, 2°22'W) in January 2013. The hepatopancreases isolated from crabs infected with either *Ent. canceri* or *H. eriocheir canceri* were crushed with a sterile pestle and mortar in 1 x PBS. The homogenous mash was then filtered through a 100 μm mesh followed by cell sieving through 40 μm filter and the filtrate was further purified using a Percoll density gradient.

Purification of *Enterocytozoon hepatopenaei* spores: *E. hepatopenaei* infected shrimp *Penaeus vannamei* were collected from commercial shrimp ponds in Thailand in January 2015. The size of the shrimp varied from 10-15 grams. The hepatopancreases were dissected out, homogenised with a sterile ground-glass pestle, and filtered through 100 μm and 40 μm cell strainers respectively to remove tissue debris. The filtered suspension was further purified by a Percoll density gradient.

Purification of *Hepatospora eriocheir* spores: The hepatopancreas of an infected Chinese mitten crab was passed through a 200 μm mesh and the filtrate was then again passed through a 40 μm mesh. The filtrate was incubated in a 1 x PBS -0.05% Triton X-100 solution for 1 hr. The washed and pelleted filtrate was then further purified using a Percoll density gradient.

Genomic DNA extraction for sequencing: For *H. eriocheir canceri*, *E. hepatopenaei* and *Ent. canceri*, aliquots of purified spores were ground for 10 minutes in liquid nitrogen. This step was repeated three times before resuspending the resulting powder in 800 μl of phenol before proceeding with a standard phenol chloroform extraction followed by an ethanol precipitation as per Campbell et al., 2013. For *H. eriocheir*, purified spores were subjected to bead beating followed by phenol/chloroform extraction and ethanol precipitation as per Campbell et al., 2013.

Sequencing Protocols: One *E. hepatopenaei* DNA sample was processed with NextFlex Rapid protocol without PCR amplification (BIOO Scientific) and was sequenced using the MiSeq v3 reagents 300 (PE). A second *E. hepatopenaei* DNA sample was processed for PacBio Sequencing which was performed using PacBio RS II technology of a 10 kb library on a single SMRT cell by Macrogen, South Korea. For *H. eriocheir* DNA a SPRiworks fragment library (Beckman Coulter) was generated and this was sequenced on the HiSeq 2000 100 PE. *Ent. canceri* DNA was used to generate a SPRIworks fragment library (Beckman Coulter), the first of these was sequenced using MiSeq v2 reagents 250 (PE) and the second using MiSeq v3 reagents 300 (PE). *H. eriocheir canceri* DNA was used to generate a SPriworks fragment library (Beckman Coulter) which was sequenced using the MiSeq v2 reagents 250 (PE).

Genome assembly and genome annotation: *MiSeq and HiSeq data:* FASTQC was used to identify and filter/trim reads with poor quality scores. A total of 197,235,682 reads with a maximum length of 300 bp were used to assemble the *E. hepatopenaei* genome with the Spades assembly pipeline (v.3) (Bankevich et al. 2012). A total of 222,450,433 paired-end reads with a maximum length of 99 bp were used to assemble the genome of *H. eriocheir* with the Velvet package (v1.1) (Zerbino & Birney 2008). A total of 8,740,870 paired-end Illumina reads with a maximum length of 251 bp were used to assemble the *H. eriocheir canceri* genome with the A5_miseq assembly pipeline (Coil et al. 2015). A total of 4,632,816 reads with a maximum length of 301 were used to assemble the *Ent. canceri* genome with the A5_miseq assembly pipeline (Coil et al. 2015). All reads used had a mean quality score above 32. Protein prediction for sequenced genomes was performed with the MAKER annotation package (Cantarel et al., 2008). To enable the program to make the best possible predictions for the newly assembled genomes of *Ent. canceri, E. hepatopenaei, H. eriocheir* and *H. eriocheir canceri*, the ab-initio protein prediction program incorporated within MAKER was trained on the published genome of *E. bieneusi* as per user guide.

PacBio data: Subreads for the genome of *E. hepatopenaei* were assembled *de novo* by Canu (version 1.3) (Berlin et al., 2015) with the default parameter setting

(http://canu.readthedocs.io/). Open-reading frames (ORFs), tRNAs and rRNAs were predicted using Prokka (version 1.11) (Seemann, 2014). Contaminating bacterial sequences found in the initial assembled reads were removed using a combination of both BLASTP and BLASTX searches for putative ORFs and by BLASTN searches for rRNAs against the NCBI nr and nt databases.

Identification of core metabolic genes: In order to assess how the protein repertoire encoded by the genomes of *Ent. canceri, E. hepatopenaei* and *Hepatospora* spp. sequenced in this study compared to that of other microsporidians, open reading frames (ORFs) of all four sequenced genomes and those from 19 publicly available microsporidian species were grouped into orthologous families with the command line program, OrthoMCL (Li et al., 2003) set on the following parameters: MCL inflation=1.1 and Maximum weight =100. These protein clusters were subsequently parsed for 381 core microsporidian proteins to assess completeness of newly sequenced genomes. This protein set has been previously used for a similar study by (Keeling et al., 2010). As a further verification for the presence/absence of this conserved gene set, BLASTP and TBLASTN searches were carried out on our newly predicted ORFS and newly generated genomic contigs. This revealed further predicted proteins that were not picked up by our OrthoMCL strategy or that were not predicted by our MAKER annotation.

Identification of transporter proteins from five genomes

ORFs from all four genomes sequenced in this study and from 19 publicly available microsporidian species were screened using the transmembrane domain prediction tool, TMHMM (Krogh et al., 2001) to select proteins with one or more transmembrane domains. Since the TMHMM program is known to erroneously identify signal peptides as transmembrane domains the protein set was also analysed with SIGNALP (Petersen et al., 2011). Proteins predicted to have a single transmembrane domain that was also identified to be a signal peptide were removed from the analysis. A BLASTP search with the remaining

protein set was carried out against the following databases: SGD, TCDB and NCBI databases (Saier et al., 2006; Cherry et al., 2012; Tatusova et al., 2014). Where possible, a consensus BLAST hit ID was selected for each putative transmembrane protein. WOLFPSORT (Horton et al., 2007) was subsequently used to predict a subcellular localization for these proteins. Predicted plasma membrane proteins that had four or more transmembrane domains were selected for further analysis as potential transport proteins.

For each predicted plasma membrane transporter, its OrthoMCL orthologous cluster was manually parsed to recover orthologs that may have been missed by the automated pipeline described above. The majority of the final predicted plasma membrane transporters (227/235) assigned to a functional group had four or more transmembrane domains. For predicted plasma membrane transporters that were not assigned to any functional group, only those with four or more transmembrane domains were retained. Although 18 non-Enterocytozoonidae microsporidian species were used in this analysis, only data for *Enc. cuniculi* and *T. hominis* were displayed in the final results.

Identification of secreted proteins/potential effectors

ORFs from the newly sequenced genomes and those from 19 other microsporidian species were parsed with the TMHMM command line tool (Krogh et al., 2001) to detect proteins with transmembrane domains. Proteins that were devoid of transmembrane domains or only possessed a single transmembrane domain were retained for further analyses. This protein set was passed to the signal peptide prediction program, SIGNALP (Petersen et al., 2011). Signal peptide-containing proteins were retained as potential secreted effectors. Finally, a Gene Ontology assessment of the predicted secreted proteins was performed with the graphical interphase version of BLAST2GO 3.2 (Conesa et al., 2005).

Phylogenetic analyses: Homologs for 21 universally conserved proteins: Wrs1p, Taf10p, TFIIE, Taf10p, Tfa2p, Sec62p, Abd1p, SPT16, Brn1p, Nob1p, Caf40p, Tfb2p, Bos1p, Npl4p, Tma46p, Tfa1p, Clp1p, Spt5p, Sec63p, Pri2p and Enp2p were aligned with MUSCLE

(v3.8.31) (Edgar, 2004) using the default settings and subsequently masked with the automatic command line tool TrimAl (v1.2rev59) (Capella-Gutierrez et al., 2009). The substitution model prediction option on MEGA (v6.06) (Tamura et al., 2011) was used to identify the best substitution models for the individual protein homolog sets. These models were then used to perform maximum likelihood analyses on the individual homolog protein sets using the command line program, RaxML (v7.2.7) (Stamatakis, 2014). These were pilot trees to check for unusually long-branch lengths indicative of unlikely orthologs. The individual masked alignments of the homolog protein sets were manually concatenated using Seaview (v4.5.4) (Gouy et al., 2010). A GTR+GAMMA module on RaxML (Stamatakis, 2006) was subsequently used to construct the final multi-protein concatenated phylogenetic tree.

Acknowledgements

The authors would like to thank Prof. Tim Flegel for facilitating the cooperation between Thailand and the UK and assisting with spore purification work done in Bangkok. This work was supported by Royal Society University Research Fellowships (B.A.P.W. and T.A.W.) and by a University of Exeter/Cefas strategic alliance studentship (DWB). AP and KS would like to acknowledge support from the Agricultural Research Development Agency (Arda) of Thailand (Grant number 8669), Mahidol University and the National Center for Genetic Engineering and Biotechnology (BIOTEC) of the Thai National Science and Technology Development Agency (NSTDA). OI is supported by Mahidol University. PJ is supported by Science Achievement Scholarship of Thailand (SAST).

The authors have no conflicts of interest to declare

Availability of data

The Whole Genome Shotgun projects described in this paper have been deposited at DDBJ/ENA/GenBank under the accessions LVKB00000000, LTAI00000000, LWDP00000000 and MNPJ00000000. The versions described in this paper are LVKB01000000, LTAI01000000, LWDP01000000 and MNPJ01000000.

Tables:

Table 1: Genome statistics for the newly sequenced genomes. *Gene set taken from Desjardins et al., 2015

Statistics	***Hepatospora eriocheir***	***Hepatospora eriocheir canceri***	***Enterospora canceri***	***Enterocytozoon hepatopenaei***
Assembly size (Mb)	4.57	4.84	3.10	3.26
Contigs	1300	2344	537	64
Mean coverage (X)	4477.89	63.18	288	363
Contig N50 (bp)	17583	3349	11128	125008
GC content (%)	22.44	23.16	40.15	25.45
GC content coding (%)	25.58	25.41	41.95	27.81
Coding regions (%)	42.39	40.31	59.50	71.87
Splicing machinery *	7/29 genes	7/29 genes	7/29 genes	7/29 genes
Genes	2716	3058	2179	2540

Figure legends:

Figure 1: Metabolic profiling of microsporidian genomes. Filled squares represent presence of a gene involved in a metabolic pathway whereas a blank space represents absence of a gene. The phylogenetic positions in the cladogram are derived from maximum likelihood analyses performed on a concatenated alignment of 21 conserved proteins. Values represent levels of bootstrap support (100 replicates) for the corresponding nodes.

Figure 2: Independent loss of glycolytic enzymes and retention of the hexokinase gene across the Enterocytozoonidae family. Filled circles represent the presence of a gene whereas empty circles represent the absence of a gene. The loss of different glycolytic genes in different lineages suggests that the selective pressure for the retention of glycolysis was lost at the base of clade, leading to evolutionary decay of the pathway over time.

Figure 3: Comparing the enzyme repertoire of key metabolic pathways between members of the Enterocytozoonidae family and other microsporidians (Figure adapted from Nakjang et al., 2013. Thick black arrows represent those pathways found in all Enterocytozoonidae. Where there is differential loss of the pathway across the Enterocytozoonidae, presence of a gene in each species is represented by a different coloured arrow. Predicted number of plasma membrane transporter families and their respective substrates are also colour coded for each species of the Enterocytozoonidae.

Figure 4: **Phylogeny of hexokinase reveals that it is duplicated in some microsporidian lineages.** The phylogenetic positions are derived from maximum likelihood analyses performed on a masked alignment of hexokinase proteins. Mammalian hexokinases were used to root the tree. Bootstrap support values that are above 95% (100 replicates) are represented by black dots on the respective nodes. Similarity between *S. cerevisiae*'s hexokinase active sites and those of other lineages are represented with a heat map.

References:

Akiyoshi, D.E., Morrison, H.G., Lei, S., Feng, X., Zhang, Q., Corradi, N. et al. (2009) Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. *PLOS Pathog* **5**: e1000261.

1 Alday-Sanz, V., Roque, A., and Turnbull, J.F. (2002) Clearing mechanisms of *Vibrio*
2 *vulnificu*s biotype I in the black tiger shrimp *Penaeus monodon*. *Dis Aquat Organ* **48**: 91-99.
3 Baranski, M., Gopikrishna, G., Robinson, N.A., Katneni, V.K., Shekhar, M.S.,
4 Shanmugakarthik, J. et al. (2014) The development of a high density linkage map for black
5 tiger shrimp (Penaeus monodon) based on cSNPs. *PLOS One* **9**: e85413.
6 Berg, J.M., Tymoczko, J.L., Stryer, L., and Stryer, L.B. (2007) *Biochemistry*. New York: W.
7 H. Freeman ; Basingstoke.
8 Berlin, K., Koren, S., Chin, C.S., Drake, J.P., Landolin, J.M., and Phillippy, A.M. (2015)
9 Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.
10 *Nat Biotechnol* **33**: 623-630.
11 Bu, H.F., Wang, X., Tang, Y., Koti, V., and Tan, X.D. (2010) Toll-like receptor 2-mediated
12 peptidoglycan uptake by immature intestinal epithelial cells from apical side and exosome-
13 associated transcellular transcytosis. *J Cell Physiol* **222**: 658-668.
14 Campbell, S.E., Williams, T.A., Yousuf, A., Soanes, D.M., Paszkiewicz, K.H., and Williams,
15 B.A. (2013) The genome of *Spraguea lophii* and the basis of host-microsporidian interactions.
16 *PLOS Genet* **9**: e1003676.
17 Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B. et al. (2008) MAKER:
18 An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome*
19 *Research* **18**: 188-196.
20 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for
21 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-
22 1973.
23 Castillo-Juarez, H., Campos-Montes, G.R., Caballero-Zamora, A., and Montaldo, H.H. (2015)
24 Genetic improvement of Pacific white shrimp [*Penaeus* (*Litopenaeus*) *vannamei*]:
25 perspectives for genomic selection. *Front Genet* **6**: 93.
26 Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T. et al.
27 (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic*
28 *Acids Res* **40**: D700-D705.

Chilmonczyk, S., Cox, W.T., and Hedrick, R.P. (1991) *Enterocytozoon salmonis* N. Sp. - an Intranuclear Microsporidium from Salmonid Fish. *Journal of Protozoology* **38**: 264-269.

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

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

Corsaro, D., Walochnik, J., Venditti, D., Muller, K.D., Hauroder, B., and Michel, R. (2014) Rediscovery of *Nucleophaga amoebae*, a novel member of the Rozellomycota. *Parasitol Res* **113**: 4491-4498.

Cuomo, C.A., Desjardins, C.A., Bakowski, M.A., Goldberg, J., Ma, A.T., Becnel, J.J. et al. (2012) Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res* **22**: 2478-2488.

Desjardins, C.A., Sanscrainte, N.D., Goldberg, J.M., Heiman, D., Young, S., Zeng, Q. et al. (2015) Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun* **6**: 7121.

Desportes, I., Lecharpentier, Y., Galian, A., Bernard, F., Cochandpriollet, B., Lavergne, A. et al. (1985) Occurrence of a New Microsporidan: *Enterocytozoon bieneusi* n. g., n. sp., in the Enterocytes of a Human Patient with AIDS. *Journal of Protozoology* **32**: 250-254.

Dolgikh, V.V., Senderskiy, I.V., Pavlova, O.A., Naumov, A.M., and Beznoussenko, G.V. (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.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 1-19.

Franzen, C., Muller, A., Hartmann, P., and Salzberger, B. (2005) Cell invasion and intracellular fate of *Encephalitozoon cuniculi* (Microsporidia). *Parasitology* **130**: 285-292.

1 Freeman, M.A., and Sommerville, C. (2009) *Desmozoon lepeophtherii* n. gen., n. sp.,
2 (Microsporidia: Enterocytozoonidae) infecting the salmon louse *Lepeophtheirus salmonis*
3 (Copepoda: Caligidae). *Parasites & Vectors* **2**.
4 Freeman, M.A., Kasper, J.M., and Kristmundsson, A. (2013) *Nucleospora cyclopteri* n. sp.,
5 an intranuclear microsporidian infecting wild lumpfish, *Cyclopterus lumpus* L., in Icelandic
6 waters. *Parasit Vectors* **6**: 49.
7 Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView Version 4: A Multiplatform
8 Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol*
9 *Evol* **27**: 221-224.
10 Haag, K.L., James, T.Y., Pombert, J.F., Larsson, R., Schaer, T.M., Refardt, D., and Ebert, D.
11 (2014) Evolution of a morphological novelty occurred before genome compaction in a lineage
12 of extreme parasites. *Proc Natl Acad Sci U S A* **111**: 15480-15485.
13 Hacker, C., Howell, M., Bhella, D., and Lucocq, J. (2014) Strategies for maximizing ATP
14 supply in the microsporidian Encephalitozoon cuniculi: direct binding of mitochondria to the
15 parasitophorous vacuole and clustering of the mitochondrial porin VDAC. *Cell Microbiol* **16**:
16 565-579.
17 Hedrick, R.P., Groff, J.M., and Baxa, D.V. (1991) *Enterocytozoon salmonis* Chilmonczyk,
18 Cox, Hedrick (Microsporea): an intranuclear microsporidium from chinook salmon
19 *Oncorhynchus tshawytscha*. *Dis Aquat Organ* **10**: 103-108.
20 Heinz, E., Williams, T.A., Nakjang, S., Noël, C.J., Swan, D.C., Goldberg, A.V. et al. (2012)
21 The Genome of the Obligate Intracellular Parasite *Trachipleistophora hominis*: New Insights
22 into Microsporidian Genome Dynamics and Reductive Evolution. *PLOS Pathog* **8**: e1002979.
23 Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai,
24 K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**: W585-W587.
25 Keeling, P.J., Corradi, N., Morrison, H.G., Haag, K.L., Ebert, D., Weiss, L.M. et al. (2010)
26 The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for
27 core carbon metabolism. *Genome Biol Evol* **2**: 304-309.

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

Kuser, P., Cupri, F., Bleicher, L., and Polikarpov, I. (2008) Crystal structure of yeast hexokinase PI in complex with glucose: A classical "induced fit" example revised. *Proteins* **72**: 731-740.

Leiro, J., Iglesias, R., Parama, A., Sanmartin, M.L., and Ubeira, F.M. (2001) Respiratory burst responses of rat macrophages to microsporidian spores. *Exp Parasitol* **98**: 1-9.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**: 2178-2189.

Liu, D., Yang, J., and Wang, L. (2013) Cadmium induces ultrastructural changes in the hepatopancreas of the freshwater crab *Sinopotamon henanense*. *Micron* **47**: 24-32.

Lom, J., and Dykova, I. (2002) Ultrastructure of *Nucleospora secunda* n. sp (Microsporidia), parasite of enterocytes of *Nothobranchius rubripinnis*. *Eur J Protistol* **38**: 19-27.

Lu, Q., Li, S., and Shao, F. (2015) Sweet talk: Protein glycosylation in bacterial interaction with the host. *Trends Microbiol* **23**: 630-641.

Michel, R., Müller, K.-D., and Hauröder, B. (2009) A novel microsporidian endoparasite replicating within the nucleus of *Saccamoeba limax* isolated from a pond. *Endocytobios Cell Res* **19**: 120–126.

Michel, R., Schmid, E.N., Boker, T., Hager, D.G., Muller, K.D., Hoffmann, R., and Seitz, H.M. (2000) *Vannella* sp. harboring Microsporidia-like organisms isolated from the contact lens and inflamed eye of a female keratitis patient. *Parasitol Res* **86**: 514-520.

Moharir, A., Peck, S.H., Budden, T., and Lee, S.Y. (2013) The role of N-glycosylation in folding, trafficking, and functionality of lysosomal protein CLN5. *PLOS One* **8**: e74299.

Mueckler, M. (1994) Facilitative glucose transporters. *Eur J Biochem* **219**: 713-725.

Nakjang, S., Williams, T.A., Heinz, E., Watson, A.K., Foster, P.G., Sendra, K.M. et al. (2013) Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol* **5**: 2285-2303.

1 Neal, M.D., Leaphart, C., Levy, R., Prince, J., Billiar, T.R., Watkins, S. et al. (2006)
2 Enterocyte TLR4 mediates phagocytosis and translocation of bacteria across the intestinal
3 barrier. *J Immunol* **176**: 3070-3079.
4 Newman, S.G. (2015) Microsporidian impacts shrimp production-industry efforts address
5 control, not eradication. *Glob Aquac Advocate*: 16-17.
6 Nylund, S., Nylund, A., Watanabe, K., Arnesen, C.E., and Karlsbakk, E. (2010)
7 *Paranucleospora theridion* n. gen., n. sp (Microsporidia, Enterocytozoonidae) with a Life
8 Cycle in the Salmon Louse (*Lepeophtheirus salmonis*, Copepoda) and Atlantic Salmon
9 (*Salmo salar*). *J Euk Micro* **57**: 95-114.
10 Pereyre, S., Sirand-Pugnet, P., Beven, L., Charron, A., Renaudin, H., Barre, A. et al. (2009)
11 Life on arginine for *Mycoplasma hominis*: clues from its minimal genome and comparison
12 with other human urogenital mycoplasmas. *PLOS Genet* **5**: e1000677.
13 Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating
14 signal peptides from transmembrane regions. *Nature Methods* **8**: 785-786.
15 Rajendran, K.V., Shivam, S., Praveena, P.E., Rajan, J.J.S., Kumar, T.S., Avunje, S. et al.
16 (2016) Emergence of *Enterocytozoon hepatopenaei* (EHP) in farmed *Penaeus* (*Litopenaeus*)
17 *vannamei* in India. *Aquaculture* **454**: 272-280.
18 Richards, T.A., Hirt, R.P., Williams, B.A., and Embley, T.M. (2003) Horizontal gene transfer
19 and the evolution of parasitic protozoa. *Protist* **154**: 17-32.
20 Saier, M.H., Jr., Tran, C.V., and Barabote, R.D. (2006) TCDB: the Transporter Classification
21 Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**:
22 D181-D186.
23 Salachan, P.V., Jaroenlak, P., Thitamadee, S., Itsathitphaisarn, O., and Sritunyalucksana, K.
24 (2017) Laboratory cohabitation challenge model for shrimp hepatopancreatic
25 microsporidiosis (HPM) caused by *Enterocytozoon hepatopenaei* (EHP). *BMC Vet Res* **13**: 9.
26 Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-
27 2069.

Shental-Bechor, D., and Levy, Y. (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A* **105**: 8256-8261.

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Stentiford, G.D., and Bateman, K.S. (2007) *Enterospora* sp., an intranuclear microsporidian infection of hermit crab *Eupagurus bernhardus*. *Diseases of Aquatic Organisms* **75**: 73-78.

Stentiford, G.D., Bateman, K.S., Longshaw, M., and Feist, S.W. (2007) *Enterospora canceri* n. gen., n. sp., intranuclear within the hepatopancreatocytes of the European edible crab *Cancer pagurus*. *Dis Aquat Organ* **75**: 61-72.

Stentiford, G.D., Becnel, J.J., Weiss, L.M., Keeling, P.J., Didier, E.S., Williams, B.A. et al. (2016) Microsporidia - Emergent Pathogens in the Global Food Chain. *Trends Parasitol* **32**: 336-348.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol and Evol* **28**: 2731-2739.

Tangprasittipap, A., Chouwdee, S., Phiwsaiya, K., Laiphrom, S., Senapin, S., Flegel, T.W., and Sritunyalucksana, K. (2012) Structure and expression of a shrimp prohormone convertase 2. *Gen Comp Endocrinol* **178**: 185-193.

Tangprasittipap, A., Srisala, J., Chouwdee, S., Somboon, M., Chuchird, N., Limsuwan, C. et al. (2013) The microsporidian *Enterocytozoon hepatopenaei* is not the cause of white feces syndrome in whiteleg shrimp *Penaeus* (*Litopenaeus*) *vannamei*. *BMC Vet Res* **9**: 139.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K., and Tolstoy, I. (2014) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res* **42**: D553-D559.

Thitamadee, S., Prachumwat, A., Srisala, J., Jaroenlak, P., Salachan, P.V., Sritunyalucksana, K. et al. (2016) Review of current disease threats for cultivated penaeid shrimp in Asia. *Aquaculture* **452**: 69-87.

Tonks, N.K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat Rev Mol Cell Biol* **7**: 833-846.

Trentmann, O., Horn, M., van Scheltinga, A.C., Neuhaus, H.E., and Haferkamp, I. (2007) Enlightening energy parasitism by analysis of an ATP/ADP transporter from chlamydiae. *PLOS Biol* **5**: e231.

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

Unal, C.M., and Steinert, M. (2014) Microbial peptidyl-prolyl cis/trans isomerases (PPIases): virulence factors and potential alternative drug targets. *Microbiol Mol Biol Rev* **78**: 544-571.

Vagin, O., Kraut, J.A., and Sachs, G. (2009) Role of N-glycosylation in trafficking of apical membrane proteins in epithelia. *Am J Physiol Renal Physiol* **296**: F459-469.

Vavra, J., and Lukes, J. (2013) Microsporidia and 'the art of living together'. *Adv Parasitol* **82**: 253-319.

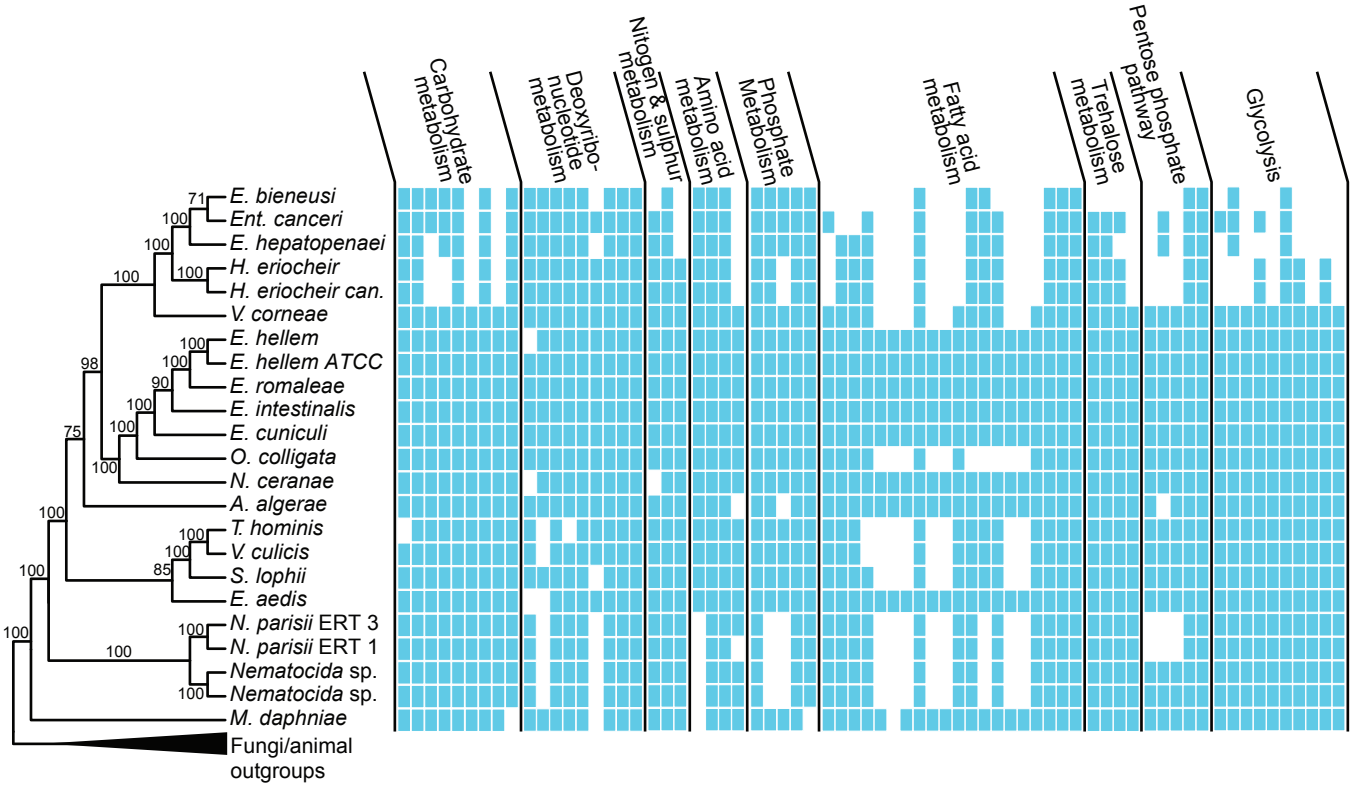
Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M. et al. (2003) The genome of *Nanoarchaeum equitans:* insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A* **100**: 12984-12988.

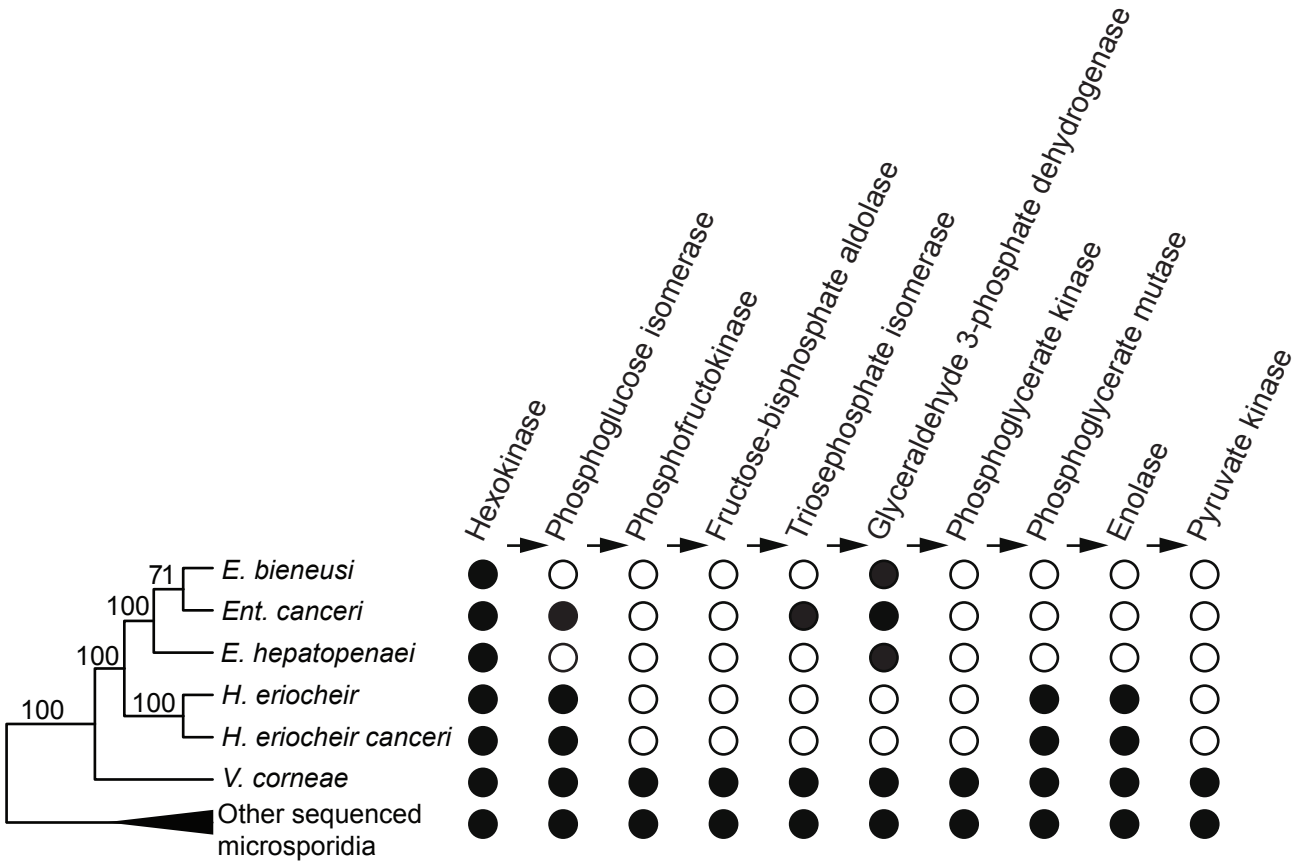
Williams, B.A., Hirt, R.P., Lucocq, J.M., and Embley, T.M. (2002) A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* **418**: 865-869.

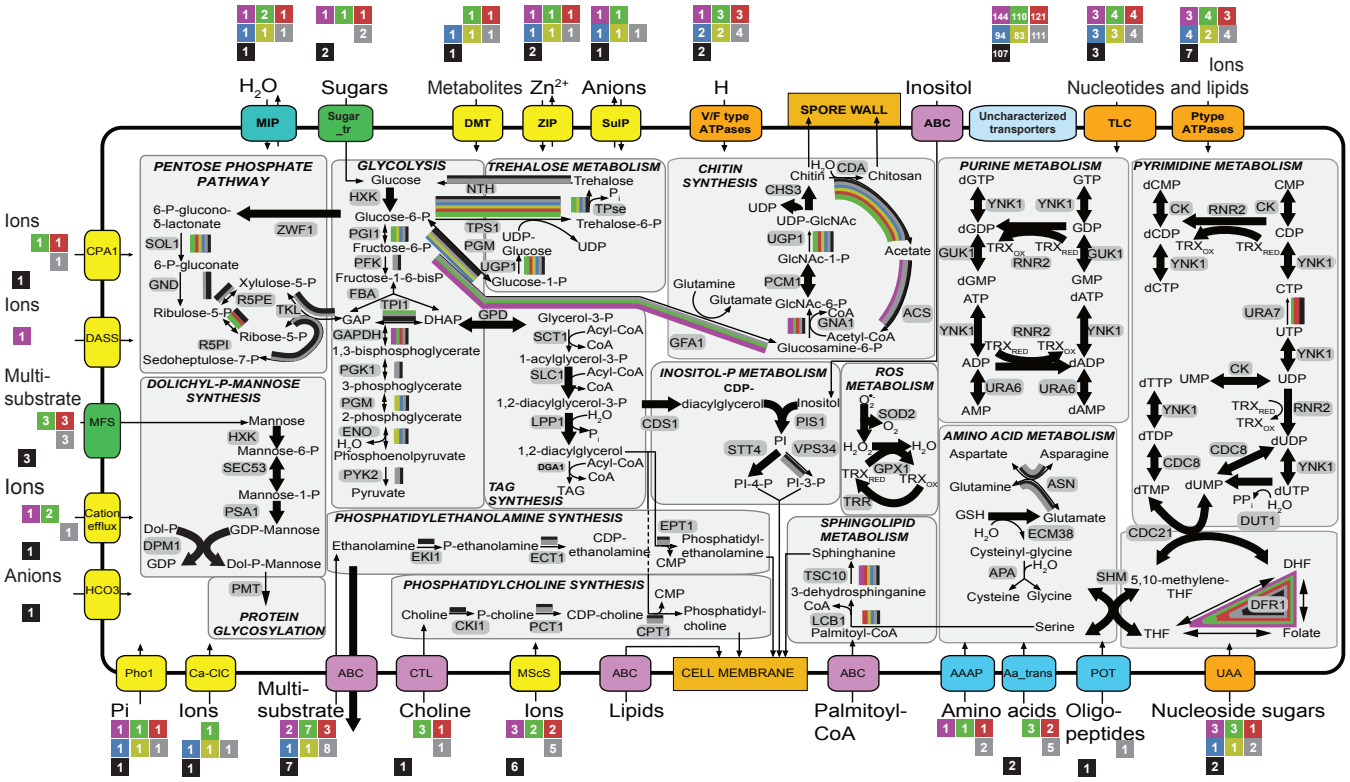
Wujek, P., Kida, E., Walus, M., Wisniewski, K.E., and Golabek, A.A. (2004) N-glycosylation is crucial for folding, trafficking, and stability of human tripeptidyl-peptidase I. *J Biol Chem* **279**: 12827-12839.

You, E.M., Liu, K.F., Huang, S.W., Chen, M., Groumellec, M.L., Fann, S.J., and Yu, H.T. (2010) Construction of integrated genetic linkage maps of the tiger shrimp (Penaeus monodon) using microsatellite and AFLP markers. *Anim Genet* **41**: 365-376.

1 Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y. et al. (2015) Genome survey and high-
2 density genetic map construction provide genomic and genetic resources for the Pacific White
3 Shrimp *Litopenaeus vannamei*. *Sci Rep* **5**: 15612.
4


Carbohydrate metabolism
Deoxyribo-nucleotide metabolism
Nitrogen & sulphur metabolism
Amino acid metabolism
Phosphate Metabolism
Fatty acid metabolism
Trehalose metabolism
Pentose phosphate pathway
Glycolysis
E. bieneusi
Ent. canceri
E. hepatopenaei
H. eriocheir
H. eriocheir can.
V. corneae
E. hellem
E. hellem ATCC
E. romaleae
E. intestinalis
E. cuniculi
O. colligata
N. ceranae
A. algerae
T. hominis
V. culicis
S. lophii
E. aedis
N. parisii ERT 3
N. parisii ERT 1
Nematocida sp.
Nematocida sp.
M. daphniae
Fungi/animal outgroups

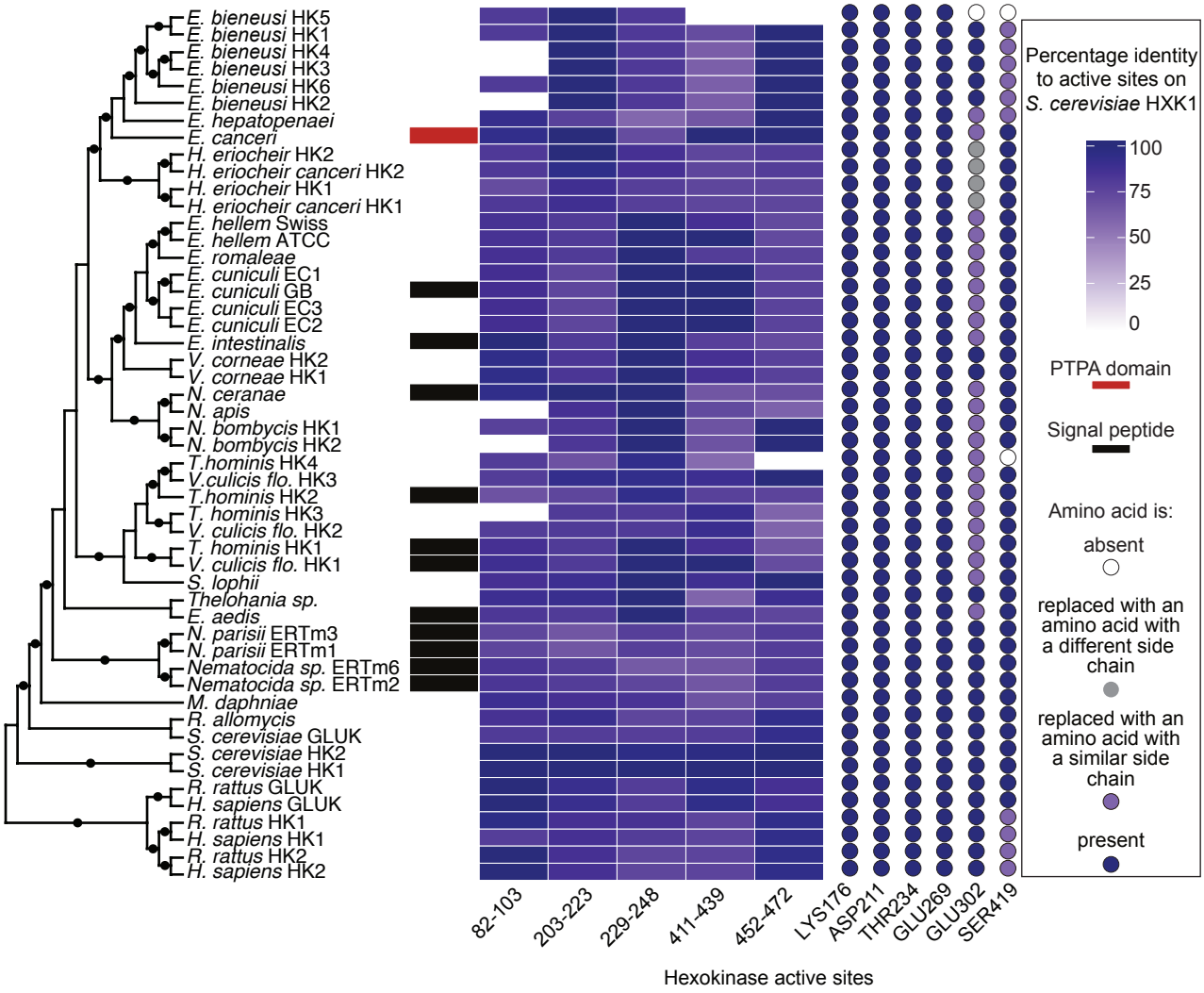

Hexokinase
Phosphoglucose isomerase
Phosphofructokinase
Fructose-bisphosphate aldolase
Triosephosphate isomerase
Glyceraldehyde 3-phosphate dehydrogenase
Phosphoglycerate kinase
Phosphoglycerate mutase
Enolase
Pyruvate kinase
71
100
100
100
100
E. bieneusi
Ent. canceri
E. hepatopenaei
H. eriocheir
H. eriocheir canceri
V. corneae
Other sequenced microsporidia



- *E. bieneusi*
- *Ent. canceri*
- *E. hepatopenaei*
- *H. eriocheir*
- *H. eriocheir canceri*
- *Enc. cuniculi*
- *T. hominis*

Dol - Dolichyl
TAG -Triacylglycerol
GAP - Glyceraldehyde-3-phosphate
DHAP - Dihydroxiacetone phosphate
GlcNAc - N-acetlyglucosamine
PI - phosphatidylinositol
GSH - Glutathione
THF - Tetrahydrofolate
DHF - Dihydrofolate
TRX - Thioredoxin

Cation efflux - Cation Efflux Pump
HCO3 - Anion Transport Protein
MIP - Major Intrinsic Protein
MFS - Major Facilitator Superfamily
DMT - Drug/Metabolite Transporter
ZIP - Zinc/Iron Permease
SulP - Sulfate Permease
AAAP - Amino Acid/Auxin Permease
V/F typeATPases - Vacuolar Vesicular Proton Pump
P type ATPases - E1-E2 ATPases
Sugar_tr - Glucose transporter/Sugar transporter family
Cation efflux - Cation Efflux Pump
Ca CIC - Calcium Dependent Chloride Channels
MscS - Small Conductance Mechanosensitive Ion Channel
TLC - ATP/ADP Translocase
UAA - UDP-N-acetylglucosamine Transporter
ABC - ATP-Binding Casette
CTL - Choline Transporter
POT - Proton-Dependent Oligopeptide Transporter
Pho1 - Phosphate transporter


E. bieneusi HK5
E. bieneusi HK1
E. bieneusi HK4
E. bieneusi HK3
E. bieneusi HK6
E. bieneusi HK2
E. hepatopenaei
E. canceri
H. eriocheir HK2
H. eriocheir canceri HK2
H. eriocheir HK1
H. eriocheir canceri HK1
E. hellem Swiss
E. hellem ATCC
E. romaleae
E. cuniculi EC1
E. cuniculi GB
E. cuniculi EC3
E. cuniculi EC2
E. intestinalis
V. corneae HK2
V. corneae HK1
N. ceranae
N. apis
N. bombycis HK1
N. bombycis HK2
T.hominis HK4
V.culicis flo. HK3
T.hominis HK2
T. hominis HK3
V. culicis flo. HK2
T. hominis HK1
V. culicis flo. HK1
S. lophii
Thelohania sp.
E. aedis
N. parisii ERTm3
N. parisii ERTm1
Nematocida sp. ERTm6
Nematocida sp. ERTm2
M. daphniae
R. allomycis
S. cerevisiae GLUK
S. cerevisiae HK2
S. cerevisiae HK1
R. rattus GLUK
H. sapiens GLUK
R. rattus HK1
H. sapiens HK1
R. rattus HK2
H. sapiens HK2
82-103
203-223
229-248
411-439
452-472
LYS176
ASP211
THR234
GLU269
GLU302
SER419
Hexokinase active sites
Percentage identity to active sites on S. cerevisiae HXK1
100
75
50
25
0
PTPA domain
Signal peptide
Amino acid is:
absent
replaced with an amino acid with a different side chain
replaced with an amino acid with a similar side chain
present