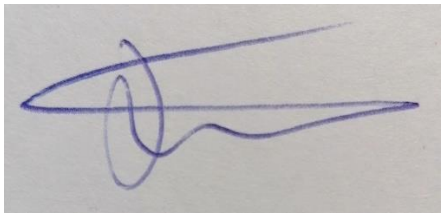
The Enterocytozoonidae: the emergence of a microsporidian clade into the aquatic and terrestrial food-chain.

Submitted by Jahcub Trew, to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy Biological Sciences
in February 2021

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.



(Signature)

……

ABSTRACT

Pathogens of livestock have proven to be a major concern, in terms of human health, economic sustainability and food availability. They have been shown to be a limiting factor in regard to these three factors and, as such, will become a growing problem as the global population continues to expand. Controlling and predicting pathogen outbreaks is vital to the sustained growth of populations and has become an important topic of study. However, it is a complex process with a multitude of varying factors, that cannot be solved through one method.

A specific family of Microsporidia, the Enterocytozoonidae, has received a lot of attention in regard to their effect on human health and aquaculture. Enterocytozoonids are largely found in marine environments and mainly infect aquatic hosts. Many of the organisms infected by enterocytozoonids are economically important to aquaculture, with a number infecting wild-caught fish and farmed crustacea. Two species, in particular, have received a lot of academic and medical attention, *Enterocytozoon bieneusi* and *Enterocytozoon hepatopenaei*. However, the family as a whole is understudied and the diversity described is hypothesised not to be a true representation of the family's diversity or distribution.

The aim of this thesis is to assess the potential risk the Enterocytozoonidae have on aquaculture. This has been done in three different ways: i) it described and catalogued the current diversity and distribution in South West UK estuaries and a greater spatial scale using metagenomic databases, ii) using population genomics, it investigates the biogeography of, *E. hepatopenaei*, a pathogen of two shrimp (*Penaeus vannamei* and *Penaeus monodon*) widely farmed in South East Asia, with the aim to infer transmission routes into shrimp ponds, iii) and lastly, it investigated the loss and gain of orthogroups within the Enterocytozoonidae associated with the lineage wide host-shift to aquatic host from terrestrial hosts.

Overall, this thesis found a greater distribution and abundance for *Enterospora canceri* and *E. hepatopenaei* in the locations sampled, and greater diversity

within the Enterocytozoondiae than was previously described. Suggesting an alternative host (s) for *En. canceri*, likely planktonic crustacea, as prevalence in the current described host was very low (3%). Use of metagenomic databases in this study also expanded upon the known distribution and diversity of the Enterocytozoonidae, showing a largely marine-based distribution. Suggesting an Atlantic-based radiation. It found that *E. hepatopenaei* is likely to be endemic to the countries examined, due to strong geographic signals. However, analysis suggests some transmission between Thailand India and China. Lastly, it identified candidate genes, unique to lineages within the Enterocytozoonidae, that could contribute to the family's success in the invasion of their respective host cells.

A greater diversity and abundance for novel sequences more related to the *Enterocytozoon/Enterospora* branch of the Enterocytozoondiae was also observed, suggesting, at least in estuaries in the Southwest of the United Kingdom, that this clade may be more diverse. Possibly due to their described main hosts being planktonic crustaceans. Which may lend to the success of the two most prevalent species in the family, *E. hepatopenaei* and *E. bieneusi*.

This study found that the Enterocytozoonidae likely pose a continued threat to aquatic livestock, largely due to their widespread nature, overlapping with human influenced environments, and evident opportunistic propensity to host-shift. In addition, also being present inland in freshwater makes preventative measures difficult, as they could be found in a range of water sources.

Contents

TABLES

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisors, Dr. Bryony Williams, and Prof. Grant Stentiford for giving me the opportunity to undertake this Ph.D. and for the support and direction throughout the studentship. With special thanks to Dr. David Studholme, who stepped in for the majority of my first year as a primary supervisor and provided help throughout the Ph.D. on bioinformatics, checking work, and CV writing.

To my previous supervisors Dr. Todd Rae (who introduced me to phylogenetics), Prof. Beth Okamura and Dr. Hannah Hartikainen (who introduced me to the application of phylogenetics in molecular ecology), I would not be here without you.

A huge thanks to the collaborators at Mahidol University, to Dr. Diva January Aldama-Cano and Dr. Natthinee Munkongwongsiri for collecting samples for me and aiding in collecting samples while I was there, and for providing a warm welcome environment to work in. Also, to Dr.Ornchuma Itsathitphaisarn and Dr.Kallaya Dangtip, for allowing access and use of their labs and resources. Furthermore, thanks to Dr. Tom Williams at Bristol University, for taking two days out of his schedule to help me with the analysis in Chapter 5 and for his advice in Chapter 4. Lastly, much of this work would not have been possible without help from members at CEFAS, namely Dr. Elisa Capuzzo, Dr. David Bass, and Dr. Ronnie van Aerle.

I am grateful for the laughs, entertainment, and 'support' provided by colleagues within Geoffrey Pope, Dr. Thomas Jenkins, Dr. Josh Dyer, Dr. Cameron Hubert, Hazel Knipe, and most importantly, the only other member of the Williams' group, (soon to be Dr.) Roland (Ronald) Pevsner. Special mention for Dr. Ellis Langley for extra motivation throughout my write-up and the encouraging words "don't fuck it up" when going for my interview.

Lastly, and most importantly, I would not be here without the inspiration for nature that my parents (Sareta Trew and Chamberlain Trew) instilled within me at a young age. Especially my sisters (Walatta and Makeda Trew) that encouraged me to pursue the study of nature and bought me countless nature books.

Chapter 1: Current burden of the Enterocytozoondiae on aqua/agriculture

1.1 The role of food in upcoming population growth

With a global population increase of more than 2 billion estimated in the next 30 years and the majority of the growth centered around food-deficit countries, the demand for global food production will increase with it (Hangmann, 2009; Tomley & Shirley, 2009; United Nations, 2019). This will require food production to increase by 70% within the next 30 years, almost doubling the current production for food in developing countries, and this does not take into account the possible rise in agricultural products required for biofuels (Hangmann, 2009). With terrestrial space being at a premium, and a global state of overfishing, the food industry has been expanding into fields that have historically received less investment. More investment is going into the farming of insects, a great source of fat, protein, and fibre; a system that does not take up much terrestrial space and is efficient in terms of energy requirement (A. van Huis *et al*., 2013). Another system receiving more interest is aquaculture: Aquaculture is the fastest-growing livestock-sector, supplying 17% of the globally consumed protein in 2017, and will likely make up a large portion of the world's intake of protein in the coming decades (Bayliss *et al*., 2017; Garcia & Rosenberg, 2010; G. D. Stentiford *et al*., 2017; United Nations, 2015). It is also predicted that aquaculture will account for 59% of global fish consumption (FAO, 2020). This will make aquaculture indispensable in feeding a growing global population (G. D. Stentiford *et al*., 2013, 2017; Subasinghe *et al*., 2009). Farmed seafood will also lessen the stress that large-scale net fishing has on wild fish and cage caught crustacea, with 46% of the world's consumed seafood and 52% of the world's consumed fish having been from farms in 2016-2018, (FAO, 2020). The rapid expansion of aquaculture-based farming has largely come about through need, with wild-caught numbers decreasing (not rising to meet the population's needs) (Bayliss *et al*., 2017; Garcia & Rosenberg, 2010). Most of the world's aquacultural produce comes from South-East Asia, producing 90%, with 60% of that coming from China alone (Bayliss *et al*., 2017; G. D. Stentiford *et al*., 2017).

Of the globally produced seafood, tropical shrimp is becoming a far more successful venture in aquaculture than the farming of fish (G. D. Stentiford, 2012). An undertaking that is largely being pursued, successfully, by South-East Asia and Central America (G. D. Stentiford, 2012), providing much-needed economic support to these areas. South-East Asia, China, and Central America made up 3.5 of the 4.3 million tons of the world's exportation of shrimp, in 2017 (*GOAL 2019: Global Shrimp Production Review « Global Aquaculture Advocate*, n.d.).

1.1.1 The role of pathogens in upcoming population growth

Inherent in the risk of a larger-scale monoculture of livestock and crops is a joint growth in pathogens that take advantage of these monocultures. Pathogens are already a major global problem to address, greatly reducing the efficiency of crop production with a ranging global loss of 8.1%-40.9% across five staple crops (Savary *et al*., 2019). The case is also very similar to animal-based foods, with a global loss of finfish ranging up to 5.8%-20% (Shinn *et al*., 2015). There have been several global pathogens that have swept through and caused major problems for a number of different livestock species, for example, African swine fever, a virus affecting farmed pigs that has been attributed to the loss of 50% of exports in Eastern Europe (Sánchez-Cordón *et al*., 2018). Even within more 'developed' countries, outbreaks of pathogens can cause major loss of livestock; for example, the outbreak of foot and mouth disease in the UK, 2001 led to a loss of £3 billion (Thomson *et al*., 2003). With globalization and the widespread export of livestock and crops making up a large portion of the economy of many countries, the spread of pathogens has become and will continue to be, an important issue. As such, threats from novel pathogens, as well as those thought to have been in decline (such as African swine fever) are a continual concern.

The loss of animal-based stock is heavily felt in aquaculture, with an estimated $6 billion lost per annum, due to various pathogens (G. D. Stentiford *et al*., 2017). For example, in 2016, a virulent strain of the gram negative bacterium, *Aeromonas hydrophila,* has caused a regional loss of 1.9 million catfish that amounts to a monetary loss of $2.6 million in East Mississippi alone (Peterman

& Posadas, 2019). Also in 2016, an outbreak of the causative agent of proliferative kidney disease (PKD), *Tetracapsuloisdes bryosalmonae*, in Yellowstone River resulted in an estimated loss of $500, 000 (Hutchins *et al*., 2021). This is likely to continue, especially given the bias towards research on terrestrial-based pathogens compared to that of aquatic pathogens. (G. D. Stentiford *et al*., 2017). Due to this lack of knowledge of aquatic-based pathogens and parasites, many symptoms may be overlooked allowing many of these pathogens to spread. Coupled with this is a dramatic increase in the number of species farmed in aquaculture, compared to that in agriculture (Bayliss *et al*., 2017; FAO, 2016). Due to this influx of new species, such as the Atlantic halibut (*Hippoglossus hippoglossus*) and the pike perch (*Sander lucioperca*) (Mylonas *et al*., 2019), depending on the state of biosecurity for the farm (s) in question, the introduction of new species to aquaculture can also come with inherent risk, in the form of pathogen spillover (Power & Mitchell, 2004). Pathogen spillover is the transmission of a pathogen (s) from its reservoir host species to a new species. Though any pathogens present in new farmed species would not pose an immediate threat in single species ponds, pathogens introduced to an environment through aquaculture have been shown to have the capacity to transmit regionally to wildlife and non-infected farmed ponds by persisting in the environment (Afonso *et al*., 2012; MG *et al*., 2015; Oidtmann *et al*., 2018).

Due to farming environments not always being ideal for animal wellbeing, pathogens may also become more virulent in hosts they are 'endemic' to, as the host may become stressed/malnourished (Earley *et al*., 2017; Manteca *et al*., 2013). These emergent pathogens have the potential to be even more detrimental in intensive aquaculture, due to high densities of potential hosts which increase transmission (Bayliss *et al*., 2017; Sundberg *et al*., 2016). Disease in certain crustacea, shrimp, for example, have sustained devastating losses estimated at over 40% of the global production of shrimp (G. D. Stentiford, 2012). There are a number of pathogens afflicting shrimp, largely viral and bacterial (Flegel, 2012), an example of some of the most impactful include Monodon baculovirus (MBV) (Flegel, 2006), Yellow-head virus (YHV) (Boonyaratpalin *et al*., 1993), White-spot syndrome virus (WSSV) (Flegel, 1997), *Vibrio* sp. (Soto-Rodriguez *et al*., 2015), and *Flavobacterium* sp. (Sheu *et al*., 2011). The types

of viral infection affecting shrimp have changed over the years, with white spot syndrome virus and yellow head virus being the most lethal among them (Flegel, 2012).

The problem with pathogens in animal-based foods is also compounded, as many pathogens of livestock are capable of zoonosis (H. Li *et al*., 2019; Tomley & Shirley, 2009). Many of the pandemics afflicting humans have come from foodborne pathogens, with most introductions to these pathogens from ingestion (H. Li *et al*., 2019). Current known foodborne pathogens have resulted in 600 million diseases in 2010 (Hoffmann & Scallan, 2017). The inherent risk to animal-based foods is not equal among all food types, with different types of pathogens being a higher risk dependent on the countries they are farmed in (Hoffmann & Scallan, 2017). Another example of disproportionate risks in animal-based food types are that the majority of emerging human diseases come from mammals (Cleaveland *et al*., 2001; Han *et al*., 2016; Woolhouse & Gowtage-Sequeria, 2005). However, this is not a trend that is seen in virus-based zoonotic infections, instead, displaying a host-neutral preference with greater infection rates seen in the diversity of viruses held by the reservoir group (Mollentze & Streicker, 2020). There is also a disproportionate burden/severity of pathogens on livestock and crops, with some crops less affected by their respective pathogens, likely due to the lack of resources in the food-deficit countries these crops are grown (Savary *et al*., 2019). This is likely to become a problem, with much of the predicted population growth coming from food-deficit countries (Tomley & Shirley, 2009).

1.1.2 Prevention of pathogen transmission

Prevention of outbreaks of the scale that have been seen many times in aquaculture and agriculture generally takes the form of computer-based modelling, vaccination, and control measures, for example, burning millions of cows to slow down the spread of the foot and mouth pandemic (Prempeh, 2001; Thomson *et al*., 2003). Another preventative measure is producing genetically modified stock, more resistant to pathogens (Pal & Chakravarty, 2020), and selective breeding of disease resistance (Pal & Chakravarty, 2020). Selective breeding for disease resistance has been used in aquaculture for close to 30

years (Gjøen & Bentsen, 1997). The general premise being that along with other beneficial characteristics, such as quick growth rates and greater retention/production of fat/muscle, they will also be resistant to a range of pathogens currently afflicting a range of similar livestock. Unfortunately, an element of human error and a lack of understanding of the diversity of pathogens and transmission routes often means biosecurity falls short. One of the elements is solved with continued surveys to discover new threats. One such way that is becoming widely used is metabarcoding for animal pathogens on environmental DNA (eDNA) (Huver *et al*., 2015; Peters *et al*., 2018). By identifying possible pathogens in the environment, this allows for more effective prevention of pathogens spread and vector control. Farmside detection methods are also being worked on for various known pathogens, as well as disseminating knowledge of known symptoms for particular harmful pathogens (Minardi *et al*., 2019).

Some preventative methods may prove to, in some cases, be detrimental in the future. For example, the use of cleaner fish has become quite useful in the removal of ectoparasites, like the salmon lice (*Lepeophtheirus salmonis*), the infection of which can lead to a reduction in growth, bodily fluids, and a raised susceptibility to secondary infections from other parasites (Gargan *et al*., 1993; Thorstad *et al*., 2015; Thorstad & Finstad, 2018; Whelan, 2010). Pathogens could also come from the cleaner fish brought in to remove the ectoparasites, as they are known to hold opportunistic infections too (Freeman *et al*., 2013; Hjeltnes *et al*., 2019). A danger that is likely to grow with the rising use of cleaner fish in aquaculture (Winfield, 2018).

1.1.3 Rise of opportunistic pathogens

Opportunist parasites are organisms that primarily rely on susceptibility and the weakened immune system in the host to infect, and that are rarely found in healthy individuals (S. P. Brown *et al*., 2012). Their ecological niches can vary widely, some largely being facultative, existing as commensal symbionts in the host-microbiome until such a time as the immune system is compromised (Taur & Pamer, 2013), or acquired from the environment (Anttila *et al*., 2015; S. P. Brown *et al*., 2012; Pamer, 2007). Opportunistic parasites make implementing preventative measures highly difficult, as they are non-specialist parasites, they can be found in a number of different hosts. This makes specific prevention

difficult, such as vaccination for endoparasites, as it would require prior knowledge of a parasite that has, as of yet, been described infecting the host in question. This would require in-depth research on individual parasites, which is extremely time and monetarily costly. Due to this, the majority of all bacterial diseases that arise in aquaculture are brought about by opportunist pathogens that are found in the marine environment (De Schryver & Vadstein, 2014; Defoirdt, 2016). This is because many opportunist pathogens are environmentally ubiquitous and wide-ranging and can proliferate in aquaculture monocultures (Defoirdt, 2016). Another contributing factor to the high levels of infection in farmed monocultures is genomic plasticity, or the comparative lack thereof. It has been shown that populations with higher genomic diversity show a greater resistance to infection than those populations with lower genomic diversity (Ekroth *et al*., 2019; King & Lively, 2012). Unfortunately, though selective breeding programs do promote genomic diversity (Goecke *et al*., 2020), the complexity of how natural population diversity arises is a difficult model to employ, and thus the selective breeding programs need to be further developed (Goecke *et al*., 2020; G. D. Stentiford *et al*., 2017). This also explains why the disease toll in food-deficit countries is higher, as they do not have the resources to properly test and filter the water, allowing opportunistic pathogens to enter the system. Through lack of resources, stress caused by suboptimal conditions is also more likely to occur in food-deficit countries. As is well known, stress can compromise the immune system which would allow the opportunistic pathogens that have found a way into the system to proliferate (Lafferty & Holt, 2003). Opportunistic pathogens also take advantage of hosts that are being weakened by a 'professional' pathogen of that host leaving the door open for opportunistic pathogens (Karvonen *et al*., 2010; Louhi *et al*., 2015).

1.2 Microsporidia, opportunistic pathogens

Microsporidia are a group of obligate intracellular, parasites with 1300-1500 described species across roughly 200 genera (Weiss & Becnel, 2014), though it has been remarked that this is likely not representative of their diversity (Ardila-Garcia *et al*., 2013). Microsporidia have a fairly cosmopolitan distribution, and have been described on all continents, except Antarctica (P. Keeling, 2009), and inhabit a number of diverse environments and hosts (González-Tortuero *et al*.,

2016; Klee *et al*., 2007; Lom & Dykoá, 2002; G. D. Stentiford *et al*., 2013). In humans, they cause varying conditions with varying symptoms, some more detrimental than others, ranging from AIDS-associated diarrhea and dysentery (Didier *et al*., 1995; Matos *et al*., 2012) to fatal deep tissue infections (Coyle *et al*., 2004; Nadelman *et al*., 2020). Among their hosts are a number of commercially and environmentally important species, such as honey bees, salmon (El Alaoui *et al*., 2006a; Klee *et al*., 2007; Palenzuela *et al*., 2014), pets (such as cats, dogs, or guinea pigs) (Cama *et al*., 2007; Mathis *et al*., 1999), and farmed livestock (Mathis *et al*., 1999; Q. Zhang *et al*., 2018).

1.2.1 The history of microsporidian phylogeny

The phylogenetic position of Microsporidia had been difficult to determine and they have been repositioned several times as a result. They were first considered to be a part of the Sporozoa, a group of parasitic protozoans, and then more specifically within the Sporozoa, in the Cnidosporidia (P. Keeling, 2009; Vossbrinck *et al*., 2014). Due to the lack of observed mitochondria at the time, they were more recently considered to be Archezoa, a group of eukaryotes that diverged prior to the acquisition of a mitochondrion (Cavalier-Smith, 1983). However, due to the discovery of their reduced mitochondrion, referred to as a mitosome (B. A. P. Williams *et al*., 2002), and evidence from increasing molecular sequence data (Capella-Gutiérrez *et al*., 2012; Vossbrinck *et al*., 2014; Vossbrinck & Debrunner-Vossbrinck, 2005) they are now positioned within a clade of divergent eukaryotes called the Opisthosporidia (including Cryptomycota and Aphelida), a sister group to true fungi (Figure 1.1) (Karpov *et al*., 2013, 2014; Torruella *et al*., 2015). Historically, microsporidian taxonomy was heavily based on phenotypic, developmental, and ecological characters, and though useful, the phylogenetic importance of these characters is not yet known (Vossbrinck *et al*., 2014). They have a number of derived features that set them apart from other eukaryotes (P. Keeling, 2009; Vossbrinck & Debrunner-Vossbrinck, 2005), though the most characteristic among them is a structure called a polar filament (Bigliardi & Sacchi, 2001; Franzen, 2005), a structure involved in the infection of host cells.

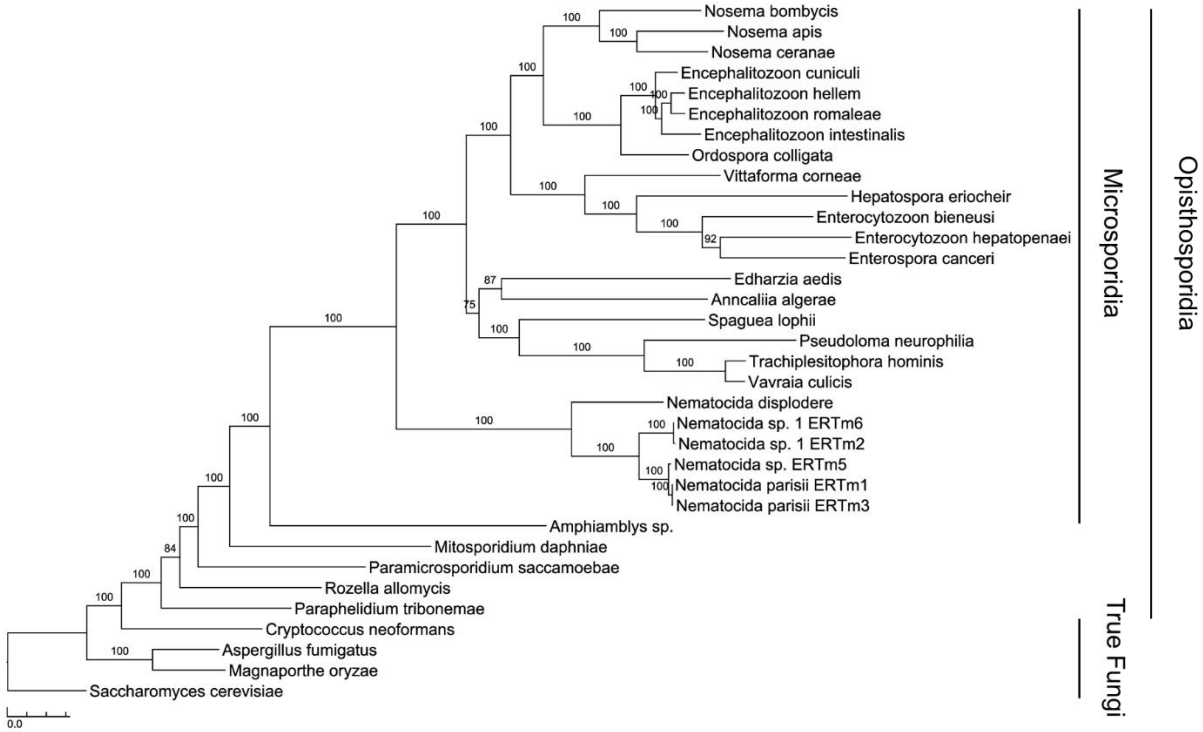


Figure 1.1: Position of Microsporidia within the Opisthosporidia. Phylogeny was inferred from 25 single-copy orthologs shared by all taxa (34 species) (Chapter 5), built using maximum likelihood in RAxML-HPC (Stamatakis, 2006).

1.2.2 Microsporidian transmission and ecology

The polar filament is a unique structure crucial for infection of the host. It is used to pierce the host's cell membrane, whereupon they extrude their sporoplasm in the host's cell (Bigliardi & Sacchi, 2001; Franzen, 2005) (Figure 1.2). Within the cell, the sporoplasm develops mature spores through a process called sporogony in which it undergoes multiple rounds of fission (Bigliardi & Sacchi, 2001; Franzen, 2005). When enough mature spores are formed, they rupture the cell membrane moving on to infect a new host or new cells within the same host and to survive harsh periods out of a host (Bigliardi & Sacchi, 2001; Franzen, 2005). These resistant chitinous spores are their sole means of propagation.

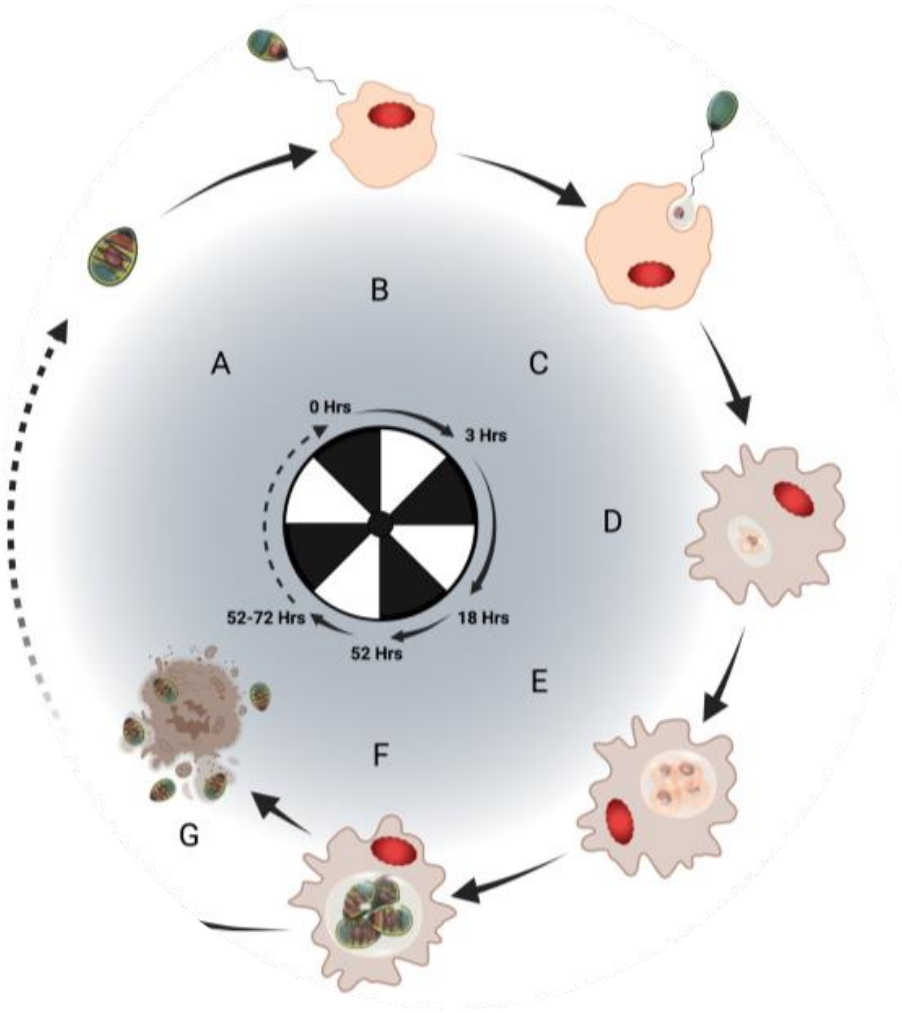


Figure 1.2: General life cycle of a microsporidian spore. A=free spore; B=spore comes into contact with the host cell, ejects polar filament and interacts with the host's cell membrane, C/D= contents of spore (sporoplasm) is extruded though polar filament into host's cell, E=proliferative stage whereupon sporoplasm undergoes fission, F= mature spores are formed, G= replicating spores disrupt cell membrane and are dispersed into the surroundings.

These spores are transmitted in a number of ways, many species of Microsporidia have been shown to transmit horizontally (Haag *et al*., 2019; Karthikeyan & Sudhakaran, 2019a; Tangprasittipap *et al*., 2013a; Wang-Peng *et al*., 2018), picked up from other hosts through ingestion or another interaction, though some species have been found to transmit vertically (Haine *et al*., 2007; Poley *et al*., 2017) and some are capable of both (Haag *et al*., 2019). There are two ways (modes) that a parasite can be transmitted to a new host, either vertically or horizontally. Transmissions that are brought about through reproductive routes, parent-to-offspring transmission, are known as vertical (Antonovics *et al*., 2017).

Horizontally is the broader of the two modes and includes transmission routes that are not from parent-to-offspring transmission. Though not exhaustive, some routes include environmental transmission (via food and water (W. Li *et al*., 2017; Zhao *et al*., 2014)), from vectors (e.g. tsetse flies and bats (Geiger *et al*., 2018; Wang-Peng *et al*., 2018), and through respiration (Antonovics *et al*., 2017; Kutter *et al*., 2018). However, waterborne transmission is likely the most frequent route of transmission in human infecting species , as microsporidian infections are most reported in low-income countries with lower water health (Javanmard *et al*., 2018; G. D. Stentiford *et al*., 2016), and there are a number of studies indicating the presence of spores in both raw and treated water (Galván *et al*., 2013; Izquierdo *et al*., 2011; Javanmard *et al*., 2018). The proposed frequent route of transmission also concurs with a greater prevalence of microsporidian infection in lower-income countries (Z. D. Wang *et al*., 2018), as the water quality in lower-income countries is less regulated and more likely to contain pathogens (Deshpande *et al*., 2020).

Transmission modes have large effects on the life-cycle of pathogens and host-parasite dynamics and have even been found to dictate how virulent some Microsporidia are (Dunn & Smith, 2001). The mode of transmission has also been attributed to bottlenecks in some species, which leads to inefficient purifying selection in some species of microsporidia (Haag *et al*., 2019). Showing that microsporidia that have mixed-mode transmission are more likely to have large genomes, with more genes and longer intergenic regions (Haag *et al*., 2019).

Host specificity is routinely observed within the Microsporidia, with many species only being found in certain groups of animals, like *Nematocida* (only described in nematodes- *Caenorhabditis elegans*) (Luallen *et al*., 2016; Reinke *et al*., 2017; Wadi & Reinke, 2020) and some *Nosema* (honey bees- *Apis mellifera* and *Apis cerena*) species (M. J. F. Brown, 2017; Klee *et al*., 2007; Pan *et al*., 2013). However, there are many instances of more generalist species that infect more than one host species (Hinney *et al*., 2016; Jeong *et al*., 2007; Zhao *et al*., 2015), with some hosts possibly acting as vectors (Wang-Peng *et al*., 2018). The range of hosts a microsporidian species has is rarely fully understood, due to the

difficulty of researching intracellular pathogens, especially those highly capable of host-switching.

1.2.3 Microsporidia, emerging public health concern

There are 17 known species on Microsporidia that infect humans, belonging to eight genera (*Encephalitozoon* (Chabchoub *et al*., 2009; Didier & Khan, 2014)*, Enterocytozoon* (Desportes *et al*., 1985)*, Pleistophora* (Cali & Takvorian, 2003)*, Tachipleistophora* (Vávra *et al*., 1998)*, Nosema* (Cali *et al*., 2010)*, Vittaforma* (Shadduck *et al*., 1990)*, Brachiola* (synonym *Nosema), and Microsporidium* (Sharma *et al*., 2014)). One of the main reasons Microsporidia have received more attention, besides their interesting evolution and the greater access to more advanced biochemical methods, is their impact on human health. Many studies carried out on the phylum are based on human infecting species, contributing to the available genome sequence data. Although the species that have been found to infect humans have done so in an opportunistic fashion, infecting those that are immunocompromised, microsporidian infections are also found in individuals with healthy immune systems (Tabatabaie *et al*., 2015). The majority of the species that affect humans cause intestinal problems (Kotler & Orenstein, 1998; Weiss, 2014), but there are some that infect other organs, such as the kidney (Nagpal *et al*., 2013), the lungs (Teachey *et al*., 2004), and the eyes (Van Gool *et al*., 2004). Though the associated symptoms are generally not life-threatening while under care, diarrhea without treatment can and does lead to death. These species are suspected to be transmitted into humans through water sources or irrigation of vegetable farms, wild and domestic animals, and through livestock (Didier, 2005; Javanmard *et al*., 2018).

1.2.4 Microsporidia in aquaculture/agriculture

Besides the human-to-human transfer of Microsporidia and through environmental sources, like unsterilised drinking water, a possible route to human infection is through ingestion of infected tissue (G. D. Stentiford *et al*., 2016). Microsporidia have been described in livestock as early as 1969, in aquaculture (salmonid fish (PUTZ *et al*., 1965)) and from 1999 in agriculture (cattle (Halánová *et al*., 1999)and swine (Jeong *et al*., 2007)). With the expansion of

aquaculture and greater access to pathogen screening tools and methods, the last two decades have seen an increase in described cases of microsporidiosis. Microsporidia have been found to affect livestock in many different aquaculture methods. In net pen (off-shore farming of fishes in net-based enclosures) three microsporidian species, in particular, have become associated with mortality in the farming of salmonid fish: *Nucleospora salmonis* infects the hematopoietic cells of the Atlantic salmon (*Salmo salar*) and Chinook Salmon (*Oncorhynchus tshawytscha*), causing proliferation of cells which can lead to anemia (Hedrick *et al*., 2012); *Paranucleospora theridion* infects the epithelial cells of Atlantic salmon and has been associated with 80% mortality in some farms, however, its role in mortality has not yet been confirmed (Nylund *et al*., 2010); finally, *Loma* (formerly *Pleistophora) salmonae* infects the endothelial and pillar cells of the gills in Chinook salmon, which leads to inflammation and can cause proliferative bronchitis (Becker & Speare, 2007). While microsporidian pathogens are still described in semi-intensive and intensive farming in fish (*Glugea spp*. (Naghashyan *et al*., 2018), for example, *Pleistophora hyphessobryconis* (Winters *et al*., 2015)), Microsporidia are largely a problem for farmed crustacea (*Thelohania spp*. (Bacela-Spychalska *et al*., 2018; Grabner, 2017; Moodie *et al*., 2003; Voronin, 1986), *Nosema spp*. (Quiles *et al*., 2019; Voronin, 1986), *Perezia spp*. (Sokolova & Hawke, 2016; Voronin, 1986) and *Pleistophora spp*. (MacNeil *et al*., 2003)).

1.2.5 Microsporidia used as biological control

Vector-borne pathogens have had a large impact on human health, with mosquito-based malarial death as high as 435,000 people in 2017 (CDC, 2019), and that is just one of the many highly impactful pathogens that mosquitoes are known to carry in one vector species. Due to the current and possible future burden of vector-borne pathogens (Dorn *et al*., 2017), and rising resistance to insecticides (Denholm & Devine, 2013; Ranson *et al*., 2011), a relatively new research field is being explored, looking at biological ways to control vector populations and thus control the spread of vector-borne pathogens. There are a number of biological control methods used in controlling vectors (and thus vector-borne pathogens): using natural competitors (predators or competitors in the same ecological niche) to control vector density (Crowder *et al*., 2019), vector

transgenesis, a method used to genetically modify a vector species making them incapable as operating as vectors (Shane *et al*., 2018), and a promising route of study investigating the effect other microbes have on transmission (Frentiu *et al*., 2014; Moreira *et al*., 2009; Walker *et al*., 2011). The use of Microsporidia as biological vector control has been considered for the last five decades (Becnel & Andreadis, 2014; Henry, 1971), as their infections can cause population decline and help in the regulation of pests (Becnel & Andreadis, 2014). Studies to this effect have been noted for pest grasshopper species (Henry, 1971) and moths (Lynch & Lewis, 1976). Some research has also shown that vector-pathogen relationships can be affected by symbiotic Microsporidia. A recent study shows a large decrease in *Plasmodium falciparum* prevalence in the presence of *Microsporidia MB* within *Anopheles gambiae* (Herren *et al*., 2020).

1.2.6 Microsporidian genome evolution

With the advances in sequencing technology and increased interest in the phylum, the rate at which microsporidian genomes are being sequenced is gaining speed. The first complete genome that was sequenced was that of *Encephalitozoon cuniculi* in 2001 (Katinka *et al*., 2001). A large source of interest in the Microsporidia is their reduced genomes; some species have some of the smallest known genomes of any eukaryote ranging from 2.3 Mb (*Encephalitozoon intestinalis*) - 51.38 Mb (*Edhazardia aedis*) (Desjardins *et al*., 2015). Following the trend of many other obligate intracellular parasites, Microsporidia have reduced their genomes, instead, relying upon the host for many of its requirements, such as metabolic and transcriptional regulator pathways (Katinka *et al*., 2001; P. J. Keeling & Slamovits, 2004b; McNamara-Bordewick *et al*., 2019; Vivarès *et al*., 2002). As a result, genes found within the Microsporidia have been shortened with a loss of most introns (Katinka *et al*., 2001), intergenic regions have been shortened or removed (P. J. Keeling & Slamovits, 2004b), as well as a reduction in duplicated segments (Katinka *et al*., 2001)and (retro)transposable elements (Heinz *et al*., 2012; Peyretaillade *et al*., 2012). The presence/absence of these non-coding regions in the genome account more for genome size variation within Microsporidia than gene counts/size (Cuomo *et al*., 2012). As an example, the genome size between *Enc. intestinalis* (2.3 Mb) and *Ed. aedis* (51.38 Mb) is greater than 20-fold, while the

difference in gene count is just above 2-fold (1,848 genes and 4,190 genes, respectively). Genes that remain are conserved genes associated with basic cellular functions; genes that are highly expressed in yeast and present throughout the Eukaryota, as a reduction of these genes would have negative effects on the organism (Nakjang *et al*., 2013). Recent genomic analysis has suggested that a bottleneck in the last common microsporidian ancestor (LCMA) may explain the genome reduction having lost 1,590 protein families (Nakjang *et al*., 2013), suggesting that only a small set of conserved genes made it through the bottleneck, with subsequent lineage-specific gene-family expansions. The lineage-specific gene-family expansions are shown by the novel gene families that are present in a large proportion in assembled microsporidian genomes (Nakjang *et al*., 2013). Another contributing factor to the reduction of the nuclear genome is the loss of a functional mitochondrion, and with it, the loss of many of the genes within the metabolic pathway that allow them to produce ADP/ATP (Embley & Martin, 2006). Instead, microsporidian genomes are rich in transporter proteins that allow them to take resources that they need from the host (P. J. Keeling & Corradi, 2011; Nakjang *et al*., 2013; Wiredu Boakye *et al*., 2017).

Likely contributing to the rapid lineage-specific gene-family expansions are the high mutation rates attributed to the Microsporidia. Higher mutation rates would have a greater probability of producing errors in DNA replication and repair machinery, which would, in turn, produce more gene duplications, and gene duplications are strongly associated with gene diversification (Costello *et al*., 2020; Hoekstra & Coyne, 2007). Mutation rates are estimated to be high in Microsporidia due to the lack of tumor-suppressing genes that repair DNA mutations (Haag *et al*., 2014) and a reduction in DNA repair genes has also be noted in *Enc. cuniculi* (Gill & Fast, 2007; Haag *et al*., 2014; P. J. Keeling & Slamovits, 2004a). These trends have also been found to be associated with reduced genomes brought about by population bottle-necks in bacteria (Nilsson *et al*., 2005).

1.3 Economically important parasites, Enterocytozoonidae

There are a number of key families and genera within the Microsporidia that have garnered more interest than others, due to their economically important hosts. The *Nosema* genus has long received attention due to the role they played in the collapse of the silkworm industry (Canning *et al*., 1999; Kellen *et al*., 1977; Solter *et al*., 2002) and their current association with the colony collapse disorder in honeybees (Antúnez *et al*., 2009; Bromenshenk *et al*., 2010; Pettis *et al*., 2012). *Encephalitozoon* has three species that infect humans, as well as other mammals, rabbits among them. Both of these genera have been well researched for a number of decades, but more recently a family of predominantly aquatic pathogens (Enterocytozoonidae) has emerged as a group of interest. The family Enterocytozoonidae comprises ten described species within a monophyletic clade of microsporidians that mostly infect crustaceans and fish (Palenzuela *et al*., 2014), with the exception of *Enterocytozoon bieneusi*, which has only been observed infecting terrestrial vertebrates. Conversely, the Enterocytozoonidae are placed in a larger clade of microsporidian species ("Terresporidia") that infect terrestrial hosts (Vossbrinck *et al*., 2014; Vossbrinck & Debrunner-Vossbrinck, 2005). This is particularly interesting, as the majority of described enterocytozoonids infect aquatic hosts, suggesting the Enterocytozoonidae are derived in their selection of host. Although the classification of this group is largely through molecular data, all currently described species have enlarged sporogonial syncitia, a structure in which the polar filament develops prior to sporoblast division (Freeman & Sommerville, 2009). Also, in all species, infection has been found specifically in the gastrointestinal tract, and homologous structures (G. D. Stentiford *et al*., 2019). The reason for the raised awareness of this group is its emergence into species of economic importance. *E. bieneusi* (infections in livestock and humans) and *Nucleospora salmonis* (infections in farmed salmon) were the two first described species and have been described for a number of decades, however, subsequent species have continued to be described in farmed fish and crustacea (Freeman *et al*., 2013; G. D. Stentiford *et al*., 2007; Sveen *et al*., 2012; Tourtip *et al*., 2009). Though none of the infections have been reported as fatal, they siphon energy away from the host likely leading to more serious conditions, sometimes leading to a reduction in growth (Tourtip *et al*., 2009). They have been described as sentinels of animal health because of this (G. D. Stentiford *et al*., 2019). This family, and other Microsporidia that infect

livestock, likely pose the greatest threat to commercial fisheries, where fish and crustacea are reared in high densities (Foltz *et al*., 2009).

1.3.1 Enterocytozoonidae phylogeny/diversity

Broadly speaking, the Enterocytozoonidae are split into two clades, the crustacean/terrestrial vertebrate infecting clade (*Enterospora/Enterocytozoon*) and the fish (occasional crustacean) infecting clade ((*Para)Nucleospora/Obruspora*). The most researched enterocytozoonids reside within the former clade, whereas many of the fish infecting species have only been noted in an initial description paper. The only species in the former clade to receive a moderate amount of research is *N. salmonis* (El Alaoui *et al*., 2006a; Foltz *et al*., 2009; Sakai *et al*., 2009). *Hepatospora eriocheir* is tentatively included within the Enterocytozoonidae in some studies (including this thesis), although its intermediate features and greater genetic divergence suggest a more basal positioning among the Enterocytozoonidae. Some researchers have erected an intermediate clade between that of the Enterocytozoonidae and the *Enterocytospora*-like species called the Hepatosporidae. Currently, the Hepatosporidae is composed of *H. eriocheir*, *Parahepatospora carcini,* a recently discovered pathogen *of Carcinus maenas* (the European shore crab) (Bojko *et al*., 2017), and an even more recently described, but unnamed, hepatosporid found in Nemertea (ribbon worms) (Robbins *et al*., 2021).

Variation within species or strains/genotypes is not well understood within the Microsporidia, due to the aforementioned difficulty in identifying host ranges for species and difficulty in culturing many of the species of interest. As mentioned, some work done on two species of enterocytozoonid, *N. salmonis* and *E. bieneusi,* has shown a multitude of host species. Sequencing of specific (ssu-rDNA and the ITS gene respectively) markers in these two species have shown a wide diversity for both species (El Alaoui *et al*., 2006b; Leelayoova *et al*., 2006). *N. salmonis* has received less research on this and a different, more conserved, marker is used, so it shows less variation than *E. bieneusi*.

1.3.2 Transmission routes within the Enterocytozoonidae

So far, transmission in the Enterocytozoonidae has been largely inferred to be horizontal, and in the case of studies looking at the transmission in *E. hepatopenaei*, a pathogen of *Penaeus vannamei* (white legged shrimp), it has been proven in a number of studies to be transmitted horizontally in ponds (Karthikeyan & Sudhakaran, 2019b; Salachan *et al*., 2017a). Though the mode of transmission is not directly known, some species have also been found to parasitise both the crustacean parasites as well as the fish host (Diamant *et al*., 2014; Freeman & Sommerville, 2009; Nylund *et al*., 2010). This may indicate a route of transmission to fish being through parasitic, crustacea, as their first two larval stages (nauplius 1 and 2) are planktonic (Johnson & Albright, 1991) and would be widely dispersed by marine currents and likely to come into contact with pelagic fauna.

1.3.3 Evolutionary novelties: Cellular and nuclear parasites

Besides the economic importance of the family, it has also received a lot of interest for some evolutionary novelties within the Microsporidia if not throughout the animal kingdom. Outside of the Enterocytozoonidae, only one species has been described to develop within the nucleus, whereas there are six enterocytozoonids (*Enterospora nucleophila* (Palenzuela *et al*., 2014), *Enterospora canceri* (G. D. Stentiford *et al*., 2007), *Paranucleospora theridion* (Nylund *et al*., 2010), *Nucleospora cyclopteri* (Freeman *et al*., 2013; Mullins *et al*., 1994), *Nucleospora salmonis* (Chilmonczyk *et al*., 1991) and *Nucleospora secunda* (Lom & Dykoá, 2002)). No work has been published looking at the genomic basis for this, and it promotes further biological queries. For example, given the high number of intranuclear species within the Enterocytozoonidae, on two separate clades within the family, is this an ancestral state that has been lost in some lineages or gained separately, and what genomic processes help facilitate this.

1.3.4 Evolutionary novelties: Loss of glycolysis

One of the more unique characters of the Enterocytozoonidae is their puzzling loss of the glycolytic pathway (Wiredu Boakye *et al*., 2017). A further reduction in

metabolic pathways from a lack in the ability for Microsporidia to produce their own ATP, largely reliant on the host via glycolysis and ATP transport that accounted for 7% of ATP intake (Berg *et al*., 2007; Richards *et al*., 2003; Tsaousis *et al*., 2008). The primary interest with this further loss is that it is unknown how enterocytozoonids could gather enough ATP to function, and how do they do so extracellularly. So far, only four enterocytozoonid genomes have been sequenced (*H. eriocheir, E. bieneusi, E. hepatopenaei,* and *En. canceri* (Desportes *et al*., 1985; G. D. Stentiford *et al*., 2007, 2011; Tourtip *et al*., 2009)*).* These sequenced genomes are of those enterocytozoonids that infect the most economically relevant hosts. Their genomes represent some of the smallest described Microsporidia, ranging from 3.1 Mb to 4.57 Mb. All published genomes, besides *H. eriocheir*, are from the *Enterocytozoon/Enterospora* clade. Though it is also likely that the fish infecting clade *((Para)Nucleospora/Obruspora*) has a similar loss of glycolytic pathways and a similarly reduced genome. This has been suggested as *H. eriocheir* has a more basal relationship within the Enterocytozoonidae (likely part of a sister clade) and shares in the loss of glycolytic characters. However, *H. eriocheir* only has a partial loss of the glycolytic genes that the remaining enterocytozoonids display (Wiredu Boakye *et al*., 2017).

1.4 Aims and Objectives

The overall aim of this thesis is to assess the potential risk the Enterocytozoonidae poses to aquaculture, and to a lesser degree, agriculture, looking at the factors that may facilitate their emergence into our current or future livestock. This is important, as, with a growing global human population, there will be a need for growth in stable sources of food (Jennings *et al*., 2016; G. D. Stentiford *et al*., 2013). To help address this, this study was undertaken to aid in predicting new pathogenic species, which may arise either through zoonotic potential or expanding farms into environments that harbor potentially pathogenic organisms. Identification of possible threats and routes of transmission will allow for appropriate precautions to be made. This study approached this in three ways:

- This was first investigated by screening the environment for the presence of enterocytozoonids. This has not been done before, and an understanding of how widespread the Enterocytozoonidae are will play a part in enabling

inferences about the pathogenic threat of this family; based on their overlap with human influenced areas, such as farms and thus their likelihood to come into contact with humans of farm life. This was investigated over two data chapters, (Chapter 2 and 3) both exploring the hypothesis that the Enterocytozoondiae are more diverse than previously described. Chapter 2 also addresses the use of the highly variable inter transcribed spacer (ITS1) region for assessing genotypes in *E. bieneusi.*

- The second way this study aims to assess the risk of the Enterocytozoondiae is to investigate the possible transmission routes that facilitated the spread of *E. hepatopenaei*. In doing so, ascertain the recent geographic origin of *E. hepatopenaei* in shrimp ponds and highlight the ways in which undescribed enterocytozoonids may also infect farmed organisms. The hypothesis being tested is that the pathogen is endemic and naturally widespread but has not been noticed due to lower levels of infection (pathogenicity/ virulence), and a change in farming practices/ genetic variant has allowed the pathogen to become prevalent. An alternative hypothesis is that it represents a recent spread epidemic that has quickly become widespread. In either scenario, this will enable farmers of the best practices to stop the spread and inhibit their fast expansion. This chapter is a good opportunity to showcase what should likely be the next steps after the first and second data chapters, where possible pathogens have been identified before they have become a problem for aqua/agriculture.

- The third way this study aims to assess the risk of enterocytozoonids is to investigate how readily the family has host switched in their lineage, looking at the genes that facilitated such an event. In so doing, this will help to predict the likelihood of further host-switching events. A secondary aim of this chapter is to look into the evolution of the small genome size within the Enterocytozoonidae. Although the Microsporidia, as a clade, display small genomes, some groups have taken this further. The Enterocytozoondiae have some of the smallest describe genomes, however, the evolution of this family has not been studied.

- The overall aim of Chapter 6 is to bring together the results from the preceding 4 data chapters to make an assessment, based on the results, on the risk that the Enterocytozoonidae pose to aquaculture and human health. This

will enable risk assessments to be made on ways in which to mitigate the risk of this clade, enabling farmers to put in measures to minimise economic loss.

Chapter 2: Investigating the diversity of the Enterocytozoonidae using metagenomics

2.1 Introduction

Metagenomic analysis has been rising in use and application since 1998 (Handelsman *et al*., 1998) and has revolutionised the way that ecological studies are carried out. Metagenomics has allowed for quicker more cost-effective ways to run ecological studies on diversity (Deiner *et al*., 2017) or community assemblages/associations (Willis & Gabaldón, 2020), primarily on microorganisms. Metagenomics is also used outside of ecological studies, now widely applied to medical (cancer (Purcell *et al*., 2017) and immunological (Duplouy *et al*., 2020)) and agricultural (gut microbiome associated studies (Davis *et al*., 2016)) research. This has led to a rise in publicly available metagenome/metabarcoding databases, with more data from sequencing runs being produced than is needed to answer initial questions. This has allowed researchers to expand the type of research being undertaken and the range of data used (Donovan *et al*., 2018), making more data accessible to a wider group of researchers.

Metagenomic analysis has contributed hugely to biodiversity studies in a number of applied ways and is increasingly being applied to help solve/mitigate anthropogenic problems. Among many other ways, they are used in conservation, to investigate indicators of healthy environments (Aylagas *et al*., 2014; Vasselon *et al*., 2017) and are vital in understanding the role of microorganisms in carbon, nitrogen, and other biogeochemical cycles (Hölker *et al*., 2015; Treonis *et al*., 2018). However, with the rise of high-throughput sequencing in the past 15 years (Lightbody *et al*., 2019; Reuter *et al*., 2015), there is a greater push for use of metagenomics in many of the studies stated above. Diagnostic metagenomics, a term coined by Pallen (Pallen, 2014), is the use of shotgun metagenomics in the discovery and detection of pathogens. The term was primarily concerned with clinical samples, though it is also applicable to non-human samples too (Studholme *et al*., 2011). Diagnostic metagenomics has been used heavily to study organisms that are difficult to culture, so heavily used in virus discovery (Alavandi & Poornima, 2012; Schulz *et al*., 2018). It has been

used to catalogue, in humans, a normal virome of species that have not been associated with disease (Popgeorgiev *et al*., 2013). This has and will allow future research to use this as a baseline for any work looking at human-based viral infections that are not a part of the 'healthy' human virome. In the case of environmental pathogens, this same method could be used to assess the potential threats of pathogens to livestock. Cataloguing endemic endoparasitic species in an environment would allow for better assessments on potentially pathogenic, and invasive species and allow for investigation of ecological patterns. Pathogen discovery in the environment would also be informative to placement of farms and fisheries, as the presence of particular pathogens is likely to influence such decisions.

Due to greater variation often found in universal markers (Tedersoo *et al*., 2015), Microsporidia-based sequences tend not to get amplified in eukaryote-wide metabarcoding studies; so Microsporidia-specific primers are often needed (Ardila-Garcia *et al*., 2013). Even shotgun genomics may be inadequate unless specialized extraction methods are used, as microsporidians have a chitin-based cell wall, structures not easily ruptured by lysis buffers alone. Meaning that many metagenomic/metabarcoding studies are likely missing Microsporidia diversity. For Microsporidia, large-scale metabarcoding studies are lacking; an oversight as they likely have a great influence on ecological networks, like many pathogens. Microsporidia have been identified in some indirect studies from host-based metagenomics: in honeybees, two of four fungal species (18s rDNA) (Cox-Foster *et al*., 2007), in giant pandas, where they made up 2% of the fungal diversity (Yang *et al*., 2018) and in pigs (Donovan *et al*., 2018) to name a few. However, host-based metagenomic studies miss any environmental occurrences of the pathogen, an equally important part of the parasite's life cycle. Especially if the microorganism is not an obligate parasite, and can feed and reproduce outside of a host. Single host-based metagenomics can also be more costly in terms of time and money to get a full picture of their distribution and abundance. More costly in terms of time, as taxonomic specific trapping techniques will need to be used to acquire host species, whereas environmental samples require less taxonomic expertise and time.

Many widespread obligate pathogens require intermediate hosts for transmission into their 'main' hosts, to fulfill their lifecycle (Balloux & van Dorp, 2017; Moore,

2010). Planktonic crustacea have been found to host many species of Microsporidia that have been described in a 'main' host (Palenzuela *et al*., 2014; Sveen *et al*., 2012; Vossbrinck *et al*., 1998), indicating that planktonic crustacea may play a role as intermediate hosts (Vossbrinck *et al*., 1998) promoting host-shifting (Quiles *et al*., 2019). Intermediate hosts have also been suggested to play a part in the radiation of Microsporidian species, as a result of the host going through diversification (Bacela-Spychalska *et al*., 2018).

There are currently ten species of Enterocytozoonidae described: *Enterocytozoon bieneusi* (Desportes *et al*., 1985), *Nucleospora salmonis* (Chilmonczyk *et al*., 1991), *Nucleospora secunda* (Lom & Dykoá, 2002), *Enterospora canceri* (G. D. Stentiford *et al*., 2007), *Paranucleospora theridion* and its synonym *Desmozoon lepeotherii* (Freeman & Sommerville, 2009; Nylund *et al*., 2011), *Enterocytozoon hepatopenaei* (Tourtip *et al*., 2009), *Nucleospora cyclopteri* (Freeman *et al*., 2013), *Enterospora nucleophila* (Palenzuela *et al*., 2014), *Obruspora papernae* (Diamant *et al*., 2014) and *Nucleospora braziliensis* (da Cunha *et al*., 2017). The only species with no 18s rDNA sequence data is *N. secunda*. The number of described species has more than doubled in size in just over a decade. Such a rapid discovery of species is suggestive of a greater diversity yet to be discovered in this family. This is a situation true of the Microsporidia as a whole (Ardila-Garcia *et al*., 2013). There is already evidence for greater diversity from several species that have enterocytozoonid-like 18s rDNA sequences, but no histology to verify (Arundell *et al*., 2015). Members of this family have a wide geographic distribution, having been described on every major continent, and throughout much of the Atlantic. But they have so far only been described in marine and estuarine environments, though there are a number of 18s rDNA sequences that are closely related to the Enterocytozoonidae but have been from freshwater (JN938583.1; KT777455.1-unpublished) (Arundell *et al*., 2015).

The most researched enterocytozoonid is *E. bieneusi*, with ten papers published on their genotypes in sheep alone since 2015 (Fiuza *et al*., 2016; Jiang *et al*., 2015; W. Li *et al*., 2014; Shi *et al*., 2016; Ye *et al*., 2015; Zhao *et al*., 2015). It is the most prevalent human infecting microsporidian (Akiyoshi *et al*., 2009; Widmer *et al*., 2013), though mainly an opportunistic parasite in this respect, mostly infecting the immunocompromised (W. Li *et al*., 2017). It infects the

intestinal epithelial cells, causing AIDS-associated diarrhoea (Desportes *et al*., 1985). It appears to be highly opportunistic in nature, having also been found to infect a plethora of other terrestrial vertebrates. These include live-stock (Mathis *et al*., 1999; Zhao *et al*., 2014), pets (guinea pigs, cats and dogs) (Cama *et al*., 2007; Santín & Fayer, 2011; Y. Zhang *et al*., 2019), and wild animals (Santín & Fayer, 2011). It has a very wide distribution, having been described on major continents, ranging from Africa to China and South America (Espern *et al*., 2007; Rinder *et al*., 2000). Though no natural reservoirs or routes of infection have been found, some studies have inferred infection from pets (Cama *et al*., 2007).

So far, *E. bieneusi* has been described as having over 240 genotypes that have been sorted into nine genotype groups that have been associated with a host, using the ribosomal ITS1 region (Q. Zhang *et al*., 2018). Genotype groups appear to be homoplastic in host choice, in that a genotype group will have genotypes from different, phylogenetically distant host species. However, many appear to be host-specific and have only been found to infect the organisms in which they were first described. There are several different human infecting genotypes (Akiyoshi *et al*., 2009; Widmer *et al*., 2013), unsurprising, considering the number of domesticated animals infected, it is likely that humans would encounter many different genotypes. *E. bieneusi* is not the only enterocytozoonid to have been shown to have a propensity for opportunistic transmission. Both *P. theridion* and *O. papernae* infect the copepod parasite of their respective "main" hosts (Diamant *et al*., 2014; Gunnarsson *et al*., 2017; Sveen *et al*., 2012), whereas *N. salmonis* infects several species of fish (El Alaoui *et al*., 2006a; Hedrick *et al*., 2012; Sakai *et al*., 2009). *E. bieneusi* has also been found in mussels in the River Shannon, Ireland, though the infection was not verified with histology, so it could just be a result of filtering water. It is likely that many more unknown enterocytozoonid infections will be attributed to described species.

Despite a great amount of research being carried out on the family (relatively), no work has been carried out investigating the diversity, distribution, or ecological role of the group. Primarily, descriptions of infections are in host species of economic importance, likely underestimating vast biodiversity in host species that are not farmed. The main aim of this study was to test the hypothesis that the Enterocytozoonidae were more diverse than currently described, using pre-existing data, and to better understand the breadth of host types and

environments that they might inhabit. Existing metagenomic databases, like the National Center for Biotechnology Information (NCBI) and The Integrated Microbial Genomes and Microbiomes (IMG), are a great resource for studying the distribution of intracellular parasites, as the size and life cycle make it difficult and time-consuming to identify specimens by microscope analysis. It also allows for the investigation of worldwide patterns in a more time and supply-efficient manner.

Another aim of this chapter was to look at the diversity found in the *E. bieneusi* genome and assess the appropriateness of using ITS as a marker to genotype "strains" of the species found in different hosts.

2.2 Methods and analysis

2.2.1 Metagenome screening

To test the hypothesis that the biodiversity of the family is greater than described, publicly available metagenome databases were screened for matches to find evidence of enterocytozoonid-like sequences. The database predominantly used was the Integrated Microbial Genomes & Microbiomes database (IMG/MER) (Markowitz *et al*., 2012), which has an in-built BLASTN function (Altschul *et al*., 1990). An *E. hepatopenaei* 18s rDNA sequence (KF362129.1) was used as the query for BLASTN searches against different datasets, focusing on freshwater, marine, and aquatic invertebrate metagenomes. Terrestrial soil /invertebrate metagenomes were also screened, to get a better understanding of the distribution of Enterocytozoonidae. However, it was expected that the aquatic metagenomes would provide the majority of the positive enterocytozoonid sequences. The E-value threshold was set to $\leq$ 1E-50, as searches with a less strict E-value recovered non-enterocytozoonid-like nucleotide sequences. Though this may not be representative of the group, without histology, sequences omitted this threshold cannot be confidently included.

2.2.2 Data sets, alignment, and tree building

Due to the varying lengths and locations of the recovered enterocytozoonid-like scaffolds relative to the 18s rDNA gene, three different datasets spanning different regions within the 18s rDNA region, with no overlap, were made (start, mid, end of 18s rDNA) to get a better idea of the biodiversity found in these datasets (Table 2.1). All sequences were initially aligned using MAFFT version 7 (Katoh *et al*., 2018) under default parameters, viewed and trimmed by eye using Bioedit version 7.2.5 (Hall, 2013).

Table 2.1: Three metagenomic datasets, split across the 18s rDNA to cover varying lengths of metagenomic scaffolds

Dataset	Aligned length	Number of sequences	Phylogenetically informative sites	Conserved sites
600-980	380	37	34.2%	39.5%
985-1470	487	55	36.3%	44%
1470-2109	540	26	36.5%	44.1%

To ensure that the different datasets did not vary greatly in the phylogenetic signal that they gave, levels of phylogenetic signal were looked at in MEGAX (Kumar *et al*., 2018). Though this would not inform on whether these gave signals for the same evolutionary relationship, it would show any datasets that had much lower/higher levels of signal. Once trimmed into non-overlapping datasets, they were aligned again, using MAFFT ver 7 under default parameters and checked by eye using Bioedit ver 7.2.5. Phylogenies were built using maximum-likelihood implemented by RAxML-HPC version 8.2 (Stamatakis, 2016). The tree was built using the General Time Reversible rate of heterogeneity, with Gamma distribution (GTR-GAMMA) assessed with 1000 bootstrap replicates. The substitution model of best fit was assessed using MEGAX, by selecting the model with the lowest Akaike information criterion (AIC). Reference Enterocytozoonidae 18s rDNA sequences from all ten described species were also added to the analysis.

To investigate any relationship between genetic distance and geographic distance, and to look at the divergence of the novel sequences from the described

species, pairwise distance was assessed for each dataset using the Tajima-Nei model (Tajima & Nei, 1984) implemented in MEGAX. The correlation between genetic distance and geographic distance between sites was tested using Mantel tests implemented in QIIME1 version 1.9.1 assessed over 1000 permutations (Caporaso *et al*., 2010).

2.2.3 Operational Taxonomic Units (OTUs) and clustering

To get a better understanding of the relative abundance of enterocytozoonid-like sequences, compared to that of other microsporidian clades, all novel sequences (725 sequences) were grouped based on sequence identity. Datasets were pooled and clustered using USEARCH version 11 (Edgar, 2010), to form operational taxonomic units (OTUs) based on a sequence identity of 97%, to get an estimate of the number of 'species' present in the metagenome. Although the sequence identity cut off of 97% is unlikely to represent the species delineations found within the Microsporidia, it offers a good starting point to start explaining the diversity and ecology of understudied groups. The representative OTUs were then clustered, using UCLUST version 11 (Edgar, 2010), with 43 18s rDNA reference sequences that were picked from throughout the known microsporidian diversity (Table 2.1-supplementary). A threshold of 80% nucleotide sequence identity was chosen to look at the diversity of species within family-like clusters, used as a proxy to see how diverse the enterocytozoonids are relative to other microsporidian "families". Eighty percent was chosen, as this is the lowest sequence similarity found between described Enterocytozoonidae. This includes *H. eriocheir*, a species sometimes regarded as a sister/basal enterocytozoonid. This analysis was also carried out on four smaller datasets for the same reasons mentioned above, as many of the shorter sequence fragments may falsely inflate the number of OTUs. Some diversity may be lost, but it is a precaution against false inflation of diversity.

2.2.4 ITS as a marker for genotyping E. bieneusi

To look at the suitability of ITS1 as a marker for genotyping *E. bieneusi*, the *E. bieneusi* reference genome was downloaded (GCA_000209485.1) from NCBI. This is currently the only assembled genome available for *E. bieneusi*, and it is

based on spores isolated faecal samples from a single infected human. To test the variability of ITS1 of and *E. bieneusi* within a host, bioinformatically, BLASTN (Altschul *et al*., 1990) was used with an ITS1 sequence (AY237209.1) as the query against the published *E. bieneusi* genome, using CLC Genomics Workbench 11.0 (https://www.qiagenbioinformatics.com/). Any hits greater than 1E-10 were taken and aligned with the published dataset the query sequence came from (Baroudi *et al*., 2018). As variation is being looked at, the E-value cannot be too strict, otherwise, possible variants would be missed. To help make a direct comparison with the current diversity of *E. bieneusi* genotypes, 42 sequences that were used in a recently published tree (Baroudi *et al*., 2018), were downloaded and used in the analysis (Table 2.2-supplementary).

2.2.5 Genetic/Genomic diversity

To look at the genetic distance between the ITS1 sequences from the genome, as above, Tajima and Nei's pairwise distance was assessed using MEGA X. To also give a measure of genomic diversity found among *E. bieneusi* spores within a host, as a comparison to the ITS1, the raw reads (Leinonen *et al*., 2011) for the genome were mapped to a genome-wide subset of 128 single-copy genes (refer to OrthoMCL method Chapter 5) from the *E. bieneusi* reference genome, using BWA MEM version 0.7.17 (H. Li & Durbin, 2010)under default parameters. The raw reads were filtered using Sickle version 1.3 (Joshi & Fass, 2011) using default parameters, SNP calling was done using GATK version 4.1.8.0 (McKenna *et al*., 2010) best practices. Intrapopulation diversity was estimated using nucleotide diversity ($\boldsymbol{\pi}$) as a proxy, estimated using VCFtools version 0.1.16 (Danecek *et al*., 2011). This will give an estimate of how diverse the genomes are, relative to the pairwise estimates for ITS sequences.

2.2.6 Tree building

Alignment and tree building was carried out using the same parameter values as described in the Baroudi *et al.* (2018) paper, to avoid any variations that may result from using an alternative alignment and tree building algorithm. The final dataset, including the ITS sequences from the BLASTN analysis and the

published Baroudi *et al*. (2018) dataset. The tree was visualised in TreeGraph 2.15.0 (Stöver & Müller, 2010).

2.3 Results

430 metagenome projects comprised of 7,641 datasets were screened using an *E. hepatopenaei* 18s rDNA query sequence using BLASTN. 176 from aquatic environments (2751 datasets), 128 from terrestrial environments (4043 datasets), and 126 from host-specific metagenomes (840 datasets). This identified 46 metagenome projects (29 from aquatic metagenomes, 12 terrestrial metagenomes (Figure 2.1), and 0 host-associated metagenomes) (Table 2.2) containing scaffolds sharing at least 75% identity to the 18s rDNA query sequence and eight metagenome projects shared at least 85% identity to the query sequence.

Figure 2.1: Locations of origin for metagenomic datasets with sequences highly similar (BLASTn ≤ 1E-50) to a *E. hepatopenaei* 18s rDNA sequence. (Yellow: marine; Blue: freshwater; Green: terrestrial)

Table 2.2 Positive blast metagenomes screened using the IMG database, *E. hepatopenaei* 18s rDNA as query

IMG Study ID	Location	Metagenome type	Sequencing platform	Positives/ Enterocytozoonid-like
Gs0053074	Atlantic Ocean	Marine	Illumina	1/1
Gs0053074	West of El Salvador, Pacific Ocean	Marine	Illumina	17/8
Gs0116197	Illinois, USA	Freshwater	Illumina	22/0
Gs0126301	Trout Lake, Wisconsin, USA	Freshwater	Illumina	8/0
Gs0126301	Sparkling Lake, Wisconsin, USA	Freshwater	Illumina	23/0
Gs0126301	Mendota Lake, Wisconsin, USA	Freshwater	Illumina	39/0
Gs0114433	Delaware River, USA	Estuarine	Illumina	25/12
Gp0156537	Lake Montjoie, Canada	Freshwater	Illumina	7/0
Gs0114818	Cold Stream Run, Pennsylvania	Freshwater	Illumina	4/0
Gs011481	Alex Branch, Pennsylvania	Freshwater	Illumina	4/0
Gs011481	Straight Creek, Pennsylvania	Freshwater	Illumina	5/0
Gs0114433	Chesapeake Bay, USA	Marine	Illumina	41/21
Gs0118433	Oldwoman Creek, Ohio	Freshwater	Illumina	5/0
Gs0053068	Lake Ontario, Ontario	Freshwater	Illumina	1/0
Gs0118430	Lake Croche, Canada	Freshwater	Illumina	9/0
Gs0118430	Lake Simoncouche, Canada	Freshwater	Illumina	9/0
Gs0114443	Lake Erie, Pennsylvania	Freshwater	Illumina	4/0

Gs0103597	Klosterneuburg, Austria	Freshwater	Illumina	49/0
Gs0017769	Twitch Island, California	Freshwater	Illumina	10/0
Gs0063447	Adventfjord, Svalbard Archipelago, Norway, station 2	Marine	454 and Illumina	4/4
Gs0110190	Louisana Shelf, Hypoxic Zone, Gulf of Mexico	Marine	Illumina	20/18
Gs0114292	Southern Atlantic Ocean	Marine	Illumina	2/2
Gs0114292	Southern Atlantic Ocean	Marine	Illumina	10/9
Gs0121595	Arthur Harbor ice station, Antarctica	Marine	Illumina	1/1
Gs0046785	Saanich Inlet, British Columbia	Marine	Illumina	17/7
Gs0053074	Southern Atlantic Ocean	Marine	Illumina	18/11
Gs0121595	Palmer Station, Antarctica	Marine	Illumina	1/0
Gs0114511	Milwaukee, Wisconsin	Freshwater	Illumina	10/0
Gs0126301	Wisconsin, USA	Freshwater	Illumina	30/0
Gs0114516	Oregon, USA	Freshwater	Illumina	2/0
Gs0116197	Asahikawa, Japan	Freshwater	Illumina	10/0
Gs0063124	Bonanza Creek, Alaska	Terrestrial	454, Illumina	5/0
Gs0120350	Kohala Peninsula, Hawaii	Terrestrial	Illumina	1/0
Gs0103008	Amazon Forest, Brazil	Terrestrial	Illumina	1/0
Gs0110119	Angelo Coastal Reserve, California	Terrestrial	Illumina	6/0
Gs0117433	Alaska, USA	Terrestrial	Illumina	21/0

Gs0085736	Massachusetts, USA	Terrestrial	Illumina	2/0
Gs0135153	Indiana, USA	Terrestrial	Illumina	1/0
Gs0053071	Weissenstadt, Germany	Terrestrial	Illumina	265/0
Gs0134627	Abisko, Sweden	Terrestrial	Illumina	10/0
Gs0114298	Alaska, USA	Terrestrial	Illumina	13/0
Gs0135149	Colorado, USA	Terrestrial	Illumina	5/0
Gs0128948	Maridalen valley, Oslo, Norway	Terrestrial	Illumina	7/0

725 (398 from aquatic metagenomes, 337 from terrestrial metagenomes, and 0 from host-associated metagenomes) scaffolds in total were returned from the BLASTN analysis, of these, 92 (92 from aquatic metagenomes, 0 from terrestrial metagenomes, and 0 from host-associated metagenomes) are $\geq$ 90% in identity with eight having a very high identity (94-100%) to the reference *E. hepatopenaei* sequence. Two sequences were 100% identical to the *E. hepatopenaei* sequence used as a query: one from Saanich Inlet, Canada (IMG accession number: Ga0008278_1196390) and one from the Atlantic Ocean, west of Cape Town, South Africa (IMG accession number: Ga0005504_1233069). The percentage of sequences that were highly Enterocytozoonidae like in relation to other microsporidian sequences was 12.7%. The proportion of locations screened that had evidence of enterocytozoonids using these search criteria was 17.4%.

2.3.1 Assessment of suitability for each dataset

The length of the three alignments ranged from 380-540bp long (Table 2.1), and the number of sequences ranged from 26-55 of a possible 92. Due to their short lengths, two datasets each lacked one of the enterocytozoonid representatives: the first dataset (dataset-600) lacks the newest addition, *N. braziliensis*, while the third dataset (dataset-1470) excludes *E. bieneusi*. The percentage of phylogenetically informative sites per dataset ranged from 36.3% - 38.9%, conserved sites ranged from 39.5% - 44.1%. With the datasets having similar levels of conservation and parsimony-informative sites, all were retained for subsequent analysis.

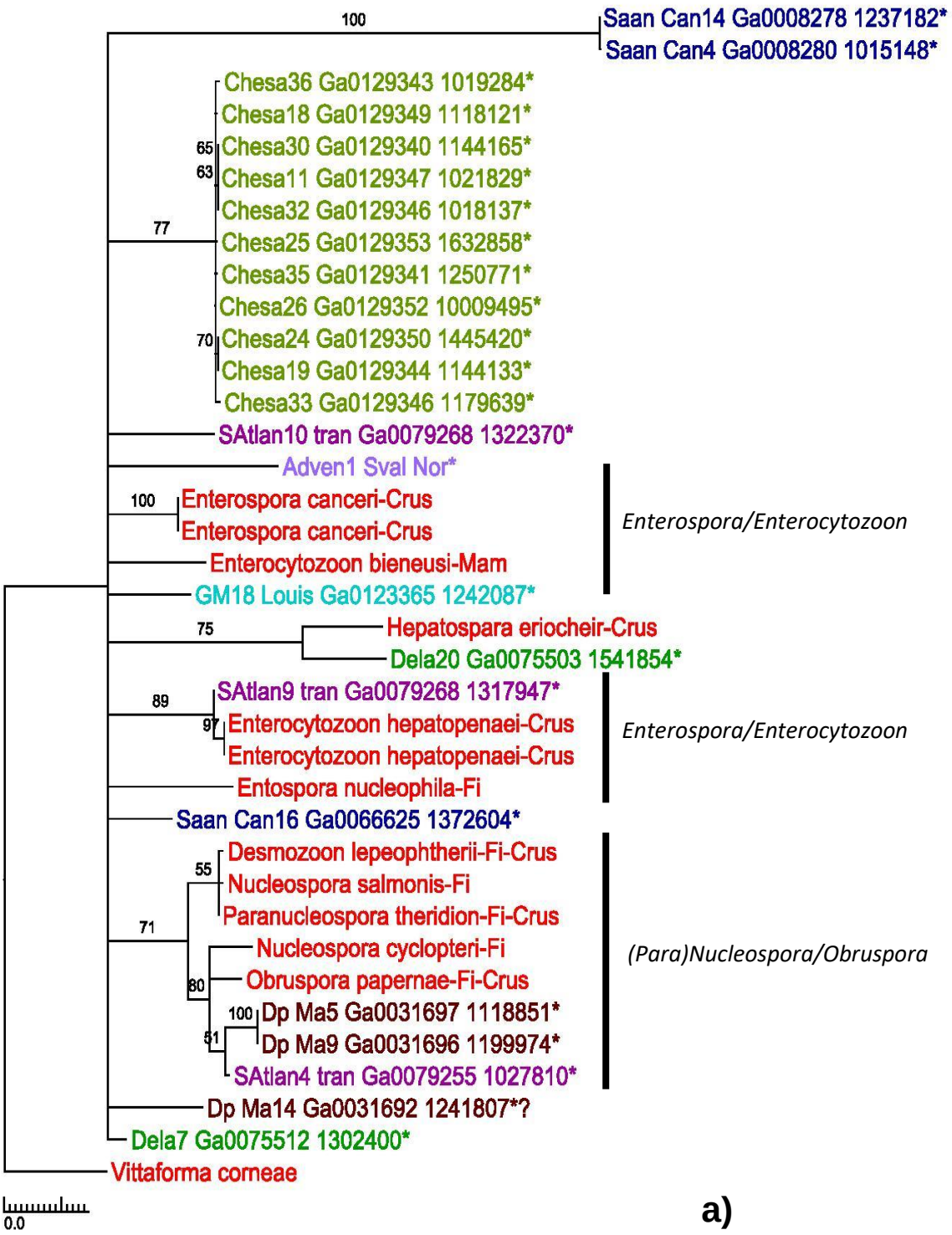
2.3.2 Enterospora/Enterocytozoon clade: more diverse?

The evolutionary relationships in all three data sets displayed in this study concur with those found in previous phylogenies (Figure 2.1a-c). Forming two clades, *En. canceri* and *En. nucleophila* group with the *Enterocytozoon* species and the *Nucleospora* species groups more closely with *P. theridion* and *O. papernae*. However, in the first and third datasets (Figure 2.1a, c), the midlevel branch support was low and has collapsed the *Enterocytozoon/Enterospora* genera. The novel enterocytozoonid sequences identified in this study were phylogenetically diverse and were distributed throughout the Enterocytozoonidae phylogeny. They were found on both enterocytozoonid clades and have also produced two novel branches within the *Enterocytozoon/Enterospora* clade (Figure 2.1b), a mix of sequences from the Chesapeake Bay and the Gulf of Mexico. There was some phylogeographical grouping when looking at the novel sequence diversity, with the samples from the Gulf of Mexico (GM Louis) and Chesapeake Bay (Chesa) mainly grouping together. Whereas other datasets appear more diverse; the Saanich Inlet (Saan Can) and the South Atlantic Transect (SAlantran) ranging throughout *Enterocytozoon/Enterospora* clade and (*Para*)*Nucleospora/ Obruspora* clade (just SAtlantran). The majority of the newly discovered diversity is in the *Enterocytozoon/Enterospora* clade. Pairwise distance estimates for the latter dataset (dataset 1470) also support the relationship of the novel sequences being closer to the *Enterospora/Enterocytozoon clade* (6/8 datasets) (Table 2.3c). However, the former two datasets show more of an even split between the two clades (5/8 and 4/8 datasets respectively) (Table 2.3a,b). Average pairwise distance between the metagenomic scaffolds and *Enterospora*/ *Enterocytozoon* clade over the three datasets is 0.154 (range 0.064-0.258), 0.163 (range 0.087-0.220), 0.191 (range 0.104-0.475). While the average pairwise distance between the metagenomic scaffolds and the (*Para*)*Nucleospora/Obruspora* clade was greater for the first and last datasets being 0.167 (range 0.117-0.306), 0.157 (range 0.094-0.225), 0.213 (range 0.111-0.527).

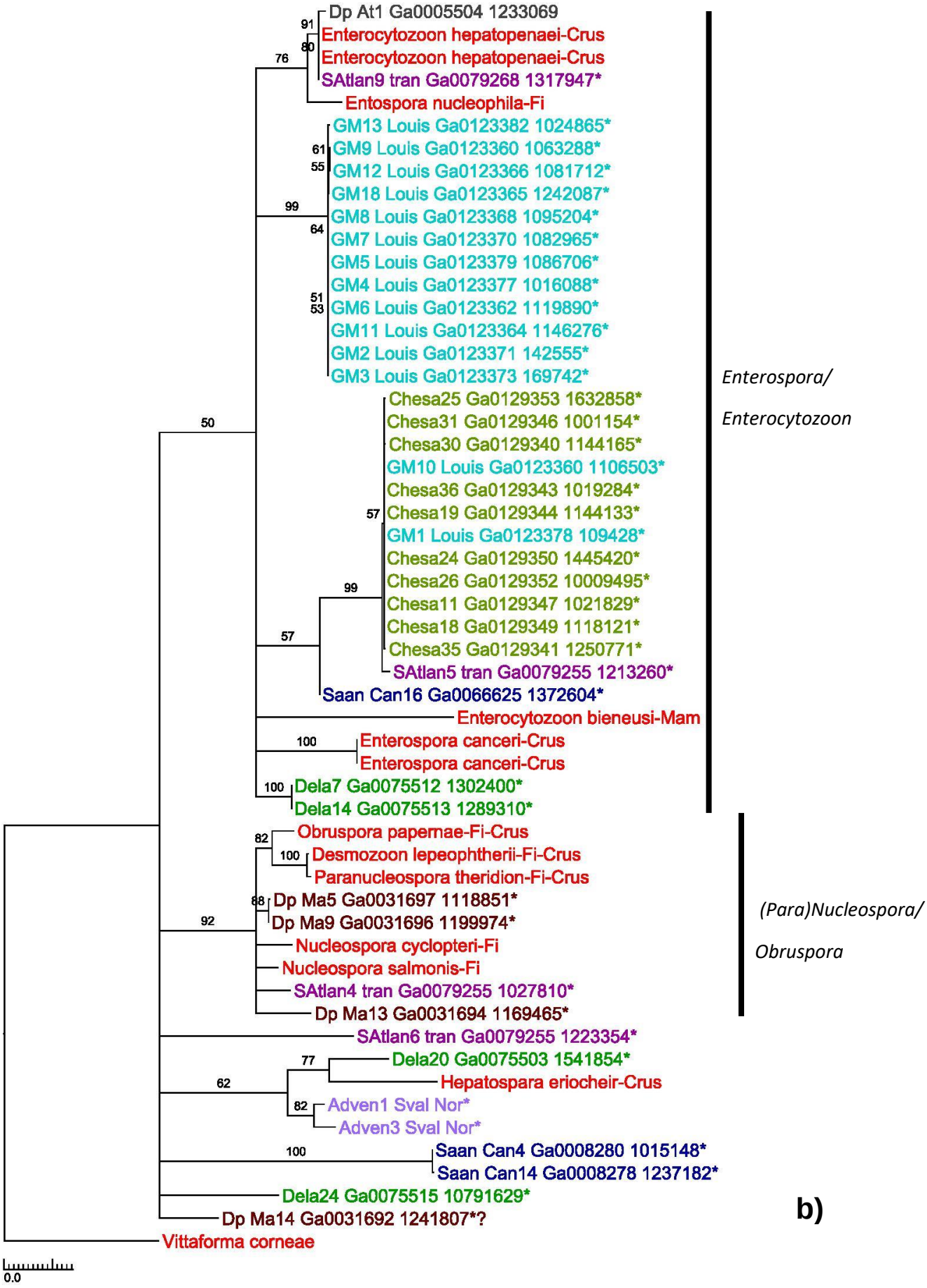
2.3.3 The relationship between geographic distance and pairwise distance between enterocytozoonid-like metagenomic scaffolds

Geographic distance between sample sites did not correlate with pairwise distance in any of the three datasets (Figure 2.2a-c -supplementary), with Mantel r statistics consistently close to zero (-0.142, -0.101, and -0.074) and no

significant p-values (0.716, 0.767, and 0.773 respectively). However, samples from Saanich Inlet were the most geographically and genetically divergent, with an average pairwise distance of 0.273 (range 0.247-0.306), 2.28 (range 0.208-0.215), and 0.507 (range 0.451-0.571), across all three datasets (600, 985 and 1470 respectively). Saanich Inlet samples also had the greatest intrapopulation pairwise distance 0.237 (range 0.225-0.249), Chesapeake Bay samples had the lowest intrapopulation distance estimates 0.037 (range 0.001-0.105).


100
Saan Can14 Ga0008278 1237182*
Saan Can4 Ga0008280 1015148*
Chesa36 Ga0129343 1019284*
Chesa18 Ga0129349 1118121*
65
Chesa30 Ga0129340 1144165*
63
Chesa11 Ga0129347 1021829*
Chesa32 Ga0129346 1018137*
77
Chesa25 Ga0129353 1632858*
Chesa35 Ga0129341 1250771*
Chesa26 Ga0129352 10009495*
70
Chesa24 Ga0129350 1445420*
Chesa19 Ga0129344 1144133*
Chesa33 Ga0129346 1179639*
SAtlan10 tran Ga0079268 1322370*
Adven1 Sval Nor*
100
Enterospora canceri-Crus
Enterospora canceri-Crus
Enterocytozoon bieneusi-Mam
GM18 Louis Ga0123365 1242087*
Enterospora/Enterocytozoon
75
Hepatospara eriocheir-Crus
Dela20 Ga0075503 1541854*
89
SAtlan9 tran Ga0079268 1317947*
97
Enterocytozoon hepatopenaei-Crus
Enterocytozoon hepatopenaei-Crus
Enterospora/Enterocytozoon
Entospora nucleophila-Fi
Saan Can16 Ga0066625 1372604*
Desmozoon lepeophtherii-Fi-Crus
55
Nucleospora salmonis-Fi
Paranucleospora theridion-Fi-Crus
71
Nucleospora cyclopteri-Fi
80
Obruspora papernae-Fi-Crus
100
Dp Ma5 Ga0031697 1118851*
51
Dp Ma9 Ga0031696 1199974*
SAtlan4 tran Ga0079255 1027810*
(Para)Nucleospora/Obruspora
Dp Ma14 Ga0031692 1241807*?
Dela7 Ga0075512 1302400*
Vittaforma corneae
0.0

a)


Dp At1 Ga0005504 1233069
Enterocytozoon hepatopenaei-Crus
Enterocytozoon hepatopenaei-Crus
SAtlan9 tran Ga0079268 1317947*
Entospora nucleophila-Fi
GM13 Louis Ga0123382 1024865*
GM9 Louis Ga0123360 1063288*
GM12 Louis Ga0123366 1081712*
GM18 Louis Ga0123365 1242087*
GM8 Louis Ga0123368 1095204*
GM7 Louis Ga0123370 1082965*
GM5 Louis Ga0123379 1086706*
GM4 Louis Ga0123377 1016088*
GM6 Louis Ga0123362 1119890*
GM11 Louis Ga0123364 1146276*
GM2 Louis Ga0123371 142555*
GM3 Louis Ga0123373 169742*
Chesa25 Ga0129353 1632858*
Chesa31 Ga0129346 1001154*
Chesa30 Ga0129340 1144165*
GM10 Louis Ga0123360 1106503*
Chesa36 Ga0129343 1019284*
Chesa19 Ga0129344 1144133*
GM1 Louis Ga0123378 109428*
Chesa24 Ga0129350 1445420*
Chesa26 Ga0129352 10009495*
Chesa11 Ga0129347 1021829*
Chesa18 Ga0129349 1118121*
Chesa35 Ga0129341 1250771*
SAtlan5 tran Ga0079255 1213260*
Saan Can16 Ga0066625 1372604*
Enterocytozoon bieneusi-Mam
Enterospora canceri-Crus
Enterospora canceri-Crus
Dela7 Ga0075512 1302400*
Dela14 Ga0075513 1289310*
Obruspora papernae-Fi-Crus
Desmozoon lepeophtherii-Fi-Crus
Paranucleospora theridion-Fi-Crus
Dp Ma5 Ga0031697 1118851*
Dp Ma9 Ga0031696 1199974*
Nucleospora cyclopteri-Fi
Nucleospora salmonis-Fi
SAtlan4 tran Ga0079255 1027810*
Dp Ma13 Ga0031694 1169465*
SAtlan6 tran Ga0079255 1223354*
Dela20 Ga0075503 1541854*
Hepatospara eriocheir-Crus
Adven1 Sval Nor*
Adven3 Sval Nor*
Saan Can4 Ga0008280 1015148*
Saan Can14 Ga0008278 1237182*
Dela24 Ga0075515 10791629*
Dp Ma14 Ga0031692 1241807*?
Vittaforma corneae
0.0
Enterospora/
Enterocytozoon
(Para)Nucleospora/
Obruspora
b)

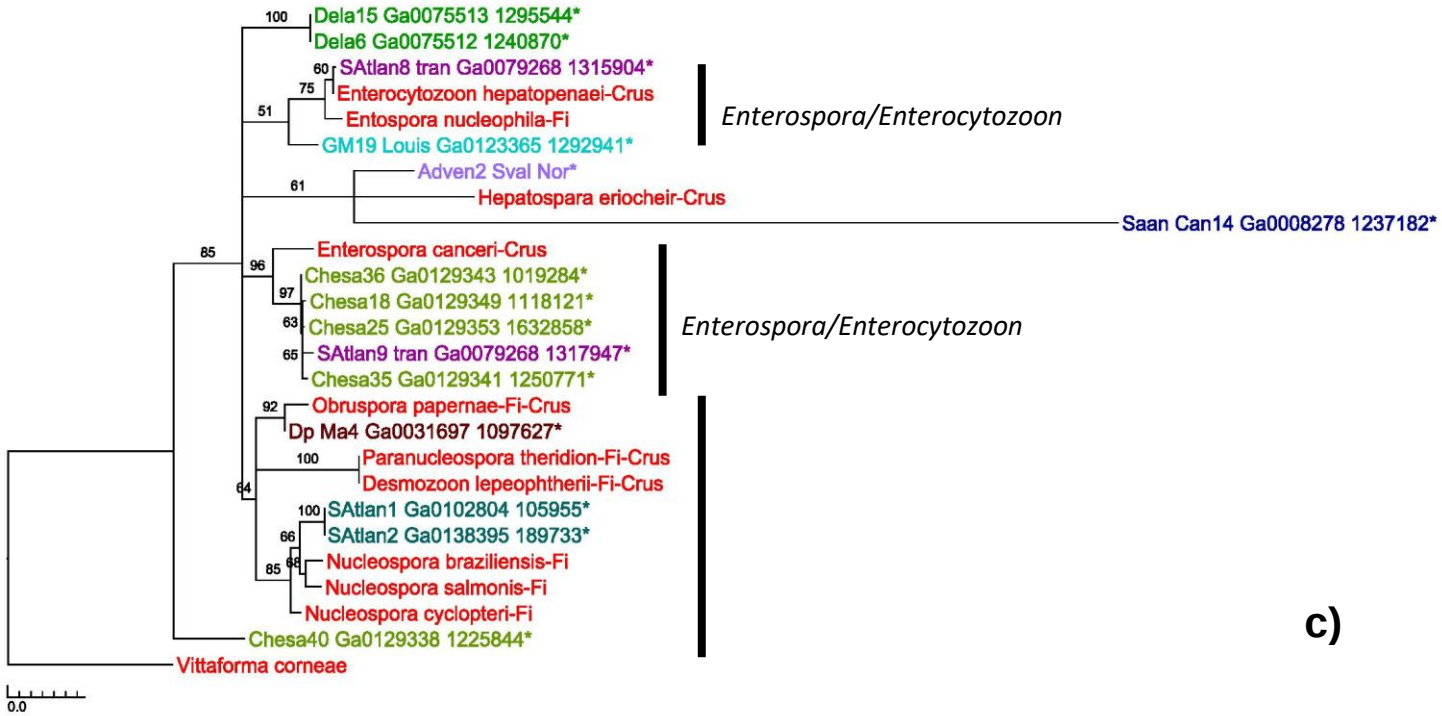


Figure 2.2: a-c: Enterocytozoonidae phylogenies from three different datasets (a-600; b-985; c-1470), built in RAxML-HPC (GTRGAMMA), showing the diversity of the 37/55/26 newly discovered microsporidian sequences respectively, detected in this study from metagenome databases, which matched the same region of the 18s rDNA gene. Scale bar intervals represent 0.01 substitutions per site for branch length. Numbers above nodes indicate bootstrap support values. Nodes with a value less than 50 were collapsed. (Red names: described species; Green names: from Delaware (Dela; Pale green names: from Chesapeake Bay (Chesa); Light blue names: Gulf of Mexico (GM); Dark blue names: Saanich Inlet (Saan Can); Purple names: South Atlantic (SAtlan); Light purple names: Adventfjord (Adven); Burgundy names: Pacific Ocean (DpMa); Grey names: Atlantic Ocean (Dpart))

Table 2.3a-c: Average pairwise distance between described enterocytozoonid taxa and enterocytozoonid-like sequences identified in this study, estimated in MEGAX from 18s rDNA for dataset 600 (Key: Saan=Saanich Inlet, USA; Dela=Delaware River, USA; GM=Louisana Shelf, USA; Dpma= West of El Salvador, Pacific Ocean; Chesa=Chesapeake Bay, USA; Satlan_tran=South Atlantic Ocean; Adven=Adventfjord, Norway)

	***Enterocytozoon* clade**	***Nucleospora* clade**	**Saan**	**Dela**	**GM**	**Dpma**	**Chesa**	**Satlan_tran**	**Adven**
Enterocytozoon clade	0.111								
***Nucleospora* clade**	0.153	0.053							
Saan	0.258	0.306	0.249						
Dela	0.158	0.179	0.273	0.236					
GM	0.064	0.117	0.27	0.143	N/A				
Dpma	0.15	0.118	0.291	0.188	0.115	0.082			
Chesa	0.132	0.142	0.247	0.171	0.133	0.154	0.007		
Satlan_tran	0.124	0.12	0.26	0.172	0.1	0.122	0.108	0.146	
Adven	0.193	0.186	0.285	0.209	0.193	0.184	0.154	0.183	N/A

2.3b

Average pairwise distance between described enterocytozoonid taxa and enterocytozoonid-like sequences identified in this study, estimated in MEGAX from 18s rDNA for dataset 985 (Key: Saan=Saanich Inlet, USA; Dela=Delaware River, USA; GM=Louisana Shelf, USA; Dpma= West of El Salvador, Pacific Ocean; Chesa=Chesapeake Bay, USA; Satlan_tran= South Atlantic Ocean; Adven=Adventfjord, Norway; Dpart=Atlantic Ocean)

	***Enterocytozoon* clade**	***Nucleospora* clade**	**Saan**	**Dela**	**GM**	**Dpma**	**Chesa**	**Satlan_tran**	**Adven**	**Dpart**
***Enterocytozoon* clade**	0.141									
***Nucleospora* clade**	0.168	0.061								
Saan	0.208	0.225	0.225							
Dela	0.184	0.167	0.221	0.188						
GM	0.131	0.153	0.215	0.172	0.047					
Dpma	0.176	0.094	0.227	0.170	0.146	0.112				
Chesa	0.146	0.149	0.217	0.18	0.119	0.172	0.001			
Satlan_tran	0.152	0.141	0.224	0.164	0.142	0.137	0.118	0.168		
Adven	0.22	0.169	0.244	0.189	0.206	0.176	0.235	0.2	0.039	
Dpatl	0.087	0.16	0.272	0.191	0.065	0.142	0.149	0.122	0.234	_

2.3c

Average pairwise distance between described enterocytozoonid taxa and enterocytozoonid-like sequences identified in this study, estimated in MEGAX from 18s rDNA for dataset 1470 600 (Key: Saan=Saanich Inlet, USA; Dela=Delaware River, USA; Satlan=South Atlantic Ocean GM=Louisana Shelf, USA; Dpma=South Atlantic Ocean; Chesa=Chesapeake Bay, USA; Satlan_tran=South Atlantic Ocean; Adven=Adventfjord, Norway)

	***Enterocytozoon* clade**	***Nucleospora* clade**	**Saan**	**Dela**	**Satlan**	**GM**	**Dpma**	**Chesa**	**Satlan_tran**	**Adven**
***Enterocytozoon* clade**	0.120									
***Nucleospora* clade**	0.172	0.121								
Saan	0.475	0.527	0							
Dela	0.123	0.191	0.504	0						
Satlan	0.171	0.111	0.571	0.174	0					
GM	0.123	0.166	0.513	0.129	0.143	_				
Dpma	0.144	0.126	0.51	0.14	0.139	0.139	_			
Chesa	0.155	0.181	0.526	0.149	0.173	0.165	0.156	0.105		
Satlan_tran	0.104	0.164	0.488	0.133	0.169	0.131	0.150	0.112	0.143	
Adven	0.234	0.239	0.451	0.205	0.241	0.215	0.221	0.244	0.219	_

2.3.4 OTU dataset assessment

As mentioned, five different datasets were compiled to capture the diversity of the varied coverage and length of all metagenomic scaffolds across the 18s rDNA region (Table 2.4). These were clustered into 27-167 OTUs (Table 2.4). The number of OTUs formed were closely associated with how many sequences were included in the dataset, showing a linear relationship between the two (Figure 2.2-supplementary). This suggested no bias for a particular dataset (18s rDNA region), as such, all datasets were used in further analysis.

2.3.5 Enterocytozoonidae: a dominant, environmental microsporidian clade?

To get a better representation of microsporidian diversity found within the metagenomes, and to give a comparative look at the abundance/ratio of the Enterocytozoonidae among the rest of the Microsporidia, sequences were clustered into operational taxonomic units (OTUs) based on identity as a threshold (97%). OTUs were then clustered into 'family'-based clusters in UCLUST, also using an identity-based threshold (80%). The Enterocytozoonidae made up a relatively large proportion of OTUs, with an average of 18.04% of each dataset made up of enterocytozoonid-like sequences (Table 2.4). A proportion made starker, as enterocytozoonid-like 'family' clusters only made up, on average, 12.95% of all 'family' clusters in this study (Table 2.4). On average the enterocytozoonid-like clusters also show more diversity than the other clusters in the dataset, with the average number of enterocytozoonid-like OTUs across all five datasets (Table 2.4) in a 'family' cluster greater (3-10.5 OTUs) than the average number of OTUs in a non-enterocytozoonid 'family' cluster (range 2.3-4.6 OTUs).

Both the difference between the number of clusters and the number of OTUs in the cluster were explained by the size of the dataset (Figure 2.3 and 2.4-supplementary), with a linear relationship between the number of OTUs in the dataset and the number/size of clusters. Showing that there was an even spread of diversity along the datasets, with no apparent bias. 'Family' clustering of representative, known species followed the known phylogeny for the sequences involved, adding support for the observations and clustering threshold made above.

Table 2.4: Summary of OTU and cluster analysis results for the five datasets, using USEARCH and UCLUST from all metagenome sequences from this study. Sequence identity threshold for OTU clustering was set to 80% identity. Based on 18s rDNA.

Dataset	Alignment length	Unique sequences	OTUs/of which are enterocytozoonid-like (percentage)	'Family' clusters/of which are enterocytozoonid-like (percentage)	Largest number of OTUs in 'family' cluster/average	largest enterocytozoonid-like cluster/average
300	180	126	65/11 (16.9%)	24 (46)/1 (4.16%)	11/2.3	3/3
600	395	273	167/19 (11.4%)	28 (46)/2 (7.14%)	42/4.6	13/9.5
900	295	308	114/18 (15.8%)	27 (45)/4 (14.81%)	25/3.5	15/10.5
1200	300	191	79/13 (16.5%)	22 (39)/3 (13.63)	14/3.1	6/3.7
1368	170	48	27/8 (29.6%)	8 (30)/2 (25%)	12/2.3	12/7
All	1493	712	258/52 (20.1%)	36 (55)/4 (11.11%)	51/5.5	22/14.5

For example, 'family' clustering using UCLUST (80% identity) clustered two *Vairimorpha* and *Nosema* species together and had clustered the species *Vavraia culicis* and *Pleistophora typicalis*. Species that have shown close relationships in prior phylogenetic analyses. However, not many of the known sequences are included in clusters with metagenome sequences, with the majority of the clusters including nothing but newly found sequences.

2.3.6 E. bieneusi diversity and ITS variability

To investigate the variation seen in the ITS1 regions of *E. bieneusi* within one host, BLASTN searches were performed against the published reference genome, using an ITS1 query sequence. This was done with the aim to assess whether ITS1 was an appropriate marker to genotype *E. bieneusi*. BLASTN searches using ITS1 against the *E. bieneusi* reference genome recovered 24 hits. The recovered BLASTN results were placed into a published dataset (Baroudi *et al*. 2018) of ITS1 sequences, to look at their position among known genotypes. The final dataset consisted of 66 sequences, with 422 sites, of which 273 are variable and 190 are parsimony informative. Although all 24 ITS1 sequences that were pulled from the published genome of *E. bieneusi* had grouped within genotype group 1 (Baroudi *et al*., 2018) (Figure 2.3), variation was seen between ITS1 sequences with the majority (13 sequences) forming a polytomy within the group. The overall relationship between the genotypes and genotype groups has

remained the same as Baroudi *et al* (2018), with the new addition of the ITS1 sequences. However, some of the 24 ITS1 sequences did show grouping, six have joined the pre-existing group 1a (Baroudi *et al*., 2018), with four sequences forming two more supported branches (71, 50 bootstrap support). Pairwise distance for the ITS1 sequences showed similar levels of variation (Table 2.6), ranging from 0% - 45.6%, with an average distance between the sequences being 6.6%. The sequence consistently most divergent from other ITS1 sequences was ABGB01001610.1 ranging from 25% - 45.6% pairwise distance (average 29.2%).

To examine whether the variation found in ITS1 sequences was representative of variation across the whole genome in *E. bieneusi* (therefore, if it was a good region for genotyping of *E. bieneusi*), within a single host, reads were mapped to 128 single-copy genes. Within the host, genome-wide diversity for *E. bieneusi* in this study was very low, using a measure of nucleotide diversity (π) as a proxy for within-population (within-host) diversity, within-population nucleotide diversity was on average 0.0015 (Table 2.5). Contrasting the high variability found for the ribosomal ITS1 pairwise distance estimates.

Table 2.5: Single copy genes from OrthoMCL analysis, used for with SNP calling in *E. bieneusi*

NCBI Reference	Characterisation	SNP count	Nucleotide diversity (π)
EDQ31221	DNA primase small subunit	1	0.000
EDQ31229	glycyl-tRNA synthetase	9	0.002
EED43701	deoxyhypusine synthase	3	0.001
EED43766	cation transport ATPase	18	0.003
EED44056	ATP-dependent 26S proteasome regulatory subunit	5	0.001
EED44079	arginyl-tRNA synthetase	9	0.001
EED44544	LSU ribosomal protein L3P	5	0.001
EED44563	protein kinase kin1	12	0.002

A relatively small number of reads were mapped to the single-copy genes (8.32%). *E. bieneusi* was conserved throughout the majority of the mapped reads, only showing 62 SNPs called along with eight (out of 128) single-copy reference genes. The eight genes were all described, though mainly characterised through homology-based analysis (Table 2.5). There was no bias in the distribution of reads over the single-copy genes, following a linear relationship (Figure 2.5a-supplementary), neither in SNP counts, also showing a linear relationship with reference length (Figure 2.5b-supplementary).

Figure 2.3: Phylogeny showing the relationship between *E. bieneusi* genotypes, built-in RAxML-HPC (GTRGAMMA), using the ribosomal ITS. Scale bar

intervals represent 0.01 substitutions per site for branch length. Numbers above nodes indicate bootstrap support values. Nodes with a value less than 50 were collapsed. (Those highlighted blue are the sequences pulled from the published *E. bieneusi* genome)

Table 2.6: Pairwise distance matrix showing distance (p) between the sequences included in Baroudi *et al* (2018), and ITS1 sequences included in this study from the *E. bieneusi* reference genome.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1																								
2	0.000																							
3	0.000	0.000																						
4	0.000	0.000	0.000																					
5	0.002	0.002	0.002	0.002																				
6	0.047	0.047	0.047	0.047	0.045																			
7	0.021	0.021	0.021	0.021	0.021	0.021																		
8	0.005	0.005	0.005	0.005	0.007	0.052	0.027																	
9	0.005	0.005	0.005	0.005	0.007	0.052	0.027	0.000																
10	0.005	0.005	0.005	0.005	0.007	0.052	0.027	0.000	0.000															
11	0.005	0.005	0.005	0.005	0.007	0.052	0.027	0.000	0.000	0.000														
12	0.002	0.002	0.002	0.002	0.005	0.050	0.024	0.002	0.002	0.002	0.002													
13	0.013	0.013	0.013	0.013	0.013	0.022	0.027	0.008	0.008	0.008	0.008	0.011												
14	0.000	0.000	0.000	0.000	0.002	0.047	0.021	0.005	0.005	0.005	0.005	0.002	0.013											
15	0.015	0.015	0.015	0.015	0.017	0.055	0.027	0.020	0.020	0.020	0.020	0.017	0.024	0.015										
16	0.006	0.006	0.006	0.006	0.009	0.064	0.035	0.012	0.012	0.012	0.012	0.009	0.023	0.006	0.024									
17	0.021	0.021	0.021	0.021	0.025	0.102	0.072	0.029	0.029	0.029	0.029	0.025	0.049	0.021	0.046	0.021								
18	0.009	0.009	0.009	0.009	0.014	0.096	0.060	0.000	0.000	0.000	0.000	0.004	0.016	0.009	0.036	0.009	0.009							
19	0.025	0.025	0.025	0.025	0.029	0.094	0.067	0.032	0.032	0.032	0.032	0.028	0.048	0.025	0.046	0.004	0.021	0.009						
20	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.054	0.054	0.054	0.054	0.050	0.054	0.047	0.047	0.070	0.127	0.120	0.115					
21	0.091	0.091	0.091	0.091	0.091	0.091	0.097	0.097	0.097	0.097	0.097	0.094	0.097	0.091	0.091	0.116	0.165	0.170	0.175	0.088				
22	0.122	0.122	0.122	0.122	0.125	0.144	0.096	0.127	0.127	0.127	0.127	0.124	0.114	0.122	0.130	0.156	0.233	0.238	0.204	0.098	0.145			
23	0.073	0.073	0.073	0.073	0.073	0.073	0.070	0.080	0.080	0.080	0.080	0.077	0.080	0.071	0.077	0.104	0.185	0.186	0.160	0.096	0.112	0.096		
24	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.253	0.253	0.253	0.253	0.250	0.253	0.247	0.257	0.342	0.515	0.569	0.456	0.260	0.274	0.253	0.266	

(Key: ABGB01000374.1 = 1, ABGB01001290.1 = 2, ABGB01001369.1 = 3, ABGB01001147.1 = 4, ABGB01001626.1 = 5, ABGB01000947.1 = 6, ABGB01001084.1 = 7, ABGB01000781.1 = 8, ABGB01000732.1 = 9, ABGB01001533.1 = 10, ABGB01001612.1 = 11, ABGB01000756.1 = 12, ABGB01000820.1 = 13, ABGB01001462.1 = 14, ABGB01001242.1 = 15, ABGB01001534.1 = 16, ABGB01001482.1 = 17, ABGB01001406.1 = 18, ABGB01001496.1 = 19, ABGB01001498.1 = 20, ABGB01001206.1 = 21, ABGB01001053.1 = 22, ABGB01001211.1 = 23, ABGB01001610.1 = 24)

2.4 Discussion

This is the first systematic survey of Enterocytozoonidae diversity, using metagenome databases. The appropriateness of using the ribosomal ITS1 as a marker for genotyping *E. bieneusi* was also assessed. This study strongly suggests that the family is more diverse and widely distributed than is currently described. Furthermore, it reveals minimal variation between spores within a single host, implying that different methods/multiple regions need to be used when barcoding studies are carried out on some species.

2.4.1 Distribution of undescribed Enterocytozoonidae

With all the enterocytozoonid-like sequences coming from estuarine/marine-based metagenomes (Figure 2.4), it suggests that the Enterocytozoonidae are primarily marine-based aquatic infecting parasites, which may have a large reservoir of diversity found around Atlantic-based coasts, except the Saanich Inlet (which was the most divergent). This may also be an artifact of sampling bias, as the majority of the institutions and sampling locations were North American. However, there is also evidence for this inference found in the already described species, with the only terrestrial instances of described enterocytozoonid infection being found in human-related infections: *E. bieneusi* infects humans/livestock (locations: Czech Republic (Sak *et al*., 2010), Uganda (Akiyoshi *et al*., 2009), Switzerland (Mathis *et al*., 1999), Haiti (Desportes *et al*., 1985), Peru (Cama *et al*., 2007), China (Zhao *et al*., 2014), Ireland (Graczyk *et al*., 2004)) and *E. hepatopenaei* infects livestock (locations: Thailand (Tourtip *et al*., 2009), China (Y. M. Liu *et al*., 2018), India (Rajendran *et al*., 2016a), Vietnam (Ha *et al*., 2010), Australia (possibly) (Hudson *et al*., 2001), Venezuela (Tang *et al*., 2017), Indonesia (Tang *et al*., 2016a) and Brunei (Tang *et al*., 2015a). The other described species were found in Atlantic-based hosts (Table 2.7), or Seas adjacent to (*P. theridion*-North Sea) (Nylund *et al*., 2010). However, there are two species described from the Arabian Sea/Indian Ocean (*O. papernae* and *N. secunda*) (Diamant *et al*., 2014; Lom & Dykoá, 2002), showing that there are estuarine/marine instances outside of the Atlantic.

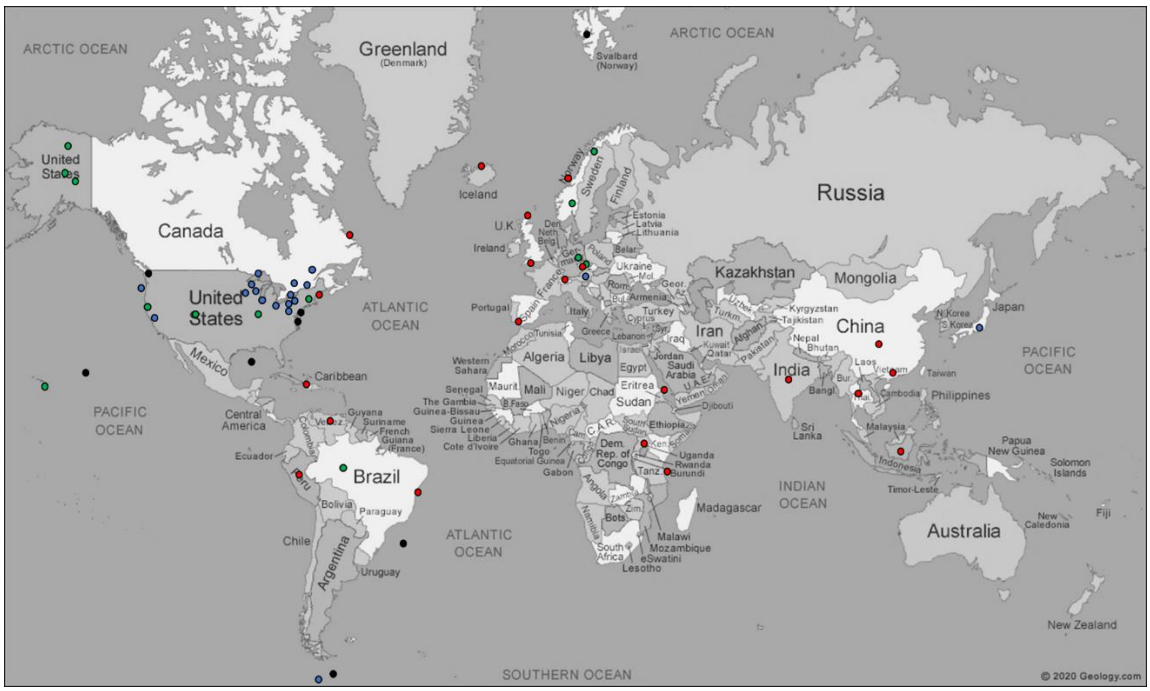


Figure 2.4: Locations of enterocytozoonid occurrences. (Red: described, published species; blue: freshwater metagenome sequences; green: terrestrial metagenome sequences; black: positive enterocytozoonid-like sequences)

2.4.2 Phylogenetic positioning

The phylogenetic positioning of the novel sequences (Figure 2.2a-c) and on average smaller pairwise distance estimates for the *Nucleospora* clade compared to the *Enterocytozoon/Enterospora* clade (average p-distance across all datasets: *Enterocytozoon/Enterospora* 0.169; *Nucleospora* 0.179) may indicate a greater radiation within the *Enterospora/Enterocytozoon* clade (or a clade that is sister to). This is also suggested by the majority of the novel sequences being placed within the *Enterospora/Enterocytozoon* clade and not within the *Nucleospora* clade. However, this could also be partially explained by the filtering bias, as these filtering techniques are designed to filter passive dispersing planktonic organisms. Whereas it would be less likely that the filtering-based techniques used would capture as many active dispersers, such as fish. As the *Nucleospora* clade has been predominately described infecting fish, it's unlikely that the diversity of this clade would be represented as well *Enterospora/Enterocytozoon* in these samples. This portion of the study did not reveal any sequences that were similar to *E. bieneusi.* This suggests, given that the diversity of the novel sequences span throughout the tree, that *E. bieneusi*

may have more terrestrial-based “intermediate” species and not be present in these geographic locations. This concurs with the evidence of *E. bieneusi* in the River Shannon (Graczyk *et al*., 2004), as estuaries would be an intermediate environment between marine life and terrestrial life.

2.4.3 OTU and clustering: the abundance of Enterocytozoonidae

The rate at which the Enterocytozoonidae are being discovered (first species described in 1985, six species out of 10 described in the past decade) suggests a very diverse family. However, it is not known whether this has been the result of selective sampling, as many of the enterocytozoonid species described infect human-affected species or a large evolutionary radiation. Evidence for the latter is inferred from a strong representative biodiversity for enterocytozoonid-like sequences in this study, making up, on average across all datasets, 18.04% of the total OTUs. While taking up a smaller proportion of ‘family’ clusters that are enterocytozoonid (Table 2.4), with an average of 2.4 (range of 1-4) clusters per data set, making up an average of 12.91% (4.16%-25%) of the clusters per dataset. This suggests that the Enterocytozoonidae sequences are likely more diverse than non-enterocytozoonid sequences. This percentage has risen from the proportion the Enterocytozoonidae made up with all sequences before the OTU analysis (12.7%). As the OTU analysis groups sequences that are 97% similar, a rise in the proportion of enterocytozoonid-like sequences would suggest greater pairwise differences between these sequences. Mirroring this are the sizes of the clusters, with enterocytozoonid-like clusters having more OTUs in every dataset. Considering the only enterocytozoonid-like sequences were found in eight (of 46) estuarine/marine datasets, it suggests a very diverse family, in the locations sampled. However, it is likely that some diversity has not been accounted for, as the 18s rDNA in general shows less taxonomic resolution than other markers, such as ITS1,2 (Tedersoo *et al*., 2015).

2.4.4 Spatial distance and genetic diversity

A lack of correlation between geographic distance and pairwise distance (Figure 2.2a-c-Supplementary) suggests that spatial distance is not a dominant factor in the genetic distance between locations sampled in this study. As is the case with

many marine systems, due to such wide dispersal capabilities in a marine medium (Cooke *et al*., 2016). What is likely a factor in the genetic distance between these locations is host availability/marine currents (Huyghe & Kochzius, 2018; White *et al*., 2010). As the primary means through which many of these organisms transmit/disperse is likely zooplankton, they would heavily rely upon the ocean currents. However, the consistent relatively low-level p-distance (0.146 average across all datasets) from sequences that originate from the Atlantic suggests a 'local' (local being the Atlantic) radiation. With the larger p-distances coming from the more isolated locations (Saanich Inlet, Canada, and Adventfjord, Svalbard Archipelago, Norway).

2.4.5 Intermediate host/host range

The current understanding of intermediate hosts for the Enterocytozoonidae is severely lacking. Currently, five described species have been noted in different species. *E. bieneusi* (humans and various vertebrates), *Nucleospora salmonis* (various species of salmon) (El Alaoui *et al*., 2006a; Foltz *et al*., 2009; Sakai *et al*., 2009), *P. theridion/D. lepeophtherii* (Atlantic salmon (*Salmo salar*) and its copepod parasite (*Lepeophtheirus salmonis*) (Freeman & Sommerville, 2009), *O. papernae* (blotchfin dragonet (*Callionymusfilamentosus*) and its copepod parasite (*Lernanthropuscallionymicola*)) and *E. hepatopenaei* (*Penaeus vannamei* and *Penaeus monodon* (Tang *et al*., 2015b). With all enterocytozoonid-like sequences found in this study coming from marine-based filtered water samples, this study suggests that marine environments may hold many undescribed hosts for the Enterocytozoonidae. This would also concur with the inference that copepods may act as a reservoir for this family (Freeman & Sommerville, 2009; Nylund *et al*., 2010).

2.4.6 Implications of the variability of ITS in the published genome

The ITS1 region has been used extensively in genotyping studies for *E. bieneusi* in various different, largely vertebrate, organisms. For this reason, it was a good reference marker to use when looking at the intraspecific diversity that can be found in a single host. However, the variability inherent in ITS regions and the diversity found among the *E. bieneusi* genotypes calls into question its validity as

a marker for infraspecific variation. This study suggests high intraspecific variability with 24 blast results for the ITS1 region, which has a wide range of p-distances from 0 to 0.456 (average 0.066). This is also shown through phylogeny, with four of the genome sequences forming two new groupings and six joining an already described genotype (Figure 2.3). The remaining 12 were too divergent to form groupings with either sequences from the Baroudi*et al* (2018) study, or the 10 mentioned. However, all fall within the genotype group1, which has the majority of the human infecting genotypes (Baroudi *et al*., 2018). This group shows a large amount of variability, it is the most diverse genotype group, holding most of the human infecting genotypes (Baroudi *et al*., 2018). Using the ITS1 estimates alone, it suggests that the individual was infected by multiple variants, with a small number of (if any) clones, shown by a high variability in the ITS1 sequences pulled from the genome. However, single-copy genes show low variability for *E. bieneusi* within a host, with an average nucleotide diversity of 0.0015 (Table 2.5). The analysis also only showed eight of 128 genes that have SNPs called (Table 2.5), suggesting high conservation, which could indicate a highly clonal sample with areas of low variance; likely showing a functional variance under selective pressure. The contrasting variation seen between ITS1 pairwise estimates and single-gene copy nucleotide diversity suggests that ITS1 mutates at a quicker rate than functional genes, which is to be expected. So, inferences drawn from genotyping of *E. bieneusi* should be cautious, and likely use a cloning/next generation sequencing approach to avoid overestimation of genotypes in one host. Though due to their apparent ability to host-shift, there is no reason to suppose that they would retain a host specific genotype, unless there are environmental barriers preventing further host-shifting. It is possible that genotype group 1 (majority human genotypes) is the result of humans coming into contact with other humans more frequently, leading to sequences that are similar and have mostly been described in humans. Further research into this should involve methods that allow a look at total DNA of a single spore or genetically identical spores, produced by one spore. Though this is not currently feasible, as *E. bieneusi* spores are nota viable option for culturing.

2.4.7 Enterocytozoonid risk to aquaculture

With such a diverse group of novel sequences coming from American-based coastal samples, it is feasible that these enterocytozoonid-like organisms could pose a threat to mariculture (cultivation of marine organisms in the open ocean). Cases of enterocytozoonid infections have already been described in farmed Atlantic salmon (*Salmo salar*) in Scotland and Norway, causing an economic loss through associated gill diseases and growth retardation (Freeman *et al*., 2013; Gunnarsson *et al*., 2017; Mullins *et al*., 1994; Nylund *et al*., 2011) and, though shown *S. salar* can be infected through spores in the water column, it has also been inferred that they may also acquire the infection through another parasite, the salmon lice (*Lepeophtheirus salmonis*), a salmon parasite that *D. lepeophtherii* hyperparasitise. Though crustacea are farmed to a lesser extent in North America, as they are primarily farmed indoors, the country is taking a bigger step towards sustained farming of crustacea (Pulidindi & Pandey, 2020). These farmed animals are, however, caught in the wild and kept in indoor farms (Zmora *et al*., 2005), which could facilitate the transmission of Microsporidia into a closed system. One such example is *Callinectes sapidus* (Chesapeake Blue Crab). As mentioned, these are wild-caught then raised to a sellable size in indoor farms (Zmora *et al*., 2005). The native range of this species is the West Atlantic Ocean namely the Gulf Mexico, Chesapeake Bay, and Delaware. Three regions that showed some of the greatest numbers of novel enterocytozoonid-like sequences, in this study. Though there has not been any recorded enterocytozoonid infection in *C. sapidus*, it has been shown that the family does infect crustacea, specifically crabs (*H. eriocheir*) and sometimes multiple crabs (*En. canceri*). There have also been past records of non-enterocytozoonid microsporidian infections in blue crabs, from *Ameson michaelis* and *Glugea stephani* (Sprague, 1965, 1970; Overstreet, 1977, 1988) (Overstreet, 1988; Overstreet & Howse, 1977; Sprague, 1965, 1970).

2.4.8 Summary

Online metagenomes are a good resource for parasite-based risk assessment studies when looking at the various ways areas may be impacted. Though it is a resource that needs prior physical research to be utilised properly (making inferences about the rest of the Enterocytozoonidae and what they infect and how

that may relate to non-described species) and needs to be followed up with further collection of samples to verify these results. However, this study has suggested that the Enterocytozoonidae have a wide, Atlantic centred, distribution. It also suggests a greater diversity, than currently described, found in planktonic communities. There is some evidence to suggest that the evolution of the Enterocytozoonidae is linked to the Atlantic Ocean, however, some species have been described in non-Atlantic oceans and terrestrially.

2.4.9 Future research and limitations of this type of study:

Though this approach is useful in getting estimates of diversity and distribution, further research needs to be carried out collecting physical samples from the locations noted here and surrounding areas, in order to get that a better estimate of the threat this family may pose to the farming industry in these locations. The inferences made from purely metagenomic data are limited, especially when looking at intracellular parasites. Limitations of this type of study also lay with sampling locations, as data are limited to where prior researchers have shown interest and their sampling methods.

2.5 Supplementary tables and figures:

Table 2.1

Described representatives for UCLUST clustering.

Species	Accession Number
Amblyospora stimuli	AF027685.1
Edhazardia aedis	AF027684.1
Amblyosporaopacita	AY090052.1
Hazardiamilleri	AY090067.1
Gurleya daphniae	AF439320.1
Vairimorpha sp.	KP208681.1
Hamiltosporidiummagnivora	AJ302318.1
Weiseria	AF132544.1
Polydispyreniasimuli	AJ252960.1
Paranosemaloctstae	AY305324.1
Amblyosporabracteata	AY090068.1
Pleistophoraanguillarum	AJ278953.1
Pleistophoramirandellae	AJ295327.1
Vavraiaculicis	AJ252961.1
Pleistophoratypicalis	AF044387.1
Glugeaanomala	AF056016.1
Pseudolomaneurophilia	AF322654.1
Loma salmonae	HM626203
Microsporidium prosopium	AF151529.1
Dictyocoelaberillonum	AJ438957.1
Spraguealophii	AF104086.1
Glugeaamericants	AF056014.1
Amesonmichaelis	L15741.1
Thelohaniacontejeani	AF492593.1
Nosema granulosis	FN434087.2
Nosema bombycis	AY259631.1
Nosema ceranae	LC510190.1
Nosema aespula	U11047.1
Vairimorpha imperfecta	AJ131645.1
Nosema apis	U26534.1
Encephalitozoon cuniculi	L07255.1
Ordospora colligata	AF394529.1
Cystosporogenesoperophterae	AJ302320.1
Vittaforma corneae	U11046.1

Glugoidesintestinalis	AF394525.1
Hepatospora eriocheir	HE584635.1
Nucleospora salmonis	AF185987.1
Enterocytozoon bieneusi	ABGB01000919.1
Anncaliia algerae	AY230191.1
Janacekiadebaisieuxi	AJ252950.1
Pseudonosemacristatellae	AF484694.1
Trichonosemapectinatellae	AF484695.1
Schroederaplumatellae	AY135024.1

Table 2.2

All sequences used for ITS1 tree, from BLASTN analysis and Baroudi *et al* study (Baroudi *et al*., 2018).

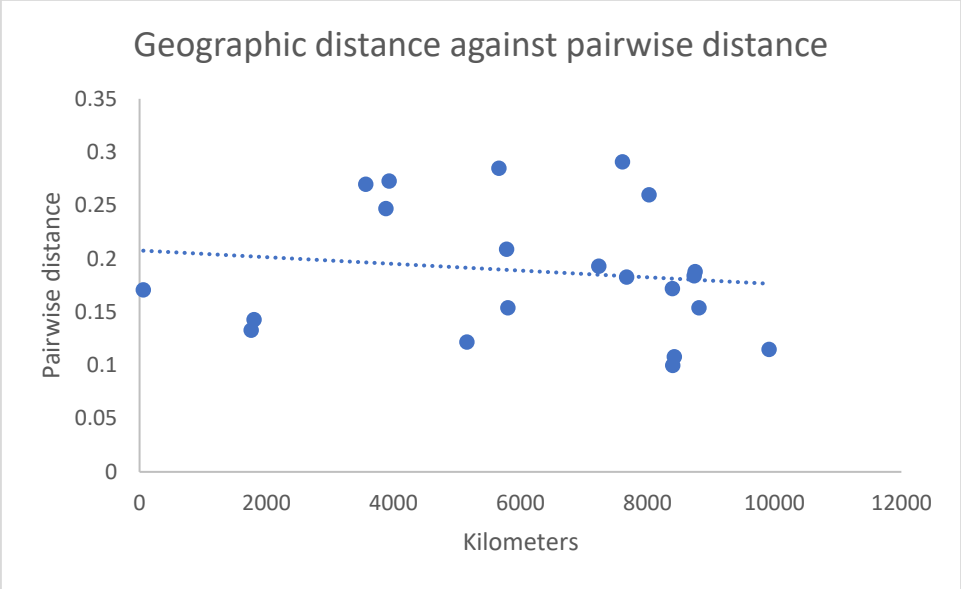
Sample ID	Accession Code	Genotype	Host Species
N/A	ABGB01001206.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001053.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001211.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001498.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001406.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001496.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000732.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001533.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001612.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000756.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000820.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001462.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001242.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001482.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001610.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000374.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001290.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001369.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001147.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001626.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000947.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01001084.1	Group 1	Human **E. bieneusi* genome
N/A	ABGB01000781.1	Group 1	Human **E. bieneusi* genome

N/A	ABGB01001534.1	Group 1	Human **E. bieneusi* genome
CHN4	HM992511	Group 1	Human *Baroudi study
IV	AF242478	Group 1	Human *Baroudi study
Peru2	AY371277	Group 1	Human *Baroudi study
CAF2	DQ683747	Group 1	Human *Baroudi study
CAF1	DQ683746	Group 1	Human *Baroudi study
HAN1	EF458627	Group 1	Human *Baroudi study
EbpC	AF076042	Group 1	Pig *Baroudi study
Donkey 1	KJ941131	Group 1	Donkey *Baroudi study
D	AF101200	Group 1	Human *Baroudi study
Linked-D	DQ836345	Group 1	Cat *Baroudi study
KB-3	JF681177	Group 1	Baboon *Baroudi study
WL14	AY237222	Group 1	Muskrat *Baroudi study
C	AF101199	Group 1	Human *Baroudi study
Q	AF267147	Group 1	Human *Baroudi study
Peru16	EF014427	Group 1	Human *Baroudi study
BEB3	AY331007	Group 2	cattle *Baroudi study
BEB6	EU153584	Group 2	cattle *Baroudi study
BEB7	EU153585	Group 2	cattle *Baroudi study
WL5	AY237213	Group 3	Muskrat *Baroudi study
WL4	AY237212	Group 3	Muskrat *Baroudi study
WL6	AY237214	Group 3	Muskrat *Baroudi study
Nig4	JN997480	Group 7	Human *Baroudi study
Nig3	JN997479	Group 7	Human *Baroudi study
Gorilla 3	JQ837796	Group 8	Gorilla *Baroudi study
Macaque1	JX000572	Group 8	Macaque *Baroudi study
KB-5	JF681179	Group 8	Baboon *Baroudi study
Horse 2	KX276713	Group 8	Horse *Baroudi study
Horse 2	KJ941139	Group 8	Horse *Baroudi study
Horse 2	GQ406054	Group 8	Horse *Baroudi study
XJH1	KU194603	Group 8	Horse *Baroudi study
YNH1	KX276712	Group 8	Horse *Baroudi study
CE01	KU847350	Group 8	Squirrel *Baroudi study
KB-6	JF681180	Group 5	baboon *Baroudi study
CAF4	DQ683749	Group 5	Human *Baroudi study
WL2	AY237210	Group 4	Raccoon *Baroudi study
WL3	AY237211	Group 4	Raccoon *Baroudi study
WL1	AY237209	Group 4	Raccoon *Baroudi study
WL26	KF591687	Group 4	Raccoon *Baroudi study
WW6	JQ863274	Group 6	Wastewater *Baroudi study
WW7	JQ863275	Group 6	Wastewater *Baroudi study

AF059610	AF059610	N/A	Dog *Baroudi study
NED4	KM870518	N/A	Dog *Baroudi study

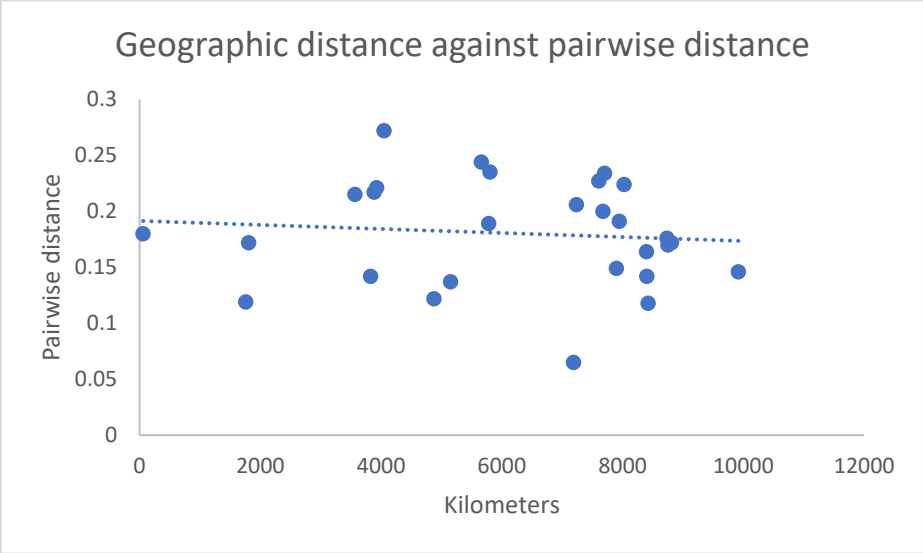
Figure 2.1a

Geographic distance against pairwise distance estimates for dataset 600



2.1b

Geographic distance against pairwise distance estimates for dataset 985



2.1c

Geographic distance against pairwise distance estimates for dataset 1470

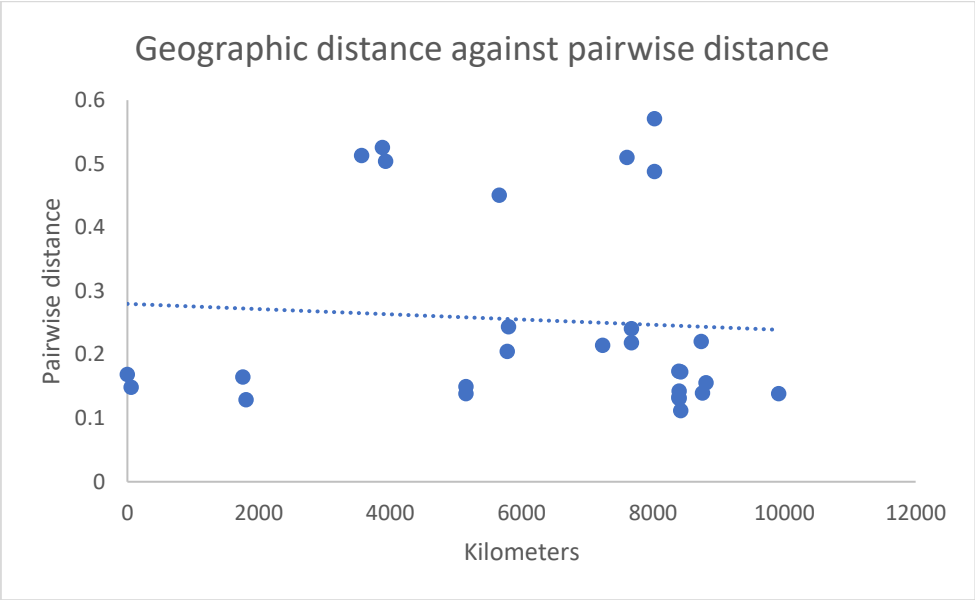


Figure 2.2

Correlation between number of sequences and OTUs

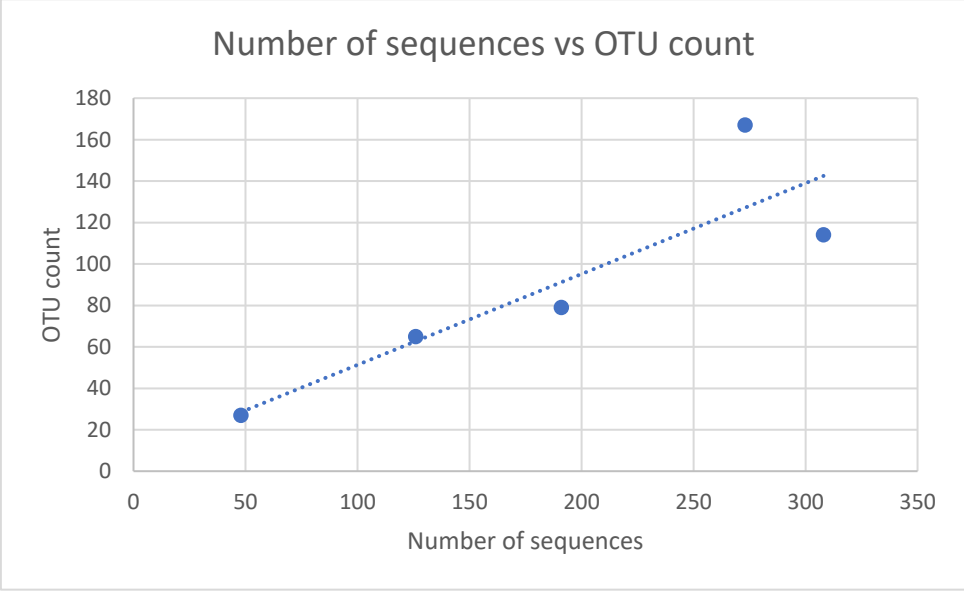


Figure 2.3

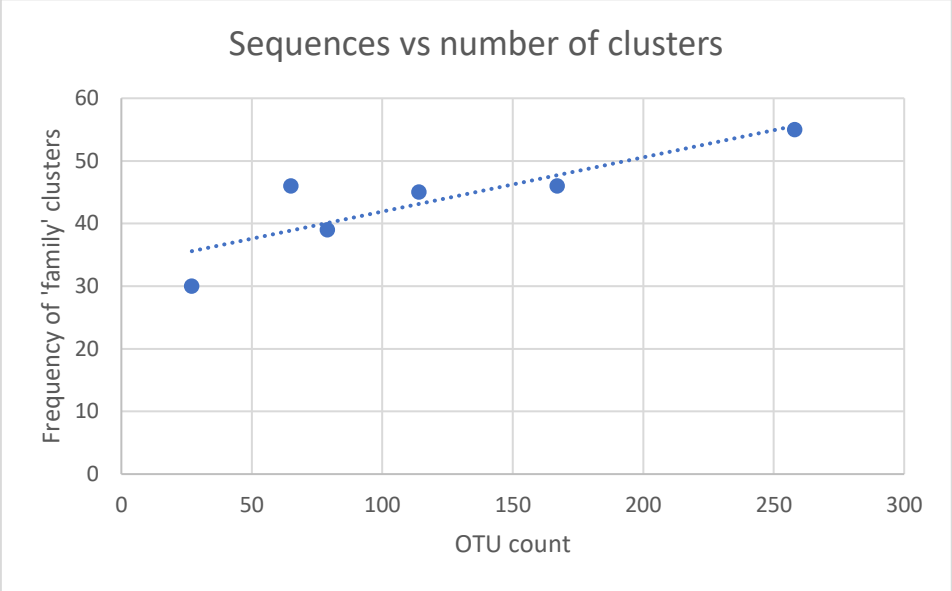


Figure 2.4

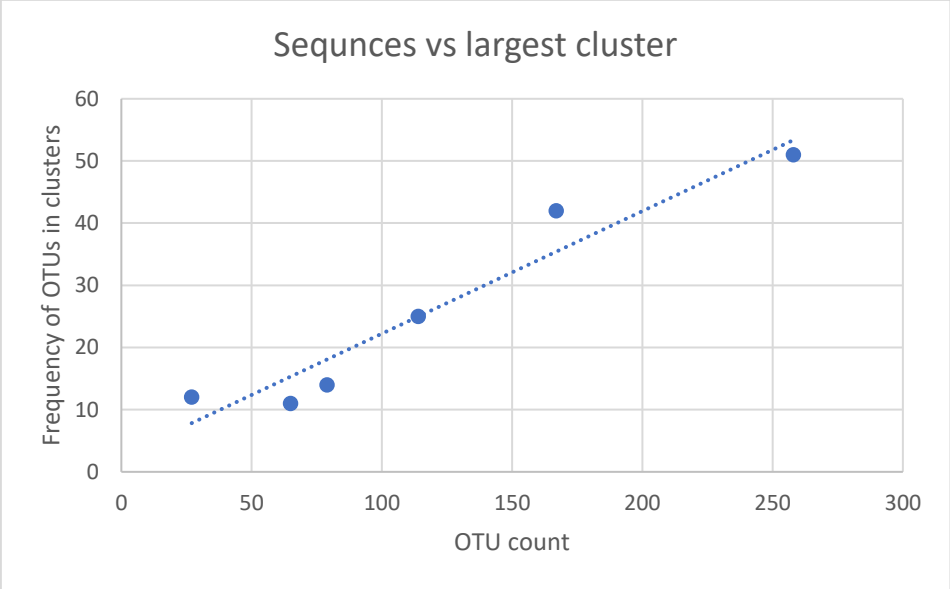
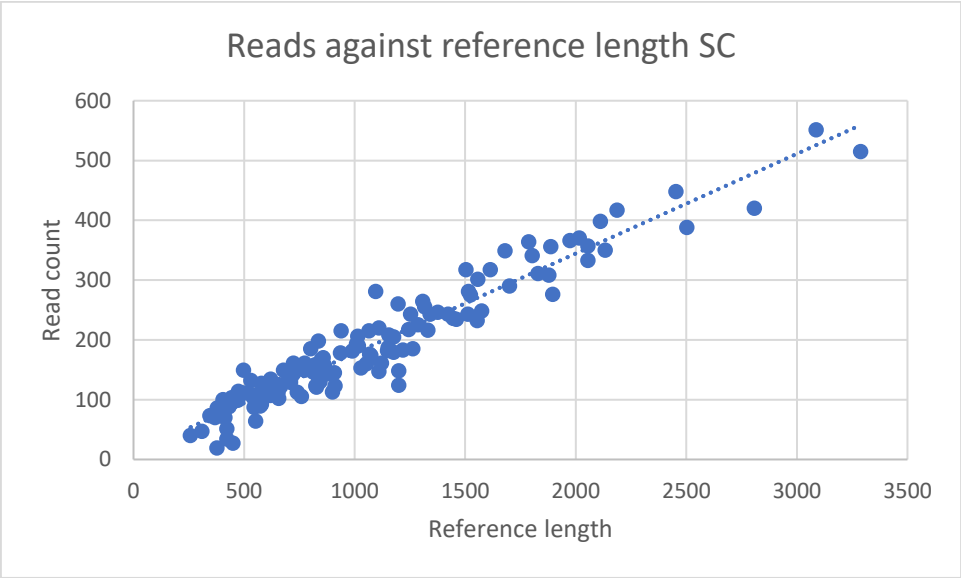
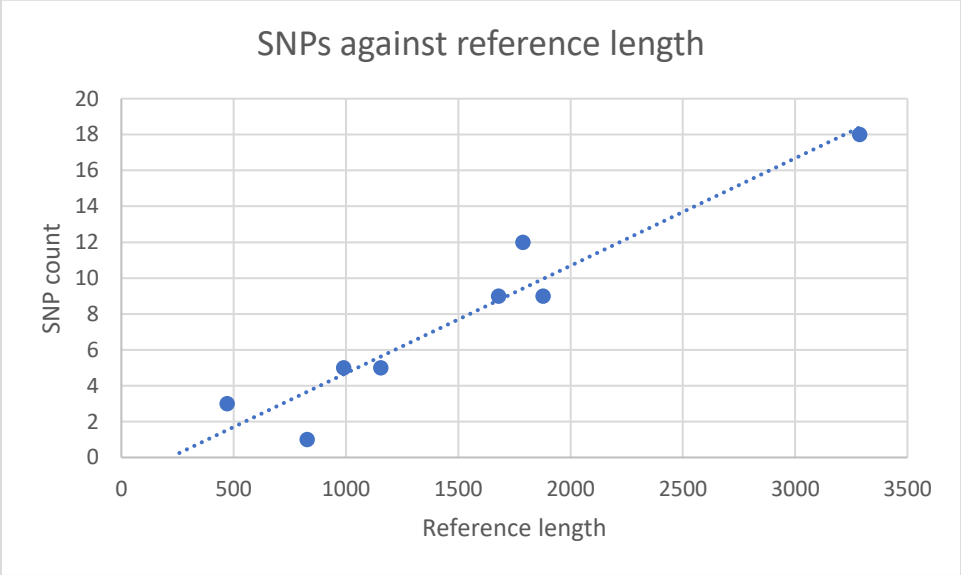


Figure 2.5a



2.5b



Chapter 3: Investigating the diversity of Enterocytozoonidae using environmental samples

3.1 Introduction

Plankton have been filtered from water samples for over a century, predominantly for classification and biodiversity studies (W. J. Clark & Sigler, 1963). These studies were often laborious, and time-consuming, as filtered plankton would need to be sorted and classified by experts in their field. However, with the advancement in molecular techniques and technologies, DNA barcoding has allowed researchers to amplify taxa-specific genetic regions (Blanco-Bercial *et al*., 2014; Djurhuus *et al*., 2017). This has led to a great number of possibilities in the way that environmental samples can be analysed. It has allowed researchers to efficiently screen for organisms at a fraction of the cost and time when compared to direct observation, via surveys and sampling of whole organisms. Its uses include, but are not limited to, detecting invasive or rare species (Furfaro *et al*., 2020; Madden *et al*., 2019) looking at distribution and diversity (Hartikainen, Ashford, *et al*., 2014; Ward *et al*., 2018) as well as community analysis (Blanco-Bercial *et al*., 2014; Djurhuus *et al*., 2017). DNA barcoding is also used in a very applied manner too, and is also used to help make and police policy, critical in ecological studies (L. F. Clark, 2015; Floyd *et al*., 2010). Environmental DNA (In this study, the term environmental DNA refers to the process of filtering environmental samples (soil or water for example) and extracting DNA from what remains on the filter (filtride)) has become very useful in studies looking at organisms that are otherwise difficult to study on a wide scale. For example, it is sometimes used to study the distribution of intracellular parasites (Hartikainen, Ashford, *et al*., 2014; Hartikainen, Stentiford, *et al*., 2014; Ward *et al*., 2018). Research that historically would be undertaken via histological means. A process far more time consuming than sequencing of multiple species.

Amplicon-based sequencing of environmental DNA and the rise of generic primers have proved useful in the investigation of intracellular parasites in the environment. As much of their lifecycle is linked and they are reliant on hosts to propagate, the spatial and temporal distributions of endoparasites are often strongly linked to those of their hosts (Byers *et al*., 2019; González & Poulin,

2005). This allows parasitologists to get a good understanding of the ecology and distribution of some endoparasites by collection and pooling of possible hosts (Ward *et al*., 2018). This is also likely to be very useful for parasites that are able to infect multiple hosts, as collecting known hosts (that will have their own habitat ranges) of the parasite of interest would only give a partial understanding of the parasite's role and effect on an ecosystem. This includes parasites like Microsporidia, frequently described as, and shown to be opportunistic with multiple hosts.

Pooling of possible hosts is undertaken in a number of different ways, from filtering environmental samples, such as water or soil (Ardila-Garcia *et al*., 2013; Hartikainen, Stentiford, *et al*., 2014), to using capture/lure based methods, such as a malaise traps (Shimabukuro *et al*., 2016). However, filtering of environmental samples may also include propagules of pathogens in the environment (Gunnarsson *et al*., 2017) which may, depending on the aim, falsely inflate pathogen diversity in hosts. Pooling also removes many of the limitations placed on single host-based metabarcoding, where samples can include hundreds to thousands of organisms. In many cases, intermediate hosts are integral in the transmission of pathogens to their 'main' host where they can reproduce (Otranto *et al*., 2006), and in some cases multiple hosts are required to finish the pathogens life-cycle (obligate multi-host pathogens) (Bowden, S. E. & Drake, 2013). In these cases, identification of intermediate hosts for pathogens of note would take time. Pooling can also be used to screen for intermediate hosts for pathogens of larger organisms as well as gain a better understanding of a pathogens full life-cycle.

Routes of enterocytozoonid, and Microsporidia as a whole, discovery are predominantly through histological and molecular characterisations of infected economically important hosts (Nylund *et al*., 2010; G. D. Stentiford *et al*., 2011; Tourtip *et al*., 2009). However, outside of initial descriptions of infection of a particular host of importance, enterocytozoonids are infrequently discovered in environmental studies. This is likely the result of no direct studies looking at enterocytozoonid distribution, and the aforementioned (Chapter 2) variation found in universal marker regions. This makes them unlikely to be picked up by broad eukaryotic, or even fungal primers. As a result, the current understanding

of the family's distribution and abundance could be biased towards the locations of commercial hosts. The only two species to get a thorough investigation of abundance are *Enterocytozoon bieneusi* and *Enterocytozoon hepatopenaei,* though these studies are still based around human-influenced areas, leaving their 'wild' distribution largely unknown. However, there have been studies investigating environmental biodiversity that have come across enterocytozoonid-like sequences (Arundell *et al*., 2015; Shen *et al*., 2017) that were highly similar to described species. But considering how diverse and abundant they are likely to be (Chapter 2), the lack of more enterocytozoonid-like sequences in large-scale eDNA/generic amplicon studies showcases the need for direct/specialised study of this family and indeed, Microsporidia.

Four enterocytozoonids (*Hepatospora eriocheir, E. bieneusi*, *Desmozoon lepeotherii,* and *Enterospora canceri*), have been described within UK waters (Freeman & Sommerville, 2009; Graczyk *et al*., 2004; G. D. Stentiford *et al*., 2007, 2011). These four species are quite divergent, *D. lepeotherii* and *En. canceri* are on different branches of the two major enterocytozoonid lineages (*Nucleospora*-like species and *Enterocytozoon*-like species, respectively), *D. lepeotherii* infects fish (Freeman & Sommerville, 2009; Gunnarsson *et al*., 2017), *E. bieneusi* infects vertebrates (Desportes *et al*., 1985; Jiang *et al*., 2015), and *En. canceri* infects crustacea (G. D. Stentiford *et al*., 2007, 2011). Although *H. eriocheir* also infects crabs, it is considered a sister taxa/basal enterocytozoonid. With such divergent enterocytozoonid species described in UK waters, it is possible that there are undescribed intermediate species between *H. eriocheir, E. bieneusi*, *D. lepeotherii,* and *En. canceri*.

The aim of this study is to investigate the unknown diversity and distribution of the Enterocytozoonidae in the South West of the UK. Primarily screening environments that could come into contact with humans. This will involve urban areas in which water bodies reside around or run through, or wild areas close to farms. This will help give a better idea of their distribution and any patterns therein, such as any associations with the type of environment/invertebrate host. The identification of environments and hosts associated with enterocytozoonids will also enable the prediction of possible future threats to aquaculture, and human health. Screening of human-influenced areas allows for more direct

predictions, as the pathogenic species identified would be the most likely to enter the human food-chain. The reason this study is being carried out in the South West of England is that two species of Enterocytozoonidae have been described on the South coast (*En. canceri*-Weymouth, UK; *H. eriocheir*-River Thames, UK). This provides a good starting place to look for diversity, as these species are quite genetically divergent, which would lead to reason that there may be greater diversity found between these two evolutionary points. A second reason for the location of sampling is that many of the species described and many of the undescribed sequences discussed in Chapter 2, are found in, or directly connected to the Atlantic Sea. The South West would be under the most influence from the Atlantic Sea. There is also a study that has identified, through 18s rDNA amplicon, an Enterocytozoonidae-like organism found in crustacea from River Seine, France (Arundell *et al*., 2015). Showing that they are also found inland, in freshwater rivers.

3.2 Methods and analysis

Though the large majority of enterocytozoonid discovery has been based on direct histological identification (Chilmonczyk *et al*., 1991; Desportes *et al*., 1985; Lom & Dykoá, 2002; Tourtip *et al*., 2009), due to the difficult nature of collecting endoparasites and the exploratory nature of this study, it would prove more time-efficient and productive to extract DNA from water samples. To this end, a sampling plan was carried out that would enable efficient screening of enterocytozoonids in the aquatic environment. Enterocytozoonidae-like sequences from a previous unpublished freshwater study (KJ019847-50.1) from the UK were also added to downstream analysis to investigate if the species discovered were similar.

3.2.1 *Primer design*

To investigate enterocytozoonid biodiversity and distribution in South West UK, enterocytozoonid specific primers were designed with the aim of using them to detect the presence of the Enterocytozoonidae by PCR-screening filtered water samples. The primers were designed to amplify the 18s rDNA region, being the region most routinely used for microsporidian phylogenies, and would provide the best comparison. To pick the best region for a primer, an alignment was made of

all the described enterocytozoonids (Table 3.1-supplementary), novel sequences gained from the metagenome search (in Chapter 2), non-enterocytozoonid microsporidians found in aquatic organisms, enterocytozoonid hosts, and representatives from aquatic organisms. When appropriate regions were identified, primers were checked using Netprimer (http://www.premierbio soft.com/netprimer/netprlaunch/netprlaunch.html) to test for self-annealing and compatible annealing temperatures for primer pairs.

3.2.2 *PCR optimisation*

Three primer sets (Table 3.1) were designed with the second two being nested to ensure specificity. They were first tested on an extraction of *En. canceri* and *N. cyclopteri* to see if they worked on a range of known enterocytozoonids. Primers were optimised using a temperature gradient (52.9ºC, 53.7ºC, 54.6ºC, 55.7ºC, 56.6ºC, 57.6ºC, 58.6ºC, and 59.3ºC), to ensure the continued use of the best temperature for each particular primer pair. PCRs were carried out on 25µl reactions (12.5µl of gotaq master mix, from promega, 2µl of the forward, reverse

Table 3.1: Novel Enterocytozoonidae specific primers designed to amplify the 18s rDNA region

Primer ID	**Sequence 5'-3'**	**Expected length of amplicon (with erv1751)**
Efw118-2	GGCTCAGTAATRTTGCGVT	950bp
(Nested)Efw457	ATGGCTCCYACGTCCAA	787bp
(Nested) Efw691	TGYCYATKGTGGRTGCTGC	597bp
Erv1751	ATTGTATTGCRCTTGCDGC	N/A

primers and the template and 6.5µl of sterile water), consisted of an initial 10 minutes denaturing at 95ºC, then 35 cycles of [denature at 95ºC for 35 seconds, annealing at 55ºC for 45 seconds and extension at 72ºC for 1 minute 30 seconds], final extension at 72ºC for 10 minutes and a rest period at 4ºC. Amplicons were run on a 1.5 % agarose gel, at 100 volts for 30minutes, to verify amplification, and were stored at -20ºC. Lastly, the sensitivity of the primers was tested by taking Six 1/10 serial dilutions of an *En. canceri* extraction, starting at a starting concentration of 8.5 ng/µl, as they were designed to be used on environmental samples, and any target DNA is likely to be of low concentration. To test their

efficacy on pooled filtered samples, they were tested on past UK environmental samples that were collected in 2011 (Table 3.2) for a separate project (B. A. P. Williams *et al*., 2018) but have been collected and prepared using similar methods proposed for this study. The primers were also tested against other non-enterocytozoonid microsporidian extractions to ensure they would not amplify more closely related organisms.

Table 3.2: Table of 'test' samples screened in this study

Location	Sample ID	Number of samples	Sample type
Pond, Exeter, UK	PS	27	Invertebrates
Burrator reserve, Dartmoor, UK	O	11	Filtrate fresh water
River Exe, Exeter, UK	ES	35	Water
Plymouth, UK	CPR	30	Sediment
Ria Formosa, Portugal	ASW	32	Plankton
Southampton, UK	DB	16	Copepods
River Teign, Teignmouth, UK	EST	11	Marine sediment
Not available	BW	9	Water
Torquay, UK	R	28	Marine water
Exeter, UK	PN	20	Pond water
Burrator reserve, Dartmoor, UK	BG	43	Bog water
Not available	ESC	16	Marine water

3.2.3 *Sampling sites*

To gather a better understanding of the range and distribution of the Enterocytozoonidae, samples were taken along an estuarine transect. The transect locations for this study focused on the five rivers running through the largest urban areas in the South West: Exeter/Exmouth-River Exe, Plymouth-River Tamar/Plym, Teignmouth-River Teign, and Barnstaple-River Taw (Figure 3.1). The River Exe is a 96 km river, that has a large fishing community centred around it. The River Exe has also garnered scientific interest over the years, with regards to the sustainability of salmon stocks (Edbrooke, 2004; Nott & Beale, 1968). The River Teign is a 50 km long river and used to be a commercially important fishing port for Teignmouth (Bielby, 1963) but is still used for personal fishing for finfish and shellfish (Inshore Fisheries and Conservation Authority, 2019). The River Taw is a 72km river that is used for commercial fishing (RTFA (River Taw Fishing Association), 2020), that supports trout and salmon all year-

round. It also acts as one of only two refuges to the endangered freshwater pearl mussel (West Country River Trust, n.d.), an indicator species on the decline. The River Tamar is a 98km river is one of the three index rivers in England and Wales, with a long record of salmon and trout fishing (Environment Agency, n.d.). It is a SSSI (Site of Specific Scientific Interest) due to its unique conditions that support rocky reefs in low salinity conditions (Plymouth Gov, n.d.). All rivers also support a large number of wildfowl and wader species, providing ample means of dispersal to surrounding water bodies. Primarily, this study aims to filter pelagic crustacea as, besides fish, most species of Enterocytozoonidae seem to infect crustacea as a primary host or intermediate host. For this reason, crustacea are likely the most important hosts in the life cycle and in the evolution of the group as a whole. Coastal plankton were also sampled, as all transect rivers mentioned above lead into the English Channel, which may show associations as a result. Here marine copepods were the main target of the sampling as they have been shown to be intermediate hosts for other species within the family (*P. theridion* and *O. papernae*). These samples were collected while on the CEFAS ship, Endeavour, as part of its annual PELTIC survey. This survey covers the Western Channel and part of the Celtic Sea (Figure 3.1). Samples were also collected from the North Sea, from Dowsing, Dogger, West Gabbard, and Warp. These were collected as part of a smaller survey, so fewer samples were collected. However, it offered the opportunity to extend the survey of enterocytozoonid diversity to the North Sea.

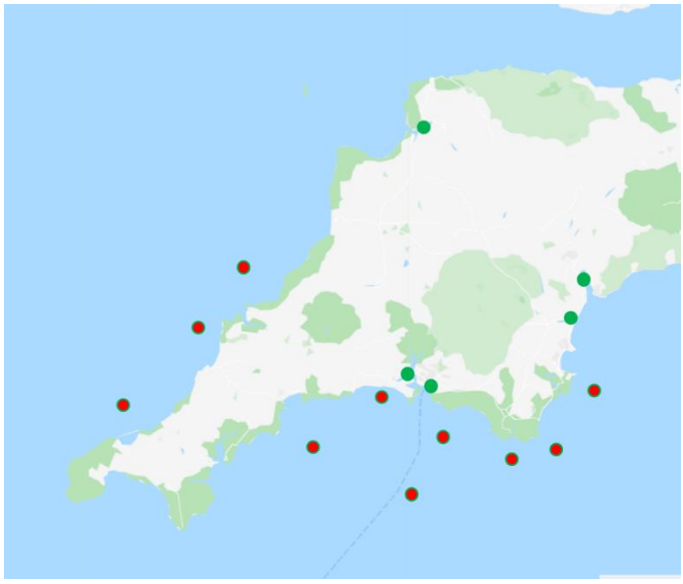


Figure 3.1: Sampling locations of major South West UK estuaries, and coastal transects. (Estuarine samples (green dots) include the River Taw, Tamar, Teign, Exe, and Plym. Coastal transects (red dots) follow around the “foot” of the UK)

3.2.4 Transect sampling:

3.2.4.1 *Estuarine transects*

Water samples were collected in Autumn 2018, Spring, Summer, and Autumn 2019, to try to capture the varying spawning times of different aquatic invertebrates. Transects moved up from the mouth of the estuary, with an initial sample at the mouth and subsequent samples were taken roughly after every 1 km (terrain allowing) for 4 km (Table 3.3). 10 L of water were collected along the transect from approximately the top 30 cm of the water body, aiming for pelagic organisms and avoiding too much sediment, as this would clog the filter. This was then filtered through a 100 µm cell strainer, using a modified 50 ml falcon tube as a funnel (Figure 3.2). The filtrate on the cell strainer was then washed into a clean 50 ml falcon tube, using absolute EtOH. Three samples were taken per sampling effort (12 in total for each estuary). Samples were stored at -20^{o}C until DNA extraction. Due to the high sensitivity of environmental DNA methods, there is a high risk of false positives due to contamination. Additionally, while on the transect, when seen, larger invertebrates were collected to give a better understanding of the types of invertebrates the Enterocytozoonidae may be infecting in the same location as the transects. Larger invertebrates were only collected from two locations a shore crab (*Carcinus maenas*) from the River Exe and mussels (*Mytilus* sp.) from the River Tamar and the River Thames, due to availability.

Several studies have looked at minimising these risks, by taking steps to avoid contamination in the field and the lab (Goldberg *et al*., 2016). To avoid contamination between transect sites, a new cell strainer was used for every site, and the falcon funnel was sterilised using a strong detergent, Virkon, and washed using absolute EtOH. To assess the risk of contamination through this method, sterile water was taken through every step per sampling effort.

Table 3.3: Estuarine transect coordinates of South West Rivers.

Site	Location	1st Transect	2nd transect	3rd transect	4th transect
River Exe	Exeter, UK	50.653758, -3.455638	50.656828, -3.457201	50.662887, -3.467046	50.627318, -3.447105
River Teign	Teignmouth, UK	50.53861, -3.575150	50.539618, -3.545451	50.540356, -3.525879	50.540672, -3.517612
River Taw	Barnstaple, UK	51.061599, -4.181147	51.068661, -4.178822	51.073233, -4.167554	51.075457, -4.160918
River Tamar	Plymouth, UK	50.389492, -4.192594	50.397406, -4.203709	50.407329, -4.201536	50.416634, -4.196347
River Plym	Plymouth, UK	50.360869, -4.127425	50.360846, -4.116228	50.365865, -4.109044	50.374796, -4.104569
River Thames	London, UK	51.487829, -0.246872	N/A	N/A	N/A
Lumpini Park lake	Bangkok, Thailand	13.731659, 100.540661	N/A	N/A	N/A
Queen Sirikit Lake	Bangkok, Thailand	13.806168, 100.550150	N/A	N/A	N/A

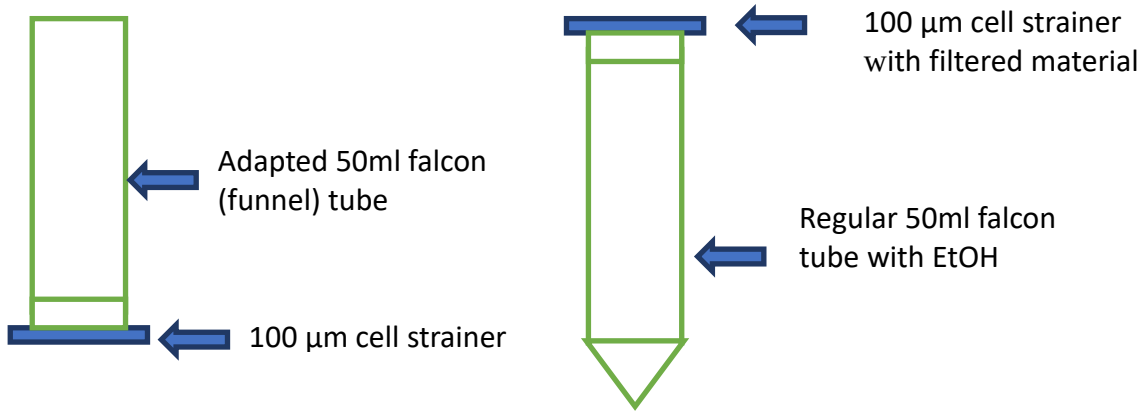


Figure 3.2:Diagram showing filtration method employed in this study. 10L of estuarine water is passed through the adapted 50ml falcon tube (on the left) with a 100 µm filter attached, the filter is then placed on a regular 50ml falcon tube (right) and the filtride is washed into the tube with absolute EtOH.

3.2.4.2 *Coastal transect*

All coastal plankton sampling was carried out between the 6th of October 2018 and the 10th of November 2018. A weighted 0.5 m diameter ring net with a filter diameter of 80 µm was used to filter plankton from the water column (Figure 3.3). Once the ship (Endeavour) was stationary, the ring net was deployed to the maximum depth of the sampling location (Table 3.4) and pulled up at a quicker rate to capture the zooplankton in the water column. All sampling was done after 12:00 am as that is when the zooplankton rise to the surface of the water column. When the ring net was brought to the surface, the outside of the net was washed, to push any plankton that may be stuck on the net down to the filter. The filtride was washed out of the filter using filtered seawater. The samples were then equally split using a Fulsom splitter, with one half being placed in absolute EtOH for DNA extraction and the other placed in Davidson's fixative for histology. The subset for histology was left in the Davidsons for 24h before being placed in 70% IDA.

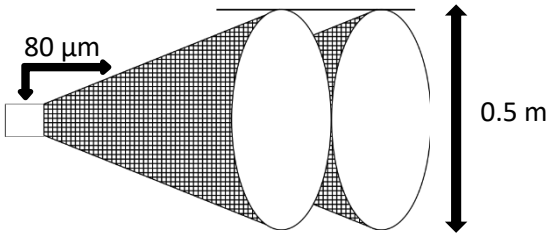


Figure 3.3: 0.5 m ring net used to filter zooplankton from the water column, attached with an 80 µm diameter filter.

Table 3.4: Sampling locations and depth of PELTIC coastal transect

Sample ID	Station number	Prime station	Coordinates	Depth (m)
PEL 1	46	50	50.48035, -5.088	25
PEL 2	48	79	50.70438, -4.932	37.6
PEL 3	61	82	50.35027, -5.541	42.3
PEL 4	140	23	50.16788, -4.602	56.6
PEL 5	143	21	50.00543, -4.334	69
PEL 6	151	20	50.32048, -4.332	21
PEL 7	154	17	50.16042, -4.077	50.8
PEL 8	214	15	50.1652, -4.802	53.8
PEL 9	218	13	50.1646, -3.570	52.5
PEL 10	228	10	50.33942, -3.303	50.5

3.2.5 Additional estuarine/invertebrate sampling:

Though out of the sample scope for this study, a chance arose to collect samples from the River Thames at Chiswick (Table 3.3) and freshwater sources in Bangkok, Thailand. Though it was not possible at the time to collect samples in a comparable way to the rest of the study, screening samples from waterbodies going through the most populated cities in the UK and Thailand (meaning a higher perceived risk, due to the density of people in the city leading to a raised likelihood of interaction and transmission to livestock), would be relevant to the study. Three water samples were taken from UK and Thai locations, using the same protocol mentioned above, along with 40 *Gammarus* sp., 20 Nereid worms, and five Chinese mitten crabs (*Eriocheir sinensis*) just from the UK. All were placed in absolute EtOH for DNA extraction.

3.2.6 *Sample preparation and DNA extraction*

50ml falcon tubes (with filtrates) were placed in a centrifuge for 10 minutes at 3000 x g to pellet the filtrate. The supernatant was discarded and 500 µl of TE buffer was added. The re-suspended filtrate was then transferred to a 2 ml screw-cap vial and placed back in -20^{o}c, defrosting was a part of the tissue lysis. Crabs were dissected and the muscle tissue and hepatopancreas were taken. *Gammarus sp*. were homogenised using a sterile mortar and pestle with 2 ml 10/1 TE buffer, and the resultant homogenate was placed in 2 ml screw-cap vials for extraction. Cross-sections were taken from mussels the same day as collection, placed into absolute EtOH immediately, and stored in -20^{o}c for later extraction. All DNA from this study was extracted from 500 µl of samples using phenol-chloroform extraction. Samples were resuspended, and if needed, transferred to a 2 ml screw-cap tube along with 10 µlof 10% SDS, 2 µl of 10 mg/ml RNAse, 400 µl of (0.17 - 0.18 mm) glass beads (Fisher Scientific™), and 800 µl of pH 8.0 phenol. The samples were then placed into a tissue lyser at 4500 rpm for 45 seconds and centrifuged at 10000 x g for 10 minutes. The top aqueous layer was transferred to a fresh Eppendorf and 400 µl of chloroform was added. This was mixed by inversion (x5), and the mixture was centrifuged at 10000 x g for 10 minutes and the top aqueous layer was transferred to a new 1.5 ml Eppendorf.

3.2.7 *EtOH precipitations*

1/10 of the final volume of 3M sodium acetate was added followed by 2/1 of the final volume of ice-cold absolute EtOH. The solution was mixed through inversion. This was stored on ice for 1hour and centrifuged (at -4ºC) for 30 minutes at 10,000 x g. The supernatant was discarded, and the pellet was washed in 300µl of ice-cold 75% EtOH. The solution was centrifuged for a further 15 minutes at 10,000rpm (at -4ºC). The supernatant was removed and spun at full speed to remove any residual EtOH. The pellet was left to dry for 10-20 minutes in a 60ºc incubator. The dried DNA was resuspended in 40 µl of sterile, nuclease-free water.

3.2.8 *PCR purification, sequencing, and cloning: (Refer back to PCR optimisation)*

Positive amplicons were cut out of the gel at the size range expected and purified using QIAquick PCR Purification Kit and sent to Eurofins, for initial sequencing. Sequencing reactions were carried out in one direction, using the nested efw691-erv1751 amplicons, giving a product that should be around 650bp. All sequences were assessed by eye in Sequencher version 5.4 (Nishimura, 2000) and viewed and edited in BioEdit version 7.2.5 (Hall, 2013). Those that were enterocytozoonid were cloned, to capture the full diversity found among the amplicons. Samples were ligated into a pGEM®-T Vector, following the Promega standard protocol for T4 DNA ligase, and transformed into One Shot TOP 10 *Escherichia coli* competent cells. 50 µL of purified amplicon and 50 µL of X-Gal Solution (20 mg/mL) were then placed onto a 1% ampicillin (50 µg/ml) agar plate (1:1000) and left in a 37ºc incubator overnight. Five colonies were taken from each plate and placed into separate vials of liquid media, in a rocking incubator at 37ºC overnight. These were pelleted at 10,000 x, in a tabletop centrifuge and plasmids were isolated from cells using a QIAprep Spin Miniprep kit. Plasmids are sent for sequencing using T7 plasmid primers.

3.2.9 *Sequence analysis*

Initial sequence identification was carried out using BLASTN (Altschul *et al*., 1990), under default parameters using NCBI's Nucleotide collection database. To better understand the relationship of these novel sequences to described

enterocytozoonids, all sequences produced in this study were aligned to published enterocytozoonid 18s rDNA (*E. hepatopenaei, En. canceri, E. bieneusi, H. eriocheir, En. nucleophila, P. theridion, D. lepeotherii, N. cyclopteri, N. salmonis, O. papernae, and N. braziliensis*) (Table 3.1-supplementary) sequences to create a phylogenetic tree. All sequences in this study were aligned in MAFFT version 7 (Katoh & Standley, 2013), under default parameters, and tested for the best fitting substitution model using MEGAX (Kumar *et al*., 2018). Evolutionary relationships were inferred using maximum likelihood, implemented in RAxML-HPC ver 8.2.4 (Stamatakis, 2014) [Generalised time-reversible (GTR) with the GAMMA rate of heterogeneity and invariable sites (GAMMA+I)]; 1000 bootstrap replicates were run on all trees. The tree GUI TreeGraph version 2.1.5 (Stöver & Müller, 2010)was used to view the trees built. A pairwise divergence analysis was also done to assess the difference between closely related sequences, using the Tajima-Nei (Tajima & Nei, 1984) model in MEGAX. (Kumar *et al*., 2016).

3.3 Results

In total, 232 of the 2011 test environmental samples (from twelve separate locations, a mix of both estuarine and fresh water and marine), 60 estuarine transect samples (from five different estuaries), and 14 coastal transect samples (from 14 different locations) and seven invertebrate samples (from two different locations) were screened (Table 3.6).

3.3.1 *Validation/optimisation of family-specific PCR primers*

To test the specificity and effectiveness of the primers, the primers were tested on enterocytozoonid, non-enterocytozoonid, and filtered environmental samples. The primers successfully amplified the *En. canceri* control, while not amplifying any product from any of the non-enterocytozoonid samples tested so far. They have amplified all DNA concentrations from the serial dilution series (Figure 3.4), and have currently amplified three described species, *En. canceri*, *N. cyclopteri* and *E. hepatopenaei*. Though the primer sets were successful over a range of annealing temperatures, the one that gave the greatest concentration of amplicon

across all primer sets was 55ºc. This temperature was used for all PCRs in this study.

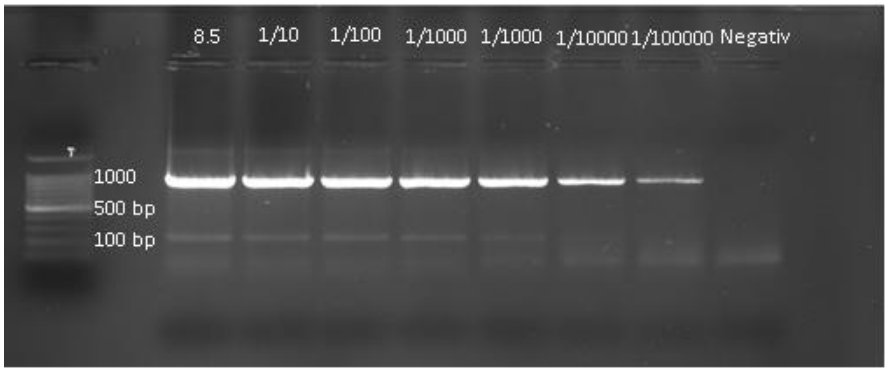


Figure 3.4: Gel image of amplification of serial dilutions of *En. Canceri* DNA, using primer pair efw118-2 – erv1751, amplifying the 18s rDNA region. Run on a 1.5% agarose gel run at 100V for 30 minutes.

3.3.2 *Sequencing of negative controls*

To test for possible routes of contamination, negative controls (sterile water) were taken at each sampling effort and put through each step the filtered samples were (per sampling effort). All negative controls showed a lack of amplification for all sampling efforts (Figure 3.5). As such, all positives are taken as true positives for the rest of the analysis.

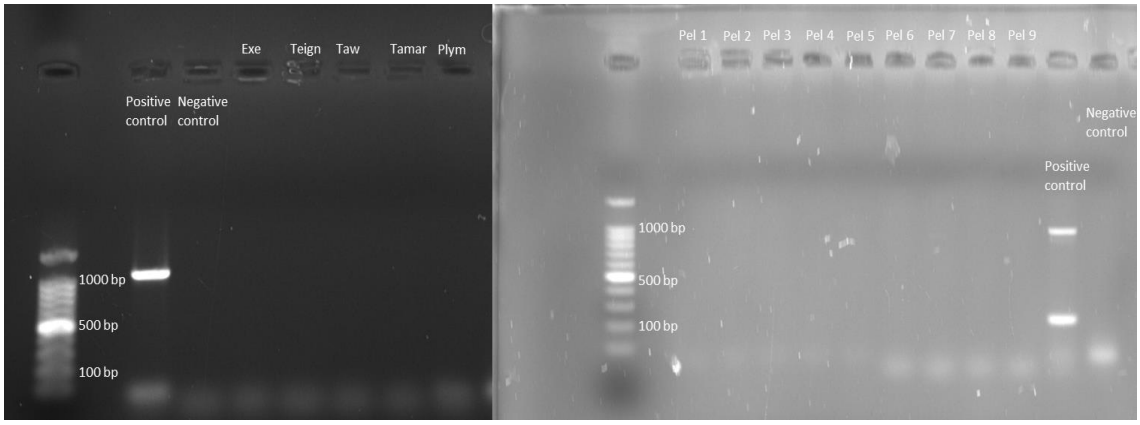


Figure 3.5: Negative control amplicons for estuarine transects (left) and for PELTIC coastal transect (right) samples taken through the extraction and amplification. Run on a 1.5% agarose gel run at 100V for 30 minutes.

3.3.3 *Prevalence of enterocytozoonids in the environment*

The prevalence of positive amplicons varied depending on the source/location of the samples. Larger invertebrate samples showed the highest percentage of enterocytozoonid-like amplicons, with four (57.1%) sequenced amplicons. 50 (21.5%) were positive from the 2011 samples across twelve of the locations, nine (15%) from the estuarine transect samples across three locations (two rivers had no positive amplicons), three (28.6%) from the Celtic Sea/English Channel coastal transect and one (25%) from the Northern Sea transect. The River Exe transect showed the greatest number of positive amplicons (33.33% of River Exe samples), and diversity (three species). Cloning was largely unsuccessful, of the 34 clones sequenced only four colonies were successfully sequenced. Time constraints prevented further cloning. Two of these (River Exe Exeter UK-Et2-clone2, River Exe Exeter UK-Et2-clone4) are from the same isolate, so may provide some indication towards diversity found in one sample.

3.3.4 *Datasets composition*

As some sequences varied in length (sequencing errors), covering different regions of 18s rDNA, two final aligned datasets were constructed spanning the range of all sequences included in this study (Table 3.5). The aligned datasets (range: 241-395bp) showed a similar amount of phylogenetically informative and variable sites (Table 3.5), showing no bias for a particular region of the 18s rDNA. Datasets have ranging availability of certain sequences because of the variability in sequence lengths, even among the described species. As such *E. bieneusi* is not represented in the analysis of dataset 2.

Table 3.5: Aligned datasets used for phylogeny and pairwise distance matrices

Dataset	Number of sequences	Aligned length	Phylogeny informative sites/percentage	Conserved Sites/percentage	Variable sites/percentage
Dataset 1	57	395	116/29.37%	178/45.06%	195/49.37%
Dataset 2	38	241	59/24.48%	132/54.77%	95/39.42%

3.3.5 *Taxonomic composition of enterocytozoonids*

Sequences were BLASTN searched to ascertain the identity of sequences obtained in this study. Sequences fall most closely to two known enterocytozoonids, *En. canceri* and *E. hepatopenaei* (Table 3.6). The most prevalent is *En. canceri*, being found in 12 (66.67%) out of 18 positive locations. Both species show a greater distribution than previously described, notably, *En. canceri*-like sequences were also found in one of the samples from Thailand (Lumpini Park, Bangkok) and the shore crab (*Carcinus maenas*) from the River Exe (Table 3.6). *E. hepatopenaei*-like sequences were found in five (21.74%) UK estuarine, freshwater, marine samples (Sample ID: BW, ESC, River Exe, Thames River, English Channel) and one of the freshwater Thai samples (Queen Sirikit Park), marking the first instances of *E. hepatopenaei*-like sequences outside of a shrimp farm. *En. canceri*-like sequences showed more sequence similarity (range: 98.56%-100%) than *E. hepatopenaei* (range: 92.7%-100%). There were additional enterocytozoonid-like amplicons, *E. bieneusi* (River Exe, UK) and *N. cyclopteri* (English Channel, UK), though both were short sequences (<150bps) so were not included in the analysis. The records for the locations of some samples (Sample ID: BW, ESC), have been lost. However, they were included as it is known they were from the UK, and would still provide data on the diversity found in the UK.

Table 3.6: Samples screened in this study and their sequence identity to described enterocytozoonids assessed using BLASTN on 18s rDNA sequences. (*denotes 2011 samples)

Location	**Sample ID**	**Number of samples**	**Sample type**	**Positive PCR**	**Sequence similarity (I= Percent identity; C=Query cover)**
***Pond, Exeter, UK**	PS	27	Filtrate-soil	Yes	*En. canceri* (I:100% C:100%)
***Burrator reserve, Dartmoor, UK**	O	11	Filtrate-fresh water	Yes	N/A
***River Exe, Exeter UK**	ES	35	Filtrate-brackish water	Yes	*En. canceri* (I:99.15-99.69% C:69-82%)

***Clyde Estuary, Plymouth, UK**	CPR	30	Filtrate-brackish water	Yes	*En. canceri* (I:99.45% C: 61%)
***Ria Formosa, Portugal**	ASW	32	Filtrate-marine water	Yes	N/A
***Southampton, UK**	DB	16	Filtrate-marine water	No	N/A
*** River Teign, Teignmouth, UK**	EST	11	Filtrate-brackish water	Yes	*En. canceri* (I:98.90-99.68% C:61-100%)
***Not available**	BW	9	Water	Yes	*E. hepatopenaei* (I:94.41-99.55% C: 99-100%)
***Torquay, UK**	R	28	Filtrate-marine water	Yes	N/A
***Pond, Exeter, UK**	PN	20	Filtrate-fresh water	Yes	*En. canceri* (I:99.58% C:100%)
***Burrator reserve, Dartmoor, UK**	BG	43	Filtrate-fresh water	Yes	*En. canceri* (I:99.46% C:100%)
***Not available**	ESC	16	Filtrate-marine water	Yes	*E. hepatopenaei* (I:100% C:98%)
Dowsing, North Sea UK	DW1	4	Filtrate-marine water	Yes	*En. canceri* (I:99.4%/C:60%)
River Tamar, Plymouth, UK	P1-4 (a-c)	12	Filtrate-brackish water	Yes	*En. canceri* (I:94.67-99.66% C:58-100%)
River Plym, Plymouth, UK	S1-4 (a-c)	12	Filtrate-brackish water	Yes	*En. canceri* I:97.09-99.31%/C:100%)
River Exe, Exeter, UK	Et1-4 (a-c)	12	Filtrate-brackish water	Yes	*E. hepatopenaei* (I: 92.7-99.5% C:98-100%); *E. bieneusi* (I:98.36%/C:96%); *En. canceri* (I:99.5%/C:41%)
River Exe, Exeter, UK	Etc1	1	Invertebrates	Yes	*En. canceri* (I:98.56%/C:41%)
River Teign, Teignmouth, UK	Tt1-4 (a-c)	12	Filtrate-brackish water	No	N/A

River Taw, Barnstaple, UK	Tat1-4 (a-c)	12	Filtrate-brackish water	No	N/A
English Channel/Celtic Sea, UK	Pel1-10	10	Filtrate-marine water	Yes	*E. hepatopenaei* (I:98-99.56%C: 78-97%); *En. canceri* (I:98.18-98.82%/C:51-52%); *N. cyclopteri* (I:100%/C:20%)
Thames River, London, UK	TM	1	Invertebrates	Yes	*E. hepatopenaei* (I:100%C: 94%)
Thames River, London, UK	TP	40	Invertebrates	Yes	*En. canceri* (I:99.4%/C:49%)
Lumpini Parke, Bangkok, Thailand	LP	5	Filtrate-fresh water	Yes	*En. canceri* (I:98.76% C: 96)
Queen Sirikit Park, Bangkok, Thailand	QS	3	Filtrate-fresh water	Yes	*E. hepatopenaei* (I:99.74% C: 99%)

3.3.6 *Phylogeny and pairwise estimates: diverse Enterospora genus*

In all datasets, the accepted relationship between described enterocytozoonids has remained the same (Figure 3.6 and 3.7). With the *Enterocytozoon* being more closely related to *Enterospora* (crustacean infecting enterocytozoonids), and *Nucleospora* being more closely related to *Paranucleospora theridion* (synonym: *Desmozoon lepeotherii*) and *Obruspora papernae* (fish infecting enterocytozoonids). As was noted with the BLASTN results (Table 3.6), the majority of the diversity found is around the *En. canceri* node (Figure 3.6 and 3.7). There was a lack of geographic-based grouping, the majority of the sequences forming a polytomy within the *En. canceri* group. There are low levels of genetic divergence observed between *En. canceri* and *En. canceri*-like sequences, with pairwise differences (p-values) ranging from 0-0.011 (average: 0.008) on dataset 1 and ranging 0-0.041 (average: 0.014) on dataset 2. However, *E. hepatopenaei*-like sequences showed a bit more genetic divergence from the described *E. hepatopenaei* sequence for dataset 1, p-values ranging from 0-0.040 (average: 0.014), while average estimates for dataset 2 are similar ranging from 0.018-0.019 (average:0.018).

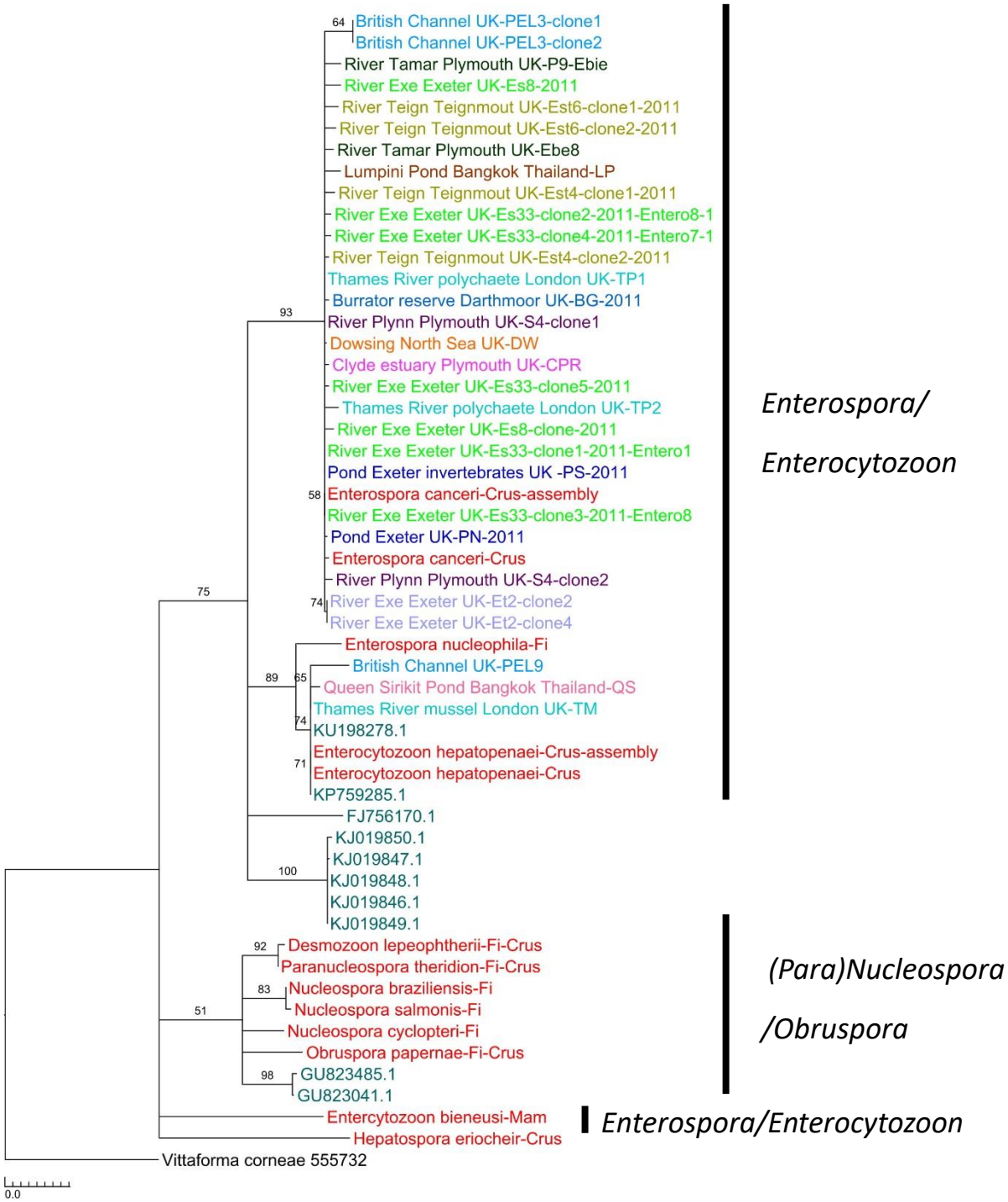


Figure 3.6: Enterocytozoonidae phylogeny from dataset 1, built-in RAxML-HPC (GTRGAMMA), showing the relationships between environmental amplicons and described species. DNA was amplified based on the 18s rDNA gene. Scale bar intervals represent 0.01 substitutions per site for branch length. Numbers above nodes indicate bootstrap support values, assessed from 1000 bootstrap replicates. Nodes with a value less than 50 were collapsed. (The red names are

from described species while all others are coloured according to the location they were sampled from)

A greater sequence divergence is also seen between *E. hepatopenae*i-like sequences ranging from 0-0.040 (average: 0.020) on dataset 1 and a p-value of 0.019 for dataset 2, with only two *E. hepatopenaei*-like. Whereas the pairwise differences between *En. canceri*-like sequences range from 0-0.025 (average:0.006) on dataset 1 and ranging from 0- 0.14 (average: 0.002) on dataset 2. Although p-values were largely low, there were outliers for *En. canceri*-like and *E. hepatopenaei*-like sequences on both datasets. The sequence from Lumpini, Bangkok (LP) shows the greatest divergence from *En. canceri* (p-value: 0.025), . Dataset 2, a clone amplicon from the River Exe (Et2-clone2) had the greatest divergence from *En. canceri* (p-value: 0.041) 0.025 more than the average. The isolate from the British Channel (PEL9) had the greatest divergence from *E. hepatopenaei* (p-value: 0.040).

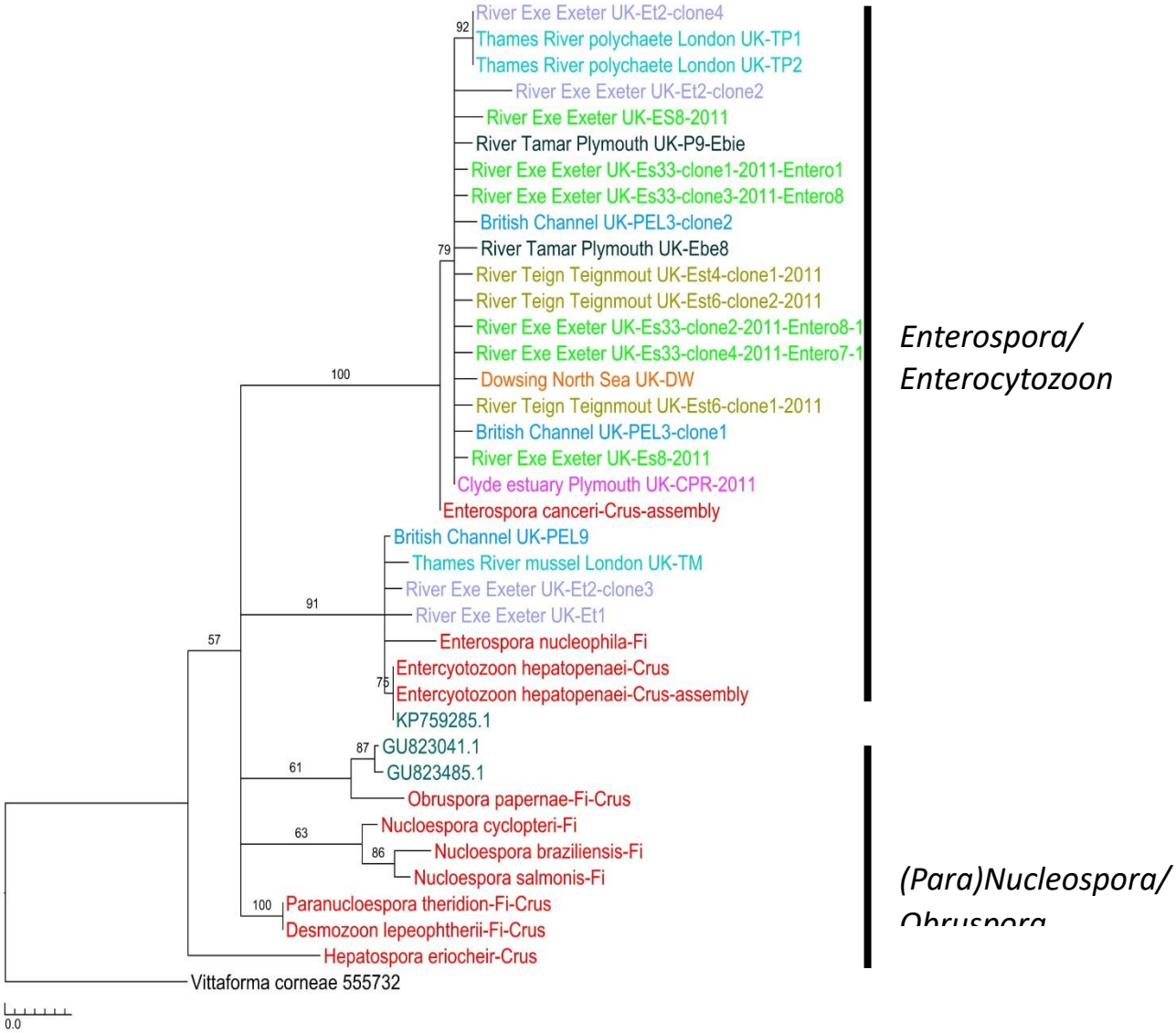


Figure 3.7: Enterocytozoonidae phylogeny from dataset 2, built-in RAxML-HPC (GTRGAMMA), showing the relationships between environmental amplicons and described species. DNA was amplified based on the 18s rDNA gene. Scale bar intervals represent 0.01 substitutions per site for branch length. Numbers above nodes indicate bootstrap support values. Nodes with a value less than 50 were collapsed. (The red names are from described species while all other colours are according to the location they were amplified from)

3.3.7 Genetic change in *En. canceri over time*

The discovery of *En. canceri* in the 2011 'test' samples and the 2018 transect samples in the same/close locations, provided an opportunity to look at possible temporal differences between *En. canceri*. Pairwise distances between *En.*

canceri sequences from the same estuaries, Tamar River, UK (EST (2011); P9 (2018)) showed no divergence, with an average p-value of 0. However, samples from the River Exe, UK (Es33 (2011); Et2 (2018)) showed some divergence with an average p-value of 0.01. Dataset 2 showed the same trend for both Tamar River and River Exe; however, River Exe shows much less divergence (p-value: 0.002).

3.4 Discussion

This is the first study looking at the distribution and diversity of the Enterocytozoonidae, in the South West of the UK. It primarily aims to catalogue any enterocytozoonid diversity present and assess the risk of the parasite's location in association with surrounding farms and human residence. Novel primers were designed and used in this study to address the proposed aims. Implicit in the study's design is the assumption that the primers used are specific enough to make inferences about absence screenings. The validation steps suggest that the primers are capable of amplifying a range of available Enterocytozoonidae and DNA concentrations as low as 1/1000000 of 8.5 ng/ul, thus likely being able to amplify trace amounts of enterocytozoonid DNA in filtered samples.

3.4.1 *Enterocytozoonid prevalence and distribution*

The distribution of the Enterocytozoonidae displayed in this study is greater than that currently described. This has been shown from the widespread occurrence of *En. canceri*-like sequences, having prior only been described in Weymouth. Additionally, the prevalence of *En. canceri*-like sequences across the samples in this study were surprisingly high, considering prior observations. The described prevalence for *En. canceri* in the European edible crab (*Cancer pagurus*) in its original description was less than 4%, of 330 crabs (G. D. Stentiford *et al*., 2007). Whereas in this study, sequences are highly similar to that of *En. canceri* are present in 66.67% of the successfully screened samples. This suggests that *En. canceri* could be far more prevalent than first estimated and that it is highly possible that European edible crabs are not its main host (which was suggested,

because of its low prevalence). The presence of *E. hepatopenaei*-like sequences in UK waters is novel and somewhat unexpected as this has previously solely been described from South East Asia and Central America. However, it has been found to infect other invertebrates (Tang *et al*., 2015b), and there is evidence for its occurrence (if only similar sequences) in marine-based environments (Chapter 2). In addition, its closest described relative (*En. canceri*) has only been described and characterised in the UK, so it is likely that *E. hepatopenaei*-like Microsporidia will be found in, if not endemic to, the UK. Despite this, the *E. hepatopenaei*-like sequences were less abundant and less conserved (average p-value: 0.020). *E. hepatopenaei*-like had nearly 3 x as many sequence differences with the described *E. hepatopenaei* sequence than is found between the described *En. canceri* sequence and the *En. canceri*-like sequences (average p-value: 0.006). This suggests a greater diversity in the *E. hepatopenaei*-like lineage, in the samples tested in this study. The occurrence of both *En. canceri*-like and *E. hepatopenaei*-like sequences in only one location (5.88% of positive locations) may suggest that there is not much of an overlap in host range/type. Although distribution has been shown to be quite wide for the *Enterocytozoon/Enterospora* clade of the Enterocytozoondiae, prevalence was not equally as high. Differences in prevalence between samples could be due to a number of factors, for example, host species abundance may not have been representative in sampling efforts. Prevalence found in this study may also indicate the actual prevalence of enterocytozoonids in the environment. More work, incorporating data on invertebrates in the transects samples, will have to be incorporated to control for the effect of the (possible) host-species composition of the transect.

With regards to distribution, there did not appear to be any clear patterns based on substrate type (Freshwater or marine/estuarine water) or geography. For example, *En. canceri*-like sequences were found across all sample substrates and in all transects that were positive. This suggests, at least in the case of *En. canceri*, that it may have a wide range of hosts (adding more support for their opportunistic nature). The two samples that showed the greatest level of sequence difference were from a shore crab sample (River Exe, UK: Etc1) and from Thailand (Lumpini Park, Thailand: LP). This would make sense given the geographic distance; regarding the samples from the shore crab, it is possible that it indicates a shore crab infecting variant. Suggesting a possible change in

host range, allowing for a greater divergence in a conserved gene, possibly due to genetic drift due to less geneflow.

The discovery of enterocytozoonid-like sequences in freshwater is also novel and it had been thought that this clade was an exclusively marine host-infecting parasite (with the exception of *E. bieneusi*). Previous studies have also found enterocytozoonid-like sequences in crayfish in France (Arundell *et al*., 2015), though none were as similar to described species as the sequences identified in this study. This could mean that, though there was no sign of freshwater-based enterocytozoonids in Chapter 2 (metagenome analysis), the Enterocytozoondiae, in particular close relatives of *En. canceri* and *E. hepatopenaei* may also be capable of infecting freshwater invertebrates. This could vastly expand the possible distribution of the Enterocytozoonidae and would allow for expansion of the species inland.

Finding *En. canceri*-like sequences so widespread and genetically variable (when using 18s rDNA), and *E. hepatopenaei* to a lesser extent, throughout human-influenced areas, suggests a possible danger for both aqua/agriculture. Both are known to infect farmed or wild-caught crustacea, with *E. hepatopenaei* rapidly reaching a very high prevalence in farmed shrimp over recent years. This should be considered, as the UK starts to plan more farming of crustaceans. In addition, some of the rivers/estuaries sampled in this study are used for recreational use (Exe estuary has water sports facilities), which could facilitate passive ingestions of water and possibly spores. As mentioned above, these rivers are also used for recreational fishing, meaning that improper preparation of fish could also result in the ingestion of spores. The fact that many Microsporidian species have a broad host range and the potential to infect humans, opportunistically, adds to the possible dangers.

3.4.2 *Phylogenetic classification*

With the majority of identities falling in-between 97% and 100%, it is likely that these samples are highly similar in nature to two species of Enterocytozoondiae: *E. hepatopenaei* and *En. canceri*. Phylogenetic analysis (Figure 3.5 and 3.6) also suggests this, with all sequences retrieved by this study grouping within these two species.

The lack of (*Para*)*Nucleospora*-like sequences is likely the result of the sampling methods used which had the aim of collecting small filtered planktonic life. This strategy, as noted in Chapter 2, was not designed to sample fish, the described host species for the enterocytozoonids in that lineage. It could also suggest a lack of occurrence of members of this branch of the Enterocytozoonidae in SW UK estuaries, and perhaps less diversity in general.

3.4.3 *Summary*

In summary, this study has shown widespread *En. canceri*-like sequences, and with how widespread and close to human influence the Enterocytozoonidae are, they could become a greater risk to farming than they currently are. It is already likely that the Enterocytozoonidae infect a range of invertebrates around UK shores and estuaries. Their host-shifting propensity has been shown in numerous species within the Enterocytozoonidae, making it likely that the continued expansion of human populations and farming monoculture, will border on environments in which enterocytozoonids naturally occur.

This study also suggests (given the sequence similarities) that there may be a frequent source of dispersal for these species, for them to have such high similarities across the South West, both in estuaries and coastal habitats. The presence of *E. hepatopenaei*-like sequences in the UK and *En. canceri*-like sequences in Thailand also suggest host assisted/anthropogenic dispersal between the two distant countries.

The 18S rDNA is likely too conserved to be able to define novel species, in organisms that are as widespread and opportunistic as the Microsporidia. A new marker needs to be used when looking at environmental samples, as there may be missed diversity not being picked up by the 18S rDNA region. An example of this is the considerable amount of work using the ITS marker for genotyping *E. bieneusi* (though there are misgivings on how this is too widely used). However, comparing these environmental sequences to described species would require widespread use of alternative markers on described species, and the currently widely used region is 18S rDNA. Until this happens, a proper estimation of the

diversity of enterocytozoonid/Microsporidia, using amplicon-based methods, cannot truly be made.

3.4.4 *Further research and limitations*

Further research into the diversity of this family further inland in freshwater bodies would be able to tell us if the composition of the species of enterocytozoonid changes as there are more freshwater hosts. This would help assess the risk to terrestrial livestock that are given un-treated fresh water to drink. Though there has only been one reported enterocytozoonid species infecting terrestrial mammals, there are likely many more undescribed. The investigation into associated host species would also aid in assessing the risk to inland livestock. Although pooling filtered environmental samples has many benefits, especially when regarding pathogens, there are limitations to the inferences that can be made on the origin of the spores sequenced due to a lack of direct histological work. As a result, host discovery is not directly possible, which makes it difficult to make further suggestions on how particular species may disperse using a host likely more well known distribution. Additionally, the 18s rDNA region is not the most variable region, as such, it may not be the best region to delimit species. However, it does provide for the best comparison, in terms of the species described.

3.5 Supplementary tables and figures

Table 3.1: Enterocytozoonid reference species

Species	NCBI Accession
Enterocytozoon hepatopenaei	KX981865.1
Enterocytozoon bieneusi	KF148056.1
Enterospora nucleophila	JX101917.1
Enterospora canceri	HE584634.1
Paranucleospora theridion	FJ594982.1
Desmozoon lepeotherii	HM800847.2
Nucleospora salmonis	AF186006.1
Nucleospora cyclopteri	KC203457.1
Nucleospora braziliensis	KT777455.1

Chapter 4: Biogeography of the shrimp pathogen, Enterocytozoon hepatopenaei: human-driven dispersal or repeated local acquisition?

4.1 Introduction

In the last ten years, thanks to decreasing costs of next-generation sequencing and a greater accessibility to population-based analyses, molecular ecologists have increasingly used population genomics to resolve more applied research problems. The application of next-generation sequencing technologies has revolutionised our understanding of molecular ecology across taxonomic scales, trophic levels, and temporal and geographic habitat gradients (Lefort *et al*., 2017; Richardson *et al*., 2017; Saupe *et al*., 2019). Whereas in the past they have predominantly been used to explain the complexities of speciation (McCormack *et al*., 2012; Wagner *et al*., 2013) and evolutionary processes (Prunier *et al*., 2016; Twyford & Ennos, 2012), they are now routinely used in a more applied fashion. Examples include the use of population genomics in investigating transmission routes, geographic origin, and distribution of parasites (Auburn & Barry, 2017; González-Candelas *et al*., 2018), and geographic patterns of susceptibility to drugs across populations (Heinz *et al*., 2019). The latter is especially important when looking at drug resistance in certain populations (Wilson, 2015).

These techniques are also extensively used in studying the molecular ecology of pathogens, ranging from detection of pathogens in the environment and their contribution to host-based metagenomes (Andersen *et al*., 2017; Andersen & Hoorfar, 2018) to monitor the global rise in antimicrobial resistance and the emergence of novel human pathogens (Desjardins *et al*., 2017; Heinz *et al*., 2019).

Amongst Microsporidia, the Enterocytozoonidae are a particularly important group in terms of its economic impact. However, studies investigating molecular ecology/comparative genomics are severely lacking for species in the Enterocytozoonidae, and for the Microsporidia as a whole. Likely contributing to this dearth is that the study of intracellular parasites comes with a number of difficulties, identification of hosts chief among them. Studies that have used comparative genomics on species from this Phylum have done so to answer

related questions; looking at ploidy and sexual reproduction (Selman *et al*., 2013), interstrain diversity, in relation to diagnostics (Pombert *et al*., 2013), and transmission mode and genome evolution (Haag *et al*., 2019). Despite so few of these studies directly addressing ecology and population structure, they have added knowledge on intraspecific diversity, showing, in the species studied, great diversity in the genomic variation within species (Pombert *et al*., 2013; Selman *et al*., 2013). This type of research is economically important, as it aids in understanding the ecological dynamics and impact a parasite has. However, more work is needed to monitor the movement of particular species, especially when it comes to trying to control and mitigate further infection, as stated in Brown (2017) (M. J. F. Brown, 2017) in regard to Microsporidia in bees.

Enterocytozoon hepatopenaei (Enterocytozoonidae: Microsporidia) has received a lot of attention. Originally detected as a rare and relatively inconsequential infection in farmed penaeid shrimp (Tourtip S, 2005), it became a major production problem in Asian shrimp farm systems subsequent to its formal description in 2009 (Tourtip *et al*., 2009). Although it is also possibly described earlier than that, in Malaysia and Australia (Hudson *et al*., 2001). It infects the hepatopancreatic epithelial cells of two species of farmed penaeid shrimp, *Penaeus monodon* (black tiger shrimp) and *Penaeus vannamei* (white leg shrimp), two major exports for Thailand (Chaijarasphong *et al*., 2020) that contribute $5 billion and $19 billion (respectively) to global annual market sales (FAO, 2016). *P. monodon* used to be more prevalently farmed throughout South East Asia, however, a supposed ‘specific pathogen-free’ (SPF) (*P. vannamei*) broodstock from Hawaii was introduced (Wyban, 2003). It was first introduced to Taiwan and China in the 1990s and then to Thailand, India, and other coast-based South East Asin countries in 2000-1 (Briggs *et al*., 2004). This change was due to a faster-growing time and lower rearing costs of *P. vannamei*, and due to the fact that the SPF status allowed brooding stocks to be imported. *P. monodon* had to be wild caught and grown, meaning stocks could likely facilitate spread of pathogens (Briggs *et al*., 2004; Wyban, 2003). It is partly suspected that it was the change to *P. vannamei* from *P. monodon* that has allowed *E. hepatopenaei* to become so prevalent. *E. hepatopenaei* has been associated with a number of disorders that are affecting these farmed shrimps, chiefly monodon slow growth syndrome (MSGS) (Tourtip *et al*., 2009), but also acute hepatopancreatic

necrosis disease (AHPND) (Aranguren *et al*., 2017) and white faeces syndrome (WFS) (Rajendran *et al*., 2016b). It has mainly been implicated in causing MSGS, which affects the rate at which the infected shrimp matures. As shrimp are sold in batches of the same size, batches of larger shrimp earn more, the presence of *E. hepatopenaei* has affected the potential income of infected harvests. Since its first description in Thailand, *E. hepatopenaei* has subsequently been detected in shrimp farmed in India (Rajendran *et al*., 2016a), China (Y. M. Liu *et al*., 2018), Vietnam (Ha *et al*., 2010; Tang *et al*., 2017), Venezuela (Tang *et al*., 2017), Indonesia (Tang *et al*., 2016a), and Brunei (Tang *et al*., 2015a). However, relatively low sequence similarity for spore wall protein 1 (91%) and B-tubulin (93%) for the Venezuelan isolate suggests pathogens that mimic the pathogenetic signs of *E. hepatopenaei* (but are taxonomically distinct parasites) may be present in shrimp from certain regions (Chaijarasphong *et al*., 2020). Despite its distribution across Asia, it is not known how *E. hepatopenaei* has become so widespread, and though it has been suggested to be endemic to the Australasian region (Salachan *et al*., 2017b), no definitive studies have been carried out to test this (Chaijarasphong *et al*., 2020).

Intraspecific transmission of *E. hepatopenaei* between farmed shrimp can be both vertical (Vu-Khac *et al*., 2018) and horizontal (Karthikeyan & Sudhakaran, 2019a; Salachan *et al*., 2017a); the latter via spores shed through faecal matter to the water column, and through cannibalism (Tang *et al*., 2016b; Tangprasittipap *et al*., 2013b). However, the route through which *E. hepatopenaei* is being (interspecifically) transmitted into shrimp ponds is not known. There are a number of viable routes, as many shrimp ponds are accessible to environmental/anthropogenic processes. Research looking at possible vectors has shown that lab-infected *Artemia salina* has been able to transmit the enterocytozoonid to *P. vannamei* (Karthikeyan & Sudhakaran, 2019a). While work looking at possible pond reservoirs of *E. hepatopenaei* (in sediment) found through PCR and histology, that two species of polychaete worm were infected (Desrina *et al*., 2020). Although no route of transmission has been verified, some research suggests a possible endemic Australasian distribution for *E. hepatopenaei* in (Salachan *et al*., 2017b), based on the parasite's widespread distribution and older studies describing similar histopathology in *Penaeus japonicus* from Australia (Hudson *et al*., 2001).

Control of *E. hepatopenaei* (and other important pathogen hazards) in shrimp farming nations is based upon the application of appropriate biosecurity protocols that limit the likelihood of contact between the hazard and susceptible hosts (E. Peeler, 2005; E. J. Peeler *et al*., 2015). Furthermore, for the efficient application of biosecurity protocols, an appropriate diagnostic strategy must be in place to ensure that the hazard under surveillance can be accurately detected and discriminated from other pathogens that may be present within the system (G. D. Stentiford *et al*., 2014). The application of improved tools for pathogen detection (and discrimination), based on whole-genome analyses, has the potential to revolutionise our understanding of pathogen transmission (and translocation) in high-trade industries such as aquaculture. The resolution provided by these tools may discriminate between strains of the same pathogen species that may exist in different geographic regions (Cissé *et al*., 2018) or show that the pathology associated with a certain disease state may be caused by taxonomically different pathogenic agents (G. D. Stentiford *et al*., 2018). When applied to transmission/translocation questions such as for *E. hepatopenaei* in shrimp, more accurate assessments can be made on whether disease outbreaks occurring in one location are related to pathogen incursion from another region (e.g. via trade in live animals or products) or alternatively, due to exposure of susceptible hosts (in this case, shrimp) to local infected hosts or reservoirs. Furthermore, by improving diagnostic resolution, the investigation can be made into whether the disease state is described as '*E. hepatopenaei*' may in fact be due to different pathogens that cause the same pathological outcome. Such knowledge has fundamental implications for the subsequent application of biosecurity strategies for controlling disease on farms.

Since its description, *E. hepatopenaei* has become one of the most well-researched enterocytozoonids, joining the more high-profile microsporidians infecting important organisms. However, publications have largely focused on new more sensitive, less invasive detection methods (Cruz-Flores *et al*., 2019), associated conditions (as mentioned above), descriptions in new countries (as mentioned above), and studies looking at possible vectors (as mentioned above). None have looked at the population structure across different countries for this species, to give a broader understanding of geographic origin.

The aim of this study is to, using populations genomics, identify the possible routes between three countries in Asia (China, India and Thailand) through which *E. hepatopenaei* is being dispersed. In so doing give a broader understanding of its geographic origin to ascertain its likely route of transmission into shrimp farms. Analysis of transmission routes will also enable the testing of three hypotheses: 1. *E. hepatopenaei* is endemic in Asia but now visible due to a change in farming practice or a more pathogenic variant; 2. Translocation of *E. hepatopenaei* between shrimp farming nations has facilitated distribution and establishment from a single original point of origin, or 3. The disease associated with '*E. hepatopenaei'* is not of single-agent aetiology – with numerous enterocytozoonids able to elicit disease in susceptible hosts. Defining the likelihood of these scenarios will aid farmers and managers to apply appropriate biosecurity practices to minimise the impact of *E. hepatopenaei* in shrimp aquaculture.

4.2 Methods and analysis:

4.2.1 *Sample collection:*

Samples were collected from Thailand in 2018. They were collected from different *P. vannamei* farms from Chantaburi, Samut Sakhon, Suphan Buri, and Pathum Thani (Table 4.1). A single bucket of shrimp was drawn from each pond and 30 shrimp were taken from each bucket for spore extraction. Further samples of hepatopancreas and gut (stomach and intestine) tissues were dissected from *P. vannamei* collected from farm sites in India (Table 4.1) in 2016 and 2017; these samples were processed whole for metagenomic analysis.

Table 4.1: Isolates used in the final analysis, collected from Thailand, China, and India (*indicates isolates sequenced for this study, +indicates reads from the reference genome, and ^indicates reads from metagenomic studies)

Isolate ID	Region, Country	Spore concentration	Total reads	Reads mapped	Unmapped reads	Average Coverage % /depth	SNPs pre-calibration	SNPs post-calibration
***Thai-Chan1-1**	Chantaburi, Thailand	$1.95*10^4$	75836	59,462	16, 374	96.79/4.18	118535	118872

***Thai-Chan1-2**	Chantaburi, Thailand	$7.2*10^6$	235268	232,792	2, 476	99.84/17.27	418886	424296
***Thai-Chan2-1**	Chantaburi, Thailand	$1.28*10^6$	220871	217,608	3, 263	99.74/16.20	353820	356415
***Thai-Pathun**	Pathum Thani, Thailand	$3.88*10^5$	170566	151,383	19, 184	99.74/11.29	302062	304561
***Thai-Suphan 1-1**	Suphan Buri, Thailand	$2.69*10^6$	231500	211,573	19, 927	98.63/12.89	156769	158289
***Thai-Suphan 1-2**	Suphan Buri, Thailand	$3.52*10^6$	493900	471,081	22, 819	99.48/29.92	185778	187419
+Thai-Chao	Chachoengsao, Thailand	(Wiredu Boakye *et al*., 2017)	7,218,583	7,209,872	8, 711	99.98/442.47	34820	34885
^Indi-Saf1	Safale, India	N/A	348, 232	335292	12, 940	83.29/4.27	12019	11807
^Indi-Saf2	Safale, India	N/A	242, 994	231793	11, 201	88.91/3.93	76378	75030
^Indi-Saf3	Safale, India	N/A	402, 372	387784	14, 998	98.67/8.85	214038	212239
^Indi-Dah	Dahanu, India	N/A	346, 876	329043	17, 833	87.64/4,34	8299	8069
^Chin-Qing1	Qingdao, China	N/A	57,803,233	185330	57,617,903	83.94/1.81	85144	87136
^Chin-Qing2	Qingdao, China	N/A	58, 966, 483	1, 289, 365	57, 677, 118	98.68/38.54	389877	413378
^Chin-Qing3	Qingdao, China	N/A	62, 763, 693	317, 061	62, 446, 632	93.37/5.10	196175	202923
^Chin-Qing4	Qingdao, China	N/A	55, 738, 161	251, 139	55, 487, 022	88.74/2.88	132385	136687

4.2.2 *Purification of spores:*

For samples collected in Thailand, hepatopancreai were dissected from each set of 30 collected whole shrimp and pooled for each separate pond, and stored in ice for later processing. The pooled hepatopancreai were homogenised with a sterile glass pestle in a 15 ml falcon tube in 1 x PBS. The homogenised samples were transferred to a 50 ml falcon tube and topped up to 50 ml with 1 x PBS. These were then filtered, first through a sterile 100 µm cell strainer (Fisherbrand). The resulting filtrate was filtered a second time, with a 40 µm cell strainer to remove the remaining tissue debris. The spores were pelleted by centrifugation at 10,000 x g for 10 mins. The pelleted spores were resuspended in 500 µl of 1 x PBS. The pelleted spores were further purified through Percoll density gradient:

An 8ml discontinuous Percoll concentration (25%, 50%, 75%, 100%) was made to concentration with 1 x PBS. A volume of 1.3 ml of the resuspended, filtered spores was pipetted onto the Percoll solution. These were separated through centrifugation, using an ultracentrifuge at 40,000 x g for 30 mins at 15°C. The spore band was removed with a syringe and washed 3 times with 1 x PBS and spun down at 14,000 x g for 30 mins at 4°C. Purified spores were resuspended in a 1ml solution of 1 x PBS, 10% glycerol, 100µg/ml of ampicillin and kanamycin, and stored at -80^{o}c.

4.2.3 *Cell lysis and gDNA extraction for spores from Thailand:*

Samples were washed three times with sterile water (spun down at 10, 000 x g for 10 mins and supernatant discarded) and resuspended in 400 µl of 10/1 TE buffer. The resuspension was transferred to a 2 ml screw-cap tube along with 10µl of 10% SDS, 2ul of 10 mg/ml RNAse A and 400µl of (0.17 - 0.18 mm) glass beads. The spore solution was initially incubated on a heat block for 2 minutes at 55°C and then bead beaten for 1 minute at 2000 rpm. The solution was then incubated for a further minute at 55°C and transferred back to the bead beater for 1 minute at 2000 rpm. This was repeated three times. A volume of 800µl of pH 8.0 phenol was added to the bead beaten cells and these were mixed by inversion (x5) and centrifuged at 10000 x g for 10 minutes. The top aqueous layer was transferred to a fresh Eppendorf and 400 µl of chloroform was added. This was mixed by inversion (x5), and the mixture was centrifuged at 10000 x g for 10 minutes and the top aqueous layer was transferred to a new 1.5 ml Eppendorf. The sample was then cleaned using a standard EtOH precipitation protocol (Chapter 3). DNA from Indian samples was extracted from homogenised hepatopancreas and gut samples and purified using a CTAB/EDTA DNA extraction protocol (Fontes *et al*., 2017).

4.2.4 *Sequencing and quality control:*

The genomic DNA was quantified using Qubit, dsDNA broad range (BR), which uses a fluorescent tag. Those that were of high enough concentration were run on a tape station, to ensure the DNA is not too degraded. Samples of high enough quality were sent off for library preparation, using Nextera XT and sequencing on the Illumina MiSeq platform, 300bp paired reads. The raw reads were sent through a pipeline to quality check the reads, removing/trimming low-quality

reads (reads with Phred score lower than 20/reads under a certain length) and adaptor sequences, using the default parameters on Sickle version 1.33 (Joshi & Fass, 2011). Samples from India were prepared usingNextSeq Series Mid-Output kit (Illumina) sequence libraries and were sequenced on an Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA; 2x150bp). Low-quality bases and remaining adapter sequences were removed using fastp version 0.20.0 (S. Chen *et al*., 2018) (using the poly-G tail trimming and default parameters).

4.2.5 *Metagenomic screening:*

To help give a better picture of international population structure, further genome data for *E. hepatopenaei* was extracted from the publicly available online database, National Centre for Biotechnology Information (NCBI) SRA database (Leinonen *et al*., 2011; National Center for Biotechnology Information, 2020), and Integrated Microbial Genomes and Microbiomes (IMG) (I. M. A. Chen *et al*., 2019) database. Three genetic regions were initially used to screen these databases, 18s rDNA (KU179095.1), Spore Wall Protein1 (KX258197.1), and Beta-tubulin (KX258197.1). Hepatopancreatic metagenomes were selected on NCBI, using the search parameters “Penaeus vannamei AND hepatopancreas”. Any positive SRA datasets were downloaded using the SRA toolkit program (Leinonen *et al*., 2011), and then quality checked and filtered using Sickle 1.33, default parameters. IMG was screened in a similar fashion, screening datasets that were from aquatic samples, both marine, and freshwater, any positive results were treated as below.

4.2.6 *Mapping reads:*

Trimmed reads were aligned to the *E. hepatopenaei* reference genome (GCA_002081675.1)using the Burrows-Wheeler Aligner version 0.7.17 (BWA-MEM) (H. Li & Durbin, 2010), under default settings. BWA-MEM was used as it was appropriate for the length of the reads used in this study.

4.2.7 *Assessment of “multiple strains” of E. hepatopenaei using Metaspades+CONCOCYT+Busco:*

As the spores that were collected for this study were collected from multiple individual shrimp there is a possibility that rather than being clonal and the result

of a single strain per pond and shrimp, multiple “strains” of *E. hepatopenaei* may be infecting a single shrimp pond or even single shrimp. To assess whether multiple or single strains were present, metagenomic contigs were assembled from each pond using MetaSPAdes version 3.11.1 (Nurk *et al*., 2017) under default parameters, designed to detect and assemble related strains of single cells. Contigs from this assembly were cut into segments 10,000 bp long and clustered into taxonomic units using the binning program CONCOCT version 1.1.0 (Alneberg *et al*., 2014). The resultant number of clustered contigs (bins) were used as an estimate of the number of different taxonomic units in the sample. To test the validity of the bins (implicitly assuming that “completeness” infers valid bins), Busco version 3 (Seppey *et al*., 2019) was used to test the “completeness” of the bins as a means of checking the validity of the bins being taken as “strains”. As the analysis uses a database of known conserved single-copy genes to assess genome completeness, both the eukaryotic and fungal datasets were used.

4.2.8 *Assessing ploidy level:*

The ploidy levels of each *E. hepatopenaei* sample were estimated using k-mer frequencies. Smudgeplot version 0.2.2 (Ranallo-Benavidez *et al*., 2020) was used to build a bar chart based on heterogenous k-mer pairs. K-mer counts were carried out using KMC version 3 (Kokot *et al*., 2017).

4.2.9 *SNP calling and initial analysis:*

Once aligned, variant calling was performed on a dataset of all isolates using Genome Analysis Toolkit version 4.0.2.1 (GATK) (McKenna *et al*., 2010), following their best practices pipeline. To ensure that all likely genuine SNPs were identified, sequencing quality assessment was recalibrated using GATK using initial SNP calling as a reference. SNPs were counted using Samtools-BCFtools version 0.1.19 (H. Li *et al*., 2009) to see if there was an increase in the SNPs called per recalibration. This was done until recalibration stopped yielding increased SNP identification. Lastly, SNPs across all samples were filtered by missing genotype (20%), minor allele count (3), and quality Phred score of 20 using VCFtools version 0.1.13 (Danecek *et al*., 2011) to reduce the effect of false positives.

4.2.10 *Population history and selection patterns:*

As the samples in this study are essentially pooled, investigating the intrapopulation history and possible demographic events using conventional analysis would be inadequate. This was investigated using Popoolation 2 version 1.2 (Kofler *et al*., 2011) to carry out Tajima's D test of neutral evolution. Popoolation 2 was used, as it allowed for analysis of pooled populations. Tajima's D and was carried out with a sliding window of 1000, step size of 500, minimum count of 2, minimum coverage of 4 (coverage has to be double of minimum count), and a variable pool size depending on the isolate. A negative genome-wide Tajima's D is indicative of an expansion after a bottleneck, whereas a positive D suggests a decrease in population size.

4.2.11 *Population structure:*

To look at population structure for the *E. hepatopenaei* isolates, principal components (PCA) analysis was carried out using the R package adegenet (glPCA) version 2.1.3 (Jombart, 2008) and a phylogeographic tree based on SNPs was built using Bayesian inference implemented in BEAST2 package version 2.6.2 (Bouckaert *et al*., 2019), SNAPP version 1.5.1 (Bryant *et al*., 2012). Coalescence and mutation rates (U and V) were set to 1, with a chain length (MCMC) of 50,000 (10% burn-in required for a plateau in posterior). Tracer version 1.7.1 (Rambaut *et al*., 2018) was used to view and analyse MCMC trace files. Analysis looking at genetic admixture was calculated using the three population test (f3), implemented in admixr version 0.9.1 (Petr *et al*., 2019). F3 stats assess the admixture between three populations; one target population and two source populations. Fixation indices (Fst) and nucleotide diversity (π) estimates, per isolate, were used to investigate interpopulation and intrapopulation dynamics, respectively. Both were also implemented using Popoolation 2 version 1.

4.2.12 *Assessing intrapopulation diversity using single-copy genes analysis:*

Given the nature of sample collection, in that it is made up of millions of individuals collected from multiple host individuals, it is possible that the microsporidia in each sample could be either clonal or the result of multiple infecting spores. In either case, estimates of intrapopulation heterozygosity would

be unreliable, as multi-copy genes would add false heterozygosity. To account for this possibility heterozygosity was assessed using single-copy genes within each sample. For this, single-copy genes were identified using a collection of all available microsporidian genomes. Orthogroups were constructed using OrthoMCL version 2.0 (L. Li *et al*., 2003) and filtered for single-copy genes (refer to Chapter 5 for full method). Using the same protocol as above for genome-wide SNP calling reads from all isolates were then mapped to the single gene copy reference using BWA, making a single gene copy dataset, and SNPs were called using GATK. Missing genotypes were filtered at 20%, using VCFtools version 0.1.13. Both interpopulation and intrapopulation divergence estimates diversity were assessed as above, using VCFtools version 0.1.13, using the het parameter to assess heterozygosity (He/Ho) on a per-sample basis, and Weir and Cockerham estimates of Fst.

4.3 Results

4.3.1 *Dataset description:*

Sequencing and bioinformatic efforts resulted in 15 datasets being taken forward for analysis (Table 4.1). Four each from China and India and seven from Thailand. Depth of coverage and percentage coverage ranged from 1.81-442.47 and 83.29%-99.98% respectively (Table 4.1). The Indian samples were among the ones that had the least coverage (Indi-Saf1), and Chinese with the least depth (Chin-Qing1). The distribution of reads mapped to contigs followed the same trend over most samples (Figure 4.1-supplementary). With the largest contigs having the most mapped reads, following a linear relationship. Two Thai samples from Suphan Buri (Thai-Suphan1-1/1-2) and all Indian samples varied from this at several datapoints, with two outliers with a greater number of mapped reads than would be expected following the trend (Figure 4.1B and C-supplementary).

4.3.2 *SNP calling and filtering:*

There was a linear relationship between the number of SNPs called and the size of the contig (Figure 4.2a-supplementary), with a similar correlation between the number of reads mapped and the number of SNPs called pre-base-recalibration (Figure 4.2b-supplementary). Showing no bias in the data, SNPs were called equally, on average, across the genome. Only three iterative base-recalibrations

were performed before the SNP calling no longer improved (Table 4.2). Filtration for a minimum allele count of 3 led to a reduction of more than 1000 SNPs, from 1680 shared SNPs to 598 shared SNPs. Leaving 598 SNPs for further analysis.

Table 4.2: SNP counts for before and after recalibration and filtering of missing genotypes for all datasets used in this study.

Dataset	SNPs pre-calibration	SNPs post-recalibration	SNPs 30% missing genotype filtered	SNPs 20% missing genotype filtered	SNPs 10% missing genotype filtered	SNPs 0% missing genotype filtered
All combined dataset	10768	10395	2826	1680	168	20
Thai dataset	6345	6340	5263	4233	2867	2867
Indian dataset	4933	4583	624	79	79	79
Chinese dataset	3554	3553	1379	604	604	604
Single copy dataset	340	328	185	133	15	1

4.3.3 *Metaspades test for multiple strains in samples:*

Looking at the possibility of multiple strains in sample sets collected in this study, a combination of Metaspades, CONCOCT, and BUSCO were used to assess this. The number of contigs produced by Metaspades for each dataset ranged from 35 (Chin-Qing1) to 865 (Indi-Saf3) (Table 4.3). These were clustered into bins by CONCOCT for each isolate, based on operational taxonomic units (OTUs). The number of bins (each bin representing an OTU) per isolate, produced by CONCOCT ranged from 4 (Chin-Qing1) to 79 (Thai-Chao), with most bins containing a single sequence (Table 4.3).

Table 4.3: MetaSPAdes, concoct and BUSCO results per isolate (*indicates isolates sequenced for this study, +indicates reads from the reference genome, and ^indicates reads from metagenomic studies

Samples	MetaSPAdes contigs	Concoct bins	Number of sequences in 5 biggest bins	Number of bins with complete BUSCO	Number of complete BUSCO	Number of complete BUSCO

				genes (Eukaryotic/Fungal)	genes-eukaryotic	genes-fungal
***Thai-Chan1-1**	710	5	695/9/3/2/1	1/1	14	25
Thai-Chan1-2	310	10	143/115/23/14/2	1/2	18	64/2
***Thai-Chan2-1**	261	46	108/76/12/9/3	1/2	19	76/1
***Thai-Pathun**	280	23	131/111/11/2/2	1/2	18	71/2
***Thai-Suphan1-1**	773	7	640/88/41/1/1	1/1	6	20
***Thai-Suphan1-2**	803	12	728/42/18/4/3	1/1	11	61
⁺Thai-Chao	256	79	150/11/4/2/2	1/1	18	73
^Indi-Saf1	-	N/A	N/A	N/A	N/A	N/A
^Indi-Saf2	-	N/A	N/A	N/A	N/A	N/A
^Indi-Saf3	865	11	576/266/9/5/3	1/2	7	6/4
^Indi-Dah	-	N/A	N/A	N/A	N/A	N/A
^Chin-Qing1	35	4	32/1/1/1	1/0	1	N/A
^Chin-Qing2	638	8	344/288/1/1/1	0/2	N/A	3/1
^Chin-Qing3	192	11	181/2/1/1/1	0/0	N/A	N/A
^Chin-Qing4	75	9	62/6/1/1/1	1/0	1	N/A
All	250	23	104/101/17/2/2	1/3	18	77/2/2

Using the eukaryotic database for Busco, no isolate had more than one bin that had complete copies of single-copy genes (Table 4.3). Of the 303 single-copy genes looked for in the eukaryotic database, only a maximum of 19 (Thai-Chan2) were found among the Thai samples, the others ranged from 6-18. There was a greater representation of the single-copy genes when using the fungal database, the number of complete genes ranged from 20-73, with 5 of 7 Thai samples having more than 60 (Table 4.3). Chinese and Indian samples had fewer complete BUSCO single-copy genes. The fungal database also showed more bins per sample that had complete copies of the single-copy genes. 5 out of 12 samples had 2 bins with complete copies of single-copy genes from Busco's database, however, the second bins have fewer copies, ranging from 1-4. As there was no evidence for multiple strains, having multiple bins (OTUs) with comparable numbers of complete reference genes, it has been assumed that the

isolates in this study are only composed of one *E. hepatopenaei* strain. Of the Indian isolates, only Indi-Saf3 passed the coverage threshold for Metaspades and was the only one carried forward with the rest of the multi-strain analysis.

4.3.4 *Estimates of ploidy levels:*

Ploidy estimates suggested that most (12 of 15) isolates were diploid, although two may be triploid (Thai-Chan2-1 and Indi-Saf3) and ploidy could not be confidently determined for several others (Table 4.4). As the majority of the isolates were diploid, for analyses that required a ploidy level, diploid was selected.

Table 4.4: Ploidy estimates per isolate (*indicates isolates sequenced for this study, +indicates reads from the reference genome, and ^indicates reads from metagenomic studies)

Isolate	**Portion of kmer pairs /Dominant ploidy**	**Portion of kmer pairs /Secondary ploidy**	**Kmer pairs 1n**
***Thai-Chan1-1**	0.39/AB	0.31/AAB	31
***Thai-Chan1-2**	0.37/AB	0.24/AABB	32
***Thai-Chan2-1**	0.51/AAB	0.2/AB	30
***Thai-Pathun**	0.45/AB	0.26/AAABB	28
***Thai-Suphan1-1**	0.45/AB	0.16/AAB	36
***Thai-Suphan1-2**	0.83/AB	0.08/AAABB	36
+Thai-Chao	0.79/AB	0.08/AAABBB	176
^Indi-Saf1	0.91/AB	0.06/AAAABB	40
^Indi-Saf2	0.96/AABB	0.04/AAABB	50
^Indi-Saf3	0.37/AAB	0.25/AB	43
^Indi-Dah	0.34/AB	0.22/AAB	37
^Chin-Qing1	0.48/AB	0.18/AAB	56
^Chin-Qing2	0.84/AB	0.07/AAABBB	59
^Chin-Qing3	0.54/AB	0.21/AABB	51
^Chin-Qing4	0.72/AB	0.14/AAAB	55

4.3.5 *Endemic population structure:*

PCA and phylogeographic analysis (Figures 4.1a, b) for the dataset encompassing all isolates showed some clustering based upon geographic location. There was strong support for the tree, posterior probabilities ranged from 0.7-1 (average 0.9). An isolate from both China and India grouped more

closely with Thai isolates, Chin-Qin2 forming a group with Thai-Chao and Indi-Saf3 more distantly forming a group with the Thai isolates. A combination of both PC1 (explaining 36% of the variance observed) and PC2 (explain 16% of the variance observed) separated the three countries. PC1 alone only separated Thailand from China, and only partially from India. Where PC2 separated Thailand and India, but cluster with the Chinese isolates. Thai isolates were however clustered much more densely and showed less divergence than either the Indian or the Chinese isolates.

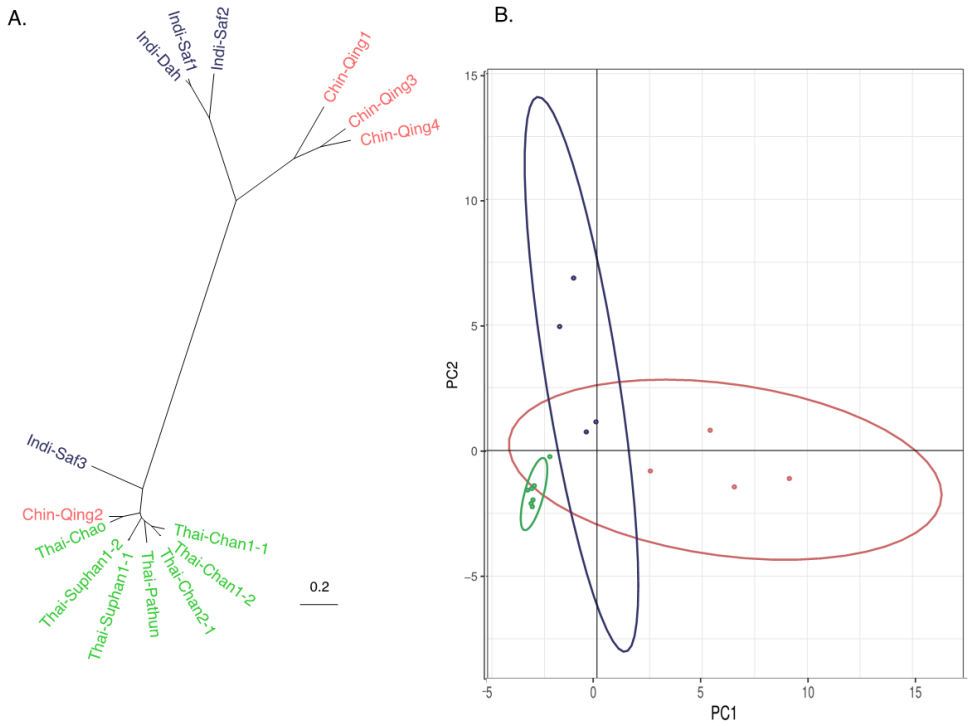


Figure 4.1: Patterns of geographic structure: A) Bayesian tree built in BEAST2 using package SNAPP, B) principal components analysis (glPCA) both analyses carried out with 598 SNPs. *China (in red), Thailand (in green), and India (in blue).

The Thai only dataset showed some clustering by provinces, both in the tree (Figure 4.2a) and the PCA (Figure 4.2b). PC1 (explaining 41% of the variance observed) separated by region, with Thai-Chao and Thai-Pathun as intermediates. However, Thai-Chan 1 (EHP18) did not cluster most closely with

the other samples from Chantaburi (Thai-Chan1 (38), Thai-Chan2 (29)). Estimates of interpopulation genetic divergence showed a similar trend, with F_{st} ranging from 0.138-0.223 (average 0.183) (Table 4.5). The greatest divergence was between Thailand and India (0.223) the least between China and Thailand (0.138).

Table 4.5: Matrix of Weir and Cockerham's mean interpopulation estimates of Fst - by country, estimated from 598 genome-wide SNPs

	Thai	**India**
Thai		
India	0.223	
China	0.138	0.2187

Estimates within countries (Table 4.6) between regions showed a greater range of divergence for Thailand, in between regions ranging from 0.025-0.116 (average 0.078), (not applicable with Chinese samples, from one region) than between Indian regions (0.057). Divergence from Thai regions mirror results from the PCA and tree, showing Suphan was the most divergent among the regions sampled. Estimates from regions in different countries were more pronounced, as would be expected. Ranging from 0.09-0.507 (average 0.147), the Indian region, Dahanu, showing the greatest divergencies ranging from 0.204- 0.507.

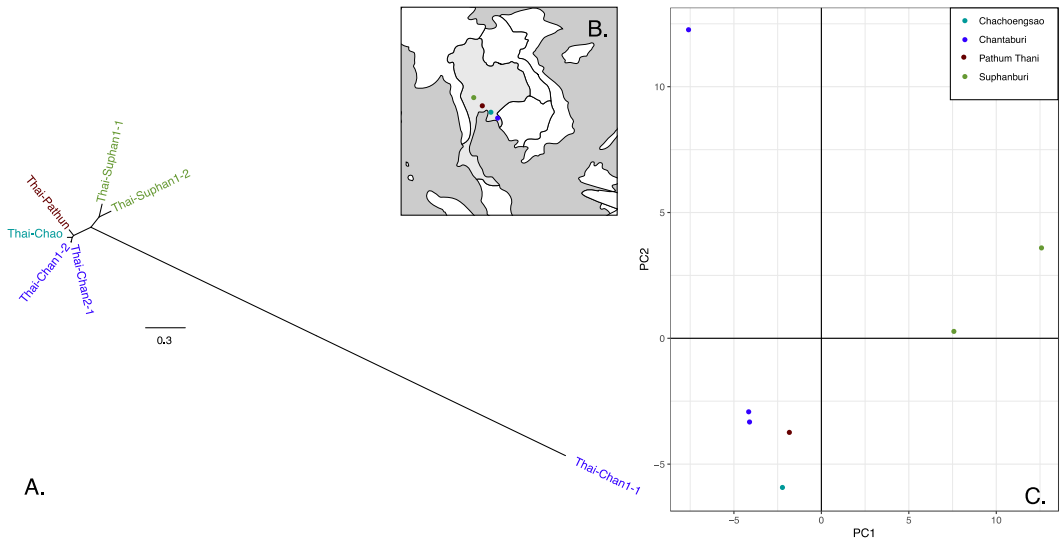


Figure 4.2: Patterns of Thai geographic structure: A) Bayesian tree built in BEAST using package SNAPP, B) principal components analysis (glPCA) both analyses carried out with 4233 SNPs.* Chachoengsao (in light blue), Pathum Thani (in brown), Suphan Buri (green), and Chantaburi (in dark blue)

Table 4.6: Weir and Cockerham's mean interpopulation estimates of Fst - by region estimated from 598 genome-wide SNPs

	Thai-Chanta	**Thai-Pathun**	**Thai-Chao**	**Thai-Suphan**	**Chin-Qing**	**Indi-Safale**	**Indi-Dahanu**
Thai-Chanta							
Thai-Pathun	0.025						
Thai-Chao	0.035	0.019					
Thai-Suphan	0.116	0.103	0.114				
Chin-Qing	0.170	0.085	0.039	0.201			
Inidi-Safale	0.159	0.009	0.031	0.198	0.227		
Indi-Dahanu	0.348	0.204	0.197	0.507	0.234	0.057	

The sample sites Thai-Pathun and Thai-Chacho have the lowest divergence estimates for any Thai isolate against a non-Thai isolate ranging from (0.009).

4.3.6 *Population history: E. hepatopenaei in shrimp ponds*

). Across all 15 isolates, nucleotide diversity (π) was moderate to high for the dataset encompassing the whole genome (ranging from 31.3% (Indi-Saf1) - 42.5% (Thai-Pathun)) (Table 4.7). The samples from India did not have enough coverage to be tested for Tajima's D. The average results of the analysis from the remaining isolates, genome-wide, ranged from --0.002 (Thai-Chan2-1) to 0.131 (Thai-Chao) (average 0.004) (Table 4.7). Of the eleven isolates, eight had positive Tajima's D, with both countries having positive and negative Tajima's D. Though for all samples, Tajima's D was fairly neutral.

Table 4.7: Intrapopulation indices per isolate, estimated from 598 genome-wide SNPs isolate ("Sc" indicates analysis on 126 single-copy genes, *indicates isolates sequenced for this study, +indicates reads from the reference genome and ^indicates reads from metagenomic studies)

Isolate	***π***	***π* (sc)**	**Tajima's D**
***Thai-Chan1-1**	0.406	0.187	0.000
***Thai-Chan1-2**	0.408	0.195	0.002
***Thai-Chan2-1**	0.401	0.164	-0.002
***Thai-Pathun**	0.425	0.188	-0.001
***Thai-Suphan1-1**	0.414	0.191	0.008
***Thai-Suphan1-2**	0.414	0.199	0.025
+Thai-Chao	0.398	0.181	0.131
^Indi-Saf1	0.313	0.126	N/A
^Indi-Saf2	0.356	0.190	N/A
^Indi-Saf3	0.367	0.164	N/A
^Indi-Dah	0.367	0.155	N/A
^Chin-Qing1	0.409	0.217	0.002
^Chin-Qing2	0.410	0.173	0.006
^Chin-Qing3	0.413	0.157	0.003
^Chin-Qing4	0.378	0.216	-0.001

4.3.7 *Population admixture-F3 statistics:*

F3 statistics were applied to populations on a large geographic scale (by country), failed to show any admixture between the populations tested in this study (Table 4.8), as all values were positive. F3 values ranged from 0.070 (Thailand)-0.170 (India) (standard deviation range 0.004-0.021).

Table 4.8: Three population admixture analysis (F3 statistics), implemented in admixr

Source populations	Target populations		
	Thailand	India	China
Thailand	India		0.084 (+-0.0215)
China	Thailand	0.167 (+-0.0212)	
India	China	0.070 (+-0.004)	

4.3.8 *Single copy gene analysis:*

The single-copy gene dataset still showed geographic divergence shown by Weir and Cockerham Fst (Table 4.9) following the same trend as in the dataset incorporating all SNPs. The greatest divergence was found between Thai and Indian isolates, and the least between Thai and Chinese Isolates (ranging from 0.035-0.126 (average 0.07)).

Table 4. 9: Weir and Cockerham's mean interpopulation estimates of F_{st} – for single-copy genes by country, estimated from 133 single-copy gene SNPs

	Thai	**India**
India	0.126	
China	0.035	0.040

However, the single-copy gene dataset showed less divergence than the Weir and Cockerham analysis on the genome-wide 598 SNPs. Similar to the Weir and Cockerham Fst estimates, the nucleotide diversity (π) observed in the single-copy gene dataset was lower than that of the genome-wide dataset (ranging from 12.6% (Indi-Saf1) - 21.7% (Chin-Qing1) (average 18%)).

4.4 Discussion

Little is known about the ecology and transmission routes of *E. hepatopenaei,* as studying the ecology of microorganisms is difficult, and in parasitic microorganisms, further still. Here, the goal was to improve the knowledge of the

routes through which *E. hepatopenaei* may be being transmitted into shrimp farms and its biogeographical origin. With three main hypotheses (endemic, anthropogenically dispersed or different causative enterocytozoonid species) explored. Thailand, India, and China were focused on as these were among the first countries where *E. hepatopenaei* was detected also major shrimp producers, with the majority of the subsequent papers coming from these three countries.

4.4.1 *Biogeography of E. hepatopenaei*

The SNPs obtained in this study, through both a phylogenomic tree, principal components analysis, and population differentiation, predominantly showed separated lineages based on geography (Figure 4.1a, b and Table 4.5). Suggesting populations mostly restricted to local transmission. It is likely that dispersal was also infrequent enough to show within-country spatial structure (Figure 4.2a, b and Table 4.6). This suggests that *E. hepatopenaei* could be endemic to countries observed in this study and already abundant and widespread across South East Asia. It is unlikely that this is by anthropogenic means, as even though Microsporidia may be fast-evolving (Cuomo *et al*., 2012), the process by which they were initially spread would likely be ongoing, so less geographic-based population structure would be evident.

A slightly lower fixation index (F_{st}=0.138) between isolates from China and Thailand suggests slightly greater levels of population connectivity between these countries (Table 4.5). This is also shown in the unrooted tree (Figure 4.1a), with the grouping formed between Chin-Qin2 and Thai-Chao and the lower population differentiation with Indian and Chinese isolates (F_{st}=0.039, 0.085, 0.170 and0.201). Levels of differentiation are similar to and lower than between some Thai isolates (Thai-Suphan x Thai-Chao (F_{st}=0.116). Suggesting that some regions between China and Thailand have similar or greater levels of gene flow with regions within Thailand. Indi-Saf3 also grouped with Thai samples (Figure 4.1a), also suggesting some population connectivity between India and Thailand. Whether this dispersal is through natural means of dispersal, via migratory birds, or anthropogenic means, like trade, is unknown. This suggests that there may have been movement from Thailand to China and India, at least from an area close to the Chachoengsao (Thai-Chao) region to the Qingdao region of China. However, Chachoengsao (Thai-Chao) used to be one of the biggest hatcheries

in Thailand in the 1970s (Patmasiriwat *et al*., 1998). So, it is possible that, in this case, the dispersal of *E. hepatopenaei* was facilitated through anthropogenic means. This has been suggested, as, if it was dispersed through a natural medium, the means through which they were dispersing (either through a vector or zoochory) would likely have occurred for a longer time than more recent anthropogenic-based dispersal. As such, non-native genotypes would be more likely to be found in native isolates. With the Thai-Chao isolate being the closest related (in terms of fixation indices and being grouped on the tree) Thai sample to the Indian and Chinese isolates, it is possible that the widespread dispersal of this regions shrimp may have facilitated the spread of *E. hepatopenaei* to some regions in India and China.

Thai-Chanta's consistent divergence from non-Thai isolates suggests that the farms sampled from Chantaburi may be the most secluded, receiving the least amount of gene flow. Similarly, Thai-Suphan, has equally high levels of differentiation with Thai samples, with a more pronounced differentiation from Indian samples (F_{st}=0.507). A naturally occurring vector for *E. hepatopenaei* has yet to be discovered, so it is not yet known how it disperses between countries. As this species has been found to infect *A. salina* under lab conditions and polychaete worms in shrimp ponds (Desrina *et al*., 2020; Tang *et al*., 2015b), it may have many aquatic vectors. So, it could possibly be through migratory birds that feed on these vectors that have helped in the initial spread of *E. hepatopenaei* and continued the spread in some cases. In the case of within region transmission, it is likely that isolates are the results of local distribution/transmission, as, if spores were from a region reservoir, the structure would reflect this and would not cluster regionally.

4.4.2 *Intrapopulation diversity and population bottlenecks*

It seems feasible that with a moderate level of genomic diversity (evidenced with the estimates of nucleotide diversity (π) found), that isolates analysed in this study are either likely receiving continual gene flow from wild reservoirs or a particularly diverse set of spores initially entered the ponds. Both inferences are also backed up by fairly neutral Tajima's D values for ten out of eleven isolates, which also suggest a 'stable' population, with mutations largely explained by neutral genetic drift. This would make sense, as the ponds sampled in this study

were ‘open’ (not indoors), and would allow for frequent opportunities for transmission of *E. hepatopenaei* from the wild into ponds. Alternatively, if there is no ongoing gene flow from outside shrimp ponds, the populations may have settled long enough for stabilising selection to have taken effect. Though the latter seems less likely, given that the shrimp are harvested 2-6 times yearly (Crespi & New, 2009). In the former scenario, these results also suggest that this may be a true representation of the genotypes found within the surrounding wild population. Whatever the source of the gene flow in shrimp ponds sampled in this study, the levels of nucleotide diversity and neutral Tajima’s D suggest it may be universal among all isolates. Likely meaning that the spores were introduced to ponds in a similar fashion, with similar amounts of geneflow (if any) from the ‘wild’ population. However, there may also be build up of genomic reservoirs of *E. hepatopenaei* in shrimp ponds, as many of the isolates obtained in this study were from earthen ponds, which would allow spores to build up in the sediment over time. Although Thai farming practices do generally incorporate sediment treatment between harvests (Yuvanatemiya *et al*., 2011), these are unlikely to be 100% effective as sediment has been shown to partially protect some microorganisms from disinfection methods (Rokunuzzaman *et al*., 2016; Y. Wang *et al*., 2021). The genetic effects on these isolates may be quite transient, as incoming geneflow from the surrounding environment may be ever-present, like many natural systems. However, it is less likely that gene flow from these ponds is reciprocated as frequently, so pond-based sequences may become more and more divergent from ‘wild’ populations. This is because, as mentioned above, shrimp are harvested 2-6 times a year, meaning that there would not be a long-standing population. This would reduce the chance for farm-based spores to be transmitted back into the ‘wild’ host species, and thus reduce the chance of farm-based genotypes making their way into ‘wild’ populations.

4.4.3 *Connectivity*

Estimates of admixture (f3 statistics) between countries suggest a lack of significant admixture between the three countries involved, overall (from a larger geographic scale). This is in line with PCA, SNP tree, and F_{st} estimates, indicating to relatively isolated populations without any recent, significant (as in enough geneflow to affect genomic structure) movement between those of any other ‘ancestral’ populations. The much lower f3 estimates for Thailand and China as

the target populations, (China, India: Thailand; Thailand, India: China respectively) is likely due to the isolate Chin-Qing2 being very similar to two Thai Isolates (Thai-Pathun, Thai-Chao). This also suggests, if only minimal, that there might be some dispersal between China and Thailand. A better understanding of dispersal routes (whether they are anthropomorphic or not) may give a clearer picture.

4.4.4 *Varying ploidy of E. hepatopenaei*

The ploidy of many Microsporidia is unknown, largely due to isolated nuclei through the lifecycle of such taxa, making it difficult for microscopy-based methods (Cali *et al*., 2017). High heterozygosity has been described in some *Encephalitozoon* species, indication polyploidy in those species (Selman, 2014), and other species have also been suggested to be diploid (Cuomo *et al*., 2012; Haag, Sheikh-Jabbari, *et al*., 2013; Haag, Traunecker, *et al*., 2013). Although it is not known whether this is a biological feature shared among all Microsporidia. The consistent (12/15) diploid identification for the isolates used in this study, using kmer distribution, does suggest that *E. hepatopenaei* is also diploid. However, in some cases (Thai-Chan1-1; Thai-Chan1-2; Thai-Pathun, Indi-Dah; Indi-Saf3) the proportion of kmers with a diploid distribution is comparable to the second highest estimated ploidy. Interestingly, the second highest ploidy represented in the kmer distribution varies among the samples, with no visible pattern. One possibility for this, is that during different cell stages, *E. hepatopenaei* cells vary in ploidy (depending on the cell's developmental stage). A state described in other pathogenic organisms, especially when dealing with stress (Y. Li *et al*., 2017), though not one yet described in Microsporidia.

4.4.5 *Conclusions and applications in aquaculture*

In conclusion, this study shows that it is unlikely that *E. hepatopenaei's* recent "expansion" across South East Asian *P. vannamei* farms was due to recent (2009) anthropogenic movement. More likely that, given the geographic structure seen between isolates, *E. hepatopenaei* was widespread prior to *P. vannamei* and *P. monodon* being used in aquaculture; and it is possible that the change to the more susceptible *P. vannamei* in 2000/1 allowed it to become so prevalent. This has implications for the continued farming of penaeid shrimp. However, there may be signs of dispersal between Thailand, China and India. To prevent

continued infection of existing ponds or infection in new ponds, methods will have to be enforced that prevent outside environmental factors from influencing the farm. This also should inform the aquaculture industry, as Microsporidia, among many other intracellular parasites, are opportunistic. So, without sufficient prior screening of potential farming locations, for either expansion or new farms, farmers run the risk of picking up new infections. This is important, as the intermediate hosts of *E. hepatopenaei* have not been characterised. So, it is not yet known how spores get into ponds, or disperse between countries. Knowledge that would be required for more targeted prevention.

Though the route of transmission into ponds from local environments is still unknown, the fact that it is occurring across several countries indicates that it is via a medium that is very widespread but local enough for there to be population structure within countries. Spores may enter the pond through the water that is locally sourced, the soil (wide microsporidian diversity in soil (Ardila-Garcia *et al*., 2013)) of inland shrimp ponds and may be transmitted via an intermediate host that is widespread but does not range far (like insects/small birds, that would occur through all of these countries). As *E. hepatopenaei* has been found in polychaetes and crabs, it is possible that there are multiple aquatic vector species. The initial spread of *E. hepatopenaei* could have been facilitated by migratory birds that feed on these vectors, as is the case for many pathogenic species (Okamura *et al*., 2019).

4.4.5 *Further research and limitations*

Further research should be carried out on the rate of sexual reproduction (if not asexual), as it is not yet known whether these organisms solely reproduce asexually, and if sexually, in what proportion. The patterns observed in this study could be less pronounced depending on the predominant mode of reproduction. For example, if they primarily reproduce asexually the pattern seen could just be from a low number of clonal individuals from genetically divergent individuals. Though it is unlikely to be caused by a few initial introductions into ponds (unless they have been subsequently distributed in a local fashion, thus keeping similar genotypes just between ponds) as isolates are still closely related within countries on a basis of distance.

Another line of investigation that should follow this work is if there is a difference between farm-based isolates and “wild” isolates and if there is a greater selective force on those infecting shrimps. Thus, allowing for greater variability found in the “wild” types. This would help inform on the type of genomic variants that are able to thrive in shrimp ponds (if there is a variant), and whether these variants could be a danger to other farmed crustaceans. To do this, more environmental samples need to be collected from around the areas that have been selected in this study from a true comparison. The single-copy set of PCR regions will also help with this, to consider evolutionary origins.

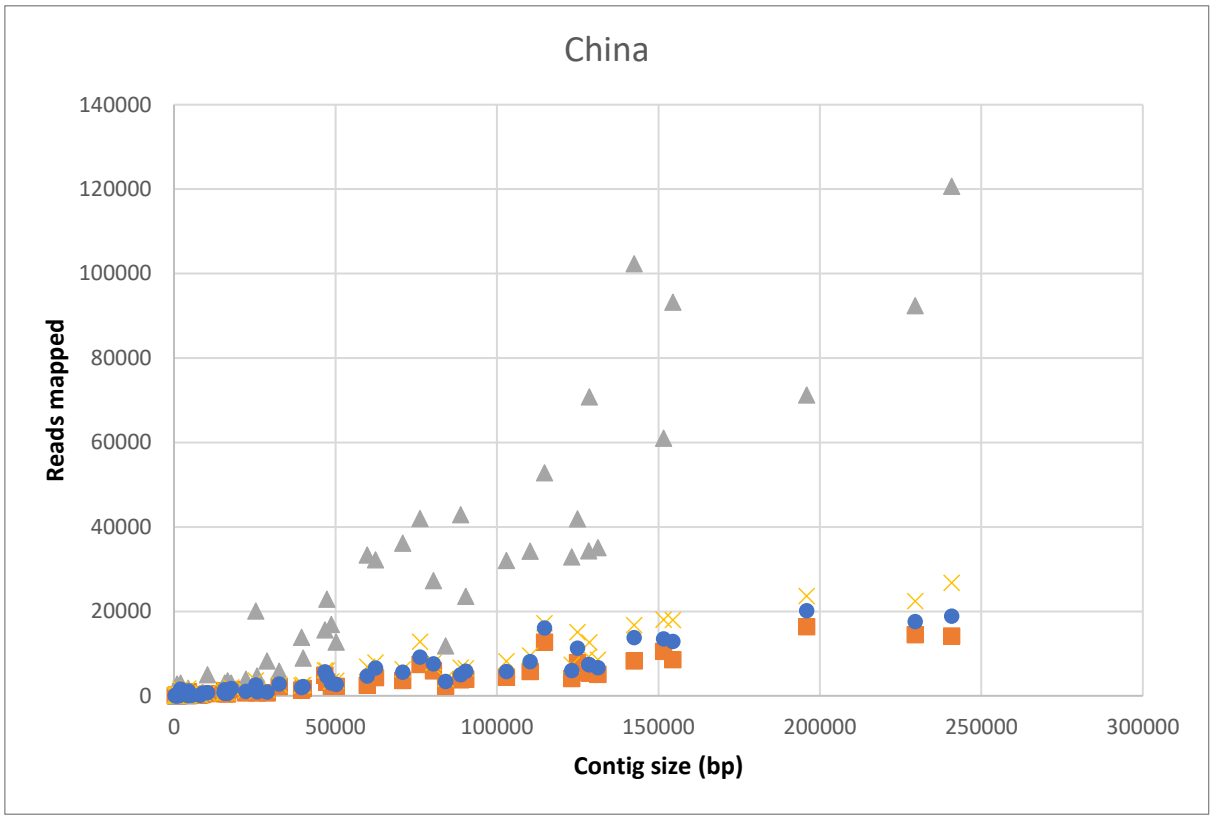
A limiting factor in this chapter is that the spores had to be pooled in order to get enough gDNA for sequencing. The pooling puts limitations on the type of analysis that can be carried out on these samples, due to not being able to attribute genomic characters to individuals. Due to this, although the analysis carried out was designed for pooled-seq, the population genomic indices estimated in this study are conservative and likely underestimates the complexity found between these isolates. Single cell sequencing would be a better tool to answer these questions more thoroughly.

4.5 Supplementary tables and figures

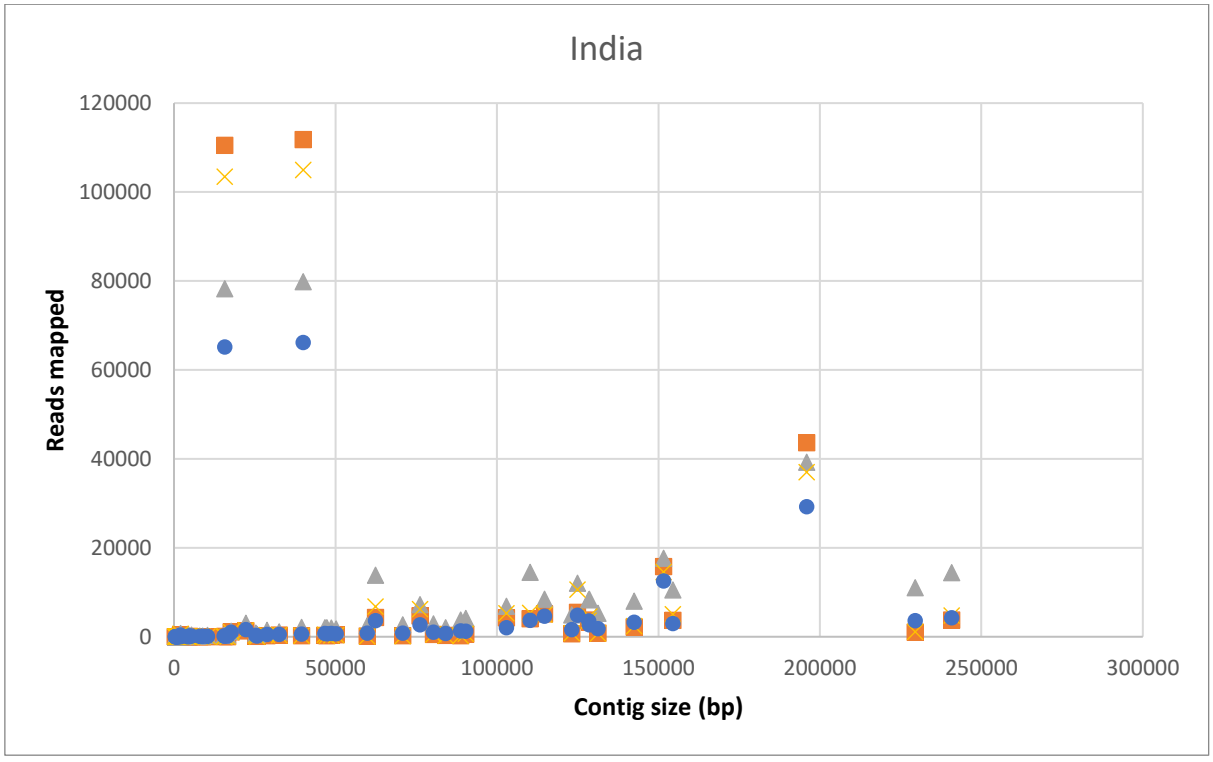
Figure 4.1a-c

Relationship between number of reads mapped and contig size

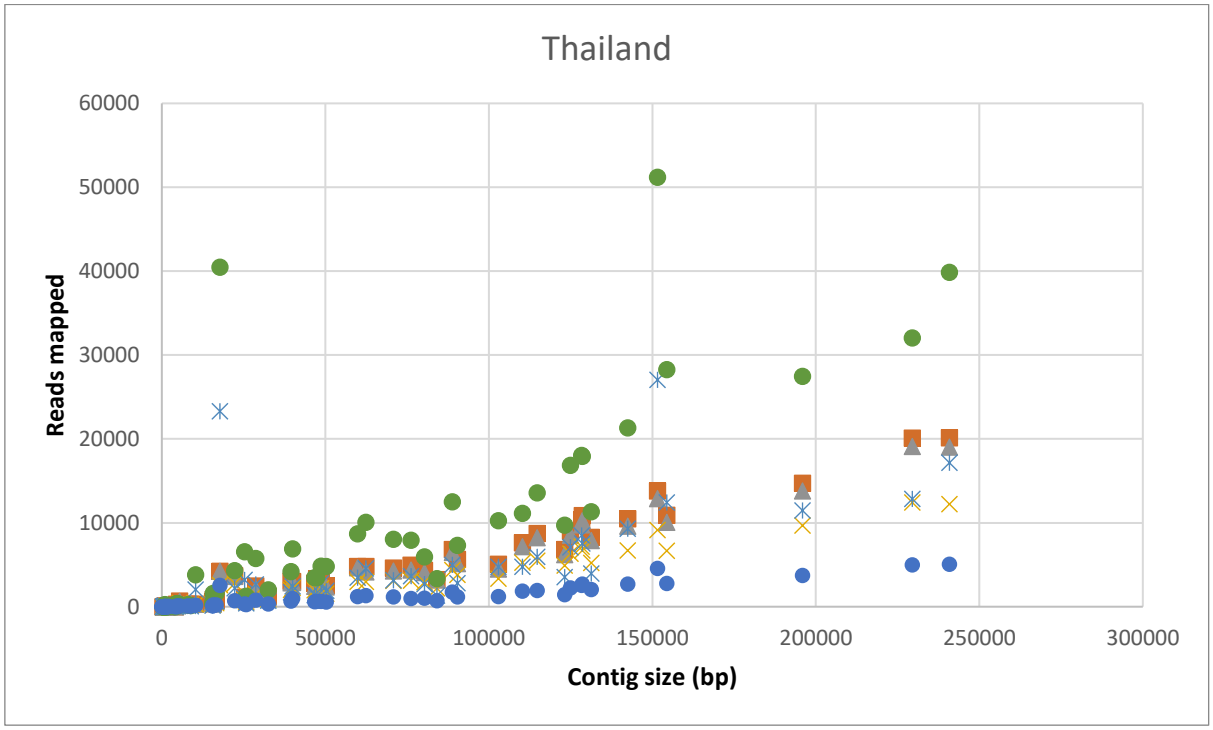
1a



4.1b



4.1C


Thailand
Reads mapped
60000
50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000
250000
300000
Contig size (bp)

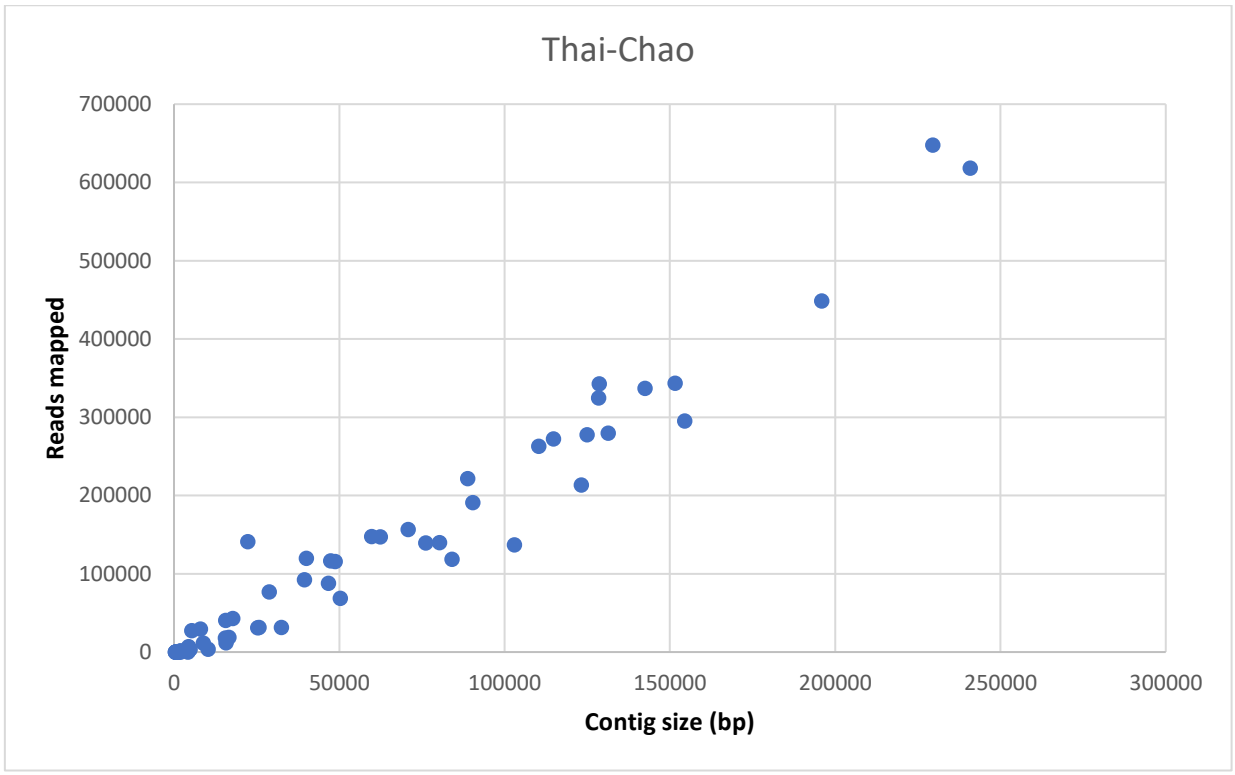
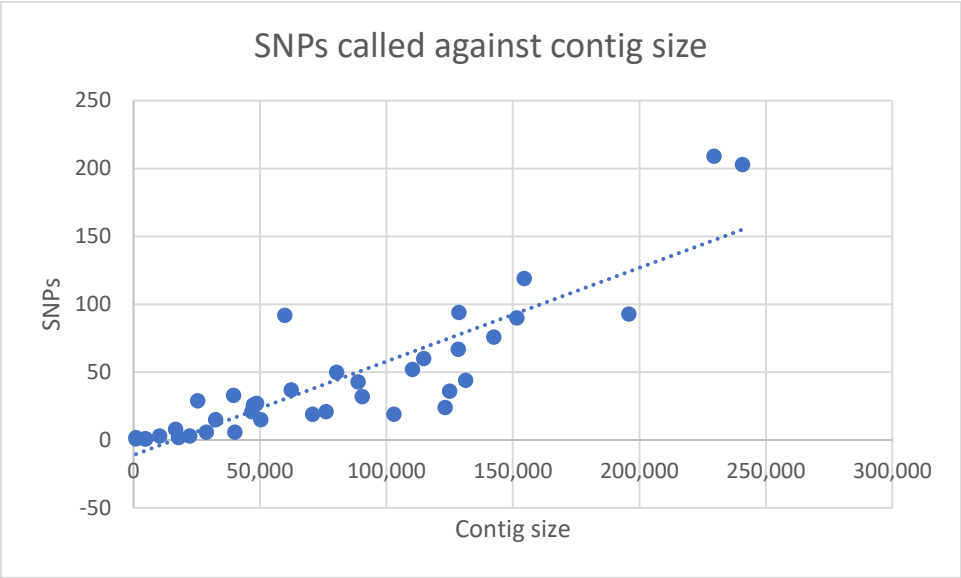
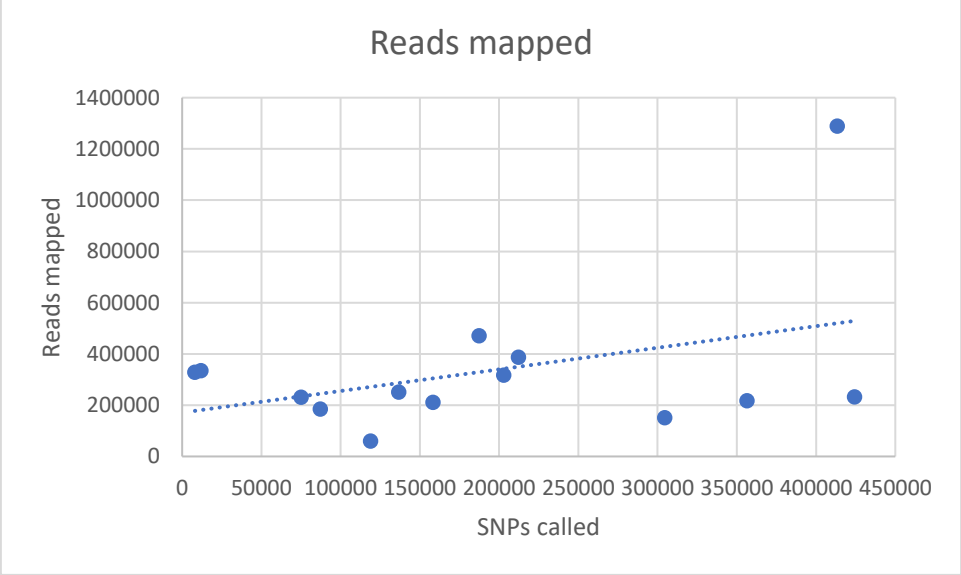

Thai-Chao
Reads mapped
700000
600000
500000
400000
300000
200000
100000
0
0
50000
100000
150000
200000
250000
300000
Contig size (bp)

Figure 4.2 a, b

4.2a). SNPs called against contig size



4.2 b). Relationship between the amount of SNPs called and reads mapped



Chapter 5: Patterns in Enterocytozoonidae infection and evolution

5.1 Introduction

Multi-gene phylogenies are very useful, adding greater resolution and confidence to phylogenies (Gontcharov *et al*., 2004) They have become very commonplace in the last two decades. They provide a more stable base for evolutionary assumptions, as inferences are being drawn from the plethora of varied evolution rates across loci all contributing to a more reliable picture of the evolutionary relationships between taxa under study. Multi-gene phylogenies are frequently used in resolving deep evolutionary histories, like eusociality in vespid wasps (Hines *et al*., 2007), diversification of plants on land (Finet *et al*., 2010), and evolution of tetrapod olfactory receptor genes (Kishida, 2008). It is the greater deep branch evolutionary resolution and reliability that make multi-gene analysis especially relevant to microsporidians. However, phylogenomic datasets can still result in incongruent trees, depending on the genes used for the phylogeny (Huang *et al*., 2010; Rosenberg & Tao, 2008). Conflicting results that would likely be minimised by using a set of conserved single-copy genes. More recently, within the last 10 years, multigene phylogenies investigating the relationship between Microsporidia and Microsporidia-like organisms have given new insight into the relationship of the phylum to other eukaryotes (Haag *et al*., 2014; Quandt *et al*., 2017).

The reduced costs in next-generation sequencing and advances in bioinformatics (more expertise and greater accessibility to high-performance computers) have allowed researchers to also investigate the evolution of proteins within the Microsporidia. Understanding the changes that have made this phylum of intracellular parasites so successful is of interest from both an academic and clinical perspective. This type of research has shown a set of Microsporidia-specific genes that are core across all species of Microsporidia, and likely associated with their parasitic lifestyle (Heinz *et al*., 2012). This addition of Microsporidia-specific gene families was accompanied by a loss of metabolic pathways and gene length reduction, attributed to the phylum-wide genome reduction (Heinz *et al*., 2012). Some of the core gene families that were novel to the microsporidia have bacterial-like nucleotide transport domains, likely

associated with ADP/ATP sequestration from the host (Cuomo *et al*., 2012; Heinz *et al*., 2012) and hexokinases with secretion signal sequences shown to be able to export proteins out of the cell (Cuomo *et al*., 2012; Pombert *et al*., 2012).

Host-shifting, an event when a parasite manages to infect and establish itself in a new host species, is a well-studied interaction in the evolution of parasitic taxa (Marston *et al*., 2017; Sakwinska *et al*., 2011). A large factor in the success of host-shifting, is the close proximity of two phylogenetically related organisms, as the probability of host shifting decreases with increased phylogenetic distance between hosts (Braga *et al*., 2020; Engelstädter & Fortuna, 2019). This is why the emergence of pathogens that influence humans, are generally either through agriculture (livestock) (D'Agostino & Cook, 2015; Mughini-Gras *et al*., 2018)or environmental/domestic interactions with mammals (cats and dogs) (Baneth *et al*., 2016; Damborg *et al*., 2016; Tan, 1997). Largely, host-shifting studies looking at parasite trends, are carried out on RNA viruses, due to their high propensity to jump hosts, a trait that is associated with high mutation rates (Longdon *et al*., 2014; Loverdo & Lloyd-Smith, 2013). However, it is a widespread phenomenon occurring across both eukaryotic (Giraud *et al*., 2010; Navaud *et al*., 2018) and bacterial life (Bonneaud *et al*., 2019; Mrochen *et al*., 2018). Gene mutations associated with host-shifts are gaining more attention (Anishchenko *et al*., 2006; Linster *et al*., 2014; Loverdo & Lloyd-Smith, 2013; Woolhouse *et al*., 2005), as a better understanding of them could impact therapeutics and enable researchers to better predict pathogenic host-shifts that could become problematic (Pimentel *et al*., 2021). For example, research done on the Avian A/H5N1 influenza virus showed that only five amino acid changes are required to be able to transmit between different host ferret species (Linster *et al*., 2014).

Opportunistic parasites are also more likely to have a host-shifting event, but this is a different process to the one posed for the parasites mentioned above, dubbed as 'professional' pathogens (Martínez, 2014). For opportunistic parasites, it is the host's inability to stop an infection, that allows it to proliferate, while professional pathogens may actively inhibit or evade host immune responses. This can come about due to a suppressed immune system as a result of ill health, or temporary effects, like prolonged stress due to non-optimal environment conditions, or lack of nutrients (Glaser & Kiecolt-Glaser, 2005; Webster Marketon & Glaser, 2008). Generally, opportunistic parasites are picked up from the

environment and many are able to parasitise the host asymptomatically. It is the reduction in the immune system that allows them to take advantage of a weakened host. This allows opportunistic pathogens to take a broader approach to infection, in terms of infection-based genes and pathways. For example, *Candida albicans* binds to phospholipids (Tams *et al*., 2019), molecules that are widely found across many epithelial cell types (Shewan *et al*., 2011), allowing it to infect a large range of hosts once immunocompromised. All members of the Enterocytozoonidae (Microsporidia: Enterocytozoondiae) follow a similar opportunistic pattern, all infecting the easily accessed epithelial/enterocyte cells of the digestive tract (G. D. Stentiford *et al*., 2019). Infection in all species of the Enterocytozoonidae have not been shown to progress to other organs/tissue types within the host, showing a transmission strategy to infect the tissue of first contact.

With the well-described presence of multi-host infections discussed above, it has been shown frequently that some microsporidian species have a number of different hosts (Quiles *et al*., 2019; K. Snowden *et al*., 1999; K. F. Snowden *et al*., 2009). However, it is not yet known if there is a drastic change in infection dynamics when microsporidian taxa change host. There are two possible reasons why they are able to host-shift so effectively: one being that due to the high mutation rates, they are able to adapt quickly to new host cells, a factor found to be associated with emerging pathogens (Alexander & Day, 2010); second, the fact that they are largely generalist parasites and infect similar tissue types in different species, for example, *C. albicans*. The Enterocytozoonidae is a family that generally infects the same tissue and host types: with exception of *E. bieneusi*, they infect aquatic hosts, and these are generally crustacea. With the Enterocytozoonidae in a largely terrestrial infecting clade (Terrasporidia), this family may represent a good group in which to investigate the molecular basis of any change in infection processes in a relatively more homogeneous background (in terms of host and tissue preference) compared to other lineages.

To investigate how genomics can be used to try to identify the molecular basis of changes in infection strategy this study investigates the proteins associated with host-shifting in the Enterocytozoondiae, using patterns of loss and gain of protein families (orthogroups) within the Enterocytozoonidae, at the level of the whole clade and for the individual. The analysis will focus on previously described genes

associated with pathogenicity. Genes associated with pathogenicity are typically geared in one of two ways: to enable quicker invasions of the host (Josling *et al*., 2015) or to better resist or evade host immunity (Thammavongsa *et al*., 2015). Linked to the parasite's ability to proliferate in a new host is the host's susceptibility to the novel pathogen (Beldomenico & Begon, 2010). A phylogenomic approach to analyse whether there are any particular protein families gained that are associated with infection will be adopted, looking for patterns between the nodes they were gained on and the hosts of the enterocytozoonids in that lineage. There are currently no published genomes sequences from the second fish infecting branch of the Enterocytozoonidae ((*Para)Nucleospora/Obruspora*), as such this clade is omitted from this study.

With the Enterocytozoondiae having some of the most reduced genomes among Microsporidia that have an available genome, besides those in the *Encephalitozoon* genus, another aim of this study is to investigate the orthogroups associated with the reduction of genome size within the Enterocytozoonidae. Genome reduction is a widely researched phenomenon within the Microsporidia, and this study may identify characteristics of particular enterocytozoonid lineages in which gene loss has been more extreme than in others.

5.2 Method and analysis

5.2.1 *Building the species tree/orthogroups assignment*

To build a robust phylogeny, microsporidian proteomes were acquired from the NCBI genome database (Table 5.1-supplementary). However, the genome sequence of the aphelid *Paraphelidium tribonemae* was acquired from a private database (Karpov *et al*., 2017). Some of the genomes are partially assembled (*Amphiamblys sp*.); those of the *Encephalitozoon* species are assembled to chromosome-level. Fungal outgroups were taken from major fungal clades (Table 5.1-supplementary).

OrthoMCL version 2.0 (Fischer *et al*., 2011) was then used to assign all proteins from all proteomes (Table 5.1-supplementary) into orthogroups. orthomclFilterFasta was used to filter out poor quality proteins (less than 20 amino acids long) from all proteomes, the remaining set of proteins were used for the rest of the analysis. To create the all-versus-all blast file that is needed for OrthoMCL, a database was made from the dataset output produced by orthomclFilterFasta, using BLAST+ (makeblastdb) version 2.6.0 (Altschul *et al*., 1990; Camacho *et al*., 2009) under default parameters. This database was then used to perform a BLASTP search against the filtered proteomes used to make the database, using BLAST+ (blastall) with a tabulated output [-m 8] (Altschul *et al*., 1990; Camacho *et al*., 2009). The output from OrthoMCL was used to build the species tree, using only single-copy genes. To get only single-copy genes, a custom script was used on the tabulated file outlining the frequency of that gene per species per orthogroup. Only genes that were single copy and present in all species in the dataset were selected. The PSI-Coffee version 11.0 (Floden *et al*., 2016)algorithm was used to align the individual data sets using homology extension. trimAl version 1.2 (Capella-Gutiérrez *et al*., 2009), using the [-gappyout] (a more lenient trimming algorithm, as the sequences, are so divergent) was used to remove the highly heterogeneous sections in all orthogroups. Model selection tests were carried out using IQ-TREE version 1.6.1 (Nguyen *et al*., 2015). Models were selected using the Akaike information criterion (AIC) (Table 5.2-supplementary). The ortho-groups were then concatenated into one dataset for partitioned analysis. A partition file was subsequently made for the analysis. Maximum likelihood trees were built using RAxML-HPC version 8.2 (Stamatakis, 2014) assessed with 1000 bootstrap replicates, using the partitioned model file created previously. The tree was viewed and edited in TreeGraph version 2.0 (Stöver & Müller, 2010).

5.2.2 *Mapping orthogroups to a species tree*

Investigation of protein families gained and lost in the enterocytozoonid lineage was carried out by mapping the loss and gain of protein families to the species tree. All protein families produced from the OrthoMCL analysis, prior to the single-copy gene filtering, were mapped onto the tree using amalgamated likelihood estimation (ALE) version 0.4 (Szöllosi *et al*., 2013). ALE uses likelihood

estimation to amalgamate gene trees to estimate their probabilities, testing different ways to map gene trees onto a selected species tree. ALE requires 1000 trees per orthogroup as burn-in for the analysis, as such, to produce the gene trees that will be mapped on the species tree, IQ-TREE was used to build 4000 trees for each orthogroup. All of which are mapped to the species tree to best estimate that orthogroups gain and loss among different lineages. Datasets for tree building were prepared using MAFFT version 7.310 (Katoh & Standley, 2013) to align sequences under default parameters, instead of PSI-Coffee, due to computing/time restrictions (the number of trees required), and trimAl [-gappyout] was used to mask the highly heterogeneous regions. The analysis was run following the protocol outlined on the ALE GitHub webpage (https://github.com/maxemil/ALE), the ALE objects were built using the script ALEobserve discarding 1000 trees (1/4 of the trees in the tree list). ALEdated was used to reconcile the trees, with outgroups included nodes were ordered relative to each other. The reconciliation event files were compiled using a custom script, and the gain and loss events were then mapped onto the species tree by eye. Gene duplications events and gene copy number were also assessed in ALE, as duplication events have a strong association on the size of genomes (Sheridan *et al*., 2020).

5.2.3 *Novel enterocytozoonid orthogroups*

To look at the possible functional distribution of orthogroups gained in the Enterocytozoonidae, orthogroups that were gained or lost at and after the node representing a common ancestor for *H. eriocheir*, *E. bieneusi*, *En. canceri* and *E. hepatopenaei* were put into multiple datasets depending on the node at which they were gained or lost. There are seven node-dependent datasets in total, one incorporating all members, then one without *H. eriocheir*, one without *H. eriocheir* and *E. bieneusi* and four separate detests for each individual species (Table 5.2). InterProScan version 5.36 (Jones *et al*., 2014) was used to search for protein domains against a number of protein domain databases with the aim of inferring possible functions for any gained and lost protein orthogroups of interest. However, it is also likely that none of the enterocytozoonid genome assemblies are complete, which could result in false negatives for lost/gained orthogroups, in regard to an absence of a gene (s). To help better understand the function of

these selected orthogroups, subcellular localisation was predicted using WoLF PSORT version 0.2 (runWolfPsortSummary) (Horton *et al*., 2007), run against the fungal database. Only results with a confidence value of 10 and above were considered. WoLF PSORT is an extension of PSORT 2, which uses amino acids to predict protein localisation based on amino acid composition, functional motifs, and sorting signals.

5.3 Results

There was a total of 132571 protein sequences from all species involved in this study. The filtering for the protein database only removed one sequence that was less than 20 amino acids long. The OrthoMCL analysis resulted in 15,369 ortho-groups. The biggest (OG1000) was solely composed of *Amphiamblys* sp. proteins and consisted of 544 uncharacterised proteins. These likely represent the repetitive elements spoken about in (Mikhailov *et al*., 2016). A large portion of orthogroups were not like this, however, with 7590 of 15, 369 orthogroups being made up of two or fewer genes. The single-copy gene script used on the ortho-groups resulted in 125 single-copy gene orthogroups that appeared in every species used in this study (Table 5.2-supplementary). The gene identity and models were chosen for each orthogroup can be seen in Supplementary Table 5.1. Across all seven enterocytozoonid nodes (datasets), 728 orthogroups in total were estimated to have been gained and 2, 401 lost at different nodes throughout the enterocytozoonid clade. The number of gained/lost orthogroups varied among the seven nodes (Figure 5.1). The greatest number of novel ortho-groups gained for a single enterocytozoonid species is *H. eriocheir*, with 215 gains and the least was *E. hepatopenaei*, gaining 55 orthogroups (Figure 5.1). There is evidence for a greater relatedness, based on patterns of shared orthogroups, between the enterocytozoonids excluding *H. eriocheir*; The node experiencing the greatest number of lineage-specific of orthogroups within the Enterocytozoonidae was the node encompassing *En. canceri*, *E. hepatopenaei,* and *E. bieneusi* (193) (Figure 5.1). In all nodes, more orthogroups were lost than gained. The greatest loss was seen from *H. eriocheir* (447). This was followed by *E. bieneusi* (403) and then by the node estimated to be the enterocytozoonid ancestor (393).

5.3.1 *History of Genomic reduction in Enterocytozoonidae*:

A phylogeny (Figure 5.1) based on 125 genes (Table 5.2-supplementary) from 34 taxa showed strong support (>70 bootstrap support on all nodes (average 98)) for the current understanding of microsporidian phylogenetic relationships, based on phylogenomics (Mikhailov *et al*., 2016; Quandt *et al*., 2017; Torruella *et al*., 2018)) and the 18s rDNA (Vossbrinck *et al*., 2014; Vossbrinck & Debrunner-Vossbrinck, 2005).

To investigate the role the rate of orthogroup gain and loss has in genome reduction, a relative measure of how many orthogroups were gained and lost was taken relative to the evolutionary rate as inferred by the number of substitutions (branch length). This was done by dividing the cumulative (cumulative from the ancestral polar tube forming microsporidian, marked on Figure 5.1*) orthogroups gained/lost by the cumulative branch length (cumulative from the same point), which was multiplied by ten ((orthogroups gained/branch length) x 10). The resultant value represents the number of orthogroups gained/lost per 0.1 substitutions (Table 5.1), which was chosen as the species tree (Figure 5.1) scale bar has intervals representing 0.1 substitutions. When looking at nodes including genera/families, of the taxa included in this tree, the Enterocytozoonidae showed the most conservation, gaining the fewest orthogroups with 54.06 orthogroups gained per 0.1 substitutions, the remaining

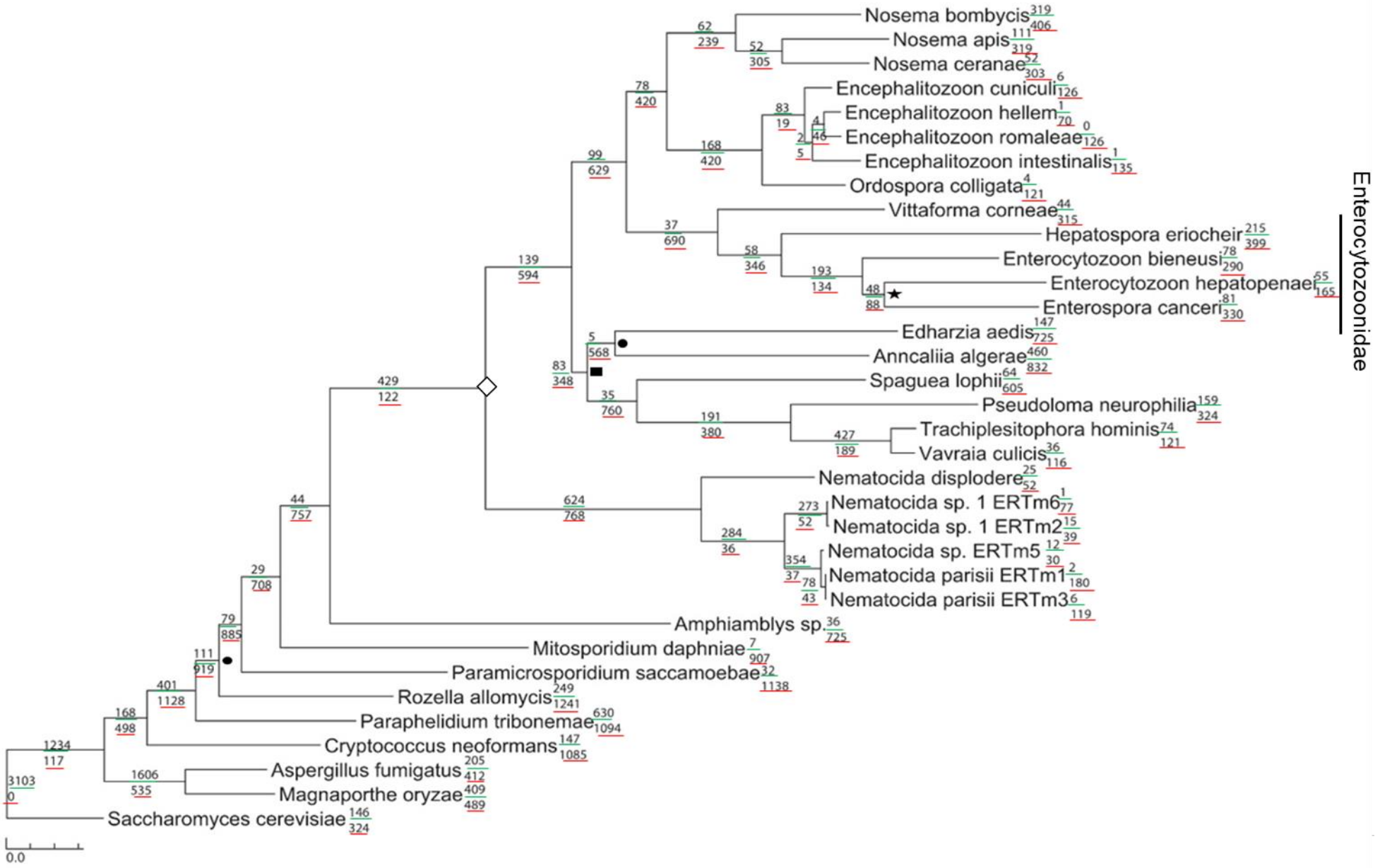

Nosema bombycis
Nosema apis
Nosema ceranae
Encephalitozoon cuniculi
Encephalitozoon hellem
Encephalitozoon romaleae
Encephalitozoon intestinalis
Ordospora colligata
Vittaforma corneae
Hepatospora eriocheir
Enterocytozoon bieneusi
Enterocytozoon hepatopenaei
Enterospora canceri
Edharzia aedis
Anncaliia algerae
Spaguea lophii
Pseudoloma neurophilia
Trachiplesitophora hominis
Vavraia culicis
Nematocida displodere
Nematocida sp. 1 ERTm6
Nematocida sp. 1 ERTm2
Nematocida sp. ERTm5
Nematocida parisii ERTm1
Nematocida parisii ERTm3
Amphiamblys sp.
Mitosporidium daphniae
Paramicrosporidium saccamoebae
Rozella allomycis
Paraphelidium tribonemae
Cryptococcus neoformans
Aspergillus fumigatus
Magnaporthe oryzae
Saccharomyces cerevisiae
0.0
Enterocytozoonidae

Figure 5.1: Gain and loss of orthogroups of microsporidian genomes. The ancestral state of orthogroups were estimated using Amalgamated likelihood estimation (ALE). Gains (green underlined) and losses (red underlined) are mapped onto a species tree based on 125 single-copy orthologs shared by all taxa (34 species). Inferred using maximum likelihood in RAxML-HPC (Stamatakis, 2006). All nodes have bootstrap values of 100 unless marked with a star (90-99), a circle (80-89), or a square (70-79). *Polar filament-forming ancestral node marked with a diamond

taxa (*Nematocida*, *Nosema,* and *Encephalitozoon*) ranged from 72.69-13.98 (Table 5.1). The inverse is true for orthogroups lost, with the Enterocytozoonidae having lost the most, losing 365.26 orthogroups per 0.1 substitutions, the remaining taxa ranging from 170.60-360.19. The same pattern is maintained looking at the relative gain at a per species level (Figure 5.2, Table 5.3-supplementary), though the relative loss of orthogroups is not as pronounced among the Enterocytozoonidae.

Despite being the most genetically divergent (longest branch length) family/genera in this study, in terms of the genetic distance from the Microsporidian polar filament-forming root, the Enterocytozoonidae show some of the lowest rates of orthogroup gain. However, there did not appear to be a positive correlation, if any, between orthogroup gain and the length of the branch (divergence of that taxa) (Figure 5.1a-supplementary). Whereas orthogroup loss showed a linear positive correlation (Figure 5.1b-supplementary). When the outliers are removed (*Nematocida*), the relationship between orthogroups lost and branch length remained positive (Figure 5.1c-supplementary), whereas that between orthogroups gained and branch length seem to show no relationship (Figure 5.1d-supplementary).

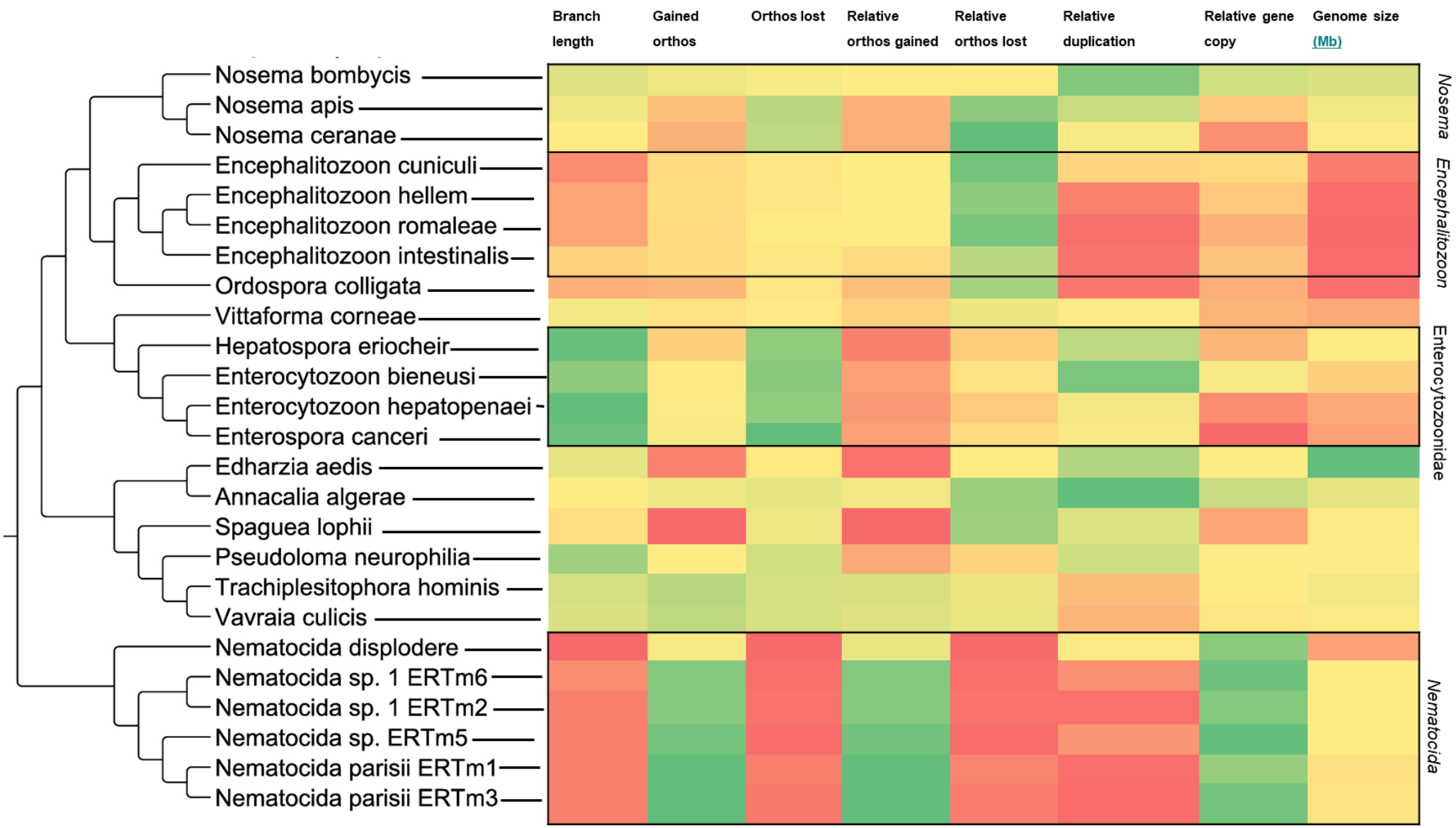
5.3.2 *Genome expansion*

To get a better understanding of whether enterocytozoonid genomes are continuing to reduce in size, a measure of duplications per 0.1 substitutions (a relative measure of duplications and gene copies were also taken in the same way as above, with orthogroups gained and lost) was taken. Duplications per 0.1 substitutions varied widely across nodes and taxa (range: 0-0.28 duplications per 0.1 substitutions) (Figure 5.2, Table 5.3-supplementary). Comparatively, across the four clades represented in this tree, the Enterocytozoonidae had the highest duplication rates averaging at 0.12 (range: 0.05-0.25) (Figure 5.2, Table 5.3-supplementary). The lowest were those of the *Nematocida* (average: 0.01, range: 0-0.04),and *Encephalitozoon* (average: 0.01, range: 0-0.03), with *Nosema* (average: 0.13, range: 0.05-0.23) in between. Conversely, across the four clades being compared in this tree, the Enterocytozoonidae had the lowest number of gene copies (average: 0.08, range: 0.06-0.11), relative to branch length and

Nematocida has the greatest (average: 0.26, range: 0.23-0.29) (Figure 5.2, Table 5.3-supplementary).

Table 5.1: Number of accumulative orthogroups gained for each clade being compared. Orthogroup loss and gain were assessed using the Amalgamated likelihood estimation (ALE)

Clade	Branch length	Gained orthos	Orthos lost	Relative orthos gained	Relative orthos lost	Relative duplication	Relative gene copy
Enterocytozoonidae	0.62	333	2250	54.06	365.26	0.00	0.07
Encephalitozoon	0.66	567	2073	85.52	312.67	0.01	0.10
Nosema	0.52	378	1873	72.69	360.19	0.01	0.10
Nematocida	0.45	624	766	138.98	170.6	0.00	0.22


Branch length
Gained orthos
Orthos lost
Relative orthos gained
Relative orthos lost
Relative duplication
Relative gene copy
Genome size (Mb)
Nosema bombycis
Nosema apis
Nosema ceranae
Encephalitozoon cuniculi
Encephalitozoon hellem
Encephalitozoon romaleae
Encephalitozoon intestinalis
Ordospora colligata
Vittaforma corneae
Hepatospora eriocheir
Enterocytozoon bieneusi
Enterocytozoon hepatopenaei
Enterospora canceri
Edharzia aedis
Annacalia algerae
Spaguea lophii
Pseudoloma neurophilia
Trachiplesitophora hominis
Vavraia culicis
Nematocida displodere
Nematocida sp. 1 ERTm6
Nematocida sp. 1 ERTm2
Nematocida sp. ERTm5
Nematocida parisii ERTm1
Nematocida parisii ERTm3
Nosema
Encephalitozoon
Enterocytozoonidae
Nematocida

Heatmap scale: Branch length: 0.69-1.17; Gained orthos: 321-1346; Orthos lost:818-2802; Relative orthos gained: 40.53-189.82; Relative orthos lost: 119.24-310.51; Relative duplication: 0-0.28; Relative gene copy: 0.06-0.29; Genome size: 2.19-51.31Mbp

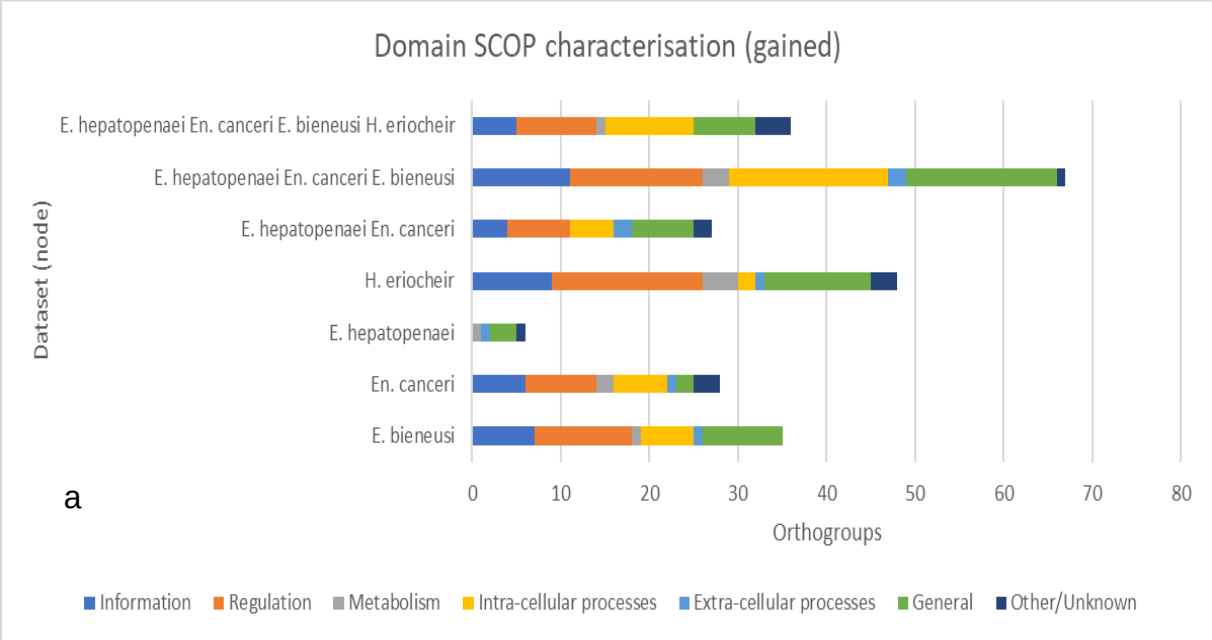
Figure 5.2: Heatmap representing orthogroups gained and lost for each microsporidian species included in this tree, relative to the accumulative branch length. (Heatmap key: factor dependent scale of relative values represented by a range of colours between red and green: red= relatively low; green= relatively high). Heatmap values are relative to each factor (column) being considered (for example, *Nematocida displodere*, relatively, has the shortest branch length but has on the highest values for relative gen copies). Clades being compared are emboldened to highlight clade-based patterns. Orthogroup loss and gain were assessed using the Amalgamated likelihood estimation (ALE).

5.3.3 *Enterocytozoonid Structural Characterisation of orthogroups (proteins) (SCOP)*

To investigate any signatures of enterocytozoonid pathogenicity, InterPro was used to analyse the functional and structural domains of the gained and lost orthogroups. To further categorise the gained and lost genes in the orthogroups, the seven general Structural Characterisation of Protein (SCOP) functional categories were implemented (Vogel, n.d.), across all lineages. This was also to allow for inferences on the function of the genes lost, in relation to genome reduction. Characterisation of genes gained within the Enterocytozoonidae followed similar patterns of lineage-specific gene family expansions (LSEs), including a large number of structural (SCOP characterisation: general), nucleotide-binding proteins, and kinases (SCOP characterisation: regulation) (Lespinet *et al*., 2002) (Figure 5.3). There were 155 different characterisations for functional/structural domains/motifs gained across all enterocytozoonid nodes, with 132 of these being unique to a single lineage, 23 were found in multiple lineages (Table 5.4-supplementary). Domains/motifs that are described as 'like', having a similarity to other described domains, made up a greater percentage in gained orthogroups (20%-34.4%) than orthogroups lost (14.4%-20.6%). Many of the domains gained but unique to a lineage are largely predicted to be housekeeping regulatory proteins, in the form of ribosomal proteins, cell surface channels/pumps, DNA replication, and membrane proteins; Indeed, the majority of the orthogroups gained are regulatory (Figure 5.3a). SCOP characterisation of InterPro analysis was largely automated, however, not all domains have been included in the SCOP database. For the domains missing from the SCOP database, manual searches for functional characterisation were carried out using the EMBL-EBI domain search webpage (Madeira *et al*., 2019), with the domain as the search criteria. Per dataset, proportionally, the base node of the Enterocytozoonidae had the greatest addition of intracellular process domains (41.7%, other enterocytozoonid nodes: 0%-29.8%).

5.3.4*Genome reduction through the loss of (retro)transposable elements and metabolic genes*

The loss of transposable elements has been associated with reduced genomes, and therefore the loss of transposable elements was investigated in this study. There were 23 (retro)transposable elements lost (Table 5.4-supplementary) throughout the enterocytozoonid lineage. A small number of domains associated with (retro)transposable elements were gained by one enterocytozoonid lineage, *H. eriocheir*. *H. eriocheir* gained a further four domains associated with (retro)transposons (Reverse transcriptase/ retrotransposon-derived protein, RNase H-like; Tc1-like transposase, DDE domain; Transposase InsF-like, Transposase, Tc1-like) (Table 5.4-supplementary). Across all lineages, the majority (range: 24.43%-41.80%) of domains lost were metabolism associated and mainly domains involved with energy production, transportation, and 'other' metabolic enzymes. The second-highest losses through all nodes were genes categorised as 'Information' (range: 23.71%-39.68%), in the form of genes that are associated with translation and DNA repair. Genes categorised as metabolic made up a much smaller percentage of orthogroups gained (0%-16.67%), with the largest percentage from *E. hepatopenaei* (16.67%).



a

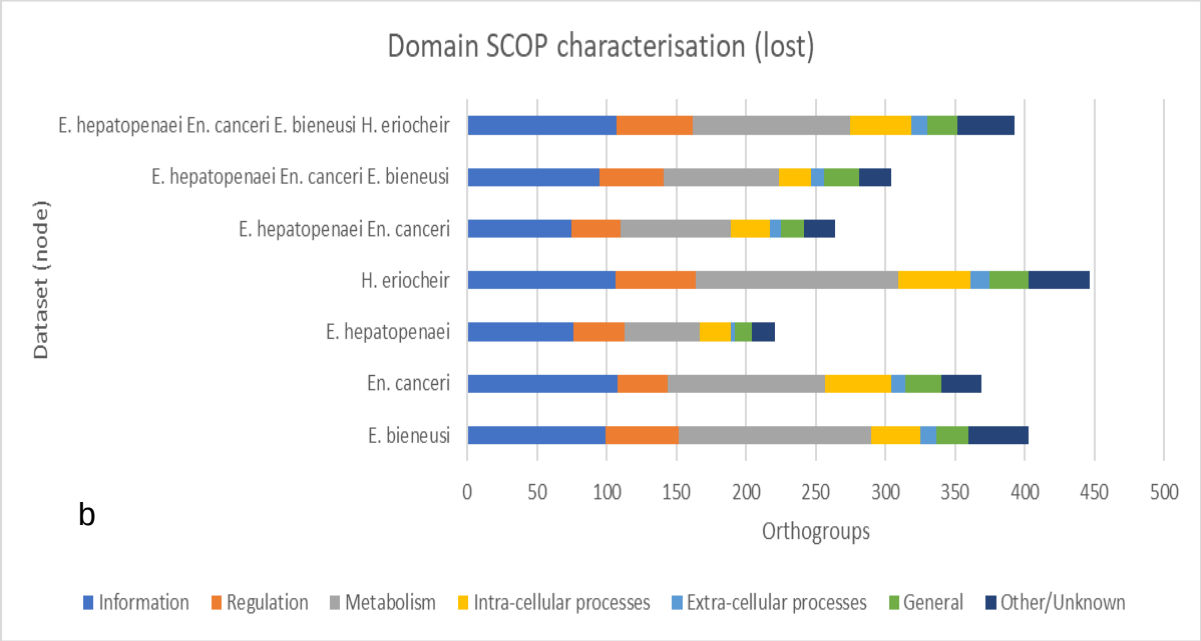


Figure 5.3: Functional annotation (7 SCOP general definitions) of orthogroups gained (a) and lost (b) throughout the Enterocytozoonidae

5.3.5 *Ricin B lectins and extracellular protein orthogroups, integral for infection?*
Ricin B- (like) lectin domains were focused on for orthogroups gained, as they have been associated with infection in Microsporidia (*N. bombycis*), as they were differentially regulated after spore germination (H. Liu *et al*., 2016). Lectins are also associated with binding in multiple parasites (Loukas & Maizels, 2000; Petri *et al*., 2002). Nine orthogroups returned with a Ricin B- (like) lectin domain, all of which were Enterocytozoonidae lineage-specific in that the orthogroups do not include species outside of the Enterocytozoonidae (Table 5.2, Figure 5.4a). However, this may also represent high genetic divergence instead of novel, specific proteins; one orthogroup (OG9557) had a 24-35% identity with *N. bombycis* Ricin-B Lectins. 55.5% were also species-specific, while 66.6% were specific to the exclusion of *H. eriocheir* (Figure 5.4).

To further characterise the orthogroups gained and lost within the Enterocytozoonidae, subcellular localisation was predicted using WoLF PSORT. The aim was to identify pathogenic genes possibly associated with a change in host-type. Working on the assumption that many of the changes to infection-based tools were likely to involve extracellular proteins (not always), this part of

the study mainly focused on gained and lost orthogroups comprised of proteins predicted to be extracellular.

Figure 5.4 a, b: Ricin B-Lectins (a) and extracellular (b) orthogroups gained and lost (1) throughout the enterocytozoonid lineage. (Green boxes=gained, red boxes=lost and white boxes=not present)

As most orthogroups were made up of multiple protein sequences, localisation predictions were only considered if the prediction was in the majority (>50%) of the genes in the orthogroup. Gained gene domains characterised and predicted to be extracellular made up a very small percentage of (SCOP) InterPro characterised proteins (range:0.00%-16.67, average: 5.08%), and the WoLF PSORT predicted (0.00%-3.45%, average: 0.88%) proteins. These domains were also largely uncharacterised on NCBI and via InterPro (Table 5.2), with the exclusion of three orthogroups (OG5226, OG9682, and OG9678). These three orthogroups, though were not characterised using InterPro likely due to them being too genetically divergent. Although they have been described in previous studies and have been associated with infection (OG9682:Enp1- has been reported to be involved in host adherence; OG9678: PTP3- function not entirely known, though has no proven adherent ability so probably structural). Although all of the orthogroups estimated to be localised extracellularly were estimated to be lineage-specific to the Enterocytozoonidae, as with the Ricin-B Lectin, it is likely that the two orthogroups (OG9678 and OG9682) were homologous to their respective BLAST results, but were too genetically divergent to be grouped in orthogroup analysis. Especially as Enp1 has not been described for the Enterocytozoonidae.

Table 5 2: Gained enterocytozoonid orthogroups predicted to be extracellular/Ricin B lectins, using WoLF PSORT and InterPro.

Orthogroup ID	**NCBI accession (s)**	**Node/taxa that have Gained orthogroup**	**Node/taxa that have lost orthogroup**	**Blast similarities/ characterisation (% identity)**	**InterPro Domain/motif**	**WoLF PSORT Localisation**
OG9522	EED43039.1 EED44094.1 OQS54814.1	*Enterocytozoon/ Enterospora*	*En. canceri*	Same as queries/ uncharacterised	Ricin B, lectin domain	Mitochondrial
OG7845	EED44527.1 ORD92901.1 ORD93447.1 OQS55637.1	*Enterocytozoon/ Enterospora*	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Cytoplasmic/ Nuclear
OG7879	ORD94928.1 ORD93353.1	*En. canceri*	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Nuclear
OG8012	OQS54094.1 OQS54098.1	*E. hepatopenaei*	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Nuclear
OG13751	ORD94924.1 OQS54085.1	*E. hepatopenaei and En. canceri* node	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Cytoplasmic/ Nuclear
OG9557	OQS54829.1 ORD95179.1	*E. hepatopenaei and En. canceri* node	N/A	*Nosema bombycis*/ Ricin B lectin (24%-35%)	Ricin B-like lectins	Nuclear
OG14085	ORD98380.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Nuclear
OG4037	ORD98562.1 ORD99258.1 ORD98371.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	Ricin B, lectin domain	Nuclear

OG9851	ORD95215.1 ORD99256.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	Ricin B-like lectins	Nuclear
OG9682	ORD93056.1 ORD94312.1 OQS54765.1	*E. hepatopenaei and En. canceri* node	N/A	Various taxa/ EnP1 (29%-100%)	N/A	Extracellular
OG9514	EED43108.1 EED42011.1 EED42639.1	*E. bieneusi*	N/A	Same as queries/ uncharacterised	N/A	Extracellular
0G13675	ORD93086.1 ORD94353.1	*En. canceri*	N/A	Same as queries/ uncharacterised	N/A	Extracellular
OG5226	ORD93182.1 ORD93203.1 ORD93455.1 ORD93603.1 ORD94121.1 ORD94676.1	*En. canceri*	N/A	Various taxa/ M2K4 (25%-100%)	Protein kinase domain	Extracellular
OG9678	ORD92812.1 OQS53456.1 OQS53386.1	*En. canceri* and *E. hepatopenaei* node	N/A	*E. hepatopenaeil* PTP3 (98.48%)	N/A	Extracellular
OG14015	ORD93180.1 ORE00586.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	N/A	Extracellular
OG140147	ORD95562.1 ORD98663.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	N/A	Extracellular
OG14120	ORD99456.1 ORD99458.1	*H. eriocheir*	N/A	Same as queries/ uncharacterised	N/A	Extracellular

The lack of gained extracellular orthogroups is similarly the case with lost gene domains for (SCOP) characterised (range:1.36%-4.23%, average:2.88%) and localised domains (range:0.00%-4.55%, average: 1.28%). This is contrasted by the most commonly localised cellular substrate across all seven nodes, for both gained and lost proteins, the nucleus (Table 5.3, Figure 5.5). Conversely, the second most frequent localisation predictions among the datasets were 'mitochondria', with the exception of genes lost for the *H. eriocheir* node and the node representing the ancestor of all enterocytozoonids.

Table 5.3: Predicted percentage of total orthogroups gained or lost among different nodes, for the two most predicted localisations.

Node	Nucleus gained	Nucleus lost	'Mitochondria' gained	'Mitochondria' lost
E. hepatopenaei	62.5%	72.44%	22.5%	11.02%
E. canceri	63.79%	70.23%	17.24%	111.83%
E. bieneusi	51.79%	62.72%	19.64%	10.09%
H. eriocheir	75.32%	62.82%	17.09%	8.33%
E. hepatopenaei* + *En. canceri	43.24%	51.52%	16.22%	16.67%
E. hepatopenaei* + *En. canceri* + *E. bieneusi	63.31%	70.95%	13.67%	10.14%
E. hepatopenaei* + *En. canceri* + *E. bieneusi* + *H. eriocheir		75%/74.1%	13.64%/5.40%	

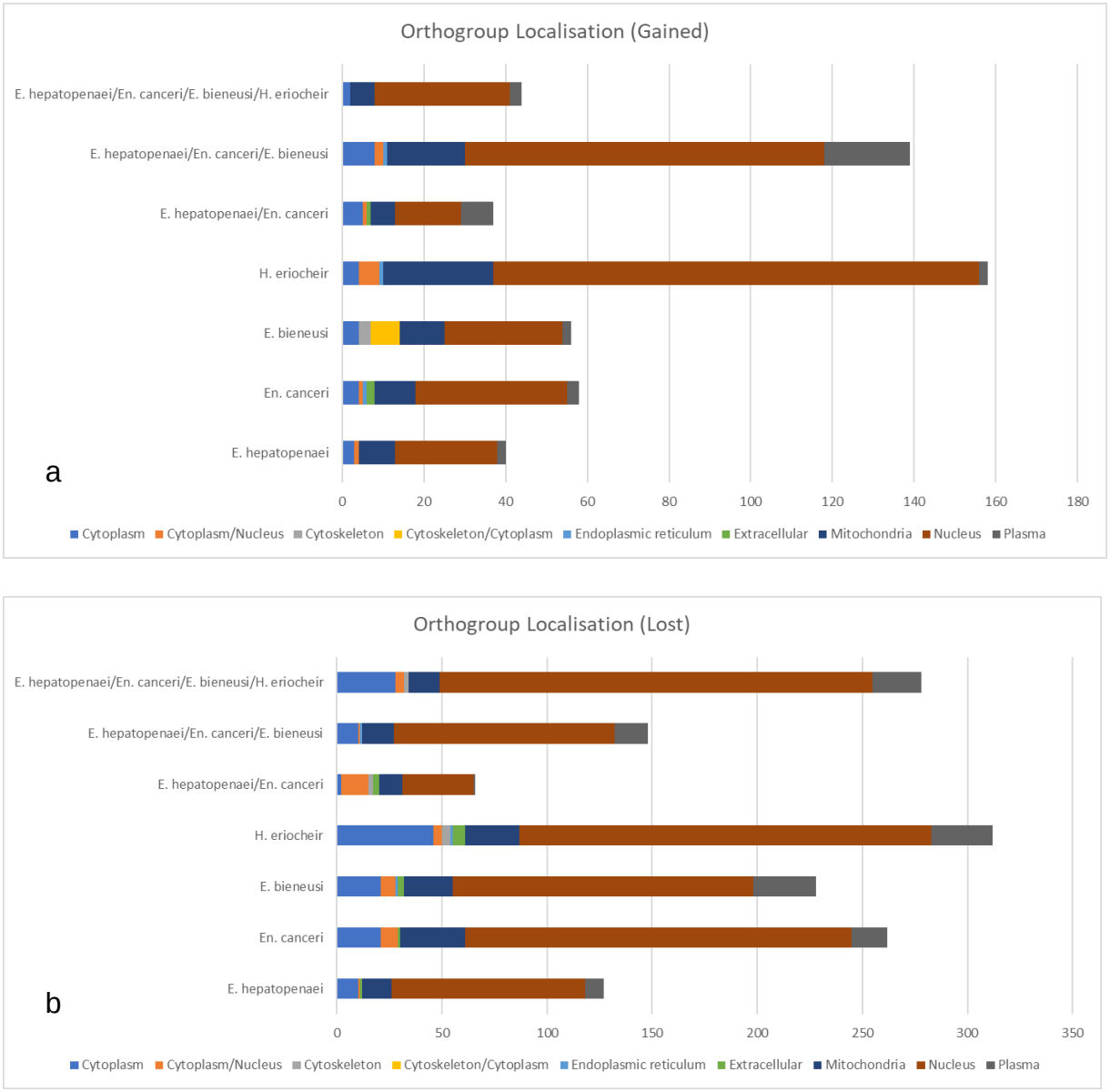


Figure 5.5: Predictions of subcellular localisation for orthogroups gained (a) and lost (b) within the Enterocytozoonidae, using WoLFPSORT

5.4 Discussion

This study bioinformatically investigated the proteins gained in the enterocytozoonid lineage, with a view to identify candidates that may be associated with infection and the host-shift to aquatic hosts in a clade composed largely of terrestrial infecting parasites. This study secondarily aimed to shine light on genome reduction in the Enterocytozoonidae, by investigating general trends in novel proteins gained/lost since the emergence of the enterocytozoonid clade. Important, as the Enterocytozoonidae, like the intensively studied

Encephalitozoon genus, have similarly drastically reduced genomes. Recent multi-gene phylogenies that include an extensive range of Microsporidia are currently lacking. Although the one shown here was primarily produced as a means to provide a robust tree to look at evolutionary processes (orthogroup gain and loss), to date (to the best knowledge) it includes the most single-copy protein-coding genes across the largest selection of microsporidian species, providing a strong framework for further research on within microsporidian evolution.

5.4.1 *Genetically divergent infection pathways*

Ricin B-lectin has been suggested in *N. bombycis* to play a role in infection, showing that inhibition of Ricin B-lectins lowers infection rates (H. Liu *et al*., 2016). With putative Ricin B- (like) lectins present throughout the Enterocytozoonidae (Table 5.2, Figure 5.5a), it may have a similar role in infection in the clade. Nine orthogroups were estimated to have Ricin B- (like) lectin domains being unique to the Enterocytozoonidae (not forming orthogroups with any Ricin B-like lectins in any non-enterocytozoonid species and not having any similar BLAST results). However, BLASTP results for one orthogroup (OG9557) showed genetic similarity to a Ricin B-Lectin identified in *N. bombycis* (Table 5.2). This suggests, at least in the case of this orthogroup, that these proteins may be homologous and that it is likely that the enterocytozoonid Ricin B-Lectin is genetically divergent. It is possible, though, that the remaining eight Ricin B- (like) lectins found in this study are Enterocytozoonidae lineage-specific and gained as a way to infect their respective hosts. The presence of Ricin B-lectins in the Enterocytozoonidae, *Nosema*, and *Encephalitozoon* (Brosson *et al*., 2005, 2006) also adds support for the suggestion that these may be ancestrally retained and were present in an ancestor (Campbell *et al*., 2013).

Analysis also identified eight orthogroups predicted to be localised extracellularly, that were gained at/throughout the Enterocytozoonidae lineage (Table 5.2 Figure 5.5b). As with the Ricin B-Lectin orthogroups, three have BLASTP results that show similarity to genes that have been previously described (M2K4, EnP1, and PTP3). This suggests that similar to the Ricin B- (Like) proteins, some of these

‘gained’ orthogroups may represent homologous and retained but highly divergent proteins, so have appeared as ‘novel’ in the analysis as they have not been grouped in their respect ancestral orthogroups. However, again, with no BLASTP results, the remaining extracellular proteins could be Enterocytozoonidae lineage-specific. EnP1 has been found to play a major role in adhesion to host cells for infection (Southern *et al*., 2007), and PTP3 is one of a small number of structural proteins that make up the polar filament (Peuvel *et al*., 2002). This provides support for the inference that remaining extracellular genes within the gained orthogroups may also play a role in infection. As with such a genetically conserved (in terms of ‘superfluous’ genomic regions) phylum, it is unlikely that genes gained and retained are vital to the parasite’s survival. Homologous representatives for neither EnP1 nor PTP3 have been found *in vitro* for members of the Enterocytozoonidae and only PTP3 has been identified in *E. hepatopenaei in silico* through genome annotation (Wiredu Boakye *et al*., 2017).

It is possible that if both of these proteins are involved in infection, the large divergence found from these homologous representatives may also be linked to a change in host type. This is also shown in the orthogroup for non-enterocytozoonid EnP1 (orthogroup ID: OG4002), which has representatives from taxa surrounding the Enterocytozoonidae: from *Ordospora*, *Encephalitozoon*, *Vittaforma,* and *Nosema*. Likewise, PTP3 (orthogroup ID: OG2282) has an even wider representation of species within the PTP3 orthogroup. The lack of enterocytozoonid EnP1 and PTP3 genes in the orthogroups of the wider range of species for their respective homologous representatives of these genes is likely because of such a large genetic divergence. Given the position of *E. bieneusi* within the Enterocytozoonidae, it is believed that it has secondarily shifted back to infecting terrestrial vertebrates, as it is the most parsimonious theory. In this regard, *E. bieneusi* may present another interesting host-shift from what would have likely been aquatic hosts to a plethora of terrestrial hosts. However, there is only one orthogroup (OG9514: uncharacterised and no BLASTP results) that is unique (Table 5.2, Figure 5.5b) to *E. bieneusi* that is predicted to be extracellularly localised. As OG9514 is the only extracellular identified orthogroup not possessed by the remaining described Enterocytozoonidae, all of which are aquatic-based, it is possible that the gene could be involved in the host-shift from aquatic hosts back to terrestrial

vertebrates. Possibly involved with interacting with different cell surfaces. However, further research on this candidate gene would need to be carried out *in vitro*, likely in the form of a gene knockout. Although this is currently not feasible, as *E. bieneusi* has not been successfully cultured and Microsporidia are lacking a protocol for genome modification.

Another domain found that may be involved in infection for *E. hepatopenaei* are putative hemolysins. Hemolysins have been hypothesised to be a mechanism of host cell lysis in Microsporidia (B. A. P. Williams, 2009), facilitating the rupture of cells to allow for the release of mature spores (Leonard, 2013). It has also been described in *O. colligata* (Pombert *et al*., 2015) and in *Nosema* species (Chetia *et al*., 2017). In terms of orthogroups gained that point at a shift towards crustacean/aquatic-based hosts, the hemolysin orthogroup gained within this lineage appears to be *E. hepatopenaei* specific. This suggests that if hemolysin proteins are involved in infection dynamics, that the hemolysin gained in this lineage was not a result of the clade's initial host type-shift.

With the Enterocytozoonidae monophyletically grouped within the Terresporidia (one of the five microsporidian branches, largely made up of parasites that infect terrestrial hosts), it would be expected that among the proteins gained at the node representing the ancestral enterocytozoonid, would be those that indicated at a major host shift from mammals to aquatic invertebrates. None of the orthogroups of interest (extracellular, Ricin B- (like) Lectins) were present in all enterocytozoonids at the base enterocytozoonid node (Figure 5.4a, b), indicating that the ancestral enterocytozoonid did not possess all the genes identified in this study. Suggesting that the host change is largely opportunistic in nature (not 'specialised') and that more specialised species-specific Ricin B- (like) Lectins, hemolysin, and extracellular genes were gained subsequently. An inference that concurs with the opportunistic nature of the clades infection route, as all the enterocytozoonids included in this study have only been noted infecting epithelial cells in the digestive tract (G. D. Stentiford *et al*., 2019), infecting the cells they first come into contact with. However, this pattern could also be indicative of fast-evolving infection-based genes that are present across the phylum. As such, these genes would differ greatly between lineages not infecting similar hosts and would likely result in a lack of conserved infection-based genes, as seen in enterocytozoonids.

5.4.2 *Evolution of the Enterocytozoonidae: ancestral genome reduction and current genome expansion*

Given how divergent enterocytozoonid species are from the 'ancestral' microsporidian node (Figure 5.2, Table 5.3-supplementary) (node after *Amphiamblys* sp. (Figure 5.1)) as illustrated by branch lengths, it would be expected that under neutral evolution, they would have gained the most orthogroups of the species included in this study. This however does not seem to be the case, with some of the lowest figures orthogroups gained (47.32-56.86), relative to their nearest, short branched relative (*Vittaforma corneae*) having 70.43 (Figure 5.2, Table 5.3-supplementary). The only species in this study with fewer orthogroups gained per 0.1 substitutions are *S. lophii* and *Ed. aedis* (40.53 and 43.54 respectively). However, this assumes similar mutation rates across Microsporidia, and the difference in substitution rates across Microsporidia has not been investigated. This suggests that in the evolution of the enterocytozoonid lineage, despite being the most genetically divergent on this tree, they were more conserved in terms of gene duplications. This possibly reflects differences in the repertoire and efficiency of DNA mismatch repair mechanisms (see below). This is also shown by a relatively low gene copy count, among enterocytozoonids (average Enterocytozoonidae: 0.08; *Nosema*: 0.11; *Encephalitozoon*: 0.9; *Nematocida*: 0.26) (Figure 5.2, Table 5.3-supplementary). It is likely that the factors involved in the gain and loss of genes are lineage-related, as closely related species have similar relative gains/losses of orthogroups. *Ne. displodere* and *N. bombycis* being the only two outliers from their respective clades, where both (*No. displodere*: 0.04; *N. bombycis*: 0.23) estimated gene duplication rates are more than double (*Nematocida* range: 0-0.01; *Nosema* range: 0.05-0.12) the rates in their respective clades.

Despite evidence above for the history of genome reduction in this family, with historically low duplication rates evidenced by relatively low gene copy numbers (Figure 5.2, Table 5.3-supplementary), the lineage may be undergoing a genome expansion. This is shown with the relatively high rate of duplications in the members within the family being double that of other microsporidians in this study (enterocytozoonid average: 0.12; average of remaining Microsporidia: 0.06), especially in *E. bieneusi* (0.25). A similar trend has also been described in plants,

the hypothesis being that functional changes to proteins are more likely to happen with recent duplications (Costello *et al*., 2020). This suggests that the gain and loss of protein families is higher after a recent duplication.

5.4.3 *Reduction in genome size through the loss of transposable elements and conservation*

The loss of 23 (retro)transposable elements throughout the Enterocytozoonidae follows the trends in genome reduction found in other microsporidian species (De Albuquerque *et al*., 2020; B. A. P. Williams *et al*., 2008). However, there is still evidence of active (retro)transposable elements within the Enterocytozoonidae. For example, the integrases, retrotransposon gag, and reverse transcriptases domains gained in *En. canceri* and *H. eriocheir* lineages (Table 5.4-supplementary). Reverse transcriptases have also been found in *S. lophii* by Hinkle *et al*. (1997) being associated with retrotransposons (Hinkle *et al*., 1997). Since Hinkle's paper, retrotransposons have also been found in *Anncaliia algerae* (previously *Brachiola algerae*), *Edhazardia aedis* (Gill *et al*., 2008; B. A. P. Williams *et al*., 2008), *Nosema bombycis* (Xu *et al*., 2006), and *V. corneae* (Mittleider *et al*., 2002). The presence of (retro)transposons in the Enterocytozoonidae also concurs with the hypothesis posed by Gill and Fast,2008; Williams *et al*, 2008, that the genome of *Enc. cuniculi* has lost these elements, as surrounding sister lineages (now including the Enterocytozoonidae) have it present. What is surprising is that it was supposed that *Enc. cuniculi* had lost it due to compaction of the genome (Gill *et al*., 2008) (*Encephalitozoon* having the smallest recorded genomes), however, enterocytozoonid genomes are also some of the most compact (though not as compact as Encephalitozoon). Suggesting that the remaining (retro)transposons may have been lost in the slight further compaction the *Encephalitozoon* have gone through and that more may have been lost in the course of enterocytozoonid evolution. Adding to this inference is that, of the Enterocytozoonidae, only *H. eriocheir* (having the largest genome in the family) has gained retrotransposon gag.

5.4.4 *MMR DNA repair*

It has been noted that during the process of compaction of the Microsporidian genome, DNA repair genes (Gill & Fast, 2007; Haag *et al*., 2014; P. J. Keeling & Slamovits, 2004a) and tumour suppressing genes (Haag *et al*., 2014) have been lost and this loss has been suggested as the reason for the hypothesised high mutation rates in the phylum. However, the node including *E. hepatopenaei* and *En. canceri* has been predicted to have gained DNA mismatch repair proteins (MMR), noted as partially missing *Enc. cuniculi*. This suggests that in this lineage, there may be a selection to retain more genes involved in DNA repair in these two species, at least DNA mismatch repair. Which, as stated above, could explain the low relative orthogroups gained.

5.4.5 *Selective pressure on the addition and loss of genes*

There is no correlation between the number of orthogroups gained and genetic divergence (branch length) (Figure 5.1a, c-supplementary), suggesting a role for adaptation in the retention of new orthogroups rather than being the result of a clock-like accumulation of new copies. It seems intuitive that for a group of organisms where there is potentially a strong selective pressure driving genomic reduction, the gaining of new proteins would have to have a positive effect on fitness to be retained/fixed. However, there seems to be no such effect on loss of orthogroups (Figure 5.1b, d-supplementary), suggesting that the loss of orthogroups is less affected by selective pressures in most cases, apart from the *Nematocida* outliers.

5.4.6 *Enterocytozoonid evolution relationships*

As would be expected, orthogroups are more frequently shared by species that had the closest shared ancestor (Table 5.2, Figure 5.5a, b). Both Ricin B-Lectins and predicted extracellular orthogroups support a closer split for *En. canceri* and *E. hepatopenaei* to the exclusion of *E. bieneusi*. However, *H. eriocheir* shows a very different evolutionary history, in terms of the proteins focused on in this study, with no shared gaining of proposed infection-based orthogroups. Supporting suggestions that *H. eriocheir* not be included within the Enterocytozoondiae (Bojko *et al*., 2017).

5.4.7 *Summary*

In conclusion, the addition of at least eight putative (Ricin B-lectins) genes associated with infection throughout the Enterocytozoonidae could show species-specific genes used for infection in new host types. This is similarly the case with the addition of at least five out of eight extracellular localised genes, two (EnP1 and PTP3) of which are possibly homologous to genes that have also been associated, *in vitro*, with infection. However, under the search criteria used in this study, the majority of the infection-based orthogroups were not retained by all enterocytozoonids and therefore presumed not to be present in the ancestral enterocytozoonid lineage. So, it is possible that the initial host shift to crustacea was not accompanied by an initial expansion/evolution of gene families specialised for pathogenicity in crustaceans. This is also a trait seen in the bacterial opportunistic pathogen *Pseudomonas aeruginosa*, that the gain and loss of genes were not required for virulence in this opportunistic pathogen (Martínez, 2014); a species that is capable of infecting different kingdoms of life (Carilla-Latorre *et al*., 2008; Mahajan-Miklos *et al*., 2000; Navas *et al*., 2007). This suggests that it is possible that the virulence-based genes that allowed the Enterocytozoonidae to infect phylogenetically divergent hosts evolved earlier in Microsporidian evolution (Martínez, 2014).

This study has generated a robust phylogenetic framework with which to investigate patterns of loss and gain of genes and gene families in the microsporidia., Here it has been used to investigate general trends of loss and gain of orthogroups within the Enterocytozoonidae. Whilst this has highlighted several orthogroups of interest further research to verify the validity of the candidates highlighted in this study could both improve understanding of the evolution of this unique phylum and help identify pathogenic genes that could be used to predict pathogenicity within species. This would also allow for targeted preventative drugs, and shed further light on how some Microsporidia are able to host-shift. As is done widely with better known pathogenic species.

5.4.8 Further research and limitations

Due to a lack of sequencing data, this study does not include any of the fish infecting species ((*Para)Nucleospora*). Interestingly, there is also one fish infecting species, *Enterospora nucleophila*, that is found with the crustacean infecting (Enterospora/Enterocytozoon) clade. Next steps should involve sequencing these species to investigate if the (Para)Nucleospora lineage has genes that are unique to infecting fish, and if so, are they orthologous to those possessed by *En. nucleophila*.

As with many forms of in silico research (especially when involving the prediction of gene function), analysis carried out in this study will need to be verified in vitro for further research. However, as mentioned, this is not currently feasible with Microsporidia. Leaving further in silico work to verify findings made in this chapter, such as transcriptomics, which would allow for the identification of genes that are upregulated during infection.

As theorised and observed in Chapter 2 and partially, Chapter 3, the Enterocytozoonidae are very likely to be more diverse than the ten species currently described for the family. As a result, there are limitations on the inferences that can be made about the patterns observed in this study, based on the small number of enterocytozoonids sequenced. For example, if the predicted extracellular gene unique to *E. bieneusi* is identified in novel aquatic enterocytozoonid species, it would be unlikely to be involved in the host-shift to terrestrial hosts.

5.5 Supplementary tables and figures:

Table 5.1: Reference genomes used in OrthoMCL and species tree building

Species	GenBank Accession
Amphiamblys sp.	GCA_001875675.1
Anncaliia algerae	GCA_000385875.2
Aspergillus fumigatus+	GCF_000002655.1
Cryptococcus neoformans+	GCF_000149245.1
Enterocytozoon bieneusi	GCA_000209485.1
Enterocytozoon hepatopenaei	GCA_002081675.1
Enterospora canceri	GCA_002087915.1
Edhazardia aedis	GCA_000230595.3
Encephalitozoon cuniculi	GCA_000091225.2
Encephalitozoon hellem	GCA_000277815.3
Encephalitozoon intestinalis	GCA_000146465.1
Encephalitozoon romaleae	GCA_000280035.2
Hepatospora eriocheir	GCA_002087885.1
Magnaporthe oryzae+	GCA_000002495.2
Mitosporidium daphniae	GCA_000760515.2
Nosema apis	GCA_000447185.1
Nosema bombycis	GCA_000383075.1
Nosema ceranae	GCF_000988165.1
Nematocida displodere	GCA_001642395.1
***Nematocida* sp. 1 ERTm2**	GCA_000250695.1
***Nematocida* sp. ERTm5**	GCA_001642415.1
***Nematocida* sp. 1 ERTm6**	GCA_000738915.1
***Nematocida parisii* ERTm1**	GCA_000250985.1
***Nematocida parisii* ERTm3**	GCA_000190615.1
Ordospora colligata	GCA_000803265.1
Paramicrosporidiumsaccamoebae	GCA_002794465.1
Paraphelidium tribonemae	N/A
Pseudolomaneurophilia	GCA_001432165.1
Rozella allomycis	GCA_000442015.1
Saccharomyces cerevisiae+	GCF_000146045.2
Spraguealophii	GCA_001887945.1
Trachipleistophora hominis	GCA_000316135.1
Vavraiaculicis subsp.	GCA_000192795.1
Vittaforma corneae	GCA_000231115.1

*+denotes fungal outgroups

Table 5.2

Orthogroups (arbitrary ID), gene ID (inferred from homologous Saccharomyces cerevisiae genes),and models of best fit

Orthogroup ID	Gene	Best Model
0G2_1109	Deoxyhypusine synthase	LG+G4
0G2_1120	Ribosomal 60S subunit protein L3	LG+I+G4
0G2_1147	Glycine--tRNA ligase	LG+I+G4
0G2_1150	Ribosomal 60S subunit protein L26B	LG+R3
0G2_1180	Serine/threonine-protein kinase KIN2	LG+F+I+G4
0G2_1183	Putative aminophospholipid-translocating P4-type ATPase NEO1	LG+F+I+G4
0G2_1187	DNA primase subunit PRI1	LG+G4
0G2_1189	Proteasome regulatory particle base subunit RPT3	LG+I+G4
0G2_1191	Gamma-tubulin	LG+I+G4
0G2_1198	Arginine--tRNA ligase MSR1	LG+F+I+G4
0G2_1200	Phenylalanine--tRNA ligase subunit beta	LG+F+I+G4
0G2_1202	Hsp90 family chaperone HSC82	LG+F+I+G4
0G2_1216	Proliferating cell nuclear antigen	LG+F+R3
0G2_1218	S-adenosylmethionine-dependent methyltransferase	LG+I+G4
0G2_1219	Xdj1p	LG+I+G4
0G2_1223	Transcription factor TFIIIB subunit BDP1	LG+G4
0G2_1224	Ribosomal 60S subunit protein L12A	LG+G4
0G2_1231	ATPase-activating ribosome biosynthesis protein	LG+G4
0G2_1241	Transcription factor TFIIIB subunit BRF1	LG+G4
0G2_1246	Hap5p	LG+G4
0G2_1256	Serine/threonine-protein kinase CDC7	LG+G4
0G2_1257	Ribosomal 40S subunit protein S4B	LG+I+G4
0G2_1263	Utp7p	LG+I+G4
0G2_1267	DNA-directed DNA polymerase alpha subunit POL12	LG+F+I+G4
0G2_1269	Arc1p	LG+I+G4
0G2_1276	CCR4-NOT core exoribonuclease subunit CCR4	LG+I+G4
0G2_1286	Ribosomal 60S subunit protein L16A	LG+G4
0G2_1287	Ribosomal 60S subunit protein L20A	LG+I+G4
0G2_1288	Gdi1p	LG+I+G4

0G2_1289	Cyclin-dependent serine/threonine-protein kinase CDC28	LG+I+G4
0G2_1294	NuA4 histone acetyltransferase complex catalytic subunit ESA1	LG+I+G4
0G2_1296	Ribosomal 60S subunit protein L15A	LG+G4
0G2_1298	TATA-binding protein-associated factor TAF6	LG+I+G4
0G2_1299	tRNA (guanine) methyltransferase	LG+I+G4
0G2_1300	DNA- (apurinic or apyrimidinic site) lyase APN1	LG+I+G4
0G2_1309	3'-5'-exodeoxyribonuclease	LG+G4
0G2_1313	TPA: H (+)-transporting V0 sector ATPase subunit d	LG+G4
0G2_1317	AAA family ATPase SEC18	LG+F+I+G4
0G2_1318	Cleavage polyadenylation factor subunit YSH1	LG+I+G4
0G2_1319	Septin SPR3	LG+G4
0G2_1323	14-3-3 family protein BMH1	LG+I+G4
0G2_1324	Chaperonin-containing T-complex subunit CCT7	LG+F+I+G4
0G2_1327	Signal recognition particle receptor subunit alpha	LG+F+I+G4
0G2_1328	TFIIH/NER complex ATPase/helicase subunit SSL2	LG+I+G4
0G2_1332	Ribosomal 40S subunit protein S2	LG+I+G4
0G2_1335	Phosphoacetylglucosamine mutase PCM1	LG+G4
0G2_1336	DNA ligase (ATP) CDC9	LG+I+G4
0G2_1337	Ribosomal 40S subunit protein S16A	LG+R3
0G2_1338	Ribosomal 60S subunit protein L13A	LG+I+G4
0G2_1341	Alpha-tubulin TUB1	LG+I+G4
0G2_1344	rRNA-processing protein MPP10	LG+I+G4
0G2_1345	Ribosomal 60S subunit protein L8A	LG+G4
0G2_1346	tRNA (guanine26-N2)-dimethyltransferase	LG+F+I+G4
0G2_1349	Diphthine synthase	LG+I+G4
0G2_1352	Proteasome core particle subunit beta 3	LG+I+G4
0G2_1353	Methionine aminopeptidase	LG+G4
0G2_1354	Aurora kinase	LG+I+G4
0G2_1355	Chaperonin-containing T-complex alpha subunit TCP1	LG+I+G4
0G2_1356	Ribosomal 40S subunit protein S0A	LG+I+G4
0G2_1358	Asparagine--tRNA ligase DED81	LG+I+G4

0G2_1364	Ribosomal 60S subunit protein L1A	LG+G4
0G2_1368	Chaperonin-containing T-complex subunit CCT4	LG+I+G4
0G2_1376	Proteasome core particle subunit alpha 1	LG+I+G4
0G2_1377	Putative AAA family ATPase RIX7	LG+R4
0G2_1378	Putative dimethyladenosine transferase	LG+G4
0G2_1380	Phosphomannomutase SEC53	LG+I+G4
0G2_1381	Pseudouridine synthase CBF5	LG+I+G4
0G2_1382	Bifunctional dITP/dUTP diphosphatase	LG+I+G4
0G2_1384	Cysteine desulfurase	LG+G4
0G2_1386	Calmodulin	LG+I+G4
0G2_1387	1-phosphatidylinositol 4-kinase	LG+F+G4
0G2_1391	Condensin subunit SMC2	LG+I+G4
0G2_1393	Arf family GTPase SAR1	LG+I+G4
0G2_1395	Polo kinase CDC5	LG+F+I+G4
0G2_1397	Chaperonin-containing T-complex subunit CCT3	LG+G4
0G2_1399	Ribosomal 60S subunit protein L18A	LG+R4
0G2_1400	TATA-binding protein	LG+I+G4
0G2_1401	H (+)-transporting V1 sector ATPase subunit A	LG+F+I+G4
0G2_1405	Serine/threonine-protein kinase HRR25	LG+G4
0G2_1408	Translation initiation factor eIF2 subunit alpha	LG+R3
0G2_1409	Nuclear protein localization protein 4	LG+G4
0G2_1416	Ribosomal 60S subunit protein L5	LG+G4
0G2_1418	Recombinase RAD51	LG+I+G4
0G2_1419	Proteasome regulatory particle base subunit RPN10	LG+I+G4
0G2_1422	tRNA adenylyltransferase	LG+F+R5
0G2_1423	Mismatch repair ATPase MSH2	LG+F+I+G4
0G2_1424	Proteasome regulatory particle lid subunit RPN7	LG+I+G4
0G2_1425	Palmitoyltransferase YKT6	LG+G4
0G2_1426	Rab family GTPase YPT31	LG+F+G4
0G2_1428	Syntaxin-binding protein	LG+F+I+G4
0G2_1430	Lhp1p	LG+I+G4
0G2_1431	rRNA (cytosine-C5-)-methyltransferase NOP2	LG+F+I+G4
0G2_1433	Proteasome regulatory particle lid subunit RPN3	LG+I+G4

0G2_1438	Iron-sulfur cluster assembly protein CIA2	LG+G4
0G2_1442	DNA-directed RNA polymerase III subunit C34	LG+I+G4
0G2_1443	Transcription elongation factor DST1	LG+G4
0G2_1444	Ribosomal 40S subunit protein S11A	LG+F+I+G4
0G2_1448	Glutamate--tRNA ligase GUS1	LG+I+G4
0G2_1449	DNA primase subunit PRI2	LG+F+G4
0G2_1451	TFIIH/NER complex ATP-dependent 5'-3' DNA helicase subunit RAD3	LG+I+G4
0G2_1454	Lysine--tRNA ligase KRS1	LG+F+G4
0G2_1457	Translation elongation factor EF-1 alpha	LG+G4
0G2_1463	Ribosomal 40S subunit protein S22A	LG+I+G4
0G2_1467	Proteasome core particle subunit beta 4	LG+I+G4
0G2_1468	Chromatin-remodeling protein SPT16	LG+I+G4
0G2_1471	Histone acetyltransferase GCN5	LG+I+G4
0G2_1472	2- (3-amino-3-carboxypropyl)histidine synthase	LG+F+I+G4
0G2_1479	Coatomer subunit alpha	LG+I+G4
0G2_1486	tRNA (guanine46-N7)-methyltransferase	LG+I+G4
0G2_1488	TATA-binding protein-associated factor TAF11	LG+R3
0G2_1490	Ribosome biosynthesis protein NIP7	LG+I+G4
0G2_1491	Translation initiation factor eIF3 subunit i	LG+I+G4
0G2_1516	Ribosomal 40S subunit protein S3	LG+I+G4
0G2_1519	rRNA methyltransferase NOP1	LG+G4
0G2_1528	MCM DNA helicase complex subunit MCM5	rtREV+F+I+G4
0G2_1531	snoRNA-binding rRNA-processing protein IMP4	LG+I+G4
0G2_1532	RNA-processing protein NOP58	LG+I+G4
0G2_1542	Translation initiation factor eIF2 subunit beta	LG+G4
0G2_1544	DNA-directed RNA polymerase II core subunit RPB3	LG+I+G4
0G2_1549	Ribosomal 40S subunit protein S5	LG+I+G4
0G2_1556	Alanine--tRNA ligase	LG+I+G4
0G2_1560	Peptide alpha-N-acetyltransferase complex B subunit NAT3	LG+F+I+G4
0G2_1587	Guanine nucleotide exchange factor SDO1	LG+I+G4
0G2_1595	Kar3p	LG+G4
0G2_1596	Transcription factor TFIIE subunit TFA1	LG+I+G4

Table 5.3 Number of accumulative orthogroups gained for each microsporidian species included in this tree, relative to the accumulative branch length

Species	**Branch length**	**Gained orthos**	**Orthos lost**	**Relative orthos gained**	**Relative orthos lost**	**Relative duplication**	**Relative gene copy**	**Genome size**
Nosema apis	0.84	514	2497	61.19	297.26	0.12	0.09	8.57
Nosema ceranae	0.8	482	2481	60.33	310.51	0.05	0.07	5.69
Nosema bombycis	0.88	697	2279	79.38	259.57	0.23	0.16	15.7
Encephalitozoon hellem	0.74	574	2194	77.67	296.89	0.01	0.09	2.25
Encephalitozoon romaleae	0.74	573	2250	77.43	304.05	0.00	0.08	2.19
Encephalitozoon intestinalis	0.78	570	2213	72.98	283.35	0.00	0.09	2.22
Encephalitozoon cuniculi	0.72	573	2199	79.47	304.99	0.03	0.09	2.5
Ordospora colligata	0.75	488	2175	65.07	290	0.00	0.08	2.3
Vittaforma corneae	0.83	586	2219	70.43	266.71	0.04	0.08	3.21
Enterocytozoon bieneusi	1.07	604	2674	56.45	249.91	0.25	0.11	3.86
Enterospora canceri	1.15	655	2802	56.86	243.23	0.05	0.06	3.1
Enterocytozoon hepatopenaei	1.17	629	2637	53.85	225.77	0.05	0.07	3.25
Hepatospora eriocheir	1.16	548	2644	47.32	228.32	0.14	0.08	4.57
Edhazardia aedis	0.86	374	2235	43.54	260.19	0.16	0.10	51.31
Spaguea lophii	0.79	321	2313	40.53	292.05	0.09	0.08	5.76
Anncaliia algerae	0.8	687	2342	85.88	292.75	0.28	0.16	12.16
Vavraia culicis	0.89	911	2393	101.9	267.67	0.02	0.10	6.12
Trachipleistophora hominis	0.9	949	2398	105.8	267.34	0.02	0.10	8.49
Pseudoloma neurophilia	1.03	607	2412	59.16	235.09	0.12	0.10	5.25
Nematocida displodere	0.69	649	818	94.61	119.24	0.04	0.24	3.1

Nematocida sp. ERTm2	0.72	1196	893	167.04	124.72	0.01	0.28	4.7
Nematocida sp. ERTm6	0.71	1182	931	166.01	130.76	0.00	0.25	4.28
Nematocida sp. ERTm5	0.71	1274	869	180.71	123.26	0.01	0.29	4.39
Nematocida parisii ERTm1	0.71	1342	1062	189.82	150.21	0.00	0.23	4.1
Nematocida parisii ERTm3	0.71	1346	1001	189.58	140.99	0.00	0.27	4.15

Table 5.4

Interpro characterisation of enterocytozoonid orthogroups

*Colour key: Black= domains/motifs unique to that species/node; colours outline domains/motifs between species and nodes

E. bieneusi	*En. canceri*	*E. hepatopenaei*	*H. eriocheir*	*E. hepatopenaei* *En. canceri*	*E. hepatopenaei* *En. canceri* *E. bieneusi*	*E. hepatopenaei* *En. canceri* *E. bieneusi* *H. eriocheir*
-Armadillo-type fold	-ATPase, AAA-type, core	-Bacterial hemolysins superfamily	-Alpha/Beta hydrolase fold	-AAA+ ATPase domain	-ADP/ATP carrier protein	-AH/BAR domain superfamily
-Beta tubulin	-Cation efflux protein	-Coils	-Armadillo-type fold	-ABC transporter type 1, transmembrane domain superfamily	-Amino acid/polyamine transporter I	-Alpha/Beta hydrolase fold
-Beta tubulin, autoregulation binding site	-Cation efflux transmembrane domain -superfamily	-Snare region anchored in the vesicle membrane C-terminus	-Aspartic peptidase domain superfamily	-Arrestin, C-terminal	-Aquaporin transporter	-Amino acid transporter, transmembrane domain
-Chitin synthase	-Coils	-Malarial early transcribed membrane protein (ETRAMP)	-Coils	-Coils	-Armadillo-type fold	-Aminoacyl-tRNA synthetase, class Ia, anticodon-binding
-Coils	-Endonuclease/ exonuclease/ phosphatase	-Prokaryotic membrane lipoprotein lipid attachment site profile	-Enolase	-DNA mismatch repair protein MutS, core domain	-ATPase, AAA-type, core	-Coils
-Nuclear pore complex protein NUP98-NUP96	-Endonuclease/ exonuclease/ phosphatase superfamily	- Ribonuclease H like	-Enolase, C-terminal TIM barrel domain	-DNA mismatch repair protein MutS, C-terminal	-Chaperone J-domain superfamily	-E3 ubiquitin ligase Bre1
-Nucleotide-diphospho-sugar transferases	-Integrase zinc-binding domain		-Glycosyl hydrolase family 32, N-terminal	-Histone H2A/H2B/H3	-Coils	-LIS1 homology motif
-Peptidase C19, ubiquitin carboxyl-terminal hydrolase	-Integrase, catalytic core		-Integrase, catalytic core	-Histone H3/CENP-A	-Cullin homology domain superfamily	-Methionyl/Leucyl tRNA synthetase
-P-loop containing nucleoside			-MCM domain	-Histone-fold	-DnaJ domain	
					-Dopey, N-terminal	
					-E3 ubiquitin-protein ligase listerin	

triphosphate hydrolase -Ribonucleotide reductase small subunit family -Ribonucleotide reductase small subunit, active site -Ribonucleotide reductase small subunit -Ribosomal protein L22/L17 -Ribosomal protein S3Ae -Ribosomal protein L22/L17, eukaryotic/archaeal -Ribosomal protein L22/L17 superfamily -Signal transduction response regulator, receiver domain -Signal transduction	-LSM domain superfamily -Mechanosensitive ion channel MscS -Mechanosensitive ion channel MscS domain superfamily - Phosphoribulokinase/uridine kinase -P-loop containing nucleoside triphosphate hydrolase -Protein kinase domain -Reverse transcriptase domain -Ribosomal protein L10e -Ribosomal protein L10e/L16 -Ribosomal protein L10e/L16 superfamily -SCAN domain -SCAN domain superfamily	-Membrane bound O-acyl transferase, MBOAT -Methionyl/Leucyl tRNA synthetase -Mini-chromosome maintenance protein -Myb domain -Nop domain superfamily -Nucleic acid-binding, OB-fold - Phosphoribulokinase/uridine kinase -P-loop containing nucleoside triphosphate hydrolase -Pseudouridine synthase, catalytic domain superfamily -PUA domain -PUA domain superfamily	-Immunoglobulin E-set -MFS transporter superfamily -OTU domain -Peptidase C65, otubain, subdomain 1 -Peptidase C65, otubain, subdomain 2 - Phosphatidylinositol N-acetylglucosaminyltransferase subunit C -P-loop containing nucleoside triphosphate hydrolase -THUMP domain -Type I protein exporter -WD40-repeat-containing domainsuperfamily - ABC transporter-like	-EF-hand domain -ER lumen protein retaining receptor -GIT, Spa2 homology (SHD) domain -Homeobox domain -Leucine-rich repeat -LSM domain superfamily -Major intrinsic protein -Major intrinsic protein, conserved site -Matrin/U1-C, C2H2-type zinc finger -Mechanosensitive ion channel MscS -Mechanosensitive ion channel MscS domain superfamily -Mediator complex subunit 15, KIX domain -MFS transporter superfamily	-Nucleic acid-binding, OB-fold -Peptidase C50, separase -Protein kinase domain -Protein kinase, ATP binding site -SEPARIN core domain -Serine/threonine-protein kinase, active site -Spc7 kinetochore protein domain -SUN domain -Tetratricopeptide repeat-containing domain -Tetratricopeptide repeat -Thioredoxin domain -Thioredoxin, conserved site -TIP49, P-loop domain -Zinc finger, RING-type

response regulator, C-terminal effector -Transcription regulator LuxR, C-terminal -Transcription factor, GTP-binding domain -Tubulin -Tubulin/FtsZ, GTPase domain -Tubulin/FtsZ, GTPase domain superfamily -Ubiquitin specific protease domain -WD40-repeat-containing domain superfamily -CheY-like superfamily -Ferritin-like superfamily -Galactose-binding-like domain superfamily	-Serine/threonine-protein kinase, active site -Ulp1 protease family, C-terminal catalytic domain - ERAP1-like C-terminal domain -Papain-like cysteine peptidase superfamily -Protein kinase-like domain superfamily -Ribonuclease H-like superfamily -Ricin B-like lectins -SNase-like, OB-fold superfamily	-Retrotransposon gag domain -Reverse transcriptase domain -Ribosomal biogenesis NSA2 family -Ricin B, lectin domain -SANT/Myb domain -Serine/threonine-specific protein phosphatase/bis (5-nucleosyl)-tetraphosphatase -SRP-independent targeting protein 2/TMEM208 -Tetratricopeptide repeat-containing domain -Tetratricopeptide repeat -Thymidylate synthase/dCMPhydroxymethylase superfamily	-FAR-17a/AIG1-like protein -Arrestin-like, N-terminal -ISXO2-like transposase domain -Ricin B-like lectins -Rnp2-like domain superfamily -Papain-like cysteine peptidase superfamily	-Nucleotide-binding alpha-beta plait domain superfamily -Palmitoyltransferase, DHHC domain -P-loop containing nucleoside triphosphate hydrolase -PNPase/RNase PH domain superfamily -Protein dopey -Recombination protein RecR -Ribonuclease P/MRP, subunit p29 -Ribonuclease P/MRP, subunit p29 superfamily -Ribosomal protein L10P -Ribosomal protein S5 domain 2-type fold -Ricin B, lectin domain -RNA recognition motif domain	-Zinc finger, RING-CH-type -Zinc finger, RING/FYVE/PHD-type -Zinc finger, RING-type, conserved site -Zinc/iron permease - FAR-17a/AIG1-like protein -Fungal lipase-like domain -Galactose-binding-like domain superfamily -Polymerase/histidinol phosphatase-like -Protein kinase-like domain superfamily -Quinoprotein alcohol dehydrogenase-like superfamily -Rossmann-like alpha/beta/alpha sandwich fold -RuvB-like

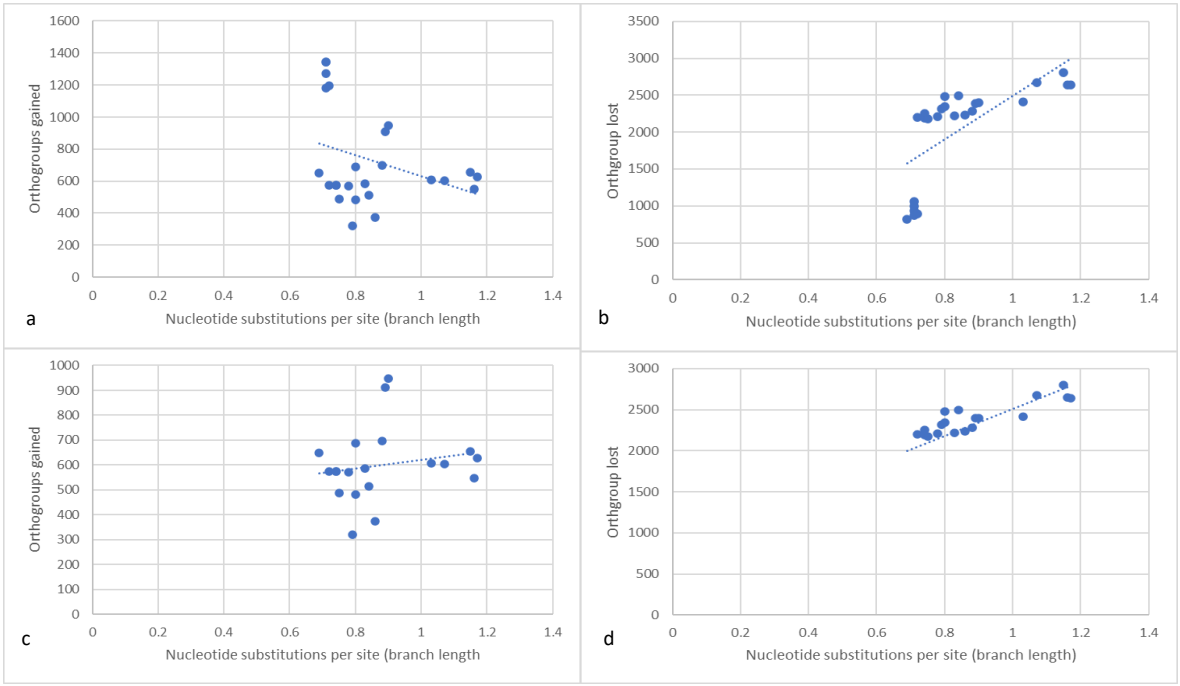
-Nucleoporin peptidase S59-like -Papain-like cysteine peptidase superfamily -Quinoprotein alcohol dehydrogenase-like superfamily -Ribonucleotide reductase-like	-Transcription Factor IIF, Rap30/Rap74, interaction -tRNA pseudouridine synthase B family -tRNA pseudouridylate synthase B, C-terminal -Type IIA DNA topoisomerase subunit A, alpha-helical domain superfamily -Zinc finger C2H2-type -ABC transporter-like -Armadillo-like helical -Calcineurin-like phosphoesterase domain, ApaH type -DNA topoisomerase, type IIA-like	-RNA-binding domain superfamily -SUI1 domain -SUI1 domain superfamily -Target SNARE coiled-coil homology domain -Tetratricopeptide repeat -Tetratricopeptide repeat-containing domain -Translocation protein Sec66 -Ubiquilin -Vacuolar (H+)-ATPase G subunit -WD40-repeat-containing domainsuperfamily -Zinc finger C2H2 superfamily -Zinc finger C2H2-type -Zinc finger, double-stranded RNA binding	-Tetratricopeptide-like helical domain superfamily -WD40/YVTN repeat-like-containing domain superfamily -RuvB-like helicase 2, domain II -Thioredoxin-like superfamil

	domain superfamily -Enolase-like, C-terminal domain superfamily -Homeobox-like domain superfamily -Metallo-dependent phosphatase-like -PUA-like superfamily -Reverse transcriptase/retrotransposon-derived protein, RNase H-like -Ribonuclease H-like superfamily -Ricin B-like lectins -Tc1-like transposase, DDE domain -Tetratricopeptide-like helical domain superfamily	-Zinc finger, RING/FYVE/PHD-type -Zinc finger, RING-CH-type -Zinc finger, RING-type - Rhodanese-like domain -ABC transporter-like -Major facilitator, sugar transporter-like -FAR-17a/AIG1-like protein -Rad21/Rec8-like protein, C-terminal, eukaryotic -Rad21/Rec8-like protein, N-terminal -Homeobox-like domain superfamily -UBA-like superfamily -Longin-like domain superfamily -Quinoprotein alcohol

	-Transposase InsF-like -Transposase, Tc1-like	dehydrogenase-like superfamily -Tetratricopeptide-like helical domain superfamily -Mg2+ transporter protein, CorA-like/Zinc transport protein ZntB -WD40/YVTN repeat-like-containing domain superfamily -Cullin repeat-like-containing domain superfamily -Aquaporin-like -Rof/RNase P-like te -Ubiquitin-like domain superfamily -Ricin B-like lectins -SKP1-like, dimerisation domain superfamily -Rhodanese-like domain superfamily -Cyclin-like superfamily

Figure 5.1a-d

Relationship between branch length (substitutions per nucleotide) and orthogroups gained (a) and lost (b). Figures c and d show relationships, respectively, with the *Nematocida* outliers, removed.



Chapter 6: General Discussion

Microsporidians have been shown to have a diverse set of transmission routes, and an array of mixtures of transmission modes (Haag *et al*., 2019; Quiles *et al*., 2019; Vu-Khac *et al*., 2018), as well as being widespread and ubiquitous ((Ardila-Garcia *et al*., 2013) Chapters 2 and 3). The relationships shared between enterocytozoonids and their hosts are not well known, but they are likely to be complex given their broad assemblage of described hosts (Fiuza *et al*., 2016; Foltz *et al*., 2009; Sakai *et al*., 2009; Tabatabaie *et al*., 2015), and ability to host-shift. It has been proposed that the presence of these pathogens likely poses a threat to modern agriculture and aquaculture via their ability to infect a range of immunocompromised and susceptible hosts when animals are reared under high density/high-stress conditions.

The aim of this thesis was to assess the risk posed to aquaculture, agriculture, and ultimately, human health by this clade (Enterocytozoonidae) of Microsporidia. This thesis covered this in three broad themes in four data chapters. The first theme addressed the investigation of undescribed enterocytozoonid diversity in the environment (Chapters 2 and 3), focused around areas under human influence, to better understand the distribution of uncharacterized pathogens with the potential to infect livestock. The second theme addressed this by examining the transmission routes and possible endemicity of the shrimp parasite, *Enterocytozoon hepatopenaei* (Chapter 4), to better understand how an enterocytozoonid has and may continue to enter and thrive in aquaculture. The third looked at trends in the evolution of the Enterocytozoonidae, focusing on the identification of potential gene families involved with host-shifting and virulence (Chapter 5), to better understand the changes that underlie the jump from terrestrial animals to marine crustacea.

6.1 Summary of chapters

To characterise undescribed diversity and distribution of the Enterocytozoonidae, this thesis involved surveying samples for enterocytozoonid-like sequences using metagenomes databases (Chapter 2) finding 92enterocytozoonid-like novel sequences. This research was continued in a more focused fashion in Chapter 3: by screening transect samples from estuaries in South West UK. This found a greater abundance and a wider distribution for *Enterospora canceri* than previous studies indicated, uncovering the first instance of freshwater enterocytozoonids and *Enterocytozoonidae hepatopenaei*-like sequences in the UK. Although Chapter 2 showed no evidence of enterocytozoonid-like sequences in freshwater metagenomes. The novel enterocytozoonid-like sequences found in Chapter 2 showed a skew towards the crustacean infecting clade (*Enterospora/Enterocytozoon*), with far less for the fish infecting clade, (*Para)Nucleospora/Obruspora* and sequencing of transect samples in Chapter 3 lacked sequences for the fish-infecting clade all together, with the majority of the diversity found around *En. canceri* and to a lesser extent, *E. hepatopenaei*. The results of Chapters 2 and 3 suggested both greater diversity and distribution for the Enterocytozoonidae than currently described. This work showed that though they can be found in freshwater, a far greater proportion of the diversity seems to be found in marine/estuarine environments. Which confirmed that the Enterocytozoonidae seem marine-based. The widespread prevalence of *En. canceri* suggests a more ubiquitous 'primary' host than adult crabs, likely finding widespread intermediate host (s) in planktonic crustacea.

In Chapter 4, genome-wide population genomics were employed to analyse the biogeography of *E. hepatopenaei*. Samples collected from shrimp ponds in China, India, and Thailand revealed geographically structured populations for the shrimp pathogen, with some evidence of migration between shrimp ponds in Thailand and India. Within country region-based population structure was also

observed within Thailand, the best sampled country in this study, with differentiation between samples less than 75 km (the shortest distance between Pathum Thani and Suphan Buri) away. Intrapopulation diversity was moderate, with an average observed heterozygosity of 20.4%. Greater estimated observed (Ho) than expected (He) suggested a recent bottleneck for *E. hepatopenaei* in the shrimp ponds samples in this study. It was concluded that it was likely that *E. hepatopenaei* had endemic populations, long-standing enough to have within-country structure; with what is likely to be repeated local transmission into ponds.

Analysis looking at the addition of gene families to investigate changes in genes associated with host-shifting and infection (Chapter 5) showed minimal addition of lineage-specific genes (2 out of 17 orthogroups) associated with extracellular activity (i.e host affecting). Orthogroups gained were largely undescribed, without any functional annotation: BLASTP analysis of the Ricin B-like Lectin families added within the Enterocytozoondiae showed no results, except a Ricin B-Lectin described in the microsporidian, *Nosema bombycis*. Results from BLASTP analysis on predicted extracellular orthogroups produced results for 3 (EnP1, PTP3, and M2K4) out of 8. With Ricin-B-lectin domains, EnP1 and PTP proteins repeatedly associated with cell binding and germination of spores, the Enterocytozoonid specific gene families with Ricin-B-lectin domains suggest these proteins may be associated with the host-shift of host type (terrestrial vertebrates to marine invertebrates and crustaceans). The lack of lineage-wide conserved genes suggests two evolutionary strategies: as the Enterocytozoonidae have only been observed infecting the epithelial cells of the gastrointestinal tract, that only a small number of ‘new’ extracellular genes are required to infect phylogenetically distant hosts; or infection-based genes are fast evolving, resulting in genetically divergent, homologous genes that appear in the analysis as ‘gained’ genes.

Genome reduction within the Enterocytozoonidae was also explored in Chapter 5, finding the clade had some of the lowest relative rates of orthogroup addition and gene copy counts compared to other clades included in this analysis. Suggesting a more conserved rate of evolution, in terms of gaining of new gene families, compared to the other taxa included in this study. Characterisation of domains of orthogroups lost showed the majority of the genes lost throughout the lineage were metabolic and information (DNA replication/repair, translation/transcription, etc) based, as well as a number of (retro)transposable elements. These results corroborate those of Wiredu Boakye *et al*., showing a loss of many of the genes involved in glycolysis as well as fatty acid metabolism (Wiredu Boakye *et al*., 2017). The analysis also showed relatively high rates of gene duplication, compared to other microsporidian lineages compared in this study, contrasting with the lower rate of orthogroup gain, as gene duplication has been strongly associated with gene family expansion. However, as gene copy counts are relatively low, suggesting historically low gene duplication in the lineage, this may show the enterocytozoonid genomes are expanding.

6.2 (*Para)Nucleospora/Obruspora* Less diverse clade?

As mentioned above, screening of both metagenome datasets (Chapter 2) and estuary samples (Chapter 3) found more novel diversity in crustacean-infecting *Enterospora/Enterocytozoon* clade than other groups within the Enterocytozoonidae. It was initially suggested that this could be due to filtering bias (Chapter 2, 3), as both types of dataset were obtained from filtering zooplankton, as such, it would not include many (if any) adult fish. Considering fish appear to be the host of ‘choice’ in the (*Para)Nucleospora/Obruspora* radiation of enterocytozoonids (Diamant *et al*., 2014; El Alaoui *et al*., 2006b; Freeman *et al*., 2013; Lom & Dykoá, 2002; Nylund *et al*., 2010; Vaz Rodrigues *et al*., 2017), it is unlikely that these sampling methods would come across them. However, two species within this clade have also been found to infect crustacea (*Obruspora papernae* and *Paranucleospora theridion* (Synonym: *Desmozoon*

lepeotherii)), albeit parasitic crustacea. The larvae of many parasitic crustacea are pelagic/planktonic until they find a host , so it is likely that parasitic crustacea are also included in the screening of pelagic marine samples (Chapter 2) (E. Williams & Bunkley-Williams, 2019). This suggests one of two things: firstly, that of the two clades, the fish pathogens could be far less diverse, which could be due to a number of factors: for example, the life cycle of parasitic crustacea may not be optimal for widespread dispersal and transmission, as a large part of its (successful) lifecycle is spent attached to a host and in general only has one developmental stage (copepodid I) that is free swimming (E. Williams & Bunkley-Williams, 2019). As such, it is unlikely that *Paranucleospora/Obruspora* would have as many opportunities for transmission as the enterocytozoonids that infect free-living planktonic crustacea. Alternatively, it may be that fish are the primary hosts of this clade, so infection of parasitic crustacea is secondary after they have attached themselves to the gills and do not serve as intermediate hosts between fish and other aquatic life. The research undertaken on *Paranucleospora theridion* also suggests this, showing infection of Atlantic salmon (Salmo salar) through water-borne spores (Sveen *et al*., 2012). The latter inference is lent support by a lack of described infections of these pathogens, in planktonic crustacea not parasitising on fish. In line with this reasoning is the lack of *(Para)Nucleospora/Obruspora*-like sequences in any of the freshwater samples from either Chapter 2 or 3. Although this may also suggest that they only infect marine fish. If their transmission routes involved crustacea as an intermediate host, as suggested with *Enterospora/Enterocytozoon* clade, they would likely have a wider distribution/diversity of hosts. As planktonic life supports much of the pelagic food web (Winder & Jassby, 2011) and therefore more opportunities to transmit to a greater diversity of potential hosts. However, other species of Microsporidia have been described hyper-parasitising crustacea as a route of transmission between hosts (G. Stentiford *et al*., 2017). Further research is needed.

6.3 Significance to aquaculture

Taken together, the results of the research undertaken over the four data chapters in this study, it is concluded that the Enterocytozoonidae are in the position (both geographically and in possession of the infection-based tools) to be a possible threat to the future of outdoor aquaculture in the UK. It is a combination of factors that has led to this inference, but namely, it is based on the apparent widespread nature of the group (Chapter 2 and 3), with there being evidence of endemic widespread enterocytozoonids becoming an international problem for shrimp ponds (Chapter 4). This has been suggested as widespread occurrences of pathogens makes preventative measures difficult to put in place, which has allowed transmission of various pathogens (Elder *et al*., 2000; Gerba & Smith, 5005; Plowright *et al*., 2017). In addition, their presence in freshwater (Chapter 3), would allow them further access inland to possibly infect inland aquatic livestock, as enterocytozoonids would be exposed to a greater range of aquatic livestock (inland ponds), than is accessible in coastal-based aquaculture (net-based fishing and offshore open-net pens). This is primarily the way in which a number of pathogens infect livestock and humans (Lejeune *et al*., 2001; Lewerin *et al*., 2019), for example *Leptospira interrogans* is most frequently transmitted from the environment through ingestion or wounds (Bierque *et al*., 2020). This would potentially facilitate further transmission from their wild range of hosts to aquatic animals farmed inland, such as inland fish and crustacean farms. Causing further economic loss, and possibly unsustainable farming. This dispersal inland would most heavily affect farmers in low-income countries, as low-income countries have been shown to provide lower levels of biosecurity, due to a lack of implementation of mitigating measures (Kambey *et al*., 2021), which would allow environmental transmission into ponds. This could also explain the presence of *E. hepatopenaei* so far inland, as seen in Thai shrimp ponds. Having viable spores would also allow for enterocytozoonids to more readily come into contact with terrestrial host (livestock), likely through intake of water from bodies

of water exposed to the environment. As is already seen with *E. bieneusi in cattle and pigs* (Leelayoova *et al*., 2006; Santín & Fayer, 2011).

The trend for there to be so few species in both the transect study (Chapter 3) and metagenome study (Chapter 2) that were closely related to the fish infecting clade ((*Para)Nucleospora/Obruspora*), with most of the diversity found on the *Enterospora/Enterocytozoon* branches, suggests that the latter clade may pose a greater threat, in terms of pond-based aquaculture. Although this may be sampling bias, as this study largely sampled filtered pelagic life in inland rivers, to date, only *Enterocytozoon* and *Enterospora* species have been described infecting inland/estuarine species, farmed or wild (Santín & Fayer, 2011; G. D. Stentiford *et al*., 2007, 2011; Tourtip *et al*., 2009). It would be expected that if the (*Para)Nucleospora* clade were as widespread inland they would have been observed. Further evidence that *Enterospora/Enterocytozoon* may have a greater distribution inland and in estuaries, than (*Para)Nucleospora* taxa. This greater diversity, distribution, and abundance, is likely brought about by the fact that they infect planktonic crustacea (or effectively use them as intermediate hosts), which would suggest the *Enterospora/Enterocytozoon* clade would likely come into contact more frequently with livestock through several different routes. This is also a trend that has been noticed in a diverse set of Microsporidia that infect gammerids, suggesting that the host and parasites may have gone through co-radiations (Bacela-Spychalska *et al*., 2018), allowing for a widespread distribution. This is perhaps already displayed with the 'success' of *E. bieneusi* and *E. hepatopenaei*, two pathogens that are doing very well in human and human-influenced organisms. No enterocytozoonid from the (*Para)Nucleospora/Obruspora* clade has been so well represented in farmed/ fished species. For example, *E. hepatopenaei* has been described in over six countries in Asia (India (Rajendran *et al*., 2016a), China (Y. M. Liu *et al*., 2018), Vietnam (Ha *et al*., 2010; Tang *et al*., 2017), Venezuela (Tang *et al*., 2017), Indonesia (Tang *et al*., 2016a), and Brunei (Tang *et al*., 2015a)), in multiple instances. Whereas, besides *Nucleospora salmonis*, the other described species

in the (*Para)Nucleospora/Obruspora* clade do not seem to be as widespread and have only been described infecting farmed fish, in the location they have been initially described (Diamant *et al*., 2014; Freeman & Sommerville, 2009; Gunnarsson *et al*., 2017; Lom & Dykoá, 2002; Sveen *et al*., 2012). However, there is also a possibility that though the primers have shown they can amplify *Nucleospora cyclopteri*, there may be preferential amplification of the *Enterocytozoon/Enterospora* clade. Due to a lack of sufficient amounts of (*Para)Nucleospora/Obruspora,* this was not tested.

Currently, aquaculture in the UK is mainly skewed towards salmon, mussels, and oyster farming (Callaway *et al*., 2012), with little in the way of farmed crustaceans and with only two indoor shrimp farming facilities (Holland, 2020). This means that the presence of enterocytozoonids across South West estuaries may not be an immediate cause for concern in terms of transmission into aquaculture, as enterocytozoonids have yet to be described to infect bivalves. However, the segments of aquaculture, in the UK (and other developed countries that have their inland farming in closed facilities), most likely to be impacted by enterocytozoonids, are inshore/offshore farming in cages and pens. As these populations are farmed in the presence of marine life, they frequently encounter both endo and ectoparasites (González Poblete, 2015; Peacock *et al*., 2019), possibly allowing for transmission from wild reservoirs to farmed fish kept in pens. This is likely how *N. salmonis* became a pathogen of farmed Atlantic salmon (*Salmo salar*) in Scotland and *N. cyclopteri* a pathogen of farmed lumpfish (*Cyclopterus lumpus*) in Iceland, Canada, and Norway. Pathogen spill over into marine net-pens has also been described in other pathogen species, with strong affect, showing increased infections of parasitic crustacea closer to marine pens (Baskin, 2006). A long term study also showed the presence 39 species of salmon pathogens close to salmon farms (Shea *et al*., 2020).

With continued consumption of infected aquatic life, bivalves and fish being one of the animal-based foods frequently eaten raw in some dishes e.g. oysters and sushi, this could lead to further zoonotic events infecting the immunocompromised. Although some raw fish dishes are frozen prior to being eaten, some spores may still be viable (Maddox & Solter, 1996), though freezing up to -20^{o}c does inactivate most spores (Fenoy *et al*., 2009; Leiro *et al*., 2012). Although no offshore/inshore fisheries are held in South West UK, there are several blue mussel (*Mytilus edulis*) and Pacific oyster (*Magallana gigas*) farms with plans for expansion (Adamson *et al*., 2018; Black & Hughes, 2017). Though mussels have, as of yet, not been shown through histology to be infected, they may act as a reservoir for the Enterocytozoonidae through filter feeding. As has been shown with *E. bieneusi* in zebra mussels (*Dreissena polymorpha*) in the River Shannon, Ireland (Graczyk *et al*., 2004).

A good way to mitigate environmental-based infections is to farm stock in indoor tank cultures, which allows for more consistent levels of biosecurity and isolation (Yanong, 2013). As mentioned, two attempts have been made at starting indoor farming of shrimp (*Penaeus vannamei*) in Stirlingshire (and Great British Prawns (GBP) in Stirlingshire, UK) and Lincolnshire (FloGro Fresh in Lincolnshire, UK) in the last few years. As these are indoor facilities, it will be easier to maintain biosecurity, not having to worry about the environmental transmission of pathogens. However, with such a widespread abundance of Enterocytozoonidae in the UK (Chapter 3), there could be many routes through which to introduce infection into their farms. Many older indoor farms supply their ponds with water from natural streams and ponds (other flow through ref (Bregnballe, 2015)), which would allow for spores in the river to be circulated though these ponds (ref). However, new recirculating systems decrease this issue by using less water that is recirculated in the ponds through filters (Bregnballe, 2015).

6.4 Implications for biosecurity and policy

As well as informing government and local farmers on potentially problematic pathogens, so that the right steps can be taken towards minimising risks and planning detection strategies/techniques, this research may also impact upon importation/exportation policy. For example, if the UK had a large shrimp farming industry, there may be a reluctance to import shrimp from countries affected by *E. hepatopenaei* and any future enterocytozoonid shrimp infections. However, evidence that the pathogen is possibly endemic to/present in the UK (Chapter 3) would likely mean that the export/import of possibly EHP infected shrimp would be less problematic.

6.5 *Widespread abundance and possible origins of the enterocytozoonid clade*

The methods used between Chapter 2 and Chapter 3 differ in terms of sample collection and sequencing, so it would be unreliable to make anything but tentative comparative inferences on diversity and distribution. However, these are the first two studies to date looking at the diversity and distribution of this family. Comparisons made between these studies, though tentative, may help in understanding the factors that contribute to the diversity and distribution of this family.

Within a smaller spatial scale (South West UK: smallest and largest distance between transects is 9.7 km and 75.6 km, respectively), species distribution was largely uniform (Chapter 3), in that En. canceri-like sequences were present across all transects and E. hepatopenaei-like sequences to a lesser extent. As would be expected, this differed over greater spatial scales, a feature also noted in pathogenic fungi (Stukenbrock, 2014) (East Coast USA: smallest and largest coastal distance between sampling sites is 57.52 km and 9917.22 km, respectively), also shown through the geographic specificity of sequences throughout the metagenome data (Chapter 2). This suggests a lower level of dispersal of enterocytozoonids between the USA-based coastal sites than the

UK-based estuary transect sites (Chapter 3). This could be due to a number of factors, for example, freshwater samples from Chapter 3 showed the presence of En. canceri-like sequences, meaning possible transmission through inland freshwater bodies. In contrast, metagenomes from USA-based freshwater environments showed no evidence of enterocytozoonid-like sequences, suggesting a lower abundance of enterocytozoonids in the freshwater locations sampled in this study. The difference in species distribution could also be due to the much larger spatial distances between sites in the USA/Atlantic samples, than the UK sites, which could limit the dispersal of species between sites (Stukenbrock, 2014). However, the statistical analysis (Chapter 2) showed no association between genetic distance and geographic distance. Meaning other factors (including the difference in collection and sequencing methods) may explain the difference in genetic distances between sites, like the hypothesis stated above. The family's distribution and abundance suggest they appear to have an efficient means to distribute widely, likely through planktonic crustacea.

Looking at the greater diversity observed within the USA-based enterocytozoonid-like sequences (average: 13.8 OTUs) (Chapter 2), with the described diversity found in the UK mainly centred around the *Enterospora/Enterocytozoon* clade, it is likely that the USA-based regions investigated in this study had a greater diversity than those located in the South West of the UK (4 OTUs). However, due to the way errors are produced in the two different amplification processes (main errors in *Taq* polymerase-based PCRs are substitution errors during amplification (Potapov & Ong, 2017), while errors from various steps in Illumina can cause errors from miss identification of fluorophores in bridge amplification to enrichment PCR (Schirmer *et al*., 2015)), error rates of the two methods differ, with error rates for NGS technologies reported as higher (Glenn, 2011). So, this could account for some of the diversity observed in the USA-based metagenome sequences. Despite this, error rates seen in NGS technologies are improved through filtering and high depth and coverage of reads (Ma *et al*., 2019). So, with the caveat that further studies need

to be carried out using comparable methods to know for sure, it is still possible that there is greater diversity in the reservoir of enterocytozoonid-like organisms on the eastern USA coast/North Atlantic. Evidence of intermediate (intermediate between Enterocytozoonidae and all other Microsporidia) species (*Parahepatospora carcini* n. gen. n. sp.) in the Atlantic coast of Canada (Nova Scotia) adds further evidence for this hypothesis (Bojko*et al.* 2017).

6.6 Prevention

Possibly the best way to deal with potentially harmful and costly parasites is to prevent, as much as possible, the parasite from coming into contact with the host. The knowledge of how this emergent family spreads and routes they may take into farms would affect the way preventative measures are put in place. Inhibition of microsporidian infection, at an environmental level, is still not a well-studied topic, due to a lack of understanding of the diversity and abundance of Microsporidia. However, some chemical procedures have been devised, aimed at getting spores to expel their polar tubes prior to contact with possible hosts. For example, it has been found that freezing at -20^{o}c for 2h inactivates spores, or applying highly alkaline chemicals, such as $KMnO_4$, are able to get spores to expel polar tubes in *E. hepatopenaei* (Aldama-Cano *et al*., 2018). This was primarily inferred to be able to be used on incoming feed for shrimp ponds, by adding agents to raise the pH of the feed. However, with it being likely that *E. hepatopenaei* and other enterocytozoonids spores are local and endemic, a different approach may have to be taken. To prevent further emerging parasites from becoming pathogenic problems, a better understanding of host ranges is required. On top of that, host-species geographic boundaries need to be better understood. Further to this, maximising the health of livestock would also help to prevent infection from opportunistic pathogens (G. D. Stentiford *et al*., 2019), which could facilitate further infections from other pathogens (Aranguren *et al*., 2017).

In summary, this Ph.D. has shown the power of using a combination of molecular and bioinformatic approaches to improve the understanding of parasite diversity, transmission (molecular ecology), and evolutionary history. This approach can lead to a more well-informed risk assessment of pathogens in the environment, in a phylum of emerging pathogens, of ever increasing interest to the food industry.

References:

A. van Huis, Itterbeeck, J. Van, Klunder, H., Mertens, E., Halloran, A., Muir, G., & Vantomme, P. (2013). Edible insects. Future prospects for food and feed security. In *FAO*.

Adamson, E., Syvret, M., & Woolmer, A. (2018). *Shellfish seed supply for aquaculture in the UK report on views collected from the industry in 2017*. 1–20.

Afonso, L. O., Richmond, Z., Eaves, A. A., Richard, J., Hawley, L. M., & Garver, K. A. (2012). Use of ultraviolet c (UVC) radiation to inactivate infectious hematopoietic necrosis virus (IHNV) and viral hemorrhagic septicemia virus (VHSV) in fish processing plant effluent. *Journal of Aquaculture Research and Development*, *3* (1).

Akiyoshi, D. E., Morrison, H. G., Lei, S., Feng, X., Zhang, Q., Corradi, N., Mayanja, H., Tumwine, J. K., Keeling, P. J., Weiss, L. M., & Tzipori, S. (2009). Genomic survey of the non-cultivatable opportunistic human pathogen, Enterocytozoon bieneusi. *PLoS Pathogens*, *5* (1), 1–10.

Alavandi, S. V., & Poornima, M. (2012). Viral metagenomics: A tool for virus discovery and diversity in aquaculture. In *Indian Journal of Virology* (Vol. 23, Issue 2, pp. 88–98). Indian J Virol.

Aldama-Cano, D. J., Sanguanrut, P., Munkongwongsiri, N., Ibarra-Gámez, J. C., Itsathitphaisarn, O., Vanichviriyakit, R., Flegel, T. W., Sritunyalucksana, K., & Thitamadee, S. (2018). Bioassay for spore polar tube extrusion of shrimp Enterocytozoon hepatopenaei (EHP). *Aquaculture*, *490*, 156–161.

Alexander, H. K., & Day, T. (2010). Risk factors for the evolutionary emergence of pathogens. *Journal of the Royal Society Interface*, *7* (51), 1455–1474.

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*,

1144–1146.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215* (3), 403–410.

Andersen, S. C., Fachmann, M. S. R., Kiil, K., Nielsen, E. M., & Hoorfar, J. (2017). Gene-based pathogen detection: Can we use qPCR to predict the outcome of diagnostic metagenomics? *Genes*, *8*, 332.

Andersen, S. C., & Hoorfar, J. (2018). Surveillance of foodborne pathogens: Towards diagnostic metagenomics of fecal samples. In *Genes* (p. 14).

Anishchenko, M., Bowen, R. A., Paessler, S., Austgen, L., Greene, I. P., & Weaver, S. C. (2006). Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *13* (3), 4994–4999.

Antonovics, J., Wilson, A. J., Forbes, M. R., Hauffe, H. C., Kallio, E. R., Leggett, H. C., Longdon, B., Okamura, B., Sait, S. M., & Webster, J. P. (2017). The evolution of transmission mode. In *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Anttila, J., Kaitala, V., Laakso, J., & Ruokolainen, L. (2015). Environmental variation generates environmental opportunist pathogen outbreaks. *PLoS ONE*, *10* (12), 1–15.

Antúnez, K., Martín-Hernández, R., Prieto, L., Meana, A., Zunino, P., & Higes, M. (2009). Immune suppression in the honey bee (Apis mellifera) following infection by Nosema ceranae (Microsporidia). *Environmental Microbiology*, *11* (9), 2284–2290.

Aranguren, L. F., Han, J. E., & Tang, K. F. J. (2017). Enterocytozoon hepatopenaei (EHP) is a risk factor for acute hepatopancreatic necrosis disease (AHPND) and septic hepatopancreatic necrosis (SHPN) in the Pacific white shrimp Penaeus vannamei. *Aquaculture*, *471*, 37–42.

Ardila-Garcia, A. M., Raghuram, N., Sihota, P., & Fast, N. M. (2013).

Microsporidian diversity in soil, sand, and compost of the pacific Northwest. *Journal of Eukaryotic Microbiology*, *60* (6), 601–608.

Arundell, K., Dunn, A., Alexander, J., Shearman, R., Archer, N., & Ironside, J. E. (2015). Enemy release and genetic founder effects in invasive killer shrimp populations of Great Britain. *Biological Invasions*, *17* (5), 1439–1451.

Auburn, S., & Barry, A. E. (2017). Dissecting malaria biology and epidemiology using population genetics and genomics. In *International Journal for Parasitology* (pp. 77–85).

Aylagas, E., Borja, Á., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). *PLoS ONE*, *9* (3), 3.

Bacela-Spychalska, K., Wróblewski, P., Mamos, T., Grabowski, M., Rigaud, T., Wattier, R., Rewicz, T., Konopacka, A., & Ovcharenko, M. (2018). Europe-wide reassessment of Dictyocoela (Microsporidia) infecting native and invasive amphipods (Crustacea): Molecular versus ultrastructural traits. *Scientific Reports*, *8* (1), 6.

Balloux, F., & van Dorp, L. (2017). Q&A: What are pathogens, and what have they done to and for us? *BMC Biology 2017 15:1*, *15* (1), 1–6.

Baneth, G., Thamsborg, S. M., Otranto, D., Guillot, J., Blaga, R., Deplazes, P., & Solano-Gallego, L. (2016). Major Parasitic Zoonoses Associated with Dogs and Cats in Europe. In *Journal of Comparative Pathology* (pp. 54–74).

Baroudi, D., Zhang, H., Amer, S., Khelef, D., Roellig, D. M., Wang, Y., Feng, Y., & Xiao, L. (2018). Divergent Cryptosporidium parvum subtype and Enterocytozoon bieneusi genotypes in dromedary camels in Algeria. *Parasitology Research*, 1–6.

Baskin, Y. (2006). Sea sickness: The upsurge in marine diseases. *BioScience*, *56* (6), 464–469.

Bayliss, S. C., Verner-Jeffreys, D. W., Bartie, K. L., Aanensen, D. M., Sheppard, S. K., Adams, A., & Feil, E. J. (2017). The promise of whole genome pathogen sequencing for the molecular epidemiology of emerging aquaculture pathogens. In *Frontiers in Microbiology* (p. 121).

Becker, J. A., & Speare, D. J. (2007). Transmission of the microsporidian gill parasite, Loma salmonae. *Animal Health Research Reviews / Conference of Research Workers in Animal Diseases*, *8* (1), 59–68.

Becnel, J. J., & Andreadis, T. G. (2014). Microsporidia in Insects. In *Microsporidia: Pathogens of Opportunity: First Edition* (pp. 521–570).

Beldomenico, P. M., & Begon, M. (2010). Disease spread, susceptibility and infection intensity: vicious circles? *Trends in Ecology and Evolution*, *25* (1), 21–27.

Berg, J., Tymoczko, J., & Stryer, L. (2007). Protein Turnover and Amino Acid Catabolism. *Biochemistry*, Chapter 23.

Bielby, G. H. (1963). *River Teign Fisheries Survey*. http://aquaticcommons.org/10886/1/D3.32E (19)__River_Teign_fisheries_survey_1963.pdf

Bierque, E., Thibeaux, R., Girault, D., Soupé-Gilbert, M.-E., & Goarant, C. (2020). A systematic review of Leptospira in water and soil environments. *PLOS ONE*, *15* (1), e0227055.

Bigliardi, E., & Sacchi, L. (2001). Cell biology and invasion of the microsporidia. *Microbes and Infection*, *3* (5), 373–379.

Black, K., & Hughes, A. (2017). Future of the Sea : Trends in Aquaculture - Foresight Evidence Review. *Foresight, Government Office for Science.*, 41.

Blanco-Bercial, L., Cornils, A., Copley, N., & Bucklin, A. (2014). DNA Barcoding of Marine Copepods: Assessment of Analytical Approaches to Species Identification. *PLoS Currents*, *6*.

Bojko, J., Clark, F., Bass, D., Dunn, A. M., Stewart-Clark, S., Stebbing, P. D., &

Stentiford, G. D. (2017). Parahepatospora carcini n. gen., n. sp., a parasite of invasive Carcinus maenas with intermediate features of sporogony between the Enterocytozoon clade and other microsporidia. *Journal of Invertebrate Pathology*, *143*, 124–134.

Bonneaud, C., Weinert, L. A., & Kuijper, B. (2019). Understanding the emergence of bacterial pathogens in novel hosts. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (p. 1782).

Boonyaratpalin, S., Supamattaya, K., Kasornchandra, J., Direkbusaracom, S., Aekpanithanpong, U., & Chantanachooklin, C. (1993). Non-Occluded Baculo-like Virus, the Causative Agent of Yellow Head Disease in the Black Tiger Shrimp (Penaeus monodon). *Fish Pathology*, *28* (3), 103–109.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15* (4).

Bowden, S. E. & Drake, J. M. (2013). Ecology of Multi-host Pathogens of Animals. *Nature Education Knowledge*, *4* (8), 5.

Braga, M. P., Landis, M. J., Nylin, S., Janz, N., & Ronquist, F. (2020). Bayesian inference of ancestral host-parasite interactions under a phylogenetic model of host repertoire evolution. *Systematic Biology*, *69* (6), 1149–1162.

Bregnballe, J. (2015). A Guide to Recirculation Aquaculture. *FAO and Eurofish Report*, 100.

Briggs, M., Funge-Smith, S., Subasinghe, R., & Phillips, M. (2004). Introductions and movement of Penaeus vannamei and Penaeus stylirostris in Asia and the Pacific. In *RAP publication 2004/10*. FAO.

Bromenshenk, J. J., Henderson, C. B., Wick, C. H., Stanford, M. F., Zulich, A.

W., Jabbour, R. E., Deshpande, S. V., McCubbin, P. E., Seccomb, R. A., Welch, P. M., Williams, T., Firth, D. R., Skowronski, E., Lehmann, M. M., Bilimoria, S. L., Gress, J., Wanner, K. W., & Cramer, R. A. (2010). Iridovirus and microsporidian linked to honey bee colony decline. *PLoS ONE*, *5* (10).

Brosson, D., Kuhn, L., Delbac, F., Garin, J., Vivarès, C. P., & Texier, C. (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., Kuhn, L., Prensier, G., Vivarès, C. P., & Texier, C. (2005). The putative chitin deacetylase of Encephalitozoon cuniculi: A surface protein implicated in microsporidian spore-wall formation. *FEMS Microbiology Letters*, *247* (1), 81–90.

Brown, M. J. F. (2017). Microsporidia: An Emerging Threat to Bumblebees? In *Trends in Parasitology* (pp. 754–762).

Brown, S. P., Cornforth, D. M., & Mideo, N. (2012). Evolution of virulence in opportunistic pathogens: Generalism, plasticity, and control. In *Trends in Microbiology* (pp. 336–342).

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29* (8), 1917–1932.

Byers, J. E., Schmidt, J. P., Pappalardo, P., Haas, S. E., & Stephens, P. R. (2019). What factors explain the geographical range of mammalian parasites? *Proceedings of the Royal Society B: Biological Sciences*, *286* (1903).

Cali, A., Becnel, J. J., & Takvorian, P. M. (2017). Microsporidia. *Handbook of the Protists: Second Edition*, 1559–1618.

Cali, A., Neafie, R., Weiss, L. M., Ghosh, K., Vergara, R. B., Gupta, R., & Takvorian, P. M. (2010). Human vocal cord infection with the microsporidium Anncaliia algerae. *Journal of Eukaryotic Microbiology*, *57* (6), 562–567.

Cali, A., & Takvorian, P. M. (2003). Ultrastructure and development of Pleistophora ronneafiei n. sp., a microsporidium (Protista) in the skeletal muscle of an immune-compromised individual. *Journal of Eukaryotic Microbiology*, *50* (2), 77–85.

Callaway, R., Shinn, A. P., Grenfell, S. E., Bron, J. E., Burnell, G., Cook, E. J., Crumlish, M., Culloty, S., Davidson, K., Ellis, R. P., Flynn, K. J., Fox, C., Green, D. M., Hays, G. C., Hughes, A. D., Johnston, E., Lowe, C. D., Lupatsch, I., Malham, S., … Shields, R. J. (2012). Review of climate change impacts on marine aquaculture in the UK and Ireland. In *Aquatic Conservation: Marine and Freshwater Ecosystems* (pp. 389–421).

Cama, V. A., Pearson, J., Cabrera, L., Pacheco, L., Gilman, R., Meyer, S., Ortega, Y., & Xiao, L. (2007). Transmission of Enterocytozoon bieneusi between a child and guinea pigs. *Journal of Clinical Microbiology*, *45* (8), 2708–2710.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10* (1), 1–9.

Campbell, S. E., Williams, T. A., Yousuf, A., Soanes, D. M., Paszkiewicz, K. H., & Williams, B. A. P. (2013). The Genome of Spraguea lophii and the Basis of Host-Microsporidian Interactions. *PLoS Genetics*, *9* (8).

Canning, E. U., Curry, A., Cheney, S. A., Lafranchi-Tristem, N. J., Kawakami, Y., Hatakeyama, Y., Iwano, H., & Ishihara, R. (1999). Nosema tyriae n.sp. and Nosema sp., Microsporidian Parasites of Cinnabar Moth Tyria jacobaeae. *Journal of Invertebrate Pathology*, *74* (1), 29–38.

Capella-Gutiérrez, S., Marcet-Houben, M., & Gabaldón, T. (2012).

Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biology*, *10* (1), 47.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25* (15), 1972–1973.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pẽa, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. In *Nature Methods* (pp. 335–336).

Carilla-Latorre, S., Calvo-Garrido, J., Bloomfield, G., Skelton, J., Kay, R. R., Ivens, A., Martinez, J. L., & Escalante, R. (2008). Dictyostelium transcriptional responses to Pseudomonas aeruginosa: Common and specific effects from PAO1 and PA14 strains. *BMC Microbiology*, *8* (1), 1–15.

Cavalier-Smith, T. (1983). Eukaryote genetics: Cloning chromosome ends. *Nature*, *301* (5896), 112–113.

Chabchoub, N., Abdelmalek, R., Mellouli, F., Kanoun, F., Thellier, M., Bouratbine, A., & Aoun, K. (2009). Genetic identification of intestinal microsporidia species in immunocompromised patients in Tunisia. *American Journal of Tropical Medicine and Hygiene*, *80* (1), 24–27.

Chaijarasphong, T., Munkongwongsiri, N., Stentiford, G. D., Aldama-Cano, D. J., Thansa, K., Flegel, T. W., Sritunyalucksana, K., & Itsathitphaisarn, O. (2020). The shrimp microsporidian Enterocytozoon hepatopenaei (EHP): Biology, pathology, diagnostics and control. *Journal of Invertebrate Pathology*.

Chen, I. M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Huntemann, M., Varghese, N., White, J. R., Seshadri, R., Smirnova, T.,

Kirton, E., Jungbluth, S. P., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., & Kyrpides, N. C. (2019). IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, *47* (1), 666–677.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 884–890.

Chetia, H., Kabiraj, D., Sharma, S., & Bora, U. (2017). Comparative insights to the transportome of Nosema: A genus of parasitic microsporidians. In *bioRxiv* (p. 110809).

Chilmonczyk, S., Cox, W. T., & Hedrick, R. P. (1991). Enterocytozoon salmonis N. Sp.: An Intranuclear Microsporidium from Salmonid Fish. *The Journal of Protozoology*, *38* (3), 264–269.

Cissé, O. H., Ma, L., Wei Huang, D., Khil, P. P., Dekker, J. P., Kutty, G., Bishop, L., Liu, Y., Deng, X., Hauser, P. M., Pagni, M., Hirsch, V., Lempicki, R. A., Stajich, J. E., Cuomo, C. A., & Kovacs, J. A. (2018). Comparative population genomics analysis of the mammalian fungal pathogen Pneumocystis. *MBio*, *9* (3).

Clark, L. F. (2015). The current status of DNA barcoding technology for species identification in fish value chains. *Food Policy*, *54*, 85–94.

Clark, W. J., & Sigler, W. F. (1963). Method of Concentrating Phytoplanktonic Samples Using Membranefilters. In *Limnology and Oceanography* (pp. 127–129).

Cleaveland, S., Laurenson, M. K., & Taylor, L. H. (2001). Diseases of humans and their domestic mammals: Pathogen characteristics, host range and the risk of emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356* (1411), 991–999.

Cooke, G. M., Schlub, T. E., Sherwin, W. B., & Ord, T. J. (2016). Understanding the spatial scale of genetic connectivity at sea: Unique insights from a land

fish and a meta-analysis. *PLoS ONE*, *11* (5).

Costello, R., Emms, D. M., & Kelly, S. (2020). Gene duplication accelerates the pace of protein gain and loss from plant organelles. *Molecular Biology and Evolution*, *37* (4), 969–981.

Cox-Foster, D. L., Conlan, S., Holmes, E. C., Palacios, G., Evans, J. D., Moran, N. A., Quan, P. L., Briese, T., Hornig, M., Geiser, D. M., Martinson, V., VanEngelsdorp, D., Kalkstein, A. L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S. K., Simons, J. F., … Lipkin, W. I. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, *318* (5848), 283–287.

Coyle, C. M., Weiss, L. M., Rhodes, L. V., Cali, A., Takvorian, P. M., Brown, D. F., Visvesvara, G. S., Xiao, L., Naktin, J., Young, E., Gareca, M., Colasante, G., & Wittner, M. (2004). Fatal Myositis Due to the Microsporidian Brachiola algerae, a Mosquito Pathogen. *New England Journal of Medicine*, *351* (1), 42–47.

Crespi, V., & New, M. (2009). Penaeus vannamei. In *FAO*. http://www.fao.org/fishery/docs/DOCUMENT/aquaculture/CulturedSpecies/file/en/en_whitelegshrimp.htm

Crowder, D. W., Li, J., Borer, E. T., Finke, D. L., Sharon, R., Pattemore, D. E., & Medlock, J. (2019). Species interactions affect the spread of vector-borne plant pathogens independent of transmission mode. *Ecology*, *100* (9).

Cruz-Flores, R., Mai, H. N., Noble, B. L., Schofield, P. J., & Dhar, A. K. (2019). Detection of Enterocytozoon hepatopenaei using an invasive but non-lethal sampling method in shrimp (Penaeus vannamei). *Journal of Microbiological Methods*, *162*, 38–41.

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

22 (12), 2478–2488.

D'Agostino, M., & Cook, N. (2015). Foodborne Pathogens. In *Encyclopedia of Food and Health* (pp. 83–86).

da Cunha, M. J. R., Cury, M. C., & Santín, M. (2017). Molecular identification of Enterocytozoon bieneusi, Cryptosporidium, and Giardia in Brazilian captive birds. *Parasitology Research*, *116* (2), 487–493.

Damborg, P., Broens, E. M., Chomel, B. B., Guenther, S., Pasmans, F., Wagenaar, J. A., Weese, J. S., Wieler, L. H., Windahl, U., Vanrompay, D., & Guardabassi, L. (2016). Bacterial Zoonoses Transmitted by Household Pets: State-of-the-Art and Future Perspectives for Targeted Research and Policy Actions. *Journal of Comparative Pathology*, *155* (1), 27–40.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27* (15), 2156–2158.

Davis, M. Y., Zhang, H., Brannan, L. E., Carman, R. J., & Boone, J. H. (2016). Rapid change of fecal microbiome and disappearance of Clostridium difficile in a colonized infant after transition from breast milk to cow milk. *Microbiome*, *4* (1), 1–10.

De Albuquerque, N. R. M., Ebert, D., & Haag, K. L. (2020). Transposable element abundance correlates with mode of transmission in microsporidian parasites. *Mobile DNA*, *11* (1).

De Schryver, P., & Vadstein, O. (2014). Ecological theory as a foundation to control pathogenic invasion in aquaculture. In *ISME Journal* (pp. 2360–2368).

Defoirdt, T. (2016). Implications of Ecological Niche Differentiation in Marine Bacteria for Microbial Management in Aquaculture to Prevent Bacterial Disease. In *PLoS Pathogens* (p. e1005843).

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. In *Molecular Ecology* (pp. 5872–5895).

Denholm, I., & Devine, G. (2013). Insecticide Resistance. In *Encyclopedia of Biodiversity: Second Edition* (pp. 298–307).

Deshpande, B. D., Agrawal, P. S., Yenkie, M. K. N., & Dhoble, S. J. (2020). Prospective of nanotechnology in degradation of waste water: A new challenges. In *Nano-Structures and Nano-Objects* (p. 100442).

Desjardins, C. A., Giamberardino, C., Sykes, S. M., Yu, C. H., Tenor, J. L., Chen, Y., Yang, T., Jones, A. M., Sun, S., Haverkamp, M. R., Heitman, J., Litvintseva, A. P., Perfect, J. R., & Cuomo, C. A. (2017). Population genomics and the evolution of virulence in the fungal pathogen Cryptococcus neoformans. *Genome Research*, *27* (7), 1207–1219.

Desjardins, C. A., Sanscrainte, N. D., Goldberg, J. M., Heiman, D., Young, S., Zeng, Q., Madhani, H. D., Becnel, J. J., & Cuomo, C. A. (2015). Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nature Communications*, *6*, 21.

Desportes, I., Le Charpentier, Y., Galian, A., Bernard, F., Cochand-Priollet, B., Lavergne, A., Ravisse, P., & Modigliani, R. (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* (2), 250–254.

Desrina, Prayitno, S. B., Haditomo, A. H. C., Latritiani, R., & Sarjito, S. (2020). Detection of Enterocytozoon hepatopenaei (EHP) DNA in the polychaetes from shrimp ponds suffering white feces syndrome outbreaks. *Biodiversitas*, *21* (1), 369–374.

Diamant, A., Rothman, S. B. S., Goren, M., Galil, B. S., Yokes, M. B.,

Szitenberg, A., & Huchon, D. (2014). Biology of a new xenoma-forming gonadotropic microsporidium in the invasive blotchfin dragonet Callionymus filamentosus. *Diseases of Aquatic Organisms*, *109* (1), 35–54.

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

Didier, E. S., & Khan, I. A. (2014). The Immunology of Microsporidiosis in Mammals. In *Microsporidia: Pathogens of Opportunity: First Edition* (pp. 307–325).

Didier, E. S., Schuitema, A. R. J., & Terpstra, W. J. (1995). Genetic and Immunological Characterization of the Microsporidian Septata Intestinalis Cali, Kotler and Orenstein, 1993: Reclassification to Encephalitozoon Intestinalis. *Parasitology*, *110* (3), 277–285.

Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., Goldsmith, D. B., Michisaki, R., Breitbart, M., Boehm, A. B., & Chavez, F. P. (2017). Evaluation of filtration and DNA extraction methods for environmental DNA biodiversity assessments across multiple trophic levels. *Frontiers in Marine Science*, *4*, 314.

Donovan, P. D., Gonzalez, G., Higgins, D. G., Butler, G., & Ito, K. (2018). Identification of fungi in shotgun metagenomics datasets. *PLoS ONE*, *13* (2), e0192898.

Dorn, P. L., Justi, S., Krafsur, E. S., Lanzaro, G. C., Cornel, A. J., Lee, Y., & Hill, C. A. (2017). Genetics of Major Insect Vectors. In *Genetics and Evolution of Infectious Diseases: Second Edition* (pp. 341–382).

Dunn, A. M., & Smith, J. E. (2001). Microsporidian life cycles and diversity: The relationship between virulence and transmission. In *Microbes and Infection* (pp. 381–388).

Duplouy, A., Minard, G., & Saastamoinen, M. (2020). The gut bacterial community affects immunity but not metabolism in a specialist herbivorous

butterfly. *Ecology and Evolution*, *10* (16), 8755–8769.

Earley, B., Buckham Sporer, K., & Gupta, S. (2017). Invited review: Relationship between cattle transport, immunity and respiratory disease. *Animal*, 486–492.

Edbrooke, D. (2004). Fears over Exe Salmon. *Express and Echo*, *11*.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26* (19), 2460–2461.

Ekroth, A. K. E., Rafaluk-Mohr, C., & King, K. C. (2019). Diversity and disease: Evidence for the monoculture effect beyond agricultural systems. In *bioRxiv* (p. 668228).

El Alaoui, H., Grésoviac, S. J., & Vivarès, C. P. (2006a). Occurrence of the microsporidian parasite Nucleospora salmonis in four species of salmonids from the Massif Central of France. *Folia Parasitologica*, *53* (1), 37–43.

El Alaoui, H., Grésoviac, S. J., & Vivarès, C. P. (2006b). Occurrence of the microsporidian parasite Nucleospora salmonis in four species of salmonids from the Massif Central of France. *Folia Parasitologica*, *53* (1), 37–43.

Elder, R. O., Keen, J. E., Siragusa, G. R., Barkocy-Gallagher, G. A., Koohmaraie, M., & Laegreid, W. W. (2000). Correlation of enterohemorrhagic Escherichia coli 0157 prevalence in feces, hides, and carcasses of beef cattle during processing. *Proceedings of the National Academy of Sciences of the United States of America*, *97* (7).

Embley, T. M., & Martin, W. (2006). Eukaryotic evolution, changes and challenges. In *Nature* (pp. 623–630).

Engelstädter, J., & Fortuna, N. Z. (2019). The dynamics of preferential host switching: Host phylogeny as a key predictor of parasite distribution*. *Evolution*, *73* (7), 1330–1340.

Environment Agency. (2017). *Tamar Index River Monitoring 2017 Report*. www.gov.uk/environment-agency

Espern, A., Morio, F., Miegeville, M., Illa, H., Abdoulaye, M., Meyssonnier, V., Adehossi, E., Lejeune, A., Phung, D. C., Besse, B., & Gay-Andrieu, F. (2007). Molecular study of microsporidiosis due to Enterocytozoon bieneusi and Encephalitozoon intestinalis among human immunodeficiency virus-infected patients from two geographical areas: Niamey, Niger, and Hanoi, Vietnam. *Journal of Clinical Microbiology*, *45* (9), 2999–3002.

FAO. (2016). The State of World Fisheries and Aquaculture. Contributing to food security and nutrition for all. The State of World Fisheries and Aquaculture 2016. In *FAO*.

FAO. (2020). The State of World Fisheries and Aquaculture: Sustainability in Action. *FAO*.

Fenoy, S., Rueda, C., Higes, M., Martín-Hernández, R., & Del Aguila, C. (2009). High-level resistance of Nosema ceranae, a parasite of the honeybee, to temperature and desiccation. *Applied and Environmental Microbiology*, *75* (21), 6886–6889.

Finet, C., Timme, R. E., Delwiche, C. F., & Marlétaz, F. (2010). Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology*, *20* (24), 2217–2222.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., & Stoeckert, C. J. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, *35* (1), 6.12.1-6.12.19.

Fiuza, V. R. da S., Lopes, C. W. G., Cosendey, R. I. J., de Oliveira, F. C. R., Fayer, R., & Santín, M. (2016). Zoonotic Enterocytozoon bieneusi genotypes found in brazilian sheep. *Research in Veterinary Science*, *107*, 196–201.

Flegel, T. W. (1997). Major viral diseases of the black tiger prawn (Penaeus monodon) in Thailand. In *World Journal of Microbiology and Biotechnology* (pp. 433–442).

Flegel, T. W. (2006). Detection of major penaeid shrimp viruses in Asia, a historical perspective with emphasis on Thailand. In *Aquaculture* (pp. 1–33).

Flegel, T. W. (2012). Historic emergence, impact and current status of shrimp pathogens in Asia. In *Journal of Invertebrate Pathology* (pp. 166–173).

Floden, E. W., Tommaso, P. D., Chatzou, M., Magis, C., Notredame, C., & Chang, J. M. (2016). PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases. *Nucleic Acids Research*, *44* (W1), W339–W343.

Floyd, R., Lima, J., deWaard, J., Humble, L., & Hanner, R. (2010). Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. *Biological Invasions 2010 12:9*, *12* (9), 2947–2954.

Foltz, J. R., Plant, K. P., Overturf, K., Clemens, K., & Powell, M. S. (2009). Detection of Nucleospora salmonis in steelhead trout, Oncorhynchus mykiss (Walbaum), using quantitative polymerase chain reaction (qPCR). *Journal of Fish Diseases*, *32* (6), 551–555.

Fontes, I., Hartikainen, H., Williams, C., & Okamura, B. (2017). Persistence, impacts and environmental drivers of covert infections in invertebrate hosts. *Parasites & Vectors 2017 10:1*, *10* (1), 1–15.

Franzen, C. (2005). How do microsporidia invade cells? *Folia Parasitologica*, *52* (1–2), 36–40.

Freeman, M. A., Kasper, J. M., & Kristmundsson, Á. (2013). *Nucleospora cyclopteri n. sp., an intranuclear microsporidian infecting wild lumpfish, Cyclopterus lumpus L., in Icelandic waters*. 1–13.

Freeman, M. A., & Sommerville, C. (2009). Desmozoon lepeophtherii n. gen., n. sp., (Microsporidia: Enterocytozoonidae) infecting the salmon louse Lepeophtheirus salmonis (Copepoda: Caligidae). *Parasites and Vectors*, *2*

(1), 1–15.

Frentiu, F. D., Zakir, T., Walker, T., Popovici, J., Pyke, A. T., van den Hurk, A., McGraw, E. A., & O'Neill, S. L. (2014). Limited Dengue Virus Replication in Field-Collected Aedes aegypti Mosquitoes Infected with Wolbachia. *PLoS Neglected Tropical Diseases*, *8* (2).

Furfaro, G., Chimienti, G., & Mariottini, P. (2020). DNA barcoding unveiling rare species: the case of Pruvotfolia pselliotes (Labbé, 1923) (Mollusca: Gastropoda: Nudibranchia) in the Mediterranean Sea. *European Zoological Journal*, *87* (1), 459–462.

Galván, A. L., Magnet, A., Izquierdo, F., Fenoy, S., Rueda, C., Adillo, C. F. V., Henriques-Gil, N., & del Aguila, C. (2013). Molecular characterization of human-pathogenic microsporidia and cyclospora cayetanensis isolated from various water sources in Spain: A year-long longitudinal study. *Applied and Environmental Microbiology*, *79* (2), 449–459.

Garcia, S. M., & Rosenberg, A. A. (2010). Food security and marine capture fisheries: Characteristics, trends, drivers and future perspectives. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (pp. 2869–2880).

Gargan, P., Whelan, K. F., & Tully, O. (1993). Infestation of sea trout (_Salmo trutta_ L.) by sea lice _ (Lepeophheirus salmonis_ (Kroyer)) in systems close to and distant from salmon farms in Ireland. *International Council for the Exploration of the Sea (ICES) Journal of Marine Science*.

Geiger, A., Malele, I., Abd-Alla, A. M., & Njiokou, F. (2018). Blood feeding tsetse flies as hosts and vectors of mammals-pre-adapted African Trypanosoma: current and expected research directions. *BMC Microbiology*, *18* (1), 17–29.

Gerba, C. P., & Smith, J. E. (5005). Sources of pathogenic microorganisms and their fate during land application of wastes. *Journal of Environmental Quality*, *34* (1), 42–48.

Gill, E. E., Becnel, J. J., & Fast, N. M. (2008). ESTs from the microsporidian Edhazardia aedis. *BMC Genomics*, *9* (1), 1–12.

Gill, E. E., & Fast, N. M. (2007). Stripped-down DNA repair in a highly reduced parasite. *BMC Molecular Biology*, *8* (1), 1–14.

Giraud, T., Gladieux, P., & Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, *25* (7), 387–395.

Gjøen, H. M., & Bentsen, H. B. (1997). Past, present, and future of genetic improvement in salmon aquaculture. *ICES Journal of Marine Science*, 1009–1014.

Glaser, R., & Kiecolt-Glaser, J. K. (2005). Stress-induced immune dysfunction: Implications for health. In *Nature Reviews Immunology* (pp. 243–251).

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, *11* (5), 759–769.

GOAL 2019: Global shrimp production review « Global Aquaculture Advocate. (n.d.). Retrieved January 25, 2021, from https://www.aquaculturealliance.org/advocate/goal-2019-global-shrimp-production-review/

Goecke, F., Klemetsdal, G., & Ergon, Å. (2020). Cultivar Development of Kelps for Commercial Cultivation—Past Lessons and Future Prospects. In *Frontiers in Marine Science* (Vol. 8, p. 110). Frontiers.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. In *Methods in Ecology and Evolution* (Vol. 7, Issue 11, pp. 1299–1307).

Gontcharov, A. A., Marin, B., & Melkonian, M. (2004). Are Combined Analyses Better Than Single Gene Phylogenies? A Case Study Using SSU rDNA and rbcL Sequence Comparisons in the Zygnematophyceae (Streptophyta). *Molecular Biology and Evolution*, *21* (3), 612–624.

González-Candelas, F., Patiño-Galindo, J. Á., & Valiente-Mullor, C. (2018). *Population Genomics of Human Viruses*.

González-Tortuero, E., Rusek, J., Maayan, I., Petrusek, A., Piálek, L., Laurent, S., & Wolinska, J. (2016). Genetic diversity of two Daphnia-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region. *Parasites and Vectors*, *9* (1).

González, M. T., & Poulin, R. (2005). Spatial and temporal predictability of the parasite community structure of a benthic marine fish along its distributional range. *International Journal for Parasitology*, *35* (13), 1369–1377.

González Poblete, L. (2015). *Ectoparasites and Associated Pathogens Affecting Farmed Salmon During Marine Grow Out in Chile and Australia*.

Grabner, D. S. (2017). Hidden diversity: parasites of stream arthropods. *Freshwater Biology*, *62* (1), 52–64.

Graczyk, T., Conn, D., Lucy, F., Minchin, D., Tamang, L., Moura, L. S., & DaSilva, A. (2004). Human waterborne parasites in zebra mussels (Dreissena polymorpha) from the Shannon River drainage area, Ireland. *Parasitology Research*, *93* (5), 385–391.

Gunnarsson, G. S., Blindheim, S., Karlsbakk, E., Plarre, H., Imsland, A. K., Handeland, S., Sveier, H., & Nylund, A. (2017). Desmozoon lepeophtherii (microsporidian) infections and pancreas disease (PD) outbreaks in farmed Atlantic salmon (Salmo salar L.). *Aquaculture*, *468*, 141–148.

Ha, N. T., Ha, D. T., Thuy, N. T., & Lien, V. T. K. (2010). Occurrence of microsporidia Enterocytozoon hepatopenaei in white feces disease of cultured black tiger shrimp (Penaeus monodon) in Vietnam. *Aquatic Animal*

Disease, Http://Hadong86.Wordpress.Com/, *July 3*.

Haag, K. L., James, T. Y., Pombert, J. F., Larsson, R., Schaer, T. M. M., Refardt, D., & Ebert, D. (2014). Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *111* (43), 15480–15485.

Haag, K. L., Pombert, J. F., Sun, Y., De Albuquerque, N. R. M., Batliner, B., Fields, P., Lopes, T. F., & Ebert, D. (2019). Microsporidia with Vertical Transmission Were Likely Shaped by Nonadaptive Processes. *Genome Biology and Evolution*, *12* (1), 3599–3614.

Haag, K. L., Sheikh-Jabbari, E., Ben-Ami, F., & Ebert, D. (2013). Microsatellite and single-nucleotide polymorphisms indicate recurrent transitions to asexuality in a microsporidian parasite. *Journal of Evolutionary Biology*, *26* (5), 1117–1128.

Haag, K. L., Traunecker, E., & Ebert, D. (2013). Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. *Molecular Ecology*, *22* (2), 314–326.

Haine, E. R., Motreuil, S., & Rigaud, T. (2007). Infection by a vertically-transmitted microsporidian parasite is associated with a female-biased sex ratio and survival advantage in the amphipod Gammarus roeseli. *Parasitology*, *134* (10), 1363–1367.

Halánová, M., Letková, V., MacÁk, V., Štefkovič, M., & Halán, M. (1999). The first finding of antibodies to Encephalitozoon cuniculi in cows in Slovakia. *Veterinary Parasitology*, *82* (2), 167–171.

Hall, T. (2013). BioEdit version 7.2. 5. *Ibis Biosciences, Carlsbad, CA, USA*.

Han, B. A., Kramer, A. M., & Drake, J. M. (2016). Global Patterns of Zoonotic Disease in Mammals. In *Trends in Parasitology* (pp. 565–577).

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, *5* (10), R245–R249.

Hangmann, M. (2009). How to Feed the World in 2050. *FAO*, *79* (5), 484.

Hartikainen, H., Ashford, O. S., Berney, C., Okamura, B., Feist, S. W., Baker-Austin, C., Stentiford, G. D., & Bass, D. (2014). Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians. *ISME Journal*, *8* (1), 177–186.

Hartikainen, H., Stentiford, G. D., Bateman, K. S., Berney, C., Feist, S. W., Longshaw, M., Okamura, B., Stone, D., Ward, G., Wood, C., & Bass, D. (2014). Mikrocytids are a broadly distributed and divergent radiation of parasites in aquatic invertebrates. *Current Biology*, *24* (7), 807–812.

Hedrick, R. P., Purcell, M. K., & Kurobe, T. (2012). *3.2.17 Salmonid Intranuclear Microsporidosis*. *June*, 1–8.

Heinz, E., Brindle, R., Morgan-McCalla, A., Peters, K., & Thomson, N. R. (2019). Caribbean multi-centre study of Klebsiella pneumoniae: whole-genome sequencing, antimicrobial resistance and virulence factors. *Microbial Genomics*, *5* (5), e000266.

Heinz, E., Williams, T. A., Nakjang, S., Noël, C. J., Swan, D. C., Goldberg, A. V., Harris, S. R., Weinmaier, T., Markert, S., Becher, D., Bernhardt, J., Dagan, T., Hacker, C., Lucocq, J. M., Schweder, T., Rattei, T., Hall, N., Hirt, R. P., & Embley, T. M. (2012). The Genome of the Obligate Intracellular Parasite Trachipleistophora hominis: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. *PLoS Pathogens*, *8* (10), e1002979.

Henry, J. E. (1971). Experimental application of Nosema locustae for control of grasshoppers. *Journal of Invertebrate Pathology*, *18* (3), 389–394.

Herren, J. K., Mbaisi, L., Mararo, E., Makhulu, E. E., Mobegi, V. A., Butungi, H., Mancini, M. V., Oundo, J. W., Teal, E. T., Pinaud, S., Lawniczak, M. K. N., Jabara, J., Nattoh, G., & Sinkins, S. P. (2020). A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. *Nature Communications*, *11* (1), 1–10.

Hines, H. M., Hunt, J. H., O'Connor, T. K., Gillespie, J. J., & Cameron, S. A. (2007). Multigene phylogeny reveals eusociality evolved twice in vespid wasps. *Proceedings of the National Academy of Sciences of the United States of America*, *104* (9), 3295–3299.

Hinkle, G., Morrison, H. G., & Sogin, M. L. (1997). Genes coding for reverse transcriptase, DNA-directed RNA polymerase, and chitin synthase from the microsporidian Spraguea lophii. *Biological Bulletin*, 250–251.

Hinney, B., Sak, B., Joachim, A., & Kváč, M. (2016). More than a rabbit's tale - Encephalitozoon spp. in wild mammals and birds. In *International Journal for Parasitology: Parasites and Wildlife* (pp. 76–87).

Hjeltnes, B., Britt, B.-J., Geir, B., Asle, H., & Cecilie, S. W. (2019). The health situation in norwegian aquaculture. In *Veterinaerinstituttet*.

Hoekstra, H. E., & Coyne, J. A. (2007). The locus of evolution: Evo devo and the genetics of adaptation. In *Evolution* (pp. 995–1016).

Hoffmann, S., & Scallan, E. (2017). Epidemiology, Cost, and Risk Analysis of Foodborne Disease. In *Foodborne Diseases: Third Edition* (pp. 31–63).

Hölker, F., Wurzbacher, C., Weißenborn, C., Monaghan, M. T., Holzhauer, S. I. J., & Premke, K. (2015). Microbial diversity and community respiration in freshwater sediments influenced by artificial light at night. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370* (1667).

Holland, J. (2020). *Ascending U.K. prawn farms now riding out the coronavirus storm*. The Responsible Seafood Advocate. https://www.aquaculturealliance.org/advocate/ascending-u-k-prawn-farms-

now-riding-out-the-coronavirus-storm/

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, *35* (2), W585–W587.

Huang, H., He, Q., Kubatko, L. S., & Knowles, L. L. (2010). Sources of error inherent in species-tree estimation: Impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. *Systematic Biology*, *59* (5), 573–583.

Hudson, D. A., Hudson, N. B., & Pyecroft, S. B. (2001). Mortalities of Penaeus japonicus prawns associated with microsporidean infection. *Australian Veterinary Journal*, *79* (7), 504–505.

Hutchins, P. R., Sepulveda, A. J., Hartikainen, H., Staigmiller, K. D., Opitz, S. T., Yamamoto, R. M., Huttinger, A., Cordes, R. J., Weiss, T., Hopper, L. R., Purcell, M. K., & Okamura, B. (2021). Exploration of the 2016 Yellowstone River fish kill and proliferative kidney disease in wild fish populations. *Ecosphere*, *12* (3).

Huver, J. R., Koprivnikar, J., Johnson, P. T. J., & Whyard, S. (2015). Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological Applications*, *25* (4), 991–1002.

Huyghe, F., & Kochzius, M. (2018). Sea surface currents and geographic isolation shape the genetic population structure of a coral reef fish in the Indian Ocean. *PLoS ONE*, *13* (3), e0193825.

Inshore Fisheries and Conservation Authority. (2019). *Managing Hand Working Fishing Activity. A Focus on Hand Gathering Supplementary Report for the B&PSC-Information & Evidence*. *July*.

Izquierdo, F., Castro Hermida, J. A., Fenoy, S., Mezo, M., González-Warleta, M., & Aguila, C. del. (2011). Detection of microsporidia in drinking water,

wastewater and recreational rivers. *Water Research*, *45* (16), 4837–4843.

Javanmard, E., Mirjalali, H., Niyyati, M., Jalilzadeh, E., Seyed Tabaei, S. J., Asadzadeh Aghdaei, H., Nazemalhosseini-Mojarad, E., & Zali, M. R. (2018). Molecular and phylogenetic evidences of dispersion of human-infecting microsporidia to vegetable farms via irrigation with treated wastewater: One-year follow up. *International Journal of Hygiene and Environmental Health*, *221* (4), 642–651.

Jennings, S., Stentiford, G. D., Leocadio, A. M., Jeffery, K. R., Metcalfe, J. D., Katsiadaki, I., Auchterlonie, N. A., Mangi, S. C., Pinnegar, J. K., Ellis, T., Peeler, E. J., Luisetti, T., Baker-Austin, C., Brown, M., Catchpole, T. L., Clyne, F. J., Dye, S. R., Edmonds, N. J., Hyder, K., … Verner-Jeffreys, D. W. (2016). Aquatic food security: insights into challenges and solutions from an analysis of interactions between fisheries, aquaculture, food safety, human health, fish and human welfare, economy and environment. *Fish and Fisheries*, *17* (4), 893–938.

Jeong, D. K., Won, G. Y., Park, B. K., Hur, J., You, J. Y., Kang, S. J., Oh, I. G., Lee, Y. S., Stein, B. D., & Lee, J. H. (2007). Occurrence and genotypic characteristics of Enterocytozoon bieneusi in pigs with diarrhea. *Parasitology Research*, *102* (1), 123–128.

Jiang, Y., Tao, W., Wan, Q., Li, Q., Yang, Y., Lin, Y., Zhang, S., & Li, W. (2015). Zoonotic and potentially host-adapted Enterocytozoon bieneusi genotypes in sheep and cattle in Northeast China and an increasing concern about the zoonotic importance of previously considered ruminant-adapted genotypes. *Applied and Environmental Microbiology*, *81* (10), 3326–3335.

Johnson, S. C., & Albright, L. J. (1991). The developmental stages of Lepeophtheirus salmonis (Krøyer, 1837) (Copepoda: Caligidae). *Canadian Journal of Zoology*, *69* (4), 929–950.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24* (11), 1403–1405.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30* (9), 1236–1240.

Joshi, N., & Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. *Available at Https://Github.Com/Najoshi/Sickle.*, 2011.

Josling, G. A., Petter, M., Oehring, S. C., Gupta, A. P., Dietz, O., Wilson, D. W., Schubert, T., Längst, G., Gilson, P. R., Crabb, B. S., Moes, S., Jenoe, P., Lim, S. W., Brown, G. V., Bozdech, Z., Voss, T. S., & Duffy, M. F. (2015). A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. *Cell Host and Microbe*, *17* (6), 741–751.

Kambey, C. S. B., Campbell, I., Cottier-Cook, E. J., Nor, A. R. M., Kassim, A., Sade, A., & Lim, P. E. (2021). Evaluating biosecurity policy implementation in the seaweed aquaculture industry of Malaysia, using the quantitative knowledge, attitude, and practices (KAP) survey technique. *Marine Policy*, *134*, 104800.

Karpov, S. A., Mamkaeva, M. A., Aleoshin, V. V., Nassonova, E., Lilje, O., & Gleason, F. H. (2014). Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. In *Frontiers in Microbiology* (p. 112).

Karpov, S. A., Mikhailov, K. V., Mirzaeva, G. S., Mirabdullaev, I. M., Mamkaeva, K. A., Titova, N. N., & Aleoshin, V. V. (2013). Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. *Protist*, *164* (2), 195–205.

Karpov, S. A., Tcvetkova, V. S., Mamkaeva, M. A., Torruella, G., Timpano, H., Moreira, D., Mamanazarova, K. S., & López-García, P. (2017). Morphological and Genetic Diversity of Opisthosporidia: New Aphelid

Paraphelidium tribonemae gen. et sp. nov. *Journal of Eukaryotic Microbiology*, *64* (2), 204–212.

Karthikeyan, K., & Sudhakaran, R. (2019a). Experimental horizontal transmission of Enterocytozoon hepatopenaei in post-larvae of whiteleg shrimp, Litopenaeus vannamei. *Journal of Fish Diseases*, *42* (3), 397–404.

Karthikeyan, K., & Sudhakaran, R. (2019b). Experimental horizontal transmission of Enterocytozoon hepatopenaei in post-larvae of whiteleg shrimp, Litopenaeus vannamei. *Journal of Fish Diseases*.

Karvonen, A., Rintamäki, P., Jokela, J., & Valtonen, E. T. (2010). Increasing water temperature and disease risks in aquatic systems: Climate change increases the risk of some, but not all, diseases. *International Journal for Parasitology*, *40* (13), 1483–1488.

Katinka, M. D., Duprat, S., Cornillott, E., Méténler, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoul, H., Peyret, P., Saurin, W., Gouy, M., & Weissenbach, J. (2001). Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. *Nature*, *414* (6862), 450–453.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20* (4), 1160–1166.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30* (4), 772–780.

Keeling, P. (2009). Five questions about microsporidia. *PLoS Pathogens*, *5* (9), 1–3.

Keeling, P. J., & Corradi, N. (2011). Balancing loss of function with shrinking genomes in the microsporidia. *Virulence*, *2* (1), 67–70.

Keeling, P. J., & Slamovits, C. H. (2004a). Simplicity and complexity of

microsporidian genomes. In *Eukaryotic Cell* (Vol. 3, Issue 6, pp. 1363–1369).

Keeling, P. J., & Slamovits, C. H. (2004b). Simplicity and Complexity of Microsporidian Genomes. *Eukaryotic Cell*, *3* (6), 1363–1369.

Kellen, W. R., Hoffmann, D. F., & Collier, S. S. (1977). Studies on the biology and ultrastructure of nosema transitellae sp. n. (microsporidia: Nosematidae) in the navel orangeworm, paramyelois transitella (lepidoptera: Pyralidae)1. *Journal of Invertebrate Pathology*, *29* (3), 289–296.

King, K. C., & Lively, C. M. (2012). Does genetic diversity limit disease spread in natural host populations? *Heredity 2012 109:4*, *109* (4), 199–203.

Kishida, T. (2008). Pattern of the divergence of olfactory receptor genes during tetrapod evolution. *PLoS ONE*, *3* (6), e2385.

Klee, J., Besana, A. M., Genersch, E., Gisder, S., Nanetti, A., Tam, D. Q., Chinh, T. X., Puerta, F., Ruz, J. M., Kryger, P., Message, D., Hatjina, F., Korpela, S., Fries, I., & Paxton, R. J. (2007). Widespread dispersal of the microsporidian Nosema ceranae, an emergent pathogen of the western honey bee, Apis mellifera. *Journal of Invertebrate Pathology*, *96* (1), 1–10.

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE*, *6* (1), e15925.

Kokot, M., Dlugosz, M., & Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer statistics. *Bioinformatics (Oxford, England)*, *33* (17), 2759–2761.

Kotler, D. P., & Orenstein, J. M. (1998). Clinical syndromes associated with microsporidiosis. *Advances in Parasitology*, *40*, 321–349.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X:

Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35* (6), 1547–1549.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33* (7), 1870–1874.

Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A., & Herfst, S. (2018). Transmission routes of respiratory viruses among humans. In *Current Opinion in Virology* (pp. 142–151).

Lafferty, K. D., & Holt, R. D. (2003). How should environmental stress affect the population dynamics of disease? *Ecology Letters*, *6* (7), 654–664.

Leelayoova, S., Subrungruang, I., Suputtamongkol, Y., Worapong, J., Petmitr, P. C., & Mungthin, M. (2006). Identification of genotypes of Enterocytozoon bieneusi from stool samples from human immunodeficiency virus-infected patients in Thailand. *Journal of Clinical Microbiology*, *44* (8), 3001–3004.

Lefort, M. C., Wratten, S. D., Cusumano, A., Varennes, Y. D., & Boyer, S. (2017). Disentangling higher trophic level interactions in the cabbage aphid food web using high-throughput DNA sequencing. *Metabarcoding and Metagenomics*, *1*, e13709.

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*, D19–D21.

Leiro, J. M., Piazzon, C., Domínguez, B., Mallo, N., & Lamas, J. (2012). Evaluation of some physical and chemical treatments for inactivating microsporidian spores isolated from fish. *International Journal of Food Microbiology*, *156* (2), 152–160.

Lejeune, J. T., Besser, T. E., Merrill, N. L., Rice, D. H., & Hancock, D. D. (2001). Livestock Drinking Water Microbiology and the Factors Influencing the Quality of Drinking Water Offered to Cattle. *Journal of Dairy Science*, *84*, 1856–1862.

Leonard, C. A. (2013). *Microsporidia Spore Adherence and Host Cell Infection In Vitro*. https://dc.etsu.edu/etd

Lespinet, O., Wolf, Y. I., Koonin, E. V., & Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research*, *12* (7), 1048–1059.

Lewerin, S. S., Sokolova, E., Wahlström, H., Lindström, G., Pers, C., Strömqvist, J., & Sörén, K. (2019). Potential infection of grazing cattle via contaminated water: A theoretical modelling approach. *Animal*, *13* (9).

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26* (5), 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25* (16), 2078–2079.

Li, H., Mendelsohn, E., Zong, C., Zhang, W., Hagan, E., Wang, N., Li, S., Yan, H., Huang, H., Zhu, G., Ross, N., Chmura, A., Terry, P., Fielder, M., Miller, M., Shi, Z., & Daszak, P. (2019). Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. *Biosafety and Health*, *1* (2), 84–90.

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13* (9), 2178–2189.

Li, W., Deng, L., Wu, K., Huang, X., Song, Y., Su, H., Hu, Y., Fu, H., Zhong, Z., & Peng, G. (2017). Presence of zoonotic Cryptosporidium scrofarum, Giardia duodenalis assemblage A and Enterocytozoon bieneusi genotypes in captive Eurasian wild boars (Sus scrofa) in China: Potential for zoonotic transmission. *Parasites and Vectors*, *10* (1), 1–8.

Li, W., Li, Y., Li, W., Yang, J., Song, M., Diao, R., Jia, H., Lu, Y., Zheng, J., Zhang, X., & Xiao, L. (2014). Genotypes of Enterocytozoon bieneusi in

livestock in China: High prevalence and zoonotic potential. *PLoS ONE*, *9* (5).

Li, Y., Shen, H., Zhou, Q., Qian, K., Van Der Lee, T., & Huang, S. (2017). Changing ploidy as a strategy: The Irish potato famine pathogen shifts ploidy in relation to its sexuality. *Molecular Plant-Microbe Interactions*, *30* (1), 45–52.

Lightbody, G., Haberland, V., Browne, F., Taggart, L., Zheng, H., Parkes, E., & Blayney, J. K. (2019). Review of applications of high-throughput sequencing in personalized medicine: Barriers and facilitators of future progress in research and clinical application. *Briefings in Bioinformatics*, *20* (5), 1795–1811.

Linster, M., Van Boheemen, S., De Graaf, M., Schrauwen, E. J. A., Lexmond, P., Mänz, B., Bestebroer, T. M., Baumann, J., Van Riel, D., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., Matrosovich, M., Fouchier, R. A. M., & Herfst, S. (2014). Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell*, *15* (2), P329-339.

Liu, H., Li, M., Cai, S., He, X., Shao, Y., & Lu, X. (2016). Ricin-B-lectin enhances microsporidia Nosema bombycis infection in BmN cells from silkworm Bombyx mori. *Acta Biochimica et Biophysica Sinica*, *48* (11), 1050–1057.

Liu, Y. M., Qiu, L., Sheng, A. Z., Wan, X. Y., Cheng, D. Y., & Huang, J. (2018). Quantitative detection method of Enterocytozoon hepatopenaei using TaqMan probe real-time PCR. *Journal of Invertebrate Pathology*, *151*, 191–196.

Lom, J., & Dykoá, I. (2002). Ultrastructure of Nucleospora secunda n. sp. (Microsporidia), parasite of enterocytes of Nothobranchius rubripinnis. *European Journal of Protistology*, *38* (1), 19–27.

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J., & Jiggins, F. M.

(2014). The Evolution and Genetics of Virus Host Shifts. *PLoS Pathogens*.

Louhi, K. R., Sundberg, L. R., Jokela, J., & Karvonen, A. (2015). Interactions among bacterial strains and fluke genotypes shape virulence of co-infection. *Proceedings of the Royal Society B: Biological Sciences*, *282* (1921).

Loukas, A., & Maizels, R. M. (2000). Helminth C-type lectins and host-parasite interactions. In *Parasitology Today* (pp. P333-339).

Loverdo, C., & Lloyd-Smith, J. O. (2013). Evolutionary Invasion and Escape in the Presence of Deleterious Mutations. *PLoS ONE*, *8* (7), e68179.

Luallen, R. J., Reinke, A. W., Tong, L., Botts, M. R., Félix, M. A., & Troemel, E. R. (2016). Discovery of a Natural Microsporidian Pathogen with a Broad Tissue Tropism in Caenorhabditis elegans. *PLoS Pathogens*, *12* (6), e1005724.

Lynch, R. E., & Lewis, L. C. (1976). Influence on the european com borer of Nosema pyrausta and resistance in maize to sheath-codar feeding. *Environmental Entomology*, *5* (6), 143–146.

Ma, X., Shao, Y., Tian, L., Flasch, D. A., Mulder, H. L., Edmonson, M. N., Liu, Y., Chen, X., Newman, S., Nakitandwe, J., Li, Y., Li, B., Shen, S., Wang, Z., Shurtleff, S., Robison, L. L., Levy, S., Easton, J., & Zhang, J. (2019). Analysis of error profiles in deep next-generation sequencing data. *Genome Biology*, *20* (50).

MacNeil, C., Dick, J. T. A., Hatcher, M. J., Fielding, N. J., Hume, K. D., & Dunn, A. M. (2003). Parasite transmission and cannibalism in an amphipod (Crustacea). *International Journal for Parasitology*, *33* (8), 795–798.

Madden, M. J. L., Young, R. G., Brown, J. W., Miller, S. E., Frewin, A. J., & Hanner, R. H. (2019). Using DNA barcoding to improve invasive pest identification at U.S. Ports-of-entry. *PLoS ONE*, *14* (9), e0222291.

Maddox, J. V., & Solter, L. F. (1996). Long-term storage of infective

microsporidian spores in liquid nitrogen. *Journal of Eukaryotic Microbiology*, *43* (3), 221–225.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47* (W1), W636–W641.

Mahajan-Miklos, S., Rahme, L. G., & Ausubel, F. M. (2000). Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts. In *Molecular Microbiology* (pp. 981–988).

Manteca, X., Mainau, E., & Temple, D. (2013). Stress in Farm Animals: Concept and Effect on Performance. In *The Farm Animal Welfare Fact Sheet*.

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., & Kyrpides, N. C. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, *40* (D1).

Marston, D. A., Horton, D. L., Nunez, J., Ellis, R. J., Orton, R. J., Johnson, N., Banyard, A. C., McElhinney, L. M., Freuling, C. M., Fırat, M., Ünal, N., Müller, T., de Lamballerie, X., & Fooks, A. R. (2017). Genetic analysis of a rabies virus host shift event reveals within-host viral dynamics in a new host. *Virus Evolution*, *3* (2).

Martínez, J. L. (2014). Short-sighted evolution of bacterial opportunistic pathogens with an environmental origin. In *Frontiers in Microbiology* (p. 239).

Mathis, A., Breitenmoser, A. C., & Deplazes, P. (1999). Detection of new Enterocytozoon genotypes in faecal samples of farm dogs and a cat. *Parasite (Paris, France)*, *6* (2), 189–193.

Matos, O., Lobo, M. L., & Xiao, L. (2012). Epidemiology of Enterocytozoon

bieneusi infection in humans. In *Journal of Parasitology Research* (Vol. 2012, pp. 36–45).

McCormack, J. E., Maley, J. M., Hird, S. M., Derryberry, E. P., Graves, G. R., & Brumfield, R. T. (2012). Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, *62* (1), 397–406.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20* (9), 1297–1303.

McNamara-Bordewick, N. K., McKinstry, M., & Snow, J. W. (2019). Robust Transcriptional Response to Heat Shock Impacting Diverse Cellular Processes despite Lack of Heat Shock Factor in Microsporidia. *MSphere*, *4* (3).

MG, F., M, G., KA, G., D, S., P, C., D, W., J, M., & D, T. (2015). Modelling Infectious Hematopoietic Necrosis Virus Dispersion from Marine Salmon Farms in the Discovery Islands, British Columbia, Canada. *PloS One*, *10* (6).

Mikhailov, K. V, Simdyanov, T. G., Aleoshin, V. V, & Belozersky, A. N. (2016). Genomic survey of a hyperparasitic microsporidian Amphiamblys sp. (Metchnikovellidae) Genome Biology and Evolution Advance Access. *At University of Library and Information Science On*, *9* (3), 454–467.

Minardi, D., Bateman, K. S., Kuzdzal, A., Stone, M., Avant, J., Condliffe, R., Brotherton, P., Laverick, M., Sritunyalucksana, K., Itsathitphaisarn, O., Baoprasertkul, P., & Stentiford, G. D. (2019). Testing of a pond-side molecular diagnostic tool for the detection of white spot syndrome virus in shrimp aquaculture. *Journal of the World Aquaculture Society*, *50* (1), 18–33.

Mittleider, D., Green, L. C., Mann, V. H., Michael, S. F., Didier, E. S., & Brindley, P. J. (2002). Sequence survey of the genome of the opportunistic microsporidian pathogen, Vittaforma corneae. *Journal of Eukaryotic Microbiology*, *49* (5), 393–401.

Mollentze, N., & Streicker, D. G. (2020). Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences of the United States of America*, *117* (17), 9423–9430.

Moodie, E. G., Le Jambre, L. F., & Katz, M. E. (2003). Thelohania montirivulorum sp. nov. (Microspora: Thelohaniidae), a parasite of the Australian freshwater crayfish, Cherax destructor (Decapoda: Parastacidae): Fine ultrastructure, molecular characteristics and phylogenetic relationships. *Parasitology Research*, *91* (3), 215–228.

Moore, J. (2010). Intermediate Host Behavior. *Encyclopedia of Animal Behavior*, 186–190.

Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., Rocha, B. C., Hall-Mendelin, S., Day, A., Riegler, M., Hugo, L. E., Johnson, K. N., Kay, B. H., McGraw, E. A., van den Hurk, A. F., Ryan, P. A., & O'Neill, S. L. (2009). A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium. *Cell*, *139* (7), 1268–1278.

Mrochen, D. M., Schulz, D., Fischer, S., Jeske, K., El Gohary, H., Reil, D., Imholt, C., Trübe, P., Suchomel, J., Tricaud, E., Jacob, J., Heroldová, M., Bröker, B. M., Strommenger, B., Walther, B., Ulrich, R. G., & Holtfreter, S. (2018). Wild rodents and shrews are natural hosts of Staphylococcus aureus. *International Journal of Medical Microbiology*, *308* (6), 590–597.

Mughini-Gras, L., van Pelt, W., van der Voort, M., Heck, M., Friesema, I., & Franz, E. (2018). Attribution of human infections with Shiga toxin-producing Escherichia coli (STEC) to livestock sources and identification of source-

specific risk factors, The Netherlands (2010–2014). *Zoonoses and Public Health*, *65* (1), e8–e22.

Mullins, J. E., Powell, M., Speare, D. J., & Cawthorn, R. (1994). An intranuclear microsporidian in lumpfish Cyclopterus lumpus. *Diseases of Aquatic Organisms*, *20* (1), 7–13.

Mylonas, C. C., Robles, R., Tacken, G., Banovi, M., Krystallis, A., Guerrero, L., & Grigorakis, K. (2019). New species for EU aquaculture. *Food Science and Technology*, *33* (2), 22–26.

Nadelman, D. A., Bradt, A. R., Qvarnstrom, Y., Goldsmith, C. S., Zaki, S. R., Wang, F., Smith, E. H., & Fullen, D. R. (2020). Cutaneous microsporidiosis in an immunosuppressed patient. *Journal of Cutaneous Pathology*, *47* (7), 659–663.

Naghashyan, H. Z., Hakobyan, A. R., & Martirosyan, A. A. (2018). A new disease in fish farms of Armenia. *Annals of Agrarian Science*, *16* (4), 410–412.

Nagpal, A., Pritt, B. S., Lorenz, E. C., Amer, H., Nasr, S. H., Cornell, L. D., Iqbal, S., & Wilhelm, M. P. (2013). Disseminated microsporidiosis in a renal transplant recipient: Case report and review of the literature. *Transplant Infectious Disease*, *15* (5), 526–532.

Nakjang, S., Williams, T. A., Heinz, E., Watson, A. K., Foster, P. G., Sendra, K. M., Heaps, S. E., Hirt, R. P., & Embley, T. M. (2013). Reduction and expansion inmicrosporidian genome evolution: New insights from comparative genomics. *Genome Biology and Evolution*, *5* (12), 2285–2303.

National Center for Biotechnology Information. (2020). *Sequence Read Archive : ERR1679586*. NCBI Handout Series. https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR1679586

Navas, A., Cobas, G., Talavera, M., Ayala, J. A., López, J. A., & Martínez, J. L. (2007). Experimental validation of Haldane's hypothesis on the role of

infection as an evolutionary force for Metazoans. *Proceedings of the National Academy of Sciences of the United States of America*, *104* (34), 13728–13731.

Navaud, O., Barbacci, A., Taylor, A., Clarkson, J. P., & Raffaele, S. (2018). Shifts in diversification rates and host jump frequencies shaped the diversity of host range among Sclerotiniaceae fungal plant pathogens. *Molecular Ecology*, *27* (5), 1309–1323.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32* (1), 268–274.

Nilsson, A. I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. D., & Andersson, D. I. (2005). Bacterial genome size reduction by experimental evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *102* (34), 12112–12116.

Nishimura, D. (2000). Sequencher 3.1.1. *Biotech Software & Internet Report*, *1* (1–2), 24–30.

Nott, F. J., & Beale, P. B. (1968). River Exe Fisheries Survey 1966-1967. *Devon River Authority*.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27* (5), 824–834.

Nylund, S., Andersen, L., Sævareid, I., Plarre, H., Watanabe, K., Arnesen, C. E., Karlsbakk, E., & Nylund, A. (2011). Diseases of farmed Atlantic salmon Salmo salar associated with infections by the microsporidian Paranucleospora theridion. *Diseases of Aquatic Organisms*, *94* (1), 41–57.

Nylund, S., Nylund, A., Watanabe, K., Arnesen, C. E., & Karlsbakk, E. (2010). Paranucleospora theridion n. gen., n. sp. (Microsporidia, Enterocytozoonidae) with a Life Cycle in the Salmon Louse

(Lepeophtheirus salmonis, Copepoda) and Atlantic Salmon (Salmo salar). *Journal of Eukaryotic Microbiology*, *57* (2), 95–114.

Oidtmann, B., Dixon, P., Way, K., Joiner, C., & Bayley, A. E. (2018). Risk of waterborne virus spread – review of survival of relevant fish and crustacean viruses in the aquatic environment and implications for control measures. In *Reviews in Aquaculture* (Vol. 10, Issue 3, pp. 641–669). John Wiley & Sons, Ltd.

Okamura, B., Hartikainen, H., & Trew, J. (2019). Waterbird-mediated dispersal and freshwater biodiversity: General insights from bryozoans. *Frontiers in Ecology and Evolution*, 29.

Otranto, D., Cantacessi, C., Testini, G., & Lia, R. P. (2006). Phortica variegata as an intermediate host of Thelazia callipaeda under natural conditions: Evidence for pathogen transmission by a male arthropod vector. *International Journal for Parasitology*, *36* (10–11), 1167–1173.

Overstreet, R. M. (1988). Aquatic pollution problems, Southeastern U.S. coasts: histopathological indicators. *Aquatic Toxicology*, *11* (3–4), 213–239.

Overstreet, R. M., & Howse, H. D. (1977). Some Parasites and Diseases of Estuarine Fishes in Polluted Habititats of Mississippi. *Annals of the New York Academy of Sciences*.

Pal, A., & Chakravarty, A. K. (2020). Disease resistance for different livestock species. In *Genetics and Breeding for Disease Resistance of Livestock* (pp. 271–296).

Palenzuela, O., Redondo, M. J., Cali, A., Takvorian, P. M., Alonso-Naveiro, M., Alvarez-Pellitero, P., & Sitjà-Bobadilla, A. (2014). A new intranuclear microsporidium, Enterospora nucleophila n. sp., causing an emaciative syndrome in a piscine host (Sparus aurata), prompts the redescription of the family Enterocytozoonidae. *International Journal for Parasitology*, *44* (3–4), 189–203.

Pallen, M. J. (2014). Diagnostic metagenomics: Potential applications to bacterial, viral and parasitic infections. *Parasitology*, *141* (14), 1856–1862.

Pamer, E. G. (2007). Immune responses to commensal and environmental microbes. *Nature Immunology*, *8*, 1173–1178.

Pan, G., Xu, J., Li, T., Xia, Q., Liu, S. L., Zhang, G., Li, S., Li, C., Liu, H., Yang, L., Liu, T., Zhang, X., Wu, Z., Fan, W., Dang, X., Xiang, H., Tao, M., Li, Y., Hu, J., … Zhou, Z. (2013). Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics*, *14* (1), 1–14.

Patmasiriwat, D., Kuik, O., & Pednekar, S. (1998). The Shrimp Aquaculture Sector in Thailand: A Review of Economic, Environmental and Trade Issues. *CREED Working Paper 19*, 1–41. http://www.iied.org/creed

Peacock, S. J., Bateman, A. W., Connors, B., Godwin, S., Lewis, M. A., & Krkošek, M. (2019). Ecology of a marine ectoparasite in farmed and wild salmon. In *Wildlife Disease Ecology: Linking Theory to Data and Application* (pp. 544–573). Cambridge University Press.

Peeler, E. (2005). The Role of Risk Analysis and Epidemiology in the Development of Biosecurity for Aquaculture. *Fisheries (Bethesda)*, *5*, 35–45.

Peeler, E. J., Reese, R. A., & Thrush, M. A. (2015). Animal Disease Import Risk Analysis - a Review of Current Methods and Practice. In *Transboundary and Emerging Diseases* (pp. 480–490).

Peterman, M. A., & Posadas, B. C. (2019). Direct Economic Impact of Fish Diseases on the East Mississippi Catfish Industry. *North American Journal of Aquaculture*, *81* (3), 222–229.

Peters, L., Spatharis, S., Dario, M. A., Dwyer, T., Roca, I. J. T., Kintner, A., Kanstad-Hanssen, Ø., Llewellyn, M. S., & Praebel, K. (2018). Environmental DNA: A new low-cost monitoring tool for pathogens in

salmonid aquaculture. *Frontiers in Microbiology*, *0*, 3009.

Petr, M., Vernot, B., & Kelso, J. (2019). Admixr-R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, *35* (17), 3194–3195.

Petri, W. A., Haque, R., & Mann, B. J. (2002). The bittersweet interface of parasite and host: Lectin-carbohydrate interactions during human invasion by the parasite Entamoeba histolytica. *Annual Review of Microbiology*, *56*, 39–64.

Pettis, J. S., Vanengelsdorp, D., Johnson, J., & Dively, G. (2012). Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. *Naturwissenschaften*, *99* (2), 153–158.

Peuvel, I., Peyret, P., Méténier, G., Vivarès, C. P., & Delbac, F. (2002). The microsporidian polar tube: Evidence for a third polar tube protein (PTP3) in Encephalitozoon cuniculi. *Molecular and Biochemical Parasitology*, *122* (1), 69–80.

Peyretaillade, E., Parisot, N., Polonais, V., Terrat, S., Denonfoux, J., Dugat-Bony, E., Wawrzyniak, I., Biderre-Petit, C., Mahul, A., Rimour, S., Gonçalves, O., Bornes, S., Delbac, F., Chebance, B., Duprat, S., Samson, G., Katinka, M., Weissenbach, J., Wincker, P., & Peyret, P. (2012). Annotation of microsporidian genomes using transcriptional signals. *Nature Communications*, *3*, 1137.

Pimentel, A. C., Beraldo, C. S., & Cogni, R. (2021). Host-shift as the cause of emerging infectious diseases: Experimental approaches using drosophila-virus interactions. *Genetics and Molecular Biology*, *44* (1), e20200197.

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. In *Nature Reviews Microbiology* (Vol. 15, Issue 8).

Plymouth Gov. (n.d.). *Tamar-Tavy Estuary (Ecological SSSI)* | *PLYMOUTH.GOV.UK*. Retrieved January 27, 2021, from

https://www.plymouth.gov.uk/parksnatureandgreenspaces/sitesspecialscientificinterest/tamartavyestuaryecologicalsssi

Poley, J. D., Sutherland, B. J. G., Fast, M. D., Koop, B. F., & Jones, S. R. M. (2017). Effects of the vertically transmitted microsporidian Facilispora margolisi and the parasiticide emamectin benzoate on salmon lice (Lepeophtheirus salmonis). *BMC Genomics*, *18* (1), 630.

Pombert, J. F., Haag, K. L., Beidas, S., Ebert, D., & Keeling, P. J. (2015). The Ordospora colligata genome: Evolution of extreme reduction in microsporidia and host-to-parasite horizontal gene transfer. *MBio*, *66* (1).

Pombert, J. F., Selman, M., Burki, F., Bardell, F. T., Farinelli, L., Solter, L. F., Whitman, D. W., Weiss, L. M., Corradi, N., & Keeling, P. J. (2012). Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *109* (31), 12638–12643.

Pombert, J. F., Xu, J., Smith, D. R., Heiman, D., Young, S., Cuomo, C. A., Weiss, L. M., & Keeling, P. J. (2013). Complete genome sequences from three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian Encephalitozoon cuniculi. *Eukaryotic Cell*, *12* (4), 503–511.

Popgeorgiev, N., Temmam, S., Raoult, D., & Desnues, C. (2013). Describing the silent human virome with an emphasis on giant viruses. *Intervirology*, *56* (6), 395–412.

Potapov, V., & Ong, J. L. (2017). Examining sources of error in PCR by single-molecule sequencing. *PLoS ONE*, *12* (1).

Power, A. G., & Mitchell, C. E. (2004). Pathogen spillover in disease epidemics. *American Naturalist*, *164*.

Prempeh, H. (2001). Foot and mouth disease: the human consequences. *BMJ*,

322, 565–566.

Prunier, J., Verta, J. P., & Mackay, J. J. (2016). Conifer genomics and adaptation: At the crossroads of genetic diversity and genome function. *New Phytologist*, *209* (1), 44–62.

Pulidindi, K., & Pandey, H. (2020). *U.S. Crustacean Market size to exceed $10.2bn by 2026*. Global Market Insights. https://www.gminsights.com/pressrelease/us-crustacean-market?utm_source=globenewswire.com&utm_medium=referral&utm_campaign=Paid_globenewswire

Purcell, R. V., Visnovska, M., Biggs, P. J., Schmeier, S., & Frizelle, F. A. (2017). Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Scientific Reports*, *7* (1), 1–12.

PUTZ, R. E., HOFFMAN, G. L., & DUNBAR, C. E. (1965). Two New Species of Plistophora (Microsporidea) from North American Fish with a Synopsis of Microsporidea of Freshwater and Euryhaline Fishes. *The Journal of Protozoology*, *12* (2), 228–236.

Quandt, C. A., Beaudet, D., Corsaro, D., Walochnik, J., Michel, R., Corradi, N., & James, T. Y. (2017). The genome of an intranuclear parasite, Paramicrosporidium saccamoebae, reveals alternative adaptations to obligate intracellular parasitism. *ELife*, *6*.

Quiles, A., Bacela-Spychalska, K., Teixeira, M., Lambin, N., Grabowski, M., Rigaud, T., & Wattier, R. A. (2019). Microsporidian infections in the species complex Gammarus roeselii (Amphipoda) over its geographical range: Evidence for both host-parasite co-diversification and recent host shifts. *Parasites and Vectors*, *12* (1), 1–20.

Rajendran, K. V., Shivam, S., Ezhil Praveena, P., Joseph Sahaya Rajan, J., Sathish Kumar, T., Avunje, S., Jagadeesan, V., Prasad Babu, S. V. A. N. V., Pande, A., Navaneeth Krishnan, A., Alavandi, S. V., & Vijayan, K. K. (2016a). Emergence of Enterocytozoon hepatopenaei (EHP) in farmed

Penaeus (Litopenaeus) vannamei in India. *Aquaculture*, *454*, 272–280.

Rajendran, K. V., Shivam, S., Ezhil Praveena, P., Joseph Sahaya Rajan, J., Sathish Kumar, T., Avunje, S., Jagadeesan, V., Prasad Babu, S. V. A. N. V., Pande, A., Navaneeth Krishnan, A., Alavandi, S. V., & Vijayan, K. K. (2016b). Emergence of Enterocytozoon hepatopenaei (EHP) in farmed Penaeus (Litopenaeus) vannamei in India. *Aquaculture*, *454*, 272–280.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Bioliology*, *67* (5), 901–904.

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11* (1), 1–10.

Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., & Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: What are the implications for malaria control? *Trends in Parasitology*, *27* (2), 91–98.

Reinke, A. W., Balla, K. M., Bennett, E. J., & Troemel, E. R. (2017). Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. *Nature Communications*, *8* (1), 1–11.

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, *58* (4), 586–597.

Richards, T. A., Hirt, R. P., Williams, B. A. P., & Embley, T. M. (2003). Horizontal gene transfer and the evolution of parasitic protozoa. *Protist*, *154* (1), 17–32.

Richardson, R. T., Bengtsson-Palme, J., & Johnson, R. M. (2017). Evaluating and optimizing the performance of software commonly used for the taxonomic classification of DNA metabarcoding sequence data. *Molecular Ecology Resources*, *17* (4), 760–769.

Rinder, H., Thomschke, A., Dengjel, B., Gothe, R., Löscher, T., & Zahler, M.

(2000). Close genotypic relationship between Enterocytozoon bieneusi from humans and pigs and first detection in cattle. *J Parasitol*, *86* (1), 185–188.

Robbins, K. M., Maslakova, S. A., & Dassow, G. von. (2021). A New Genus Of Microsporidian Parasite (Hepatosporidae; Micro-Sporidia) Found In The Oocytes Of Ribbon Worms From The North Pacific Genus Maculaura (Heteronemertea; Nemertea). *BioRxiv*, 2021.04.04.438387.

Rokunuzzaman, M., Hayakawa, A., Yamane, S., Tanaka, S., & Ohnishi, K. (2016). Effect of soil disinfection with chemical and biological methods on bacterial communities. *Egyptian Journal of Basic and Applied Sciences*, *3* (2), 141–148.

Rosenberg, N. A., & Tao, R. (2008). Discordance of species trees with their most likely gene trees: The case of five taxa. *Systematic Biology*, *57* (1), 131–140.

RTFA (River Taw Fishing Association). (2020). *River Taw Fish Records - RTFA*. http://www.rivertawfisheries.co.uk/html/fishing_records.html

Sak, B., Kučerová, Z., Kváč, M., Květoňová, D., Rost, M., & Secor, E. W. (2010). Seropositivity for enterocytozoon bieneusi, Czech Republic. *Emerging Infectious Diseases*, *16* (2), 335–337.

Sakai, M., Baxa, D. V., Kurobe, T., Kono, T., Shivappa, R. B., & Hedrick, R. P. (2009). Detection of Nucleospora salmonis in cutthroat trout (Oncorhynchus clarki) and rainbow trout (Oncorhynchus mykiss) by loop-mediated isothermal amplification. *Aquaculture*, *288* (1–2), 27–31.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A., & Moreillon, P. (2011). Staphylococcus aureus host range and human-bovine host shift. *Applied and Environmental Microbiology*, *77* (17), 5908–5915.

Salachan, P. V., Jaroenlak, P., Thitamadee, S., Itsathitphaisarn, O., & Sritunyalucksana, K. (2017a). Laboratory cohabitation challenge model for

shrimp hepatopancreatic microsporidiosis (HPM) caused by Enterocytozoon hepatopenaei (EHP). *BMC Veterinary Research*, *13* (1), 1–7.

Salachan, P. V., Jaroenlak, P., Thitamadee, S., Itsathitphaisarn, O., & Sritunyalucksana, K. (2017b). Laboratory cohabitation challenge model for shrimp hepatopancreatic microsporidiosis (HPM) caused by Enterocytozoon hepatopenaei (EHP). *BMC Veterinary Research*, *13* (1), 1–7.

Sánchez-Cordón, P. J., Montoya, M., Reis, A. L., & Dixon, L. K. (2018). African swine fever: A re-emerging viral disease threatening the global pig industry. *Veterinary Journal*, *223*, 41–48.

Santín, M., & Fayer, R. (2011). Microsporidiosis: Enterocytozoon bieneusi in domesticated and wild animals. In *Research in Veterinary Science* (Vol. 90, Issue 3, pp. 363–371).

Saupe, E. E., Myers, C. E., Townsend Peterson, A., Soberón, J., Singarayer, J., Valdes, P., & Qiao, H. (2019). Spatio-temporal climate change contributes to latitudinal diversity gradients. *Nature Ecology and Evolution*, *3*, 1419–1429.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*, *3*, 430–439.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research*, *43* (6), e27.

Schulz, F., Alteio, L., Goudeau, D., Ryan, E. M., Yu, F. B., Malmstrom, R. R., Blanchard, J., & Woyke, T. (2018). Hidden diversity of soil giant viruses. *Nature Communications*, *9* (1), 1–9.

Selman, M. (2014). Genomic Analysis of Encephalitozoon species. In *PhD*

Thesis. uOttawa.

Selman, M., Sak, B., Kváč, M., Farinelli, L., Weiss, L. M., & Corradi, N. (2013). Extremely reduced levels of heterozygosity in the vertebrate pathogen Encephalitozoon cuniculi. *Eukaryotic Cell*, *12* (4), 496–502.

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, *1962*, 227–245.

Shadduck, J. A., Meccoli, R. A., Davis, R., & Font, R. L. (1990). Isolation of a microsporidian from a human patient. *Journal of Infectious Diseases*, *162* (3), 773–776.

Shane, J. L., Grogan, C. L., Cwalina, C., & Lampe, D. J. (2018). Blood meal-induced inhibition of vector-borne disease by transgenic microbiota. *Nature Communications*, *9* (1), 1–10.

Sharma, S., Balne, P. K., & Das, S. (2014). Ocular Microsporidiosis. In *Microsporidia: Pathogens of Opportunity: First Edition* (pp. 403–419).

Shea, D., Bateman, A., Li, S., Tabata, A., Schulze, A., Mordecai, G., Ogston, L., Volpe, J. P., Frazer, L. N., Connors, B., Miller, K. M., Short, S., & Krkošek, M. (2020). Environmental DNA from multiple pathogens is elevated near active Atlantic salmon farms. *Proceedings of the Royal Society B*, *287* (1937), 20202010.

Shen, H., Zang, Y., Song, K., Ma, Y., Dai, T., & Serwadda, A. (2017). A meta-transcriptomics survey reveals changes in the microbiota of the Chinese mitten crab Eriocheir sinensis infected with hepatopancreatic necrosis disease. *Frontiers in Microbiology*, 732.

Sheridan, P. O., Raguideau, S., Quince, C., Holden, J., Zhang, L., Gaze, W. H., Holden, J., Mead, A., Raguideau, S., Quince, C., Singer, A. C., Wellington, E. M. H., Zhang, L., Williams, T. A., & Gubry-Rangin, C. (2020). Gene duplication drives genome expansion in a major lineage of

Thaumarchaeota. *Nature Communications*, *11* (1), 1–12.

Sheu, S. Y., Chiu, T. F., Young, C. C., Arun, A. B., & Chen, W. M. (2011). Flavobacterium macrobrachii sp. nov., isolated from a freshwater shrimp culture pond. *International Journal of Systematic and Evolutionary Microbiology*, *61* (6), 1402–1407.

Shewan, A., Eastburn, D. J., & Mostov, K. (2011). Phosphoinositides in cell architecture. *Cold Spring Harbor Perspectives in Biology*, *3* (8), a004796.

Shi, K., Li, M., Wang, X., Li, J., Karim, M. R., Wang, R., Zhang, L., Jian, F., & Ning, C. (2016). Molecular survey of Enterocytozoon bieneusi in sheep and goats in China. *Parasites and Vectors*, *9* (1).

Shimabukuro, P. H. F., Moreira, J. A. C., & Costa, T. S. da. (2016). Sand Flies (Diptera: Psychodidae) From the Brazilian Atlantic Forest Domain Collected With Malaise Traps. *Journal of Medical Entomology*, *53* (6), 1488–1491.

Shinn, A. A. P., Pratoomyot, J., Bron, J. E., Paladini, G. G., Brooker, E., & Brooker, A. J. (2015). Economic impacts of aquatic parasites on global finfish production. *Global Aquaculture Advocate*.

Snowden, K. F., Lewis, B. C., Hoffman, J., & Mansell, J. (2009). Encephalitozoon cuniculi infections in dogs: A case series. *Journal of the American Animal Hospital Association*, *45* (5), 225–231.

Snowden, K., Logan, K., & Didier, E. S. (1999). Encephalitozoon cuniculi strain III is a cause of encephalitozoonosis in both humans and dogs. *Journal of Infectious Diseases*, *180* (6), 2086–2088.

Sokolova, Y., & Hawke, J. P. (2016). Perezia nelsoni (Microsporidia) in Agmasoma penaeiinfected Atlantic white shrimp Litopenaeus setiferus (Paenaidae, Decapoda) and phylogenetic analysis of Perezia spp. complex. *Protistology*, *10* (3).

Solter, L. F., Siegel, J. P., Pilarska, D. K., & Higgs, M. C. (2002). The impact of mixed infection of three species of microsporidia isolated from the gypsy

moth, Lymantria dispar L. (Lepidoptera: Lymantriidae). *Journal of Invertebrate Pathology*, *81* (2), 103–113.

Soto-Rodriguez, S. A., Gomez-Gil, B., Lozano-Olvera, R., Betancourt-Lozano, M., & Morales-Covarrubias, M. S. (2015). Field and experimental evidence of Vibrio parahaemolyticus as the causative agent of acute hepatopancreatic necrosis disease of cultured shrimp (Litopenaeus vannamei) in northwestern Mexico. *Applied and Environmental Microbiology*, *81* (5), 1689–1699.

Southern, T. R., Jolly, C. E., Lester, M. E., & Hayman, J. R. (2007). EnP1, a microsporidian spore wall protein that enables spores to adhere to and infect host cells in vitro. *Eukaryotic Cell*, *6* (8), 1354–1362.

Sprague, V. (1965). Nosema sp. (Microsporida, Nosematidae) in the Musculature of the Crab Callinectes sapidus. *The Journal of Protozoology*, *12* (1), 66–70.

Sprague, V. (1970). Some protozoan parasites and hyperparasites in marine decapod Crustacea. *American Fisheries Society Special Publication*, 511–526.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30* (9), 1312–1313.

Stamatakis, A. (2016). The RAxML v8.2 Manual. *Manual/Tutorial*.

Stentiford, G. D. (2012). Diseases in aquatic crustaceans: Problems and solutions for global food security. *Journal of Invertebrate Pathology*, *110* (2), 139.

Stentiford, G. D., Bass, D., & Williams, B. A. P. (2019). Ultimate opportunists—the emergent enterocytozoon group microsporidia. *PLoS Pathogens*, *15* (5), e1007668.

Stentiford, G. D., Bateman, K. S., Dubuffet, A., Chambers, E., & Stone, D. M. (2011). Hepatospora eriocheir (Wang and Chen, 2007) gen. et comb. nov.

infecting invasive Chinese mitten crabs (Eriocheir sinensis) in Europe. *Journal of Invertebrate Pathology*, *108* (3), 156–166.

Stentiford, G. D., Bateman, K. S., Longshaw, M., & Feist, S. W. (2007). Enterospora canceri n. gen., n. sp., intranuclear within the hepatopancreatocytes of the European edible crab Cancer pagurus. *Diseases of Aquatic Organisms*, *75* (1), 61–72.

Stentiford, G. D., Becnel, J. J., Weiss, L. M., Keeling, P. J., Didier, E. S., Williams, B. A. P., Bjornson, S., Kent, M. L., Freeman, M. A., Brown, M. J. F., Troemel, E. R., Roesel, K., Sokolova, Y., Snowden, K. F., & Solter, L. (2016). Microsporidia – Emergent Pathogens in the Global Food Chain. In *Trends in Parasitology* (Vol. 32, Issue 4, pp. 336–348).

Stentiford, G. D., Feist, S. W., Stone, D. M., Bateman, K. S., & Dunn, A. M. (2013). Microsporidia: Diverse, dynamic, and emergent pathogens in aquatic systems. *Trends in Parasitology*, *29* (11), 567–578.

Stentiford, G. D., Feist, S. W., Stone, D. M., Peeler, E. J., & Bass, D. (2014). Policy, phylogeny, and the parasite. *Trends in Parasitology*, *30* (6), 274–281.

Stentiford, G. D., Ross, S., Minardi, D., Feist, S. W., Bateman, K. S., Gainey, P. A., Troman, C., & Bass, D. (2018). Evidence for trophic transfer of Inodosporus octospora and Ovipleistophora arlo n. sp. (Microsporidia) between crustacean and fish hosts. *Parasitology*, *145* (8), 1105–1117.

Stentiford, G. D., Sritunyalucksana, K., Flegel, T. W., Williams, B. A. P., Withyachumnarnkul, B., Itsathitphaisarn, O., & Bass, D. (2017). New Paradigms to Help Solve the Global Aquaculture Disease Crisis. *PLoS Pathogens*, *13* (2).

Stentiford, G., Ramilo, A., Abollo, E., Kerr, R., Bateman, K., Feist, S., Bass, D., & Villalba, A. (2017). Hyperspora aquatica n.gn., n.sp. (Microsporidia), hyperparasitic in Marteilia cochillia (Paramyxida), is closely related to crustacean-infecting microspordian taxa. *Parasitology*, *144* (2), 186–199.

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11* (7).

Studholme, D. J., Glover, R. H., & Boonham, N. (2011). Application of high-throughput DNA sequencing in phytopathology. *Annual Review of Phytopathology*, *49*, 87–105.

Subasinghe, R., Soto, D., & Jia, J. (2009). Global aquaculture and its role in sustainable development. *Reviews in Aquaculture*, *1* (1), 2–9.

Sundberg, L. R., Ketola, T., Laanto, E., Kinnula, H., Bamford, J. K. H., Penttinen, R., & Mappes, J. (2016). Intensive aquaculture selects for increased virulence and interference competition in bacteria. *Proceedings of the Royal Society B: Biological Sciences*, *283* (1826).

Sveen, S., Øverland, H., Karlsbakk, E., & Nylund, A. (2012). Paranucleospora theridion (Microsporidia) infection dynamics in farmed Atlantic salmon Salmo salar put to sea in spring and autumn. *Diseases of Aquatic Organisms*, *101* (1), 43–49.

Szöllosi, G. J., Rosikiewicz, W., Boussau, B., Tannier, E., & Daubin, V. (2013). Efficient exploration of the space of reconciled gene trees. *Systematic Biology*, *62* (6), 901–912.

Tabatabaie, F., Abrehdari Tafreshi, Z., Shahmohammad, N., & Pirestani, M. (2015). Molecular detection of microsporidiosis in various samples of Iranian immunocompromised patients. *Journal of Parasitic Diseases*, *39* (4), 634–638.

Tajima, F., & Nei, M. (1984). Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, *1* (3), 269–285.

Tams, R. N., Cassilly, C. D., Anaokar, S., Brewer, W. T., Dinsmore, J. T., Chen, Y. L., Patton-Vogt, J., & Reynolds, T. B. (2019). Overproduction of phospholipids by the kennedy pathway leads to hypervirulence in Candida albicans. *Frontiers in Microbiology*, *10* (FEB).

Tan, J. S. (1997). Human zoonotic infections transmitted by dogs and cats. *Archives of Internal Medicine*, *157* (17), 1933–1943.

Tang, K. F. J., Aranguren, L. F., Piamsomboon, P., Han, J. E., Maskaykina, I. Y., & Schmidt, M. M. (2017). Detection of the microsporidian Enterocytozoon hepatopenaei (EHP) and Taura syndrome virus in Penaeus vannamei cultured in Venezuela. *Aquaculture*, *480*, 17–21.

Tang, K. F. J., Han, J. E., Aranguren, L. F., White-Noble, B., Schmidt, M. M., Piamsomboon, P., Risdiana, E., & Hanggono, B. (2016a). Dense populations of the microsporidian Enterocytozoon hepatopenaei (EHP) in feces of Penaeus vannamei exhibiting white feces syndrome and pathways of their transmission to healthy shrimp. *Journal of Invertebrate Pathology*.

Tang, K. F. J., Han, J. E., Aranguren, L. F., White-Noble, B., Schmidt, M. M., Piamsomboon, P., Risdiana, E., & Hanggono, B. (2016b). Dense populations of the microsporidian Enterocytozoon hepatopenaei (EHP) in feces of Penaeus vannamei exhibiting white feces syndrome and pathways of their transmission to healthy shrimp. *Journal of Invertebrate Pathology*, *140*, 1–7.

Tang, K. F. J., Pantoja, C. R., Redman, R. M., Han, J. E., Tran, L. H., & Lightner, D. V. (2015a). Development of in situ hybridization and PCR assays for the detection of Enterocytozoon hepatopenaei (EHP), a microsporidian parasite infecting penaeid shrimp. *Journal of Invertebrate Pathology*.

Tang, K. F. J., Pantoja, C. R., Redman, R. M., Han, J. E., Tran, L. H., & Lightner, D. V. (2015b). Development of in situ hybridization and PCR assays for the detection of Enterocytozoon hepatopenaei (EHP), a microsporidian parasite infecting penaeid shrimp. *Journal of Invertebrate Pathology*.

Tangprasittipap, A., Srisala, J., Chouwdee, S., Somboon, M., Chuchird, N., Limsuwan, C., Srisuvan, T., Flegel, T. W., & Sritunyalucksana, K. (2013a).

The microsporidian Enterocytozoon hepatopenaei is not the cause of white feces syndrome in whiteleg shrimp Penaeus (Litopenaeus) vannamei. *BMC Veterinary Research*, *9*.

Tangprasittipap, A., Srisala, J., Chouwdee, S., Somboon, M., Chuchird, N., Limsuwan, C., Srisuvan, T., Flegel, T. W., & Sritunyalucksana, K. (2013b). The microsporidian Enterocytozoon hepatopenaei is not the cause of white feces syndrome in whiteleg shrimp Penaeus (Litopenaeus) vannamei. *BMC Veterinary Research*.

Taur, Y., & Pamer, E. G. (2013). The intestinal microbiota and susceptibility to infection in immunocompromised patients. In *Current Opinion in Infectious Diseases*.

Teachey, D. T., Russo, P., Orenstein, J. M., Didier, E. S., Bowers, C., & Bunin, N. (2004). Pulmonary infection with microsporidia after allogeneic bone marrow transplant. *Bone Marrow Transplantation*.

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, R. H., Hildebrand, F., Bork, P., & Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*.

Thammavongsa, V., Kim, H. K., Missiakas, D., & Schneewind, O. (2015). Staphylococcal manipulation of host immune responses. In *Nature Reviews Microbiology*.

Thomson, G. R., Vosloo, W., & Bastos, A. D. S. (2003). Foot and mouth disease in wildlife. In *Virus Research*.

Thorstad, E. B., & Finstad, B. (2018). Impacts of salmon lice emanating from salmon farms on wild Atlantic salmon and sea trout. *NINA Report*.

Thorstad, E. B., Todd, C. D., Uglem, I., Bjørn, P. A., Gargan, P. G., Vollset, K. W., Halttunen, E., Kålås, S., Berg, M., & Finstad, B. (2015). Effects of salmon lice Lepeophtheirus salmonis on wild sea trout salmo trutta - A

literature review. In *Aquaculture Environment Interactions*.

Tomley, F. M., & Shirley, M. W. (2009). Livestock infectious diseases and zoonoses. In *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Torruella, G., De Mendoza, A., Grau-Bové, X., Antó, M., Chaplin, M. A., Del Campo, J., Eme, L., Pérez-Cordón, G., Whipps, C. M., Nichols, K. M., Paley, R., Roger, A. J., Sitjà-Bobadilla, A., Donachie, S., & Ruiz-Trillo, I. (2015). Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology*.

Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S. A., Burns, J. A., Sebé-Pedrós, A., Völcker, E., & López-García, P. (2018). Global transcriptome analysis of the aphelid Paraphelidium tribonemae supports the phagotrophic origin of fungi. *Communications Biology*.

Tourtip, S., Wongtripop, S., Stentiford, G. D., Bateman, K. S., Sriurairatana, S., Chavadej, J., Sritunyalucksana, K., & Withyachumnarnkul, B. (2009). Enterocytozoon hepatopenaei sp. nov. (Microsporida: Enterocytozoonidae), a parasite of the black tiger shrimp Penaeus monodon (Decapoda: Penaeidae): Fine structure and phylogenetic relationships. *Journal of Invertebrate Pathology*, *102* (1), 21–29.

Tourtip S. (2005). Histology, ultrastructure and molecular biology of a new microsporidium infecting the black tiger shrimp Penaeus monodon. *Department of Anatomy, Faculty of Science. Mahidol University, Bangkok*.

Treonis, A. M., Unangst, S. K., Kepler, R. M., Buyer, J. S., Cavigelli, M. A., Mirsky, S. B., & Maul, J. E. (2018). Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. *Scientific Reports*.

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

Twyford, A. D., & Ennos, R. A. (2012). Next-generation hybridization and introgression. In *Heredity*.

United Nations. (2015). World Population Prospects: The 2015 Revision. In *United Nations Economic and Social Affairs*.

United Nations. (2019). World Population Prospects 2019: Highlights. In *United Nations Publication*.

Van Gool, T., Biderre, C., Delbac, F., Wentink-Bonnema, E., Peek, R., & Vivarès, C. P. (2004). Serodiagnostic studies in an immunocompetent individual infected with Encephalitozoon cuniculi. *Journal of Infectious Diseases*.

Vasselon, V., Rimet, F., Tapolczai, K., & Bouchez, A. (2017). Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecological Indicators*.

Vávra, J., Yachnis, A. T., Shadduck, J. A., & Orenstein, J. M. (1998). Microsporidia of the genus Trachipleistophora - Causative agents of human microsporidiosis: Description of Trachipleistophora anthropophthera n. sp. (Protozoa: Microsporidia). *Journal of Eukaryotic Microbiology*.

Vaz Rodrigues, M., Juliana Francisco, C., Silva David, G., José da Silva, R., Pessoa Araújo Júnior, J., & Marianna Vaz Rodrigues, C. (2017). A new microsporidium species, Nucleospora braziliensis n. Sp. infecting Nile tilapia (Oreochromis niloticus) from brazilian aquaculture. *~ 496 ~ International Journal of Fisheries and Aquatic Studies*, *5* (1), 496–505. www.graphpad.com.

Vivarès, C. P., Gouy, M., Thomarat, F., & Méténier, G. (2002). Functional and evolutionary analysis of a eukaryotic parasitic genome. In *Current Opinion in Microbiology*.

Vogel, C. (n.d.). *Function Annotation of SCOP Domain Superfamilies*. Retrieved January 21, 2021, from http://supfam.org/SUPERFAMILY/function.html

Voronin, V. N. (1986). Microsporidia of Crustacea. *Protozoologiya Mikrosporidii*.

Vossbrinck, C. R., Andreadis, T. G., & Debrunner-Vossbrinck, B. A. (1998). Verification of intermediate hosts in the life cycles of microsporidia by small subunit rDNA sequencing. *Journal of Eukaryotic Microbiology*, *45* (3), 290–292.

Vossbrinck, C. R., & Debrunner-Vossbrinck, B. A. (2005). Molecular phylogeny of the Microsporidia: Ecological, ultrastructural and taxonomic considerations. *Folia Parasitologica*, *52* (1–2), 131–142.

Vossbrinck, C. R., Debrunner-Vossbrinck, B. A., & Weiss, L. M. (2014). Phylogeny of the Microsporidia. *Microsporidia: Pathogens of Opportunity: First Edition*, 203–220.

Vu-Khac, H., Thanh, T. N. T., Thu, G. N. T., Le, C. H., & Nguyen, V. D. (2018). Vertical transmission and early diagnosis of the microsporidian Enterocytozoon hepatonaei in whiteleg shrimp Penaeus vannamei. *Journal of Pure and Applied Microbiology*, *12* (3), 1125–1131.

Wadi, L., & Reinke, A. W. (2020). Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLoS Pathogens*.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*.

Walker, T., Johnson, P. H., Moreira, L. A., Iturbe-Ormaetxe, I., Frentiu, F. D., McMeniman, C. J., Leong, Y. S., Dong, Y., Axford, J., Kriesner, P., Lloyd, A. L., Ritchie, S. A., O'Neill, S. L., & Hoffmann, A. A. (2011). The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature*.

Wang-Peng, S., Zheng, X., Jia, W. T., Li, A. M., Camara, I., Chen, H. X., Tan, S.

Q., Liu, Y. Q., & Ji, R. (2018). Horizontal transmission of Paranosema locustae (Microsporidia) in grasshopper populations via predatory natural enemies. *Pest Management Science*.

Wang, Y., Jin, Y., Han, P., Hao, J., Pan, H., & Liu, J. (2021). Impact of Soil Disinfestation on Fungal and Bacterial Communities in Soil With Cucumber Cultivation. *Frontiers in Microbiology*, *0*, 2242.

Wang, Z. D., Liu, Q., Liu, H. H., Li, S., Zhang, L., Zhao, Y. K., & Zhu, X. Q. (2018). Prevalence of Cryptosporidium, microsporidia and Isospora infection in HIV-infected people: A global systematic review and meta-analysis. In *Parasites and Vectors*.

Ward, G. M., Neuhauser, S., Groben, R., Ciaghi, S., Berney, C., Romac, S., & Bass, D. (2018). Environmental Sequencing Fills the Gap Between Parasitic Haplosporidians and Free-living Giant Amoebae. *Journal of Eukaryotic Microbiology*, *65* (5), 574–586.

Webster Marketon, J. I., & Glaser, R. (2008). Stress hormones and immune function. In *Cellular Immunology*.

Weiss, L. M. (2014). Clinical Syndromes Associated with Microsporidiosis. In *Microsporidia: Pathogens of Opportunity: First Edition*.

Weiss, L. M., & Becnel, J. J. (2014). Microsporidia: Pathogens of Opportunity: First Edition. In *Microsporidia: Pathogens of Opportunity: First Edition*.

West Country River Trust. (n.d.). *Freshwater Pearl Mussels | Westcountry Rivers Trust*. Retrieved January 27, 2021, from https://wrt.org.uk/project/freshwater-pearl-mussels/

Whelan, K. (2010). A Review of the Impacts of the Salmon Louse , Lepeophtheirus salmonis (Krøyer , 1837) on Wild Salmonids. *Atlantic Salmon Trust*.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure.

Proceedings of the Royal Society B: Biological Sciences.

Widmer, G., Dilo, J., Tumwine, J. K., Tzipori, S., & Akiyoshi, D. E. (2013). Frequent occurrence of mixed enterocytozoon bieneusi infections in humans. *Applied and Environmental Microbiology*.

Williams, B. A. P. (2009). Unique physiology of host-parasite interactions in microsporidia infections. In *Cellular Microbiology*.

Williams, B. A. P., Hamilton, K. M., Jones, M. D., & Bass, D. (2018). Group-specific environmental sequencing reveals high levels of ecological heterogeneity across the microsporidian radiation. *Environmental Microbiology Reports*.

Williams, B. A. P., Hirt, R. P., Lucocq, J. M., & Embley, T. M. (2002). A mitochondrial remnant in the microsporidian Trachipleistophora hominis. *Nature*, *418* (6900), 865–869.

Williams, B. A. P., Lee, R. C. H., Becnel, J. J., Weiss, L. M., Fast, N. M., & Keeling, P. J. (2008). Genome sequence surveys of Brachiola algerae and Edhazardia aedis reveal microsporidia with low gene densities. *BMC Genomics*.

Williams, E., & Bunkley-Williams, L. (2019). Life Cycle and Life History Strategies of Parasitic Crustacea. In *Parasitic Crustacea* (Vol. 3, pp. 179–266). Nature Publishing Group.

Willis, J. R., & Gabaldón, T. (2020). The human oral microbiome in health and disease: From sequences to ecosystems. *Microorganisms*, *8* (2), 308.

Wilson, J. W. (2015). Bacterial pathogens. *Cancer Treatment and Research*, *161*, 91–128.

Winder, M., & Jassby, A. D. (2011). Shifts in Zooplankton Community Structure: Implications for Food Web Processes in the Upper San Francisco Estuary. *Estuaries and Coasts*, *34* (4).

Winfield, I. J. (2018). Cleaner Fish Biology and Aquaculture Applictions. *Journal*

of Fish Biology, *93* (5), 1027–1027.

Winters, A. D., Langohr, I. M., Souza, M. D. A., Colodel, E. M., Soares, M. P., & Faisal, M. (2015). Ultrastructure and molecular phylogeny of Pleistophora hyphessobryconis (Microsporidia) infecting hybrid jundiara (Leiarius marmoratus × Pseudoplatystoma reticulatum) in a Brazilian aquaculture facility. *Parasitology*, *143* (1), 41–49.

Wiredu Boakye, D., Jaroenlak, P., Prachumwat, A., Williams, T. A., Bateman, K. S., Itsathitphaisarn, O., Sritunyalucksana, K., Paszkiewicz, K. H., Moore, K. A., Stentiford, G. D., & Williams, B. A. P. (2017). Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae microsporidia. *Environmental Microbiology*, *19* (5), 2077–2089.

Woolhouse, M. E. J., & Gowtage-Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, *11* (12), 1842–1847.

Woolhouse, M. E. J., Haydon, D. T., & Antia, R. (2005). Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology and Evolution*, *20* (5), 238–244.

Wyban, J. (2003). Penaeus vannamei seedstock production: Recent developments in Asia. *Global Aquaculture Advocate*, 78–79.

Xu, J., Pan, G., Fang, L., Li, J., Tian, X., Li, T., Zhou, Z., & Xiang, Z. (2006). The varying microsporidian genome: Existence of long-terminal repeat retrotransposon in domesticated silkworm parasite Nosema bombycis. *International Journal for Parasitology*, *36* (9), 1049–1056.

Yang, S., Gao, X., Meng, J., Zhang, A., Zhou, Y., Long, M., Li, B., Deng, W., Jin, L., Zhao, S., Wu, D., He, Y., Li, C., Liu, S., Huang, Y., Zhang, H., & Zou, L. (2018). Metagenomic Analysis of Bacteria, Fungi, Bacteriophages, and Helminths in the Gut of Giant Pandas. *Frontiers in Microbiology*, *9*, 1717.

Yanong, R. P. E. (2013). Biosecurity in Aquaculture , Part 3 : Ponds. *Southern Regional Aquaculture Centre*.

Ye, J., Xiao, L., Wang, Y., Guo, Y., Roellig, D. M., & Feng, Y. (2015). Dominance of Giardia duodenalis assemblage A and Enterocytozoon bieneusi genotype BEB6 in sheep in Inner Mongolia, China. *Veterinary Parasitology*, *210* (3–4), 235–239.

Yuvanatemiya, V., Boyd, C. E., & Thavipoke, P. (2011). Pond bottom management at commercial shrimp farms in chantaburi province, Thailand. *Journal of the World Aquaculture Society*, *42* (5), 618–632.

Zhang, Q., Cai, J., Li, P., Wang, L., Guo, Y., Li, C., Lei, M., & Feng, Y. (2018). *Enterocytozoon bieneusi genotypes in Tibetan sheep and yaks*. *2*.

Zhang, Y., Koehler, A. V., Wang, T., Cunliffe, D., & Gasser, R. B. (2019). Enterocytozoon bieneusi genotypes in cats and dogs in Victoria, Australia. *BMC Microbiology 2019 19:1*, *19* (1), 1–8.

Zhao, W., Zhang, W., Yang, D., Zhang, L., Wang, R., & Liu, A. (2015). Prevalence of Enterocytozoon bieneusi and genetic diversity of ITS genotypes in sheep and goats in China. *Infection, Genetics and Evolution*, *32*, 265–270.

Zhao, W., Zhang, W., Yang, F., Cao, J., Liu, H., Yang, D., Shen, Y., & Liu, A. (2014). High prevalence of enterocytozoon bieneusi in asymptomatic pigs and assessment of zoonotic risk at the genotype level. *Applied and Environmental Microbiology*, *80* (12), 3699–3707.

Zmora, O., Findiesen, A., Stubblefield, J., Frenkel, V., & Zohar, Y. (2005). Large-scale juvenile production of the blue crab Callinectes sapidus. *Aquaculture*, *244*, 129 – 139.