Myxozoan parasite diversity in relation to point source pollution using environmental DNA

Emma Sweeney

A thesis submitted to the University of Nottingham

For the degree of Masters in Research (MRes) Ecology

Supervisors:
Hanna Hartikainen
Tom Reader

Abstract

Monitoring freshwater species diversity is essential to understanding the effects that anthropogenic activities have on aquatic environments. Currently, there is limited literature available in relation to the adverse effects of effluent waste on aquatic environments. Parasitic species can act as successful indicators for ecosystem health, as their biodiversity is beneficial to an ecosystem by creating links between trophic levels and increasing species diversity. Environmental DNA (eDNA) metabarcoding is an emerging molecular tool that facilitates non-invasive sampling and detection of communities of species. Myxozoa is an exceptionally large and diverse collective of microscopic obligate cnidarian endoparasites, with multiple waterborne transmission phases, making them ideal subjects for eDNA analysis. Combined sewage overflow sites that overflowed for >100 and <2,500 hours per annum (2020/2021) were selected and 60 (upstream and downstream from CSO) samples were collected from 15 locations during July 2021. A further 30 samples were collected from 15 sites on the river Erewash in July and September 2021. eDNA was extracted from each sample and ideal PCR program was optimised by qPCR for species-specific primers for myxosporeans, malcosporeans and macroinvertebrates and presence of each taxon was confirmed by agarose gel electrophoresis. Amplicons were sequenced on Illumina MiSeq 300bp paired end chemistry in-house at Nottingham DeepSeq facility. Data analysis was conducted using R version 2021.09.0+351 in RStudio version "Ghost Orchid". The results reported in the current study relate only to the two parasite groups, Myxosporea and Malacosporea, as it was found after sequencing that the invertebrate primers yielded insufficient metazoan reads for reliable analyses. The results suggest that the number of individual species of myxosporeans and malacosporeans per site were higher at upstream sites, or sites that received lower levels of pollution throughout the year. From this it can be inferred that there may be a reduction in levels of biodiversity of Myxozoans downstream from CSOs. The current study has demonstrated that it is possible to detect and map diversity of Myxozoan species through eDNA approaches, in relation to environmental stressors such as effluent waste. [1

Table of Contents

1. Introduction

1.1 Pollution in aquatic environments

1.2 Parasitic importance in healthy ecosystems

1.3 eDNA as a method for measuring ecosystem health

1.4 Myxozoans for testing eDNA approaches

2. Materials and Methods

2.1 Sampling, filtration and eDNA extraction

2.2 qPCR optimisation and gradient PCR

2.3 PCR amplification

2.4 SPRI bead cleaning and index PCR

2.5 Quantification and sequencing

2.6 Data analysis

3. Results

3.1 Myxosporean and malacosporean occurrence

3.2 Myxosporean communities across pollution impacted rivers

3.3 Malacosporean communities in pollution impacted rivers

3.4 Myxosporean and malacosporean communities in the River Erewash

4. Discussion

4.1 Detection of Myxozoans by eDNA

4.2 Diversity and distribution of Myxozoans in UK rivers

4.3 CSO impacts of Myxozoan communities

4.4 Myxozoan diversity along a river continuum in the River Erewash

5. Conclusion

Acknowledgements

1. Introduction

1.1 Pollution in aquatic environments

The population decline and rapid reduction in biodiversity within freshwater organisms worldwide (83% between 1970 and 2014, an annualised index of 3.9%) has been attributed to five leading causes: 'overexploitation', 'water pollution', 'flow modification', 'destruction or degradation of habitat' and 'invasion by exotic species' (Dudgeon *et al*. 2006, Reid *et al*. 2018). Freshwater environments such as lakes, reservoirs and rivers cover approximately 2.3% of land surface area on Earth, yet these ecosystems host almost 9.5% of all described animal species (Freshwater Animal Biodiversity Assessment, Balian *et al*. 2007). While factors such as destruction and degradation of habitat are leading and persistent causes for population declines in freshwater systems, new threats are emerging as the global human population expands. For example, in general, water pollution is already well established and acknowledged in the degradation of freshwater ecosystems, but the emerging contaminants and new anthropogenic processes involved are changing rapidly (Reid *et al*. 2018).

Freshwater environments directly receive pollution from point-source discharges from anthropogenic activities such as agriculture and aquaculture, oil production and paper production. Another significant contributor to poor water quality is intermittent discharges of storm sewage from combined sewer overflows (combined sewer overflows, hereafter CSOs, are sewers that are designed to collect rainwater runoff, domestic sewage, and industrial wastewater in the same pipe). Sudden and unexpected rainfall, increased rainfall and in more recent years, dramatic increase in residential housing without the construction of more sewage treatment works (hereafter STWs) and CSOs is resulting in the increased overflow of untreated sewage directly into rivers across the United Kingdom. This not only affects urban city centre rivers, but also pristine chalk streams and protected aquatic environments alike (Reynolds *et al*. 2012). With the number of annual sewage overflows that occur ever increasing, emerging contaminants of larger volumes are causing newer concerns for ecosystems and freshwater species biodiversity. These include active pharmaceutical ingredients (such as hormonal supplements, antibiotics, antivirals, antidepressants), illicit drugs, and endocrine disruptors (Dudgeon *et al*. 2018), all of which have unknown effects on aquatic environments and have sparked countless related research studies over the last two decades. These effluents, when discharged into receiving water bodies without adequate treatment are thought to induce mild to severe toxic effects in the aquatic fauna. According to the United Nations, 80% of the current wastewater produced is discharged into water bodies without any adequate treatment (Bhanot and Hundal 2020). A study conducted by Bernet *et al*. (2004) assessed the effects of wastewater disposal on histological alterations in fish. The study found that trout exposed to wastewater from STWs showed higher

histopathological indices than trout caught upstream of the discharge point of the STW, indicating a negative effect of wastewater on the histopathological status of the examined organs of trout, such as the skin, liver and gills. Yet, the broader effects of these individual compounds or combined with each other on aquatic populations, communities and biodiversity, as well as ecosystem function, remain undetermined (Dudgeon *et al*. 2018).

1.2 Parasitic importance in healthy ecosystems

When considering ecosystem health, it is naturally assumed that an ecosystem with high levels of parasitic diversity is an unhealthy one. However, recent studies have introduced theories of how parasitism can shape communities, ecosystems and ecology. Accumulating evidence indicates that instead, as parasite diversity increases, ecosystem functioning in-turn, improves (Hudson *et al*. 2006). Whilst initially this may seem counterintuitive as parasites can threaten endangered species and negatively impact host fitness, they can in fact be favourable in the elevation of biodiversity. 'Ecosystem health' focuses on the functioning of an entire community, and therefore must encapsulate the entire performance of the system. When looking at successful ecosystems, most free-living animals are hosts to a large diversity and prevalence of parasites, at an individual, a population or a species level (Hudson *et al*. 2006, Bell *et al*. 2018, Bordes and Morand 2009) such as ticks, bacteria and nematodes. This variation of parasitic species not only creates links between trophic levels within the ecosystem (for example, tapeworm larvae parasitising herbivorous prey species in the muscle and developing into an adult within the predatory carnivorous species, known as a predator-parasite link), but also creates chains of multispecies connections which can stabilise the structure of a community, in-turn, increasing resilience and aiding persistence (Hudson *et al*. 2006, Lafferty *et al*. 2006). Furthermore, when parasites are considered as a significant species within an ecosystem, the diversity within that ecosystem immediately increases. Consequently, the previously mentioned ongoing global changes are altering the structure of the parasitic diversity that host species may face, mostly due to the reshaping of the geographical distributions of parasites due to climate change, habitat fragmentation and destruction, and pollution by foreign elements (Bordes and Morand 2009).

With the importance of parasite biodiversity being considered, they can therefore be positive indicators for ecosystem health. Until recently, the identification of species present within an environment has relied on the trapping, capture or direct observation of individuals. The observation of parasitic species often requires the capture of the host species, with invasive histological sampling. Although such approaches have significantly improved over time, they remain limited by factors such as identification errors, sampling biases, false negatives and

coarse taxonomic resolution (Keck *et al*. 2022). Furthermore, these methods are not suitable for endangered or reclusive species.

1.3 eDNA as a method for measuring ecosystem health

Environmental DNA (hereafter eDNA) refers to extracellular DNA that is present and can be extracted from an environmental sample, such as water, soil and sediment (Rees *et al*. 2014). eDNA is utilised for the monitoring of aquatic and semi-aquatic species through the analysis of DNA by several methods. The approach facilitates non-invasive sampling and detection based solely on the collection of water samples when there is no visible presence of the target species (Gomes *et al*. 2017). Research conducted over the last decade has confirmed the power of eDNA approaches for accurately surveying a wide range of different populations in finer detail (Bruce *et al*. 2021).

Early demonstrations of the significance of eDNA include the detection of the incipient invasive species of Asian carp (bighead carp, *Hypophthalmichthys nobilis* and silver carp, *Hypophthalmichthys molitrix*) in the Great Lakes (Erie and Michigan, USA) (Jerde *et al*. 2013). With the species' presence only being identified previously in the Great Lakes through capture, three times in 1995 and once in 2010, eDNA techniques allowed for the early detection of the Asian carp, and the implementation of the Great Lakes surveillance plan for protecting imperilled species (Jerde *et al*. 2013). Following initial intense scrutiny, eDNA as a method was eventually accepted and in turn, resulted in a wave of research employing the technique. Another such example includes the monitoring of great crested newts (*Triturus cristatus*) in the UK. The national citizen science-based study, conducted by Biggs *et al*. (2015), compared the effectiveness of eDNA monitoring with that of traditional methods; torch counts, bottle trapping and egg searches, and relied on the ability of volunteers to collect eDNA samples throughout known locations of the newts. A total of 35 ponds were visited four times throughout the breeding season, and eDNA successfully detected newts on 139 out of 140 occasions, equating to 99.3% in success rate in contrast to torch counts (75% success rate), bottle traps (76% success rate) and egg searches (44% success rate) (Biggs *et al*. 2015). In conjunction with the rapid increase in the use of eDNA for biomonitoring, many research studies began to use high-throughput sequencing, where entire communities of different organisms from a variety of species can be described, and this approach was termed eDNA metabarcoding (Bruce *et al*. 2021).

Due to its detection power and relative ease of application, eDNA techniques have now been enhanced to focus on queries related to microbial community diversity, evolution,

ecology and host-pathogen interactions (Bastos Gomes *et al*. 2017). The development of these techniques is especially significant for parasitology because parasites can be laborious to locate, are morphologically uncooperative and genetically divergent (Bass *et al*. 2015). A study conducted by Huver *et al*. (2015) developed the application of an eDNA method to detect and quantify the pathogenic parasitic nematode *Ribeiroia ondatra*e known to infect North American amphibians. The study detected high specificity and sensitivity through quantitative PCR (qPCR) to *Ribeiroia ondatra*e from 15 field sites with 90% accuracy compared to classical amphibian host necroscopy. The study also examined the detectability of *R. ondatra*e in water samples subjected to different sample degradation conditions such as temperature and time. It is important to note that a vast majority of parasite-based eDNA studies focus on the qPCR amplification of the target species. Whilst qPCR is successful in the detection of species, the data provided is limited to presence/absence and does not infer relative abundance or the presence of multiple species within one sample. The importance of improved parasitological eDNA research is clear because parasitology requires the identification of some of the smallest and most cryptic eukaryotes known (Bass *et al*. 2015). The current study will focus on not only the detection of species specific eDNA within samples but also the sequencing of the amplified eDNA with species specific primers to infer greater species resolution and data within each sample.

1.4 Myxozoans for testing eDNA approaches

Myxozoa is an exceptionally large and diverse collective of microscopic obligate cnidarian endoparasites consisting of approximately 2600 species currently. Myxozoans are distributed globally in both marine and freshwater habitats, and they fluctuate between primary invertebrate hosts (annelid and bryozoan) and secondary vertebrate hosts (predominantly fish) (Alama-Bermejo and Holzer, 2021). It is estimated that Myxozoans emerged approximately 600-700 million years ago, and by observing their phylogeny and life history, it can be assumed that they first settled as solely parasites of aquatic invertebrates, and approximately 200-300 million years later implement their secondary fish hosts, with the occurrence of aquatic invertebrates on Earth and sequence analysis of Myxozoans has showed clear phylogenetic grouping according to fish host species (Lisnerová *et al*. 2020). These circumstances aided extensive host-associated biodiversification, and potentially suggests the main reason for the unmistakable success of Myxozoa, compared to other parasitic cnidarians (Lisnerová *et al*. 2019)

For these reasons, the Myxozoan group of parasites is immensely appealing from an evolutionary perspective, given that they are some of the earliest multicellular parasites on Earth. They are heavily diversified and opportunistic, with the capability to infect many hosts

from different animal groups, switching to distant and isolated groups such as molluscs, monogeneans, and trematodes (Holzer *et al*. 2018, Alama-Bermejo and Holzer, 2021) as well as amphibians, reptiles, waterfowl and shrews in freshwater, marine and terrestrial environments (Hartikainen *et al*. 2014).

Extremely reduced in size as an adaptation to parasitism, Myxozoans are composed of very few cell types throughout their lifecycle stages (Yokoyama *et al*. 2012). The phylum Myxozoa is divided into two classes, myxosporeans and malacosporeans. Transmission of myxosporeans occurs via microscopic water-borne spores that consist of shell valves, sporoplasms and polar capsules that contain coiled polar filaments (Yokoyama *et al*. 2012, Americus *et al*. 2020), and their cnidarian resemblance is supported by genetic relatedness and the presence of these polar capsules in Myxozoan specific research (Americus *et al*. 2020). Polar capsules are organelles found in myxosporean spores that are responsible for the first stage of infection between invertebrate and vertebrate hosts. Myxosporean spore stages (myxospores) are ingested by their primary hosts, annelids, where the polar filaments project to ‘anchor’ said spore to the gut epithelium. Shell valves open and sporoplasms penetrate the epithelium. Reproduction and development occur in the gut tissue, and eventually produces usually eight actinospres within a pansporocyst (Yokoyama *et al*. 2012). Mature actinsospores are then released from the host where they then float within the aquatic environment. It is here that spores will eventually make contact with the skin or gills of the secondary vertebrate host species and penetrate through the epithelium, to begin development of the myxosporean stage. Multiplication occurs followed by migration via the nervous or circulatory system to the final site of infection. Initially, identification at the species level of myxosporeans was based on the dimensions of spores and polar capsules, in relation to width and length (relative to guidelines of Lom and Arthur, 1989) (Yokoyama *et al*. 2012). This method of identification requires great morphological knowledge and highlights the need for eDNA high-throughput sequencing (HTS) methods for the rapid and reliable identification and observation of Myxozoan species.

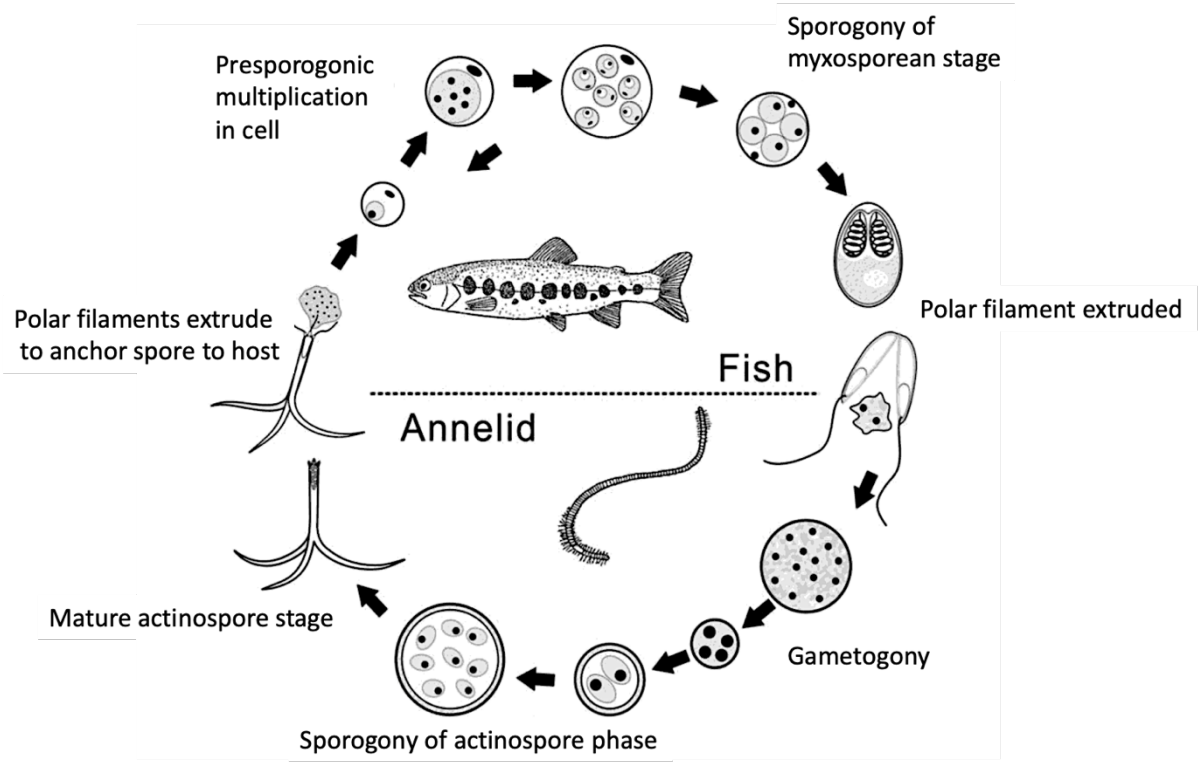

Fig 1. Diagram of myxosporean invertebrate and vertebrate host alternating lifecycle as described by Yokoyama *et al*. 2012 with added labels for each phase.

Malacosporean lifecycle still remains considerably secretive as they are a recently discovered group, with only five species being described to date (*Buddenbrockia plumatellae, Buddenbrockia bryozoides, Tetracapsuloides bryosalmonae, Tetracapsuloides vermiformis* and *B. allmani*) (Patra *et al*. 2017) All species discovered are known to parasitise bryozoans, but the primary life cycle described is that of *Tetracapsuloides bryosalmonae*. Development begins in the body cavity of freshwater bryozoans, in sac-like stages, with the release of malacospores to the surrounding aquatic environment. Spores then come into contact again with the skin or gills of their Salmonid host. The sporoplasm then penetrates epithelium where it is transported to the kidney interstitium, via the bloodstream, causing proliferative kidney disease (PKD). Mature spores are then released into the surrounding water with urine after migration through the kidney tubules (Yokoyama *et al*. 2012).

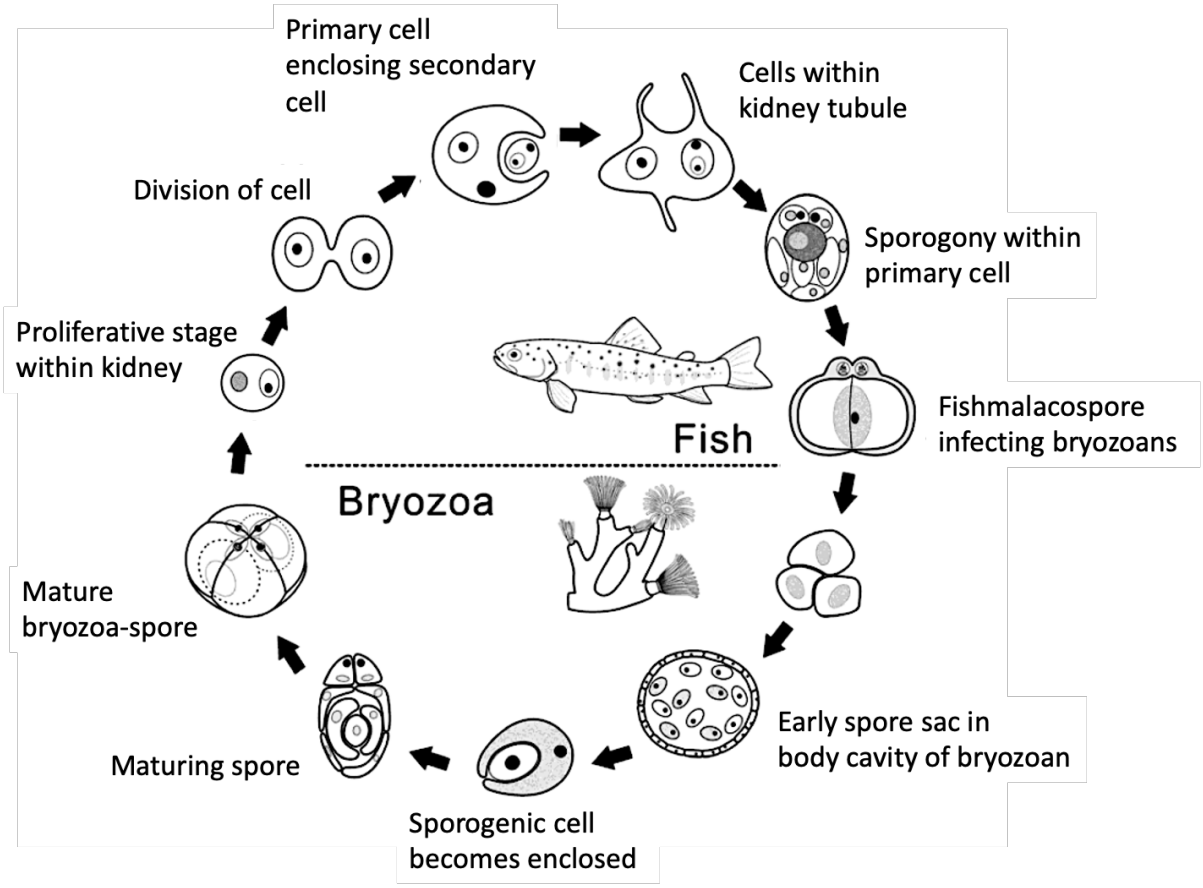

Fig 2. Diagram of malacosporean invertebrate and vertebrate host alternating lifecycle as described by Yokoyama *et al*. 2012 with added labels for each phase.

It is the above multiple associated waterborne transmission phases of Myxozoans that makes them ideal subjects for eDNA analysis for a deeper understanding in their importance for the preservation and elevation of biodiversity. With the ever-increasing development of anthropogenic activity across the planet, it is crucial that the surveillance of ecosystems, their inhabitant species, populations and biodiversity are equally increasing. The need to establish efficient monitoring programmes of parasites in aquatic ecosystems is crucial, whether it be in the light of dramatic global biodiversity shifts, or the emergence of parasitic diseases. The traditional methods of parasite identification are time consuming, require invasive sampling methods and high levels of taxonomic expertise. In the present study, we aim to establish how the collection of eDNA water samples and lineage-specific amplicon sequencing using myxosporean, malacosporean and macroinvertebrate targeted PCR can be suitable for detecting and mapping parasites, in relation to potentially destructive effluent point source pollution sites, and how this may lead to a reduction in aquatic biodiversity. With the combined use of eDNA techniques for the rapid biomonitoring of parasitic communities, there is the potential to be able to use parasites as indicators for ecosystem health.

2. Materials and Methods

2.1 Sampling, filtration and eDNA extraction

Sampling locations were selected using the website ‘Is My River Fit to Play in?’ (https://experience.arcgis.com/experience/e834e261b53740eba2fe6736e37bbc7b/page/2021/), which highlights effluent waste storm drain sites across the UK that overflow annually. The sampling sites were all located within 3h travel distance from Nottingham. Only sites that overflowed for >100 and <2,500 hours per annum (2020/2021) were selected. Four samples were collected from each site, two replicate water samples upstream from the storm drain, and two replicate water samples downstream from the storm drain (depending on access, 20-100m distance up/downstream from the point source). Sixty samples were collected from 15 storm drain locations during July 2021. A further 30 samples were collected from 15 sites on the river Erewash in July and September 2021. Water samples were collected using a disposable 50ml syringe, submerged into water with a gloved hand. 250ml of water was filtered through 25mm diameter cellulose nitrate filter membranes, comprising a 0.45µm membrane and a 5µm pre-filter membrane. The 0.45µm membrane was removed and a further 250ml of water was filtered through the 5µm membrane. Both filter membranes were removed from the filter holders using sterile forceps and placed in the same 2ml screw top centrifuge tube. Filter papers were preserved in 800ul Longmire's buffer at -20C.

Fig. 3 Map of all sampling locations within England. The central collection of sampling sites represents the sites along the River Erewash (flowing south from Kirkby to Golden Brook) and sites surrounding Nottingham.

eDNA extraction steps were carried out at a specified DNA extraction workbench to prevent cross-contamination with PCR products. Protective equipment (lab coat, nitrile gloves, medical grade face mask) were worn and surfaces were sterilised with 1% distel before every work cycle. General preparation included the preparation of fresh 8M GuHCl (8 x 95.53g [MW of GuHCl] = 764,24g per 1L H2O = 0,76g GuHCl per 1ml H2O), ethanol (100%) and proteinase K (20mg/µl in sterile 50 mM TRis pH8, 1.5mM calcium acetate, aliquoted and frozen). A Qiagen DNeasy Blood and Tissue kit including AW1 and AW2 buffer and AE buffer was used with a modified protocol as described below. Throughout, all sample IDs were photographed and documented with codes corresponding to sampling information and locations.

0.5ml of 0.1 mm diameter glass beads was added to each original 2ml sample tube containing the two filter membranes and 800ul Longmire's buffer. All samples were processed in a Precellys bead-beater at 10,000 rpm for 3 cycles of 20s shaking, 40s rest to avoid overheating and then spun down in a centrifuge at 6000 rpm for 1 min. This procedure disrupted the filter papers and allows DNA extraction from even hardened parasitic spores. 10µl of proteinase K was added to each sample (1µl per 100µl initial buffer volume). All samples were then vortexed for 1 min and incubated in a thermomixer for 1 hour (shaking gently) at 56˚C. Post incubation, samples were vortexed for 15 seconds. 500µl 8M GuHCl and 500µl ethanol (100%) was added to each original tube to ensure all product was rinsed from the tube and attained, then vortexed and spun down. All product was then added to a fresh 2ml eppendorf tube and processed according to manufacturer's protocol (Qiagen Blood and Tissue kit). Final purified eDNA was eluted into 100µl of AE buffer after incubation in a thermomixer for 5 min at 37˚C shaking gently to maximise yields. 50µl of DNA extract was archived for later use, and all pure DNA samples were labelled A (final voucher) and B (backup), and placed in the freezer for storing.

2.2 qPCR optimisation and gradient PCR

A qPCR was run on 6 randomly selected samples, one positive control and one negative control for all primers to optimise the ideal number of cycles per primer. qPCR reactions consisted of 5µl of a premade buffer solution (750 µl GoTaq® Probe qPCR Master Mix with 1.25µl Promega CRX reference dye, suitable for ABI fast qPCR and 15µl EvaGreen dye), 0.5µl forward and reverse primer, 1µl template and 3µl 4.6µl nuclease free H_20. qPCR was performed on ABI Fast at Nottingham Life Sciences with the following program for SYBR Green detection: 2 min activation at 95˚C, 45 cycles of 3 sec denaturation at 95˚C, annealing and extension for 30 sec at 60˚C followed by a melting curve from 60-95˚C. A gradient PCR

was run on 12 of the same randomly selected sample for all primers to optimise the ideal annealing temperature for each primer. Each gradient PCR reaction consisted of 7.5µl NEBNext Ultra II Q5 Master Mix Taq (1X concentration), 0.45µl forward primer (1µM), 0.45µl reverse primer (1µM) and 4.6µl nuclease free H_20. 2µl DNA extract was added to each corresponding well at a separate post-PCR hood designated bench, for the Myxozoan targets a second nested gradient PCR was conducted using 1µl PCR product from the ERIB PCR and 5.6µl nuclease free H_20. Gradient PCR program consisted of an increase in annealing temperature across 12 samples, from 55-67˚C. Optimal number of cycles and annealing temperature for each primer is described in the following section.

2.3 PCR amplification

A two step PCR approach was applied to prepare metabarcoding libraries for three targets, comprising 1) myxosporean parasites (FR, pers. comm. I. Fiala), 2) malacosporean (malaco, pers. comm. I. Fiala) parasites and 3) aquatic macroinvertebrates (BF) (Leese *et al*. 2020).

Table 1. Pooling of myxosporean 2nd step primers and volume added per primer.

Forward Primer	Volume in pool	Sequence 5'-3'	Length	Reverse Primer	Volume in pool	Sequence 5'-3'	Length
FR_for1_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCGATGAGWAACWACTGGAGG	90bp	FR_rev1_nex0	200µl	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTATYAACATTCAAGC	89bp
FR_for3_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTMAAYGAGWAACAACTGGAGG	90bp	FR_rev2_nex0	50µl	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTAYTAACSTTCAAGC	89bp
FR_for5_nex0	100µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGTCGAGTAACAACTGRAGG	90bp	FR_R6_nex0	50µl	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTRTAWCATTCAGGC	88bp
FR_for6_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTRWTTTGAGTAACRACTGGAGG	90bp	FR_rev6_nex0	100µl	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTRTAWCATTCAGGC	88bp
FR_for4_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCGTTGAKAAACAACTAGAGG	90bp				
FR_for7_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTGTCGAGAAACAAHTRGAGG	90bp				
FR_F2_nex0	50µl	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTYCGKTGAGTAACWACTGGAGG	90bp				

A 96-well plate was used for 60 effluent waste samples, 30 Erewash samples, 4 replicates (of the first site of the first 4 columns on the plate) and 2 negative controls. Each amplification reaction in all steps consisted of a total of 15µl of prepared master mix, which was prepared in a UVed PCR hood. For 1st step PCR reactions, this consisted of 7.5µl

NEBNext Ultra II Q5 Master Mix Taq (1X concentration), 0.45µl forward primer (1µM), 0.45µl reverse primer (1µM) and 4.6µl nuclease free H_20. 2µl DNA extract was added to each corresponding well at a separate post-PCR hood designated bench. For the Myxozoan targets a second nested amplification was conducted using 1µl PCR product from the ERIB PCR and 5.6µl nuclease free H_20.

Table 2. Table of primers included in the current study, step in PCR, target species, sequence, length and specific PCR program

Forward Primer	1st/2nd Round	Target	Sequence 5′-3′	Length	Reverse Primer	Sequence 5′-3′	Length	PCR Program
BF2_nex0	1st round	Macroinvertebrates	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCHCCHGAYATRGCHTTYCC	89bp	BR2_nex0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCDGGRTGNCCRAARAAYCA	89bp	2 min at 95˚C, 35 cycles of 95˚C for 10s, 62˚C for 30s, 72˚C for 30s. Held at 72˚C and cooled to 10˚C.
ERIB1	1st round	Myxozoans	ACCTGGTTGATCCTGCCAG	19bp	ERIB10	CTTCCGCAGGTTCACCTACGG	21bp	2 min at 95˚C, 30 cycles of 95˚C for 10s, 63˚C for 30s, 72˚C for 30s. Held at 72˚C and cooled to 10˚C.
FR_for_nex0 pool	2nd round	Myxospores ans	POOL (SEE TABLE 1)	Pool	FR_rev_nex0 pool	POOL (SEE TABLE 1)	Pool	2 min at 95˚C, 35 cycles of 95˚C for 10s, 58˚C for 30s, 72˚C for 30s. Held at 72˚C and cooled to 10˚C.
malaco_envmidF_nex0	2nd round	Malacosporeans	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATTCAGGTCCATTCGTG	88bp	malaco_envmidR_nex0	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTHGAATGTCCAGGC	88bp	2 min at 95˚C, 35 cycles of 95˚C for 10s, 58˚C for 30s, 72˚C for 30s. Held at 72˚C and cooled to 10˚C.

Quality and successful amplification was confirmed by agarose gel electrophoresis (1.5%) using 1µl of product. PCR reactions were stored at -20˚C ready for cleaning.

2.4 SPRI Bead cleaning and indexing PCR (2nd step PCR)

PCR products for FR and malaco primers were purified and size-selected for target amplicons using a two sided Solid Phase Reversible Immobilisation (SPRI) protocol. SPRI beads (Speed Beads Sigma GE65152105050250) were brought to room temperature for 30 min whilst rotating gently. Fresh 80% ethanol and Tris (10mM, pH8) were prepared before each lab session. PCR products were spun down in their original plate and 16µl Tris was added to ensure there was 30µl of product per well. SPRI beads were vortexed for 2 minutes until they reached a homogenous brown colour. 15µl of SPRI beads were added to each well at a 0.5x ratio therefore binding to DNA >500bp and mixed by pipetting. Each plate was incubated at room temperature for 15 minutes and placed on a magnetic stand for 3

minutes. 30µl of the supernatant was removed per sample and placed into a new plate in corresponding wells. 30µl of beads was added for a 1.0x ratio to bind to DNA >100bp, mixed by pipetting and incubated for a further 15 minutes. Plates were placed on a magnetic stand for 3 minutes and the supernatant was removed and discarded. Whilst remaining on the magnetic stand, 200µl of ethanol was added, left for 30 seconds, removed and discarded. This wash step was then repeated. Samples were left to air dry on the magnetic stand for 15 minutes. The plates were removed from the stand and 32.5µl Tris was added for resuspension, pipetted 30 times and left to incubate at room temperature for 2 minutes. The plates were then placed back on the magnetic stand for 3 minutes, and 30µl of the supernatant was removed, placed into a new plate, sealed and stored.

For the BF PCR product, a one-sided protocol was applied to remove only bands below 200bp and any primer dimer, as there were no higher unwanted bands. 24µl of beads was added to each well for a 0.8x ratio, mixed by pipetting and incubated. Once placed on the magnetic stand for 3 minutes, the supernatant was discarded, the same wash step as previously mentioned was repeated and the plate was left to dry for 15 minutes. 32.5µl Tris was added and pipetted 30 times, and the supernatant was placed into a new plate and stored.

All plates were then indexed after initial amplification of the target region using Integrated DNA Technologies Unique Dual-Indexed indexing primers. Indexing primers allow the mixing of up to 96 samples to be sequenced as one pool. Each index combination was given to a specific library in the plate for de-multiplexing during data analyses. Index PCR master mix consisted of 7.5µl NEBNext Ultra II Q5 Master Mix (1X concentration), 5.5µl nuclease free H_20, 1µl IDT index primer and 1µl cleaned PCR product per well. Preparation of master mix was conducted in a UVed PCR hood and PCR product was added at a designated post PCR hood bench. A new pipette tip was used when adding each individual index primer to avoid between-well contamination. Index PCR program was as follows: 95˚C for 3 min, then for 10 cycles, 95˚C for 30 s, 55˚C for 30 s, 72˚C for 30 s, followed by a final extension at 72˚C for 5 minutes and cooled to 4˚C until removed from the machine. Samples were cleaned again using the same 0.8x one-sided protocol as previous. Concentrations of all samples were quantified using BioDrop TOUCH Spectrophotometer and Qubit (Thermo Fisher Scientific Qubit 4 Fluorometer), and diluted and normalised with a calculated volume of Tris per well to a concentration of 10nM. 20µl from each well was then moved to a new plate, and 12µl of each sample was combined with 18µl Tris to reach a final concentration of 4nM. Finally, 30µl of the BF library, 15µl of the FR library and 5µl of the malaco library were pooled (6:3:1) and

quantified once again. The pooled amplicons were polished by size selection to 100-600bp fraction on the BluePippin instrument (Sage Science Ltd) and sequenced on Illumina MiSeq 300bp paired end chemistry in-house at Nottingham DeepSeq facility.

2.6 Data analysis

Data analysis was conducted using R version 2021.09.0+351 in RStudio version "Ghost Orchid". Package dada2 was used for filtering, denoising and the removal of chimaeras (Callahan *et al*. 2016). ASV's (amplicon sequencing variants, referring to all unique DNA sequences recovered from the high-throughput analysis of amplified target regions) were annotated against PR2_v4_13_March2021.RData using DECIPHER package (Wright, 2016) and manually curated with reference to Hartikainen *et al*. 2016 and Liu *et al*. 2019. Additionally, ASVs were clustered into Operational Taxonomic Units (OTUs) using swarm v.2 (Mahe et al. 2015). Correspondence of OTUs and ASVs was verified by phylogenetic analyses and classified into previously described phylogenetic groups (Hartikainen et al. 2016). ASVs were searched against GenBank nt collection using blastn and sequences with >92% identity across the whole amplicon sequence were assigned the annotation of the top Blast hit. This manual curation enabled the assignment of all 75 ASVs to their closest species match for visualisation purposes. No ASVs were collapsed into OTUs based on blast results, and all community analyses were conducted at the level of ASVs. Metadata on the site of infection and fish host group/species were gathered for each database matched sequence from literature and GenBank searches.

Myxosporean ASV sequences were aligned in MAFFT v.7.407_1 (Katoh et al. 2019) with 0.123 gap extension penalty and default settings. The alignment was curated manually and with BMGE v.1.12_1 (Criscuolo and Garibaldo 2010) to remove ambiguously aligned bases. Phylogenetic relationships were inferred using MrBayes v. 3.2.7_0 (Ronquist and Huelsenbeck 2003) with 4 chains of 100000 generations sampled every 500 iterations and a burn-in of 25%. GTR nucleotide substitution model and Gamma distributed rates across sites were implemented and posterior probabilities plotted on consensus topology using Newick Display 1.6 (Lemoine et al. 2019). The resulting topology was imported into R studio. Within R studio, statistical assumptions for both myxosporean and malacosporean species diversity were assessed in relation to annual CSO overflow using packages Phyloseq (version 1.38.0), AICcmodavg (version 2.3-1) and gridExtra (version 2.3). A generalised linear model (GLM) was created with a log-link function to transform the data by exponentiating the linear predictors. Finally, a Poisson regression was run due to there being count data that was not normally distributed, and an ANOVA (Analysis of Variance) table

produced to determine statistical significance (Beckerman *et al*. 2017). For data comparing species diversity in relation to sites upstream and downstream from CSOs, excluding overflow rate, Welch's two-sample t-test was utilised, as there is a difference between the variations of the two populations and the sample sizes are unequal.

3. Results

The results reported here relate only to the two parasite groups, Myxosporea and Malacosporea. This is because it was found after sequencing in preliminary analyses that the invertebrate specific primers yielded insufficient metazoan reads for reliable analyses. Annotations revealed a large diversity of bacterial contaminants which overwhelmed the signal for metazoan targets.

3.1 Myxosporean and malacosporean occurrence

In total 2,127,210 reads were generated by the myxosporean assay and 379,117 reads by the malacosporean assay. The reads were grouped into 3989 and 80 ASV's, respectively. After annotation, 75 myxosporean and 14 malacosporean ASV were attributed to Myxozoa. For Myxosporea, this resulted in 680,113 paired end myxosporean reads (31% of total reads attributed to target, 69% non-target) and for Malacosporea, 222,037 paired-end reads (58% attributed to target, 42% non-target). The ASVs were further confirmed to be Myxozoans by separate phylogenetic analyses for Myxosporea and Malacosporea (Hartikainen *et al*. 2016, Liu *et al*. 2019) and blast searches against Genbank nt collection.

Not all samples contained both myxosporeans and malacosporeans. In the case of Myxosporea, of the 96 eDNA samples originally included, 9 did not produce any amplicons (and were not included in the sequencing run, leaving 87 libraries for sequencing), 10 contained no myxosporean reads after sequencing, leaving 77 samples for analysis. This indicated 80.2% prevalence of myxosporeans across the sites. For malacosporeans, of the 96 eDNA samples included, 54 did not produce any malacosporean amplicons, and were not included in the sequencing run. The remaining 42 samples all contained malacosporean reads, indicating that malacosporean prevalence among the sampled sites was 43.75%.

In order to standardise sampling effort for each parasite group, the myxosporean and malacosporean datasets were rarefied to an even number of reads. For myxosporeans, 38 samples containing <1100 reads were removed, leaving 39 samples with >2543 reads per library (14 rare ASVs were removed during rarefication and 99,177 total reads remained). For malacosporeans, 7 samples containing <1100 reads were removed leaving 28 samples

with >1,208 reads per sample (5 rare ASVs were removed during rarefication and 33,824 reads remained). The rarefaction was performed separately for datasets with and without the River Erewash sampling transect (see methods). The UK effluent waste sampling sites comprised a broader geographical range and a higher myxosporean diversity (50 ASV's across 28 samples) than the River Erewash sampling sites (22 ASVs across 11 samples). This was not reflected in the malacosporean data, where ASV richness did not vary between the UK effluent waste sampling dataset and the River Erewash dataset (14 ASVs in both datasets, with 17 and 11 samples, respectively).

Based on annotation and phylogenetic analyses, four of the myxosporean clades defined in a previous eDNA metabarcoding study (Hartikainen et al. 2016) were also found, comprising representatives of the freshwater and marine gall bladder clade, environmental clade and *Myxobolus* 4 (Hartikainen *et al*. 2016). The *Myxobolus* 4 clade originates from the previously divided subclades "*Myxobolus* 1-5" which provides further resolution for plotting ASV occurrence (Fig 1). All amplified malacosporean ASVs phylogenetically identified as being in one of two genera; *Buddenbrockia* or *Tetracapsuloides.* Through further phylogenetic analysis, the ASVs were suggested to closely group into four previously identified species, *Buddenbrockia plumatellae* sac, *Buddenbrockia plumatellae* worm, *Buddenbrockia* worm FS2 and Tetracapsuloides bryosalmonae (figure not shown).

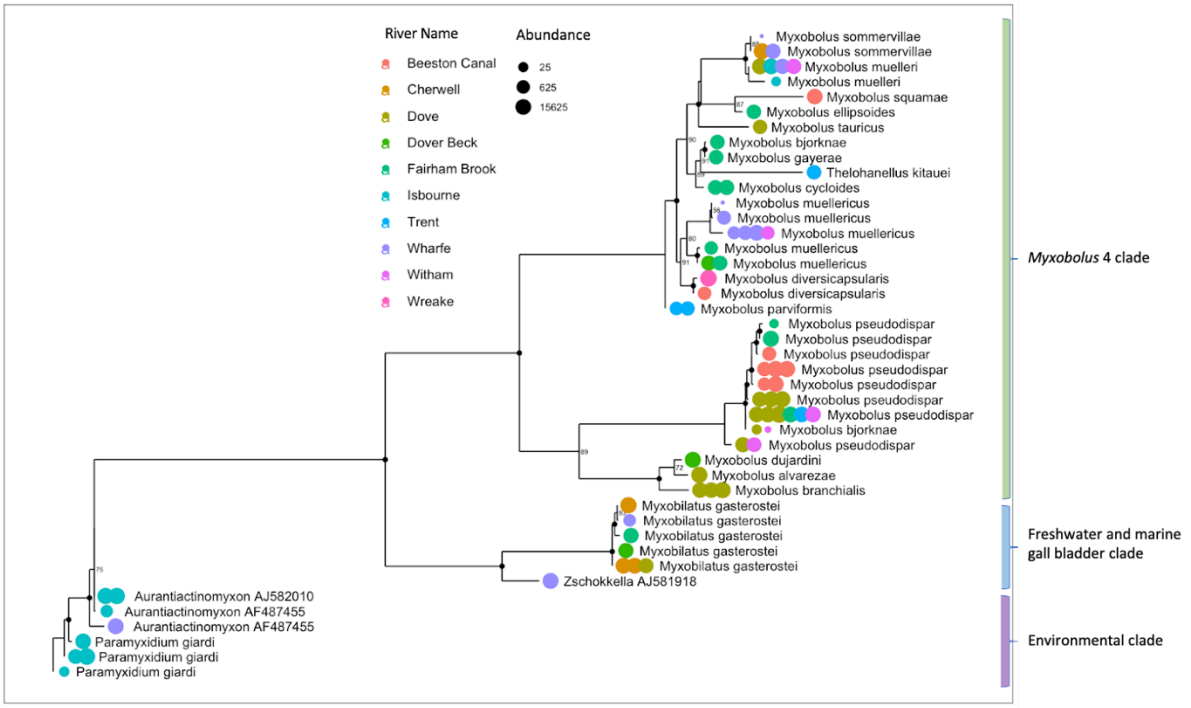

Fig 4. Phylogenetic tree illustrating the species and species complexes of myxosporean communities present in the sampled UK rivers. The relative abundance of each species is visualised by the size of each point, and the colour of each point represents the rivers in which said species were present. Each species is also categorised (right) into their phylogenetically inferred clades by colour (Hartikainen *et al*. 2016).

3.2 Myxosporean communities across pollution impacted rivers

Myxosporean ASVs belonging to the *Myxobolus* 4 Clade were most frequently found (see fig. 5). From a total of 60 effluent waste river samples, 28 samples showed amplification of myxosporea; 22 included amplification of species from the *Myxobolus* 4 Clade, 8 from the "Freshwater and Marine Gall Bladder Clade" and 4 from the "Environmental Clade". Representatives of the "Freshwater and Marine Gall Bladder Clade" and the "Environmental Clade" dominated the samples in the Rivers Cherwell and Isbourne. Strikingly few sites contained a mix of clades, despite ASV and species level variation among sites.

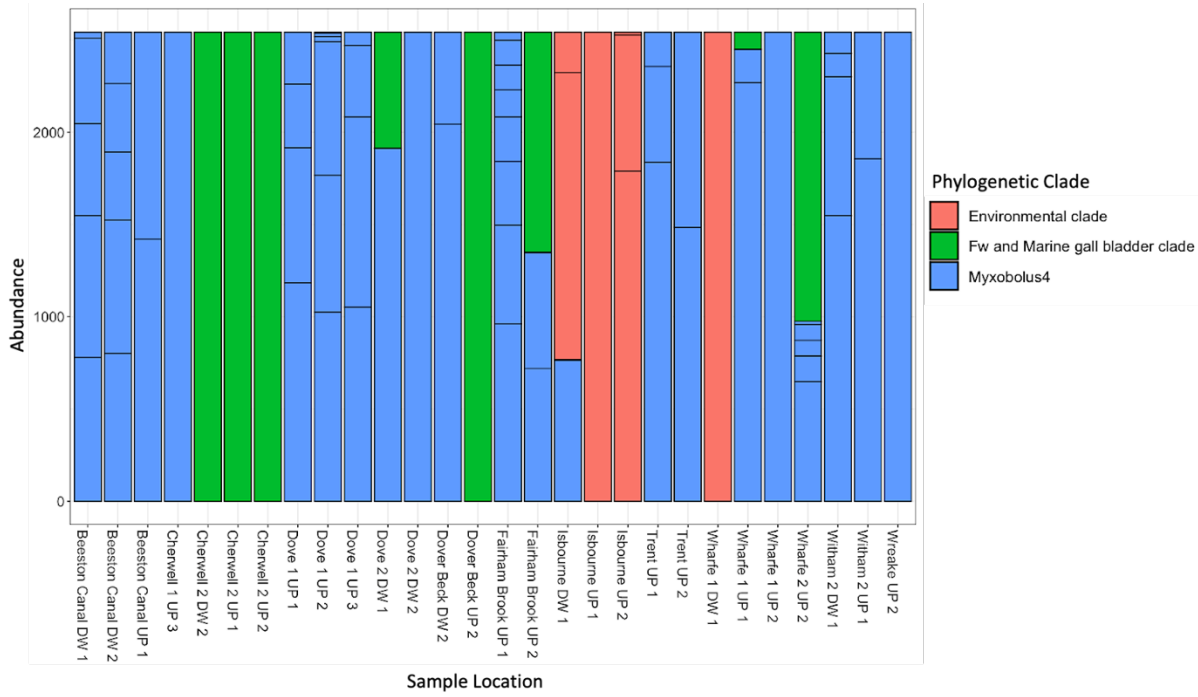

Fig 5. Barplot of all myxosporean ASVs at a phylogenetic subclade level in relation to UK river sampling location and replicate, with relative abundance after normalisation. Subclades were estimated in correspondence to previous phylogenetic data and analyses (Hartikainen *et al*. 2016, Liu *et al*. 2019).

Data was further resolved to indicate the genera of species present in samples (fig.6). Over half (15 of the 28 samples) were completely dominated by species from the *Myxobolus* genera, and 6 sites saw *Myxobolus* presence along with other genera such as *Aurantiactinomyxon*, *Myxobilatus*, *Paramyxidium*, *Thelohanellus* and *Zschokella*. When divided into sample position (upstream and downstream), all 6 genera were found to be present in at least one or more upstream samples. In contrast, only three genera were notably present in downstream samples (*Myxobolus*, *Myxobilatus* and *Aurantiactinomyxon*) and extremely low levels (approximately 100 reads) of *Paramyxidium* at one site. It is important to note that there are fewer downstreams sites (9) compared to upstream (19), as 4 downstream samples were not selected due to no amplification before sequencing, and 13

were lost during filtering post sequencing for having little to no amplification (<1100 reads). This contrasts with 2 upstream samples not being selected for sequencing, and only 8 being removed for having insufficient amplification post sequencing.

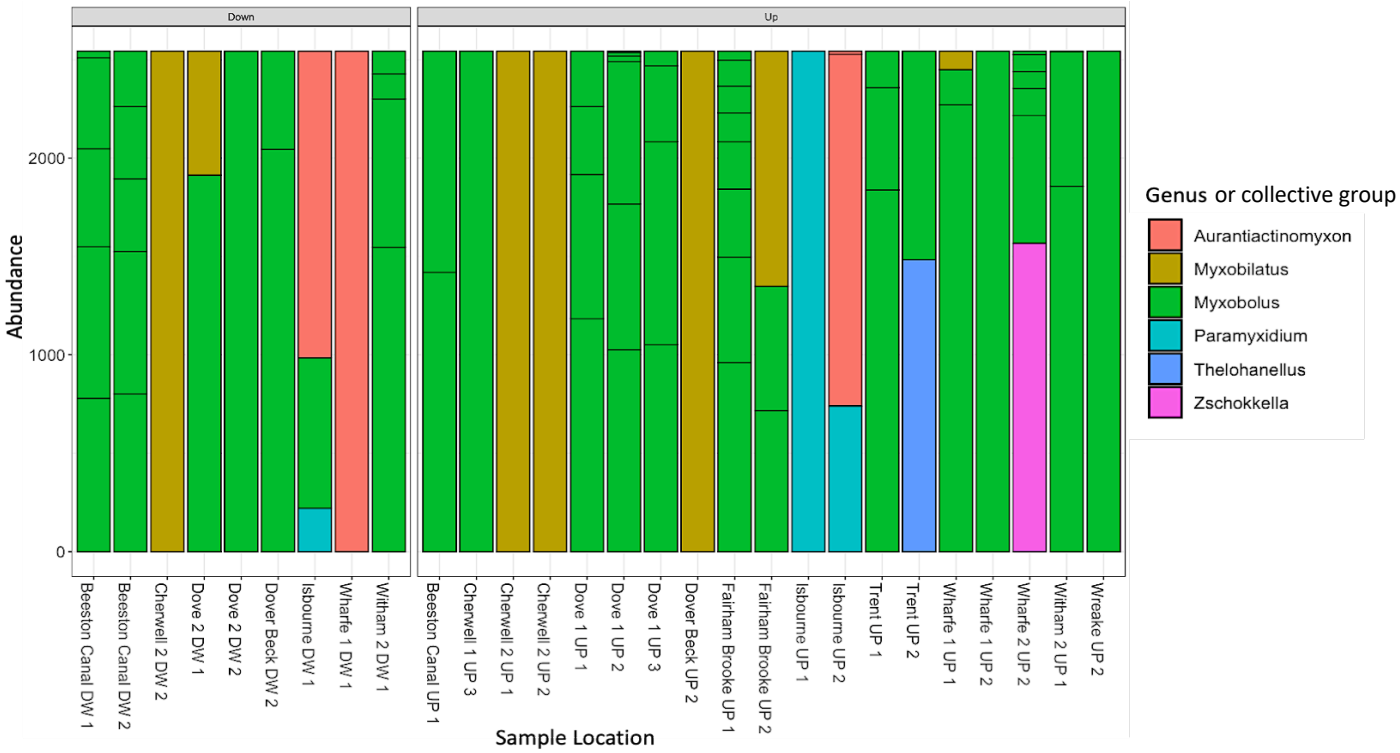

Fig. 6 Comparison barplot of myxosporean diversity at a genus/collective group level (*Aurantiactinomyxon* is not a genus but instead a collective group as described by Lom and Dykova, 2006), subdivided into downstream (left) and upstream (right) sample locations and replicates in relation to effluent waste outlets, with relative abundance after normalisation, and colour coded in correspondence to observed genus/collective group.

Data was further resolved to a species level to identify which species were present at each site and in relation to sampling position (see fig. 7). The most frequently observed species were *Myxobolus pseudodispar*, occurring in 13 sampling locations, 8 of which were upstream. *Myxobolus muellericus* (*Myxobolus 4 clade*) and *Myxobilatus gasterostei* (freshwater and marine gall bladder clade) were also commonly identified, both occurring 7 times across sites, 5 of which were upstream locations in both instances. In the 7 times that *Myxobilatus gasterostei* occurred, it was the only species present in the sample a total of 4 times. Species from the environmental clade, *Aurantiactinomyxon* and *Paramyxidium giardi* were only ever identified in the River Isbourne (*Aurantiactinomyxon* AJ582010), Wharfe (*Aurantiactinomyxon* AF487455, downstream only) and the River Trent and at no other locations.

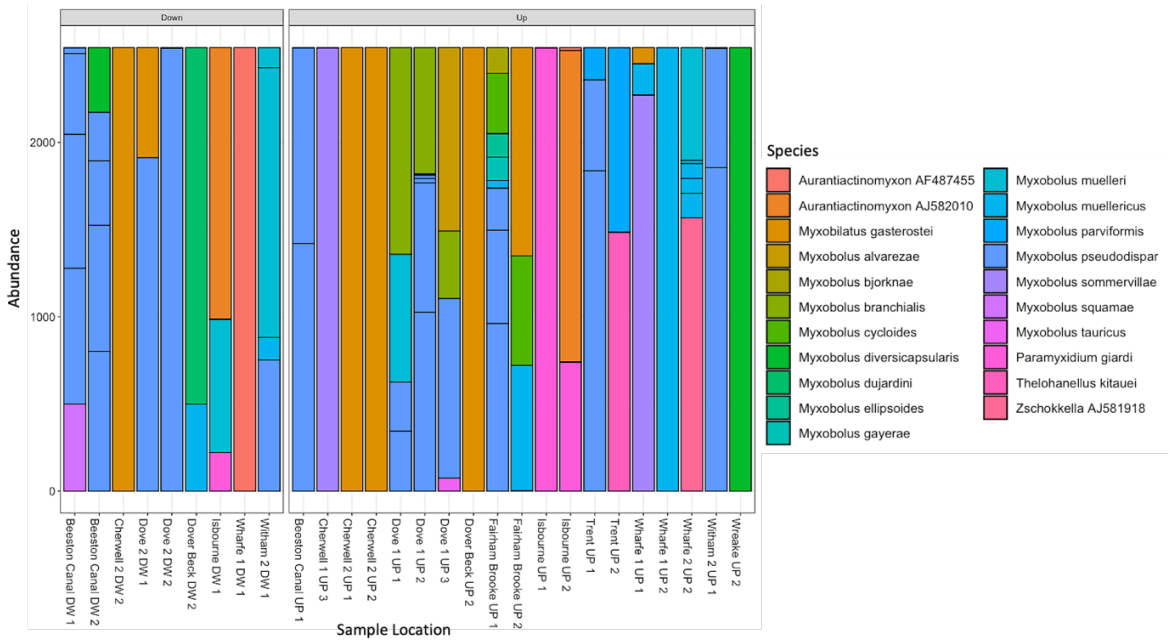

Fig. 7 Comparison bar plot of myxosporean diversity at a species level, subdivided into downstream (left) and upstream (right) sample locations in relation to effluent waste outlets, with relative abundance after normalisation, and colour coded in relation to observed species. Species names are inferred by comparing phylogenetic analysis, and matching OTUs with predefined species with more than a 92% phylogenetic match.

Fish host infection site was inferred from the species data (see fig. 8) in correspondence with Liu *et al*. (2019), and half of the species observed are responsible for infecting the gills (occurring 16 times) and muscles (occurring 12 times). The muscle infecting species included the prevalent *M. pseudodispar*. The gallbladder is also a common infection site, with gallbladder infecting species being identified at 7 sites.

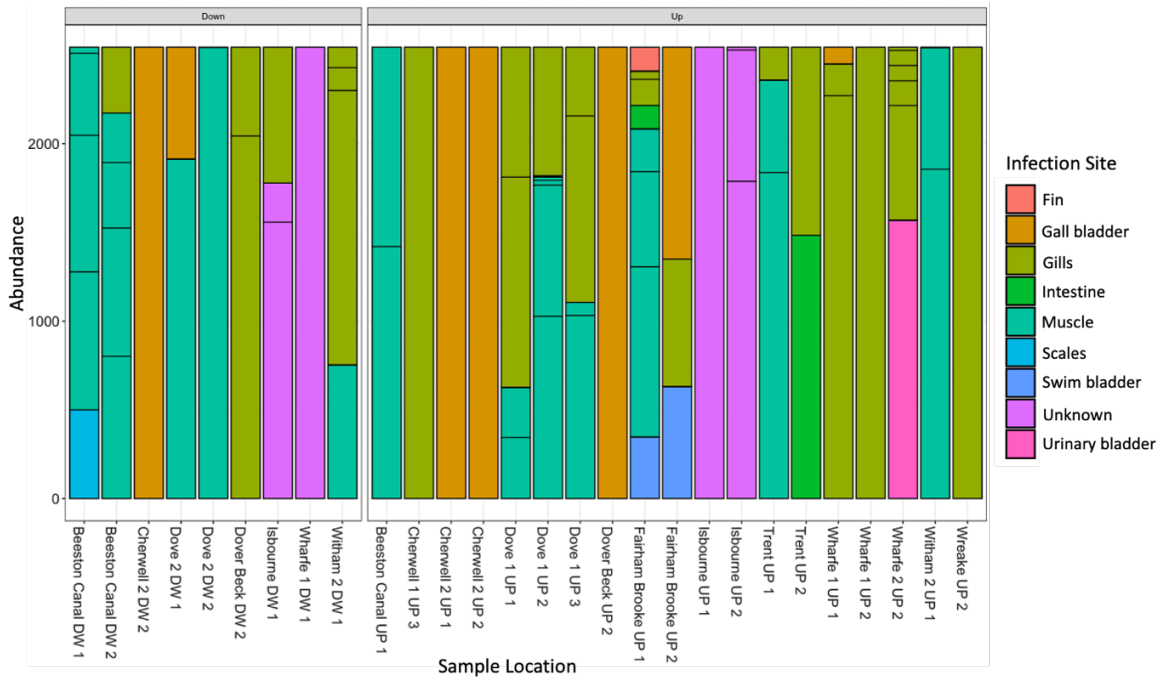

Fig. 8. Comparison barplot of myxosporean diversity inferring parasite infection site of fish host species, subdivided into downstream (left) and upstream (right) sample locations in relation to effluent waste outlets, with relative abundance after normalisation, and colour coded in relation to infection site inferred by Liu *et al*. 2019.

Known species and infection sites were then attributed to putative fish host species groups (see fig. 9), with the most frequently observed host group being cyprinids. Cyprinids were identified at 21 of the 28 sites, and sticklebacks at 6. There was also one occasion of eel being the inferred putative host group, at only one upstream Wharfe site. The putative host group at any Isbourne sites remains unknown, and this would require further life-history studies to understand the host specificity among the group of myxosporeans in the "Environmental Clade".

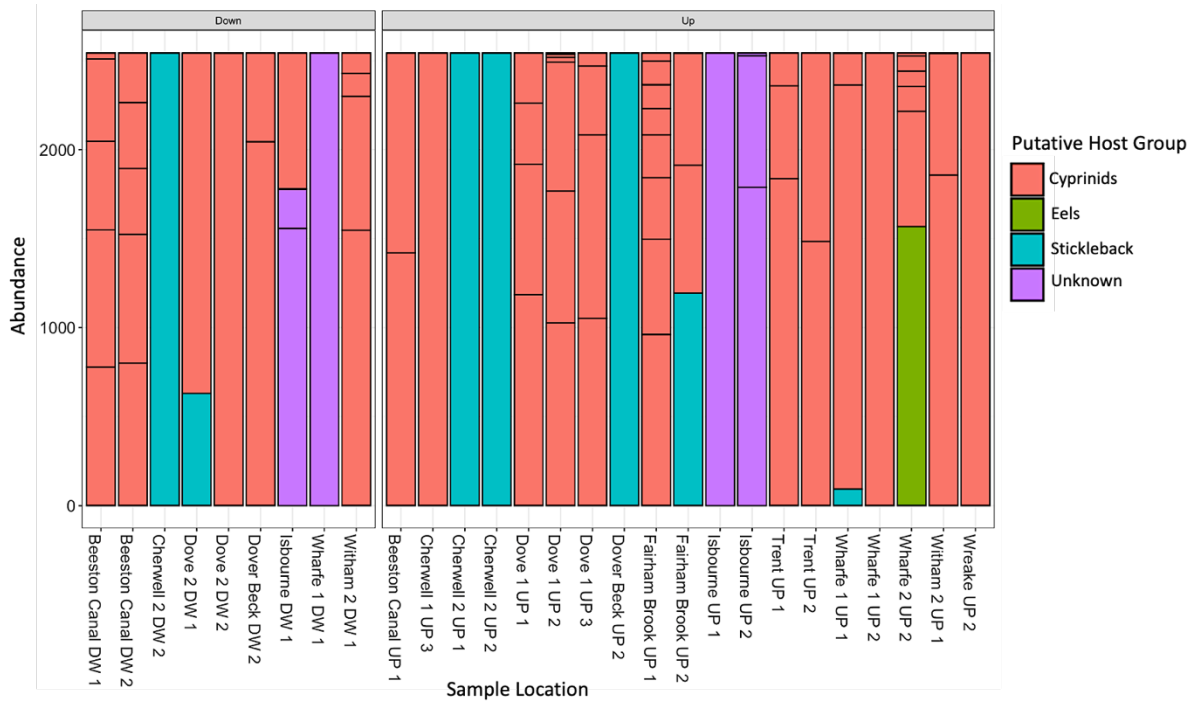

Fig. 9. Barplot of inferred putative host group across effluent waste sampling sites, with relative abundance normalised across all libraries. Each colour represents a different host group, defined by previous phylogenetic analysis (Liu *et al*. 2019).

Alpha diversity measures of myxosporean samples with Chao1 (number of species) and Shannon (number of effective species), in relation to sample position (see fig.10), upstream sites had a non-significant trend of higher levels of diversity measures in both plots.

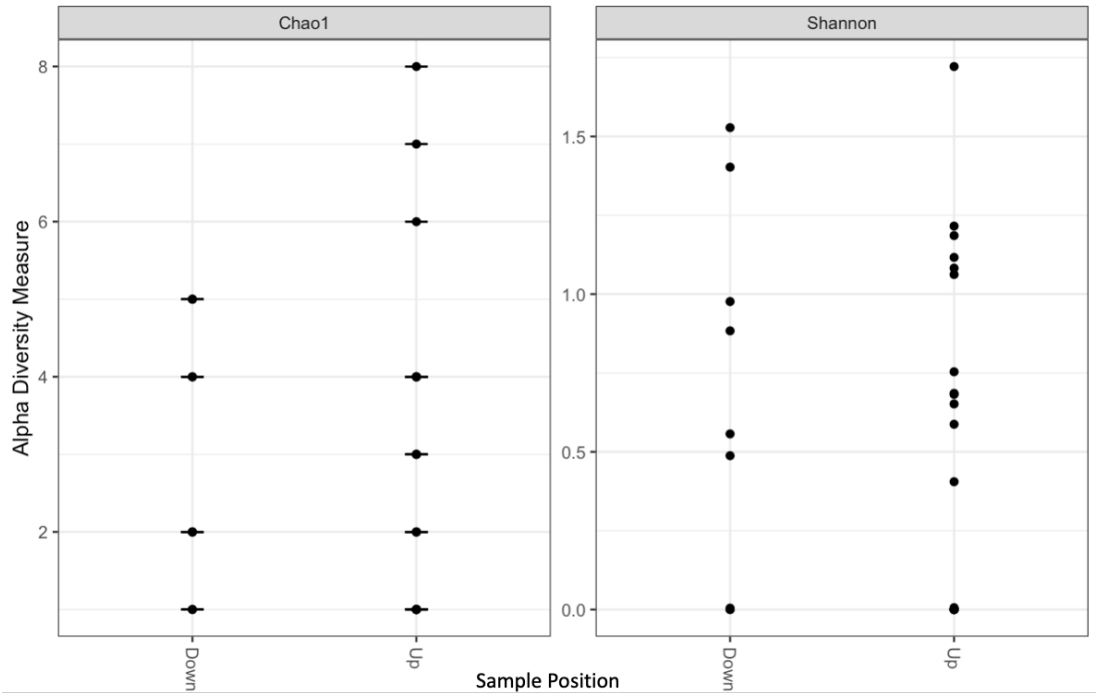

Fig. 10. Alpha diversity measures of myxosporean data using Chao1 and Shannon to infer observed species richness between upstream and downstream sites from CSOs.

Welch's two-sample t-test for myxosporean diversity data (number of individual species) at sites upstream and downstream from CSOs revealed that data was not significantly different between upstream and downstream sites. ***p = 0.7978, df = 66.869.***

Scatter plots were mapped to show the relationship between hourly annual overflow rate and number of observed individual species found at each site (see fig.11). Upstream sites with no annual overflow had significantly higher levels of diversity, with some sites having up to 6-8 individual species observed. As overflow rate increases, species diversity decreases, and any sites with >1250 hours per annum of overflow had no species present.

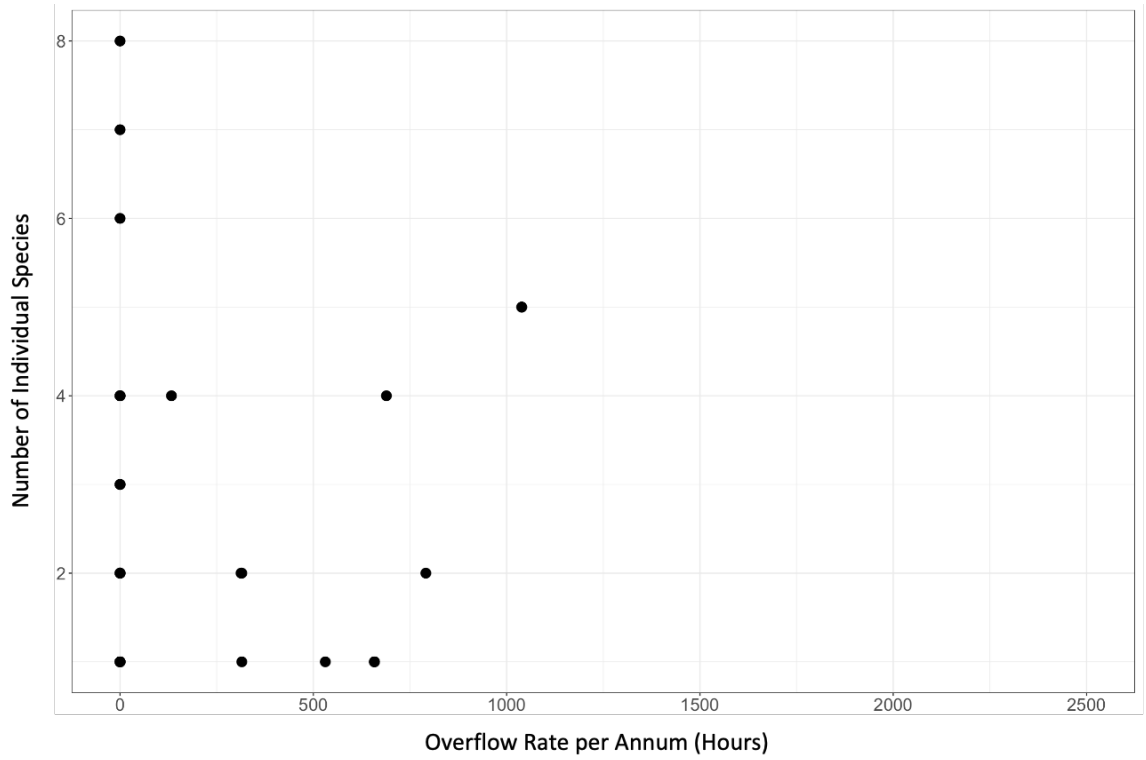

Fig. 11. Scatter plot of myxosporean diversity inferring number of observed species richness per sampling location (upstream sites receiving 0 hours of overflow, and downstream sites receiving anything above 0 hours overflow) (Chao1, Shannon) in relation to effluent waste overflow rate measured in hours per annum (2020-2021) after samples were pooled by replicates to represent a single sampling location, combining upstream and downstream sites.

The relationship between hourly annual overflow rate and number of observed individual species found at each site for myxosporeans underwent statistical analysis and found that the data was Poisson distributed, and therefore not normally distributed. After running Poisson regression and producing an ANOVA table, the relationship between the two variables was found to be statistically significant: p ***= 0.0129, df = 1.***

Principal Coordinates Analysis (PCoA) revealed no clear similarity in community composition or a role for relative abundance in driving similarity between sampling sites (see fig. 12,13). The samples from Rivers Isbourne and Wharfe appear as dissimilar from all other samples, suggesting that these sites had a different composition of parasite communities, as also suggested by the presence of the “Environmental Clade” representatives in these sites.

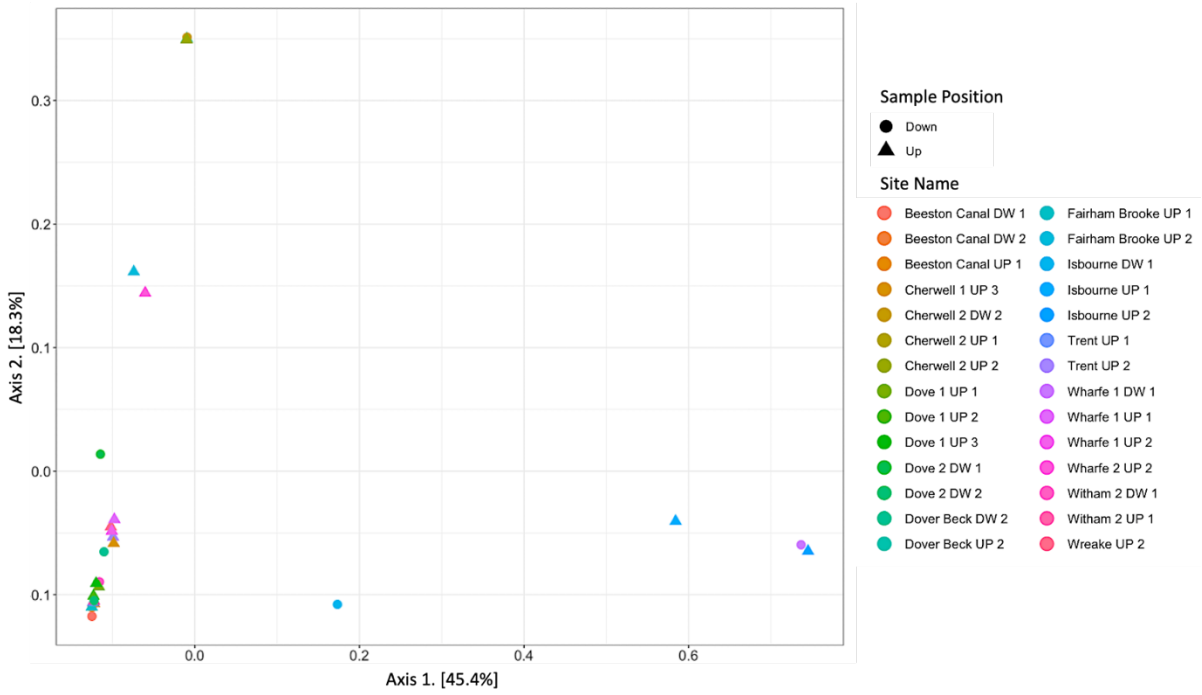

Fig 12. Principal coordinates analysis (PCoA) of relationship between sampling site and sample position at upstream and downstream sites, considering relative abundance of species shared between samples (weighted).

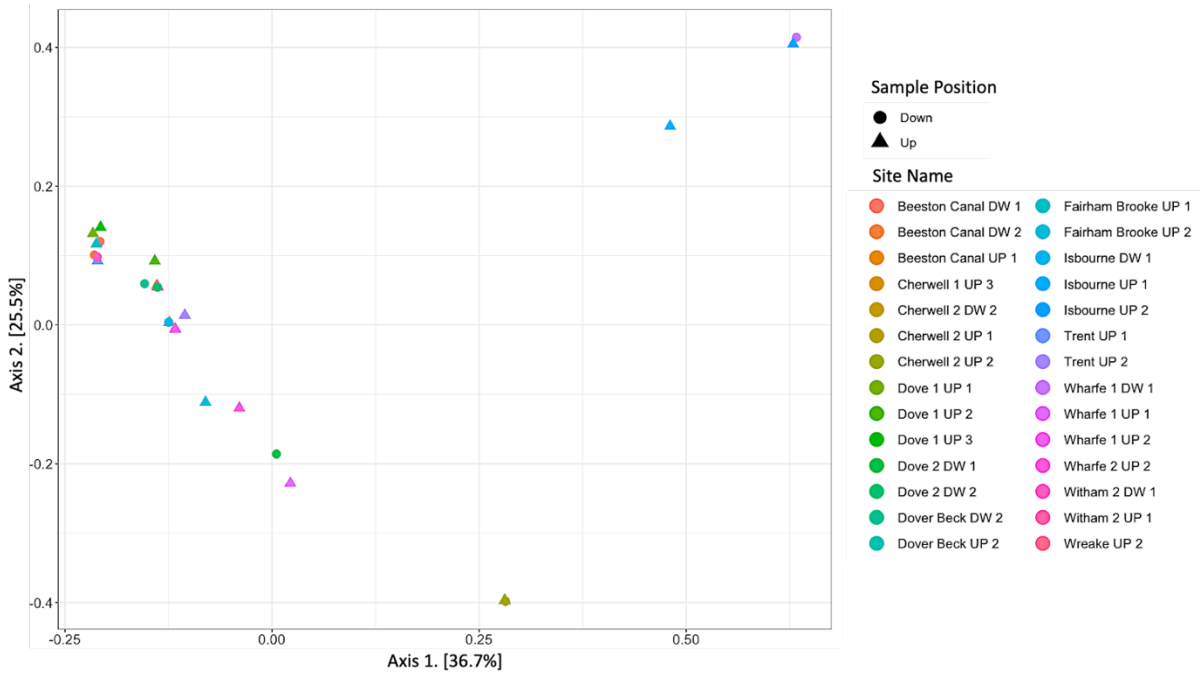

Fig 13. Principal coordinates analysis (PCoA) of relationship between sampling site and sample position, excluding relative abundance or species, and instead taking presence/absence into consideration (unweighted).

3.3 Malacosporean communities in pollution impacted rivers

Two genera of malacosporeans were inferred by phylogenetic analyses, *Buddenbrockia* and *Tetracapsuloides*, with a small contribution of unclassified genera in one site (Dove) (see fig. 14). ASV's belonging to the *Buddenbrockia* genus were strikingly prevalent across effluent

waste sites, dominating 13 of the 17 samples. In very few cases was there a mix of genera (in just 4 of the 17 sites), with low relative read counts from *Tetracapsuloides*. The ASVs attributed to *Tetracapsuloides* are all identified by phylogenetic analysis and *T. bryosalmonae*, suggesting the presence of this salmonid fish specialist parasite in the Rivers Witham and Dove. Significant variation in within-river occurrence of particular species was observed. For example, the sample taken upstream from the effluent waste overflow in the River Witham was entirely dominated by *Tetracapsuloides,* whereas the downstream sample was dominated entirely by *Buddenbrockia* (250m apart). Similarly, upstream sites of the River Dove's first sampling location consisted of *Buddenbrockia*, whereas upstream of the second sampling location (at a different waste overflow slightly further up the river) was dominated by Tetracapsuloides.

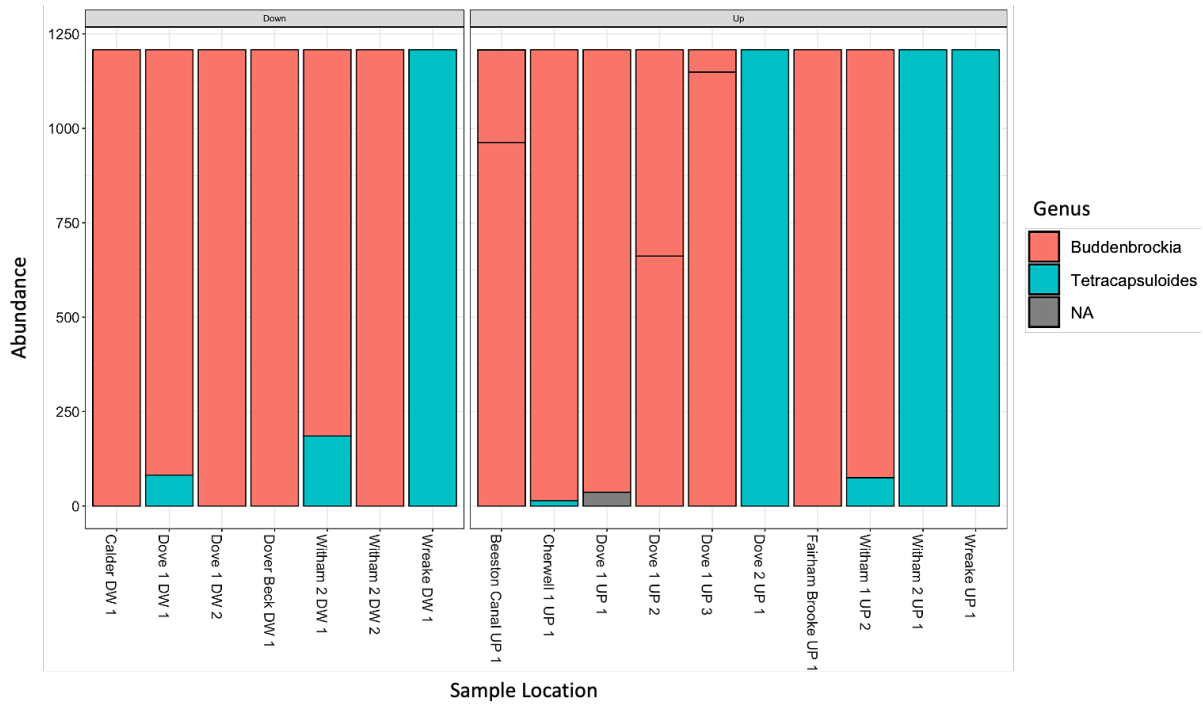

Fig. 14. Barplot of malacosporean ASVs at a phylogenetic genus level of UK river sampling location and replicate, with relative abundance after normalisation. Genera were estimated from automatic annotation against the PR2 database.

When placed in a phylogenetic context, the malacosporean ASVs formed three well supported groups, corresponding to previously sequenced isolates (phylogenetic analysis not shown, annotated ASVs plotted to sites in fig. 16). *Buddenbrockia worm fs2* (undescribed worm-like *Buddenbrockia* previously found in the bryozoan host *Fredericella sultana*), which was present in 10 of the 17 sites. *Tetracapsuloides bryosalmonae*, also known to infect *F. sultana*, was present in 9 samples. It must be noted that 5 of these observations consisted of < 250 reads per site. In 5 of the samples, ASVs grouping with *Buddenbrockia plumatellae* worm sequences were identified. Interestingly, these were the

most prevalent species of upstream River Dove sites (site 1), but were not amplified in any downstream samples of the same sampling site. The downstream sites were instead dominated by *Buddenbrockia worm fs2*. *Buddenbrockia plumatellae* sacs were also identified, but only once at one upstream site in the Beeston Canal. Due to the limited number of species that have been phylogenetically resolved for malacosporeans, it is difficult to infer whether there is more species variation upstream from effluent waste sites compared to downstream from this barplot, as the presence of *Buddenbrockia plumatellae sac* is the only notable difference between sampling locations upstream and downstream (Fig 16).

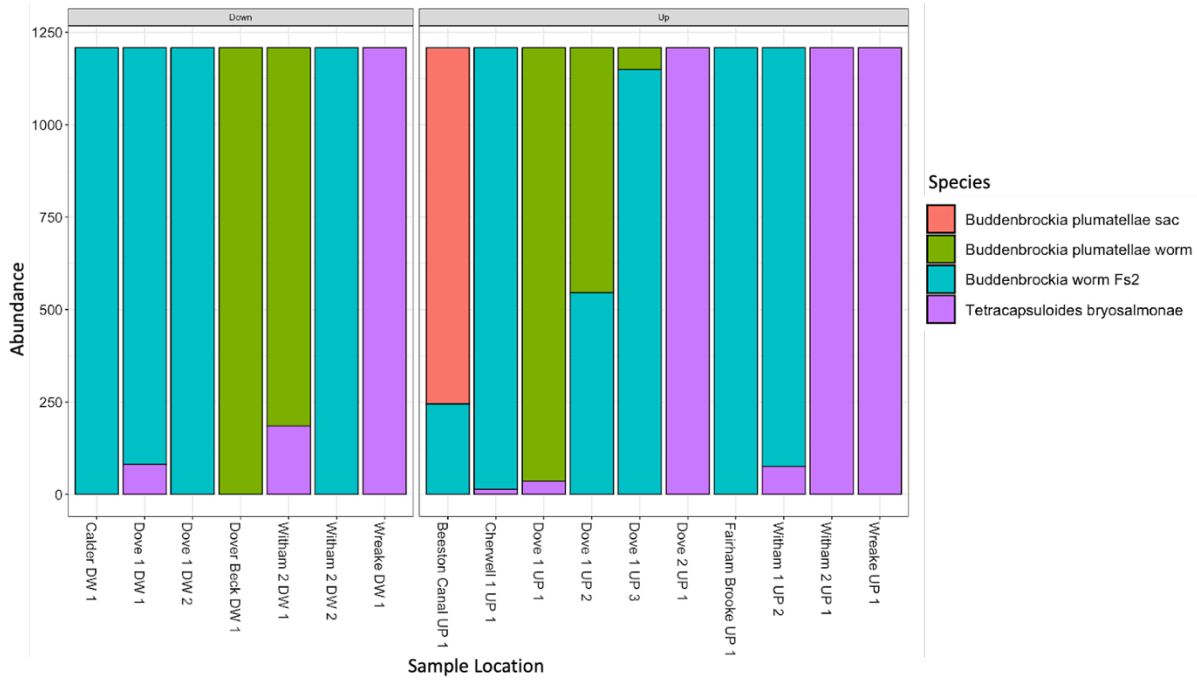

Fig. 16. Comparison barplot of malacosporean diversity at a species level, subdivided into downstream (left) and upstream (right) sample locations in relation to effluent waste outlets, with relative abundance after normalisation, and colour coded in relation to observed species. Species names are inferred by comparing phylogenetic analysis, and matching OTUs with predefined species with more than a 92% phylogenetic match.

Upstream sites had higher alpha diversity measures (Fig. 17) both in Chao1 (number of species) and Shannon (number of effective species) metrics. The relationship between hourly annual overflow rate and number of observed individual species found at each site (Fig. 18) was broadly negative. It was not possible to test the significance of the trend due to low species diversity, but sites with no annual overflow were the only sites to harbour all of the 3 individual species observed. As overflow rate increased, species diversity decreased, and any sites with >1000 hours per annum of overflow had no malacosporean species present (Fig. 18).

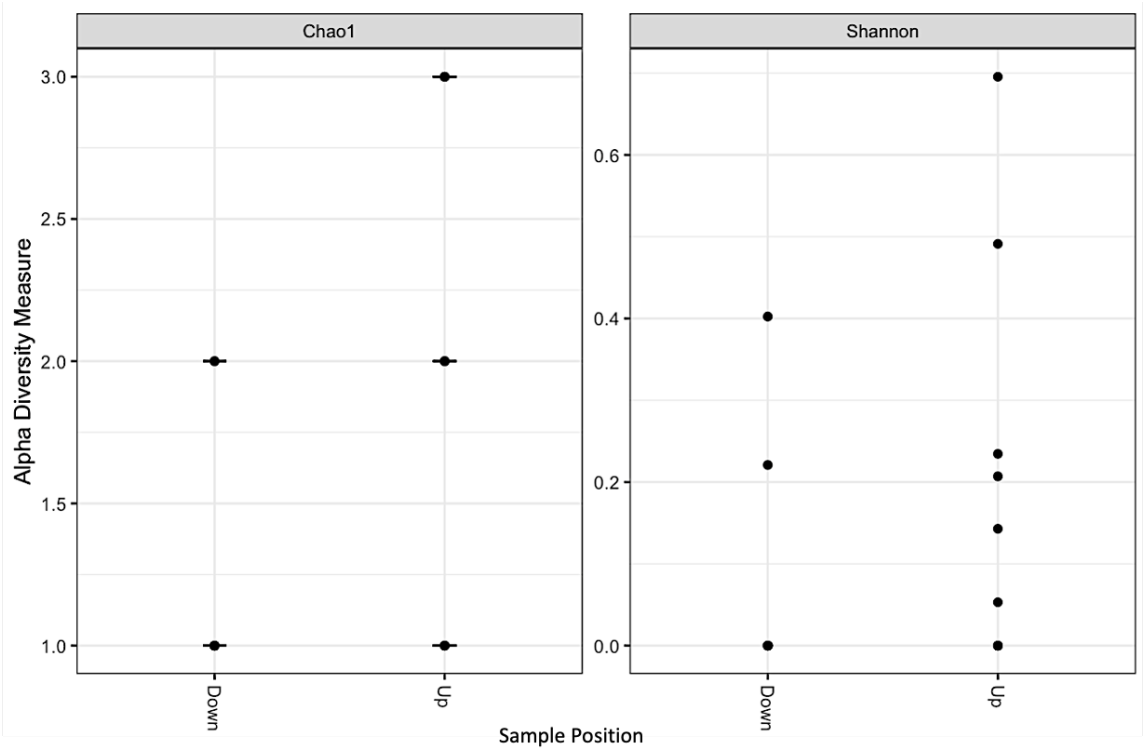

Fig. 17. Alpha diversity measures of malacosporean data from effluent waste sites using Chao1 and Shannon to infer observed species richness between up and downstream sites.

Welch's two-sample t-test for malacosporean diversity data (number of individual species) at sites upstream and downstream from CSOs revealed that data was significantly different between upstream and downstream sites: ***p = 0.004645, df = 45.741.***

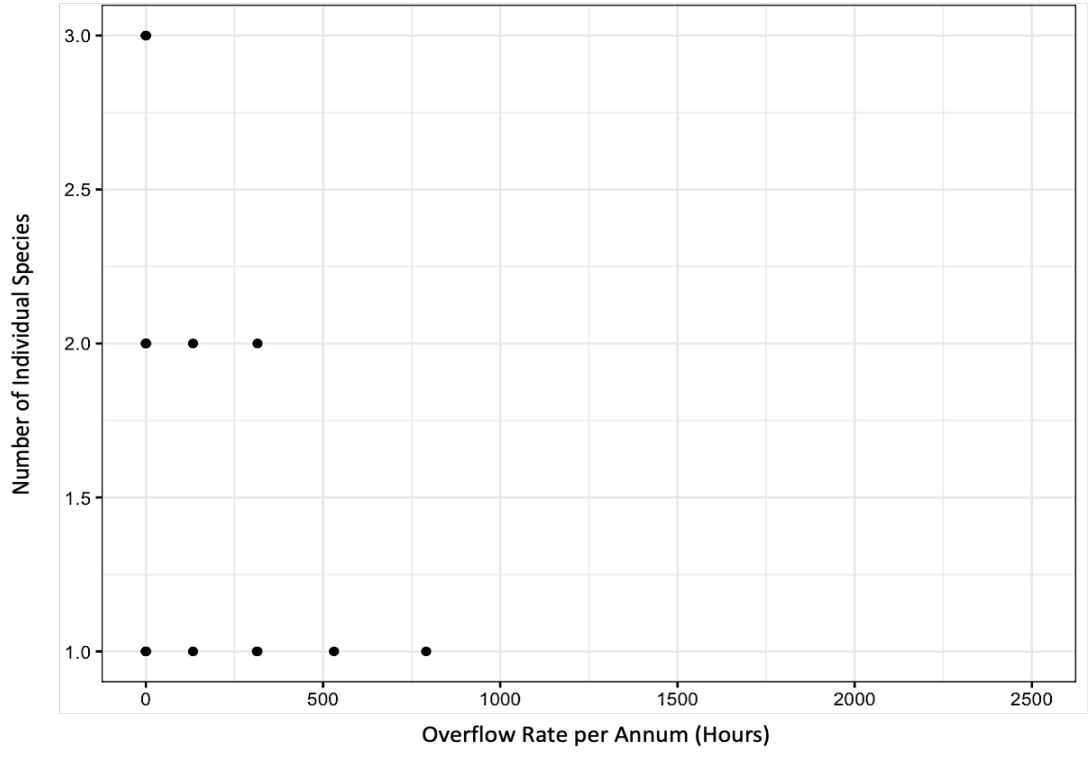

Fig. 18. Scatter plot of malacosporean diversity inferring number of observed species richness per sampling location (Chao1, Shannon) in relation to effluent waste overflow rate measured in hours per annum (2020-2021) after samples were pooled by replicates to represent a single sampling location.

The relationship between hourly annual overflow rate and number of observed individual species found at each site for malacosporeans underwent statistical analysis and found that the data was Poisson distributed, and therefore not normally distributed. After running Poisson regression and producing an ANOVA table, the relationship between the two variables was found to not be statistically significant: ***p > 0.291***, ***df = 1.***

3.4 Myxosporean and malacosporean communities in the River Erewash

In the Erewash river, myxosporean ASV's across all sampling sites other than one (Gallows 1) were accredited to the *Myxobolus* 4 clade. *Myxobolus pseododispar* was present in 9 of the 11 sites that remained after sequencing and trimming, being the dominant species in 8 of the 9 aforementioned sites. *Myxobolus alvarezae* was the only species identified at the Gallows 2 site, and *Myxobolus diversicapsularis* was the only species identified at Pinxton sites, both of which are species that were only identified twice across all effluent waste sites (see previous section, including both upstream and downstream). There were never more than 2 species present at any given Erewash site, a notable difference to the species diversity recorded at upstream effluent sites (Fig 19, plot A).

The lack of within-site diversity was even more pronounced in malacosporeans, where all ASVs per site were annotated to the same reference sequence or species (Fig 19, plot B). Malacosporeans were found in two of the side streams of the Erewash, the Beauvale Brook and the Nut Brook, which lacked myxosporean detections. ASVs with high similarity to the sac-like *B. plumatellae* sequences were found only in the September Beauvale Brook samples. Across the malacosporeans, detection in September samples were more often positive than in the July samples. The Hallam Fields STW is located downstream of the site labelled "Sandiacre". The Hallam Fields STW has some of the highest levels of annual CSO overflow, at 8,354 hours in 2021. Sites downstream from the STW were almost entirely dominated by *Myxobolus pseudodispar*, with small numbers of reads for *Myxobolus muellericus* and *Myxobolus parviformis*. Notable decrease in occurrence of malacosporeans was also observed after this point (Fig 19, plot B).

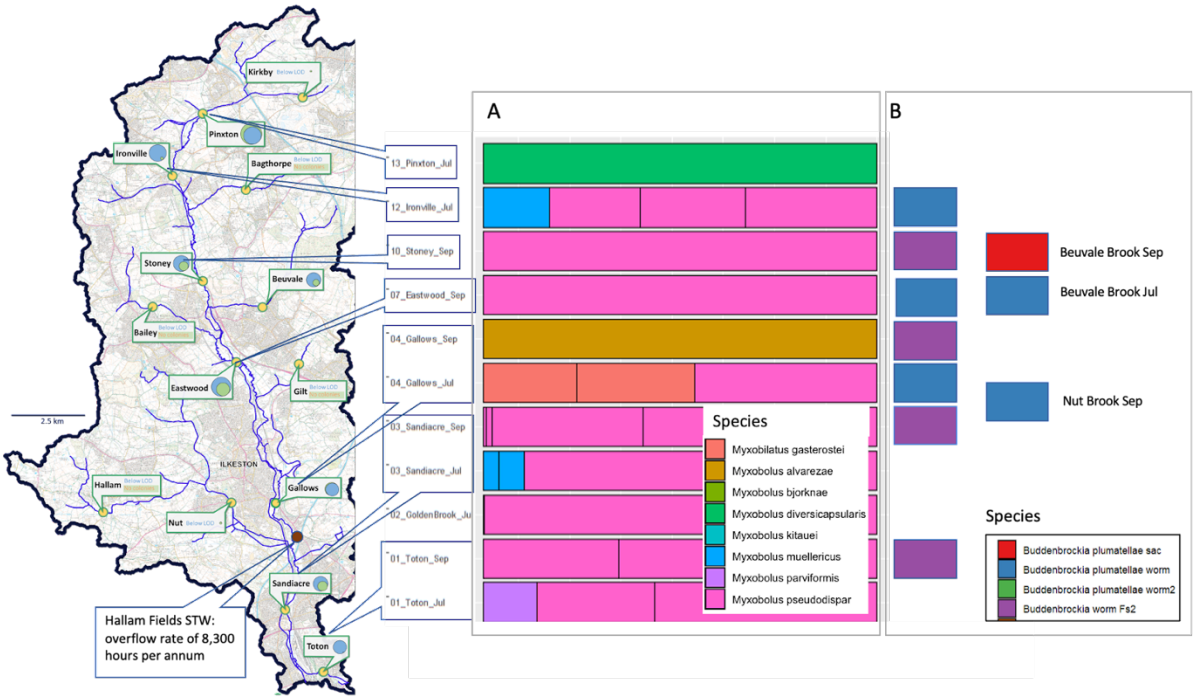

Fig. 19. Bar plot of myxosporean (A) and malacosporean (B) communities at different sampling locations of the River Erewash at a species level, with different colours relating to different species present, in relation to relative abundance, normalised throughout libraries. Each site that showed significant amplification is mapped and labelled in correspondence to its location along the River Erewash. Hallam Fields STW is also labelled in brown between sites 3: Sandiacre and 4: Gallows. The river flows from north (Kirkby) to south (Toton).

The target organ of infection from myxosporeans in the Erewash was muscle, and due to the prevalence of *Myxobolus pseudodispar* throughout the Erewash network. This contrasted with the effluent waste sites, which had a prevalence of pathogens known to infect the gills. The different dominant myxobolid species at the Pinxton sites, and at the Gallows sites were inferred to mostly infect the gill and the gallbladder of fish hosts. All of the malacosporeans identified to date were inferred to infect the kidney of the fish host, and use a bryozoan host in the life-cycle. The putative fish hosts in the Erewash consisted largely of cyprinids, with stickleback hosts inferred in the Gallows site, due to the one-time presence of *Myxobilatus gasterostei.* (see fig. 20)

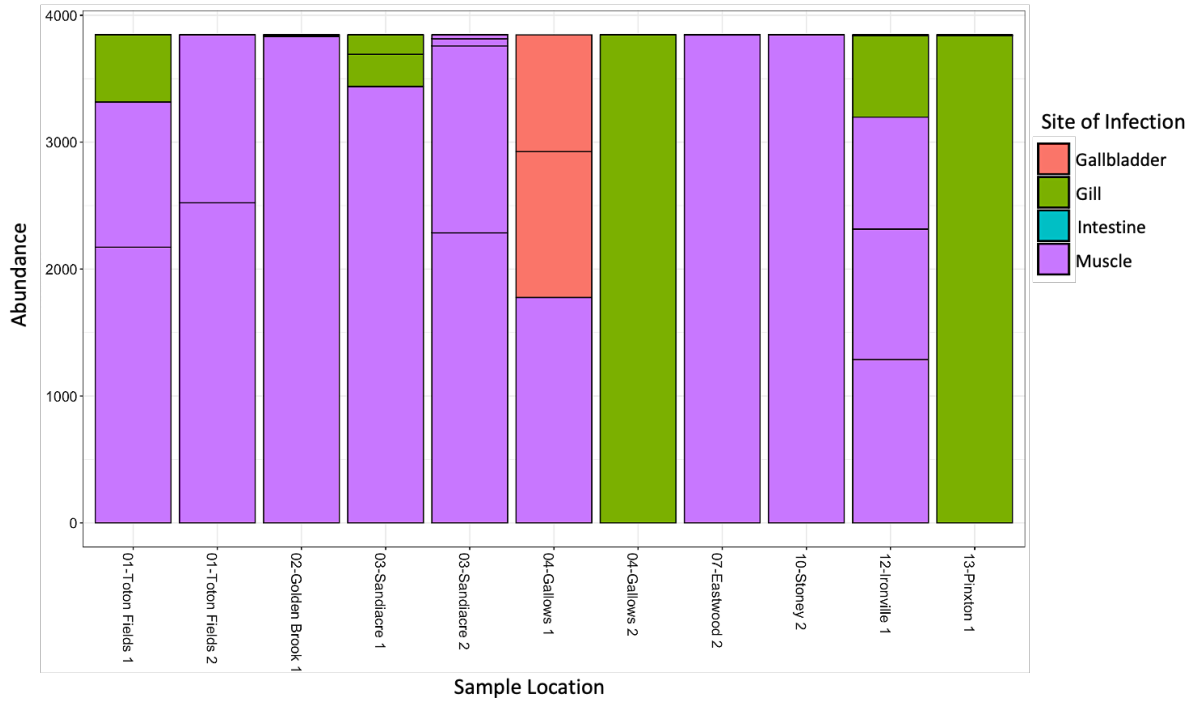

Fig. 20. Bar plot of myxosporean community data in relation to site of infection in fish host species and sampling location, with relative abundance after normalisation. Infection site was inferred from previous data and analyses (Liu *et al*. 2019)

4. Discussion

4.1 Detection of Myxozoans by eDNA

The methods utilised in the current study showed varied levels of success with environmental DNA (eDNA) and High Throughput Sequencing (HTS) in assessing myxosporean and malacosporean presence and diversity in environmental water samples in relation to combined sewage overflows (CSO's). Macroinvertebrate data was not further resolved after sequencing. In other studies, the BF/BR primer pair has resulted in 40-60% of metazoan amplicons (Leese *et al*. 2020), with variation between sites. It is possible that our samples had extremely high bacterial loads, swamping amplification of rare metazoan targets. Further optimisation of the annealing temperatures, and PCR conditions could be used to improve the performance of BF/BR primers, as well as an improved SPRI cleaning protocol as this proved less successful for these primers.

The lineage-specific myxosporean assay, however, successfully detected myxosporean ASVs across all sampling sites (CSO sites and Erewash sites). Phylogenetic placement of the ASVs indicated a moderate level of taxonomic diversity in myxosporeans, implying success in the selected pooled primer combinations, with taxa successfully detected from three different lineages. One of the most widely described and most species-rich genera within Myxozoa in the current study is *Myxobolus*, with more than 850 species identified currently worldwide (Liu *et al*. 2019) and previous studies have confirmed the prevalence of

these clades in eDNA samples from other sites in the UK (Hartikainen *et al*. 2016). This level of taxonomic diversity within the myxosporean ASVs supports and further describes more cryptic species of myxosporeans, such as those of *Myxobolus pseudodispar*, and highlights the need for lineage-specific assays such as in the current study to produce data down to the level of ASVs, to be able to understand species variation at a deeper level.

Together however, the results for both parasite primers showed high levels of non-target species and suggests that inclusion of the general eukaryote 1st step PCR should be re-evaluated, and a pre-sequencing screening with quantitative PCR or Sanger sequencing may be a cost-effective way of reducing the inclusion of libraries mostly composed of contaminant sequences.

4.2 Diversity and distribution of Myxozoans in UK rivers

The phylogenetic analysis in the current study supported the prospective assignment of sequence similarity via blast searches, as shown by the grouping of ASVs annotated to each Genbank reference sequence. For instance, *Myxobolus pseudodispar* is composed of 8 different ASVs, some site specific, and one widespread ASV. The most common *M. pseudodispar* ASV was present in 4 different rivers, the River Dove, Fairham Brook, the River Trent and the River Witham, with Beeston Canal and Fairham Brook harbouring a unique *M. pseudodispar* ASV each. *M. pseuodispar* is considered a generalist species complex, with a wide range of hosts, primarily different cyprinid species (Lisnerová *et al*. 2020). Cyprinids are common in the sampled area, although species level fish-host occurrence data is currently not available. It is notable that *M. pseudodispar* isolates from different fish samples show up to 5% differences in the SSU rDNA sequences (Forró and Eszterbauer, 2016). This is a significantly large intraspecific difference for any Myxozoans, and very few of the other muscle-dwelling myxosporean species have radiated to the extent of *M. pseudodispar* in this way (Lisnerová *et al*. 2020, Farro and Eszterbauer, 2016).

Moreover, sequence analysis of *M. pseudodispar* by Lisnerová *et al*. 2020 showed clear phylogenetic grouping according to fish host species, forming 13 well recognised clades. From this, we can infer that the ASVs of *M. pseudodispar* found in the current study, as opposed to being considered as a generalist species may in fact represent multiple host-specialised species, instead occurring according to their secondary cyprinid host species distributions, signifying host species variation, as *M. pseudodispar* has radiated to different host species. This supports the theory of a deeper split between *M. pseudodispar* species, as previously described by Lisnerová. However, further resolution of ASVs is required to support this statement.

In addition, Lisnerová *et al*. 2020 also states that although *M. pseudodispar* were initially thought to only infect muscle, infection of other tissues such as the gills also occurs. This factor may also be an explanation for the high detection probability of *M. pseudodispar* in eDNA samples and demonstrates the utility of combining host-specific and eDNA sampling for understanding Myxozoan radiations, and the significance of variation seen between ASVs.

Three different, closely related *M. pseudodispar* ASVs were present only in the Beeston Canal and no other rivers sampled. The unique occurrence of these parasites in the canal site may indicate that the host species for these pathogens favour the environmental conditions in the canal. This is supported by the occurrence of the sac-like *Buddenbrockia plumatellae* in the canal, known to infect the bryozoan *Cristatella mucedo*, which is a common species in slow moving rivers and in lakes. The Beeston canal is a small, 5-mile-long man-made canal (Canal and Rivers Trust) that starts at Beeston Lock, and runs alongside the River Trent before rejoining further downstream. Either end of the canal is frequently obstructed by two large lock gates which allow boats to reach higher levels of the canal, and travel upstream. These gates are closed whenever boats are not passing, therefore giving the canal unique hydrology and connectivity by obstructing the flow of the canal and preventing any significant flow-through from the River Trent (Morely and Lewis, 2006). A study conducted by Walker and Hassall (2020) observed the effects of water chemistry and lock-mediated connectivity on macroinvertebrate diversity and community structure in the Leeds-Liverpool canal in England and found that the community structures of macroinvertebrates varied considerably between groups of sampling sites separated by locks, suggesting that locks may act as barriers for aquatic species by restricting dispersal. The study also observed that as concentrations of chemical stressors such as metals and dissolved carbon from effluent overflows increased in the canal closer to the centre of Leeds, the diversity of macroinvertebrates decreased (Walker and Hassall 2020).

Myxozoans rely on their secondary vertebrate fish hosts for their dispersal across different aquatic systems, where they can be shed as free-living mature spores to repeat their life-cycle. A reduction in access to different parts of a river system for fish can in turn reduce access for myxosporeans. Findings such as these help to highlight the importance of emerging techniques such as the ones described in this study for the biomonitoring of aquatic ecosystems.

Only 14 malacosporean ASVs were detected in this study, reflecting the known low diversity in this clade (Hartikainen *et al*. 2014). Further malacosporean putative species have been discovered in European rivers by sampling fish tissues, and the diversity of malacosporeans is expected to be much vaster than known currently (Bartošová-Sojková *et al*. 2014). The reason for this hidden biodiversity within malacosporeans can likely be attributed to the cryptic nature of malacosporean sporogonic and presporogonic life stages and mostly asymptomatic infections in their fish hosts (Bartošová-Sojková *et al*. 2014). In this study, all the detected malacosporeans have been observed previously within their bryozoan hosts, and none of the malacosporeans found from fish hosts only were recovered from the eDNA samples. This highlights the role of the parasite life-cycle in driving detection probability in eDNA samples; fish generally release very few malacospores with their urine, whereas spore release from bryozoans occurs in large flushes of spores. Malacosporeans were also not as widely spread as myxosporeans, with only 46% of sampling locations showing presence of malacosporeans, compared to 87% showing presence of myxosporeans. This result could derive from the lower diversity of malacosporeans, but also from their reliance on phylactolaemate bryozoans as hosts, which are likely to have a more restricted distribution in the sampled rivers than the oligochaete hosts of myxosporeans. An interesting second approach to the current study would be to assess the presence of bryozoans in these sampling locations, to understand whether they are responsible for shaping the malacosporean communities that are present or whether there is an external pressure involved, such as increased secondary host diversity, or preferred hydrology.

The ordination analyses identified weak dissimilarity between sites, however, suggesting that the communities from the Rivers Cherwell and the Isbourne differed from the other sites sampled.

The Cherwell hosted a low diversity Myxozoan community at both upstream and downstream sampling locations from the CSO dominated by *Myxobilatus gasterostei*, a parasite of stickleback fish. The Cherwell catchment is a clay dominated catchment (Neal *et al*. 2006) and much of the river is associated with direct run-offs from the agricultural surrounding areas, rather than from groundwater aquifer sources (Neal *et al*. 2006), resulting in increased levels of sediments within the river system. Previous studies have demonstrated that macroinvertebrate fauna of rivers and streams is affected by sedimentation via various means, such as loss of habitat, burial of food sources, accumulation of sedimentary particles on respiratory and feeding structures and burial of the invertebrates entirely (Martin and Neely, 2001). Such conditions may also favour the

dominance of pollution tolerant taxa, such as tubificid worms and other sediment-dwelling oligochaetes, which may have acted as a selective filter on the myxosporean communities in the river. It may be the case that the only species present, *Myxobilatus gasterostei*, and their secondary hosts, stickleback (Gasterosteidae), have a unique way of persisting within such an environment. The rapid colonisation of three-spined stickleback (*Gasterosteus aculeatus*) in freshwater environments has resulted in the accelerated evolution of morphological and behavioural traits (Bjærke *et al*. 2010), one example of which includes increased benthic foraging. The other adaptation explaining the dominance of *M. gasterostei* in these environments may be the ability of its primary hosts, naidid worms, to undergo fission and persist in fluctuating environments (Atkinson and Bartholomew, 2009).

The River Isbourne sites also appear to be unique compared to a majority of other CSO sites, and sites upstream from CSOs are completely dominated by myxosporeans belonging to the " Environmental Clade" (Hartikainen et al. 2016). The ASVs present within this clade upstream from the CSO are closely related to the reference sequences for *Aurantiactinomyxon AJ582010*, and *Paramyxidium giardi*, and both species are unique to this river system within the sites sampled in this study. The vertebrate host identity for *Aurantiactinomyxon* is unclear, and could represent fish, amphibians or even birds, but the vertebrate and invertebrate hosts for *Paramyxidium giardi* are known – *Anguilla anguilla* and *Tubificoides pseudogaster*, respectively (Rocha *et al*. 2019). The River Isbourne is 14 mile long tributary of the River Avon which flows through Gloucestershire and Worcestershire (Elliott *et al*. 2022), and is significantly more southerly than the other sampling sites in the current study. The soils and geology of the Isbourne consist of shallow fine loam over limestone, fine silty soil and deep impermeable clay (Clarke *et al*. 2016). Several studies have shown a preferred primary host species of myxosporeans such as *Aurantiactinomyxon* to be the oligochaete, *Tubifex tubifex* (Hallett *et al*. 2002), which is widely spread in most UK rivers. The absence of this clade from all other sites sampled in this study suggests that the representatives of the "Environmental clade" use a vertebrate host with a limited distribution among the study sites, that may favour the unique geology and soils of the Isbourne.

4.3 CSO impacts of Myxozoan communities

The malacosporean community composition did not show clear differences between sites upstream and downstream from CSOs, mainly due to the low number of ASVs detected overall. Among the myxosporeans, every genus was found at least once upstream of CSO sites, whereas only three were found at sites downstream from CSOs (*Myxobolus, Myxobilatus* and *Aurantiactinomyxon*. *Paramyxidium* were also found but at approximately 100 reads). The number of individual ASVs of myxosporeans and malacosporeans identified

per site was plotted in relation to annual CSO overflow and found that diversity and number of individual ASVs per site was higher at sites upstream from CSOs, or sites that received lower levels of pollution throughout the year, for both classes. From this, and the increasing trends in alpha diversity, we can infer that there may be increased levels of biodiversity of Myxozoans at upstream sites from CSOs.

It is important to note that despite more sites upstream from CSOs being selected for sequencing, the same number of upstream and downstream sites were screened initially. Several downstream sites showed no amplification of myxosporeans or malacosporeans, or showed very few myxosporean reads. For example, some sites upstream from the CSO, such as the Fairham Brook, hosted eight individual myxosporean ASVs, whereas the downstream samples of the same site were not selected for sequencing due to no evidence of myxosporean amplification. Initially, it was assumed that the lack of amplification in said samples was due to inhibition by an extremely high concentration of inhibitors present in the samples, such as effluent agricultural run-off, or simply the sewage itself. However, after an inhibition test was run by qPCR and inhibition was confirmed for a vast majority of the samples, all samples were diluted to a 1:10. It was after dilution and PCR amplification that it was noticed that still no Myxozoan amplification occurred at many downstream sites. From this, we can infer that there was a real absence of Myxozoans at downstream sites. A study conducted between 2006 and 2007 on the relationship between metazoan parasite species diversity and organic pollution in the upper Manyame catchment in Zimbabwe, found that parasite species richness in all unpolluted sites was high, while in polluted sites a maximum of one parasite species was encountered (Madanire-Moyo and Barson, 2010).

Statistical analyses in the current study both support and contradict the theory of a reduction in biodiversity downstream from effluent pollution. For myxosporeans, there was a significant correlation in the reduction of number of individual species as annual overflow increased, but not in the case of malacosporeans. In the case of number of species at sites upstream and downstream from CSOs, this was reversed, malacosporeans saw a significant relationship between upstream and downstream groups and species diversity, whereas myxosporeans did not. To improve these results for a more significant statistical outcome, it would be important to include all samples for sequencing and therefore all samples for analysis, to highlight presence and absence between the two groups (upstream and downstream). This would then also reveal the real absence of Myxozoans at sites downstream from CSOs.

The sites upstream from CSOs hosted several ASVs absent from the downstream sites, including *Myxobolus sommervillae* (River Cherwell and Wharfe), *Myxobolus branchialis* (River Dove), *Myxobolus alverazae* (River Dove), *Myxobolus cycloides* (Fairham Brook), *Myxobolus ellipsoides*, (Fairham Brook) and *Myxobolus gayerae* (Fairham Brook). The lack of such species downstream further supports the hypothesis that CSO effluent waste is associated with reduced parasitic diversity. The same cannot be said for malacosporeans, as there is only one significant difference in species identification between sites upstream and downstream from CSOs: the presence of *Buddenbrockia plumatellae sac* at one upstream site (Beeston Canal).

Myxobolus are taxonomically classified by their preference of infecting host tissues and are primarily responsible for infecting the gills and muscle tissues of their secondary fish hosts. They produce large polysporic plasmodia within said tissues, with the potential to cause severe disease (Liu *et al*. 2019, Lom and Dyková 2006). The previously mentioned study by Bernet *et al*. (2004) that focused on the effects of effluent pollution from STW on brown trout (*Salmo trutta*) by histopathological examination, found the gills to be the most sensitive organ to the effects of effluent waste water in fish caught downstream from and effluent outlet, in contrast to fish caught upstream. Such histopathological alterations included epithelial cell lifting, infiltration of the gill epithelium, alterations to the epithelial cytoplasm, deformations of lamellae and fusion to adjacent lamellae (Bernet *et al*. 2004). A typical, healthy histopathological cross section of most fish gills without *Myxobolus* infection or pollution related degradation indicates a neat order of primary and secondary lamellae, which can be easily differentiated (Maftuch *et al*. 2018). An infected gill shows congestion within the gill in the form of blood accumulation which forms a significant swell of thrombocytes, caused by venous blood stream decrease (Martuch *et al*. 2018) as well as edema within the lamellae, making the differentiation of lamellae more difficult.

With the majority of CSO sites in the current study being dominated by gill infecting species of *Myxobolus*, it would be interesting to consider whether the modification of gill structures by effluent pollution is in turn creating easier access to gill tissues for infection by *Myxobolus* spores. The polar filaments of mature myxosporean actinospores extrude to anchor to the various tissues of the host, and *Myxobolus* species have also developed spores with lateral flattening that enables easier access to tissues than their precursors that lived in body cavities. This very evident degradation of gill tissues along with improved physical attributes of *Myxobolus* may benefit the anchoring stage of infection, increasing the prevalence of *Myxobolus* species at CSO locations. The identification of many gill infecting species may

also be attributed to spores being more easily shed from the gills into the surrounding water column than alternative infection sites within the host's body, and therefore more easily collected within eDNA samples. Further studies are required to distinguish between these hypotheses. Such work could be extended to investigate the radiation of the *M. pseudodispar* species complex, which has a propensity for infecting muscles but also varies in the identity of the target organ. Bhanot and Hundal, 2020 found in muscular tissue of carp fish (*Labeo rohita*), following exposure to untreated sewage water, destruction of essential fatty acids. This raises the question of whether untreated effluent waste in general is acting destructively in terms of fish anatomy and physiology, reducing body condition and general fitness and increasing the prevalence of specific species of myxosporeans.

The complexity and diversity of parasitic life history strategies implies that their communities may incorporate adverse stresses that influence other aspects of the ecosystem, such as effluent waste. Therefore, parasites can be regarded as potentially sensitive indicators for ecosystem health and stability (Dzikowski *et al*. 2003, Hudson *et al*. 2006). Parasites with simple, single-host life cycles are known to persist perturbed environments, whereas species with complex, multiple host life cycles, such as Myxozoans, are more likely to only persist only within stable ecosystems, where both their free-living stages and secondary hosts are capable of survival (Dzidowski *et al*. 2003). Many studies have demonstrated the negative impacts of pollution on parasitic communities, which is reflected in a decline in species richness and diversity (Dzidowski *et al.* 2003, Dudgeon *et al.* 2018), as seen in the current study.

When observing reduced Myxozoan species diversity downstream from CSOs, it may be assumed that this reduction in richness is due to a change in macroinvertebrate communities. Several studies have found macroinvertebrate abundance to increase downstream from CSOs. Wright and Burgin (2009) conducted a study observing macroinvertebrate abundance and family richness in relation to two waste discharges, a coal-mine drainage system and CSO drainage in an otherwise non-polluted river (Rover Grose, Southeastern Australia). The study found that while abundance declined downstream from the coal mine drainage, it increased immediately after the CSO. A group of six taxa (Ancylidae, non-Ancylidae gastropods, Nemertea, Simuliidae, Hydroptilidae and Corbiculidae) were particularly abundant and strongly influenced the polluted macroinvertebrate community. This group of biota mutually increased their abundance, in the presence of sewage effluent, more than three times than those found at unpolluted sites within the same river. An alternative study (Xu *et al*. 2014) found particular

macroinvertebrate species to be extensively adaptable to environments with extremely poor water quality, examples include Chironomidae, Lymnaeidae, *Tubifex* sp., and *Limnodrilus* sp.. It is at this point in the current study that sequenced data of macroinvertebrates would have been beneficial, because it may be the case that certain species thrive downstream from sewage effluent, and effects from pollution are somehow preventing Myxozoan infection. Examples of such include negatively affecting the free-living parasite stage (in this case as mature actinospores or bryozoa-spores) or the stages of development within the primary invertebrate host. In this sense, effluent pollution may be benefiting certain macroinvertebrate taxa.

4.4 Myxozoan diversity along a river continuum in the River Erewash

When observing myxosporean and malacosporean community composition within the River Erewash, it is clear that there is an overall reduction in biodiversity between ASVs. There was never any instance of more than two species identified within the samples taken, a stark contrast to other CSO sites which showed up to 8 individual species within a sample. The communities are also moderately distinct between sites, and the spatial and temporal variation is large despite the flow of the river connecting said sites. An interesting alternative approach would be to build a greater knowledge of the hydrology and geography of the Erewash, to give an understanding as to why the communities seem so distinct between species, and why there is no real diversity within sampled sites. A majority of malacosporean detection occurred in samples collected in September (2021). Bryozoan communities mature over the summer months in the UK, and juvenile fish become infected in the autumn (Morris and Adams, 2006). This suggests that the success of identification of these species was due to them having reached their free-living mature bryozoa-spore phase, and therefore were more easily collected in water samples.

The River Erewash is a tributary of the River Trent that flows roughly southwards towards Derbyshire, eventually joining the River Trent via the Attenborough Wildlife Park, Nottinghamshire. The river drains a heavily urbanised catchment and receives effluent from 28 individual combined sewage overflows of varying sizes (Vane *et al*. 2010). Records of some these CSOs date back to the early 1980s, a significant amount of time for an aquatic environment to receive point source pollution at various sites. A majority of other sampled sites only received effluent from a handful of CSOs along the length of the river system where samples were taken. Furthermore, the Erewash is not a particularly large river, suggesting that effluent waste is more likely to be concentrated within the system in contrast to previous rivers sampled such as the Trent or the Dove. The most significant CSO is

located between sampling sites 3 (Gallows) and 4 (Sandiacre), the Hallam Fields STW, which overflowed with untreated effluent for 8,354 hours in 2021, equating to 348 days of the year. A variety of myxosporean species dominated various sites upstream from this STW, as well as the smaller rivers and streams that connect and feed into the Erewash. Examples of such include *Myxobolus diversicapsularus* at the Pinxton site, *Myxobolous gasterostei* at the Gallows site in July (2021) and *Myxobolus alverazae* at the Gallows site the following September. When observing species identified in samples taken downstream from the STW (Goldenbrook, Toton, July and September), they are dominated by just one species, the frequently and largely temporally observed *Myxobolus pseudodispar*. It must be mentioned that *Myxobolus parviformis* was also identified at Toton in July, but at very low reads. These results combined with the fact that the Erewash is more polluted than sampling sites in the previous study, further supports the theory of effluent pollution being a driver for reduction in biodiversity of parasites, and therefore an indicator for compromised ecosystem health, however further analysis of effluent load within the system is required.

All species identified within the Erewash samples (other than one instance of *Myxobolous gasterostei* at the Gallows) were consistent with a preference of infecting fish gills and muscle tissues, and therefore being taxonomically classified as belonging to the *Myxobolus* genus. The overwhelming presence of muscle and gill infecting species and the highly polluted nature of the Erewash further supports the hypothesis that effluent waste is potentially causing anatomical and physiological alterations to fish within the environment, in turn creating more successful infection rates for *Myxobolus* spp..

In terms of malacosporean data, there was only one occurrence of *Buddenbrockia worm* FS2 in samples downstream from the Hallam Fields STW, and no other species were identified in these samples. There were, however, instances of multiple species being identified at various sites upstream from the STW. In three cases, two species were present within the samples. It is once again at this point in the study where the successful sequencing of macroinvertebrate communities would have been hugely beneficial, to observe the relationship between malacosporeans and their bryozoan hosts and how these relationships are shaped within a highly polluted multichannel river system.

Conclusion and future recommendations

In conclusion, the current study has demonstrated that it is possible to detect and map a diversity of myxosporean and malacosporean species through eDNA approaches. Amplicon sequencing using species-specific primers and a nested 2-step PCR approach can be

employed to gain useful insights into Myxozoan communities, and how they may respond to environmental pressures, such as effluent waste, within an ecosystem. This approach is likely to be particularly valuable for future ecological and conservation studies, as the presence of parasitic communities, which can be indicators for ecosystem health, can be identified and taxonomically classified in a much less time consuming and more cost-effective manner than traditional techniques. The methods described in this study could be utilised in a preventative manner, by offering a means for identifying factors that may be detrimental to an ecosystem before they have had significant destructive effects on biodiversity. Approaches such as these are highly significant in the present day, with the ever-increasing levels of anthropogenic activities across the globe, it is crucial that the successful biomonitoring of species and their environments is also increasing synergistically.

Suggestions for future improvement and research areas of the current study include further optimisation of the annealing temperatures and PCR conditions to improve the performance of BF/BR primers. The successful amplification and sequencing of Myxozoan primary hosts, macroinvertebrates, would present a useful secondary perspective to the study, to build a deeper understanding of the interactions and host-parasite relationships that occur within ecologically compromised aquatic ecosystems. Furthermore, the sampling of river sites that were entirely isolated from any effluent waste would be crucial in the next steps, to rule out any confounding factors that may be altering the results between sites upstream and downstream from CSOs currently, such as a swell of effluent that may have altered upstream samples, or effluent waste affecting the entire river system, such as the River Erewash.

Acknowledgements

I would like to take this opportunity to firstly give enormous thanks Hanna Hartikainen for her continued support, extensive knowledge and guidance throughout this research project, to the team of individuals who travelled to various locations in England and waded through rivers to collect samples, to Florian Melzer for his collection and extraction of eDNA samples from the Erewash river, which provided an excellent secondary perspective to the study, to the team at Nottingham DeepSeq facility for the successful sequencing of our amplicons and to Ann Lowe for her unwavering reliable support when conducting laboratory based work. I would also like to thank my friends and family for their endless support and patience during my time of study over the last 12 months.

References

Alama-Bermejo, G., Holzer, A. S., 2021. Advances in discoveries of Myxozoan genomics. *Trends in Parasitology*. Volume 37, Issue 6. [online journal] Pgs 552-568.

Americus, B., Lotan, T., Bartholomew, J. L., Atkinson, S. D., 2020. A comparison of the structure and function of nematocysts in free-living and parasitic cnidarians (Myxozoa). *International Journal for Parasitology*. Volume 50, Issue 10-11. [online journal] Pgs 763-769.

Atkinson, S.D. and Bartholomew, J.L., 2009. Alternate spore stages of Myxobilatus gasterostei, a myxosporean parasite of three-spined sticklebacks (Gasterosteus aculeatus) and oligochaetes (Nais communis). Parasitology Research, 104(5), pp.1173-1181.

Balian, E. V., Segers. H., Lévêque, C., Martens, K., 2008. The Freshwater Animal Diversity Assessment: an overview[online journal] Pgs. 627-637.

Bass, D., Stentiford, G. D., Littlewood, D. T. J., Hartikainen, H., 2015. Diverse applications of environmental DNA methods in parasitology. *Trends in Parasitology*. Volume 31. Issue 10. [online journal] Pgs 499-513.

Bartošová-Sojková, P., Hrabcová, M., Pecková, H., Patra, S., Kodádková, A., Jurajda, P., Tyml, T., Sibylle Holzer, A., 2014. Hidden diversity and evolutionary trends in malacosporean parasites (Cnidaria: Myxozoa) identified using molecular phylogenetics. *International Journal for Parasitology*. Volume 44. Issue 8. [online journal] pgs 565-577.

Beckerman, A. P., Childs, Z. D., Petchey, O. L., 2017. Getting started with R – An introduction for biologists. Second Edition. Oxford University Press.

Bell, K. C., Carlson, C. J., Phillips, A. J., 2018. Parasite collections: Overlooked resources for integrative research and conservation. *Trend in Parasitology*. Volume 34. Issue 8. [online journal] Pgs 637-639.

Bernet, D., Schmidt-Posthaus, H., Wahli, T., Burkhardt-Holm, P., 2004. Evaluation of two monitoring approaches to assess effects of waste water disposal on histological alteration in fish. *Hydrobiologica*. Issue 524. [online journal] Pgs 53-66.

Bhanot, R., Hundal, S. S., 2020. Biochemical and histopathological effects in muscular tissue of carp fish (*Labeo rohita*, Hamilton 1822) following exposure to treated and untreated sewage water. *Research in Environmental Planning and Management*. Volume 28. [online journal] pgs 63991-64013.

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, A. R., Foster, J., Wilkinson, J. W., Arnell, A., Brotherton, P., Williams, P., Dunn, F., 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*. Volume 183. . Elsevier ScienceDirect [online journal]. Pgs 19-28

Bjærke, O., Østbye, K., Lampe, H. M., Vøllestad, L. A., 2010. Covariation in shape and foraging behaviour in lateral plate morphs in three-spined stickleback. *Ecology of Freshwater Fish*. Volume 19. Issue 2. [online journal] pgs 249-246.

Bordes, F., Morand, S., 2009. Parasite diversity: An overlooked metric of parasite pressures? *Oikos*. Volume 118. Number 6. [online journal] Pgs 801-886.

Bruce, K., Blackman, R., Bourlat, S. J., Hellstrom, A. M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., Elbrecht, V., Fazi, S., Fonseca, V., Hanfling, B., Leese, F., Machler, E., Mahon, A. R., Meissner, K., Panksep, K., Pawlowski, J., Yanez, P. S., Seymour, M., Thalinger, B., Valentini, A., Woodcock, P., Traugott, M., Vasselon, V., Deiner, K., 2021. A practical guide to DNA based methods for biodiversity assessment. *Advanced Books.* Pensoft Publishers. Sofia, Bulgaria, 2021. ISBN 978-619-248-053-0 (e-book).

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. and Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), pp.581-583.

Clarke L, Short C and Berry R (2016) Isbourne Catchment Project: Scoping Study (Final Report), Report to Isbourne Catchment Partnership & Environment Agency. School of Natural and Social Sciences & CCRI, University of GloucestershireI: Gloucestershire, UK.

Criscuolo A, Gribaldo S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210.

Dudgeon, D., Arthington, A. H., Gessner, M. O., KAwabata, Z-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A-H., Soto, D., Stiassny, M. L. J., Sullivan, C. A., 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*. Volume 81. Issue 2. Wiley Online Library. [online journal] Pgs 163-182.

Dzidowski, R., Paperna, I., Diamant, A., 2003. Use of fish parasite species richness indices in analyzing anthropogenically impacted coastal marine systems. *Helgoland Marine Research*. Volume 57. [online journal] pgs 220-227.

Elliott, B., Damm, C., Nyland, A., Porr, M., Nilsson Stutz, L., Warren, G., 2022. A Mesolithic settlement on the River Isbourne and Winchcombe, Gloucestershire. *Mesolithic Miscellany*. *Decolonising the Mesolithic?* Volume 29. Issue 2. https://sites.google.com/site/mesolithicmiscellany/journal-volumes/volumes [e-Book] ISSN 0259-3548

Forró, B., Eszterbauer, E., 2016. Correlation between host specificity and genetic diversity for the muscle-dwelling fish parasite *Myxobolus pseudodispar*: examples of Myxozoan host shift? *Folia Parasitologica*. Volume 63. Issue 9. [online journal].

Gomes, G. B., Hutson, K. S., Domingos, J. A., Chung, C., Hayward, S., Miller, T, L., Jerry, D. R., 2017. Use of environmental DNA (eDNA) and water quality data to predict protozoan parasite outbreaks in fish farms. *Aquaculture*. Volume 479. [online journal] Pgs 467-473.

Hallett, S. L., Atkinson, S. D., El-Matbouli, M., 2002. Molecular characterisation of two aurantiactinomyxon (Myxozoa) phenotypes reveals one genotype. *Journal of Fish Diseases.* Volume 25. Pgs 627-631.

Hartikainen, H., Gruhl, A., Okamura, A., 2014. Diversification and repeated morphological transitions in endoparasitic cnidarians (Myxozoa: Malacosporea). *Molecular Phylogenetics and Evolution*. Volume 26. [online article] Pgs 261-269.

Hartikainen, H., Bass, D., Briscoe, A. G., Knipe, H., Green, A. J., Okamura, B., 2016. Assessing myxozoan presence and diversity using eDNA. *International Journal of Parasitology*. Volume 46. Issue 12. ScienceDirect. [online journal]

Holzer, A. S., Bartošová-Sojková, P., Born-Torrijos, A., Lövy. A., Hartigan, A., Fiala, I., 2018. The joint evolution of the Myxozoa and their alternate hosts: A cnidarian recipe for success and vast biodiversity. Volume 27, Issue 7. [online journal] Pgs 1651-1666.

Hudson, P. J., Dobson, A. P., Lafferty, K. D., 2006. Is a healthy ecosystem one that is rich in parasites? *Trends in Ecology & Evolution*. Volume 21. Issue 7. [online journal] Pgs 381-385.

Huver, J., R., Koprivnikar, J., Johnson, P. T. J., Whyard, S., 2015. Development and application of an eDNA to detect and quantify a pathogenic parasite in aquatic ecosystems. *Ecological Applications*. Volume 25. Issue 4. [online journal] Pgs 991-1002.

Jerde, C. L., Chadderton, W. L., Mahon, A. R., Renshaw, M. A., Corush, J., Budny, M. L., Mysorekar, S., Lodge, D. M., 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences*. Volume 70. Issue 4. Canadian Science Publishing. [online journal].

Katoh, K., Rozewicki, J. and Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualisation. Briefings in bioinformatics, 20(4), pp.1160-1166.

Keck, F., Hurlemann, S., Locher, N., Stamm, C., Deiner, K., Altermatt, F., 2022. A triad of kick net sampling, eDNA metabarcoding and predictive modelling to assess aquatic macroinvertebrate biodiversity. *bioRxiv*. [reprint].

Lafferty, K. D., Dobson, A. P., Kuris, A. M., 2006. Parasites dominate food web links. *PNAS*. Issue 103. Volume 30. [online journal]. Pgs 11211-11216

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., Zizka, P. M. A., 2020. Improved freshwater macroinvertebrate detection from environmental DNA through minimized non-target amplification. Special Issue Original Article. [online journal].

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S. and Gascuel, O., 2019. NGPhylogeny. fr: new generation phylogenetic services for non-specialists. Nucleic Acids Research, 47(W1), Pgs. 260-265.

Liu, Y., Lövy, A., Gu, Z., Fiala, I., 2019. Phylogeny of Myxobolidae (Myxozoa) and the evolution of myxospore appendages in the *Myxobolus* clade. *International Journal of Parasitology.* Volume 49. Issue 7. June 2019. ScienceDirect [online journal]. Pgs 523-530.

Lisnerová, M., Blabolil, P., Holzer, A., Jurajda, P., Fiala, I., 2020. *Folia Parasitologica*. Volume 76, Issue 9. doi: 10.14411/fp.2020.019. [online journal]

Lisnerová, M., Fiala, I., Cantatore, D., Irigoitia, M., Timi, J., Pecková, H., Bartošová-Sojková, P., Sandoval, C. M., Luer, C., Morris, J., Holzer, A., 2019. Mechanisms and drivers for the establishment of life-cycle complexity in Myxozoan parasites. *Biology*. Volume 9. Issue 1. [online journal].

Lom, J., Dyková, I., 2006. Myxozoan genera: definition and notes on taxonomy, life-cycle terminology and pathogenic species. *Folia Parasitologica*. Volume 53. Issue 1. [online journal] pgs 1-36.

Maftuch, M., Sanoesi, E., Farichin, I., Saputra, B. A., Ramdhani, L., Hidayati, S., Fitriyah, N., Prihanto, A. A., 2018. Histopathology of gill, muscle, intestine, kidney and liver on *Myxobolus* sp.-infected Koi carp (*Cyprinus carpio*). *Journal of Parasitic Diseases*. Volume 52. [online journal] pgs 137-143.

Mahé, F., Rognes, T., Quince, C., De Vargas, C. and Dunthorn, M., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ, 3, p.e1420.

Martin, D. C., Neely, R. K., 2001. Benthic macroinvertebrate response to sedimentation in a *Typha angustifolia* L. wetland. *Wetlands Ecology and Management*. [online journal] pgs 441-454.

Molnar, K., Eszterbauer E., Szekely, C., Dan, A., Harrach, B., 2002. Morphological and molecular biological studies in intramuscular *Myxobolus* spp. of cyprinid fish. *Journal of Fish Diseases.* Volume 25. Pgs 643-652.

Morley, N. J., Lewis, J. W., 2006. Anthropogenic pressure on a molluscan-trematode community over a long-term period in the Basingstoke canal, UK, and its implications for ecosystem health. *EcoHealth*. Volume 3. [online journal] pgs 269-280.

Morris, D. J., Adams, A., 2006. Transmission of *Tetracapsuoides bryosalmonae* (Myxozoa: Malacosporea), the causative organisms of proliferative kidney disease, to the freshwater bryozoan *Fredericella sultana*. *Parasitology*. Volume 133. Issue 6. [online journal]

Madanire-Moyo, G., Barson, M., 2010. Diversity of metazoan parasites of the African catfish *Clarias gariepinus* (Burchell, 1822) as indicators for pollution in a subtropical African river system. *Journal of Helminthology*. Volume 84. Cambridge University Press. Pgs 216-227.

Neal, C., Neal, M., Hill, L., Wickham, H., 2006. River water quality of the River Cherwell: An agricultural clay-dominated catchment in the upper Thames Basin, southeastern England. *Science of the Total Environment*. Volume 360. Issue 1-3. [online journal] pgs 272-287.

Okamura, B., Anderson, C. L., Longshaw, M., Feist, S. W., Canning, E. U., 2001. Patterns of occurrence and 18s rDNA sequence variation of PKX (Tetracapsuloides bryosalmonae), the causative agent of salmonid proliferative kidney disease. *J Parasitol*. Volume 87. Issue 2. [online journal] pgs 379-385.

Patra, S., Hartigan, A., Morris, D. J., Kodadkova, Holzer, A. S., 2017. Description and experimental transmission of Tetracapsuloides vermiformis n. sp. (Cnidaria: Myxozoa) and guidelines for describing malacosporean species including reinstatement of Buddenbrockia bryozoides n. comb. (syn. Tetracapsula bryozoides). Parasitology. Vol 144. Issue 4. Cambridge University. [online journal] pgs 497-511.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., Gough, K. C., 2014. REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a service tool in ecology. *Journal of Applied Ecology*. Volume 51. Issue 5. British Ecological Society. [online journal]. Pgs 1450-1459.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., Kidd, K. A., MacCormack, T. J., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D., Cooke, S. J., 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*. Volume 94. Issue 3. [online journa] Pgs 849-873.

Reynolds, . S., Maberly, S. C., Parker, J. E., De Ville, M. M., 2012. Forty years of monitoring water quality in Grasmere (English Lake District): separating the effects of enrichment by treated sewage and hydraulic on phytoplankton ecology. *Freshwater Biology*. Issue 57. [online journal] Pgs 384-399.

Rocha, S., Alves, A., Antunes, C., Azevedo, C., Casal, G., 2019. Molecular data infers the involvement of a marine aurantiactinomyxon in the life cycle of the myxosporean parasite *Paramyxidium giardi* (Cnidaria, Myxozoa). *Parasitology*. Volume 146, issue 12. Pgs 1555 -1563. [online journal]

Rodriguez, P., Martinez-Madrid, M., Angel Arrate, J., Navarro, E., 2001. Selective feeding by the aquatic oligochaete Tubifex tubifex (Tubificidae, Clitellata). *Hydrobiologia*. Volume 463. [online journal] pgs 133-140.

Ronquist, F. and Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12), pp.1572-1574.

Vane, C. H., Kim, A. W., McGowan, S., Leng, M. J., Heaton, T. H. E., Kendrick, C. P., Coombs, P., Yang, H., Swann, G. E. A., 2010. Sedimentary records of sewage pollution using faecal markers in contrasting peri-urban shallow lakes. *Science of the Total Environment*. Volume 409. Issue 2. Pp. 345-356.

Walker, J. R., Hassall, C., 2021. The effects of water chemistry and lock-mediated connectivity on macroinvertebrate diversity and community structure in a canal in northern England. *Urban Ecosystems*. Issue 24. [online journal] Pgs 491-500

Walsh, C. L., Kilsby, C. G., 2007. Implications of climate change on flow regime affecting Atlantic Salmon. *Hydrology & Earth System Sciences*. Volume 11. Issue 3. [online journal] pgs 1127-1143

Wright, E.S., 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R. R J., 8(1), p.352.

Wright, I. A., Burgin,S., 2009. Comparison of sewage and coal-mine wastes on stream macroinvertebrates within an otherwise clean upland catchment, Southeastern Australia. *Water, Air and Soil Pollution*. Issue 204. [online journal]

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. Volume 13. Issue 134. [online journal]

Yokoyama, H., Grabner, D., Shirakashi, S., 2012. Transmission biology of the Myxozoa. *Health and Environment in Aquaculture*. Edited by Carvalho, E. D., David, G. S., Jose de Silva, R.. Published by InTech 2012. Rijeka, Croatia. ISBN 978-95351-0479-1. Pgs 3-29. [e-Book].

Xu, M., Wang, Z., Duan, X., Pan, B., 2014. Effects of pollution on macroinvertebrates and water quality bio-assessment. *Hydrobiologia*. Issue 729. [online journal] pgs 247-259.