Manuscript Number: IJPara15_418R2

Title: Assessing myxozoan presence and diversity with environmental DNA

Article Type: Full Length Article

Keywords: eDNA, Myxozoa, lineage-specific primers, custom barcoding, faecal DNA, index misassignment, Myxosporea, piscivore faeces

Corresponding Author: Dr. Hanna Hartikainen,

Corresponding Author's Institution: Eawag

First Author: Hanna Hartikainen

Order of Authors: Hanna Hartikainen; David Bass; Andrew G Briscoe, PhD; Hazel Knipe; Andy J Green, PhD; Beth Okamura

Manuscript Region of Origin: UNITED KINGDOM

Abstract: Amplicon sequencing on a High Throughput Sequencing (HTS) platform (custom barcoding) was used to detect and characterise myxosporean communities in environmental DNA samples from marine and freshwater environments and in faeces of animals that may serve as hosts or whose prey may host myxosporean infections. A diversity of myxozoans in filtered water samples and in faeces of piscivores (otters and great cormorants) was detected, demonstrating the suitability of lineage specific amplicons for characterising otherwise difficult to sample parasite communities. The importance of using the approach was highlighted by the lack of myxosporean detection using commonly employed, broadly-targeted eukaryote primers. These results suggest that, despite being frequently present in eDNA samples, myxozoans have been generally overlooked in 'eukaryote-wide' surveys. Lineage-specific primers in contrast detected 107 OTUs that were assigned to both the "freshwater" and "marine" myxosporean lineages. Only 7% of these OTUs clustered with sequences in GenBank, providing evidence for substantial undescribed myxosporean diversity. Many new OTUs, including those found in otter faeces, clustered with a clade of myxosporeans previously characterized by sequences from invertebrate hosts and water samples only. Because myxozoan species identification is highly reliant on molecular signatures, lineage-specific amplicon sequencing offers an effective and non-destructive means of improving our knowledge of myxozoan diversity. In addition, the analysis of myxozoan DNA in faeces of piscivores offers a potentially efficient method of sampling for diversity and revealing life cycles as piscivore activities may integrate myxozoan infections in fish over relatively broad spatial scales.

Highlights

- eDNA reveals diversity across the myxosporean phylogeny in water and faecal samples
- A custom barcoding approach amplifies myxosporean SSU rDNA that is missed in ‘eukaryote-wide’ surveys
- Piscivore activities may be useful in integrating myxozoan infections in prey over broad spatial scales
- Recommendations for assessing the presence of myxozoans in eDNA are provided

Assessing myxozoan presence and diversity with environmental DNA

Hanna Hartikainen^{1,2,3*}, David Bass^{3,4}, Andrew G. Briscoe³, Hazel Knipe^{3,5}, Andy J. Green⁶, Beth
5 Okamura³

¹ Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

² Institute for Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

³ Department of Life Sciences, The Natural History Museum, Cromwell Road, London, SW7 5BD,
10 UK

⁴ Centre for Environment, Fisheries and Aquaculture Science (Cefas), Barrack Road, The Nothe,
Weymouth, Dorset, DT4 8UB, UK

⁵ Cardiff School of Biosciences, Sir Martin Evans Building, Museum Place, Cardiff, CF10 3AX,
UK

15 ⁶Department of Wetland Ecology, Estación Biológica de Doñana, EBD-CSIC, Américo Vespucio
s/n, 41092 Sevilla, Spain

*Corresponding author: Hanna Hartikainen; Eawag, Ueberlandstrasse 133, Dübendorf,
Switzerland; phone: +41 58 765 5446; Hanna.Hartikainen@eawag.ch

20 Note: Supplementary data associated with this article

Abstract

Amplicon sequencing on a High Throughput Sequencing (HTS) platform (custom barcoding) was used to detect and characterise myxosporean communities in environmental DNA samples from marine and freshwater environments and in faeces of animals that may serve as hosts or whose prey may host myxosporean infections. A diversity of myxozoans in filtered water samples and in faeces of piscivores (otters and great cormorants) was detected, demonstrating the suitability of lineage specific amplicons for characterising otherwise difficult to sample parasite communities. The importance of using the approach was highlighted by the lack of myxosporean detection using commonly employed, broadly-targeted eukaryote primers. These results suggest that, despite being frequently present in eDNA samples, myxozoans have been generally overlooked in ‘eukaryote-wide’ surveys. Lineage-specific primers in contrast detected 107 OTUs that were assigned to both the “freshwater” and “marine” myxosporean lineages. Only 7% of these OTUs clustered with sequences in GenBank, providing evidence for substantial undescribed myxosporean diversity. Many new OTUs, including those found in otter faeces, clustered with a clade of myxosporeans previously characterized by sequences from invertebrate hosts and water samples only. Because myxozoan species identification is highly reliant on molecular signatures, lineage-specific amplicon sequencing offers an effective and non-destructive means of improving our knowledge of myxozoan diversity. In addition, the analysis of myxozoan DNA in faeces of piscivores offers a potentially efficient method of sampling for diversity and revealing life cycles as piscivore activities may integrate myxozoan infections in fish over relatively broad spatial scales.

Keywords: eDNA, Myxozoa, lineage-specific primers, custom barcoding, faecal DNA, index misassignment, Myxosporea, piscivore faeces

1. Introduction

Myxozoans have recently been shown to represent a spectacular radiation of endoparasitic cnidarians (Jimenez-Guri et al., 2007; Holland et al., 2010; Nesnidal et al., 2013) with complex life cycles that involve vertebrates and invertebrates as their intermediate and primary hosts, respectively (Canning and Okamura, 2004). There are two major clades, the early diverging and species-poor Malacosporea and the highly derived and speciose Myxosporea (Fiala et al. 2015). With >2000 described species (Lom and Dykova, 2006) myxozoans currently contribute some 18% to cnidarian species diversity as it is known today (Okamura et al., 2015). The majority of vertebrate hosts are freshwater and marine fish, and transmission usually takes place via water borne spores. Only a small fraction (~2%) of myxozoan life-cycles has, however, been resolved or inferred (Atkinson and Bartholomew, 2009; Eszberbauer et al., 2015). Especially enigmatic are the life cycles of myxozoans that infect shrews, amphibians and wildfowl (Bartholomew et al., 2008; Hallett et al., 2015). Although some myxozoans cause clinical diseases that impact fisheries and aquaculture (e.g. salmonid whirling disease, proliferative kidney disease, ceratomyxosis), many infections are inapparent and have little impact on fish hosts (Lom and Dyková, 1992).

It is broadly accepted that myxozoan species diversity is at present underestimated (Lom and Dykova, 2006). Low infection prevalences, low infection intensities and patchy distributions are all likely to contribute. However, the asymptomatic nature of many myxozoan infections is probably especially important in contributing to the substantial but hidden diversity of myxozoans. In addition, growing evidence for cryptic species complexes (Whipps and Kent, 2006; Holzer et al., 2013; Hartikainen et al., 2014c) suggests further challenges for understanding species diversity. Traditional approaches of detecting the presence and diversity of myxozoans entail dissection of individual hosts followed by histological examination of an array of material to identify associated parasites. However logistical difficulties and moral issues may arise if large numbers of potential hosts require sacrificing when prevalences are low or distributions patchy. Other approaches that avoid host sacrifice may therefore be desirable.

Further challenges to revealing myxozoan diversity are presented by the general lack of informative features and by morphological convergence, both being related to the extreme morphological simplification that characterises myxozoans (Fiala et al., 2015). Sequence data are therefore crucial
75 for identification and are widely included in taxonomic studies to ensure species discrimination (Atkinson et al., 2015). The increasing availability of myxozoan Small Subunit rRNA gene sequences (SSU rDNA) can therefore, to an extent, serve as a measure of known diversity and a framework within which novel sequence types can be placed in the absence of morphological data.

Environmental sampling now provides an alternative to traditional methods for detecting parasites
80 and characterising their diversity (Bass et al., 2015). For example, DNA extraction of total DNA from environmental samples (eDNA) has been increasingly used to sequence the microbial biosphere (microscopic organisms present in the environmental sample) and to detect larger organisms via DNA shed into the environment with e.g. skin or mucous cells (Bråte et al., 2010; Thomsen et al., 2012; Mächler et al., 2014). Indeed, because many parasites specifically shed
85 transmission stages into the environment their detection in water samples may be particularly facilitated. In accord, specific PCR-based methods conducted on water samples are commonly used to indicate disease risk (Hallett and Bartholomew, 2006; Hung and Remais, 2008; Strand et al., 2011; Strand et al., 2014). Analysing environmental samples to detect myxozoans and to characterise diversity could therefore be highly complementary to identifying infections in
90 potentially many and diverse hosts. This approach, however, has remained unexplored for myxozoans.

Here we demonstrate how lineage-specific amplicon sequencing using myxozoan targeted PCR is suitable for detecting and characterising myxosporeans in environmental samples. We apply such a “custom barcoding” approach to a wide range of eDNA samples, test its replicability and compare
95 the detection of myxosporeans using specific primers versus eukaryote-wide primers (for the V4 region of SSU rRNA gene). We also examine whether myxosporeans can be detected in faeces of piscivores whose wide-ranging foraging may provide an integrated sample of myxozoans present

within systems. We reveal extensive diversity across the myxosporean phylogenetic spectrum including in clades with poor taxon sampling and little understanding of host usage and life-cycle details. We provide practical guidance for using lineage-specific amplicons in surveys of myxosporeans and consider how these may provide insights on life cycles, parasite communities, infection risks and ultimately aid in the monitoring and management of parasitic diseases.

2. Methods

2.1 Samples for eDNA analysis

Environmental samples were collected in July 2013 and September 2013 from three freshwater localities in SE England, comprising a small pond (< 20m; henceforth referred to as ‘Avon pool’) laterally situated and contiguous to the River Avon (Hampshire) (near Downton, 51°00′42″N, 1°44′44″W) through which river water slowly circulates; a meandering outflow from a fish farm that receives water further downstream on the River Avon (at Bickton, 50°54′49″N, 1°47′34″W) and; a humic, shallow lake (California Lake, 51°22′44″N, 0°52′22″W) within a country park in Berkshire. Up to 100L of water was passed serially through 55µm and 20µm meshes, and the accumulated material (filtrand) was collected and placed at 4°C. 25L of the filtered water was transported to the lab and filtered serially onto 3µm and 0.45µm filters, with filtrand collected and stored again at 4°C. Filtrands were subsequently dried in a freeze-drier at -56°C for approx. 12h to remove residual moisture and DNA was extracted using the MoBio UltraSoil kit (MoBio Laboratories, Carlsbad, CA, USA) (Hartikainen et al. 2014a, as described for the WEY sample set). The soil kit was chosen as it incorporates steps that reduce the presence of PCR inhibitors in the final DNA extracts. Additional samples were collected in July after disturbing the sediment (by kicking) in two sites (Downton and California Lake) and collecting the filtrand retained on the 55µm mesh from total of 25L.

Faecal samples were obtained from various sites in the UK and Spain from piscivores (otter spraints and great cormorant faeces), wildfowl, and birds that consume terrestrial invertebrates (including earthworms) (Table 1). We also collected earthworm casts. The rationale for examining this material is as follows. Otters and cormorants may consume a variety of fish with myxozoan infections and their faeces may present a means of integrative sampling for myxozoans present in water bodies. Wildfowl have been shown to host myxosporeans (Bartholomew et al., 2008). Myxosporean infections in earthworm-eating shrews suggest that earthworms may act as hosts (Prunescu et al., 2007). Spores of some myxosporeans mature in the intestinal epithelium and are excreted by aquatic oligochaetes (Alexander et al. 2015). Although casts are not technically faeces, spores may nevertheless be present. Henceforth we refer to all material (both casts and true faeces) as faeces. Faecal DNA was extracted using the MoBio Faecal DNA kit (MoBio Laboratories, Carlsbad, CA, USA) from subsamples of 1g or the whole sample if smaller.

Further eDNA samples collected for other studies were included (as listed in Supplementary Table S1). These samples represented a total of 386 DNA extractions of filtrands collected variously as follows (1) South Africa (SA): 80 water samples (~2-10L) collected in December 2011 from freshwater and marine sites (Hartikainen et al., 2014a), (2) Weymouth, southwest England (WEY ENV): 20 DNA extractions incorporating material from six 150L marine water samples filtered serially through >55 μ m, 20 μ m and 0.45 μ m (Hartikainen et al., 2014a), (3) Tamar Estuary, southwest England (TAM): 38 DNA extractions from three sites on the Tamar Estuary (Neals Point, Wilcove and Cremyll) derived from 60L of marine water filtered serially as above, (4) southern England and Portugal (EXE): 79 DNA extractions from freshwater and marine water samples (K Hamilton and B Williams, personal communication), (5) southeast England, River Colne estuary (EST): DNA from 32 marine-to-brackish sediment samples (~1–2 g) (Dong et al., 2009; Hawkins and Purdy, 2007), (6) River Lambourn at Boxford, Berkshire, UK: biofilm samples from experimental tiles) (BIOF) (K Lehman and A Singer, personal communication), (7) tissues and incubation water from 242 invertebrates (see samples listed in Hartikainen et al., 2014b), comprising: aeolid nudibranchs

(n=2), asteroid sp. (n=1), *Cancer pagurus* (n=7), *Carcinus maenas* (n=5), *Cerastoderma edulis* (n=6), cheilostome bryozoan (n=2), *Diogenes pugilator* (n=1), *Gibbula* sp. (n=32), *Littorina* sp. (n=33), mixed copepods (n=18), *Monodonta lineata* (n=15), *Nereis* sp. (n=12), ophiuroid sp. (n=2), *Palaemon* sp. (n=4), *Palaemonetes* sp. (n=10), *Peachia* sp. (n=1), spirorbid sp. (n=2), intertidal amphipod (n=8), and 90 filtered water samples from incubations with *Carcinus maenas*, *Cancer pagurus*, *Gibbula* sp. and *Littorina* sp..

2.1 Myxosporean lineage-specific amplicons

Primers targeting a variable section of the SSU rRNA gene were designed to detect both marine and freshwater myxozoans in the large and derived clade Myxosporea. A nested protocol was chosen to increase both detection efficiency and primer specificity, with the inner primer pair producing an amplicon of 450-490bp, depending on the species (Table 2). Although the primers were designed to be as inclusive within Myxosporea as possible, a number of primer mismatches occurred for sphaerosporids, indicating that amplification in this group may be less efficient.

116 marine water samples, 270 freshwater samples, 20 faecal samples and 242 invertebrate tissue and incubation water samples were tested using the nested PCR protocol (results of each PCR shown in Supplementary Table S2). Positive samples were then pooled according to environment type giving 12 libraries in total (summarised in Fig. 1 and Table 3). The pooled PCRs were cleaned using a QIAquick PCR purification kit (Qiagen, Hilden, Germany) following manufacturer's recommendations. For each pool an indexed library was constructed using a TruSeq Nano DNA sample preparation kit and subsequently sequenced on 1/33 of an Illumina MiSeq flowcell; v3 600 cycle, 2x300 bp paired-end. The cormorant library was run separately from others. To assess the replicability of subsampling from larger faecal samples, two subsamples from the same otter spraint were extracted and processed independently for sequencing (resulting in sequence libraries Otter1A and Otter1B). Further information for water sample volumes (ranging from 2L to >100L), sample

types (e.g. freshwater, marine, sediment, etc.) and localities are provided in Supplementary Table S1.

175 Resulting amplicons were quality controlled using prinseq v.0.20.4 and the paired ends were joined
using Pear 0.8 (Zhang et al., 2014). Reads were clustered into OTUs using the UPARSE pipeline
(Edgar, 2013). Preliminary annotations were conducted in QIIME 1.8.0 (Caporaso et al., 2010) with
reference to the PR2 database (Guillou et al., 2012) using a comparison of blast and uclust methods.
The myxozoan sequences in Supplementary Table S2 were added onto the PR2 database before
180 annotations. For phylogenetic analyses, OTUs were blasted against GenBank and it was ensured
that closest matches to each were present in the reference dataset before tree building. The reference
alignment created using MAFFT with L-INS-I algorithm with default parameters. Myxosporean
OTUs were then placed in the reference alignment using MAFFT v.7.1.2.3b localpair algorithm
with default settings and –add flag activated. All alignments were inspected and corrected manually
185 in BioEdit v.7.2.5 and doubtfully aligned positions were masked. Maximum likelihood trees were
built using RAxML (Stamatakis, 2006; Stamatakis et al., 2008) using the GTR+ Γ model with 100
bootstrap replicates to assess clade support. Lineages well supported in the ML topology but not
containing any OTUs from our study were trimmed to few representative reference sequences to
reduce alignment size. Lineages containing OTUs generated in this study were retained in their
190 entirety and ML analyses repeated with the trimmed alignment.

2.2 Detection of myxozoans with ‘general’ eukaryote primers

Existing and new amplicon data were used to test the suitability of general barcodes for assessing
myxosporean diversity. The SSU rDNA amplicon data obtained from various marine environments
using general eukaryote wide primers (targeting the V4 and V9 regions of SSU rDNA) as part of
195 the BioMarks project (Logares et al., 2014; Massana et al., 2015) (<http://www.biomarks.eu>) were
checked for myxozoan reads. The primers used to generate the BioMarks data have no mismatches
to myxozoan (including both myxosporeans and malacosporans) in the 3’ region of the primers

and a maximum of two mismatches in the full primer sequences. Theoretically, these primers should therefore have amplified any myxozoan sequence types present in the environmental samples. Local blastn searches with 162 myxozoan SSU rDNA sequences as query words (accession numbers provided in Supplementary Table S2) were used to interrogate the complete BioMarks data, which comprised millions of amplicon reads from coastal marine water column and sediment samples from around the European coast from Norway south to Spain and east to the Black Sea. The 10 best hits for each query were re-blasted against the non-redundant nucleotide database (<http://www.ncbi.us.gov>) and screened for matches to myxozoans.

New V4 SSU rDNA amplicon data was generated with the BioMarks primers and PCR protocol, with one additional degeneracy in a single nucleotide of the reverse primer (5'-CCAGCASCYGC GGTAATWCC-3', amplicon size approx. 450bp) to assess whether eDNA specifically enriched in myxozoans (otter faeces) would enable myxozoan detection by "general" primers. Other faecal samples were also sequenced using the "general" primers, to assess if further myxozoans could be detected (wildfowl and garden bird faeces, earthworm casts) (Table 1, Fig. 1). Further sequencing library preparation details are reported in Supplementary Data S1. OTU calling and annotations were conducted as described above for the lineage-specific amplicons of myxosporeans.

2.3 Data accessibility

All paired-end reads generated with the myxosporean lineage-specific primers are available in the Short Read Archive under the accession number SRP077990. The OTU table and representative sequences are available from Dryad data repository <http://dx.doi.org/10.5061/dryad.3f0d0>.

3. Results

3.1 Myxozoan detection using general eukaryote primers

No sequences of myxozoan origin were recovered by blastn searches across the V4 and V9 SSU rDNA eukaryote-wide amplicons in the BioMarKs database. In addition, no myxozoans were found in the MiSeq libraries generated in this study from faecal samples using the same broadly-targeted V4 primers (Table 1) (see Supplementary Data S1 for further information on library preparation).

225 3.2 Invertebrate samples

PCR products of correct size were produced in amphipods (n = 9) but not in any other invertebrates tested (16 taxa tested and a total of 242 samples, see Hartikainen et al., 2014b for details). Direct sequencing confirmed that the PCR products were amphipod SSU rDNA and represented a case where amplification was non-specific.

230 3.3 Myxosporean communities retrieved using lineage-specific amplicons

The myxosporean specific assay produced single bands on electrophoretic gels and those in the size range of 450 bp were found to consistently contain sequences of myxozoan origin (pilot cloning study, data not shown).

In total 5,294,827 paired end reads were generated with the myxosporean assay, of which 80%
235 remained after quality control and paired end joining. 96% of the cleaned reads (4,054,823) clustered into 112 OTUs, verified to be myxozoans via blast, uclust and phylogenetic analyses. Myxosporean OTUs with very few reads were not considered further (<10 reads across all libraries, this resulting in the removal of five OTUs). To guard against false positive detections due to leakage between libraries (see Discussion), occurrences that accounted for <0.1% of total reads per
240 OTU, were not considered. The remaining 4% of the reads were not of myxozoan origin and comprised OTUs with few reads (n=1258 OTUs, the most common sequences were from chlorophytes (*Dunaniella*), Centroheliozoa, kinetoplastids and peritrichs, data not shown). The libraries were normalized to the smallest library size (resulting in 175,508 reads per library). Coverage across the libraries was high and saturation in species accumulation curves was reached in
245 the most diverse libraries at around 50,000 reads (Supplementary Fig. S2). Species accumulation

curves using a) raw and b) normalised data from myxosporean custom barcoding.). All the major marine clades were retrieved with relationships agreeing with previous work (Fig. 3, Fig. 4A) (Fiala, 2006, Bartošová et al., 2009, Bartošová et al., 2010). In our analysis the “Freshwater and Marine Gall Bladder Clade” was not well resolved, similarly to other recent analyses with varying
250 taxon sampling (Rangel et al., in press). The molecular taxonomy of the clade in general is problematic as it comprises several cryptic species complexes (Alama-Bermejo et al., 2016). Nevertheless, this clade is shown in Fig. 4B as several novel OTUs were found grouping within the clade. For example, an OTU corresponding to *Ortholinea orientalis* (OTU 103, 100% similarity across the amplicon with the reference sequence used in the phylogenetic analysis) was found in
255 high abundance in the marine library (UK) (Fig. 4B). The speciose *Myxobolus* clade was further divided into subclades *Myxobolus* 1-5 to provide more resolution when plotting OTU occurrence. The assignment of these clades was for presentation purposes and was based on the inspection of the tree estimate and not on morphological or host-related traits. All of the *Myxobolus* groups 1-5 were well supported (>80% bootstrap support), apart from one exception. OTU 77, representing
260 *Henneguya nuesslini*, was found in high abundance in the freshwater library from South Africa. The placement of this species was not well supported within any of the *Myxobolus* clades defined here, but for presentation purposes it was grouped within *Myxobolus* 3 clade (Fig. 4C). In eight cases OTUs showed >97% similarity to a known reference sequence during the clustering step of the UPARSE algorithm. These OTUs are shown in the tree and annotated by the OTU number as well
265 as the reference sequence match. Of species identified in this manner, an OTU corresponding to *Myxobilatus gasterostei* (OTU 3) was particularly frequently found in the freshwater and otter faecal samples.

3.3.1 Faecal samples

Amplification from otter and cormorant faeces was strong and highly specific. The cormorant faecal
270 sample contained eight myxosporean OTUs, seven of which were unique to this sample (Fig. 2F). The most abundant OTU in the cormorant faeces (OTU56, 92% of all reads) was also detected in

low abundance in one of the environmental water filtrates (California Lake, October sample, n=39 reads). All of the OTUs in the cormorant faeces were placed within *Myxobolus* Clade 4 that includes *Myxobolus muelleri* (AY129314) and *Endocapsa rosulata* (AF306791), amongst others.

275 The clade also contained many new environmentally derived OTUs from other libraries, with poorly resolved positions (Fig. 4B).

Two of three otter spraints were positive for myxosporeans, with OTU richness being highest in otter spraint 2 (Table 3). This was largely reflected in the presence of myxobolids that were conspicuously rare or absent (e.g. *Myxobolus* Clades 3 and 5) in the replicates of otter spraint 1

280 (Fig. 2D-E). Over both spraints, 21 OTUs clustered in the “Environmental clade”, making it the most OTU rich lineage, even when the myxobolids were considered as a single clade (n=5, sum of representatives of *Myxobolus* Clades 1 -5). When measured in read abundance, the top three most abundant OTUs across all otter samples were OTUs 61, 55 and 85. The OTUs represented three major clades - the Urinary Bladder Clade, the Environmental Clade and the *Myxidium* Clade,

285 respectively (Fig. 2A-C).

The two replicate DNA extractions from a single otter spraint showed good concordance (Fig. 2). Across all 23 OTUs in the two replicates, two OTUs were missing from replicate B (OTU68, n=111 reads and OTU65, n=90 reads). Nine of the most abundant OTUs across the replicate otter libraries were the same (Fig. 2). The rank order of abundance varied in the two replicates, OTU85 was most

290 abundant in sample A, and OTU61 was most abundant in sample B (Fig. 2A-C).

3.3.2 Water samples

Myxosporeans frequently amplified from the eDNA in the three freshwater sites sampled in two seasons (Avon pool, Bickton and California Lake, in July and September, Supplementary Table S1). The majority of positive samples came from the 20-55µm and >55µm size fractions, with

295 lower amplification success in samples from small size fractions (3- 0.45µm and 3-20µm) (Supplementary Table S1). As samples were pooled for sequencing, it was not possible to assess

whether different myxosporean communities were present in the different size fractions or in the samples augmented by kicking the sediment prior to sample collection. However, it was observed during gel electrophoresis that samples incorporating sediment produced visibly stronger amplicon
300 bands than those without such treatment during sample collection.

The myxosporean communities in the three sites showed distinct spatial patterns with the vast majority of OTUs being unique to sites (Supplementary Fig. S3). Temporal variation was also detected in the three UK sites. Samples collected in July from two connected sites (Avon pool and Bickton) contained a high proportion of OTUs clustering in the Chloromyxid Freshwater Gall
305 Bladder Clade (Fig. 3A). Representatives of the Freshwater Gall Bladder Clade were largely absent in September samples from these sites and from California Lake. The latter were dominated by myxobolids that clustered in *Myxobolus* Clade 4 (the differences in OTU presence in the different phylogenetic groupings is further shown in Supplementary Fig. S2). In general, OTU richness was low in the California Lake samples (11 and 3 OTUs observed in July and October, respectively),
310 when compared to the river samples (Avon pool: 12 and 19 OTUs observed in July and October, respectively, Bickton: 28 and 22 OTUs observed in July and October, respectively). The number of OTUs clustering in the “Environmental Clade” was remarkably high (n=32 from all libraries, 20 of which were present in the California Lake, Bickton and Avon pool samples). 31 of these OTUs were novel and one OTU (OTU1) matched a previously characterised sequence of
315 *Aurantiactinomyxon* sp. from *Tubifex ignotus* (AF483598, under the criterion of >97% similarity during clustering with UPARSE, sequences were identical apart from a single A/T SNP site). In comparison, the much larger *Myxobolus* Clade contained 43 OTUs, two of which matched previously known sequences, namely *Thelohanellus hovorkai* (DQ231155, OTU4) and *Myxobolus diversicapsularis* (GU968199, OTU102).

320 The myxosporean communities in the two pooled samples (Library 11 from ponds in South Africa and Library 12 from UK marine locations) were diverse, containing 30 and 33 OTUs, respectively. All OTUs in the marine library were placed within known marine lineages in the reference

phylogenetic tree (Fig. 3). Despite pooling PCRs from several sites, only seven OTUs were detected in the South African sequence library and all were unique to the South African library. These were placed variously in the *Myxobolus* Clade and the “Urinary Bladder Clade”. One OTU was identified with >97% similarity to a previously characterized sequence, OTU4 with *Thelohanellus hovorkai* (DQ231155). The pooled library from UK marine environments (MAR_UK), in comparison, was relatively diverse, containing 18 OTUs, all unique to this library. An OTU (OTU5) clustering with *Myxidium gadi* (DQ377707) was identified as a match to a previously characterized sequence; others were novel using our criterion of similarity.

4. Discussion

The methods developed here proved well suited for both conventional cloning-based and HTS approaches to assess myxosporean presence and diversity in various sample types. The lineage-specific myxosporean assay detected 107 myxosporean OTUs associated with a diversity of aquatic environments. Phylogenetic placement of these OTUs indicates that a wide taxonomic range of myxosporeans was successfully amplified with the same primer set, with putative taxa detected in both the freshwater and marine myxosporean lineages (Fiala et al., 2015). New OTUs (i.e. those that did not cluster at >97% similarity with reference sequences obtained from GenBank) were detected in all sequence libraries, with only 7% of OTUs matching previously known sequences. These results suggest that there is much myxosporean diversity still to be discovered and described even in regions of the world where the myxozoan fauna is relatively well known. Especially striking was the large number of OTUs found to cluster within a clade defined by sequences derived from water samples or invertebrate hosts (mainly Oligochaeta) (the “Environmental Clade”). No fish hosts are known for the species in this clade, but our detection of sequences (OTUs) in otter faeces in large abundances suggest that members of this clade infect fish that are commonly consumed by piscivores. Potential explanations for lack of detection in fish so far include cryptic infections (although the abundance in otter faeces argues against this) or an unusual site of infection in the fish hosts.

The greater number of taxa sampled in the freshwater relative to the marine lineage likely reflects
350 bias associated with processing a larger number of freshwater samples and sample types. However,
the pooled marine library results confirm that a relatively high diversity of myxosporeans can be
retrieved across the known marine clades. The OTU richness in the pooled sample was relatively
high, with 18 unique OTUs. It is possible that greater dilution in marine environments may reduce
the likelihood of sampling spores and that sample volumes required are larger than in the freshwater
355 environments (e.g. we used 60L samples in the Tamar Estuary and 150L samples from Weymouth
shore, Supplementary Table S1).

The detection of myxosporeans in otter and cormorant faeces likely reflects infections of fish prey
and provides evidence that sampling parasites in faeces of piscivores offers a non-destructive and
potentially less time-consuming method to gain an integrative view of myxosporeans present in
360 water bodies. Molecular detection of both myxosporeans and associated prey items in faeces
additionally offers the possibility of narrowing down the range of potential myxozoan hosts
(thereby providing clues for resolving parasite life cycles) and for characterising the parasite
communities in fish species consumed by piscivores. There is currently no evidence that otters or
cormorants are hosts of myxozoans, but we cannot exclude the possibility that myxozoans detected
365 in faeces were associated with infections of otters or cormorants themselves.

There was substantial variation in the representation of myxozoan SSU rDNA types across the well
sampled freshwater libraries. Ordination analyses showed that samples from the same sites, albeit
from different time points, clustered together when abundance information was also considered. For
example, OTU3 showed 100% sequence similarity with *Myxobilatus gasterostei* and was found in
370 all Bickton and Avon pool libraries but not in the California Lake samples. It was also present in the
two replicate otter faeces (but not the cormorant faeces or otter spraint 2). OTU2, identical to
Chloromyxum truttae (AJ581916), was found in the Bickton and Avon pool samples but only in
July. These results suggest that eDNA sampling at different times may be informative for inferring
seasonality of transmission and deducing life cycles. In addition, the presence of multiple novel

375 OTUs in the otter samples, some of which matched OTUs in water samples, suggests that the broad
sampling achieved by piscivore activities offers a means of characterising diversity that would be
much more difficult to assess in other ways. More extensive joint studies of eDNA in water and
piscivore faeces over time and space could inform on transmission strategies and turnover in
parasite species and communities.

380 No myxozoans were found in the eDNA surveys of marine and freshwater environments using
general eukaryotic primers, suggesting that these primers are inappropriate for investigations that
aim to detect and characterise myxozoans. This is reinforced by the lack of detection of
myxosporean sequences by general eukaryote V4 primers when analysed by MiSeq HTS using
material (otter faeces) that was shown to be positive for myxosporeans by lineage specific PCR.

385 The lack of detection with general primers that, in theory, have no mismatches to the majority of
myxozoans, may be related to the complexity of most environmental samples. For example, the
relative rarity of myxozoan spores may explain the lack of detection. Modifications to improve the
general primer approaches with respect to myxozoan detection could include e.g. combining general
primers with blocking primers designed against highly represented taxa, or using primer designs
390 that bias against common metazoans that may otherwise dominate samples. Our preliminary results
using “antimetazoan” primers demonstrate the success of this approach (data to be published
elsewhere).

Our replicate sampling from otter spraints suggested that presence/absence based analyses are
robust to subsampling and PCR introduced noise, as for example, 21 of the 23 myxosporean OTUs
395 were found in both replicates. Such differences are likely to arise from differences in the initial
subsamples taken from the spraint, especially as the two missing OTUs from replicate B were
within the rarest 5 OTUs in replicate A. Also, the top nine abundant OTUs were the same in both
replicates, however, the rank order of the OTUs was not identical. Using read abundances from
HTS is notoriously difficult, and at best semi-quantitative due to spurious differences in
400 amplification efficiency introduced during PCR. Such effects are especially likely in the case of a

nested PCR approach, as used here. Our results suggest that read abundances could be somewhat indicative of relative abundances of taxa within samples, but that such comparisons will nevertheless be characterized by significant uncertainty (especially in the case of nested PCR approaches, as used in this study).

405 In view of our results we highlight several issues and make recommendations for future work on myxozoan eDNA. Firstly, the use of serial filtration is recommended using cloth meshes of 55 and 20 μm . Fewer positive PCRs and weaker bands were obtained from small size fraction samples ($<3\mu\text{m}$), suggesting that the majority of myxosporeans were caught in larger mesh sizes. This may partly be due to imperfect size fractionation during filtration, as the meshes clog with increased
410 volume processed, the effective pore size is reduced and smaller particles than the nominal mesh size are trapped. We note that detections are likely to be improved by disturbing sediments prior to water collection. Such disturbance is likely to resuspend spores or infected host material and was suggested by the relatively strong amplicon bands that were observed during gel electrophoresis for samples collected after kicking the sediments. Improved detection when sampling near the bottom
415 as opposed to surface water has also been noted for detecting fish eDNA (Turner et al., 2015) and for the crayfish plague agent *Aphanomyces astaci* (Strand et al. 2014). The downside of such an approach may be a reduction in water volumes that can be filtered because of clogging as well as resuspension of DNA particles preserved in sediments and potentially no longer indicative of current parasite presence (Levy-Booth et al. 2007).

420 Secondly, we recommend sampling large water volumes (5-20L at minimum) to detect the relatively rare spores of myxosporeans. However, we also caution that increasing sample volumes will also concentrate PCR inhibitors that may hinder successful amplification. We did not test the effects of inhibition and some samples may have presented false negative results. Future studies would benefit from application of internal controls to assess the level of inhibition. However, in
425 many cases, inhibition is difficult to deal with, even if its presence is detected.

Thirdly, we recommend lineage-specific amplicon sequencing using HTS platforms. Our results suggest that even within pooled DNA extractions from very large volume samples, myxosporean diversity is limited when compared to other parasites (e.g. Hartikainen et al. 2014a). For example, the large volume freshwater samples contained on average only 16 OTUs (max 28, min 4) and the species accumulation curves (see Supplementary Fig. S1) subsequently saturated at around 10-50,000 reads (we sequenced at minimum to 170,000 reads). Such considerations obviously depend on the sample type, but for most applications coverages of 25,000 reads are likely to be adequate to characterise myxosporean diversity. This would allow multiplexing many libraries into a single sequencing run and screening of multiple sites and sampling points. In fact, with the very high coverage and relatively low complexity in this study, a rarely mentioned problem of “leakage” between libraries was observed. This is also referred to as index misassignment and may interfere with high sensitivity amplicon sequencing approaches (Nelson et al., 2014). Some OTUs, very abundant in only one library (e.g. >120,000 reads), also occurred in very low read frequencies in all other libraries (1-6 reads). An obvious example was the leakage of 1-3 reads from the single marine library into the freshwater libraries and we used this to set the level of the threshold (0.1%) for accepting the presence of an OTU in a library. Such technical artifacts during the MiSeq run are often not accounted for and may result from sequencing and image analysis errors during the index sequencing phase of the run (Nelson et al., 2014). Carry-over of amplicons from previous runs was not possible as the two runs (first containing the cormorant library and second containing all other libraries) with the myxosporean barcodes and primers were run several months apart. Although the thresholding may have discarded some very rare true occurrences, this was necessary in our data as the “leakage” greatly affected presence/absence estimates and species accumulation estimates. Such effects are difficult to detect in diverse OTU datasets but, if unaccounted for, are nevertheless likely to affect the interpretation of HTS data in general.

Lineage-specific amplicon sequencing clearly improves the detection of myxosporeans, however, not all myxosporeans are likely to be detected using the methods developed here. Additional primer

development would be required to incorporate myxozoans such as sphaerosporids and malacosporeans.

In conclusion, we demonstrate that it is possible to detect a diversity of myxosporeans through
455 eDNA approaches. Amplicon sequencing using a nested primer approach (custom barcoding) can
be employed to gain insights into myxozoans within systems and can be expected to expand the
known diversity of myxozoan species. Such approaches are likely to be particularly valuable for
parasites, such as myxozoans, whose species identification is highly reliant on molecular signatures
due to cryptic speciation, morphological plasticity and convergence that hinder species recognition
460 based on morphology. The analysis of myxozoan DNA in faeces of piscivores may offer a
particularly efficient method of sampling for myxozoan diversity. The movements and activities of
these animals may provide an integrated view of diversity over much broader areas than can be
effectively sampled by characterising eDNA in water samples or by using traditional parasitological
surveys for fish infections. Focusing sampling efforts in habitats where there is a high diversity and
465 density of potential hosts will optimize sampling effort (Poulin and Morand, 2014).

Finally, we point out that these approaches are relevant for the management of parasitic diseases.
Accurate identification of disease agents is essential, and an ability to gather geographic
information on the range of parasites present in different locations is relevant for understanding
disease risk. The value of characterising parasite communities as a whole, and not just specific
470 disease agents, is often overlooked in wildlife management. However, multiple infections can
exacerbate disease. Furthermore, the distributions of many parasites are changing as a result of
environmental degradation and invasions. Detecting parasites via eDNA approaches provides a pro-
active means of recognising potential infection pressures on a range of hosts and the potential for
spill-over and disease outbreaks. Lineage-specific amplification of parasite eDNA offers a means of
475 characterising and monitoring entire communities of parasites with relevance for issues ranging
from conservation of endangered hosts to the health of fish stock and sustainable aquaculture.

Acknowledgements

Our research was supported by a grant from the SynTax award scheme supported by the Linnean
480 Society, BBSRC and NERC awarded to BO and DB and by the Natural Hazards and Resources
Initiative funded by the NHM. We thank Carl Sayer, Emily Alderton and Paul Bradley for help in
obtaining otter spraints, Colin Little for provision of robin faecal sample, Suzanne Williams, Grant
Stentiford, Stuart Ross and Matthew Green for help with filtering water samples from the Tamar
Estuary, Cecile Reed, Kevin Purdy, Katja Lehmann, Andrew Singer, Bryony Williams and Kristina
485 Hamilton for other sample collection and DNA extractions, Craig-Baker Austin and Rachel Foster
for help with filtering water samples returned to the laboratory, and Darren Butterworth for
permission to access localities located on the River Avon.

References:

Alama-Bermejo, G., Jirků, M., Kodádková, A., Pecková, H., Fiala, I., Holzer, A. S. 2016. Species
490 complexes and phylogenetic lineages of *Hoferellus* (Myxozoa, Cnidaria) including revision of the
genus: A problematic case for taxonomy. Parasit Vectors 9, 1.

Alexander, J.D., Kerans, B.L., El-Matbouli, M., Hallett, S.L., Stevens, L., 2015. Annelid-
myxosporean interactions, in: Okamura, B., Gruhl, A., Bartholomew, J. (Eds.), Myxozoan
Evolution, Ecology and Development Springer International Publishing Switzerland, pp. 217-234.

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

Atkinson, S.D., Bartošová-Sojková, P., Whipps, C.M., Bartholomew, J.L., 2015. Approaches for
characterising myxozoan species, in: Okamura, B., Gruhl, A., Bartholomew, J. (Eds.), Myxozoan
500 Evolution, Ecology and Development. Springer, Cham, Switzerland, pp. 111-123.

- Bartholomew, J.L., Atkinson, S.D., Hallett, S.L., Lowenstine, L.J., Garner, M.M., Gardiner, C.H., Rideout, B.A., Keel, M.K., Brown, J.D., 2008. Myxozoan parasitism in waterfowl. *Int. J. Parasitol.* 38, 1199-1207.
- 505 Bartošová, P., Fiala, I., Hypša, V. 2009. Concatenated SSU and LSU rDNA data confirm the main evolutionary trends within myxosporeans (Myxozoa: Myxosporidia) and provide an effective tool for their molecular phylogenetics. *Mol. Phylogenet. Evol.* 53, 81-93.
- Bartošová, P., Freeman, M. A., Yokoyama, H., Caffara, M., Fiala, I. 2011. Phylogenetic position of *Sphaerospora testicularis* and *Latyspora scomberomori* n. gen. n. sp. (Myxozoa) within the marine urinary clade. *Parasitology* 138, 381.
- 510 Bass, D., Stentiford, G.D., Littlewood, D.T.J., Hartikainen, H., 2015. Diverse applications of environmental DNA methods in parasitology. *Trends Parasitol.* in press.
- Bråte, J., Logares, R., Berney, C., Ree, D.K., Klaveness, D., Jakobsen, K.S., Shalchian-Tabrizi, K., 2010. Freshwater *Perkinsea* and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *The ISME Journal* 4, 1144-1153.
- 515 Canning, E.U., Okamura, B., 2004. Biodiversity and evolution of the Myxozoa. *Adv. Parasitol.* 56, 43-131.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* 7, 335-336.
- 520 Dong, L.F., Smith, C.J., Papaspyrou, S., Stott, A., Osborn, A.M., Nedwell, D.B., 2009. Changes in benthic denitrification, nitrate ammonification, and anammox process rates and nitrate and nitrite reductase gene abundances along an estuarine nutrient gradient (the Colne Estuary, United Kingdom). *Appl. Environ. Microbiol.* 75, 3171-3179.
- Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Meth.* 10, 996-998.
- 525

- Eszberbauer, E., Atkinson, A., Diamant, A., Morris, D., El-Matbouli, M., Hartikainen, H., 2015. Myxozoan life cycles: practical approaches and insights, in: Okamura, B., Gruhl, A., Bartholomew, J. (Eds.), *Myxozoan Evolution, Ecology and Development*. Springer International Publishing, Cham, Switzerland, pp. 175-198.
- 530 Fiala, I., Bartošová-Sojková, P., Okamura, B., Hartikainen, H., 2015. Adaptive radiation and evolution within the Myxozoa, in: Okamura, B., Gruhl, A., Bartholomew, J. (Eds.), *Myxozoan Evolution, Ecology and Development* Springer International Publishing Switzerland, pp. 69-84.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., 2012. The Protist Ribosomal Reference database (PR2): a catalog of
 535 unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, gks1160.
- Hallett, S.L., Atkinson, S.D., Bartholomew, J.L., Székely, C., 2015. Myxozoans exploiting homeotherms, in: Okamura, B., Gruhl, A., Bartholomew, J.L. (Eds.), *Myxozoan Evolution, Ecology and Development*. Springer International Publishing, Cham, Switzerland, pp. 125-135.
- 540 Hallet, S.L., Bartholomew, J., 2006. Application of a real-time PCR assay to detect and quantify the myxozoan parasite *Ceratomyxa shasta* in river water samples. *Dis. Aquat. Org.* 71, 109-118.
- Hartikainen, H., Ashford, O.S., Berney, C., Okamura, B., Feist, S.W., Baker-Austin, C., Stentiford, G.D., Bass, D., 2014a. Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians. *The ISME Journal* 8, 177-186.
- 545 Hartikainen, H., Gruhl, A., Okamura, B., 2014c. Diversification and repeated morphological transitions in endoparasitic cnidarians (Myxozoa: Malacosporea). *Mol. Phylogenet. Evol.* 76, 261-269.
- Hartikainen, H., Stentiford, G.D., Bateman, K.S., Berney, C., Feist, S.W., Longshaw, M., Okamura, B., Stone, D., Ward, G., Wood, C., 2014b. Mikrocytids are a broadly distributed and divergent
 550 radiation of parasites in aquatic invertebrates. *Curr. Biol.* 24, 807-812.

- Hawkins, R.J., Purdy, K.J., 2007. Genotypic distribution of an indigenous model microorganism along an estuarine gradient. *FEMS Microbiol. Ecol.* 62, 187-194.
- Holland, J.W., Okamura, B., Hartikainen, H., Secombes, C.J., 2010. A novel minicollagen gene links cnidarians and myxozoans. *Proc. R. Soc. Biol. Sci. Ser. B* 278, 546-553.
- 555 Holzer, A.S., Bartošová, P., Pecková, H., Tým, T., Atkinson, S., Bartholomew, J., Sipos, D., Eszterbauer, E., Dyková, I., 2013. 'Who's who' in renal sphaerosporids (Bivalvulida: Myxozoa) from common carp, Prussian carp and goldfish—molecular identification of cryptic species, blood stages and new members of *Sphaerospora* sensu stricto. *Parasitology* 140, 46-60.
- Hung, Y.W., Remais, J., 2008. Quantitative detection of *Schistosoma japonicum* cercariae in water
560 by real-time PCR. *PLoS Negl. Trop. Dis.* 2, e337.
- Jimenez-Guri, E., Philippe, H., Okamura, B., Holland, P.W.H., 2007. *Buddenbrockia* is a cnidarian worm. *Science* 317, 116-118.
- Levy-Booth, D.J., Campbell, R.G., Gulden, R.H., Hart, M.M., Powell, J.R., Klironomos, J.N., Pauls, K.P., Swanton, C.J., Trevors, J.T., Dunfield, K.E. 2007. Cycling of extracellular DNA in the
565 soil environment. *Soil Biol Biochem* 39, 2977-2991.
- Logares, R., Audic, S., Bass, D., Bittner, L., Boute, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., 2014. Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* 24, 813-821.
- Lom, J., Dykova, I., 2006. Myxozoan genera: definition and notes on taxonomy, life-cycle
570 terminology and pathogenic species. *Folia Parasit.* 53, 1-36.
- Lom, J., Dyková, I., 1992. Protozoan parasites of fishes. Elsevier Science Publishers.
- Mächler, E., Deiner, K., Steinmann, P., Altermatt, F., 2014. Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Sci.* 33, 1174-1183.

- Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R.,
575 Claverie, J.M., Decelle, J., 2015. Marine protist diversity in European coastal waters and sediments
as revealed by high-throughput sequencing. *Env. Microbiol.* in press, DOI: 10.1111/1462-
2920.12955.
- Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L., Graf, J. 2014. Analysis, optimization and
verification of Illumina-generated 16S rRNA gene amplicon surveys. *PloS One* 9, e94249.
- 580 Nesnidal, M.P., Helmkampf, M., Bruchhaus, I., El-Matbouli, M., Hausdorf, B., 2013. Agent of
whirling disease meets orphan worm: phylogenomic analyses firmly place Myxozoa in Cnidaria.
PloS One 8, e54576.
- Okamura, B., Gruhl, A., Bartholomew, J., 2015. An introduction to myxozoan evolution, ecology
and development, in: Okamura, B., Gruhl, A., Bartholomew, J. (Eds.), *Myxozoan Evolution,*
585 *Ecology and Development.* Springer International Publishing, Cham, Switzerland, pp. 1-20.
- Poulin, R., Morand, S., 2014. *Parasite biodiversity.* Smithsonian Institution.
- Prunescu, C.-C., Prunescu, P., Pucek, Z., Lom, J., 2007. The first finding of myxosporean
development from plasmodia to spores in terrestrial mammals: *Soricimyxum fegati* gen. et sp.
n.(Myxozoa) from *Sorex araneus* (Soricomorpha). *Folia Parasit.* 54, 159-164.
- 590 Rangel, L.F., Rocha, S., Casal, G., Castro, R., Severino, R., Azevedo, C., Cavaleiro, F., Santos, M.
J. in press. Life cycle inference and phylogeny of *Ortholinea labracis* n. sp.(Myxosporea:
Ortholineidae), a parasite of the European seabass *Dicentrarchus labrax* (Teleostei: Moronidae), in
a Portuguese fish farm. *J Fish Dis* doi/10.1111/jfd.12508/full
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu,
595 L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic
inference and model choice across a large model space. *Syst. Biol.* 61, 539-542.
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web
600 servers. Syst. Biol. 57, 758-771.

Strand, D.A., Holst-Jensen, A., Viljugrein, H., Edvardsen, B., Klaveness, D., Jussila, J., Vrålstad, T., 2011. Detection and quantification of the crayfish plague agent in natural waters: direct monitoring approach for aquatic environments. Dis. Aquat. Org. 95, 9.

Strand, D.A., Jussila, J., Johnsen, S.I., Viljamaa-Dirks, S., Edsman, L., Wiik-Nielsen, J., Viljugrein,
605 H., Engdahl, F., Vrålstad, T., 2014. Detection of crayfish plague spores in large freshwater systems. J. Appl. Ecol. 51, 544-553.

Thomsen, P., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Willerslev, E., 2012. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 21, 2565-2573.

610 Whipps, C.M., Kent, M.L., 2006. Phylogeography of the cosmopolitan marine parasite *Kudoa thyrsites* (Myxozoa: Myxozporea). J. Eukaryot. Microbiol. 53, 364-373.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614-620.

615 Figure Legends

Fig. 1. Summary of material examined and libraries generated for sequencing. Samples collected for this study and those available from other studies are indicated (extant samples and BioMarKs project). See Supplementary Table S1 for further details. Twelve MiSeq libraries were created after pooling positive samples per environmental type and using myxosporean primers for custom
620 barcoding. Four libraries were created using general eukaryote primers applied to faecal samples. Waterfowl are goose and mallard, garden birds are robin, pigeon and blackbird. * and ** indicate two libraries that comprised pooled positive PCRs from many freshwater localities in South Africa (**) and from several sites from coastal marine localities in southern England (UK Marine *).

Fig. 2. Twenty OTUs comprising the majority of the reads in the normalised samples. OTUs are
625 split according to the major phylogenetic clade as also shown in Figure 3.

Fig. 3. Maximum Likelihood phylogenetic estimate of the myxosporean relationships. Labelled major clades are shown in more detail in Figure 4. Bootstrap values >80 are not shown. The width of the triangles is not indicative of taxon sampling.

Fig. 4. Maximum likelihood topology with branch supports estimated with 1000 bootstrap
630 replicates. (A) Marine lineage. (B) Freshwater lineage without myxobolid clade. (C) *Myxobolus* clade. The relationships of the subtrees are shown in Figure 3. Taxa in bold were retrieved from the various environmental samples by custom barcoding. Branches with bootstrap supports >80 are labelled with black solid dots, support values are indicated for branches with >50 bootstrap support, branches with support <50 are unlabelled. Occurrence of each OTU generated in this study is
635 plotted using environment specific symbols. The filled symbols indicate particularly abundant OTUs. OTUs with number and species name indicate cases where OTU similarity was >97% to a known sequence type and clustered with a reference sequence during OTU calling.

Supplementary Fig. S1. Rarefaction curves showing species accumulation as sequence coverage
640 increases. (A) Data rarefied to smallest library size. (B) Data non-rarefied.

Supplementary Fig. S2. Abundance of reads in each library split for each main phylogenetic lineage
detected in this study. The data were normalised before plotting, making the read abundances
comparable across the libraries.

Supplementary Fig. S3. Heatmap of OTU abundance across the different samples and average
645 hierarchical cluster analysis.

Table 1. Faecal samples tested for detection of myxozoans. The letters in brackets identify samples that were pooled for sequencing in a MiSeq library (n=4) following amplification with ‘general eukaryote’ SSU rDNA primers. Column n shows the number of samples of each type, with the number of samples positive for myxosporean custom barcode primers. * indicates that the sample was subsequently used for HTS to reveal the myxosporeans present.

Organism	n	Habitat			Collection
		type	Location	County	date
domestic goose (a)	3 (0)	farmyard	Port Meadow	Oxfordshire	27.11.2012
Canada goose (a)	4 (0)	pond	Crockerton	Wiltshire	27.11.2012
mallard (a)	3 (0)	parkland	Hinksey Park	Oxfordshire	10.12.2012
otter (b)	2 (1)*	lake shore	Malham Tarn	Yorkshire	03.11.2013
otter (b)	1 (1)	pond	near Dereham	Norfolk	18.01.2014
		Lake			
otter (b)	1 (1)*	shore	Ketteringham	Northamptonshire	17.01.2014
robin (c)	1 (0)	garden	Salisbury Plain	Wiltshire	10.12.2012
pigeon (c)	1 (0)	garden	Oxford	Oxfordshire	10.12.2012
blackbird (c)	1 (0)	garden	Oxford	Oxfordshire	10.12.2012
		Playing			
worm (d)	3 (0)	field	near Erlestoke	Wiltshire	27.11.2012
great cormorant	1 (1)*	fish ponds	Isla Mayor	Sevilla	5.12.2013

Table 2. PCR and sequencing primers.

	Forward primer	Sequence 3`-5`	Reverse primer	Sequence 3`-5`	Expected amplicon size
1st	myxo_617F	CGCGCAAATTAC	myxo_23	CGTTACCGGAAT	approx.
PCR	_all	CCAMTCCA	13R_all	RRCCTGACAG	900bp
2nd	myxo_764F	CCGCGGTAATTC	myxo_18	ATTTCACCTCTCG	450-490bp
PCR	_all	CAGCTCCAG	17_v1*	CCATCGA	
			myxo_18	ATTTCACCTCTCG	450-490bp
			17_v2*	CGGCMAA	
			myxo_18	ATTTCACCTCTCG	450-490bp
			17_v3*	CTGCCAA	

* Three versions of the same reverse primer were combined at equimolar concentrations to create a degenerate reverse primer only encompassing the nucleotide differences exhibited by the myxosporeans used in the reference alignment. The pooled primer is referred to in the text as myxo_1817_mix.

Table 3.HTS libraries generated using the myxosporean custom barcodes.

Library code	Description	Number of myxosporean OTUs	Number of reads
Faecal 1	otter spraint 1A Ketteringham	23	359096
Faecal 2	otter spraint 1B Ketteringham	22	613712
Faecal 3	otter spraint 2 Malham Tarn	29	287888
Faecal 4	cormorant faeces Spain	8	466120
Water 1	Avon pool July	12	342752
Water 2	Avon pool September	19	435727
Water 3	Bicton July	28	265784
Water 4	Bicton September	22	374046
Water 5	California Lake July	11	179008
Water 6	California Lake September	4	175508
Water 7	pooled South Africa freshwater	7	229069
Water 8	pooled UK marine	18	325530

Figure

Figure 2 colour

Figure 2 bw

Figure 3

Figure 4A colour

Figure 4A bw

Figure 4B colour

Figure 4B bw

Figure 4C Colour

C

Figure 4C BW

C