



University of
Nottingham
UK | CHINA | MALAYSIA

Environmental DNA approaches and host genomics for myxozoan parasite ecology and disease prediction

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

August, 2025

Florian Melzer

20316105

Supervisors: Hanna Hartikainen, Stephen Dugdale

School of Life Sciences
University of Nottingham

Table of Contents

Abstract	1
Declaration.....	3
Funding	3
Acknowledgements.....	3
List of Figures.....	4
List of Tables	9
Introduction and background	10
Parasites in ecosystems	10
Myxozoans	11
Detection of myxozoans	13
eDNA in pathogen and host detection	14
eDNA quantification in rivers.....	15
Population genetic information from eDNA	18
Chapters and aims	20
Outcomes and contributions	23
Chapter 1: Simultaneous detection and quantification of host and parasites in riverine systems using environmental DNA	26
Abstract	26
Introduction	27
Methods.....	34
Results.....	42
Discussion	51
Conclusion.....	58
Chapter 2: Rapid direct sequencing, de-novo assembly (and annotation) of complete mitogenomes using long read technology with application in the freshwater bryozoan <i>Fredericella sultana</i>	60
Abstract	60
Introduction	61
Methods.....	64
Results.....	70
Discussion	83
Conclusion.....	92

Chapter 3: River network scale population genetics of the bryozoan	
<i>Fredericella sultana</i>	93
Abstract	93
Introduction	94
Methods.....	98
Results.....	103
Discussion	114
Conclusion.....	123
Chapter 4: Ecological similarities of fish, invertebrate and parasite	
<i>communities: concordance highlighted by eDNA, conventional count data and</i>	
<i>ecological traits</i>	124
Abstract	124
Introduction	125
Methods.....	128
Results	133
Discussion	148
Conclusion.....	152
Bibliography.....	154
Appendices	178
Chapter 1 – Supplementary	178
Sampling kit preparation & guide.....	178
eDNA extraction using GuHCl and Longmire’s buffer (after Qiagen	
DNeasy Blood & Tissue protocol)	179
Chapter 2 - Supplementary.....	183
Step-by-Step: Rapid direct sequencing, de-novo assembly and	
annotation of complete mitogenomes using long read technology	
with application in the freshwater bryozoan <i>Fredericella sultana</i> ..	183
Chapter 4 – Supplementary	204

Abstract

Host-parasite relations shape evolution, community biodiversity and disease dynamics. Myxozoa are a group of diverse, internal parasites belonging to phylum Cnidaria with a wide range of invertebrate and fish hosts. In riverine freshwater ecosystems the complex around the myxozoan parasite *Tetracapsuloides bryosalmonae* with the bryozoan *Fredericella sultana* as its primary host and salmonids as its intermediate host has been in focus because *T. bryosalmonae* can cause the high-mortality Proliferative Kidney Disease (PKD) impacting wild and farmed fish.

Spatial and temporal variation in host-parasite co-occurrences influences disease dynamics but makes research through conventional methods challenging. I present here an optimised, simultaneous capture approach of both host environmental DNA (eDNA) and parasite spores from river water. Detection and quantification of host and parasite(s) through highly sensitive, species-specific qPCR and broad-range metabarcoding reveals variation across a single river catchment. I demonstrate that eDNA-based methods work at a practical scale and will be crucial in modelling host and parasite co-occurrence and prediction of seasonal population changes.

Host population genetics let us trace community (structural) changes and can relate to parasitic infection susceptibility. The genomics of the host *F. sultana* are still largely unexplored. The mitogenome sequence, widely used in detection, quantification and population genetics for both tissue and eDNA-based studies, is unknown. We developed a direct sequencing and de-novo assembly approach for the complete mitogenome, including gene annotation and order, expression profiles and discovery of a group II intron. Knowledge of sequence, gene order and transcription profiles expands metabarcoding, direct or long-range sequencing and eRNA possibilities, and, crucially for PKD dynamics, leads to insights into population genetics of the host *F. sultana*.

Based on the complete mitogenome we establish the mitogenomic diversity of *F. sultana* on a single river-basin scale to explore intra-species variability with a simple single gene marker and the more informative full mitogenome. Both approaches reveal similar population structures with distinct haplotypes spread through the river. Additionally, we designed a mitogenomic marker panel for future tissue and eDNA work. We present a case study that can be applied in other host-parasite systems to investigate population level tolerance and resistance, spatial and temporal patterns of local adaptation, and dispersal and range expansion dynamics.

Besides population genetics biodiversity, ecology and environmental factors shared by hosts and parasites lead to similarities and co-occurrence patterns between communities. To understand if co-occurrence alone can be informative, we combined myxozoan parasite metabarcoding with physical site characteristics, biodiversity and ecological traits of fish and invertebrate species. While the comparisons highlighted concordant patterns of fish and invertebrate occurrence, they could not explain co-occurrence with parasites. Host presence and shared biodiversity are crucial determinants of parasite occurrence. However, the resulting theoretical ecological niche is not always realised leading to complex parasite distribution and prevalence of infections.

Simultaneous eDNA detection of parasites and hosts, host mitogenome population genetics and interaction study between local ecology and parasite co-occurrence are important steps to better understand host-parasite relationships and disease dynamics. This thesis contributes to the conceptual and methodological foundation to the identification of areas and local populations that are threatened by PKD.

Declaration

I hereby declare that I am the author of this PhD dissertation and that I have used only those sources and literature in the bibliography.

Nottingham, 11.08.2025



Florian Melzer

Funding

Funding for my research and this thesis has been provided by the Biotechnology and Biological Sciences (BBSRC) Doctoral Training Programme at the University of Nottingham.



**University of
Nottingham**
UK | CHINA | MALAYSIA



**Biotechnology and
Biological Sciences
Research Council**

Acknowledgements

I would like to sincerely thank my supervisors Hanna Hartikainen and Stephen Dugdale for their friendly and generous support, guidance and help in all aspects of my thesis development and research.

I would also like to thank all members of the APELab at the University of Nottingham as well as all my laboratory and office colleagues for their help with my research and their great support.

Thanks to all my family and friends for their love and support at all times.

List of Figures

Introduction and background:

Figure 1: Lifecycle of the freshwater bryozoan *F. sultana* throughout the seasons in connection with vertical and horizontal transmission of the PKD-causing parasite *T. bryosalmonae*. The bryozoan grows into colonies from statoblast that endured of the winter. Through sexual reproduction and larval dispersal, statoblast production and fragmentation new colonies can be generated. Infected bryozoan colonies produce spores during warm summer water temperatures that infect fish which in turn produce spores that can infect *F. sultana* colonies. These colonies largely die off during fall and leave statoblast behind. Statoblasts and colony fragments can harbour the parasite. 12

Chapter 1:

Figure 1: eDNA dynamics from ① release, ② admixture with different organisms and sources, ③ downstream effects like transport, degradation, deposition/resuspension, further admixture and dilution and to ④ capture and extraction in context of branched, fluvial networks ⑤. 29

Figure 2: A) Components of the basic eDNA sampling kit. This kit allows collection of three replicate water samples per site, using two filter paper sizes loaded onto re-usable filter holders, through which water is pushed in-line to provide increased capture in particle size range of $>0.45\ \mu\text{m}$. B) eDNA sampling method. 1a) The sampling person should face upstream against the flow when sampling. Water is picked up in a sweeping motion across a span of approximately 2m in front of the sampler to a volume of 50ml into the syringe (without filter), 1b) then the dual filter (top (orange): $5\ \mu\text{m}$, bottom (green): $0.45\ \mu\text{m}$) is placed onto the syringe and 1c) water is pressed through the combined filters and expelled behind the sampler. 2) This is repeated five times. 3) The $0.45\ \mu\text{m}$ filter is removed from the bottom (total volume filter: 250ml) and put to the side and the sampling is continued with only $5\ \mu\text{m}$ filter for five additional times (5x 50ml) to a total volume of 500ml. It is recommended that, if possible, the sampler moves across the width of the river when sampling changing position with each 50ml repeat. 35

Figure 3: Map of the River Erewash sampling sites of eDNA from river water (black dot or pie) in the Lower Trent Catchment (black outline, waterways in blue). The circle and diamonds depict occurrence of *F. sultana* eDNA in qPCR (orange when positive), occurrence of *F. sultana* in visual wading surveys (purple when positive) and occurrence of malacosporeans in metabarcoding (red when positive). 43

Figure 4: A) The volume of water that was possible to pass through filter, as a single filter alone (white bars), and as double filter (grey bars), where

double filter consists of a 5µm filter in combination with a smaller pore size filter. B) DNA yields achieved after extraction for the different filter combinations. All filter papers were extracted and DNA concentration measured separately.	44
Figure 5: DNA extraction method optimisation, where QB = Qiagen buffer, LO = Longmire's only, LE= Longmire's and Ethanol (100%) and LI = Longmire's and isopropanol (40%). For each sample and method, plotted are A) DNA yield, B) qPCR CTvalue of <i>F. sultana</i> DNA, where POS refers to known positive DNA samples extracted previously, and NEG comprises water only, and C) CT value of internal positive control molecule amplification. The reference value is indicated by NO, in these samples only water and target internal positive control were added.	45
Figure 6: A) Comparison of qPCR detection based on eDNA and wading survey with detection (+) or absence (-). B) Detection of <i>F. sultana</i> (Fs) and <i>T. bryosalmonae</i> (Tb) from qPCR assays and other myxozoan parasites from metabarcoding on the same samples and sites.	47
Figure 7: Correlation of <i>F. sultana</i> eDNA copy number and number of substrates with <i>F. sultana</i> colonies in the wading survey area in A) July and B) September. The eDNA sample was taken at the downstream portion of the wading survey area. Regression lines plotted in red or blue.	47
Figure 8: eDNA concentration of <i>F. sultana</i> as A) copy numbers per original sampling volume (Copy number/L) of both sampling dates on all sites before LOD filtering and B) in density as counts per eDNA concentration across all sites and dates.	48
Figure 9: Proportion of sites positive for <i>F. sultana</i> (Fs), <i>T. bryosalmonae</i> (Tb) and other myxozoans detected via metabarcoding for two sampling periods, A) July and B) September.	50
Figure 10: Comparison between <i>F. sultana</i> eDNA quantification by qPCR and read count of <i>Buddenbrockia</i> worm type, which is known to infect <i>F. sultana</i> , sampled during A) July and B) September.	51

Chapter 2:

Figure 1: A) Miniprep for plasmid-style extraction for (potential) mitoenrichment from metazoan/animal tissue with pre-lysis tissue disruption (enzymatic or mechanical) before alkaline lysis, neutralization, column binding, washing and final elution. B) Sequencing workflow with initial quantification and quality control, library preparation with the Nanopore Rapid Barcoding or Sequencing kit, sequencing on V14 flow cell and basecalling, followed by either alignment to reference genome or de-novo assembly (with flye).	66
Figure 2: Electrophoresis agarose gel of plasmid prep DNA extracts of <i>Oncorhynchus mykiss</i> , <i>Lumbriculus variegatus</i> and <i>Fredericella sultana</i> . O = Original extract, D = Depletion of linear DNA with Exonuclease V	

(RecBCD) treatment; MD = Mechanical disruption with pestle smear, LK = Enzymatic lysis with Longmire's Buffer and Proteinase K, CTAB = Standard CTAB extraction of tissue.	71
Figure 3: Read length distribution of the complete sequencing run with all combined samples A) Total sequenced reads B) Sequences that finally aligned to the assembled mitogenome and overlaps of long reads (>10k nt, 0.8%) with the green area indicating the length of the full mitogenome.	72
Figure 4: Annotations of the complete mitogenome from MITOS2 with all the genes from the OXPHOS proteins, rDNA, tRNAs, variable sections including the two mitocontrol regions including schematic of speculated initiation of mtDNA replication by 1) enzyme recruitment through motif (repeat) recognition and 2) RNA primer or direct strand synthesis	74
Figure 5: A) Smoothed coverage of short read MiSeq data (2x 150bp) (blue) (H. Hartikainen, unpublished) and average of all available RNAseq data (grey). B) Coverage of new long read sequencing using latest R10.4.1 flow cell technology (dark red) compared to older R9 sequencing (yellow; Hanna Hartikainen, unpublished).	77
Figure 6: A) Coverage of the Nanopore mitogenome sequencing across the complete mitogenome B) Expression profiles along the mitogenome from published and unpublished short read RNA-seq data of <i>Fredericella sultana</i> zooids or statoblasts with genomic annotation C) Strand specific expression profile of a single individual, positive (forward) and negative (reverse).	79
Figure 7: Pyrimidine-purine walk of the – strand (mtDNA) of <i>Fredericella sultana</i> with genetic annotation and regions coloured according to feature type and table of pyrimidine-purine counts and ratios	81
Figure 8: A) Location of the predicted group II intron B) Predicted IDDT score of the predicted IEP of the group II intron by AlphaFold2 C) Highest scoring predicted structure (rank 1, pLDDT=59.5 pTM=0.581) d) pIDDT scores over the predicted structure.	83

Chapter 3:

Figure 1: Map of all Erewash sampling sites (2021-2024). Empty diamonds: no bryozoans found by survey, red diamonds: bryozoans found. Lines below site names indicate number of nad3 Sanger sequences obtained from tissue and eDNA, circles indicate number of individuals sequenced for full mitogenomes.	97
Figure 2: A) Nucleotide diversity and B) Tajima's D along the full mitogenome with π (blue), Θ (yellow) and Tajima's D (red), nad3 marker region marked in red.	105

Figure 3: Cluster of nad3 marker region and full mitogenome ancestry proportions of sequenced individuals with column pattern for sites and colours corresponding to clusters or ancestry proportion.	107
Figure 4: A) Haplotype network of full mitogenome distances with colours as sampling site origins and number of nucleotide changes in the grey boxes on the connecting branches, B) Haplotype network of all full mitogenome sequences with colours as ancestry population proportion.	108
Figure 5: Haplotype network of all nad3 marker sequences with continuous colours as site origin downstream to upstream of the River Erewash and the Swiss outgroup (including tissue and eDNA).	110
Figure 6: Map of the River Erewash with pie charts of the combined tissue and eDNA haplotype clusters (colour) of the nad3 marker at the sampling sites.	112
Figure 7: Comparison of cluster assignments (colours) of tissue (dots) and eDNA (lines) samples for all tested sites (River Erewash, UK and Swiss outgroup) with the number of sequences for each site on the left-side scale.	113

Chapter 4:

Figure 1: Map of the eight sampled catchments Wharfe and Ouse Lower, Aire and Calder, Witham, Trent Lower and Erewash, Dove, Soar with Wreake, Avon with Isobourne and Cherwell and Ray. Green circles show eDNA parasite metabarcoding sites, blue triangles for EA fish survey and orange diamonds for EA invertebrate survey sites.	129
Figure 2: Data processing and analysis of EA survey count data, ecological species traits and parasite metabarcoding results.	131
Figure 3: A) Clustering of ecological traits for all fish species in the ecological trait database (clusters indicated by colour). B) Subset of fish present in the sampling sites with subclusters and species names indicated by colours (colours do not indicate same clusters in A and B).	134
Figure 4: 10-year average of insect (A) and non-insect invertebrate (B) abundance from EA sampling sites.	136
Figure 5: NMDS ordinations based on ecological traits of A) all invertebrate species found in the invertebrate ecological trait database, B) subset of species found in the full EA database and C) species with a species level identification in the EA database. Shape indicates trait cluster association and colour indicates taxonomy.	138
Figure 6: A) Occurrences of invertebrate species sorted by phylum to class taxonomy and trait clusters (colour), B) Total parasite read numbers of the eDNA metabarcoding in relation to relative abundance of oligochaetes on the sampling sites with correlation (R) and p-value (p), C)	

Total parasite read numbers of the eDNA metabarcoding in relation to relative abundance of the trait clusters (colours) on the sampling sites with correlation (R) and p-value (p).	139
Figure 7: Myxozoan parasite diversity from eDNA metabarcoding with total occurrences on the tested sites and species of the parasite genera.	140
Figure 8: Non-metric multidimensional scaling (NMDS) graphs of fish species and trait abundance and composition on all sites. A) Fish species abundance, B) Fish species composition, C) Fish trait abundance, D) Fish trait composition.....	142
Figure 9: Non-metric multidimensional scaling (NMDS) graphs of invertebrate species and trait abundance and composition as well as physical site parameters on all sites. A) Invertebrate species abundance, B) Invertebrate species composition, B) Invertebrate traits abundance, D) Invertebrate traits composition, E) Physical parameters.	143
Figure 10: Non-metric multidimensional scaling (NMDS) graphs of parasite species and trait abundance and composition. A) Parasite species abundance, B) Parasite species composition, C) Parasite trait abundance, D) Parasite trait composition	144

Supplementary - Chapter 1:

Figure 1: Inhibition Positive Control (expected Ct = 36.1) of all eDNA samples from both sampling dates with and without 1:2 dilution.	182
--	-----

Supplementary - Chapter 2:

Figure 1: DNA extraction using a commercial kit like NEB Monarch® Spin Plasmid Miniprep Kit.....	183
Figure 2: Direct Nanopore sequencing and de-novo assembly workflow	186
Figure 3: Example of BLAST results from a read subset identifying closest BLAST database match.....	202
Figure 4: YR graphs of all bryozoans with mitochondrial Group II introns	203

List of Tables

Chapter Introduction:

Table 1: Example literature for detection, quantification and sequencing of the host <i>F. sultana</i> and the parasite <i>T. bryosalmonae</i>	14
---	----

Chapter 1:

Table 1: Number of surveyed <i>F. sultana</i> colonies (wading) and detected copy numbers after LOD filtering in the originally sampled volume (qPCR) on the sampling dates in the River Erewash.	49
---	----

Chapter 2:

Table 1: Sampling sites on the River Erewash, Derbyshire/Nottinghamshire, England, United Kingdom on 8 th July 2024 with geographic information and samples sequenced. Nr. = number	67
Table 2: Results of initial lysis and plasmid extraction tests	70
Table 3: Genomic annotation of genes, rRNA, tRNA and repeats of the mitogenome with position, length, direction and expression levels (where applicable)	75

Chapter 3:

Table 1: Sampling sites on the River Erewash and collection dates, sample replication and locality data. Tissues indicated in bold were used in a whole mitogenome-based approach.	98
Table 2: Comparative statistics of marker sequences from all and just mitogenome sequenced individuals and of the full mitogenomes.....	103

Chapter 4:

Table 1: Procrustes analysis of ordination comparison including fish, invertebrate and physical site parameters as m^2 with better concordance indicated by blue shading and significant values in bold.....	146
Table 2: Procrustes analysis of ordination comparison of parasite parameters internally, against fish and invertebrates as m^2 with better concordance indicated by blue shading and significant values in bold.	147

Supplementary - Chapter 1:

Table 1: River Erewash sampling sites	181
--	-----

Introduction and background

1 This introduction provides background to the role of parasites in
2 ecosystems, particularly myxozoans and their bryozoan hosts. I provide
3 details of how eDNA is currently used in research to reveal host-parasite
4 dynamics, paying particular attention to factors that challenge quantitative
5 approaches given metazoan targets and complex hydrology of river
6 networks.

7 Parasites in ecosystems

8 Parasites make up a large proportion of living species across the tree of life
9 (Poulin and Randhawa 2015, Okamura et al. 2018). According to Weinstein
10 and Kuris (2016) 'half of all animals may have a parasitic lifestyle'. Host and
11 parasite interactions drive evolutionary outcomes and give rise to e.g.
12 complex life cycles, variation in host range and - in some cases - frequent
13 host transitions, which shape the often parallel evolution leading to
14 convergence and shared traits between parasitic taxa. Co-speciation
15 events in host and parasites associate with expansion of biodiversity
16 (Poulin and Randhawa 2015). However, there is a wide discrepancy across
17 parasite taxa in degree of differentiation and evolutionary variety (Hay et al.
18 2020). Incomplete or unknown evolutionary and life histories hinder full
19 understanding of parasite evolution and ecology. Parasitism can therefore
20 be challenging to study.

21 Multistage lifecycles, small size and overall cryptic nature make detection
22 and identification of parasitic taxa through traditional sampling difficult.
23 Additionally, micro-eukaryotic parasites are phylogenetically and
24 functionally extremely diverse and they are not always associated with
25 disease, yet play important biological roles as moderators or regulators,
26 contributing to the ecological equilibrium of ecosystems.

27 Myxozoans

28 Within metazoans, the Myxozoa are an important group and represent the
29 most diverse radiation of cnidarians (Okamura et al. 2011, Hartikainen et
30 al. 2014, Chang et al. 2015, Holzer et al. 2018). Their large diversity of 60
31 genera and around 2160 species (Hartikainen et al. 2014) accounts for 20%
32 of all described cnidarians (Atkinson et al. 2018). Two clades make up the
33 myxozoans: Myxosporea and Malacosporea. In multistage life cycles they
34 infect fish, reptiles and amphibians as intermediate hosts and annelid or
35 bryozoans as definitive host in terrestrial, marine and freshwater
36 environments. We focus here on the endoparasitic and trophically
37 transmitted *Tetracapsuloides bryosalmonae* (Class Malacosporea) which
38 infects phylactolaemate bryozoans, mostly *Fredericella sultana*, as well as
39 salmonid fishes (Hedrick et al. 1993) (see Fig. 1). Infection in fish can lead
40 to high mortality among wild or farmed fish via development of Proliferate
41 Kidney Disease (PKD). PKD is therefore recognised as an epidemiological
42 risk with wide-reaching ecological and economic impacts (Okamura et al.
43 2011, Svavarsdóttir 2016). Infection in bryozoans can spread systemically
44 due to covert infection strategies of the parasite (Abd-Elfattah et al. 2014),
45 and recovered fish remain chronically infected in some cases, spreading
46 the parasite apparently asymptotically (Bettge et al. 2009a, Grabner
47 and El-Matbouli 2009). The small size and internal localisation of
48 myxozoans in fish hosts makes their detection challenging, necessitating
49 extensive temporal and spatial sampling of many hosts. Comprehensive
50 data are therefore almost challenging to achieve with traditional sampling
51 (i.e.: sampling by collecting individual hosts across the life cycle), leading
52 to underestimation of parasite presence or abundance (Duval et al. 2024).
53 A better monitoring approach is required to aid in detecting imminent
54 disease risks, or delaying detection pathogen invasions (Stelzer et al.
55 2024). For containment or prevention measurements – be it in
56 conservation or industry context – a robust assessment of
57 presence/absence, as well as an accurate estimate of abundance of

58 parasites and (intermediate), host can be crucial in combating or avoiding
 59 disease spread (Carraro et al. 2018, Hutchins et al. 2021).

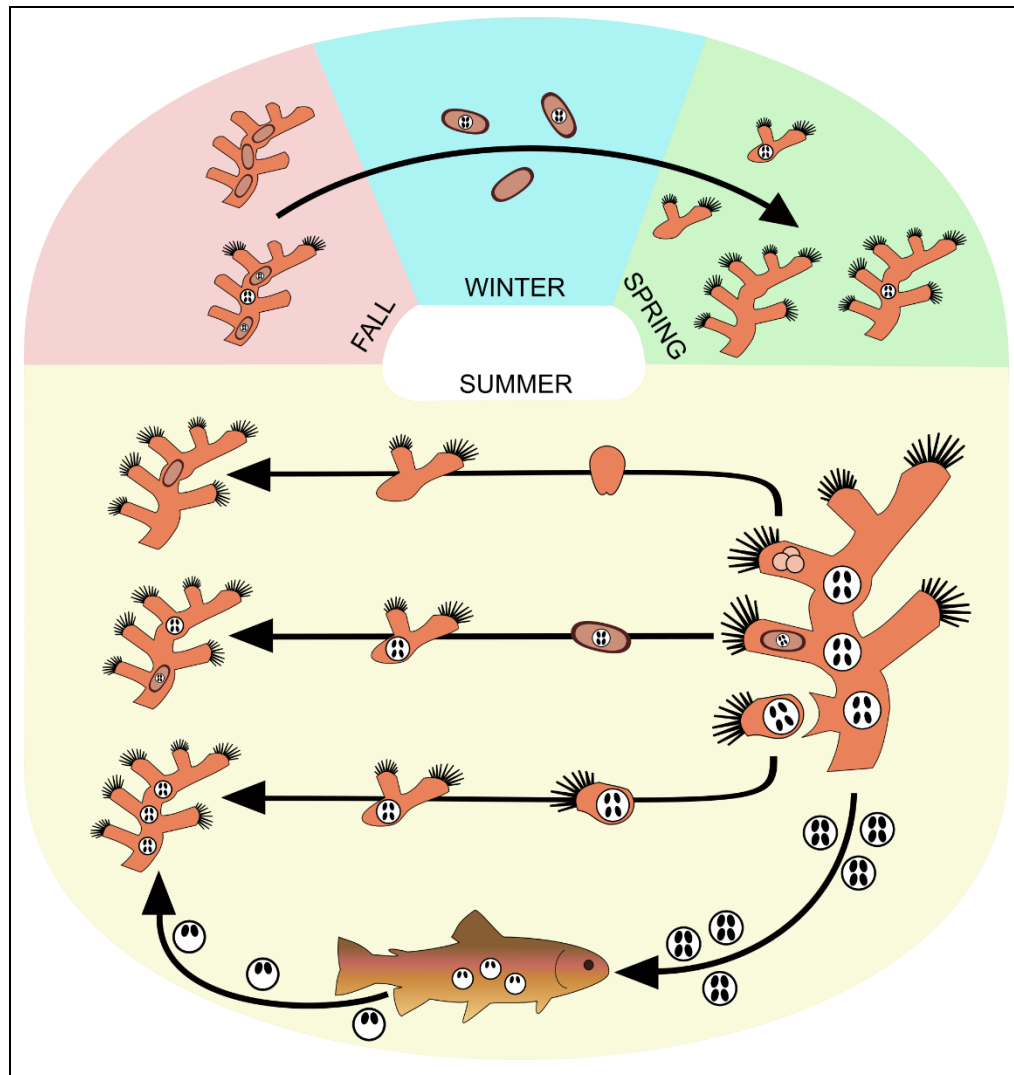


Figure 1: Seasonal lifecycle of the freshwater bryozoan *F. sultana*, with vertical and horizontal transmission of the PKD-causing parasite *T. bryosalmonae*. The bryozoan grows into colonies from overwintering statoblasts in the spring. Larval dispersal, statoblast production and colony fragmentation generate new colonies via sexual and asexual routes. Infected bryozoan colonies release parasite spores mostly during summer. Spores released from bryozoans infect fish, where PKD develops. Mature spores released to water from fish can infect *F. sultana* colonies. The parasite can disperse along the clonal growth of the bryozoan, and infect both statoblasts and colony fragments.

60 *T. bryosalmonae* and *F. sultana* follow a lifecycle influenced by seasons
 61 and water temperature (Fig. 1): Statoblasts of *F. sultana*, dormant
 62 propagules from the previous year, grow into branched, clonal colonies
 63 through budding with warmer spring temperatures (Bilton et al. 2001,

64 Okamura et al. 2019). Colonies live as filter feeders on suitable substrate
65 such as rocks and roots in slow to moderate river currents. In summer a
66 population peak occurs through sexual and asexual reproduction leading
67 to either dispersible larvae or statoblast generation and fragmentation of
68 zooids which form their own colonies. The growth of freshwater bryozoans
69 colonies can lead to biofouling issues (Mant et al. 2011, Mant et al. 2013).
70 In fall with decreasing temperatures, most of the colonies die off leaving
71 the statoblast behind which overwinter until the next spring. As competent
72 primary host, infected colonies of *F. sultana* release *T. bryosalmonae*
73 malacospores which infect salmonoid fishes via the gills which
74 themselves release spores infecting bryozoan colonies.

75 Development of parasite spores is linked to water temperature and *F.*
76 *sultana* population size (Carraro et al. 2018). Release of bryozoan infecting
77 spores by the fish through urine completes the parasite lifecycle
78 (Hartikainen and Okamura 2015). In bryozoans, the parasite also spreads
79 through vertical transmission by harbouring in statoblasts (Abd-Elfattah et
80 al. 2014) or fragmented zooids (Gorgoglione et al. 2016).

81 Detection of myxozoans

82 Bryozoans and parasite infections can be detected in numerous ways
83 (Table 1). Tissue-based methods mainly focus on infection prevalence and
84 disease progression. Assessment of infection prevalence and intensity rely
85 on collection of bryozoan colonies through wading surveys, or dissection
86 of kidneys mostly in brown trout *Salmo trutta* and rainbow trout
87 *Oncorhynchus mykiss*. Parasite presence can be determined visually
88 under a microscope and through histology (Clifton-Hadley et al. 1984,
89 Longshaw et al. 2002) or through (q)PCR amplification of extracted DNA
90 (Bettge et al. 2009a). Molecular tools play a key role in assessment of
91 parasite load (Table 1, Strepparava et al. (2018)), and different diagnostics
92 are used to determine disease severity (Clifton-Hadley et al. 1987).

93 Recently, several eDNA-based detection approaches involving qPCR or
 94 metabarcoding have been developed. eDNA targets slough genetic host
 95 material and parasite spores. Therefore, primarily disease dynamics
 96 concerning vertical transmission events and the parasite pressure
 97 (prevalence or occurrence rates in the ecosystem) on either bryozoans or
 98 salmonids fish can be studied with eDNA.

Table 1: Example literature for detection, quantification and sequencing of the host *F. sultana* and the parasite *T. bryosalmonae*

Approach	<i>F. sultana</i>	<i>T. bryosalmonae</i>
Detection (qPCR)	<u>eDNA</u> : Carraro et al. (2017), (Fontes et al. 2017), Carraro et al. (2018) <u>Survey & kidney qPCR</u> : Syrová et al. (2020), Ros et al. (2021)	
		<u>eDNA</u> : Hutchins et al. (2018), Oredalen et al. (2022) <u>kidney</u> : Bettge et al. (2009a), Bettge et al. (2009b), Strepparava et al. (2018), Hutchins et al. (2021)
Sequencing	<u>Population genetics</u> : Hatton-Ellis et al. (1998), Freeland et al. (2000), Freeland et al. (2001) <u>Microsatellites</u> : Hartikainen and Jokela (2012), Filippenko et al. (2014)	<u>Metabarcoding</u> : <u>eDNA</u> : Hartikainen et al. (2016)

99 eDNA in pathogen and host detection

100 DNA (and RNA) is continuously shed by organisms into the environment
 101 through mechanisms such as death, decay or abrasion. This leads to
 102 accumulation of genetic material in soil, sediment or water. DNA obtained
 103 from this spectrum is called environmental DNA (eDNA). While the study

104 of its complement, environmental RNA (eRNA), is more recent (Parro et al.
105 2007, McGrath et al. 2008, Poretsky et al. 2009), eDNA has been used since
106 the late 1990's to detect species, for examples see: (Taberlet and Luikart
107 1999, Ficetola et al. 2008). In contrast to most traditional sampling, eDNA
108 is non-invasive and non-destructive, and the simultaneous capture of
109 eDNA shed from all living organisms in the environment provides suitable
110 data for studying complex species interactions. eDNA is widely used to
111 study birds (Monge et al. 2020) and fish (Sato et al. 2017, Uchii et al. 2017b,
112 Landínez-García et al. 2020, Shu et al. 2020, Tsuji et al. 2020), as well as
113 invertebrates like fan worms (Von Ammon et al. 2019) and mussels
114 (Sansom and Sassoubre 2017, Marshall and Stepien 2019). Although less
115 explored, eDNA is also used for integrated approaches targeting rarer
116 species such as parasites, together with their more abundant hosts, be it
117 mobile and migratory fish or sessile invertebrates hosts such as bryozoans
118 (Carraro et al. 2017, Carraro et al. 2018).

119 eDNA quantification in rivers

120 Can eDNA be used to estimate abundance? Despite many studies
121 focussing on bacterial taxa routinely using read counts to quantify
122 abundance, in metazoans, the individual mechanisms of production and
123 emission of DNA into the environment can be difficult to assess (Sansom
124 and Sassoubre 2017, Bylemans et al. 2018, Jo et al. 2019, Jerde 2021).
125 Release rates are dependent on various non-continuous and fluctuating
126 factors. Life history differences of parasites and hosts make interspecific
127 comparison and generalities challenging. Species-specific release
128 mechanisms and seasonality shape eDNA presence and abundance
129 (Pochon et al. 2017, Sansom and Sassoubre 2017, Bylemans et al. 2018,
130 Jo et al. 2019, Zilius et al. 2021). Biomass, the number and size of
131 individuals correlate to DNA emission, but not always equally. Many
132 species are dormant for large parts of a year, for example amphibians,
133 insect and molluscs become inactive or hibernate in cold winter
134 temperatures. The consequential reduction in DNA release can make

135 detection difficult. Reproduction, including increase in activity and
136 production as well as gamete release, and rapid seasonal growth, facilitate
137 detection but may correlate eDNA read counts better with activity rather
138 than biomass. Further complications arise from different environmental
139 factors triggering active or reproductive periods in a species-specific
140 manner (Sansom and Sassoubre 2017, Jo et al. 2019).

141 Detection and quantification become patchy at the distribution range
142 edges and invaded areas, as species may only flourish in certain
143 conditions, or seasons, even though they are well established (Lawson
144 Handley et al. 2019, Seymour et al. 2021, Zilius et al. 2021). Particularly
145 bryozoans may be influenced by excessive release of DNA due to die offs
146 during drought or flood periods, and during natural die-offs at the end of
147 the growing season (Pochon et al. 2017). Additional environmental
148 aspects, such as temperature, weather events or daylength, might
149 influence activity or growth, further cloud the correlation between DNA
150 emission and biomass (Sansom and Sassoubre 2017, Jo et al. 2019).

151 These biotic and abiotic factors can lead to an uneven contribution of
152 individuals and communities to the eDNA detected in a river. Continuous
153 sampling over a longer time and on a larger scale covering the extreme
154 ends of a species in terms of survivability and activity is recommended
155 (Sigsgaard et al. 2017b, Reinholdt Jensen et al. 2021). Ideally, laboratory
156 experiments such as tank experiments can substantiate eDNA sampling
157 by assessing impacts of abiotic factors such as temperature and daylength
158 on the relationship between biomass and DNA emission.

159 The experiments could also shed light on the decay rates that affect
160 released DNA. eDNA is naturally unstable and will steadily fracture outside
161 an organism until it can no longer be detected. Abiotic factors such as
162 temperature and water chemistry (pH, presence or absence of certain
163 chemicals) can stabilize or weaken DNA resulting in longer or shorter half-
164 lives (Sansom and Sassoubre 2017, Bylemans et al. 2018). Together with

165 the dilution rate the decay rate - resulting in loss over distance and time -
166 determines the transportation and distribution of eDNA.

167 Complete sampling for eDNA aims to collect a random, unbiased
168 subsample of all DNA molecules present at a site. For aquatic riverine
169 environments this means that the local accumulation is mostly
170 established by upstream emerging DNA transported downstream by the
171 river's current (Pont et al. 2018, Shogren et al. 2019). River water samples
172 therefore consist of an only small portion of eDNA emitted in close
173 proximity and a much larger portion of molecules brought from distant
174 locations. The input of remote DNA is determined by initial quantity of DNA
175 emission connected to release rates – for example number and size of
176 individuals - as well as decay, dilution, and dispersion factors – river flow
177 speed, filtering effects or additional water admixture (Carraro et al. 2018,
178 Altermatt et al. 2020, Carraro et al. 2020a, Carraro et al. 2020c).

179 These biotic and abiotic factors decide the given species detection
180 probability from a water sample. The experimental design then influences
181 detection chance further. Initial sampling size and replicate number will be
182 hugely influential for any downstream application (Taberlet et al. 1999,
183 Waits and Paetkau 2005, Carroll et al. 2018, Lugg et al. 2018, Azarian et al.
184 2020, Carraro et al. 2020c). Consider a water body containing different
185 groups of DNA molecules with different proportions making up the overall
186 frequency. Any sampling will pick a number of molecules at random. The
187 chance of selection a molecule from a rare group such as parasites
188 increases with the number of molecules drawn. Increase volume and
189 replication of the draw also increases the chance of recovering rarer
190 molecule types. eDNA can reliably detect a range of elusive myxozoan
191 parasites in river systems at different time points (Hartikainen et al. 2016,
192 Marshall et al. 2022, Oredalen et al. 2022, Lisnerová et al. 2023).

193 Prior modelling and experiments as well as the inclusion of geomorphical,
194 hydrological and ecological parameters can contribute to a robust
195 sampling scheme (Carraro et al. 2017, Carraro et al. 2018, Carraro et al.

2020a, Carraro et al. 2020c). Presence/absence detection through eDNA has been conducted (for examples see: (Agersnap et al. 2017, Elbrecht et al. 2017, Stat et al. 2017, Lacoursière-Roussel et al. 2018, Marshall and Stepien 2020)), mostly using standard barcoding markers (16S, 18S, COI etc.) and qPCR or NGS techniques. Applying mathematical-statistical models of dispersion rates based on DNA release, decay and river flow rates allows the conversion of eDNA concentrations into parameters indicating biomass or individual numbers (Olds et al. 2016, Carraro et al. 2017, Carraro et al. 2018). It further enables the prediction of a continuous quantity distribution along the riverscape from a collection of singular datapoints.

207 Population genetic information from eDNA

208 eDNA opens new possibilities to investigate population level ecology (209 Waits et al. 2001, Andrews et al. 2018, Parsons et al. 2018, Adams et al. 210 2019, Sigsgaard et al. 2020, Tsuji et al. 2020, Zizka et al. 2020). Due to the 211 degraded nature of eDNA, marker-based analysis of genetic diversity is 212 used to study and examine patterns of gene flow. Haplotype level studies 213 allow us to better understand the distribution rates and limits of a given 214 species (Stat et al. 2017, Dufresnes et al. 2019, Marshall and Stepien 2019, 215 Andres et al. 2021). The choice of appropriate markers is critical but 216 difficult and typically must be designed and tested from the ground up. 217 Recent advances in whole genome sequencing in a wide range of taxa have 218 made this easier, however number, quality, and completeness of unique 219 genomes can be limiting.

220 Informative, universal gene region with sufficient allelic variation to use as 221 a marker for population genetics is challenging to obtain. A range of 222 markers has been shown to be beneficial (Tsuji et al. 2018, Corse et al. 223 2019, Jensen et al. 2021). High diversity of the chosen marker is often 224 traded off with universality, particularly in selection of the conserved 225 flanking regions for primer design. A multi marker approach is often 226 recommended to overcome this problem, however, phasing of the alleles

227 from different markers becomes problematic in eDNA as the DNA pool is
228 mixed, and DNA fragmented. Mitochondrial DNA (mtDNA) is the most
229 plentiful source of genetic material from eDNA (Jensen et al. 2021).
230 Metabarcoding studies sequencing the mitochondrial Cytochrome c
231 oxidase subunit I (COI or COX1) have shown that haplotype level
232 information can be extracted from existing markers (Elbrecht et al. 2018,
233 Marshall and Stepien 2019) however, the level of genetic diversity obtained
234 is low. Nuclear DNA (nDNA) is less abundant than mitochondrial genomes
235 in cells and therefore less prevalent in environmental samples. Still nDNA
236 can be readily gained for sequencing (Andres et al. 2021). With its larger
237 non-coding and genetic variable regions, especially introns and internal
238 transcribed spacers (ITS) and, nDNA could potentially offer more
239 delimiting power than mtDNA.

240 Short tandem repeats (STRs) like microsatellites (and minisatellites) are
241 the most obvious target for population genetics. They are considerably
242 more variable than other genetic regions and most underly weak selection
243 pressure. As a result, they are ideal for haplotype level analysis. A
244 collection of STRs and their count variations of the DNA repeat motifs can
245 identify single individuals in a population (De Barba et al. 2017). However,
246 in eDNA attributing such multi-locus genotypes to individuals is not
247 possible due to the mixed input DNA from multiple individuals. Instead,
248 STR signature patterns can be assessed according to sampling site (or
249 locality), akin to pooled double digest restriction-site associated DNA
250 (ddRAD)analysis.

251 Single nucleotide polymorphisms (SNPs) of coding and non-coding gene
252 regions have been widely used in genetics. They are the basis for (meta-)
253 barcoding and species delimitations. Standard markers such as COI or
254 16S have been tried and tested in population genetics but due to their
255 limited genetic variance might lack the power for studies on a small spatial
256 scale or in systems with general low intraspecific genetic diversity
257 (Elbrecht et al. 2018, Marshall and Stepien 2019). Targeting SNPs will give

258 lower information content as they are less variable but more stable as
259 mostly restricted to non-functional base changes. The same sites should
260 be variable in all alleles making sequencing errors easier to detect at the
261 cost of the degree of variation. They provide a more conservative basis and
262 could be combined with STRs for best results.

263 Due to these limitations, amplicon analysis of several mitochondrial and
264 nuclear loci was selected here for testing. In preliminary analysis, we could
265 show that mitochondrial markers designed after an intra- and interspecific
266 genome analysis and targeting the likely most variable regions can detect
267 haplotype level differences in eDNA samples from a single river. Further,
268 eDNA and traditional individual sequencing can capture the same degree
269 of environmental variance (Waits et al. 2001, Elbrecht and Leese 2015,
270 Fonseca 2018).

271 **Chapters and aims**

272 This project aims to assess the occurrence, quantity and overall
273 biodiversity of parasite and host species, focusing on the myxozoan *T.*
274 *bryosalmonae* and the bryozoan *F. sultana* during the summer months in
275 UK rivers. To understand the epidemiology of aquatic diseases and
276 associated exposure risks it is necessary to detect and monitor host and
277 pathogen dispersal, parasite infection rates and temporal or seasonal
278 variation. Non-invasive eDNA sampling enables tracking of the host and
279 parasite distribution along riverscapes. This will allow us to estimate the
280 scope of pathogen pressure on associated fish farms or local fishing
281 communities. The key aspects of the project are monitoring of presence
282 and absence, quantification, and population structure analysis. The
283 capability of eDNA for species detection or metabarcoding, including
284 density or population size assessments, have been broadly demonstrated
285 (Jerde et al. 2011, Elbrecht et al. 2017, Stat et al. 2017, Lacoursière-
286 Roussel et al. 2018, Dufresnes et al. 2019, Harper et al. 2019, Jeunen et al.
287 2019, Liu et al. 2020, Turon et al. 2020). The possibility of eDNA for studying

288 population level ecology has started to attract research in the last few
289 years (Waits et al. 2001, Andrews et al. 2018, Elbrecht et al. 2018, Adams
290 et al. 2019, Azarian et al. 2020, Sigsgaard et al. 2020, Tsuji et al. 2020, Zizka
291 et al. 2020, Jensen et al. 2021). I believe that eDNA can not only detect
292 presence and abundance of parasites and hosts but can also aid in
293 understanding genetic diversity on the haplotype level and its relation to
294 parasite disease dynamics (Bellis et al. 2021). Host distribution together
295 with environmental factors shapes parasite abundance and host-parasite
296 dynamics (Duval et al. 2024). Shared ecology and community similarities
297 can explain part of the spatial variation in host and parasite co-occurrence
298 (Grunberg 2021).

299 1) *Simultaneous detection of host (F. sultana) and parasite (T.*
300 *bryosalmonae) based on eDNA (Chapter 1)*

301 Methods for detection and quantification from eDNA have been well
302 established for the parasite *Tetracapsuloides bryosalmonae* and its host
303 *Fredericella sultana* (Hartikainen et al. 2016, Carraro et al. 2017, Carraro
304 et al. 2018). Building on these existing methods I developed an optimised,
305 simultaneous capture approach of both host eDNA and parasite spores
306 and present its results from a complete river system revealing the presence
307 of bryozoans and myxozoan parasites. Capability in exploring co-
308 occurrences allows us to understand host parasite interactions and the
309 resulting disease dynamics in riverscapes. This could lay the groundwork
310 for conservation and disease monitoring as well as commercial
311 applications.

312 2) *Sequencing and de-novo assembly of F. sultana mitogenome*
313 *(Chapter 2)*

314 The mitogenome is widely used in detection, quantification and
315 population genetic studies for both tissue and eDNA-based studies. As
316 the genomics of the host *F. sultana* are still largely unknown, past studies
317 relied on short marker sequences in the rRNA or COI region. I developed

318 and present the results of a direct sequencing and de-novo assembly
319 approach for the complete mitogenome, including gene annotation and
320 order, expression profiles and discovery of a group II intron.

321 *3) Host diversity through mitogenome-based population genetics*
322 *(Chapter 3)*

323 Ecology and population genetics of the host *F. sultana* might play a large
324 role in shaping parasite distribution and disease dynamics. I establish the
325 mitogenomic diversity of the *F. sultana* on a single river-basin scale to
326 explore intra-species variability along the full mitogenome and with a
327 single gene marker. Both approaches reveal similar population structures
328 with distinct haplotypes spread through the river. Based on these results
329 we designed a mitogenomic marker panel for future tissue and eDNA work.

330 *4) Exploration of ecology and co-occurrence of hosts (fish, invertebrates)*
331 *and parasites (myxozoans) (Chapter 4)*

332 Hosts and parasites live in a shared environment. Shared ecology and co-
333 occurrence patterns can lead to correlations and concordance between
334 sites. I investigate if the presence, number and amount of fish and
335 invertebrate species from count data matches ordinations of their
336 ecological traits and if they can be brought into concordance with parasite
337 numbers from metabarcoding. Ultimately, I determined strong
338 concordance between fish and invertebrate species parameters, but only
339 weak concordance with parasites. Local biodiversity and ecology do not
340 necessarily predict host and parasite (co-)occurrence.

341 **Outcomes and contributions**

342 I set out to investigate the host-parasite dynamics in riverine ecosystems
343 related to the PKD-causing myxozoan parasite *T. bryosalmonae* and its
344 invertebrate host *F. sultana*. The four chapters together make a
345 contribution to the assessment of parasite occurrence, exposure risk and
346 enhance our tools to reveal the temporal and spatial characteristics of
347 parasite and host distributions, as well as exploring the mechanistic basis
348 of parasite community assembly.

349 The findings leverage the use of eDNA, and I propose that eDNA is a
350 powerful tool to detect and quantify the presence of the parasite across a
351 wide range of aquatic systems, particularly informative when combined
352 with eDNA analysis of the invertebrate host on the same samples. I
353 developed an optimised and adaptable approach for the simultaneous
354 detection of host eDNA and parasite spores, and correlated data from
355 quantitative real-time PCR and myxozoan metabarcoding (Chapter 1). The
356 data are novel as they reveal variation among a small river and its
357 catchment, showing that eDNA-based analyses are sensitive enough to
358 detect variation at this practical scale for river management, and will be
359 crucial for modelling of host and parasite co-occurrence and prediction of
360 seasonal population changes.

361 The mitochondrial genome is the basis for many genetic identification
362 techniques across a broad range of taxa. From species-specific (q)PCR to
363 metabarcoding, mitochondrial gene loci such as 16S or COI are the
364 favoured targets (Stelzer et al. 2024), with long read technology
365 increasingly feasible and favoured method of tracing eDNA origins and to
366 deduct population genetics (Deiner et al. 2017, Jensen et al. 2021,
367 Takehashi et al. 2022). The complete annotated mitogenome of the
368 freshwater bryozoan, including the discovery of the group II intron (Chapter
369 2), facilitates the adoption of these approaches in PKD research.
370 Knowledge of gene order and transcription profiles expands
371 metabarcoding, direct or long-range sequencing and eRNA possibilities,

372 and, crucially for PKD dynamics, may provide insights into population
373 genetics of the host *F. sultana*.

374 My results apply and demonstrate this principle in a single river network
375 (Chapter 3), and will contribute to better predictive understanding of PKD
376 epidemiology. Hosts and parasite engage in constant co-adaptation,
377 which can be observed in disease dynamics, during, e.g. range expansions
378 and changing environmental conditions (Bellis et al. 2021). Results in
379 Chapter 3 provide a case study that can be applied in other host-parasite
380 systems to investigate population level tolerance and resistance, spatial
381 and temporal patterns of local adaptation, and dispersal and range
382 expansion dynamics.

383 To understand if co-occurrence alone can be informative, we combined
384 myxozoan parasite metabarcoding with biodiversity and ecological traits of
385 fish and invertebrate species (Chapter 4). While the comparisons
386 highlighted concordant patterns of fish and invertebrate occurrence, they
387 could not explain parasite co-occurrence. This suggests that species-
388 specific identity is more important in shaping parasite distributions than
389 ecological function. It underscores the importance of protecting
390 biodiversity if we are to conserve parasites as parts of natural, well-
391 functioning ecosystems. As parasites track specific species of hosts,
392 rather than the ecological roles of the hosts, ecosystem level conservation
393 that focuses on resilience (the presence of many different species with the
394 same functional trait) is crucial. It also highlights the idea that host
395 presence and shared biodiversity are crucial for parasite occurrence, and
396 re-enforces the idea in Chapter 1, which promotes detection of hosts at
397 the same time as parasites, however, these are not always realised leading
398 to complex parasite distribution and infection patterns.

399 Simultaneous eDNA detection of parasites and hosts, host mitogenome
400 population genetics and interaction study between local ecology and
401 parasite co-occurrence are important steps to better understand parasite
402 disease dynamics. PKD continues to have a drastic impact on wild and

403 farmed fish. Detection of the parasite is in demand as part of regular
404 biomonitoring and to trace PKD outbreaks and together the chapters in this
405 thesis contribute to the conceptual and methodological foundation to the
406 identification of areas and local populations that are threatened by PKD.

Chapter 1: Simultaneous detection and quantification of host and parasites in riverine systems using environmental DNA

KEYWORDS: parasite eDNA, water filtration, detection, quantification, rivers, PKD

Abstract

407 The distribution, frequency and movement of hosts and parasites drives
408 spatial epidemiological patterns and therefore, knowledge of how host
409 species, vectors and infection reservoirs are distributed in the environment
410 is crucial for management and control of disease outbreaks. Wildlife
411 diseases caused by pathogens with complex life cycles or broad host
412 ranges are particularly challenging as effective control strategies often
413 require assessing the distributions of a multitude of hosts species, across
414 complex landscapes. Using environmental DNA (eDNA) isolated from
415 filtered river water samples, I assessed simultaneously the occurrence of
416 an invertebrate host and a range of myxozoan fish parasites. A quick water
417 filtration method targeting specifically parasite and host eDNA capture
418 was developed using adaptable and reusable components, allowing for a
419 range of pore sizes. A large and a small pore size filters were employed at
420 the same time to drastically increase (2-3x) the filtered water volume and
421 extracted eDNA. Comparison of field surveys and eDNA detection showed
422 that detection and quantification of host species (*Fredericella sultana*)
423 eDNA performed well (~85% congruency with visual surveys). Detection of
424 parasites through myxozoan-specific metabarcoding revealed the
425 presence of a range of 5 parasite groups, and qPCR confirmed specifically
426 the non-detection of the malacosporean myxozoan *Tetracapsuloides*
427 *bryosalmonae* in the river. The results indicate that the dual filter approach,
428 with streamlined eDNA extraction procedure is highly effective at
429 capturing both host and parasite eDNA simultaneously.

Introduction

Environmental and climate change increasingly affect the biodiversity and functioning of freshwater ecosystems (Reid et al. 2019, Barbarossa et al. 2021) and are key drivers of changes in aquatic disease dynamics, including pathogen range expansions and increased disease incidence (Pulkkinen et al. 2010, Fisher et al. 2012). Understanding how spatial and temporal disease patterns respond to environmental change requires quantitative tracking of parasite and host species distributions. This, in turn, relies on the availability of sensitive and rapid monitoring and surveillance techniques (Ostfeld et al. 2005).

Surveillance is particularly challenging when infection prevalence is low, pathogens are spatially aggregated (Cable et al. 2017) or during the early stages of pathogen range and host shifts (Morales-Castilla et al. 2021). The reliability of detection methods and the capacity to identify host populations at risk are critical considerations for the implementation of new monitoring tools for wildlife management (Macaulay et al. 2022). Accordingly, the development of sensitive, scalable, and easy-to-apply detection methods, capable of identifying a pathogen presence even in the absence of overt disease is important for capturing spatial variability in disease outbreaks, assessing environmental impacts and for providing early warning of emerging threats.

eDNA techniques offer a promising approach for host-parasite detection, as both host and parasite DNA can potentially be isolated from a single environmental sample. While the use of eDNA in monitoring wildlife diseases has gained some attention (Strand et al. 2014, Huver et al. 2015, Sana et al. 2018, Hutchins et al. 2018), combined applications targeting both host and parasite are more recent (Carraro et al. 2017, Carraro et al. 2018, Robinson et al. 2018, Rusch et al. 2018, Fossøy et al. 2020, Marshall et al. 2022, Oredalen et al. 2022, Robinson et al. 2022, Bass et al. 2023, Lisnerová et al. 2023).

459 A major challenge in such multi-target systems is the potential mismatch
460 in size distribution of eDNA-containing particles from hosts and parasites.
461 For many parasites, detection depends on capturing entire spores or
462 transmission propagules (Bass et al. 2015, 2023), which are typically in the
463 size range of bacteria or protists. In contrast, host eDNA is often released
464 as sloughed cell-bound fragments, or free DNA strands, occupying much
465 smaller size ranges (Barnes and Turner 2016). The dispersion, mixing and
466 frequency of host and parasite eDNA in the environment are also likely to
467 differ (Bass et al. 2023), and eDNA decay rates may vary depending on their
468 mode of release.

469 Here I firstly review the main challenges associated with co-detection of
470 host and parasite eDNA from shared water samples. Secondly, I develop
471 and test eDNA capture and extraction methods aimed at improving
472 detection efficiency across both host and parasite targets.

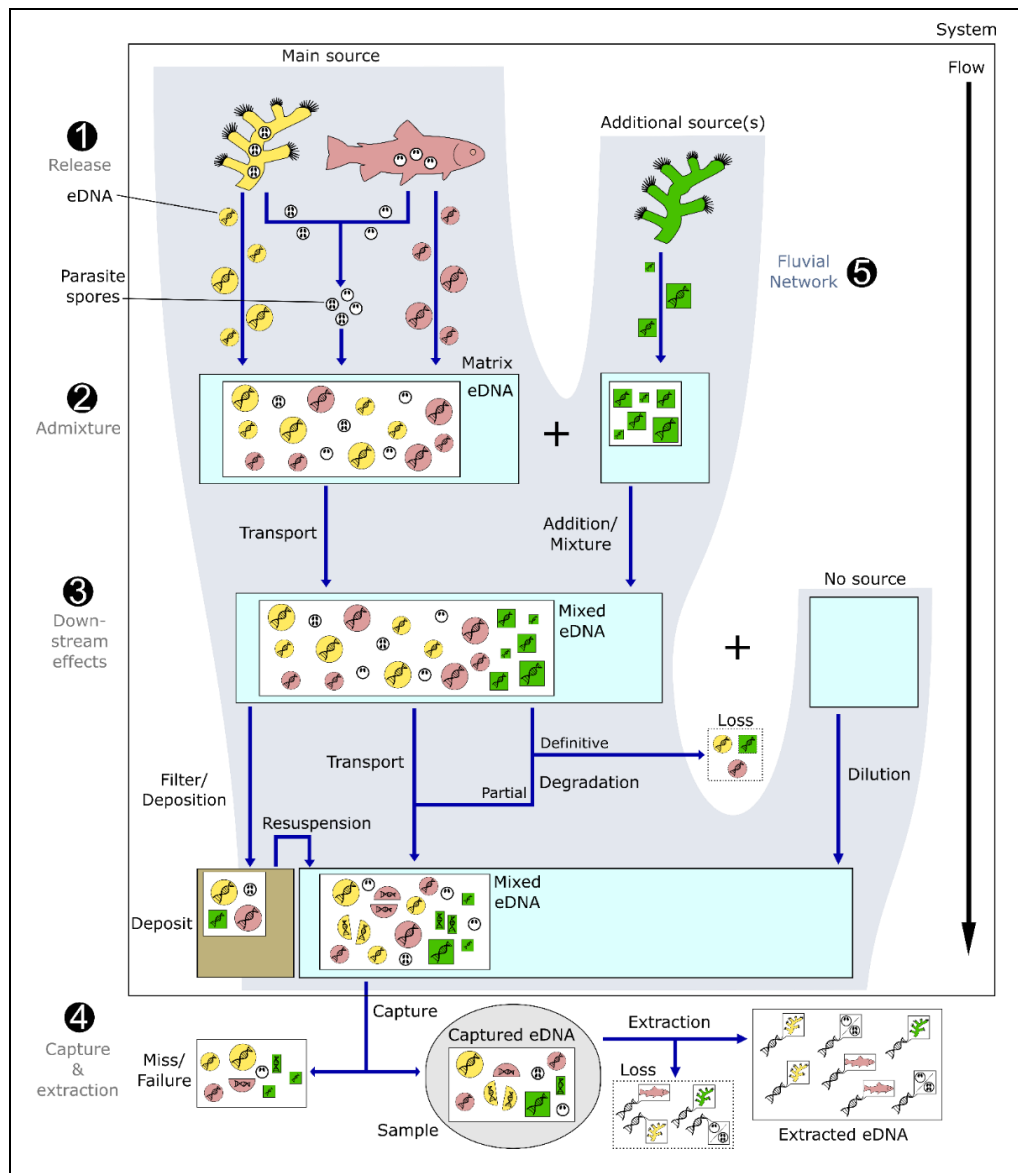


Figure 1: eDNA dynamics from ① release, ② admixture with different organisms and sources, ③ downstream effects like transport, degradation, deposition/resuspension, further admixture and dilution and to ④ capture and extraction in context of ⑤ branched, fluvial networks.

473 *eDNA dynamics and mechanisms*

474 The spatial and temporal distribution of eDNA in aquatic ecosystems,
 475 illustrated here in the context of a fluvial network, arises from
 476 unidirectional transport from the source, alongside dilution and admixture
 477 processes influenced by both network structure and hydrological flow (Fig.

1, ② and ⑤). Key processes, including release, dispersion, transport and transformation and degradation are mediated by biotic and abiotic factors, as well as species-specific biological and ecological traits (Fig. 1, ① and ③).

For hosts and parasites, eDNA release mechanisms differ markedly. Host eDNA is typically shed as a by-product of normal activity, such as movement or feeding, whereas parasites actively release transmission stages into the environment (Fig. 1, ①). These stages comprise the entire, viable organism, with traits that may enhance dispersal distance and environmental persistence (De Kinkelin et al. 2002, Morris and Adams 2007, Naldoni et al. 2019). These properties likely contribute to differences in both detectability and decay rates of host and parasite eDNA in water samples.

Parasite stages may originate from multiple different host species, which can vary widely in mobility and spatial distribution. For example, the myxozoan *Tetracapsuloides bryosalmonae* spores are released from both the sessile, suspension feeding invertebrate *Fredericella sultana* and a range of mobile salmonid fish hosts. Currently, it is not possible to identify the specific host source of parasite spores using eDNA, though eRNA (Ramón-Laca et al. 2025) or epigenetic markers may offer future solutions (Balard et al. 2024, Hirayama et al. 2024, Ruiz et al. 2025). Parasite spore release can vary seasonally (Duval et al. 2024), or follow developmental cycles (Agnew and Koella 1999, Morris et al. 2002). In contrast, host eDNA shedding is generally more stable, though still influenced by seasonal activity (Sansom and Sassoubre 2017, Wacker, Fossøy et al. 2019). Suspension feeding bryozoans, for example, continuously release faecal pellets, and although they remain untested likely contain intracellular or extracellular DNA. While the interactions between host and parasite eDNA dynamics remain unclear, the distinct biological and ecological mechanisms underlying their release and transport suggest that optimised sampling strategies are essential for improving detection.

509 Tank and mesocosm studies, as well as particle dispersion analyses, show
510 that eDNA emission in non-parasitic species depends on factors such as
511 individual number, body size, life stage and activity levels (Sansom and
512 Sassoubre 2017, Jo et al. 2019, Kirtane et al. 2021). However, there is no
513 consensus on the key drivers shaping the downstream spatial distribution
514 of eDNA post-release (Fig. 1, 3) (Jo and Yamanaka 2022). Hydrological
515 transport plays a major role in dispersal of eDNA, and although models
516 often assume complete mixing within the water column, empirical
517 evidence shows that transport is typically non-uniform and context-
518 specific (Sansom and Sassoubre 2017, Pont et al. 2018, Fremier et al.
519 2019, Shogren et al. 2019, Stoeckle et al. 2021, Wood et al. 2021a, b,
520 Carraro et al. 2022). Furthermore, differences in the eDNA “packaging”,
521 such as host-derived fragments versus intact, resilient parasite spores,
522 can influence post-release distribution. eDNA decay, often modelled as an
523 exponential function, may vary depending on a wide range of abiotic
524 factors such as pH (Strickler et al. 2015), salinity, UV radiation (Tzafesta
525 and Shokri 2025), temperature (Jo et al. 2019, Saito and Doi 2020), time
526 since shedding (Sansom and Sassoubre 2017, Troth et al. 2021), initial
527 particle size and fragment length (Bylemans et al. 2018) and species-
528 specific factors (Krehenwinkel et al. 2018).

529 Host and parasite eDNA may also be subject to distinct deposition and
530 resuspension dynamics, as seen in other species (Zhao et al. 2021). *T.*
531 *bryosalmonae* malacospores, for example, are buoyant and non-motile,
532 remaining infective for up to 24 hours (De Kinkelin et al. 2002). In many river
533 systems, this could result in transport over several kilometres, even
534 accounting for DNA decay and dilution effects (Carraro et al. 2017, Carraro
535 et al. 2018). Some myxozoan spores are highly resistant, can accumulate
536 in sediments and remain viable upon resuspension (Okamura et al. 2015).
537 Other parasites may produce transmission stages which are motile or
538 exhibit behavioural adaptations such as phototaxis, or buoyancy control
539 (reviewed in Bass et al. 2023). To improve detection of both host and

540 parasite eDNA, these physical and biological processes must be
541 incorporated into sampling design and processing protocols.

542 *eDNA capture*

543 The particle size, state and abundance of aquatic eDNA, along with water
544 characteristics such as suspended sediment load, determine the optimal
545 choice of filter material, pore size and filtration volume (Fig. 1, 4) (Spens
546 et al. 2017, Li et al. 2018, Hunter et al. 2019, Sepulveda et al. 2019, Jo et al.
547 2020, Sanches and Schreier 2020, Zaiko et al. 2022). Parasites spores may
548 be more patchily distributed than fish-derived eDNA, and increasing
549 sample replication may improve detection reliability. However, simply
550 increasing sampling volume does not always enhance detection
551 probability (Fontes et al. 2017).

552 Collecting more DNA across multiple subsamples on to the same filter
553 may help surpass per-sample quantification thresholds, yielding more
554 accurate eDNA concentration estimates. Nonetheless, larger volumes
555 increase the risk of filter clogging and PCR inhibition. Capturing parasite
556 spores may benefit from larger pore size filters, yet this may reduce the
557 capture of host extracellular eDNA, highlighting the trade-offs in designing
558 methods that work equally well for both targets.

559 For preservation, Longmire's buffer is commonly recommended (Renshaw
560 et al. 2015, Spens et al. 2017, Kumar et al. 2020a). The choice of DNA
561 extraction method balances cost, time and effort. For eukaryotic eDNA,
562 the Qiagen DNeasy kit shows high performance across studies (Deiner et
563 al. 2015, Lear et al. 2018, Li et al. 2018), although few studies have
564 compared host and parasite DNA extraction efficiency directly. Differential
565 lysis efficiency, especially of resilient parasite spores, may introduce
566 systematic biases (Bass et al. 2023).

567 *Myxozoan parasites*

568 This study aims to develop a filtration approach suitable for
569 simultaneously capturing eDNA from both freshwater bryozoan and fish
570 hosts together with buoyant parasite spores of malacosporean
571 myxozoans. These organisms release eDNA particles of different size and
572 abundance (Turner et al. 2014, Bylemans et al. 2018, Zhao et al. 2021,
573 Kumar et al. 2022).

574 Malacosporeans have a two-host life cycle, with malacospores infectious
575 to fish released by bryozoans to the water column. Bryozoan host occupy
576 broad ecological niches and exhibit extensive clonal propagation, enabling
577 extensive vertical transmission of malacosporeans such as *T.*
578 *bryosalmonae* through colony budding and resting stage formation
579 (reviewed in Hartikainen and Okamura 2015, Fontes et al. 2017c).
580 Identifying both bryozoan hosts and their parasites via eDNA would
581 contribute to understanding the spatial distribution parasite hot-spots and
582 reservoirs in bryozoans (Abd-Elfattah et al. 2014, Fontes et al. 2017b).

583 For the bryozoan hosts and malacosporean parasites, the expected eDNA
584 particle size range is 0.2 - 10µm, while parasite spores are buoyant, rare in
585 the water column, and typically 10-12µm in diameter. qPCR and
586 metabarcoding approaches have reliably detected both hosts and
587 parasites in eDNA previously (Carim et al. 2016, Carraro et al. 2018,
588 Oredalen et al. 2022, Lisnerová et al. 2023). We evaluate a two-filter, in-line
589 water filtration method that enhances host eDNA capture using a small
590 filter pore size (0.45µm), and increases parasite detection by incorporating
591 a second, larger pore size filter (5µm). Filters are combined into one
592 collection tube, yielding a single sample that integrates signals from both
593 eDNA types. This setup is low cost, manually operated and field-portable,
594 reducing waste and logistical complexity while increasing flexibility and
595 ease of use.

Methods

596 *eDNA collection kit design*

597 A flexible, reusable and cost-effective eDNA collection method via water
598 filtration was developed (Fig. 2a). The filter paper materials tested were
599 cellulose nitrate (CN) and glass fibre (GF). We tested commonly used pore
600 sizes (0.2µm CN, 0.45µm CN, 0.7µm GF) which are able to capture most
601 of the eDNA particle range. A 5µm CN filter paper was included in
602 combination with all smaller pore sizes, to retain parasite spores/larger
603 particles, whilst also preventing clogging of the smaller pore size. Typical
604 eDNA kit consumables, sampling kit setup and extraction protocol are
605 provided in Supplementary Materials.

606 The performance of the double-filter unit was tested by manually passing
607 natural river water (River Kennet, Wiltshire/Berkshire, UK) through the
608 filtration setup using the 50ml syringe (Fig 2b). Each filter paper
609 combination was tested twice. The filter papers were all saved separately,
610 and eDNA extracted separately to measure DNA capture efficiency with
611 and without the 5µm CN filter paper.

612 Subsequently, we only used the 0.45µm and 5µm CN filters in-series, the
613 water volume was standardised as follows: the first 250ml of water was
614 filtered through both size fractions, after which another 250ml was
615 processed through the 5µm filter paper only, as it remained resilient to
616 clogging.

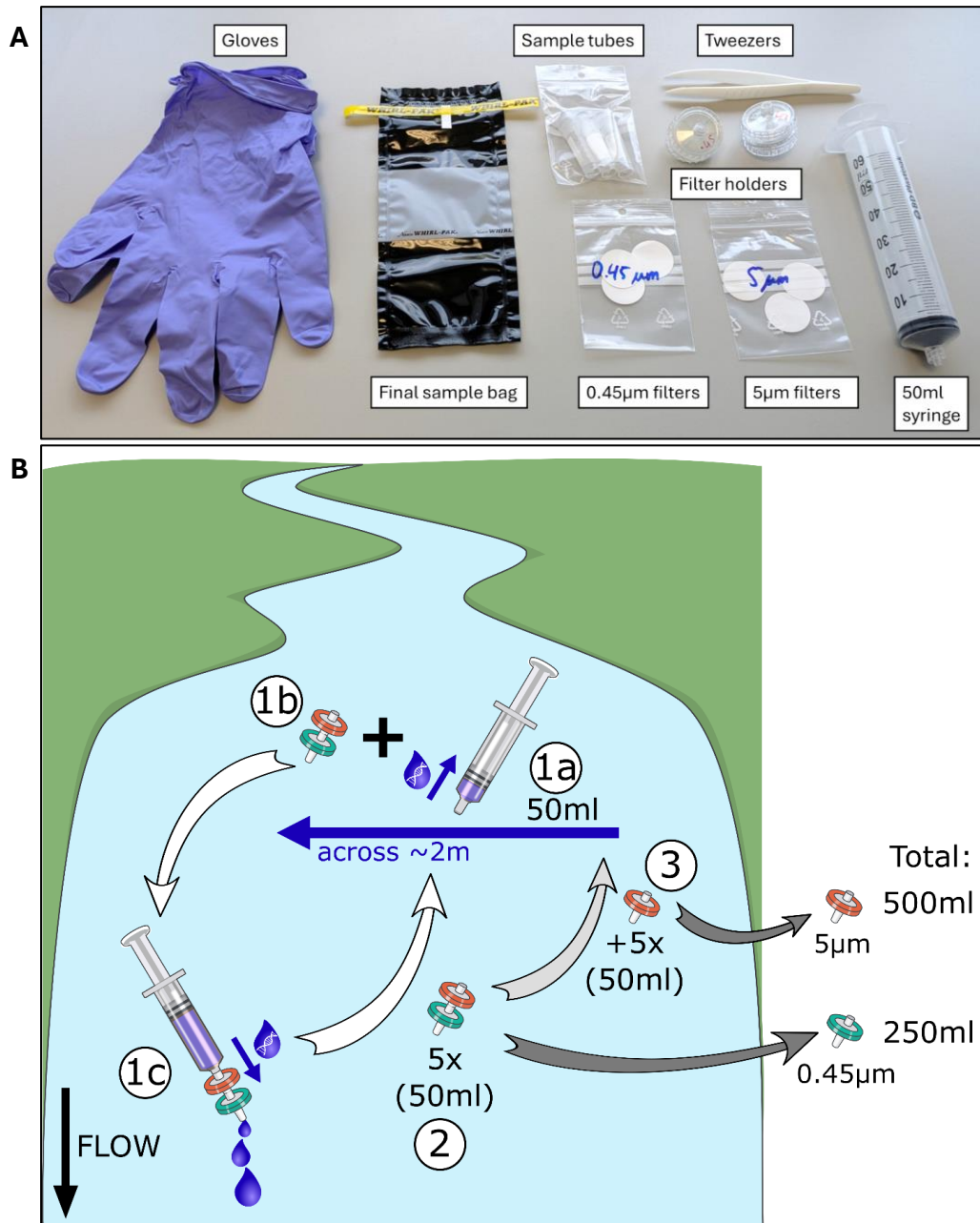


Figure 2: A) Components of the basic eDNA sampling kit. This kit allows collection of three replicate water samples per site, using two filter paper sizes loaded onto re-usable filter holders, through which water is pushed in-line to provide increased capture in particle size range of $>0.45 \mu\text{m}$. B) eDNA sampling method. 1a) The sampling person should face upstream against the flow when sampling. Water is picked up in a sweeping motion across a span of approximately 2m in front of the sampler to a volume of 50ml into the syringe (without filter), 1b) then the dual filter (top (orange): $5\mu\text{m}$, bottom (green): $0.45\mu\text{m}$) is placed onto the syringe and 1c) water is pressed through the combined filters and expelled behind the sampler. 2) This is repeated five times. 3) The $0.45\mu\text{m}$ filter is removed from the bottom (total volume filtered: 250ml) and put to the side and the sampling is continued with only $5\mu\text{m}$ filter for five additional times (5x 50ml) to a total volume of 500ml. It is

recommended that, if possible, the sampler moves across the width of the river when sampling changing position with each 50ml repeat.

617 After water filtration in the field, filter papers from both holders were
618 placed in the same collection tube for DNA extractions using sterile
619 forceps provided in the kit. If replicate samples were collected on the site,
620 the same filter holders were used by re-loading with fresh filter papers
621 included in the kit. The collection tubes comprised 2ml screw top tubes
622 containing 1ml Longmire's buffer (0.1 M Tris, 0.1 M EDTA, 10 mM NaCl,
623 0.5% (w/v) SDS) (Renshaw et al. 2015) and 200µl of 1 mm diameter glass
624 beads. When only one replicate sample was taken per filter holder setup,
625 the Longmire's buffer can be directly applied to the filter holder and sealed
626 using caps. Samples were stored in a fridge (5°C) until extraction to avoid
627 freezing and re-thawing.

628 We adapted the Qiagen DNeasy Blood & Tissue kit to accommodate
629 sample preservation with Longmire's Buffer and lysis directly in the
630 sampling tube. Using Longmire's replaces the addition of ATL buffer from
631 the kit but also increases the total extraction volume. The AL buffer from
632 the kit was also replaced as follows. According to the safety sheets, AL
633 contains 30-40% guanidine hydrochloride (GuHCl). GuHCl is a chaotropic
634 agent that together with ethanol facilitates binding of DNA to the silica
635 membrane of the spin column. By changing from AL buffer to 8M GuHCl
636 solution the total volume is halved, allowing processing in a single tube
637 using a standard centrifuge size.

638 The impact of this replacement relative to using the kit specification was
639 tested using subsamples of a single tissue sample from *F. sultana*. Mock
640 eDNA samples were prepared by taking pieces of a colony sample
641 (preserved in ethanol) and transferring it into a bead tube. Then 1ml of
642 water from the local university pond was added. The tissue was pulverized
643 with a tissue lyzer and then spun down. DNA was extracted following the
644 Qiagen DNeasy Blood & Tissue protocol with one crucial change: Instead
645 of using AL buffer and ethanol we added either one of these combinations

646 (with 2 replicates each): 1) QB - Control: AL/EtOH mix, 2) LO - GuHCl &
647 Longmire's buffer, 3) LE - GuHCl & Ethanol (100%) and 4) LI - GuHCl &
648 Isopropanol (40%). To simulate the eDNA extraction, a filter paper was
649 inserted to a bead tube and 20µl of the mock eDNA (above) added. For QB,
650 360µl ATL buffer and for all other 560µl of Longmire's buffer were added.
651 All tubes were vortexed in a tissue lyzer and 40µl proteinase K (20mg/ml)
652 added. Samples were incubated at 56°C for 3h, and protocol instructions
653 were followed to clean the DNA. For LO 200µl 8m GuHCl and 400µl
654 Longmire's, for LE 200µl 8M GuHCl and 400µl Ethanol (100%) and for LI
655 200µl and 400µl Isopropanol (40%) were added. Final volume for all
656 samples was 1200µl. The rest of the extraction test followed the Qiagen
657 DNeasy Blood & Tissue protocol. We continued with column loading and
658 fixation of DNA to membrane. Columns were eluted twice with 100µl AE
659 buffer.

660 After extraction, resulting DNA concentration was measured using Qubit,
661 the *F. sultana* specific qPCR assay (described below) and the Internal
662 Positive Control (IPC, described below) were used to measure the
663 amplification success of the extracted DNA, and the amount of inhibition
664 carried over from the different extraction methods, respectively.

665 A detailed reference filtration protocol is provided in Supplementary
666 Materials 2. All eDNA collection kit preparation was conducted in a
667 dedicated eDNA laboratory, with kit assembly taking place in a dedicated
668 UV-sterilised hood. A separate eDNA kit (Fig. 2) was prepared for each site,
669 and replicate samples per site were collected using the same syringe and
670 re-usable filter holder, after rinsing with river water. Prior to kit preparation,
671 any reusable equipment (e.g. filter holders) was disassembled to
672 components, washed in bleach (prepared as 1:10 dilution from 10%
673 laboratory grade bleach), rinsed and soaked three times with double-
674 distilled water, allowed to airdry in a loosely closed plastic container with
675 tissue paper, and sterilised under UV for 20min in a reflective UV chamber.

676 *Sampling site selection*

677 The sampling kit design was tested on field samples collected from the
678 River Erewash (Derbyshire/Nottinghamshire, England, UK), known to
679 contain the invertebrate host, *Fredericella sultana*, colonies and
680 malacosporean parasites. Following the recommendations by Carraro et
681 al. (2020) we focused sampling on the main downstream reaches to cover
682 species occurring in disjointed or independent hotspots. We expect that
683 bryozoan communities and the myxozoan parasite (Carraro et al. 2020b,
684 Carraro et al. 2023) have non-uniform occurrences along the rivers. In our
685 experience, based on previous scouting efforts, the abundance of
686 bryozoans is largely dependent on suitable substrates and can range from
687 rare to very frequent depending on the river system. If possible, we placed
688 the sampling sites on the main reach above and below confluences of
689 tributaries to control for admixture of upstream communities. Additionally,
690 at least one sampling site was included on each tributary branch, and we
691 aimed for a minimum distance between the sampling sites of 3-5km to
692 enable detection of upstream sources, while avoiding nestedness due to
693 admixture from distance sources.

694 The sampling sites on the River Erewash (15 sites) are shown in Figure 3
695 and locality information provided in Supplementary materials 3. The river
696 was sampled for eDNA on separate occasions (14-16th July, 27-30th July and
697 13-15th September 2021). Three replicate water samples were taken on
698 each site using the method described above. On the first site of each
699 sampling date, three field negative samples were generated by filtering
700 double-distilled water brought in sterile glass bottles to the field through a
701 randomly picked sampling kit.

702 *Field wading surveys*

703 A manual search to determine presence of *F. sultana* was conducted on all
704 sites on the eDNA survey date. *F. sultana* colony number on each site was
705 estimated by “catch-per-unit-effort” wading survey along approximately 50
706 m river stretch, undertaken by the same two people for 20 min. This
707 protocol standardised the search effort for each site as all available
708 potential substrata for attachment (e.g. rocks and boulders, tree roots,
709 wood pieces, debris) were examined for the presence of *F. sultana*. Both
710 banks to the middle of the stream were covered and the number of
711 substrates and substrates with colonies were recorded.

712 *DNA extraction*

713 In all cases, the two filter papers from one water sample were placed in a
714 single 2ml screwtop tube with 1ml Longmires buffer. The filter papers were
715 disrupted using a tissue homogenizer (settings: 3 x 20s at 10k rpm, 40s
716 rest) (Precellys Evolution Touch Homogenizer, Bertin Technologies, France)
717 before processing with the DNeasy Blood & Tissue kit (Cat. No. 69504,
718 Qiagen, Germany) with the following modification: to reduce the total
719 volume we replaced the addition of ATL lysis buffer with addition of 1:2 part
720 8M GuHCl solution and 1:1 part 100% ethanol to the disrupted filter in
721 Longmire’s buffer. Custom Proteinase K solution was prepared from
722 separate stocks (Pro. No. P2308, Sigma-Aldrich, USA) at concentration of
723 20mg/ml to achieve the same concentration as used in the commercial kit.
724 DNA was finally eluted in 2x 50µl PCR-grade water, with a 5min incubation
725 at 37°C to improve DNA release from the column.

726 *Quantification, inhibition and Limit of Detection - qPCR*

727 Probe-based qPCR assays were used to detect and quantify *F. sultana* and
728 *T. bryosalmonae* in the eDNA extracts. No *T. bryosalmonae* was detected
729 using methods described in Selzer et al. 2025, and hence the parasite
730 qPCR is not reported further. The 16S *Fredericella sultana* qPCR assay was
731 used to detect and quantify presence of *F. sultana* with forward primer

732 Fs_16S_F1q (300 nM): 5'-CATTGAGCTTCGGGAATGTT-3', reverse
 733 Fs_16S_R1q (900 nM): 5'-ATGAAACCTCGTCCCTTGTG-3', probe
 734 Fs_probe_16S (modified quencher, 200 nM): 5'-
 735 /5Cy5/GGGTCAGGT/TAO/TGCTAAGCCATGA/3IAbRQSp/-3' (Carraro et al.
 736 2018) and with a dilution series of gBlock of the amplicon sequence as
 737 standards (Integrated DNA Technologies Ltd). Amplifications were carried
 738 out using Promega GoTaq Probe qPCR Master Mix (Cat. No. A6102,
 739 Promega, USA) on an ABI 7500 Fast (Applied Biosystems, USA) with the
 740 settings recommended by Promega. The degree of inhibition in the DNA
 741 extraction tests using genomic DNA and in the field collected eDNA
 742 samples was assessed with an Internal Positive Control (IPC) assay:
 743 forward primer (50 nM) 5'-GTATTCCTGGTTTGTAGGTTGAGC-3', reverse
 744 primer (900 nM) 5'-ATGTGACTGGACTTCCGTATCG-3' and probe (250 nM) 5'-
 745 Cy3-CGACGGCCAGTGAATTGTAATACGA-BHQ-2-3' (Carraro et al. 2018).

746 All samples were run diluted 1:2 with PCR-grade water and three qPCR
 747 replicates were run for each sample. Carraro et al. (2018) reported a LOQ
 748 of 5.229 copies and a LOD of 0.242 for *F. sultana*. eDNA from sites with
 749 confirmed presence of *F. sultana* delivered copy numbers from 257.804 to
 750 1250. In this study, fluorescence above the LOD but below LOQ was
 751 detected in 4% of negative controls. Subsequent SYBR Green based melt
 752 curve analysis and Sanger sequencing revealed this to be due to a chimeric
 753 primer dimer formation. Therefore, samples were only counted as positive
 754 if: 1) the copy number exceeded the LOQ, and 2) copy number above LOQ
 755 was detected in two of three technical replicate reactions for each of the
 756 field replicates. We observed some inhibition at very low copy numbers
 757 (see Supp. Fig. 1) and therefore decided to dilute all samples by 1:2 as the
 758 resulting better inhibition-free amplification performance outweighed the
 759 reduced DNA input.

760 *Parasite metabarcoding*

761 To reveal the full malacosporean species range in the river, metabarcoding
762 targeting myxozoan parasites was conducted with the primer panel
763 described in Lisnerova et al. 2023, 2025. The metabarcoding libraries were
764 prepared using a two-step PCR approach, with tailed amplicons indexed
765 using IDT Unique-Dual Indexes (10bp). Pre-amplification using ERIB
766 primers (Lisnerova et al. 2023) were conducted using 7.5µl of NEBNext
767 Ultra II Q5 Master Mix Taq (1X concentration), 0.45µl forward primer (1µM),
768 0.45µl reverse primer (1µM) and 4.6µl nuclease free H₂O. DNA extract (2µl)
769 was used in final 15µl reactions conducted in duplicate and pooled. In the
770 first library preparation PCR, tailed and pooled myxozoan specific primers
771 (Lisnerova et al. 2023) were used with 1µl PCR product from the ERIB PCR
772 and 5.6µl nuclease free H₂O. As per published protocol, a separate PCR
773 reaction was conducted for malacosporean primers (Lisnerova et al.
774 2025). Duplicate PCR products for each primer set (myxozoans and
775 malacosporeans) were pooled and purified using a two-sided Solid Phase
776 Reversible Immobilisation (SPRI) protocol using 0.5x ratio, followed by 1.0x
777 ratio. Index PCR master mix consisted of 7.5µl NEBNext Ultra II Q5 Master
778 Mix (1X concentration), 5.5µl nuclease free H₂O, 1µl IDT index primer mix
779 and 1µl cleaned PCR product per well. Amplification was conducted for 10
780 cycles, and samples were cleaned 0.8x SPRI beads. After quantification
781 with Qubit (Thermo Fisher Scientific Qubit 4 Fluorometer), samples were
782 normalised, and the pooled amplicons were polished by size selection to
783 100-600bp fraction on the BluePippin instrument (Sage Science Ltd). The
784 pooled libraries were sequenced in-house at DeepSeq, University of
785 Nottingham, on a 40% of an Illumina MiSeq 300bp paired-end run at 3:1
786 ratio of myxosporean to malacosporean library. The proportions reflected
787 expected species richness of each parasite group.

Results

788 *Testing eDNA capture methods*

789 The use of the combined small and large pore size filter doubled or tripled
790 the volume of water passed over the small pore size CN filters, whilst
791 having no impact for the 0.7µm GF filter (Fig. 4). Despite the higher volume
792 passed through the GF filter alone (433ml), and in combination with the
793 5µm CN filter (404ml), the DNA yield was lower. In terms of eDNA yield, the
794 best single filter size was the 0.45µm CN filter despite the much lower
795 water resulting volume compared to the 0.7µm GF filter (0.45µm CN: 14.6
796 compared to 0.7µm GF: 10.6ng/µl, Fig. 4b). The best combination was the
797 0.45µm CN filter with the addition of the 5µm filter, with increase in the
798 filtration volume (+259ml, +250%) and enhancing the probability of
799 capture of particles in the parasite spore size range.

800 DNA extraction method tests on homogenised *F. sultana* tissue revealed
801 consistent performance of Longmire's Buffer as a lysis reagent, and
802 indicated similar or enhanced performance to the proprietary commercial
803 product (Fig. 5). The addition of Longmire's in combination with ethanol
804 reduced the overall DNA yield slightly compared to Qiagen Buffer and
805 Longmire's only (Fig 5a), but in subsequent tests with qPCR of target *F.*
806 *sultana* DNA, the Cq values were lower in extractions with ethanol or
807 isopropanol added, indicating better amplification of target DNA (Fig. 5b).
808 Similarly, the addition of ethanol, as is done in the commercial kit protocol,
809 reduced level of inhibition when compared to Longmire's buffer added
810 without the ethanol (Fig. 5c). Here it is noted that inhibition was very low
811 throughout the experiment, and the variation observed in any extraction
812 method is unlikely to affect detectability.

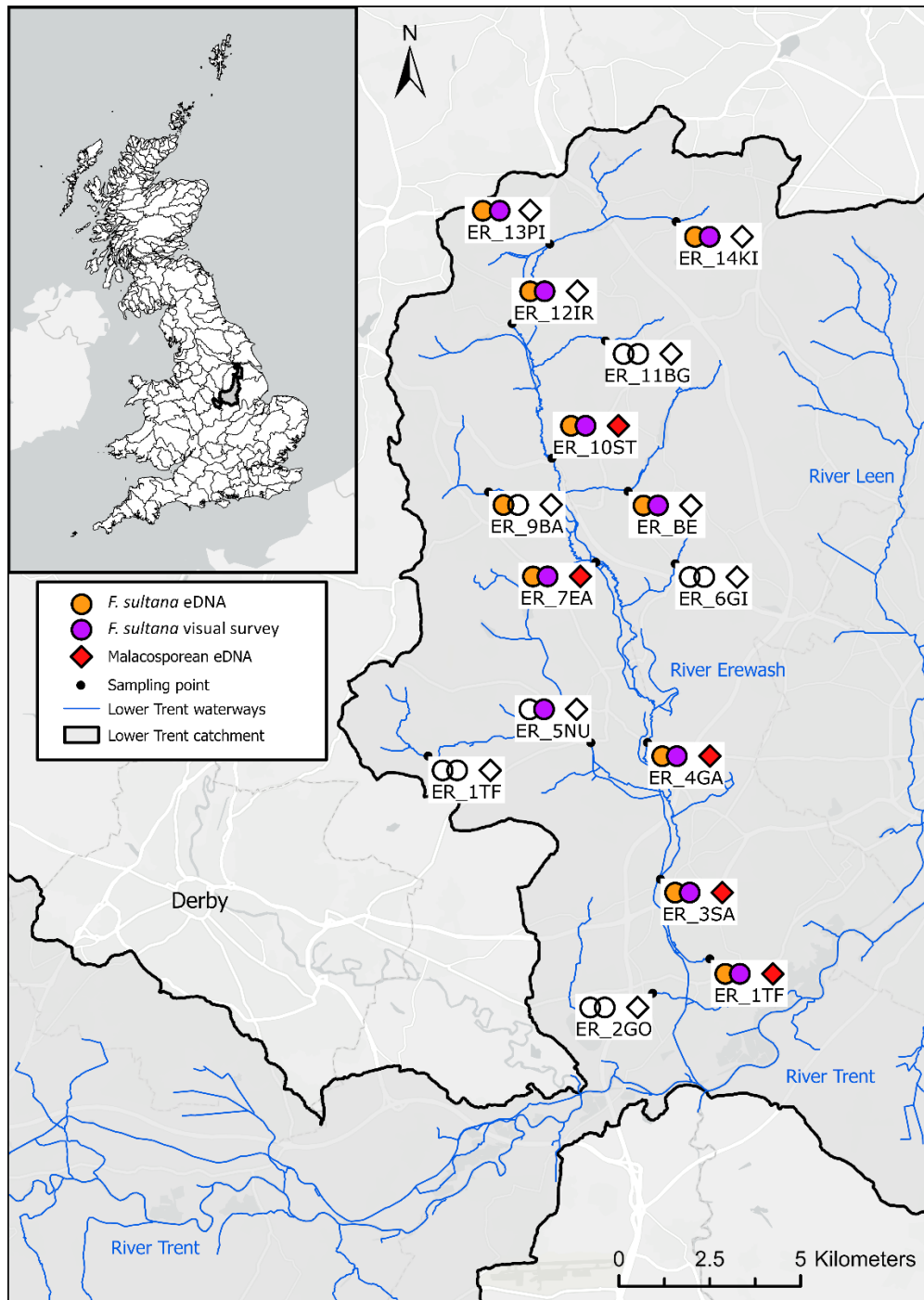


Figure 3: Map of the River Erewash sampling sites of eDNA from river water (black dot or pie) in the Lower Trent Catchment (black outline, waterways in blue). The circle and diamonds depict occurrence of *F. sultana* eDNA in qPCR (orange when positive), occurrence of *F. sultana* in visual wading surveys (purple when positive) and occurrence of malacosporeans in metabarcoding (red when positive).

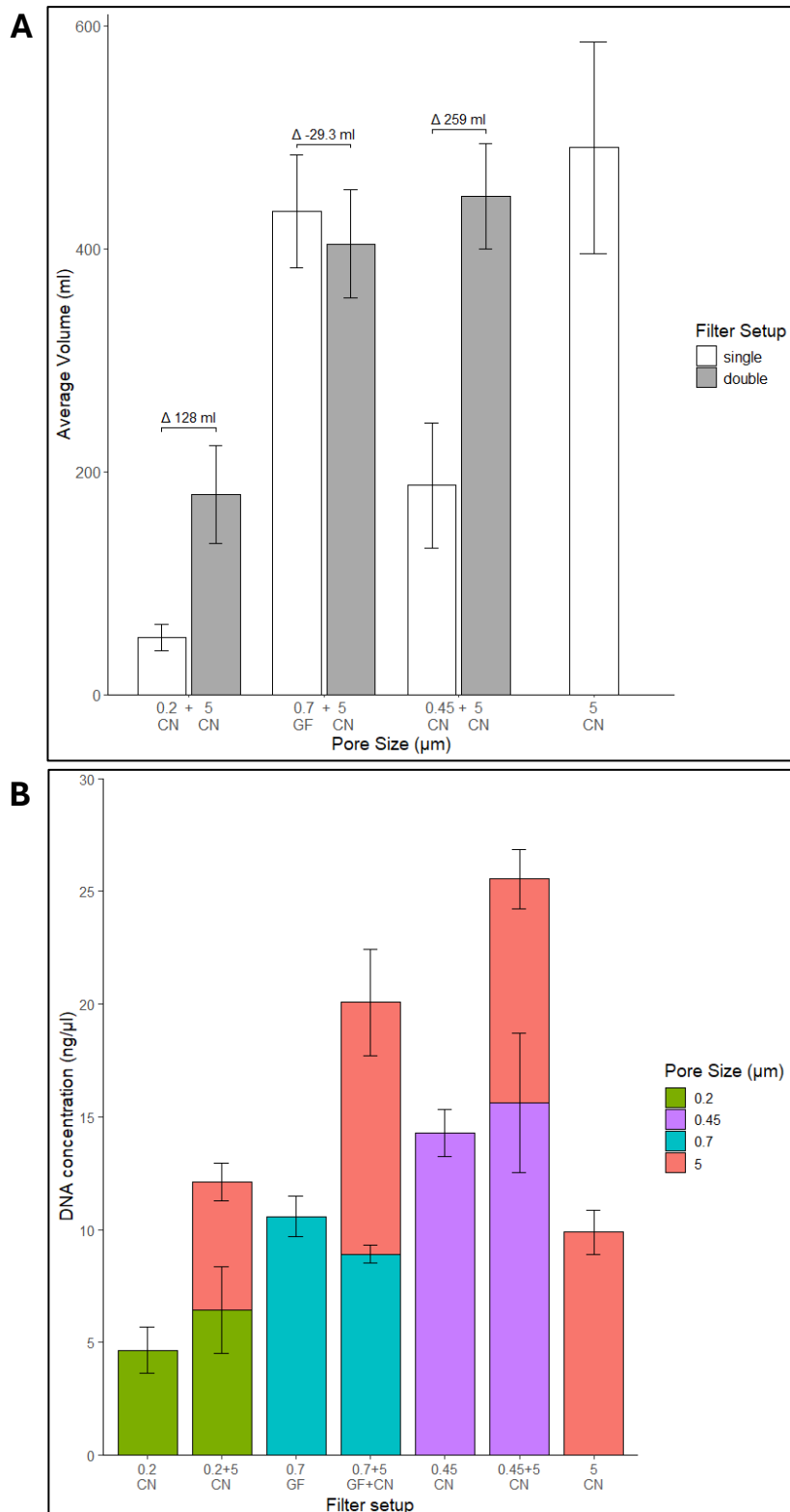


Figure 4: A) The volume of water that was possible to pass through filter, as a single filter alone (white bars), and as double filter (grey bars), where double filter consists of a 5μm filter in combination with a smaller pore size filter. B) DNA yields achieved after extraction for the different filter combinations. All filter papers were extracted and DNA concentration measured separately.

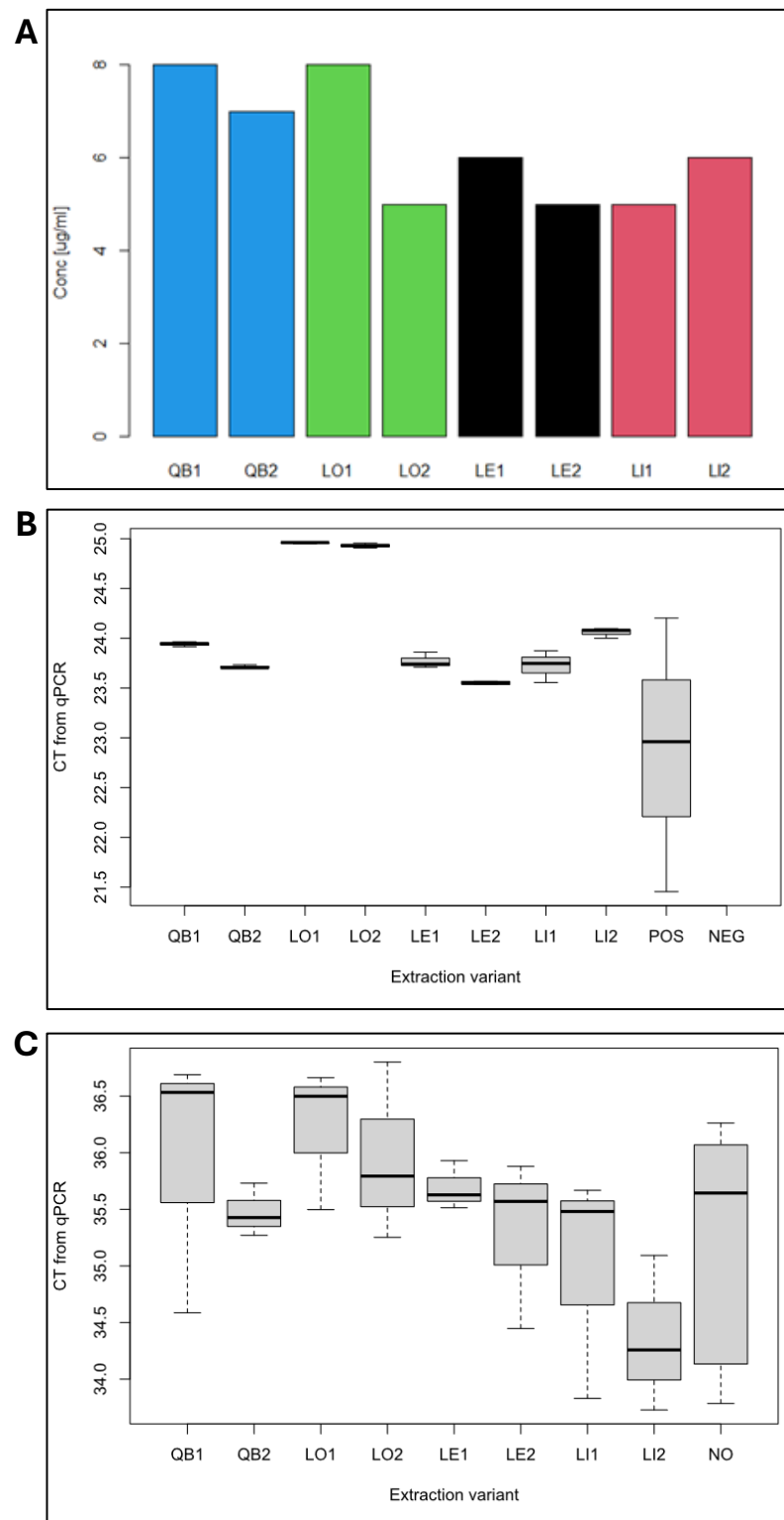


Figure 5: DNA extraction method optimisation, where QB = Qiagen buffer, LO = Longmire's only, LE= Longmire's and Ethanol (100%) and LI = Longmire's and isopropanol (40%). For each sample and method, plotted are A) DNA yield, B) qPCR CTvalue of *F. sultana* DNA, where POS refers to known positive DNA samples extracted previously, and NEG comprises water only, and C) CT value of internal positive control

molecule amplification. The reference value is indicated by NO, in these samples only water and target internal positive control were added.

813 *F. sultana* visual survey and eDNA detection

814 *F. sultana* eDNA was prevalent in the River Erewash, with 67% of samples
815 positive by qPCR (n=53, Fig. 3). Bryozoan detection frequency via eDNA
816 and wading agreed well (Fig. 5a). Of all the 15 sites surveyed with both
817 methods, 67% % (n = 10) were found to contain *F. sultana* colonies by
818 wading surveys. The accuracy of the eDNA method for detecting the
819 bryozoan presence in these ten confirmed sites was 90% over the two
820 sampling dates. In ER05 colonies were found in the wading survey, but not
821 in eDNA. In ER09 and ER14, *F. sultana* was detected in eDNA only in the
822 second sampling. The eDNA and wading surveys agreed on the absence of
823 *F. sultana* in four of five sites, with an eDNA detection without wading
824 survey detection in the fifth site (ER09, Fig. 3 and Table 1). Across all sites,
825 eDNA and wading surveys provided the same result on *F. sultana*
826 presence/absence in 86.7% of the sites. Calculated copy numbers in the
827 sampled water volume ranged from 31600 to 1003 copies/ml (Fig. 7a, Table
828 1). The highest concentration of *F. sultana* eDNA was detected in the upper
829 half of the river, upstream and including site ER_7. The smallest side
830 streams ER_02, ER_06, ER_XX and ER_11 harboured no bryozoans.
831 Significant correlation ($R^2 = 0.42$, $P = 0.005$) was found between visual
832 survey and eDNA copy number of *F. sultana* only for first sampling date (Fig.
833 7a) but not for the second sampling date (Fig. 7b).

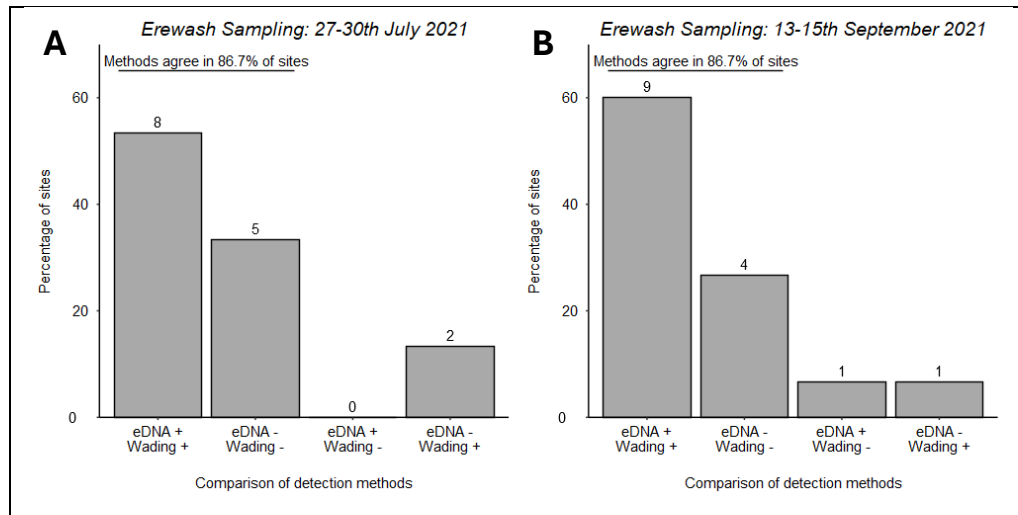


Figure 6: A) Comparison of qPCR detection based on eDNA and wading survey with detection (+) or absence (-). B) Detection of *F. sultana* (Fs) and *T. bryosalmonae* (Tb) from qPCR assays and other myxozoan parasites from metabarcoding on the same samples and sites.

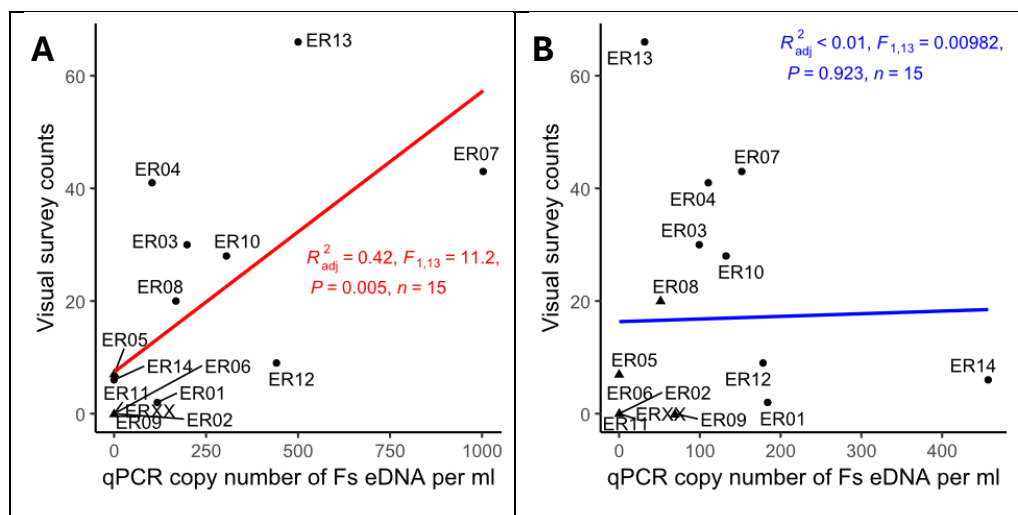


Figure 7: Correlation of *F. sultana* eDNA copy number and number of substrates with *F. sultana* colonies in the wading survey area in A) July and B) September. The eDNA sample was taken at the downstream portion of the wading survey area. Regression lines plotted in red or blue.

834 Apart from the most upstream sites, copy numbers tended to be greater
835 in the July, rather than September sampling (Fig. 8a & b), reflecting
836 seasonal growth and reproduction patterns of colonies.

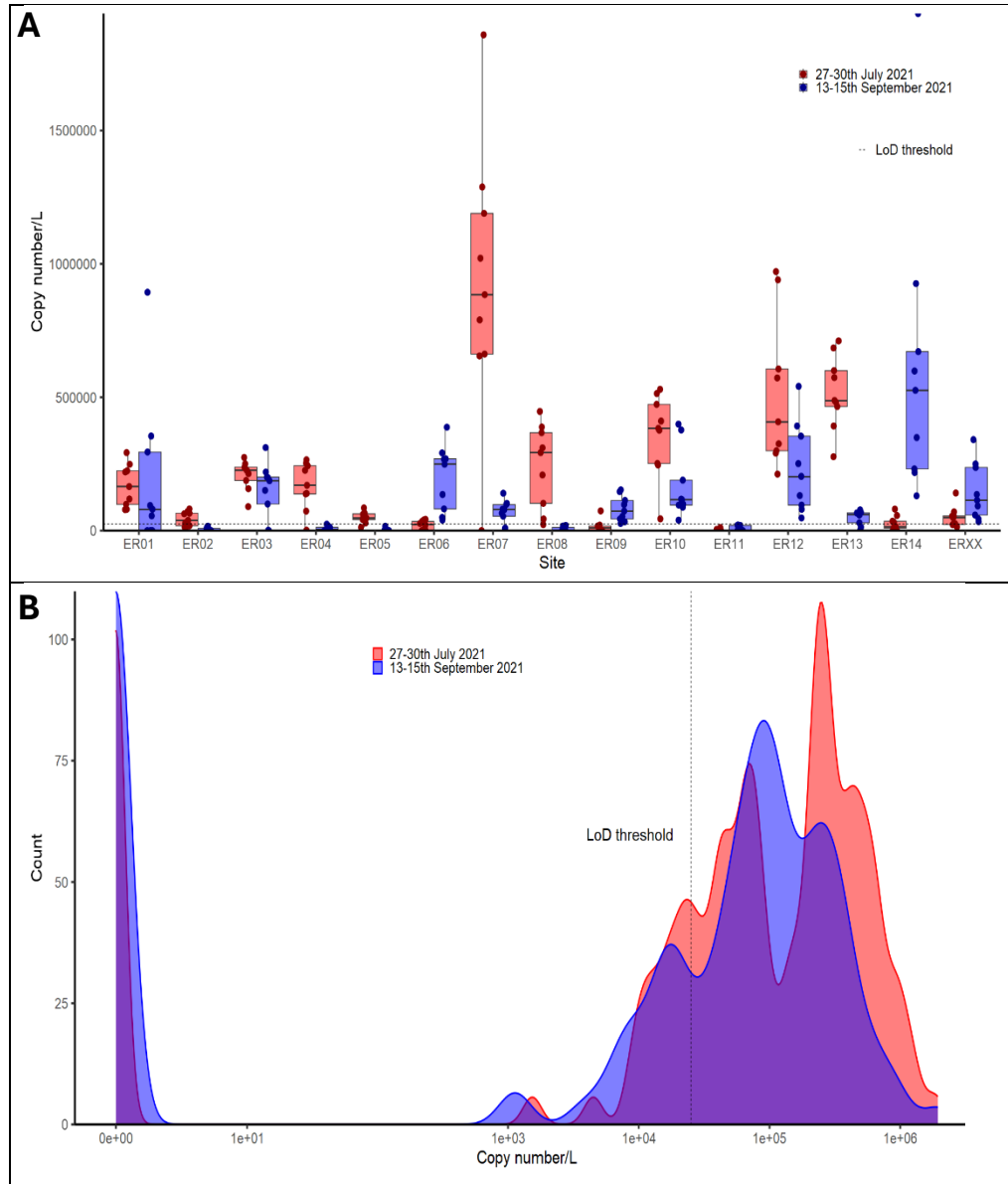


Figure 8: eDNA concentration of *F. sultana* as A) copy numbers per original sampling volume (Copy number/L) of both sampling dates on all sites before LOD filtering and B) in density as counts per eDNA concentration across all sites and dates.

Table 1: Number of surveyed *F. sultana* colonies (wading) and detected copy numbers after LOD filtering in the originally sampled volume (qPCR) on the sampling dates in the River Erewash.

Site Code	Stream type	Survey colony count 27-30 th July 2021	qPCR 27-30 th July 2021 (Copy/ml)	qPCR 13-15 th September 2021 (Copy/ml)
ER_01TF	Main	2	117.4	183.9
ER_02GO	Side	0	0.0	0.0
ER_03SA	Main	30	198.2	99.2
ER_04GA	Main	41	103.1	110.4
ER_05NU	Side	7	0.0	0.0
ER_06GI	Side	0	0.0	0.0
ER_07EA	Main	43	1003.1	151.9
ER_08BE	Side	20	168.0	51.1
ER_09BA	Side	0	0.0	69.2
ER_10ST	Main	28	305.3	132.3
ER_11BG	Side	0	0.0	0.0
ER_12IR	Main	9	441.2	178.4
ER_13PI	Main	66	500.0	31.6
ER_14KI	Main	6	0.0	457.0
ER_XXWH	Side	0	0.0	0.0

837 *Myxozoan parasite detection*

838 No *T. bryosalmonae* presence was detected in the R. Erewash in any of the
839 visually screened colonies or the eDNA samples in qPCR, suggesting the
840 absence of the parasite in the river (Fig. 9). Regardless, other
841 malacosporean parasites were detected in metabarcoding, including
842 *Buddenbrockia* sp. known to infect *F. sultana*. July sampling revealed

843 higher parasite occurrence in eDNA, with six sites, all with confirmed *F.*
 844 *sultana* presence, harbouring malacosporean parasites (Fig. 9a & b).

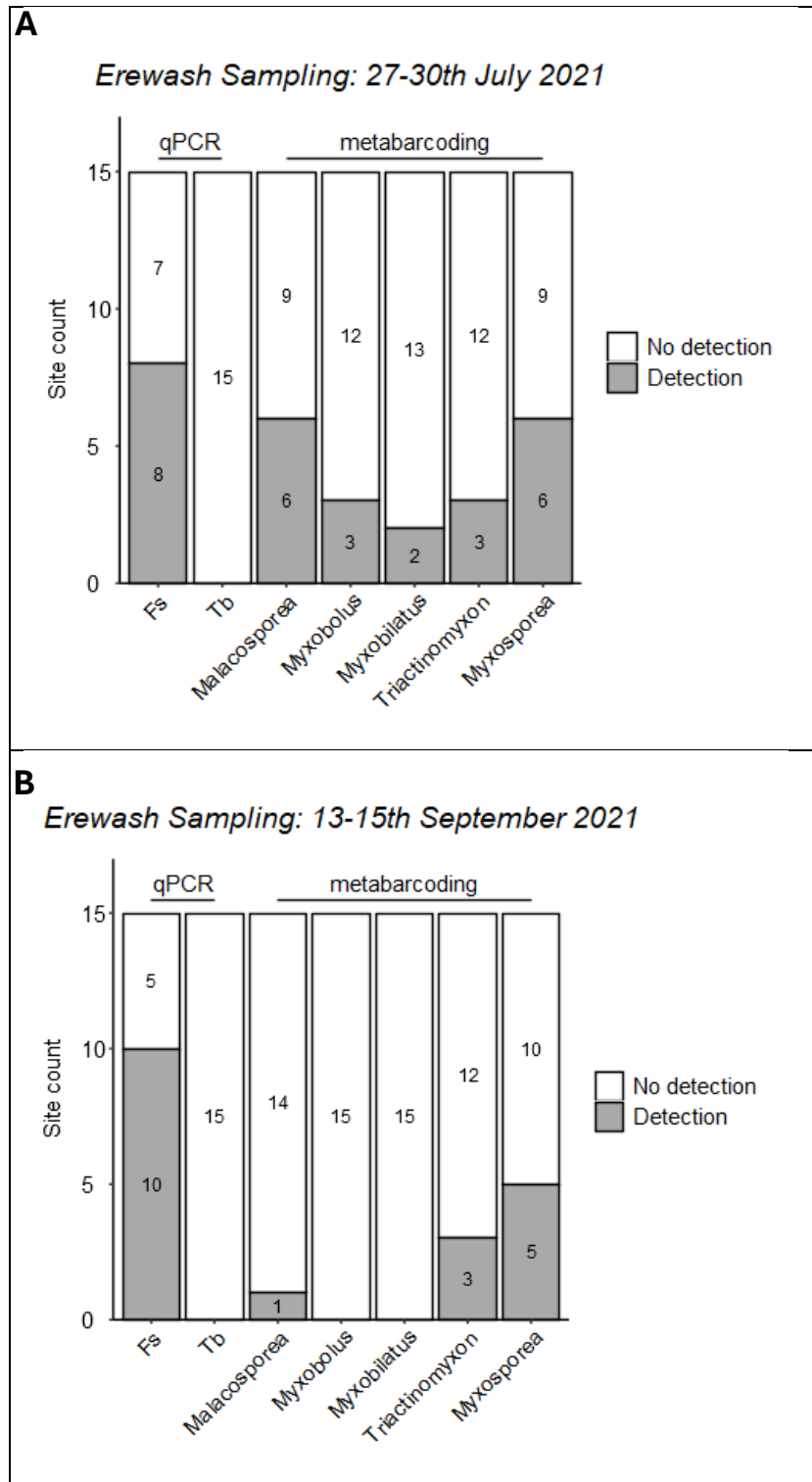


Figure 9: Proportion of sites positive for *F. sultana* (Fs), *T. bryosalmonae* (Tb) and other myxozoans detected via metabarcoding for two sampling periods, A) July and B) September.

845 *Buddenbrockia* sp1 (worm type) is an undescribed malacosporean known
 846 to infect *F. sultana* (Tops and Okamura 2003, Hartikainen et al. 2014) and
 847 found frequently in the metabarcoding. Highest levels of the
 848 *Buddenbrockia* sp1 were found on ER01, ER03, ER04 and ER10 with lower
 849 levels in ER07 (Fig 10). None were detected in the side streams and in the
 850 upstream sites (ER12, ER13, ER14). In all sites with *Buddenbrockia*, *F.*
 851 *sultana* was also found through qPCR and visual survey. Only the upstream
 852 sites and the side stream ER08 have bryozoans present but no detectable
 853 malacosporean parasites. Additionally, *Myxobolus*, *Myxobilatus*,
 854 *Triactinomyxon* and *Myxosporea* parasite species were detected in a
 855 number of sites: ER01 to ER04, ER12 and ER13.

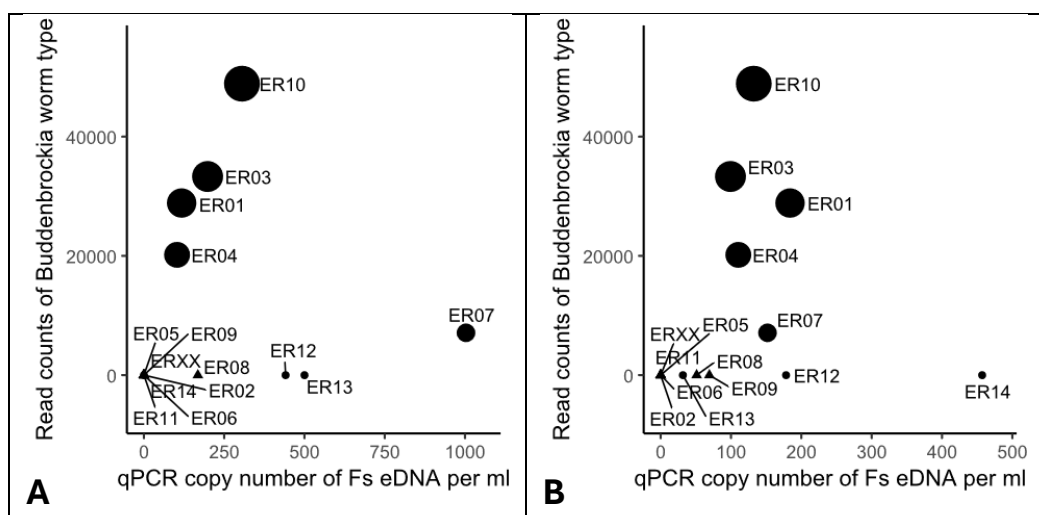


Figure 10: Comparison between *F. sultana* eDNA quantification by qPCR and read count of *Buddenbrockia* worm type, which is known to infect *F. sultana*, sampled during A) July and B) September.

Discussion

856 Myxozoan parasites of fish and bryozoans provide an opportunity to assess
 857 the co-occurrence patterns of hosts and parasites obtained from eDNA
 858 samples. The general assumptions about eDNA release, dispersal and
 859 decay have wider implications for capture of pathogen and host eDNA,
 860 however, current filtration methods may bias against either the host or
 861 parasite detection.

862 *eDNA capture for hosts and parasites*

863 The filtration setup presented here mitigates biases and provides an
864 integrated, “catch-all” sample, that is designed to capture across the
865 parasite and host eDNA particle size and state distribution (PSD). Filter size
866 and water volume choice are the two main considerations in eDNA
867 sampling setups. Filter size and sample volume can be optimised for single
868 species detection and adapted interchangeably along PSD capture
869 proportion equivalent as shown by Turner and Barnes et al. (2014).
870 However, multi-species capture approaches face a conundrum of
871 different, species-specific PSDs and resulting divergent filter size-capture
872 efficiencies relations. Theory and empirical evidence suggest that total
873 collectable eDNA is most abundant with a $<0.2\mu\text{m}$ filter size, mostly
874 attributed to DNA from bacteria. The rarer metazoan eDNA with a wider
875 PSD is best captured with pore sizes between $1\text{-}10\mu\text{m}$ (Turner et al. 2014,
876 Barnes et al. 2021, Zhao et al. 2021). Quantification via eDNA further relies
877 on capture of near all eDNA particles present, therefore, even for
878 metazoans requiring a small filter size ($<0.5\mu\text{m}$) and standardised sample
879 volume. To reach sufficient copy numbers to exceed LOQ, particularly for
880 rarer particles like parasite spores, substantially larger pore size filters are
881 required; e.g. myxozoan spores are generally $>10\mu\text{m}$ diameter, and many
882 other metazoan parasites are substantially larger (e.g. trematode
883 cercaria).

884 The filtration testing here revealed that the use of CN filter papers clearly
885 enhanced DNA yields post-extraction over glassfibre filters, despite similar
886 performance in filtered volume. The use of multiple filter papers and
887 preservation in the field often necessitates using larger preservative
888 volumes, and Longmires Buffer has been shown to provide good results
889 (Renshaw et al. 2015). Our modification presented here shows a
890 successful replacement of extraction buffers from commercial kits,
891 substantially easing the processing of larger volume samples, and
892 reducing costs of having to purchase additional reagents. We detected

893 optimal performance during extraction when using Longmires Buffer, with
894 added GuHCl and ethanol.

895 Our goal was to ensure capture of the bigger parasite spores by using a
896 bigger filter pore size and a higher water volume while still capturing a large
897 proportion of other eDNA particles on the smaller filter. The method of
898 using two separate filter cartridges allows separating the two different size
899 fractions if required, however, in this study during field sampling we opted
900 to combine both filter size in DNA extractions to reduce processing times
901 and effort. The separate filter cartridges further allow a different volume to
902 be filtered through each size if required, as this enable reaching the LODs
903 for rarer particles on the larger pore size papers. In this study, the volumes
904 were standardized for each filter size to allow direct between-sample
905 comparisons. This limits absolute comparisons of quantification (copy
906 number against water volume) to other eDNA studies and their sampling
907 methods. We found lower copy numbers (0.1 – 0.5 million copies/L)
908 compared Carraro et al. (2018) (above 6 million copies/L) while taking
909 250ml of river water on the 0.45 μ m and an additional 250ml on the 5 μ m
910 filter instead of 500ml on a single 0.45 μ m filter and using a different
911 extraction method (Qiagen Blood & Tissue versus PowerSoil DNA kit).

912 Thus far no agreed gold standard for eDNA sampling has been established,
913 with variation in methods concerning filter material, extraction method
914 and dealing with inhibitors (Renshaw et al. 2015, Spens et al. 2017, Hunter
915 et al. 2019). The Longmires Buffer based preservation and storage method
916 is ideally suited for remote and low-tech situations and is stable at ambient
917 temperatures for prolonged periods of time (Renshaw et al. 2015, Kumar et
918 al. 2020a). Due to the addition of glass beads prior to sample collection,
919 filter disruption and lysis can be conducted in the sample tube without a
920 need to transfer filter papers. The preservation buffer (Longmire's) doubles
921 as lysis buffer in DNA extraction and reduces volumes in handling and
922 processing. A modified version of the kit can be used to preserve the filter
923 paper directly in the filter holder by inserting the Longmires buffer into the

924 filter holder and sealing with caps. Good reusability and portability of our
925 eDNA kit design together with the omission of potentially harmful
926 chemicals (like ethanol or CTAB) in sample preservation make it ideal for
927 overseas collaboration, commercial application or citizen science.

928 *Patterns of parasite detection across a river network*

929 Despite abundant populations of *F. sultana* and known occurrence of
930 brown trout (*Salmo trutta*) in the River Erewash, *T. bryosalmonae* was not
931 detected on any site using qPCR, nor was it found in the metabarcoding.
932 This indicates its absence in the studied river, as significant co-occurrence
933 and co-detectability rates of hosts and parasites have been demonstrated
934 before (Duval et al. 2024). The Erewash is a heavily polluted river, that
935 receives run-off from several treated waste-water outlets along its length.
936 The nutrient levels in the river are high, and water levels fluctuate
937 seasonally. Although brown trout are recorded, their numbers are low, and
938 lack of suitable fish hosts may preclude wider distribution and detectable
939 abundance of *T. bryosalmonae* in these sites. The lack of detection of *T.*
940 *bryosalmonae* in both metabarcoding, bryozoan tissue and eDNA suggest
941 that eDNA sampling may be a convenient tool for identifying parasite-free
942 sites, that can be prioritised for conservation action, including habitat
943 improvement, stocking and rewilding of rivers to allow the return of
944 salmonids (Ros et al. 2022). Other small, parasite-free tributaries of
945 salmonid rivers have been demonstrated to have significantly better
946 spawning success of contribute to brown trout populations (Schmidt-
947 Posthaus et al. 2021). Barrier removal and movement of brown trout is
948 increasingly facilitated in the Trent catchment (for example as part of the
949 Trent Gateway project by the Environment Agency) and using eDNA to
950 monitor the spread and establishment of fish parasites may further inform
951 on how connectivity will modify disease risk in the system.

952 Other malacosporean parasites have not been investigated previously with
953 relation to bryozoan populations, and the metabarcoding approach
954 revealed strong association of known *F. sultana* infecting malacosporeans

955 with abundant bryozoan populations. Notably, particularly the sites on the
956 main river, discounting the most upstream sites, harboured more
957 malacosporeans, despite relatively lower eDNA copy number of bryozoans
958 in the most downstream sites. This may reflect either dilution effects
959 (downstream sites have more water flow and volume, without proportional
960 increase in the benthic habitat available for bryozoans), or alternatively the
961 larger availability of fish hosts in these sites, influencing the infection
962 prevalence.

963 Other myxozoan parasites were only detected in a limited number of sites
964 (up to 6). Parasite metabarcoding shows profound changes in detectable
965 parasite presence for most parasite groups between sampling dates. Taxa
966 assigned to triactinomyxon species were repeatably found across July and
967 September samplings, suggesting that the spore release from the hosts of
968 this parasite group are not limited to a single time point. The groups
969 Malacosporea, Myxobolus and Myxobilatus were almost exclusively
970 detected only on the first sampling date, suggesting distinct peaks in
971 detectable spore production. The detection of a range of myxozoans in the
972 eDNA samples suggests that the sampling method was successful in
973 capturing across a range of different parasite spores. Particularly the
974 correlation of detection of *F. sultana* and *Buddenbrockia* sp. suggests that
975 the sampling method was successful in capturing across a substantial
976 particle size range of host and parasite eDNA, whilst retaining the signal of
977 co-occurrence.

978 In future, metabarcoding of myxozoan communities on the different filter
979 sizes could provide further evidence for performance of the larger pore size
980 filter in concentrating parasite stages from the water column besides
981 facilitating easier higher volume sampling and ultimately an ideal pore size
982 to maximise capture of parasite spores. Such evidence would be important
983 to assess the statistical effect of rarity, and its effect on parasite detection
984 (Schuster et al. 2022). While *F. sultana* is detectable with an almost 90%
985 chance if upstream presence has been confirmed by survey, the detection

986 of malacosporeans is limited to the main channel of the river. Parasite
987 distributions, and consequently detectability, are a function of host
988 ecology, infection dynamics, host dependency and life cycle stage
989 (Oredalen et al. 2022, Ivers and Jha 2023). Our findings highlight the need
990 for higher and repeated sampling efforts to reliably detect parasites from
991 eDNA compared to more prevalent species (like *F. sultana*).

992 *Detection of bryozoan eDNA*

993 Detection of *F. sultana* in eDNA had a high success rate. Comparison of
994 wading survey and eDNA estimates the capability of eDNA compared to
995 more traditional sampling methods, which are highly labour intensive and
996 require experienced personnel. *F. sultana* was consistently detected in the
997 River Erewash on both sampling dates with the number of positive sites
998 differing slightly between the samplings (8 out of 15 and 10 out of 15 sites).
999 High congruency (above 85%) of survey and eDNA was found with both
1000 methods largely agreeing on presence (8 to 9 sites) and absence (4 to 5
1001 sites) with some ambiguity between survey and eDNA on two sites.
1002 Bryozoans release eDNA during their suspension feeding activities, likely
1003 associated with fecal eDNA and the survey do not reflect the same
1004 sampling space as eDNA is also brought in from upstream and the wading
1005 survey was conducted in a small and confined locality at the sampling site
1006 (Thalinger et al. 2021b, Wood et al. 2021a). Therefore, eDNA might be
1007 contributed to by many animals in considerable distance to the sampling
1008 point. The degree of eDNA released from colonies covered in the wading
1009 survey to the eDNA samples is unclear. *F. sultana* is widespread in the river
1010 with most sites showing *F. sultana* presence. A similar *F. sultana*
1011 distribution has been reported by Carraro et al. with the main reaches
1012 having the highest abundance (Carraro et al. 2017). By using the *F. sultana*
1013 distribution and copy number information, modelling the extent of *F.*
1014 *sultana* populations in the entire catchment area with habitat
1015 characteristics and flow profiles similar to the Erewash (e.g. the R. Leen,

1016 Wye, Derwent) may facilitate the mapping of host communities liable for
1017 parasite invasion in the future.

1018 Inhibition is a concern in many eDNA studies (King et al. 2009, McKee et al.
1019 2015, Hunter et al. 2019, Sidstedt et al. 2020). The IPC testing in this study
1020 showed low levels of inhibition (evidenced by non-significant shift in Cq
1021 between eDNA spiked and only IPC reactions), although dilution of
1022 samples 1:2 resulted in earlier Cq on average when using the target *F.*
1023 *sultana* assay. This can be observed when extracted eDNA yields are high,
1024 as might be with our method where more water can be processed. In other
1025 cases, dilution might have a detrimental effect, potentially forcing some
1026 samples with low copy number below the detection threshold. Therefore,
1027 checking for amplification success across a range of dilutions may be
1028 required even in the absence of detectable inhibition. Our results also
1029 highlight the benefit of using species-specific, highly sensitive methods
1030 alongside metabarcoding to find the optimal methods for processing and
1031 preparing samples. These optimisation steps are particularly beneficial
1032 when dealing with parasites or other challenging species.

1033 *Comparsion of techniques*

1034 We conducted triplicate replication in field and in laboratory to account for
1035 between sample and reaction variation. Further, we determined specific
1036 detection requirements and thresholds prior to sampling and repeated
1037 sampling on another timepoint to reduce the influence of weather and
1038 seasonality. We observed sporadic, low-level amplifications in negative
1039 controls and in otherwise negative samples. These were verified to be
1040 unspecific amplicons in melt curve or gel electrophoresis analysis.
1041 Carraro et al. (2018) showed that the qPCR assays can reliably detect less
1042 than one copy (given enough replicates) meaning that the sensitivity is
1043 easily low enough to detect true target DNA presence in most if not all
1044 replicates. To avoid false positives impacting the detection rates we
1045 implemented a rigorous detection limit: two of three reactions should
1046 show amplification to be considered positive. This allows for mistakes

1047 (error in pipetting, sample mix-ups or contamination across wells) in the
1048 sampling and laboratory process but ensures that low-rate sample cross-
1049 contaminations or unspecific amplifications do not influence detections.
1050 Since I observed high copy numbers (>100 copies/ml) in all sites where I
1051 detected *F. sultana* in wading surveys, we additionally implemented the
1052 LOQ as the copy number limit for positive detections. This ensured that any
1053 amplicons are not the result of mistakes or contamination and allowed
1054 gating out any potential non-target amplicons but may have resulted in a
1055 small number of false negative detections.

1056 It is essential to establish good standards for eDNA practices to increase
1057 the reliability and comparability of findings through eDNA (Schenekar
1058 2022). Efforts have been made to evaluate and standardize common
1059 practices used in sampling, extracting, sequencing and analysing (Geerts
1060 et al. 2018, Shu et al. 2020). Rigor similar to protocols used in ancient DNA
1061 (aDNA) are frequently demanded, particularly when working with
1062 parasites, the detection of which may have biosecurity and farming related
1063 implications (Sieber et al. 2020, Thalinger et al. 2021a). The inclusion of
1064 field and experimental negatives is seen as essential to ensure the validity
1065 of the sensitive processes (Bustin 2010, Pinfield et al. 2019, Griffin et al.
1066 2020).

Conclusion

1067 Our detection and quantification approach of host and parasite eDNA from
1068 a wide particle size spectrum offers new ways to fine tune pathogen
1069 diagnostics by specifically maximising capture and detection efficiency.
1070 Better access to the full eDNA particle size range may further advance the
1071 understanding the “ecology” of eDNA, which is largely unknown for most
1072 species. Without consideration of the various sources of eDNA, biases
1073 may be present in almost all stages of eDNA studies, with major influences
1074 (Elbrecht and Leese 2015, Krehenwinkel et al. 2017, Loos and Nijland
1075 2020, Martins et al. 2020). With full eDNA particle size spectrum
1076 assessment, it may be possible to better control for nestedness and site

1077 connectivity (Perry et al. 2024) and ultimately set ecology and disease (or
1078 parasite infection) dynamics in a broader context (Carraro et al. 2020b,
1079 Carraro et al. 2023, Carraro and Altermatt 2024). Capturing whole
1080 parasites, or potentially small dispersive stages/larger cell clusters of free-
1081 living animals have a higher probability of delivering information on allele
1082 frequencies and haplotypes. Information on intraspecies diversity could
1083 unveil more detailed population structures and the interaction between
1084 ecological communities and species distribution patterns (Parsons et al.
1085 2018, Sigsgaard et al. 2020).

1086 We highlight a highly sensitive and robust baseline sampling and detection
1087 approach and apply the obtained samples in single species eDNA
1088 quantification and parasite community metabarcoding. eDNA and its
1089 characteristics promise to be useful for developing occupancy models and
1090 population networks. The efficient, broad spectrum sampling design
1091 developed here will assist in employing eDNA-specific statistic and
1092 modelling approaches in the context of host-parasite relationships.

Chapter 2: *Rapid direct sequencing, de-novo assembly (and annotation) of complete mitogenomes using long read technology with application in the freshwater bryozoan *Fredericella sultana**

KEYWORDS: bryozoan, direct sequencing, de-novo assembly, mitogenome, genetic reference

Abstract

1093 Comprehensive genetic reference data remain scarce across much of the
1094 tree of life, including many well-studied taxa and particularly among basal
1095 metazoans and lower bilaterians. Robust genomic resources are essential
1096 for advancing research in evolutionary biology and ecology, particularly in
1097 the context of nucleic acid-based biomonitoring methods that assess
1098 species occurrence and community composition. However, the
1099 effectiveness of species-specific detection and metabarcoding from
1100 pooled or environmental samples is often hindered by incomplete
1101 reference databases. Although mitochondrial genomes (mitogenomes)
1102 are commonly used for detection assays, DNA barcoding and population
1103 genetics, studies often rely on short, conserved regions like rRNA genes or
1104 cytochrome oxidase subunit I (COI). Recent advances in Oxford Nanopore
1105 sequencing technologies have enabled long-read, high-fidelity direct
1106 sequencing at unprecedented speed and cost-efficiency. Combined with
1107 the development of optimised bioinformatics tools, these innovations
1108 facilitate the assembly of complete mitogenomes from minimal and often
1109 crude input material. Here, we present a universal, streamlined workflow
1110 for sequencing and *de-novo* assembly of mitogenomes, with step-by-step
1111 instructions including a bioinformatics pipeline. We applied this workflow

1112 to assemble and annotate a complete mitogenome of the freshwater
1113 bryozoan *Fredericella sultana* (Class: Phylactolaemata), a known host of
1114 the proliferate kidney disease parasite *Tetracapsuloides bryosalmonae*.
1115 The combined approach yields robust-coverage assemblies which
1116 evidence expected gene order and accurate resolution across repetitive
1117 mitogenome regions. We further explore mitogenomic features such as
1118 relative expression patterns, putative regulatory non-coding regions and
1119 intra-species variability. Notably, we report the presence of a group II intron
1120 within the mitogenome – only the fourth time such instance is documented
1121 in bilaterians, and the first reported in freshwater bryozoans.

Introduction

1122 Nucleic acid-based approaches using mixed, pooled or environmental
1123 samples are increasingly employed in ecological studies, particularly for
1124 biomonitoring, management and conservation purposes. Species-specific
1125 detection methods such as quantitative or digital droplet PCR
1126 (qPCR/ddPCR) are commonly used to identify and quantify taxa of interest
1127 (Agersnap et al. 2017, Carraro et al. 2017, Baker et al. 2018). These assays
1128 are especially critical for the surveillance of pathogenic, invasive or
1129 endangered species, where accurate detection can inform conservation
1130 and management decisions (Jerde et al. 2011, Hartikainen et al. 2016,
1131 Gorički et al. 2017, Uthicke et al. 2018).

1132 With increasing accessibility to massively parallel sequencing
1133 technologies, next-generation sequencing (NGS) platforms are now widely
1134 used for species identification and metabarcoding (Elbrecht et al. 2017,
1135 Corse et al. 2019, Lisnerová et al. 2023). Much of this work focuses on
1136 species that are rare, invasive, keystone or otherwise ecologically
1137 significant (Gorički et al. 2017, Dufresnes et al. 2019, Marshall and Stepien
1138 2019, Stepien et al. 2019, Crane et al. 2021). There is a growing interest in
1139 expanding these efforts into population genomics and meta-ecology to

1140 inform models of species distribution, gene flow, community connectivity
1141 and ecosystem function (Sigsgaard et al. 2017a, Uchii et al. 2017a, Adams
1142 et al. 2019, Sigsgaard et al. 2020, Tsuji et al. 2020, Zizka et al. 2020, Dugal
1143 et al. 2021).

1144 As the field evolves, there is a shift from short amplicon-based markers to
1145 longer fragments or even complete mitogenomes (Deiner et al. 2017,
1146 Doorenspleet et al. 2021, Jensen et al. 2021, Székely et al. 2021, Farrell et
1147 al. 2022, Urban et al. 2022). However, the effectiveness of detection,
1148 identification and intraspecific analyses remains constrained by the
1149 limited availability of sequence reference data, particularly among non-
1150 vertebrate taxa (Elbrecht et al. 2017). Molecular studies of invertebrates
1151 often rely on conserved mitochondrial markers such as COI or 16S/12S
1152 rRNA, despite their known limitations in resolving species-level
1153 relationships in several groups, including cnidarians and bryozoans (Stat
1154 et al. 2017, Elbrecht et al. 2018, Yoshitake et al. 2019).

1155 Vast amounts of barcoding data are being generated for invertebrates,
1156 mostly focussing on the COI gene – the “barcoding” region (Elbrecht and
1157 Leese 2017, Elbrecht et al. 2019, Giebner et al. 2020, Hintikka et al. 2022).
1158 Broadening the use of full mitogenomes allows for improved assessment
1159 of mitochondrial variability across intra- and interspecific levels. This can
1160 provide insights into mutation rates, haplotype diversity, gene flow and
1161 evolutionary dynamics. A comprehensive mitochondrial reference dataset
1162 facilitates such analyses and supports the integration of genetic data into
1163 ecological and conservation frameworks.

1164 Freshwater bryozoans, particularly those in the Class Phylactolaemata,
1165 represent a group of ecologically relevant but genomically under
1166 characterised invertebrates. These sessile, colonial suspension-feeders
1167 inhabit a wide range of lentic and lotic freshwater ecosystems and include
1168 several cosmopolitan species. Notably, *Fredericella sultana*
1169 (Blumenbach, 1779) (Fs) serves as the primary host of *Tetracapsuloides*
1170 *bryosalmonae*, the causative agent of Proliferative Kidney Disease (PKD) of

1171 salmonid fish (Clifton-Hadley et al. 1984, Anderson et al. 1999, Feist et al.
1172 2001, Okamura and Wood 2002). PKD is associated with increased
1173 mortality and threatens wild and farmed fish stocks, with bryozoans as a
1174 reservoir of infections into salmonid fish. Host-parasite coevolution and
1175 local adaptation in bryozoan populations are thought to be intrinsically
1176 linked to PKD emergence (Carraro et al. 2018).

1177 Developing robust molecular tools for accurate species identification and
1178 population level analyses in bryozoans is essential for understanding their
1179 ecological role, dispersal dynamics and host-parasite interactions. High
1180 resolution mitochondrial data enable the study of haplotype level gene
1181 flow, founder effects, population turnover and spatial structure, from
1182 individual colonies to entire catchments and biogeographic regions.

1183 Here we present a streamlined protocol for rapid mitogenome sequencing,
1184 assembly and annotation using Oxford Nanopore technology. The
1185 workflow, optimised for small invertebrate tissue samples, enables the
1186 generation of a complete, annotated mitogenome within a single working
1187 day. When multiplexed, it can yield one whole mitogenome per hour. We
1188 apply this method to generate the first complete mitochondrial genome of
1189 *F. sultana*, identifying coding and non-coding regions, gene order, and
1190 structural variation across individuals. Key steps of the pipeline are
1191 outlined in Figure 1, and detailed instructions are provided as a
1192 Supplementary File for community use.

1193 This approach leverages plasmid extraction protocols to enrich
1194 mitochondrial DNA and deplete nuclear genomic content, improving
1195 sequencing efficiency. Compared to short-read platforms, long-read
1196 sequencing simplifies de novo assembly, particularly in resolving repetitive
1197 and duplicated regions. The amplification-free protocol is fast, cost-
1198 effective, and suitable for barcoding multiple samples concurrently.
1199 Moreover, it opens avenues for the simultaneous analysis of associated
1200 symbionts or microbiota, and potentially even epigenetic modifications.
1201 While the method requires relatively high input DNA concentrations

1202 (>20ng/μl), successful long-read generation is achievable from total inputs
1203 of as low as 50-100ng, depending on DNA quality.

Methods

1204 *Optimising mitochondrial DNA extraction for speed and yield*

1205 A plasmid miniprep kit (Monarch® Spin Plasmid Miniprep Kit) was used to
1206 quickly and efficiently extract DNA from small amounts of invertebrate
1207 tissue. Previous studies have shown that minipreps or plasmid-targeting
1208 DNA extraction methods can enrich mitochondrial DNA over nuclear DNA
1209 (Defontaine et al. 1991, Peloquin et al. 1993). These kits are designed to
1210 capture bacterial plasmids, with the added benefit of being quick, simple,
1211 and size limiting (< 25kb) (e.g QIAprep Spin Miniprep Kit (Cat. No. 27104,
1212 Qiagen, Germany and Monarch Spin Plasmid Miniprep Kit (Cat. No.
1213 T1110S, New England Biolabs, USA). The circular and supercoiled
1214 mitogenomes have similar properties to plasmids, including similar size
1215 range, with most metazoan mitogenomes being 12-25kb long. Therefore,
1216 the size limitation properties of the plasmid prep kits favour the retention
1217 of metazoan mtDNA, whilst limiting the extraction of larger, unsheared
1218 chromosomal fragments. Recent results have indicated that plasmid
1219 minipreps outperform other mtDNA targeting extraction methods and
1220 produce sequencing ready DNA (Quispe-Tintaya et al. 2013, 2015,
1221 Mascolo et al. 2019). To our knowledge this study is the first to attempt this
1222 amplification-free, target enriching approach to generate DNA samples for
1223 direct sequencing of the mitogenome in invertebrates, where tissue
1224 material availability and purity is a complicating factor.

1225 The potential failure-points of the plasmid prep method for invertebrate
1226 mtDNA are a differential lysis of eukaryotic versus bacterial cells, and the
1227 much larger amount of nuclear DNA present in eukaryotic cells, potentially
1228 overwhelming lysis and the extraction column membrane binding
1229 capacity. To determine the fastest and most generally applicable lysis
1230 method, tissues of three aquatic species were used, and three pre-lysis

1231 treatments employed. The species used were *F. sultana* (with chitinous
1232 ectocyst and resistant to tissue disruption), *Lumbriculus variegatus* (easily
1233 disrupted, soft, highly polyploid oligochaete worm) and a vertebrate
1234 control, *Oncorhynchus mykiss* (rainbow trout, section of purely muscle
1235 tissue). Input volumes for each were roughly around 100mg which
1236 represents a pin head size amount: for *F. sultana* this was a ~1cm branch
1237 with 2-3 zooids, for *L. variegatus* a whole worm (1-1.5cm) and for *O. mykiss*
1238 a ~2mm² cube of tissue.

1239 The tested pre-lysis treatments were (Fig. 1a & 2): 1.) Mechanical
1240 disruption of tissue through grinding with a plastic pestle in a 1.5ml tube
1241 for around 10-20mg of tissue in resuspension buffer. 2.) Freezing (-80C)
1242 with and without addition of resuspension buffer and disruption with
1243 pestle. 3.) Digestion in Longmire Buffer and Proteinase K at 56°C for 30min
1244 after disruption with pestle. The success of the lysis was assessed visually
1245 and the amount of tissue remaining post-lysis was evaluated. After testing,
1246 the Longmire and Prot K pre-lysis was selected, as this achieved the most
1247 consistent yields. The final protocol included tissue placed in 180µl
1248 Longmire's buffer in 1.5ml tube and 20µl Proteinase K solution (20mg/ml
1249 or 600 U/ml), pestle ground against the side wall of the tube and incubated
1250 for 30min at 56°C to break down the tissue. Post-lysis, the extraction can
1251 be conducted with any plasmid isolation kit, with small modification
1252 detailed in the Step-by-Step pipeline to increase yield.

1253 We tested if linear genomic material of the nuclear chromosome can be
1254 depleted using the Exonuclease V (RecBCD) (Cat. No. M0345S, New
1255 England Biolabs, USA) leaving the intact, circular mitogenome behind
1256 (Ramón-Laca et al. 2023). However, total DNA amount was greatly reduced
1257 after digestion (see Fig. 2), and we did not use this approach in the
1258 subsequent library preparation. Reasons might be 1) star activity of the
1259 enzyme removing not only linear molecules, 2) extracted mitogenomes are
1260 nicked or linear after extraction or 3) mitogenome content is low compared
1261 to the total DNA amount.

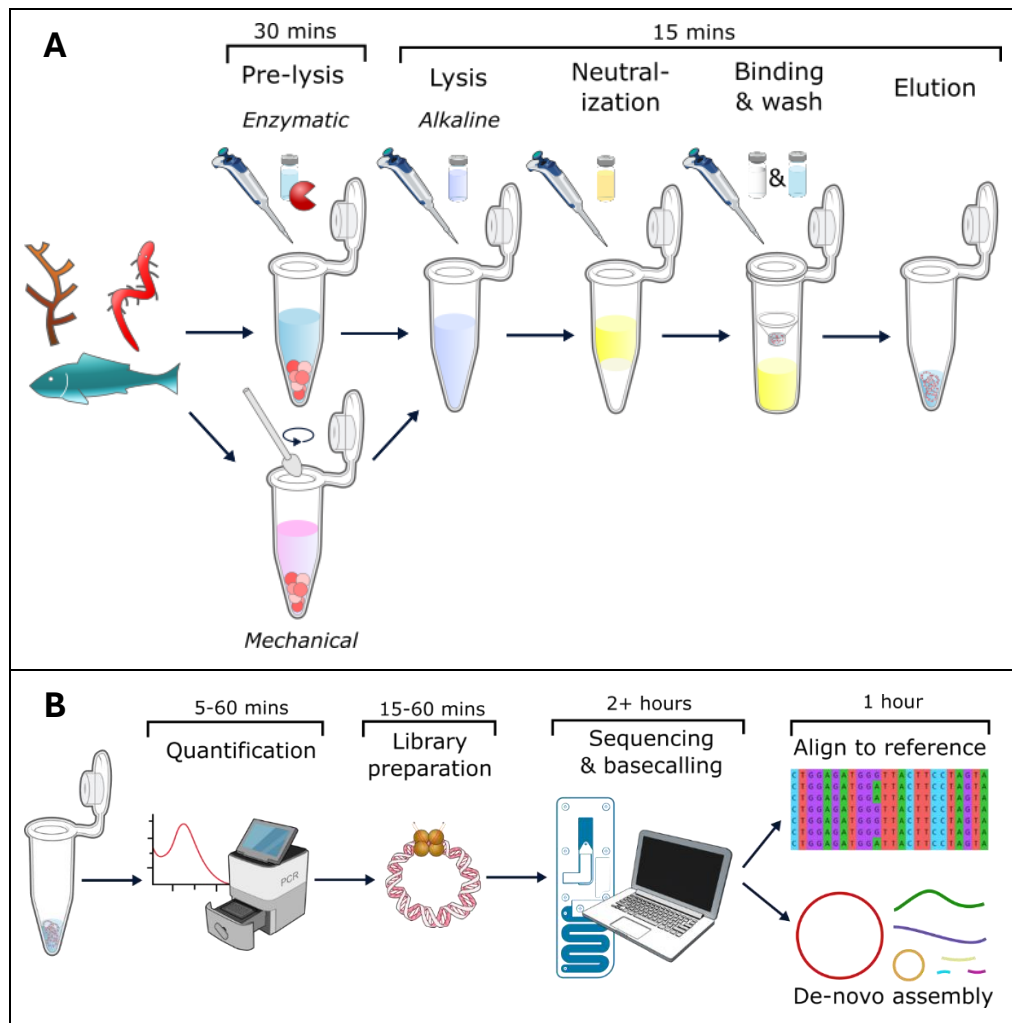


Figure 1: A) Miniprep for plasmid-style extraction for (potential) mitoenrichment from metazoan/animal tissue with pre-lysis tissue disruption (enzymatic or mechanical) before alkaline lysis, neutralization, column binding, washing and final elution. B) Sequencing workflow with initial quantification and quality control, library preparation with the Nanopore Rapid Barcoding or Sequencing kit, sequencing on V14 flow cell and basecalling, followed by either alignment to reference genome or de-novo assembly (with fly).

1262 *Sample collection and sequencing*

1263 Individual *F. sultana* colonies were sampled from submerged substrates
 1264 on four sites in the River Erewash, on 8th July 2024 (see Table 1).
 1265 Oligochaetes were taken from substrate on the same sites, while rainbow
 1266 trout tissue was taken from commercially available fish. 44 single
 1267 branches (around 1cm-5cm length) of colonies were processed for DNA
 1268 extraction. Only complete, unbroken and connected branches without
 1269 sponge, algae or other growth were prepared to maximise purity and to
 1270 ensure single-individual origin.

1271 Samples were checked for sufficient concentration (and purity) using NEB
 1272 Luna Universal qPCR master mix (New England Biolabs, USA) with a 537bp
 1273 mitochondrial nad3 marker (1µM F primer (5'-3'):
 1274 GGCCCTTGTAAGTCAAGTGA, 1µM R primer (5'-3'):
 1275 ACTCCTTGCCAGCTCTTTGT) on Applied Biosciences QuantStudio 3
 1276 (ThermoFisher Scientific, USA), using the standard fast setting for a melt
 1277 curve (initial hold: 95°C for 60s and 45 cycle of 95°C denaturation for 15s
 1278 and 60°C annealing and extension for 60s). The 24 samples with the lowest
 1279 Cq values and a single peak (at 82.5-83.1°C) in melting curve analysis were
 1280 selected for mitochondrial DNA Nanopore sequencing (Fig. 1b). Samples
 1281 that did not meet the required concentration for library preparation were
 1282 cleaned and concentrated using AMPure XP Beads (Beckman Coulter
 1283 United Kingdom, UK). The final input for the library preparation ranged from
 1284 23ng to 50ng and sequencing was carried out on a MinION flow cell (R10.4
 1285 chemistry: Oxford Nanopore Technologies, UK) with 800ng of library for
 1286 44h (for library preparation details, see Step-by-Step guide, Fig. 1b).

Table 1: Sampling sites on the River Erewash, Derbyshire/Nottinghamshire, England, United Kingdom on 8th July 2024 with geographic information and samples sequenced. Nr. = number

Site Code	Site Nr.	Site Name	Latitude	Longitude	Nr. Sequenced
ER_3SA	3	Sandiacre	52.9256233	-1.28590382	7

ER_7EA	7	Eastwood	53.00415799	-1.31115581	5
ER_10ST	10	Stoney	53.02996834	-1.32875258	6
ER_12IR	12	Ironville	53.0633666	-1.34455135	6

1287 **Mitogenome assembly, annotation and variant calling**

1288 The sequenced reads were basecalled (Fig. 1b) using dorado basecaller
1289 (v0.7.0, <https://github.com/nanoporetech/dorado/releases/tag/v0.7.0>).
1290 After testing, de-novo assembly was most successful with Flye 2.9.5
1291 (Kolmogorov, Yuan et al. 2019, Kolmogorov, Bickhart et al. 2020,
1292 <https://github.com/mikolmogorov/Flye/tree/flye>) assembler, generally
1293 suggested for use with long-read data and able to handle metagenomic
1294 inputs. Identification of mitochondrial scaffolds relies on blast searches
1295 against GenBank databases, and strategies for optimisation of search
1296 parameters in suggested in the Step-by-Step protocol. The mitochondrial
1297 genome was annotated using MITOS2 2.1.9 on galaxy.eu (Donath et al.
1298 2019). Repeats were identified with equicktandem
1299 (<https://www.bioinformatics.nl/cgi-bin/emboss/equicktandem>).
1300 Methylation data can be generated through modkit pileup v0.3.1 (Oxford
1301 Nanopore Technologies: <https://github.com/nanoporetech/modkit>).
1302 samtools (<https://www.htslib.org/>) (Danecek et al. 2021) were used to
1303 calculate coverages and consensus sequences, with variant calling on
1304 freebayes 1.3.8 (<https://github.com/freebayes/freebayes>). R code, scripts
1305 and online tools and settings are detailed in the Step-by-Step protocol to
1306 enable a user to quickly generate the results we illustrate below.

1307 To assess the quality of the compiled mitogenomes, RNA-seq data
1308 available in public databases (SRS6620694, SRS6620693, SRS6620692,
1309 SRS11894130, SRS11894127, SRS11894116, SRS11894105,
1310 SRS4788304, SRS4788303, SRS4788302) and unpublished read sets were
1311 used. Variant calling on aligned transcriptomic data can uncover highly
1312 expressed and variable regions useful for detection, quantification and

1313 population genetics based on DNA or RNA. RNA-seq data was aligned to
1314 the mitogenome assembly to check if the expression levels align with the
1315 MITOS2 annotations. The polished complete mitogenome assembly and
1316 the 18S and 28S rRNA region served as reference. The RNAseq mapping
1317 methods are detailed in the step-by-step guide, and alternative methods
1318 of polishing the mitogenome if short-read data are available are provided.
1319 To assess performance of our method we additionally compared it against
1320 existing short read (125bp) MiSeq and older (R9 chemistry) Nanopore (N50
1321 3000bp) data (Hanna Hartikainen, unpublished) in de-novo assembly and
1322 annotation (Fig. 5). Additionally, congruence with the published
1323 mitogenome sequence and annotation of two related Phylactolaemata
1324 species was checked: *Cristatella mucedo* (CM019737.2) and *Pectinatella*
1325 *magnifica* (OR460713.1, MG546680.1, NC_038192.1).

1326 We calculated counts and ratios of pyrimidine (cytosine and thymine) and
1327 purines (adenine and guanine) to identify the heavy (H or -) and light (L or
1328 +) strands of the mitogenome. Using gene annotations and strand-specific
1329 RNAseq data from mRNA transcripts, we conducted a DNA walk to
1330 investigate gene order and strand orientation. In this pyrimidine-purine
1331 walk, each pyrimidine (C or T) was assigned a step -1 and each purine (A or
1332 G) a step of +1, producing a cumulative plot that reflects strand asymmetry
1333 across the mitogenome.

1334 Unexpected transcriptomic expression between the *cox1* and *atp8* genes
1335 led to exploration of potential ORFs in this unannotated, 'non-coding'
1336 region. ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) was used to
1337 assess locations of potential ORFS. The amino acid sequences of
1338 identified ORFs were used to predict protein structure with ColabFold
1339 v1.5.5 (AlphaFold2 using MMseqs2) (Mirdita et al. 2022) and default
1340 settings. For sequence and structural comparison, the RCSB Protein Data
1341 Bank search (Berman 2000) and its pairwise structure alignment tool
1342 ([rcsb.org/alignment](https://www.rcsb.org/alignment)) with the flexible alignment methods jFATCAT and jCE-
1343 CP were used (Bittrich et al. 2024).

Results

1344 *Mitochondrial DNA extraction*

1345 Mechanical disruption with pestle resulted in complete tissue lysis for *L.*
 1346 *variegatus* worms but achieved only limited effectiveness in salmon and
 1347 bryozoans (Table 2 and Fig. 2). Freezing resulted in poor lysis effectiveness
 1348 in all taxa, and Longmires Buffer + ProtK was most effective as tissue in all
 1349 species was completely dissolved and liquid mostly clear (Table 2). If this
 1350 treatment potentially resulted in over-digestion and shearing of
 1351 chromosomal DNA was difficult to assess, however, in all treatments that
 1352 achieved good tissue disruption we observed higher DNA concentrations
 1353 in the final elute (see Fig. 2).

Table 2: Results of initial lysis and plasmid extraction tests

Organism	Tissue preparation	Concentration (ng/ul)	A260/A230	A260/A280
<i>Lumbriculus variegatus</i>	Pestle smear in resuspension buffer	60.0 (Biodrop)	2.6	2.0
<i>Fredericella sultana</i>	Pestle smear in resuspension buffer	8.0 (Biodrop)	4.0	2.0
<i>Oncorhynchus mykiss</i>	Pestle smear in resuspension buffer	1.0 (Biodrop)	-	1.0
<i>Oncorhynchus mykiss</i>	Longmire's + Prot K, no resuspension buffer	26.0 (Biodrop)	1.24	2.167
<i>Oncorhynchus mykiss</i>	Freeze + pestle, added resuspension buffer	11.0 (Biodrop)	2.2	1.83
<i>Fredericella sultana</i> (total: 24 extractions)	Longmire's + Prot K, no resuspension buffer	Avg. 7.78 (Qubit) [0.122 – 33.2]	-	-

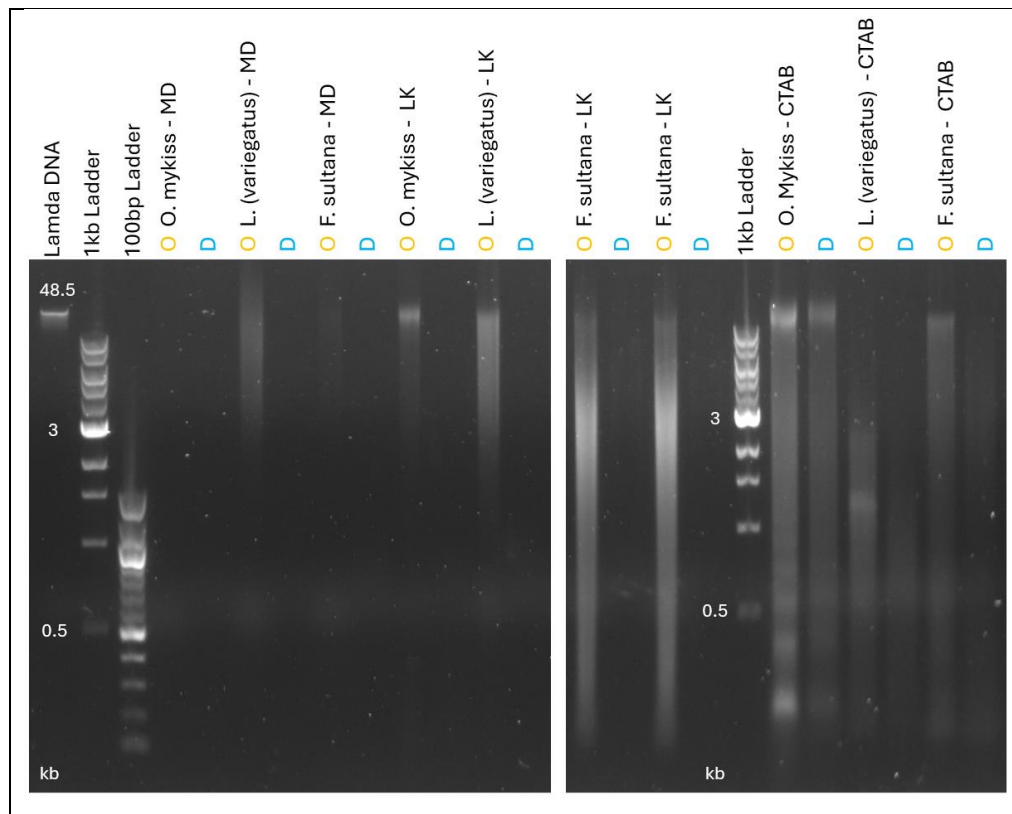


Figure 2: Electrophoresis agarose gel of plasmid prep DNA extracts of *Oncorhynchus mykiss*, *Lumbriculus variegatus* and *Fredericella sultana*. O = Original extract, D = Depletion of linear DNA with Exonuclease V (RecBCD) treatment; MD = Mechanical disruption with pestle smear, LK = Enzymatic lysis with Longmire's Buffer and Proteinase K, CTAB = Standard CTAB extraction of tissue.

1354 Nanopore sequencing

1355 A total of 2 million reads were generated, and 65.7% of the reads were
 1356 below 1000bp in length, with 34.7% in the 1,000-10,000bp bin (see Fig. 3a).
 1357 Only around 1.8% were between 10,000 to 50,000bp, the range of
 1358 mitochondrial genome sizes. The longest read is above 1 million bp. The
 1359 read length bins were re-evaluated once the mitochondrial genome was
 1360 assembled to assess the proportion of reads contributing to the target
 1361 assembly (Fig. 3b). Most of the reads sequenced and comprising the *F.*
 1362 *sultana* mitogenome were below 1000bp in length. Around 0.8% were
 1363 between 10,000 to 50,000bp, and specifically, 5 reads were around
 1364 20,000-24,000 bp, spanning the full length of the mitogenome.

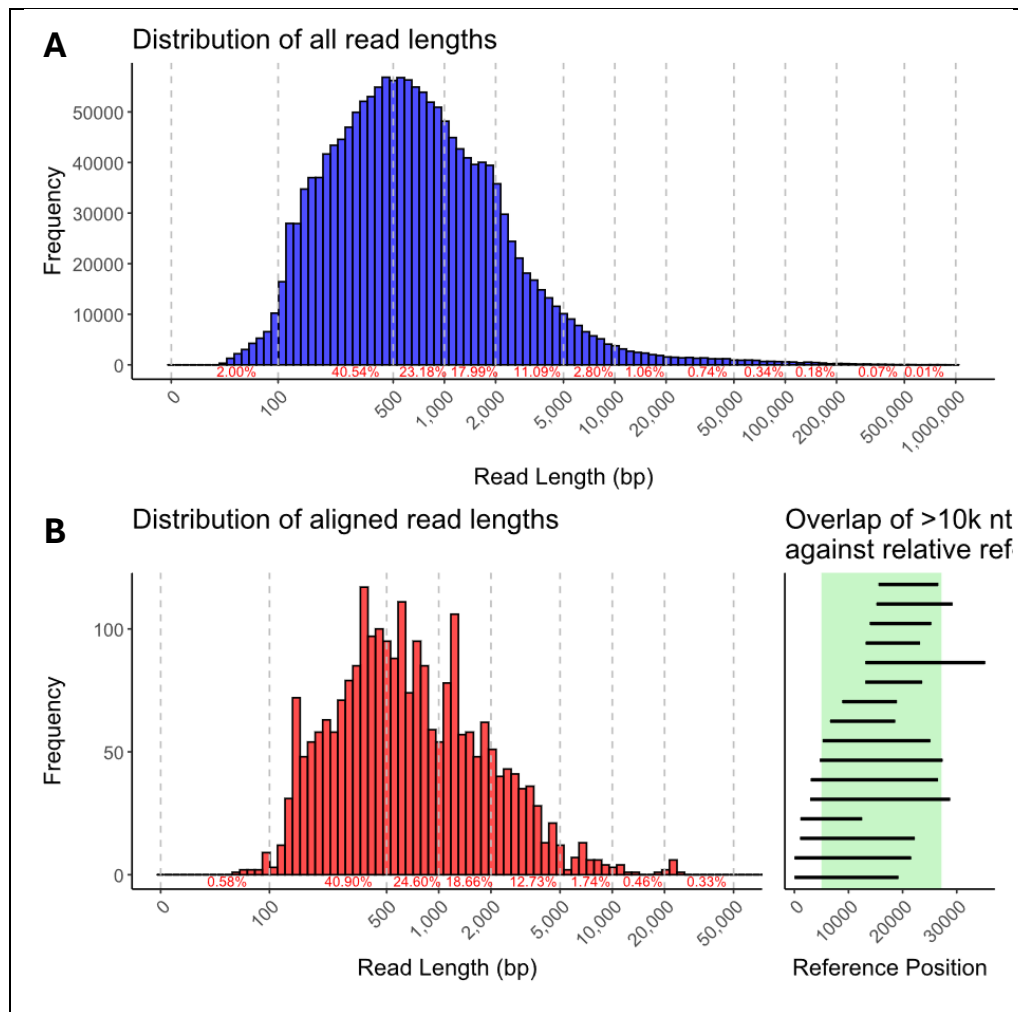


Figure 3: Read length distribution of the complete sequencing run with all combined samples A) Total sequenced reads B) Sequences that finally aligned to the assembled mitogenome and overlaps of long reads (>10k nt, 0.8%) with the green area indicating the length of the full mitogenome.

1365 Mitogenome assembly and annotation

1366 The full length of the *F. sultana* mitogenome assembly was 22,175bp (Fig.
 1367 4). The full mitogenome is longer than the proposed *Pectinatella magnifica*
 1368 mitogenome (NC_038192.1: 17,539 bp) and similar in length to *Cristatella*
 1369 *mucedo* (CM019737.2: 21,008 bp). The other freshwater bryozoan
 1370 mitogenomes were assembled from short-read sequencing data and are
 1371 of good quality, however unidentified gaps in the sequences might still
 1372 exists (especially around repeats and the potentially complex structures of

1373 the mitocontrol regions involved in D-loop formation and replication).
1374 Generally, complete, mostly marine, bryozoan mitogenomes are up to
1375 30kbp (longest: *Costaticella bicuspis* MW159929.1 with 31,788 bp). To
1376 illustrate the significance of long read data for mitogenome assembly, we
1377 mapped an existing dataset of 90.9m short-reads from *F. sultana* (Illumina
1378 250bp paired-end reads) against the newly produced reference
1379 mitogenome (Fig. 5). Although 237,475 reads (0.26%) mapped, and 6940x
1380 mean coverage on average was achieved, 0.98% of the mitogenome
1381 remained uncharacterised with a no coverage. Further, coverage is uneven
1382 with 7.31% below 100x and 2.60% below 10x coverage. Especially, repeat
1383 and non-coding regions are most affected. Our long-read data has a mean
1384 coverage of 68x with only 1.10% below 10x.

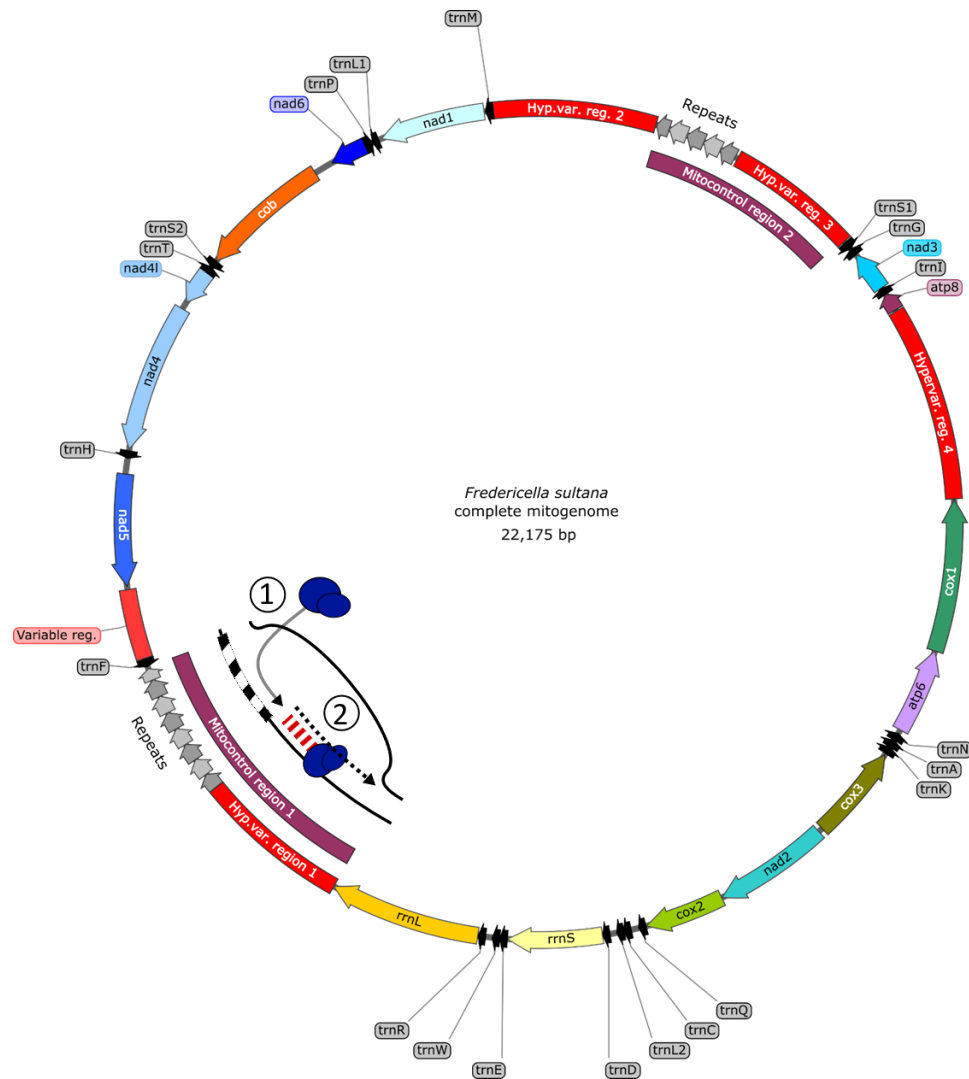


Figure 4: Annotations of the complete mitogenome from MITOS2 with all the genes from the OXPHOS proteins, rDNA, tRNAs, variable sections including the two mitocontrol regions including schematic of speculated initiation of mtDNA replication by 1) enzyme recruitment through motif (repeat) recognition and 2) RNA primer or direct strand synthesis

Table 3: Genomic annotation of genes, rRNA, tRNA and repeats of the mitogenome with position, length, direction and expression levels (where applicable)

Name	Type	Start	End	Length	Direction	Avg. Expression
trnS1(tct)	tRNA	2925	2992	67	+	211
trnG(tcc)	tRNA	2996	3060	64	-	1231
nad3	gene	3063	3416	353	-	7444
trnI(gat)	tRNA	3418	3480	62	-	1979
atp8	gene	3479	3637	158	-	2351
cox1	gene	5243	6637	1394	-	13979
atp6	gene	6971	7370	399	-	13447
trnN(gtt)	tRNA	7406	7471	65	-	1045
trnA(tgc)	tRNA	7475	7539	64	-	256
trnK(ttt)	tRNA	7541	7605	64	-	2296
cox3	gene	7619	8441	822	-	14970
nad2	gene	8481	9478	997	+	5428
cox2	gene	9481	10162	681	+	18902
trnQ(ttg)	tRNA	10169	10229	60	+	1921
trnC(gca)	tRNA	10298	10360	62	+	46
trnL2(taa)	tRNA	10361	10421	60	+	38
trnD(gtc)	tRNA	10482	10541	59	+	341
rrnS	rRNA	10546	11358	812	+	2583
trnE(ttc)	tRNA	11354	11418	64	+	151
trnW(tca)	tRNA	11419	11481	62	+	19
trnR(tcg)	tRNA	11540	11602	62	+	321
rrnL	rRNA	11606	12903	1297	+	2481
trnF(gaa)	tRNA	15398	15464	66	-	264
nad5	gene	16102	17056	954	-	2765
trnH(gtg)	tRNA	17193	17255	62	-	725
nad4	gene	17271	18549	1278	-	4175
nad4l	gene	18617	18919	302	-	2717
trnT(tgt)	tRNA	18920	18981	61	+	307
trnS2(tga)	tRNA	18983	19046	63	-	716
cob	gene	19045	20176	1131	-	8332
nad6	gene	20344	20667	323	-	5627
trnP(tgg)	tRNA	20660	20725	65	+	976
trnL1(tag)	tRNA	20727	20784	57	-	503
nad1	gene	20811	21712	901	-	5049
trnM(cat)	tRNA	21725	21788	63	-	787
repeat	repeat	1016	1127	111	-	218
repeat	repeat	1128	1286	158	-	170
repeat	repeat	1287	1441	154	-	249
repeat	repeat	1442	1596	154	-	257
repeat	repeat	1597	1735	138	-	215

repeat	repeat	14218	14354	136	+	266
repeat	repeat	14355	14509	154	+	280
repeat	repeat	14509	14663	154	+	86
repeat	repeat	14663	14817	154	+	64
repeat	repeat	14817	14970	153	+	56
repeat	repeat	14972	15126	154	+	45
repeat	repeat	15126	15278	152	+	9
repeat	repeat	15278	15378	100	+	13

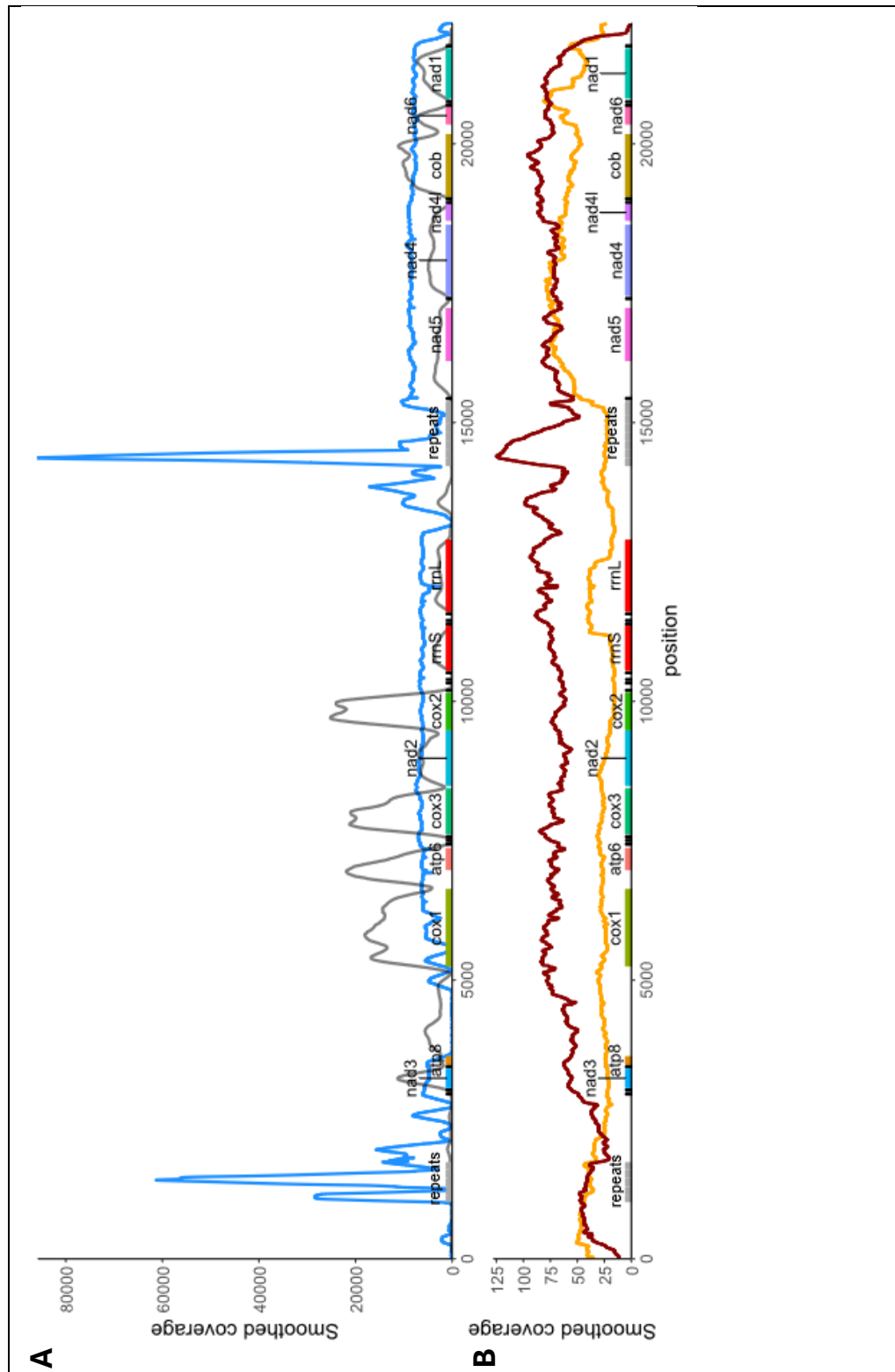


Figure 5: A) Smoothed coverage of short read MiSeq data (2x 150bp) (blue) (H. Hartikainen, unpublished) and average of all available RNAseq data (grey). B) Coverage of new long read sequencing using latest R10.4.1 flow cell technology (dark red) compared to older R9 sequencing (yellow; Hanna Hartikainen, unpublished).

1385 The tightly spaced coding regions of all the expected mitochondrial genes
1386 were identified and annotated by a single round of MITOS analysis (Fig. 4).
1387 The 12S (rrnS) and 16S (rrnL) were identified adjacent, and all tRNAs except
1388 trnY and trnV were annotated by MITOS. The presence of the two missing
1389 tRNAs was confirmed with manual alignment of the tRNAs from related
1390 species (*Cristatella mucedo* and *Pectinatella magnifica*). Open reading
1391 frame searches by ORFfinder and alignment of RNA-seq data (Fig. 6b&c)
1392 confirmed the validity of the coding gene placements. Expression was
1393 highest on the coding regions and the ribosomal DNA (Table 3 and Fig. 6).
1394 tRNAs regions showed very low or no expression. RNAseq is an optional
1395 step in the Step-by-Step protocol as the coding region annotation was
1396 successful when based only on MITOS and ORFfinder results. However, the
1397 additional RNAseq data aided in the analysis of four large variable regions
1398 discovered in the mitogenome (Fig. 4).

1399 Three of the large variable regions with no gene annotation were located
1400 between trnS1 and trnM, nad5 and rrnL (16S), cox1 and atp8 (Fig. 4).
1401 RNAseq analysis confirmed expression in these regions, and a viable,
1402 functional domain containing ORF was found between cox1 and atp8. This
1403 was determined to be a group II intron (Fig. 6 & 8, discussed below). The
1404 other two non-coding regions in this area spanned 3,312 and 2,459 bp,
1405 respectively, and contained several, identical 154 bp long repeat motifs
1406 (total 13) in the non-coding region 1 (8) and 2 (5) (Fig. 4). Direction of the
1407 repeats is reversed in the non-coding region 2 compared to 1 (Fig. 4). The
1408 last repeats of the sections are shorter (136 or 138 bp and 111 or 100 bp)
1409 than the previous full-length motifs. A low level of RNAseq transcripts
1410 mapped to the 3'-end of these regions, and on some of the repeats (Fig. 6B),
1411 with expression matching the reversed strand direction of the repeats (Fig.
1412 6C). A function of these regions could not be established.

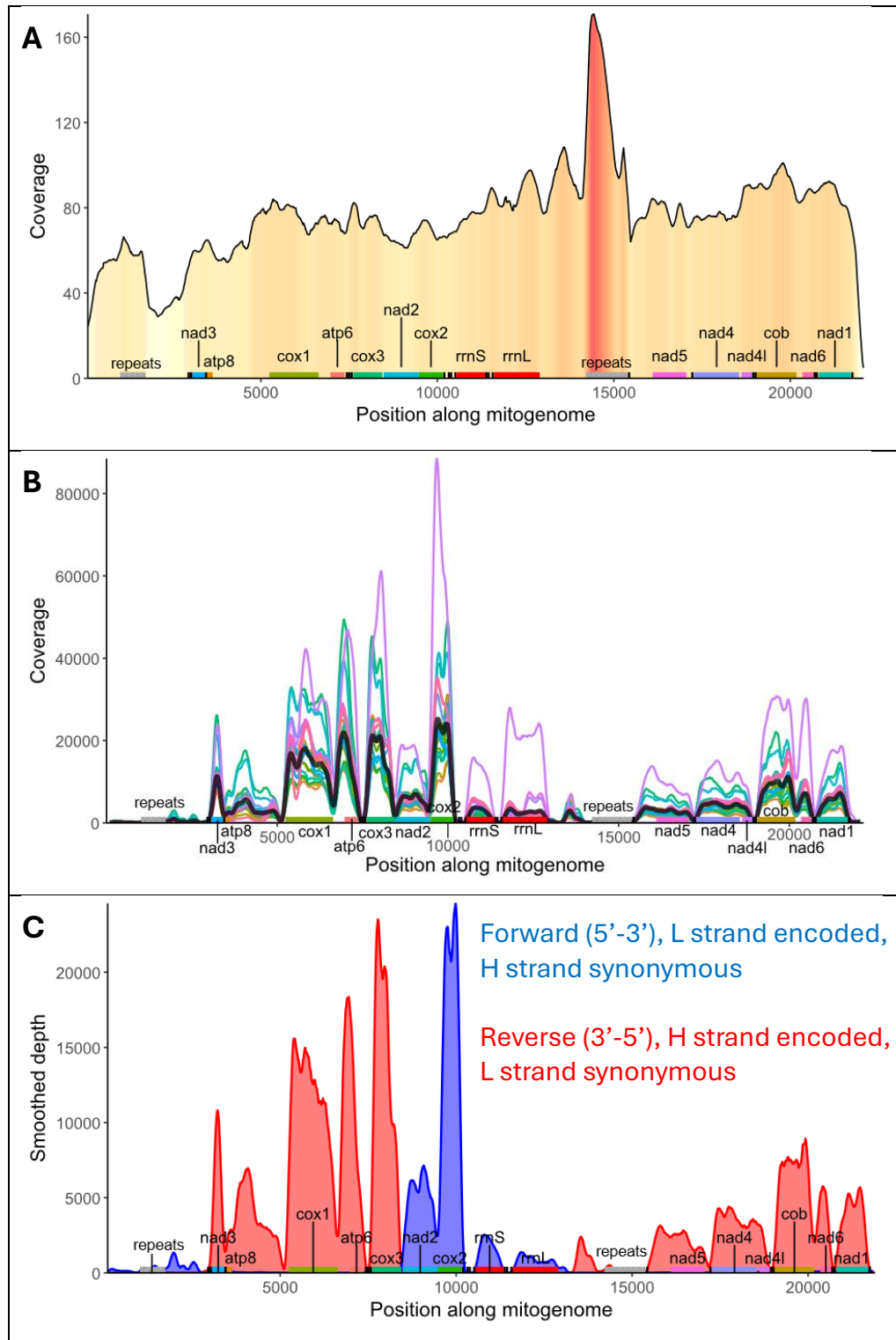


Figure 6: A) Coverage of the Nanopore mitogenome sequencing across the complete mitogenome B) Expression profiles along the mitogenome from published and unpublished short read RNA-seq data of *Fredericella sultana* zooids or statoblasts with genomic annotation C) Strand specific expression profile of a single individual, positive (forward) and negative (reverse).

1413 *Strand directionality and pyrimidine-purine ratios*

1414 Strand directionality is crucial for accurate gene prediction, the identifying
1415 co-transcribed regions, and the inference of putative promoter sites.
1416 Mitogenome annotations were based on the heavy (H or -) strand, as the
1417 reference assembly was purine rich (53%, Fig. 7). Only nad2 and cox2, and
1418 the ribosomal RNA genes are L strand-encoded and H strand synonymous;
1419 all other protein-coding genes of the respiratory complex, as well as the
1420 group II intron-encoded protein (IEP), are H strand coded and L strand-
1421 synonymous (Fig. 4, Fig. 6c).

1422 Strand switching is apparent in the MITOS2 annotations (Fig. 4), the
1423 reversal in mRNA transcript directionality (Fig. 6c), and the kink in the
1424 purine-pyrimidine walk between cox3 and nad2 (Fig. 7). rRNA is purine-
1425 rich, other protein-coding transcripts are pyrimidine-rich. Although nad2,
1426 cox2 and rRNA share the same transcriptional direction, they differ in
1427 nucleotide composition. The IEP is less pyrimidine-rich than the
1428 surrounding genes.

1429 Differences in base ratios between coding and the proposed mitocontrol
1430 regions support their distinct functional roles. While the repeat regions
1431 exhibit balanced base composition, transcripts in the anterior (3') region
1432 are highly pyrimidine-rich. Transcripts associated with the mitocontrol
1433 region 1 (beyond 13 kbp) are L strand synonymous and H strand encoded.
1434 Conversely, those associated with the mitocontrol region 2 (beyond 3 kbp)
1435 are H strand-synonymous and L strand encoded. With the exception of
1436 trnS1, trnT and trnP, all tRNAs share the same directionality as their
1437 surrounding genomic regions.

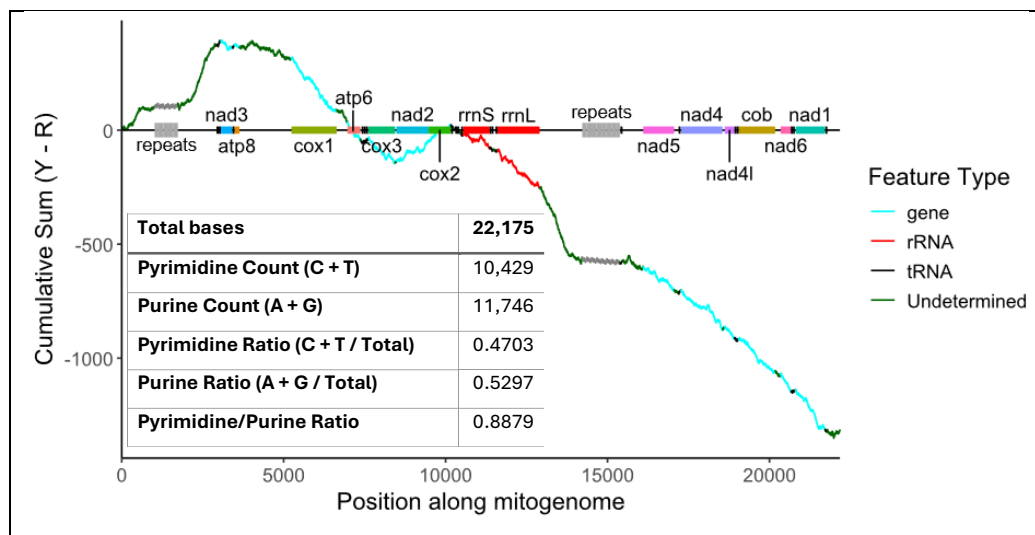


Figure 7: Pyrimidine-purine walk of the – strand (mtDNA) of *Fredericella sultana* with genetic annotation and regions coloured according to feature type and table of pyrimidine-purine counts and ratios

1438 *Discovery of a group II intron-encoding protein*

1439 Between the atp8 and cox1 genes lies a 1,606bp intergenic region that is
 1440 unannotated and appears non-coding. However, RNA-seq data show
 1441 expression levels spanning the entire region and comparable to known
 1442 mitochondrial genes (Fig. 8a). Within this region, we identified a 1446bp
 1443 ORF encoding 481aa protein containing a conserved domain X at the C
 1444 terminus – characteristic of Type II maturases – as annotated using the
 1445 Pfam database. This domain is associated with Group II introns, which are
 1446 highly divergent, self-splicing mobile genetic elements known to encode
 1447 both functional ribozymes and maturase proteins (Zimmerly and Semper
 1448 2015).

1449 Despite their functional similarity and likely shared ancestry, the
 1450 nucleotide sequences of Group II intron-encoded proteins are highly
 1451 divergent, making them difficult to detect by sequence homology alone
 1452 (Kobayashi et al. 2022). Simulation with Alphafold (Fig. 8b-d) was used to
 1453 validate the functional identity of this ORF. The resulting 3D model was well
 1454 resolved, with a predicted IDDT score >60 across most of the protein (rank
 1455 1, average pIDDT=59.5, pTM=0.581) (Fig. 8b). The structure features a

1456 canonical “hand and palm” fold, comprising predominantly alpha-helices
1457 and some beta-sheets, with lower confidence in the regions linking the
1458 domains (Fig. 8d).

1459 To further assess structural similarity, we compared the predicted model
1460 against known structures in the BSRC protein database using multiple
1461 alignment algorithms (jFATCAT and jCE-CP). These analyses revealed
1462 strong structural similarity to known Group II intron-encoded proteins. The
1463 best structural match by RMSD (lower equal better match) was to entry
1464 8H2H (*Lactococcus lactis*), with RMSD scores ranging from 3.12 to 2.79
1465 (jFATCAT) and 3.73 to 2.18 (jCE-CP). According to TM-score (higher is
1466 better), the best match was 5HHJ (*Roseburia intestinalis* XB6B4), with
1467 scores from 0.32 to 0.43 (jFATCAT) and 0.28 to 0.62 (jCE-CP). Closest
1468 identity match was best 5HHJ (*R. intestinalis* XB6B4) together with 5HHJ,
1469 8H2H and 7UIM (7UIM: *Agathobacter rectalis*), with identities from 14% to
1470 18%, (jFATCAT) and 13% to 20% (jCE-CP). High confidence structural
1471 predictions aligned and scored the best against Cryo-EM or X-ray based
1472 deposited crystal structures. Overall, the best matches were 8H2H (*L.*
1473 *ctococcus lactis*) and 5HHJ (*R. intestinalis* XB6B4), with high structure but
1474 low sequence similarity.

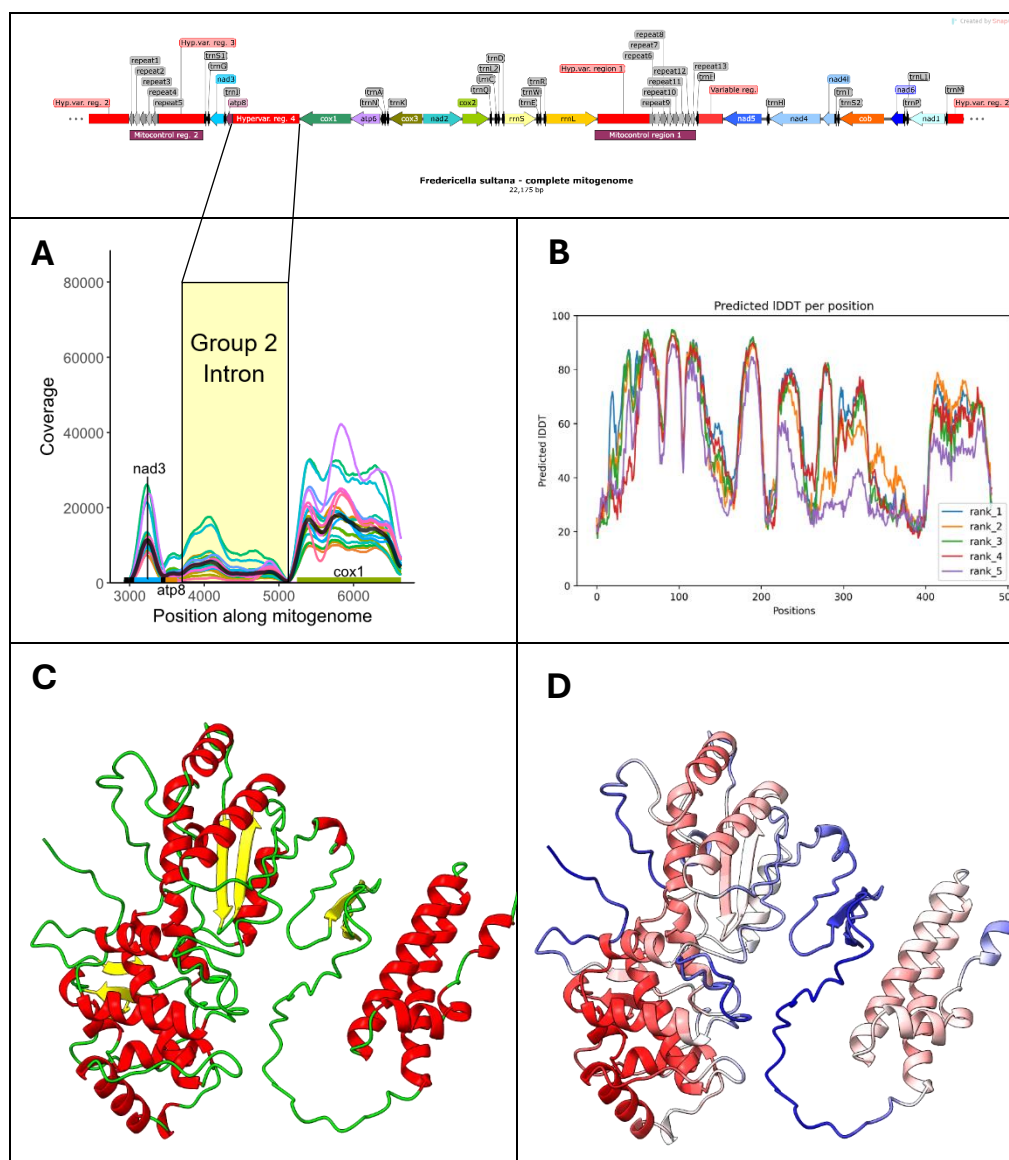


Figure 8: A) Location of the predicted group II intron B) Predicted IDDT score of the predicted IEP of the group II intron by AlphaFold2 C) Highest scoring predicted structure (rank 1, pLDDT=59.5 pTM=0.581) d) pLDDT scores over the predicted structure.

Discussion

Access to mitogenome sequences is invaluable for ecological and evolutionary research. This study demonstrates that complete reference mitogenomes can be generated directly from tissue samples without prior reference sequences or PCR amplification. Using a single Nanopore flow cell, up to 24 individuals can be sequenced within a one-day workflow. The approach offers unmatched speed and affordability: with the most cost-effective Nanopore options, sequencing costs range from £40-66 per sample, depending on consumables. Initial setup costs were £900 (as of

1483 17/07/2024) or £1,800 at the date of writing (07/02/2025) with a MinION
1484 Starter Pack. Each run processes 24 samples, with first results available
1485 within 12 hours. A 24-hour runtime is typically sufficient (maximum
1486 runtime 48h), and post-processing - basecalling, de-novo assembly and
1487 annotation - can be completed under one hour.

1488 Many tools and steps are available for mitogenome analysis, but new users
1489 often struggle to move efficiently from raw sequence data to assembly and
1490 validation. To address this, I provide a Step-by-Step protocol that enables
1491 rapid progress in generation and preparation of mitogenome data for
1492 research purposes. During the annotation, we discovered a Group II intron
1493 for the first time in freshwater bryozoan, highlighting the power of long-read
1494 sequencing compared to primer-based or short-read NGS approaches.
1495 This finding adds to a small but growing number of Group II intron reports
1496 in bilaterians, previously observed in annelids and marine bryozoans of the
1497 order Cheilostomatida (Vallès et al. 2008, Richter et al. 2015, Jenkins et al.
1498 2022, Kobayashi et al. 2022), and marks just the fourth such case reported
1499 in bilaterians to date.

1500 *Maximising speed, yield and DNA quality for mtDNA sequencing*

1501 Using an off-the-shelf plasmid Miniprep kit offers a rapid and accessible
1502 approach for extracting size-targeted, potentially mitochondria-enriched
1503 DNA from metazoan tissues, ideal for whole mitogenome sequencing.
1504 These kits are widely available in research labs, standardise workflows,
1505 and minimise user variability. The plasmid prep method is most effective
1506 when the input DNA is largely intact, as fragments within the typical
1507 mitogenome range are best retained eluted from the extraction columns.
1508 To optimise this, several pre-lysis treatments were tested. The most
1509 effective was enzymatic lysis in Longmire's buffer with added Proteinase K,
1510 omitting mechanical disruption to prevent DNA shearing. While freezing
1511 proved unsuitable, previous studies have reported success with gentle
1512 tissue disruption using glass or steel beads (Mascolo et al. 2019). A

1513 combination mild mechanical disruption and short enzymatic lysis may
1514 therefore represent the optimal compromise.

1515 Mitogenome size may also affect DNA extraction efficiency. The *F. sultana*
1516 mitogenome, at 22.5kb, falls near the optimal column capture range
1517 (~25kb), suggesting this approach may be most effective for species with
1518 smaller or similarly sized mitochondrial genomes. Library preparation can
1519 further the proportion of mitogenome reads in sequencing output. Circular
1520 mitogenomes are supercoiled, potentially limiting accessibility of the
1521 barcoded transposome complex used in the Nanopore library prep. While
1522 read coverage was generally even across the mitogenome, coverage spiked
1523 around the two large, non-coding mitochondrial regions. These are often
1524 associated with initiation of mitochondrial transcription and replication
1525 and might play a role in D-loop formation (Bernt et al. 2013). These regions
1526 differ from the rest of mitogenomes in two ways: they contain multiple
1527 ~125bp repeats, which can interfere with alignment and assembly, and
1528 they form a linear, three-stranded D-loop-like structure composed of a
1529 single unpaired DNA strand and a paired RNA/DNA hybrid strand, which
1530 may resist transposome cutting. To mitigate these biases, enzymatic
1531 linearisation (restriction enzymes or CRISPR/Cas) may increase coverage
1532 uniformity. This approach could also enable single-molecule reads
1533 spanning nearly the full mitogenome (Ramón-Laca et al. 2023).

1534 *From single reads to whole mitogenome de-novo assembly*

1535 The latest Nanopore technology using the Rapid Barcode/Sequencing kit
1536 with an R10.4.1 flow cell on MinION platform enables efficient library
1537 preparation, sequencing and basecalling of mitogenomes within hours.
1538 The amplification-free approach is well-suited for de novo mitogenome
1539 assembly, particularly when references are lacking or incomplete.

1540 Mitochondria-specific extraction (Gould et al. 2015), such as nuclear DNA
1541 depletion, or mitochondrial enrichment strategies (Ramón-Laca et al.
1542 2023), including long-range PCR and overlapping primers (King et al. 2014,

1543 Mascolo et al. 2017, Karin et al. 2022, Kobayashi et al. 2022), have shown
1544 some promise, and whole-genome amplification may increase on-target
1545 reads (Yu et al. 2022). However, such methods often require prior
1546 knowledge of the mitogenome structure and supercoiling mitigation. While
1547 genome skimming or shotgun sequencing using short reads has been
1548 successfully applied in bryozoans (Orr et al. 2019, Jenkins et al. 2022) and
1549 other taxa (Gan et al. 2014, Richter et al. 2015, Trevisan et al. 2019), it may
1550 be costlier, less efficient, and more prone to coverage bias than long-read
1551 strategies.

1552 The choice of assembler is a crucial step that often takes a long time to
1553 refine, the Step-by-Step protocol provided here simplifies this choice.
1554 Assemblers designed for long reads and those that are “error-rate” aware
1555 perform best. Of those tested, only Flye consistently produced complete,
1556 circular mitogenome assemblies. Notably, NOVOplasty and MitoHiFi are
1557 incompatible with nanopore data. Flye, designed for noisy long reads and
1558 variable coverage, performed well even under metagenomic sequence
1559 datasets (Kolmogorov et al. 2019, Kolmogorov et al. 2020) and is
1560 recommended for similar projects due to its speed, reliability and ease of
1561 use.

1562 Only ~2% of assembled contigs were confidently assigned to *F. sultana*,
1563 likely due to sparse reference data. At the time of analysis, only partial
1564 mitochondrial sequences and nuclear 18S/28S rRNA genes for *F. sultana*
1565 were available in the nt database on GenBank. Many “Unassigned” contigs
1566 (75% of a read subset) likely represent nuclear *F. sultana* sequences.
1567 Therefore, the mitochondrial enrichment employed here does not remove
1568 the nuclear components, and the data can be used for skimming for other
1569 phylogenetically informative genes, e.g. the nuclear ribosomal operon
1570 genes (18S/28S) genes were identified on a single contig providing the full
1571 operon from the nanopore data.

1572 9.1% of the reads suggested bacterial origin. As the bryozoan guts were not
1573 dissected out, this was an expected outcome, however, it may also reflect

1574 competition between abundant, short bacterial molecules and rarer,
1575 longer bryozoan DNA. Although cleaner input tissue could reduce this bias,
1576 the “quick-and-dirty”, essentially metagenomic samples, still yielded
1577 complete, high-quality mitogenomes.

1578 *Mitogenome structure, annotation and gene expression and mechanisms*

1579 Metazoan mitogenomes are diverse but typically consist of a single
1580 circular molecule, as observed in all bryozoan mitogenomes available in
1581 GenBank (Nosek and Tomáška 2003, Gissi et al. 2008, Bernt et al. 2013,
1582 Shtolz and Mishmar 2023, Veeraragavan et al. 2024). This architecture is
1583 retained in *F. sultana*, supported by long reads that span and circularise
1584 the mitogenome. Short-read alignments revealed biallelic SNPs in all
1585 individuals, with only low level of heteroplasmy (data not shown),
1586 consistent with canonical uniparental mitochondrial inheritance.
1587 Mitochondrial gene order is generally conserved among closely related
1588 metazoans. *F. sultana* shares syntenic gene blocks with the
1589 phylactolaemates *P. magnifica* and *C. mucedo*, interrupted by variable
1590 non-coding regions. Shared features include a directional inversion of the
1591 *nad2* and *cox2* genes (Fig. 4, 6c, 7), and non-coding regions between *nad3*-
1592 *nad1* and *nad5-rrnL*. These features are conserved across
1593 phylactolaemates, but are absent in the more diverse marine
1594 gymnolaemates and stenolaemates.

1595 RNA-seq data supported the identification of non-coding regions, novel
1596 coding regions and transcript directionality. The concentration of protein-
1597 coding genes on the H strand suggests transcriptional asymmetry, likely
1598 evolved for efficient mitochondrial gene expression. Together with strand
1599 biases and nucleotide asymmetries, the transcript organisation indicates
1600 a tightly coordinated, possibly modular system of genome expression and
1601 maintenance. Pyrimidine enrichment of protein-coding genes and purine-
1602 rich rRNAs are expected due to translational and amino acid chemistry
1603 based constraints (Bradshaw et al. 2005), regardless of directionality. The
1604 localisation of *nad2*, *cox2*, and rRNAs on the L strand, despite

1605 transcriptional alignment, may reflect a conserved regulatory module or
1606 remnants of an ancestral genome structure.

1607 Repeats in the mitocontrol region show balanced pyrimidine-purine ratio
1608 and are preceded by a pyrimidine-enriched, non-coding transcribed
1609 segment with unclear function, although possibly involved in mitogenome
1610 regulation and replication. Repeats or introns are absent in the other
1611 sequenced phylactolaemates, which also lack the non-coding pattern
1612 seen in *F. sultana*. Other non-coding regions in *C. mucedo* are pyrimidine-
1613 rich (based on adjacent gene direction), while *P. magnifica* shows mixed
1614 composition. The other phylactolaemate mitogenomes were assembled
1615 from short read data, and may therefore miss repeats or erroneously
1616 consider reads stemming from nuclear mitochondrial DNA segment
1617 (NUMTs). Long read data is more reliable here, as it spans problematic
1618 regions and minimises misalignments.

1619 The mechanisms of mitochondrial DNA transcription and replication in
1620 bryozoans remain unknown but are assumed to resemble those in
1621 mammals, where they are best understood (Falkenberg and Gustafsson
1622 2020, Rackham and Filipovska 2022). In this model, polycistronic
1623 transcripts - long RNAs containing multiple genes interrupted by tRNAs -
1624 are initiated at promoters in the mitocontrol region. Post-transcriptional
1625 processing follows the punctuation model, where intergenic tRNAs are
1626 cleaved to release mature rRNA, tRNA and mRNA. Replication typically
1627 begins at the heavy-strand origin (O_H) within the mitocontrol region and at
1628 the light-strand origin (O_L), formed during strand separation. A D-loop is
1629 forms when a third strand of DNA (7S DNA), primed by 7S RNA, is inserted
1630 at O_H , marking the start of heavy-strand replication (Nicholls and Minczuk
1631 2014). The two mitocontrol regions (1 and 2) are located on opposite
1632 strands and have opposing orientations, suggesting complementary roles.
1633 Region 2 shows fewer repeats and lower expression, mostly downstream
1634 of the repeats, possibly representing RNA primers for replication
1635 (Falkenberg and Gustafsson 2020). The repeats may act as motifs

1636 recognized by the enzyme producing RNA primers; more repeats could
1637 enhance enzyme recruitment and RNA production. It remains unclear
1638 whether these two mitocontrol regions alone regulate replication or if
1639 additional origins exist. No O_L has been identified in *F. sultana*, and similar
1640 regions are not reported in other bryozoan mitogenomes.

1641 *Group II introns newly discovered in freshwater bryozoans*

1642 Self-splicing introns and their associated mobile genetic elements are
1643 common in organellar genomes of some eukaryotes, including protists
1644 (e.g. diatoms), plants, and fungi (Novikova and Belfort 2017). While rare in
1645 metazoan mitochondria, they have been found in basal non-bilaterian
1646 groups such as sponges (Huchon et al. 2015, Schuster et al. 2017, Ahmed
1647 et al. 2024) and placozoans (Dellaporta et al. 2006)). More recently,
1648 mitochondrial group II introns have also been reported in bilaterians,
1649 including annelids and marine bryozoans (Vallès et al. 2008, Richter et al.
1650 2015, Jenkins et al. 2022, Kobayashi et al. 2022). Jenkins et al. (2022)
1651 identified these introns in four cheilostome bryozoans. The same type of
1652 intron now identified in *F. sultana* marks the first report for the Class
1653 Phylactolaemata. No introns have been reported in the only other two
1654 available phylactolaemate mitogenomes of *Cristatella mucedo* and
1655 *Pectinatella magnifica*.

1656 Group II introns are composed of a 2-3kb elements composed of an
1657 autocatalytic ribozyme region (500-900bp) and a distally located protein
1658 coding region. Their structure and function have been extensively reviewed
1659 (Michel et al. 1989, Palmer and Logsdon 1991, Bonen and Vogel 2001,
1660 Lambowitz and Belfort 2015, Zimmerly and Semper 2015, Zhao and Pyle
1661 2017, Xu et al. 2023, Li and Jiang 2024), and briefly summarised here. The
1662 ribozyme sequence is highly variable, but its secondary structure typically
1663 forms six conserved domains (I-VI) responsible for self-splicing. The intron
1664 encoded protein (IEP) contains a reverse transcriptase (with seven

1665 conserved blocks), as well as X, D, DNA binding and endonuclease
 1666 domains. The IEP stabilises the ribozyme and enables splicing, reverse
 1667 transcription and retromobility. In functional introns, IEP binds to the intron
 1668 RNA, promoting lariat formation. The resulting ribonucleoprotein (RNP)
 1669 targets DNA performs reverse splicing into the top strand, and primes
 1670 reverse transcription on the bottom strand. Repair mechanisms then
 1671 complete the integration. In organelles, introns are usually found in
 1672 housekeeping genes, are more stable and less truncated than bacterial
 1673 versions, and often lose mobility, retaining only splicing or regulatory roles.
 1674 They likely originated via horizontal gene transfer from bacterial
 1675 retroelements. Jenkins et al. (2022) suggest bryozoan introns may have
 1676 been acquired from symbiotic or associated microbes (e.g. bacteria, algae,
 1677 fungi) (Jenkins et al. 2022). Clonal propagation, budding, or regeneration in
 1678 somatic cells – bypassing the germline – may have enabled their
 1679 persistence (Szitenberg et al. 2010, Huchon et al. 2015). Other
 1680 regenerative, microbe associated metazoans (e.g. sponges, placozoans,
 1681 cnidarians and annelids) also host mitochondrial introns, typically of the
 1682 IIA RNA subgroups with mitochondria-like IEPs, likely acquired
 1683 independently (Dellaporta et al. 2006, Vallès et al. 2008, Huchon et al.
 1684 2015, Lavrov and Pett 2016, Schuster et al. 2017, Ahmed et al. 2024). *F.*
 1685 *sultana* meets many of these hallmarks, as among Phylactolaemata
 1686 species is considered to reproduce sexually relatively rarely, relying on
 1687 clonal budding, as demonstrated by relatively high frequently of
 1688 monoclonal populations (H. Hartikainen, pers. obs).

1689 The Group II introns discovered by Jenkins et al. (2022) include reverse
 1690 transcriptase and maturase domains primarily within or near the *cox1*
 1691 gene, alongside additional introns lacking domain structures. In *F. sultana*,
 1692 only one such intron with identifiable domains is found, located adjacent
 1693 to a complete, uninterrupted *cox1* gene, similar to observations in the
 1694 marine bryozoans *Cupuladria biporosa* and *Discoporella cookae* (Jenkins
 1695 et al. 2022). The group II intron transcript is slightly pyrimidine rich,

1696 however, less so than the neighbouring genes (see Fig. 7). In three of the
 1697 marine bryozoans from the order Cheilostomatida (*Exechonella vieirai*,
 1698 *Cupuladria biporosa*, *Discoporella cookae*) all mitochondrial introns are
 1699 distinctly purine rich (in contrast to the surrounding pyrimidine rich exons)
 1700 (Jenkins et al. 2022) (see Supp. Fig. 4). *Parantropora penelope* is an
 1701 exception where all rRNA and introns are roughly neutral. For the first time
 1702 in bryozoans, expression of the group II intron and its IEP region has been
 1703 demonstrated, with expression comparable to rRNA, nad5 and nad4.
 1704 However, no continuous expression or linkage spans between cox1 and the
 1705 IEP. Whether the expressed transcript yields a functional ribozyme and IEP
 1706 is uncertain. Structurally, the predicted IEP aligns with function bacterial
 1707 references. The protein is shorter (~1.5kb) than typical group II introns (2-
 1708 3kb), likely due to gaps or truncation, leaving limited room for a full
 1709 ribozyme function. The simulated protein structure matched well in the
 1710 “hand and palm” region with the reference structures provided by
 1711 Alphafold2. The thumb (domain X) shows less certainty and weaker 3D
 1712 alignment, suggesting limited model accuracy in that region. RNA-seq
 1713 shows that IEP has a distinct expression profile, separate from flanking
 1714 atp8 and cox1. As in other bryozoans, ribozyme domains are absent,
 1715 complicating identification. Domains V and VI – crucial for catalytic
 1716 function and IEP interaction – were not detected, raising questions about
 1717 whether the IEP remains functions in splicing, or has adopted another role.

1718 One of the many questions that remain is: Are group II introns truly rare in
 1719 metazoans, or just simply underdetected? Their increasing discovery in
 1720 lower metazoans and basal bilaterians suggests a wider, overlooked
 1721 presence. Their evolutionary origins via horizontal gene transfer, from
 1722 viruses, bacteria, fungi or algae, remain plausible, but unresolved
 1723 (Lambowitz and Belfort 2015, Richter et al. 2015, Jenkins et al. 2022,
 1724 Kobayashi et al. 2022, Ahmed et al. 2024). No higher bilaterians show
 1725 organellar introns, and genomic data for lower bilaterians are still sparse.
 1726 As regulators of mitochondrial function, group II introns are also being

1727 explored as potential drug targets (Liu and Pyle 2021), and our approach of
1728 rapidly generating mitogenomes of in underrepresented taxa may help
1729 deliver the data needed to assess their importance.

Conclusion

1730 We present a time- and cost-efficient full mitogenome sequencing
1731 approach using optimised methods for extraction, sequencing and de
1732 novo assembly. The pipeline performs robustly even with metagenomic
1733 samples. Advances in Nanopore sequencing chemistry and base calling
1734 now allow in-house sequencing that rivals industry standard. The resulting
1735 long reads outperform short-read data in mitogenome assembly, enabling
1736 the discovery of a previously unknown group II intron and improved
1737 resolution of non-coding regions. Our streamlined pipeline handles low
1738 coverage and Nanopores high error rates effectively. Access to complete
1739 mitogenomes enables long-range PCR from tissue or environmental
1740 samples (e.g. eDNA), eliminating reliance on short marker-based
1741 approaches. Developing markers that work on both DNA and RNA could
1742 deepen our understanding of eDNA production, decay, and
1743 spatiotemporal dynamics, while improving detection and quantification.
1744 Ultimately, insights into the genetics and ecology of *F. sultana* may
1745 advance our understanding of biotic interactions, such as host-parasite
1746 infection dynamics and disease emergence.

Chapter 3: *River network scale population genetics of the bryozoan *Fredericella sultana**

KEYWORDS: population genetics, river, bryozoan, haplotypes, intraspecific diversity

Abstract

1747 The freshwater bryozoan *Fredericella sultana* is the competent primary
1748 host of the parasite *Tetracapsuloides bryosalmonae* which causes
1749 proliferative kidney disease in wild and farmed salmonoid fish. Despite its
1750 crucial role in the parasite's disease dynamics ecology and population
1751 genetics of *F. sultana* are poorly understood. Amplification-free, direct
1752 Nanopore sequencing allowed me to conduct population genetic-level
1753 analysis for the full mitogenome of 17 individuals from four sites of a single
1754 river network (River Erewash, UK). I found non-uniform genetic diversity
1755 between different mitogenome regions (average: 2%) and identified 1609
1756 segregating sites (rate of 7%). The individuals stem from three different
1757 ancestry populations and form two major haplotypes. Additionally, we
1758 obtained 154 sequences of a short nad3 gene marker including eight
1759 further sites of the same river and a Swiss outgroup from tissue and eDNA.
1760 Marker sequencing matched the full mitogenome findings while also
1761 extending the haplotype network which is dominated by two major,
1762 ubiquitous haplotype groups and accompanied by two rarer haplotype
1763 groups found mainly on two separate sites. Based on these results we
1764 designed a marker set for comprehensive 18 mitochondrial loci covering a
1765 wide range of genetic diversity and suitable for future population genetic
1766 studies on tissue and eDNA.

Introduction

1767 Understanding the genetic structure of natural populations is fundamental
1768 to studying how species disperse, adapt and persist across heterogenous
1769 and changing environments. Population genetic approaches offer valuable
1770 insight into gene flow, dispersal dynamics and historical demography,
1771 underpinning the species ability to respond to environmental change
1772 (Demarche et al. 2019, Angst et al. 2022, Wadgymar et al. 2022). Recent
1773 advances in sequencing technology and molecular tools have enabled
1774 increasing resolution to these processes, particularly when paired with
1775 non-invasive approaches such as eDNA (Adams et al. 2019, Sigsgaard et
1776 al. 2020). Although eDNA has been widely adopted for species detection
1777 and biodiversity monitoring (Minamoto 2022), its application to
1778 intraspecific genetic studies remains infrequent (for examples see:
1779 Elbrecht et al. 2018, Tsuji et al. 2020, Dugal et al. 2021, Farrell et al. 2022).
1780 Progress is often hindered by lack of information on population genetic
1781 diversity in the reference populations that contribute to the eDNA pool.
1782 Reference haplotype diversity is particularly challenging to assess in
1783 species that are mobile, and sexually reproducing, meaning that the
1784 reference population shedding the eDNA is complex and dynamic. For
1785 example, assessing the number of contributors (Sethi et al. 2019), allele
1786 frequencies (Andres et al. 2021) and suitable error-correction (Macé et al.
1787 2022) are all necessary. Nonetheless, eDNA represents a promising
1788 frontier for scalable, spatially explicit assessments of population
1789 structure, especially in dendritic and hydrologically dynamic systems such
1790 as river networks (for examples see: Marshall and Stepien 2019, Stepien et
1791 al. 2019, Tsuji et al. 2020, Weitemier et al. 2021).

1792 Colonial aquatic invertebrates, particularly filter-feeding taxa, are
1793 especially well suited for exploring the integration of eDNA with population
1794 genetics. These organisms provide key ecosystem functions, including
1795 water filtration (Ostroumov 2005, Hamann and Blanke 2022) and nutrient
1796 cycling (Atkinson et al. 2013, Hopper et al. 2023) while also serving as

1797 bioindicators (Marshall and Stepien 2020) and hosts (Carraro et al. 2018)
1798 within broader ecological networks. Freshwater bryozoans are common
1799 species in rivers, they are sessile, clonal, and often patchily distributed,
1800 but frequently highly abundant. Reproduction is largely asexual by
1801 budding, resulting in large monoclonal stands that shed abundant eDNA
1802 for each genotype (Chapter 1). Furthermore, their distribution across
1803 interconnected aquatic habitats provides a natural framework for testing
1804 how environmental heterogeneity and hydrological connectivity shape
1805 population structure detectable through molecular means. Their life
1806 history characteristics and ecological roles thus make them both
1807 ecologically important and methodologically tractable targets for eDNA-
1808 based genetic analysis.

1809 The freshwater bryozoan *Fredericella sultana* (Blumenbach, 1779)
1810 exemplifies such a model bryozoan, and is especially abundant in rivers. *F.*
1811 *sultana* inhabits varied submerged substrates and is the definitive
1812 invertebrate host for the myxozoan parasite *Tetracapsuloides*
1813 *bryosalmonae* (Feist et al. 2001), the causative agent of proliferative kidney
1814 disease (PKD) in salmonid fishes. PKD poses a significant threat to both
1815 wild and farmed fish populations, with substantial ecological and
1816 economic impacts, as evidenced by mass mortality events such as the
1817 2016 Yellowstone River fish kill (Hutchins et al. 2018). *F. sultana* serves as
1818 a key reservoir host for the parasite, and therefore understanding its
1819 population genetic structure is crucial for understanding the (future)
1820 transmission dynamics of *T. bryosalmonae* (Carraro et al. 2017, Ros et al.
1821 2021). Many parasites exhibit local adaptation to host genotypes (Bellis et
1822 al. 2021), further emphasising the need for high-resolution genetic data on
1823 host populations. Host population genetics can indirectly reflect
1824 environmental gradients and demographic structure or directly influence
1825 co-infection probability and parasite fitness (Ivers and Jha 2023). Despite
1826 its important role in transmitting disease among wild and farmed fish,
1827 fundamental knowledge of *F. sultana* population biology, particularly

1828 genetic diversity, remains limited (Hartikainen and Jokela 2012, Filippenko
1829 et al. 2014).

1830 We predict that the hierarchical structure of river systems and the
1831 environmental heterogeneity they encompass directly influence the
1832 spatial distribution, connectivity, and persistence of *F. sultana*
1833 populations. However, before this can be assessed using eDNA, the
1834 reference population genetic diversity needs to be assessed. In this study,
1835 we investigate the population genetic structure of *F. sultana* within the
1836 River Erewash catchment (Nottinghamshire/Derbyshire, UK) to evaluate
1837 how spatial and environmental factors influence genetic diversity and
1838 connectivity. Widespread abundance of *F. sultana* has previously been
1839 confirmed in the River Erewash using surveys and qPCR (Chapter 1) and
1840 preliminary testing (unpublished) revealed intraspecific diversity between
1841 spatially close populations along the river. We assess whether a short
1842 mitochondrial marker is sufficient to resolve intraspecific genetic variation
1843 at a fine spatial scale, and whether matching patterns can be detected in
1844 eDNA.

1845 Building on our previous assembly of the complete *F. sultana*
1846 mitochondrial genome (Chapter 2), we examine congruence between
1847 haplotype clusters generated from nad3 and whole mitogenome data to
1848 inform whether short markers, much more readily detectable in eDNA, can
1849 be used to substitute for whole mitogenomes. We further assess whole
1850 mitogenomic variability among individuals. These analyses provide key
1851 information about the reference population genetic structure in the River
1852 Erewash, providing capacity for future assessment of population genetic
1853 structure via eDNA.

1854 Given the predominantly clonal reproduction in *F. sultana*, via budding and
1855 statoblasts (Bushnell and Rao 1974, Bilton et al. 2001, Hengherr and Schill
1856 2011), we hypothesise limited haplotype diversity and shallow population
1857 structure, with limited haplotype turnover across the river network. By
1858 integrating molecular data with ecological context, this work aims to

1859 inform our understanding of the role of bryozoans in freshwater parasite
 1860 ecology and provides the first study to track host populations across
 1861 aquatic landscapes using environmental DNA.

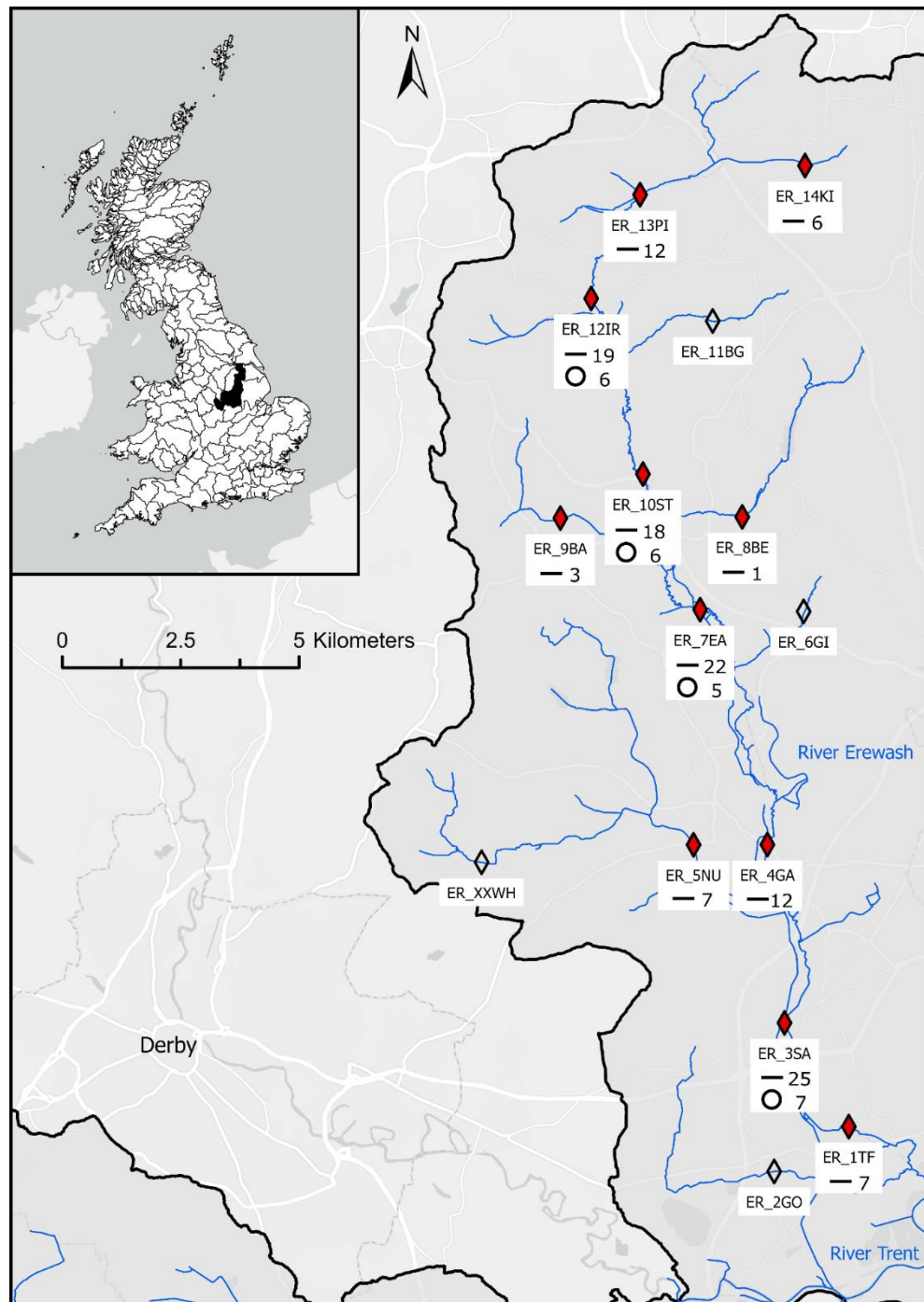


Figure 1: Map of all Erewash sampling sites (2021-2024). Empty diamonds: no bryozoans found by survey, red diamonds: bryozoans found. Lines below site names indicate number of nad3 Sanger sequences obtained from tissue and eDNA, circles indicate number of individuals sequenced for full mitogenomes.

Methods

1862 *Bryozoan and eDNA sampling*

1863 Individual *F. sultana* colonies were sampled from the River Erewash, UK,

1864 from 15 sites in July 2021 and 2024 (Fig. 1). Colonies were picked with

1865 forceps from the substrate in an approximately 10m long transect and

1866 stored in river water until extraction the next day. To ensure single

1867 genotypes were sampled for each extraction, unbroken branches, with at

1868 least 3 alive zooids, were selected. Previously extracted DNA from colonies

1869 collected in the same manner from five sites in Switzerland, and a site in

1870 Dorset, UK, were included as comparisons to the Erewash samples.

1871 Sampling locations, replication and sequencing method are indicated in

1872 Table 1. eDNA samples were collected at the same time as the tissue

1873 samples using method detailed in Chapter 1. In the Erewash, an additional

1874 eDNA sample set was collected in September 2021 exactly in the same

1875 locations as detailed for July 2021 (Table 1). Tissue samples collected in

1876 2021 were sequenced using amplicon approaches, as were the eDNA

1877 samples, and collections from 2024 were processed using whole

1878 mitogenome sequencing.

Table 1: Sampling sites on the River Erewash and collection dates, sample replication and locality data. Tissues indicated in bold were used in a whole mitogenome-based approach.

Site Code	Site Name	River	Country	Latitude	Longitude	Date (# colonies collected)	Nr. eDNA
4Urt	Urtene	Urt	Switzerland			July 2020 (1)	
8Gla	Glane	Gla	Switzerland			July 2020 (1)	-
9Cer	Cerne	Cer	UK			July 2020 (1)	-

Boi_Amont	Amont STEP	Boi-ron	Switzerla nd	46.49944	6.46536	July 2020 (10)	July 2020 (2)
Boi_Aval	Aval STEP	Boi-ron	Switzerla nd	46.49668	6.4701	July 2020 (1)	July 2020 (3)
Boi_Emboc hure	Emboch ure	Boi-ron	Switzerla nd	46.49384	6.47796	-	July 2020 (2)
ER01	Toton Fields	Ere- wash	England (UK)	52.90581 2	- 1.26603 546	July2021 (6)	July202 1 (1)
ER03	Sandiac re	Ere- wash	England (UK)	52.92562 3	- 1.28590 382	July2021 (18), July 2024 (7)	July202 1 (7)
ER04	Gallows	Ere- wash	England (UK)	52.95947 9	- 1.29072 211	July2021 (7)	July202 1 (5)
ER05	Nut Brook	Ere- wash	England (UK)	52.95957 4	- 1.31393 342	July2021 (4)	July202 1 (3)
ER07	Eastwo od	Ere- wash	England (UK)	53.00415 8	- 1.31115 581	July2021 (14), July 2024 (5)	July202 1 (8)
ER08	Beuvale Brook	Ere- wash	England (UK)	53.02166 2	- 1.29762 505	-	July202 1 (1)
ER09	Bailey Brook	Ere- wash	England (UK)	53.02176 1	- 1.35487 7	-	July202 1 (3)
ER10	Stoney	Ere- wash	England (UK)	53.02996 8	- 1.32875 258	July2021 (13), July 2024 (6)	July202 1 (5)
ER11	Bagthor pe Brook	Ere- wash	England (UK)	53.05887 9	- 1.30632 84	-	July202 1 (1)

ER12	Ironville	Ere-wash	England (UK)	53.06336 7	- 1.34455 136	July2021 (13), July 2024 (6)	July202 1 (6)
ER13	Pinxton	Ere-wash	England (UK)	53.08297 2	- 1.32883 9	July2021 (6)	July202 1 (6)
ER14	Kirkby	Ere-wash	England (UK)	53.08818 3	- 1.27674 526	July2021 (6)	-

1879 *DNA extractions*

1880 DNA from tissues was extracted after overnight lysis using a standard CTAB
1881 and chloroform-isoamyl alcohol (1:20) protocol following modifications
1882 detailed in (Tops and Okamura 2003). eDNA samples extracted as outlined
1883 in Chapter 1. 24 For mitogenome level analysis, tissue was extracted the
1884 using the Monarch® Spin Plasmid Miniprep Kit (New England Biolabs, USA)
1885 after pre-lysis treatment and sequenced on a MinION flow cell (R10.4
1886 chemistry: Oxford Nanopore Technologies, UK) with the Rapid Barcoding
1887 Kit 24 V14 as described in Chapter 2.

1888 *Full mitogenome analysis*

1889 DNA from 24 *F. sultana* tissue samples from 4 sites on the Erewash were
1890 processed for full mitogenome sequencing (Fig. 1). The barcoded library
1891 preparation and post processing methods are detailed in Chapter 2.

1892 The nanopore reads demultiplexed into BAM files using the dorado demux
1893 command and indexed with samtools index. Individuals with an average
1894 coverage over the mitogenome below 2.25x were discarded, this resulted
1895 in 17 individuals being available for downstream analysis.

1896 Allele variants were called using FreeBayes with settings: minimum
1897 coverage of 3 base reads, minimum alternate allele fraction of 0.2 and
1898 count of 2, haplotype length of 0 and no population priors. Variant Calling
1899 Files (VCFs) that contain SNPs, position and ratio information for all

1900 sequenced individuals were compressed with bgzip and samples renamed
1901 to match the BAM file before merging. Merged VCF was indexed with tabix
1902 and filtered with bcftools for quality and depth. The file was then converted
1903 to FASTA with bcftools consensus and aligned with mafft.

1904 Phylogenetic ML trees were calculated in MEGA11 and raxmlGUI 2.0.13
1905 after model testing (best model: GTR +G). Segregating sites, nucleotide
1906 diversity and Tajima's D were assessed with a sliding window (window size:
1907 250bp, shift: 100 bp) along the entire mitogenome. We used beagle to
1908 phase genotypes from the VCF file and converted the output to bed format.
1909 Eigenvectors for PCA were calculated with PLINK and visualized in R with
1910 ggplot2. Finally, haplotypes were determined as ancestry population
1911 proportions with ADMIXTURE. Testing showed that 3 populations resulted
1912 in the lowest CV error (0.48542). Ancestry proportions combined with the
1913 haplotype networks from pegas were visualized in R with ggplot2. Fst
1914 values were calculated using dnaSP after grouping individuals according
1915 to their majority ancestry proportions.

1916 *Primer design and selection for testing*

1917 Primers were designed using the 24 full mitochondrial genomes from the
1918 Erewash and all available mitochondrial sequences available from
1919 GenBank (Kumar et al. 2020b) as a reference alignment. Variable regions
1920 flanked by conserved primer sites were chosen, and primer synthesis
1921 conducted by IDT (Integrated DNA Technologies, B.V., Belgium). Target
1922 regions and primers were designed and verified via blast searcher to
1923 minimise non-target amplification. To ensure a practically feasible marker
1924 set, the primers were checked efficiency and amplification using SYBR
1925 qPCR on *F. sultana* genomic DNA at low concentration (conc.), and for off-
1926 target amplification (spurious bands) using eDNA samples. Only primer
1927 combinations with a strong single band at the expected size and low
1928 primer-dimer generation were chosen for the final short marker set
1929 analysed in-silico. PCR products were Sanger sequenced to ensure
1930 successful on target amplification. The amplicons (including primers) are

1931 220 to 417 bp long which makes them most suitable for 2x250bp
1932 sequencing on a MiSeq. The marker set covers 16S (4 primer pair
1933 combinations), 12S (2), cox3 (3), atp6 (2), cox1 (1), atp8 (1), nad3 (2), cytB
1934 (2), nad4 (1).

1935 The mitochondrial nad3 marker (forward primer:
1936 GGCCCTTGTAAGTCAAGAGTGA, reverse primer:
1937 ACTCCTTGCCAGCTCTTTGT) was chosen for testing in-situ using Sanger
1938 sequencing from tissue and eDNA samples. The remainder of the markers
1939 designed were not used in-situ due to time limitations.

1940 *Population genetics using nad3 haplotyping in Erewash*

1941 The mitochondrial nad3 primers were used in all *F. sultana* tissues and
1942 eDNA samples collected from the River Erewash and the outgroup
1943 comparison sites (Table 1). All PCRs used GoTaq® G2 Green Master Mix (5x)
1944 (Promega UK Ltd, UK) according to the manufacturer's recommendations:
1945 primers each 1 µM (final concentration), with the addition of 1 µl bovine
1946 serum albumen (BSA 1µg/ml) and 1 µl template DNA (tissue: ~100ng,
1947 eDNA: ~10-20ng). Cycling settings were an initial 2min denaturation at
1948 95°C, 45 cycles of 45s denaturation at 95°C, 30s annealing at 59°C (for
1949 nad3) and 1min extension at 72°C ,and a final extension of 5mins. For
1950 Sanger sequencing, PCR products were cleaned with PEG buffer and
1951 precipitation with 80% EtOH twice prior to Sanger sequencing in reverse
1952 directions with the downstream R primer by Eurofins Genomics Europe
1953 Shared Services GmbH (Germany). Chromatograms were checked for
1954 successful full length on target sequencing. Ambiguous bases were
1955 manually corrected.

1956 Alignment was generated in MEGA 11 with Muscle (Kumar et al. 2016). Only
1957 SNPs with two or more occurrences were considered. The first and last 20
1958 base pairs were trimmed to ensure equal length for all sequences. An
1959 evolutionary maximum-likelihood tree was generated using the Kimura 2-
1960 parameter model (Kimura 1980), an estimated gamma distribution and

1961 invariable sites rate in MEGA 11. A distance matrix was calculated from this
 1962 tree using R, factoextra and NbClust containing all tissue and eDNA
 1963 sequences obtained with the nad3 marker. A heatmap of the pairwise
 1964 distance from all samples was generated. Optimal number of clusters for
 1965 hierarchical k-means clustering was estimated with elbow, gap stats and
 1966 silhouette plots. Sequences were then sorted into haplotype groups
 1967 (clusters) according to their genetic distance. PCA and cluster assignment
 1968 were plotted. A dendrogram including cluster identities was produced as
 1969 well as a comparison of sequences from tissue and eDNA together with
 1970 their cluster. pegas 1.3 and haploNet functions were used to compute
 1971 haplotype groupings and networks for both marker and full mitogenome
 1972 sequences.

Results

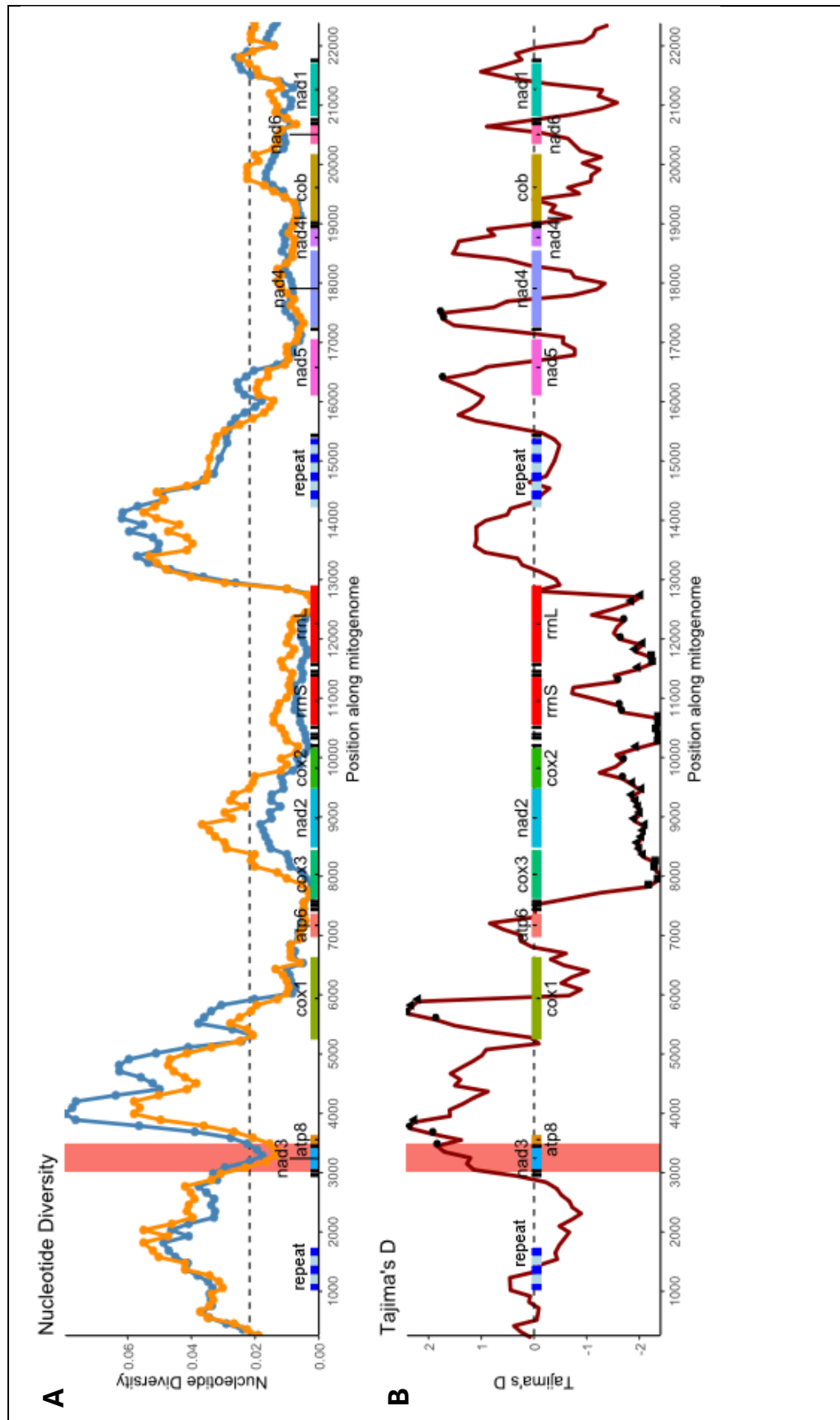
1973 *Full mitogenome*

Table 2: Comparative statistics of marker sequences from all and just mitogenome sequenced individuals and of the full mitogenomes.

Statistic	Full mitogenome	nad3 sequences of mt. seq. individuals	All nad3 sequences
Sequences	17	17	154
Total sites	21739	460	460
Segregating sites S	1609	20	63
Ratio of S/Total	0.074014	0.043478	0.136957
Average Distance	0.02157	0.017104	0.01955
Tajima's D (average)	-0.06373	1.305666	-0.7873

1974 Of the 24 *F. sultana* tissue samples across the Erewash, 17 individuals
 1975 produced sufficient coverage (> 2-3x) across the whole mitogenome.
 1976 Individuals are from a subset of samples used later for the nad3 marker
 1977 sequence set. They include four sites along the main river reaches,
 1978 downstream to upstream: ER03, ER07, ER10 and ER12 (Fig. 1). Whole
 1979 mitogenome sequences show overall comparable average distances

1980 between individuals (0.022 ± 0.002 , Table 2), but uneven nucleotide
1981 diversity along the mitogenome (Fig. 2). The mitogenome region with the
1982 highest nuclear diversity was the newly discovered group II intron and the
1983 non-coding mitocontrol regions (Chapter 2). On average, 7.4% of all sites
1984 across the mitochondrial genome are segregating (Fig. 2), with the lowest
1985 diversity found in the coding and functional regions such as *cox3*, *atp6*,
1986 *nad4*, rRNA and tRNAs. The most diverse genes are *cox1* and most of the
1987 *nad* genes (*nad1*, *nad2*, *nad3* and *nad5*). The rate of nucleotide diversity
1988 along the mitogenome regions ranged from $\pi = 0 - 0.090$ and $\Theta = 0.001-$
1989 0.058 . Similarly, Tajima's *D* ranged from -2.402 to 2.414 , indicating
1990 variation in selection strength across the mitochondrial genome (Fig. 2).
1991 Close to or significant negative Tajima's *D* can be found in the section
1992 between *cox3* and rRNA, indicating a dominance by rare variants. Group 2
1993 intron and the genes *cox1*, *nad5* and *nad4* showed positive values,
1994 indicative of maintenance of multiple alleles or loss of rare alleles and
1995 under balancing selection or a recent bottleneck.



↑ **Figure 2:** A) Nucleotide diversity and B) Tajima's D along the full mitogenome with π (blue), Θ (yellow) and Tajima's D (red), *nad3* marker region marked in red.

1996 The 17 full mitochondrial genomes were clustered based on hierarchical
 1997 clustering on genetic distances and according to their ancestral
 1998 population proportion based on the filtered VCF files (Fig. 3 & 4). The best
 1999 fitting hierarchical k-means clustering divided the full mitogenomes into
 2000 two clusters: Cluster 1 with all ER03, all ER12 and two ER07 individuals
 2001 (total 6) and Cluster 2 with one ER07 and all ER10 individuals (total 11) (Fig.
 2002 3). While there are no individuals with identical full mitochondrial genomes
 2003 (Fig. 4), they share ancestral origins (Fig. 3 & 4). Three potential ancestry
 2004 populations were identified: Pop1 containing one ER07 and four ER10
 2005 individuals, Pop2 with three ER03 and one ER10 and Pop3 with one ER03,
 2006 two ER07 and all ER12 individuals. The Fst value of Pop1 to Pop2 and Pop3
 2007 is higher than both are between
 2008 each other. Pop1, mostly ER10
 2009 individuals, is less varied with
 2010 number of SNPs between 69 to 114
 2011 in the haplotype network (Fig. 4). Biggest difference between two
 2012 individuals is 427 base changes. Distances in between individuals of Pop2
 2013 and Pop3 are as large as distance of both populations are to Pop1. Except
 2014 for individuals of ancestry population 1 the haplotype network is not
 2015 monophyletic. The ancestral populations Pop2 and Pop3 are split across
 2016 the hierarchical genetic distance tree.

Ancestry population		Fst
Pop1	Pop2	0.43189
Pop1	Pop3	0.33665
Pop2	Pop3	-0.03344

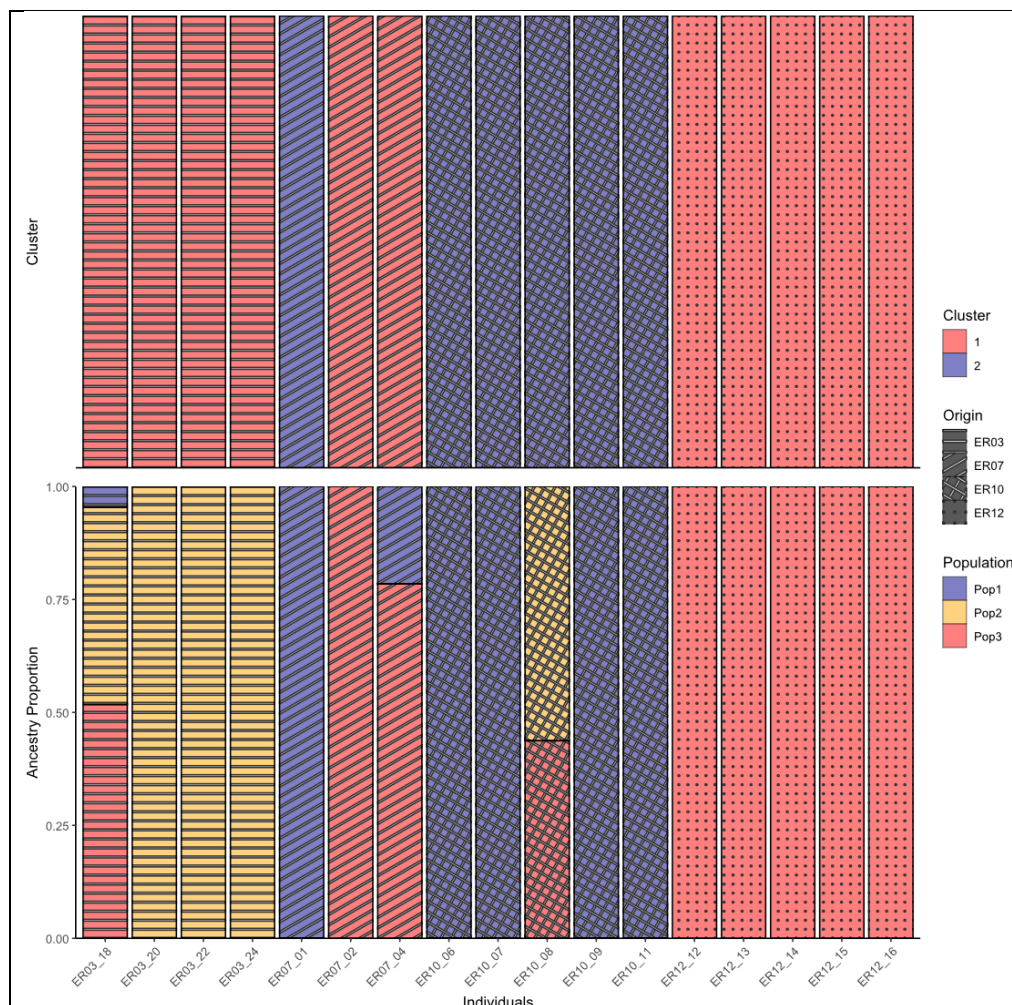


Figure 3: Cluster of nad3 marker region and full mitogenome ancestry proportions of sequenced individuals with column pattern for sites and colours corresponding to clusters or ancestry proportion.

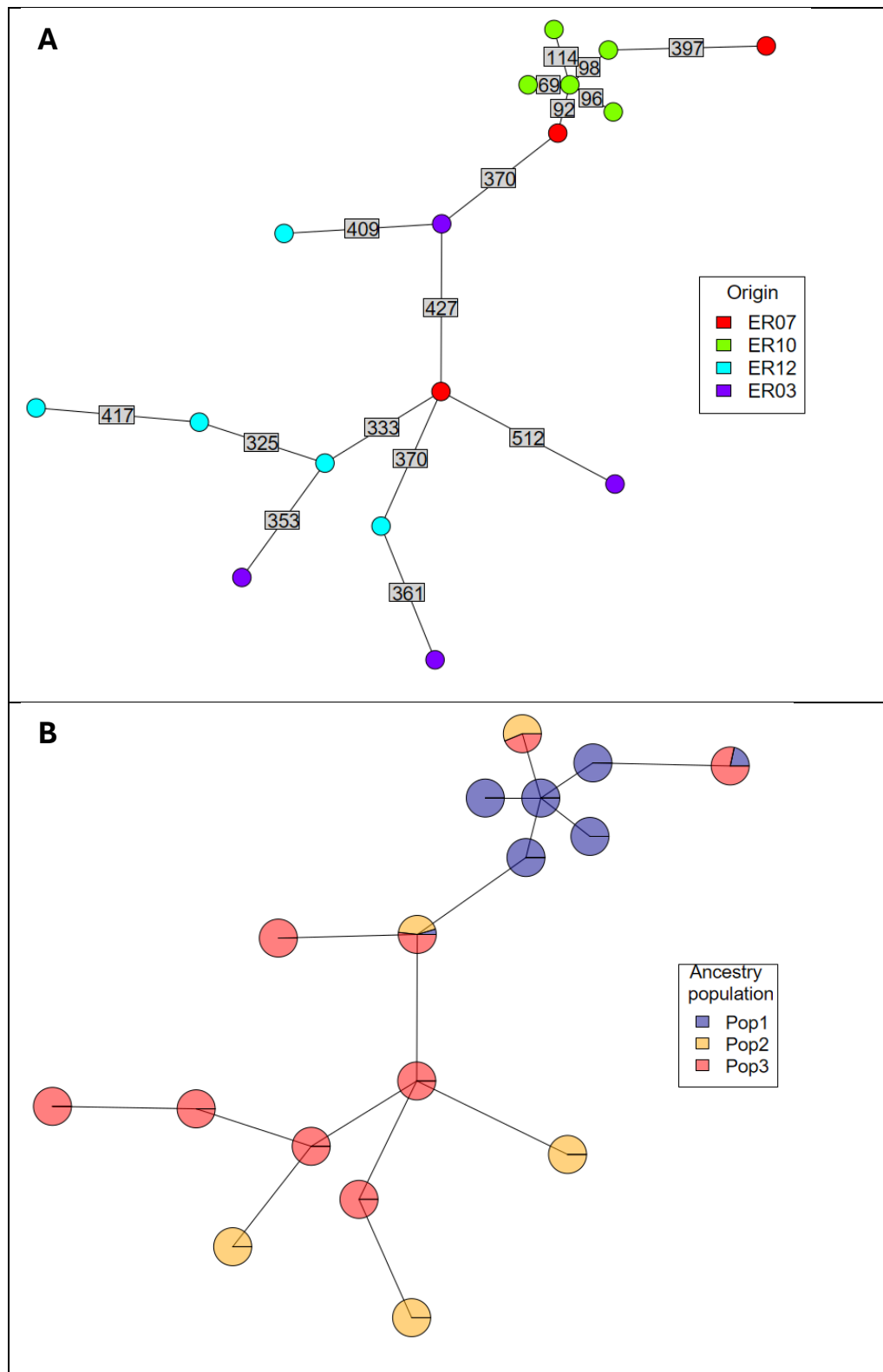


Figure 4: A) Haplotype network of full mitogenome distances with colours as sampling site origins and number of nucleotide changes in the grey boxes on the connecting branches, B) Haplotype network of all full mitogenome sequences with colours as ancestry population proportion.

2017 *nad3* marker

2018 A total of 154 *nad3* Sanger sequences were obtained from 87 UK and 14
2019 Swiss tissue samples and 46 UK and seven Swiss eDNA samples (Fig. 5).
2020 Two sequences were excluded as outliers, as they were outside the
2021 standard deviation of the overall data and lead to a drastic increase in
2022 average and maximal divergence. The average genetic distance based on
2023 the complete filtered alignment of *nad3* was 1.74%, with a range from
2024 0.00% for identical distance to the maximum of 5.68% for the most
2025 divergent. The most distance was between two individuals from ER03 to
2026 individuals from ER01 and Swiss eDNA samples. Most common pairwise
2027 distances are 0.0%, 1.4% and 4.3%, with no distance between identical
2028 individuals, median distance between individuals of the two largest
2029 haplotype groupings and high distances of the rarer, more distant
2030 individuals. The ratio of SNPs is 13.7% (Table 2), based on the ratio of
2031 segregating sites *S* compared to the total number of sites. Tajima's *D* of the
2032 marker sequence was negative with -0.79.

2033 Compared to the 7.4% of the full mitogenome segregating between the 17
2034 individuals, the *nad3* marker segregated on 13.7% nucleotide positions
2035 (Table 2). The average distance between the full length mitogenomes was
2036 2.2%, compared to 2.0% in *nad3*. The respective values for Tajima's *D* (on
2037 average) were -0.06 compared to -0.79. In comparison, the relative number
2038 of segregating sites *S* of the full mitogenome is ~70% higher than in just the
2039 *nad3* marker subset of the same, mitogenome sequenced individuals,
2040 however compared to the complete *nad3* marker set of all 154 individuals
2041 the observed ratio of *S* is ~45% lower (Table 2). Tajima's *D* on average is
2042 negative across all individuals sequenced for *nad3*, but positive in the *nad3*
2043 subset in the full mitogenomes only, and roughly neutral in the full
2044 mitogenome.

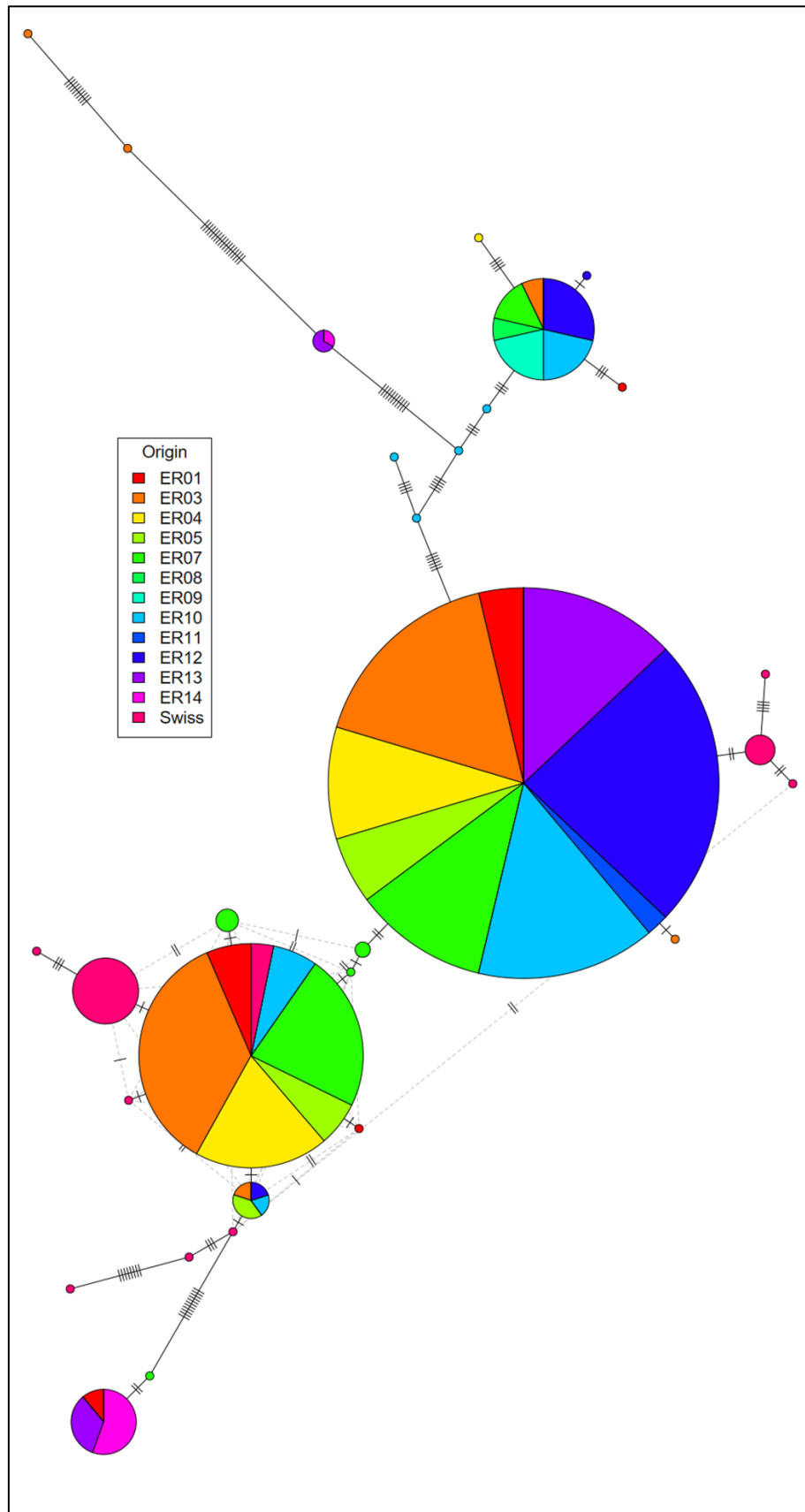


Figure 5: Haplotype network of all *nad3* marker sequences with continuous colours as site origin downstream to upstream of the River Erewash and the Swiss outgroup (including tissue and eDNA).

2045 Mapping the genetic distances of the nad3 marker sequences of all
2046 individuals reveal a complex haplotype network dominated by two closely
2047 related haplotypes made up of identical sequences and with little
2048 difference in between and a few close related sequences around it (Fig. 5).
2049 These haplotypes can be found in several sites along the river. Slight
2050 disparity in site distribution is apparent when comparing two groupings:
2051 The smaller one is dominated by downstream sites (ER03, ER04 and ER07),
2052 while the larger one has a high proportion of upstream sites (ER10, ER12
2053 and ER14) which are in part completely lacking from the other haplotype.
2054 Additionally, there are two minor haplotypes made up of mostly ER13 and
2055 ER14 or of ER09, ER10 and ER12 (excluding Swiss) with higher distance
2056 from all other sequences and several single, related but more diverse
2057 sequences. In general, there are few individuals with unique sequences,
2058 with a wide range of diversity or distance in them. ER10 is the most unique
2059 and diverse site, together with ER14 which is also split across the
2060 haplotype network. Swiss samples form no distinct, separate haplotype,
2061 but they are closely connected (mostly between 1-6 base changes) to
2062 either one of the two dominant types. Individuals with the highest genetic
2063 distance to others are two individuals from ER03 and several upstream
2064 individuals from ER13 and ER14.

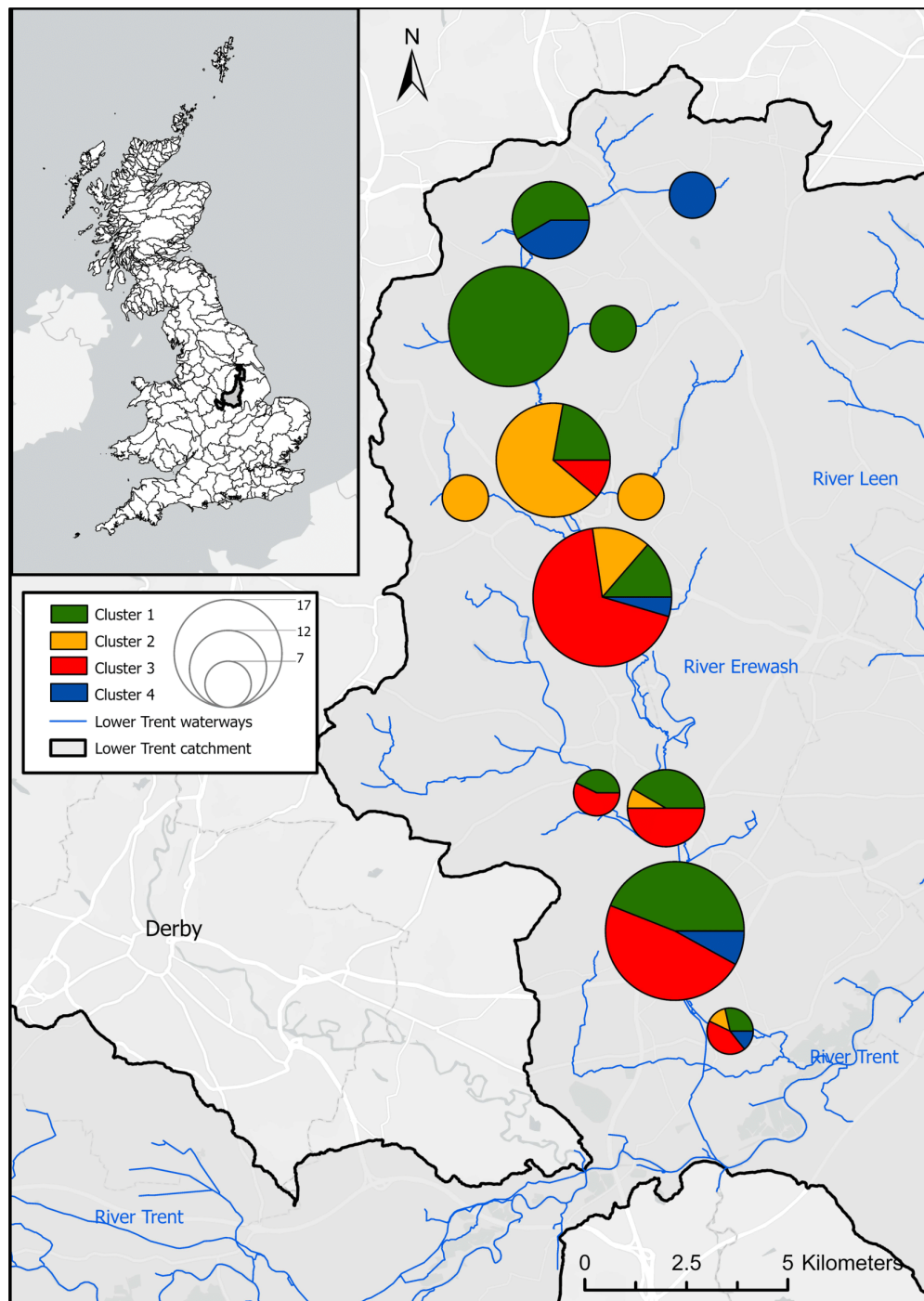


Figure 6: Map of the River Erewash with pie charts of the combined tissue and eDNA haplotype clusters (colour) of the *nad3* marker at the sampling sites.

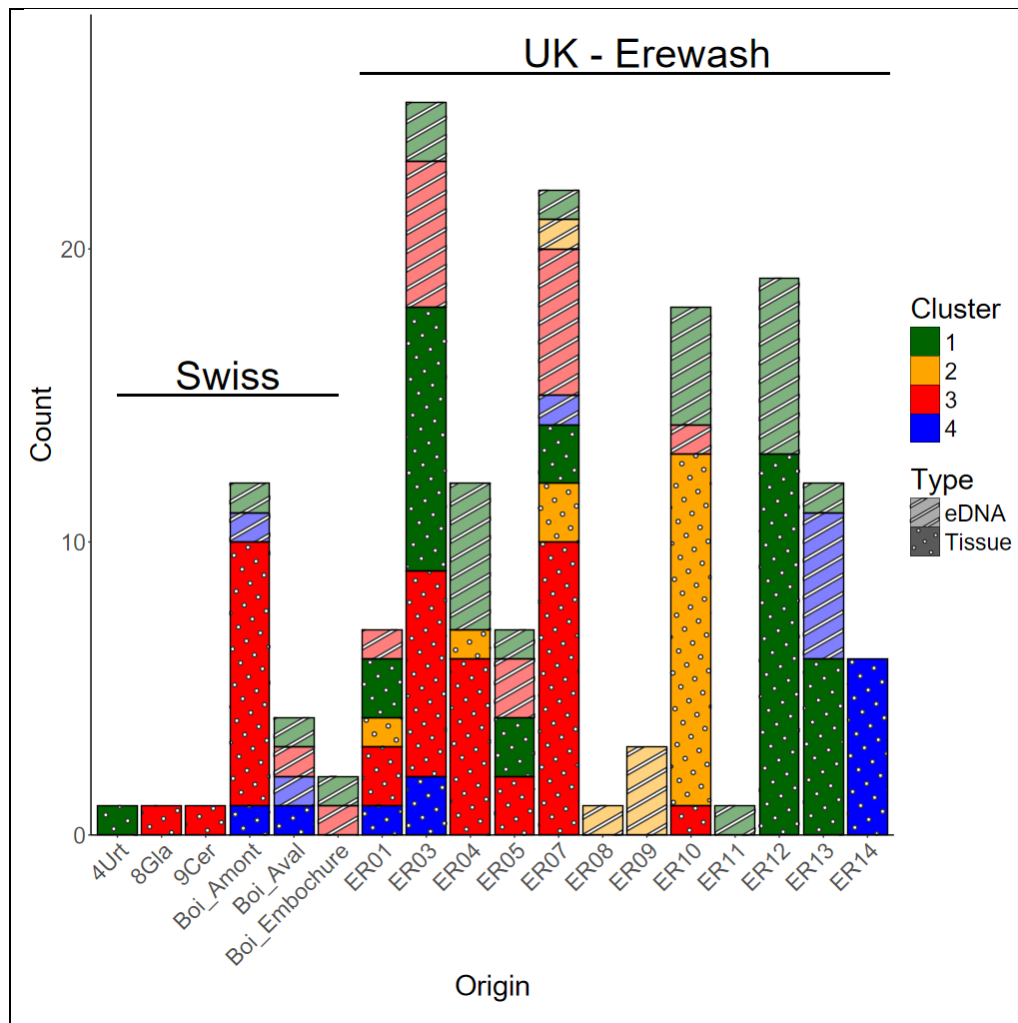


Figure 7: Comparison of cluster assignments (colours) of tissue (dots) and eDNA (lines) samples for all tested sites (River Erewash, UK and Swiss outgroup) with the number of sequences for each site on the left-side scale.

Haplotypes are not exclusive to single sites and can be detected in various, non-neighbouring regions along the river. Not counting singeltons 10 unique haplotypes (defined through unique SNPs not found in other sequences) were compiled into four haplogroup clusters (Fig. 6). These clusters are shared between the Swiss (outgroup) and UK (Erewash) samples and between tissue and eDNA. Tissue and eDNA samples show differences but also similarities: Overall comparable amounts of tissue and eDNA sequences have been assigned to each cluster, however the picture is more complex when looking at individual sites (Fig. 6 and 7). Cluster 3 is found mostly downstream in both (tissue: 34.7% of sequences, eDNA 45.3% and total: 38.3%). Cluster 1 could be observed on most sites

2076 (total: 35.7%). Cluster 2 and 4 are rare (13.6% and 12.3%), 4 is most
2077 dominant in the most upstream sites closest to the source (ER13 and
2078 ER14) while Cluster 2 was most often detected in the mid regions (ER08,
2079 ER09 and ER10) Amount of tissue found through survey and eDNA was low
2080 in the side streams (ER05, ER09 and ER11). It was high to intermediate in
2081 the main river reaches (ER03, ER07 and ER04, ER10, ER12, ER13)
2082 Collection of sequences was overall achieved to a lesser degree in eDNA.
2083 In a few cases eDNA detected haplotypes where no tissue samples were
2084 gathered (side streams ER09 and ER11). The tissue and eDNA results
2085 notably differ from each other for the sites ER04 and ER10: Cluster 1 was
2086 found in both sites in eDNA only.

Discussion

2087 *Population genetics methods (sequencing overview)*

2088 Marker-based approaches using short DNA sequences are valued for their
2089 simplicity and the ability to work for low quantity and quality samples
2090 (Turon et al. 2020). They are suitable used for mixed samples including
2091 environmental DNA (Elbrecht et al. 2018, Parsons et al. 2018, Weitemier et
2092 al. 2021) and are generally transferable to RNA sequencing (Laroche et al.
2093 2018, Von Ammon et al. 2019, Miyata et al. 2022). We assessed if a marker
2094 sequencing approach is sufficient to cover the complexity of population
2095 genetics as several restrictions apply: A marker amplicon relies on a known
2096 reference sequence and efficient primers (Elbrecht and Leese 2017,
2097 Elbrecht et al. 2017, Keck et al. 2022). They need to balance size limitations
2098 against genetic divergence (Weitemier et al. 2021). Several markers were
2099 tested. Multiple suitable gene loci were identified: nad1, nad2/cox3, nad2,
2100 nad3, nad4, atp6 (2) and atp8 (Table S1). We decided to focus on the nad3
2101 marker as it consistently showed the highest genetic diversity between
2102 sequences and proved to amplify and sequence most reliably overall. We
2103 briefly explored if our nad3 marker is suitable for eDNA as well (Fig. 5 and
2104 6). Using Sanger sequencing we assessed the potential and efficiency of
2105 our primers in eDNA samples in preliminary testing and did not expect to

2106 retrieve (quantifiable) data about all haplotypes as found in tissue also
2107 from eDNA. However, while getting sequences from the most prominent
2108 haplotypes on the sites (as expected), we also gathered sequences for the
2109 rarer haplotypes. Through Sanger sequencing of eDNA samples no
2110 additional haplotypes were identified. Groupings from eDNA and tissue
2111 matched. The proportions of haplotypes from eDNA samples are likely not
2112 appropriately reflecting those present at the sites but allow us to obtain
2113 presence and absence data of haplotypes in the eDNA samples from those
2114 sites. This shows that short markers, also suitable for eDNA, could be used
2115 for population genetics on a particular small scale, with distance between
2116 sites of around 2-4km. We can delimit haplotypes that allow us to separate
2117 communities on a 40km long river. Certain regions of the mitochondrial
2118 genome are variable enough to infer population structure, among them
2119 nad3.

2120 While a single marker allowed us to build a population network with several
2121 distinct haplotypes (Fig. 5), it also exemplifies the limitation of a confined
2122 number of SNPs compared to full mitochondrial sequences. Joining and
2123 linkage of fragmented SNPs across several markers is near impossible
2124 which hinders the ability to trace complex ancestry and reconstruct
2125 population structure with markers alone. None of sequenced individuals
2126 showed identical complete mitogenome sequences, but they mostly
2127 collapsed into identical groupings when only regarding the nad3 marker
2128 (Fig. 3). This highlights the need for a deeper understanding of
2129 mitochondrial genetic variance across a wider spatial and temporal scale
2130 and the need to build a reference base to choose appropriate markers for
2131 future studies.

2132 Recent Nanopore technology advances lower the error far enough to make
2133 the here produced low coverage sequencing viable (Chen et al. 2021,
2134 Baeza and García-De León 2022, Urban et al. 2022, Kolmogorov et al.
2135 2023). Three to four times average coverage of the mitogenome allowed
2136 reliable determination of SNPs when using a conservative approach

2137 excluding singletons (= SNPs only occurring in one individual). Our DNA
2138 extraction and Nanopore sequencing approach works well, even for tissue
2139 samples of low quantities and quality (see Chapter 2 for details). A single
2140 sample can be processed from raw tissue to basecalled and de-novo
2141 assembled sequences in three to four hours. 24 individual samples (of
2142 different species) can be fully analysed in a single working day. The whole
2143 process could be done in the field with limited equipment. We obtained 17
2144 usable full mitogenome sequences from a total of 24 samples. The main
2145 obstacle was microbial contamination that reduced the amount of on-
2146 target (Fs) sequences.

2147 *Haplotyping along the river*

2148 The nucleotide diversity along the entire mitogenome (Fig. 2) was higher
2149 than initially expected, but still comparable to previously reported
2150 maximum regional diversity of up to 4% (Hatton-Ellis et al. 1998, Freeland
2151 et al. 2000). Most diverse are the non-coding regions and the group 2 intron
2152 while genes, rRNAs and tRNAs are less diverse. The genes of the
2153 mitogenome encode for important proteins of the oxidative
2154 phosphorylation complexes that produce energy through respiration.
2155 Ribosomal and transfer RNA are also vital for cell function. Their lower
2156 genetic variance can be explained through sequence conservation and
2157 restrictive selection on mutations which limits the number of
2158 (disadvantageous) nucleotide polymorphisms. In genes most SNPs are
2159 expected to be limited to the 3rd codon position and be synonymous base
2160 changes with no change to the amino acid sequence.

2161 The nad3 marker is mainly protein coding with tRNAs at the ends.
2162 Consequently, SNPs predominantly occur in the 3rd codon position (data
2163 not shown). Most sequences show a very low diversity and SNP rates, with
2164 any base changes shared among multiple individuals and only few
2165 singletons (Table 2). Therefore, the haplotype population network is
2166 founded in closely related groupings (Fig. 5). Rare, partly more distant
2167 groupings and individuals branch from the two dominant types. The Swiss

2168 individuals do not form an outgroup to the Erewash individuals but are
2169 interlaced with them. However, the genetic diversity is based on the single,
2170 short mitochondrial marker.

2171 Since the full mitogenome sequenced belong to the two closely related
2172 major haplotypes (Fig. 3) their Tajima's D is positive (Table 2) as their nad3
2173 marker sequences are distributed between the haplotypes and lack any of
2174 the more rare and divergent SNPs. Considering all 154 marker sequences,
2175 Tajima's D is negative due to the higher number of rare variants, however
2176 overall non-significant. Across the entire mitogenome Tajima's D is non-
2177 uniform, most changes in value are on gene borders and in coding regions
2178 (Fig. 2). The mitogenome is monochromosomal without recombination
2179 and of assumed uniparental inheritance. While Tajima's D is non-zero on
2180 most sites, there is significant evidence of non-random selection on a
2181 handful of sites. Variation seems to only accumulate in permissible
2182 regions. Notable areas are the repeats of the non-coding, suspected
2183 mitocontrol regions where values are neutral (around 0), in group II intron,
2184 cox and nad4 they are significantly positive ($D > 0$, rare alleles scarce) and
2185 in cox3 to rrnL they are significant negative values ($D < 0$, rare alleles
2186 abundant). Significant values indicate population expansion after
2187 bottleneck if $D < 0$ and balancing selection or population contraction for
2188 $D > 0$ (Tajima 1989). While no clear conclusion can be drawn, regions of the
2189 mitogenome accumulate SNPs at different rates and have distinct
2190 variability. We included the most variable regions in our marker set.

2191 Two groupings of pairwise genetic distances are apparent in the marker
2192 and full mitogenome sequencing data (when ignoring identical sequences
2193 with distance = 0%). Similar individuals have a genetic difference of about
2194 1% and more divergent ones have a difference of 3-4%. 1% distances are
2195 more common in the marker approach while distances are higher overall
2196 with a quarter being identical. The full mitogenome comparison does not
2197 have identical sequences (except one) but is also split into similar two
2198 groupings.

2199 This supports the theory that there are distinct potentially genetically
2200 isolated populations forming the local communities. These populations
2201 have low internal genetic diversity but substantial divergence from other
2202 populations. We can detect and define these as haplotypes. Intra-
2203 haplotype diversity is fairly low (below 2%) but still obvious in both marker
2204 and full mitogenome approaches (Table 2), however the number of
2205 identical sequences is high in the marker sequences. Inter-haplotype is
2206 higher (above 3%) than what is commonly considered normal intraspecific
2207 diversity (98% shared identity and 2% difference) (Gómez-Rodríguez et al.
2208 2017). The mitochondrial diversity is likely not representative of the full
2209 genetic diversity including the nuclear chromosome.

2210 Haplotypes disperse along the whole river (Fig. 5 and 6). Haplotype groups
2211 are not restricted to single site or to neighbouring sites. This indicates that
2212 local population structure of bryozoans in the Erewash takes the form of a
2213 metapopulations consisting of inter-connected, locally confined
2214 populations (demes). Migration and gene flow between sites and
2215 populations are likely present, but no firm assessment can be made with
2216 the data set presented here. Since two major haplotypes dominate, the
2217 spread of clonal individuals through statoblasts and budding might have
2218 shaped the initial population genetic structure reinforced over time
2219 through the founder effect (Templeton 1980, Barton and Charlesworth
2220 1984). Both marker and full mitogenome approach result in good
2221 resolution of different populations along the river (Fig. 4 and 6). The
2222 linearity of the river due to constant transport of water and material
2223 downstream is partly reflected in the haplotype distribution pattern.
2224 Cluster 3 was most prominent in the lower downstream areas, Cluster 2 in
2225 the intermediate regions and Cluster 4 on the most upstream sites (Fig. 6).
2226 However, there are breaks in these patterns. Most obviously Cluster 1 is
2227 found on almost all sites and haplotypes most upstream where also found
2228 further downstream and vice versa. The source for tissue and eDNA differs,
2229 eDNA could be brought in from areas not covered in the traditional

2230 surveying and sampling. Different dispersal mechanism besides transport
 2231 with water current are known, for example by waterfowl, fish or humans
 2232 (Bilton et al. 2001, Figuerola et al. 2005, Okamura et al. 2019). Bryozoans
 2233 likely show a strong founder effect which is enhanced by asexual
 2234 reproduction (budding) and deposition of long-lasting spore banks into the
 2235 local river environment (Hatton-Ellis et al. 1998). This could lead to stability
 2236 of populations over long time periods despite the everchanging river
 2237 environment and seasonality effects (Freeland et al. 2001). Long term
 2238 observations are needed to confirm this. Stable metapopulation structure
 2239 with genetically distant local population could lead the local adaptations
 2240 of parasites to hosts (Bellis et al. 2021). Genetic isolation of bryozoan
 2241 demes could be a barrier to parasite spread. However, one haplotype was
 2242 detected a long the full stretch of the river. This could be an initial
 2243 colonising, highly clonal strain which could aid the spread of parasites in
 2244 the water way. Genetically very similar or even identical haplotypes (3 and
 2245 4) were the most commonly found in this study. Thus far the ecology of
 2246 bryozoans is only theoretically well understood, case studies are rare.

2247 The mitogenome has no recombination and (presumed) uniparental
 2248 inheritance. Limited sexual reproduction and propagation mainly
 2249 occurring clonally through budding and statoblasts could lead to the
 2250 domination of single or a limited number of (mitochondrial) haplotypes.
 2251 There can be considerable genetic turnover between sites and over time.
 2252 Founder effects in combination with propagule banks leads to an interplay
 2253 of population genetics with seasonality, reproductive mode, bottlenecks
 2254 and recolonisation (Freeland et al. 2001). Some bryozoans have
 2255 propagules that are mobile and able to cover long distances along with the
 2256 current, like the flotation assisted statoblasts of *Plumatella repens*,
 2257 *Pectinatella magnifica* and *Cristatella mucedo* (Bushnell and Rao 1974,
 2258 Okamura et al. 2019). Wider distribution by fish or waterfowl is also known
 2259 (Okamura et al. 2019). *Fredericella sultana* has a limited dispersal

2260 potential through rudimentary statoblasts and migrating zooids
 2261 (Gorgoglione et al. 2016).

2262 Bryozoan communities as a whole are stable against local extinction and
 2263 resistant to changing or unfavourable conditions but genetically unstable
 2264 due to changing haplotype presence and proportions (Hatton-Ellis et al.
 2265 1998, Freeland et al. 2001). Genetic diversity and genetic turnover are
 2266 defined by metapopulation structure and influenced by dispersal and
 2267 endurance mechanisms. Here we uncovered a metapopulation structure
 2268 with haplotype makeup independent of distance and limited apparent
 2269 gene flow between neighbouring sites. Haplotype distribution is congruent
 2270 with those postulated in literature (see above). *F. sultana* is found in most
 2271 reaches of the river system. A strong propagule-reinforced founder effect
 2272 leads to the domination of two widespread haplotypes with little internal
 2273 diversity. Some sites are characterized by less frequent haplotypes.

2274 Sites show interconnectivity and gene flow in the sense that no haplotype
 2275 is exclusive to single sites but can be found across the river in non-
 2276 neighbouring sites (Fig. 6). In the smaller sampling size full genome
 2277 sequencing we found higher differentiation between individuals compared
 2278 to the marker approach. The groupings in haplotype cluster and ancestral
 2279 population generally match, however individual ER07_04 was assigned to
 2280 different groups and marker cluster 2 is subdivided into two different
 2281 ancestral populations (Fig. 3). This highlights the limitations of short gene-
 2282 based markers, they do not reveal less genetic diversity.

2283 Riverine systems predominantly unidirectional and isolated upstream
 2284 branches join into mixed downstream confluences (Carraro et al. 2020a).
 2285 One could expect that the branches show genetically distinct haplotypes
 2286 that are unique in the individual upstream reaches and that downstream a
 2287 mixture of haplotypes can be found or that, alternatively, one of the
 2288 upstream haplotypes dominates downstream. The marker approach
 2289 showed a generally matching pattern in our data (Fig. 6). Downstream sites
 2290 (ER01 to ER07) have a mixture of haplotypes (nestedness) but are

2291 dominated by one. Upstream sites (ER08 to ER14) have less shared and
2292 more confined haplotypes. The full mitogenome sequencing which was
2293 more limited in number of sites and individuals matches the marker
2294 approach but did not show a clear population pattern. Individuals of the
2295 same site tend to be more closely related indicating that geography
2296 impacts the ecology of *F. sultana*. The considerable variation between
2297 individuals might accumulate over time in a predominantly asexually
2298 reproducing species.

2299 *Implication for host and parasite systems*

2300 Thus far the implication of the host population structure for the parasite
2301 ecology and overall epidemiology of the system, including its
2302 consequences for the salmonid fish hosts, can only be speculated. We
2303 currently lack the connection between infection rates and population
2304 genetics of the host(s). It is generally known that parasites exhibit host
2305 specificity stemming from local adaptations (Bellis et al. 2021). A few
2306 studies have begun to shine light onto the immune response of bryozoan
2307 and fish hosts to infection with the myxozoan (Kumar et al. 2015, Sudhagar
2308 et al. 2019, Bailey et al. 2020, Kumar et al. 2020b, c, Sudhagar et al. 2020,
2309 Kumar et al. 2021), but we currently lack a profound understanding to
2310 assess potential implications. It is assumed that the parasite might spread
2311 faster through highly clonal bryozoan communities and that more diverse,
2312 sexually reproducing populations are not as vulnerable. This is
2313 complicated by the fact that bryozoans can regenerate from propagule
2314 (statoblast) banks and vertical parasite transmission through statoblasts
2315 or colony fragmentation containing parasites (Abd-Elfattah et al. 2014).
2316 Locally successful established bryozoan strains should deposit more
2317 statoblasts than freshly arriving individuals and therefore strengthen the
2318 presence of already established types. A pattern which could be deduced
2319 from the data shown in this study. In theory an island or metapopulation
2320 structure of bryozoan in the environment is most likely (van Nouhuys
2321 2009). The parasite must overcome the geographic and genetic gap

2322 between these local populations (demes). Presently we do not know the
2323 ecological and evolutionary mechanisms and the broader consequences
2324 behind this.

2325 *Next-generation sequencing approaches*

2326 This study is the first to investigate the variability of the full mitogenome of
2327 *F. sultana* and could be a stepping stone to assemble a reference database
2328 for its population genetics. We started by designing a set of short markers.
2329 In future medium to long range amplification including whole mitogenome
2330 amplification or selective sequencing based on our findings could be
2331 possible. In recent years Oxford Nanopore sequencing has been improved
2332 and new possibilities, which might be valuable for eDNA, have emerged
2333 (Loose et al. 2016, Giesselmann et al. 2019, Doorenspleet et al. 2021,
2334 Egeter et al. 2022, Urban et al. 2022). In a recent study (Urban et al. 2022)
2335 it was shown that selective Nanopore sequencing can be used to identify
2336 animals through haplotype level information gained from soil samples. It is
2337 dependent on sufficient genetic material and good sequencing design to
2338 obtain enough information from eDNA to delineate individuals. To date,
2339 most eDNA studies use NGS with marker and adaptor sets and several
2340 rounds of amplification such as in MiSeq (Elbrecht and Leese 2017, Dugal
2341 et al. 2021, Weitemier et al. 2021). This allows for high sensitivity but limits
2342 the obtained sequences to a few genomic regions. Most researchers use
2343 established metabarcoding markers and they show some success.
2344 Metabarcoding from tissue using similar techniques have been
2345 established for decades and provides good references for such eDNA
2346 studies. The capabilities of MiSeq for population genetics has been
2347 demonstrated and haplotypes can be delimited (Sigsgaard et al. 2017a,
2348 Tsuji et al. 2018, Weitemier et al. 2021). They are generated from combined
2349 multi-loci data but the linkage between those loci is hard to estimate.
2350 Usually not enough genetic regions are covered, and biases not easily
2351 assessed to reliably measure allele frequencies across the genome.

Conclusion

2352 Amplification-free direct sequencing and latest Nanopore technology
2353 uncovered the genetic diversity along the entire mitogenome of *F. sultana*.
2354 We found substantial, but not uniform diversity along the mitogenome with
2355 large differences between different genetic regions. Full mitogenome
2356 sequencing has ample resolution to distinguish individual bryozoan
2357 colonies on the small geographic scale of a single river network.
2358 Haplotyping of the full mitogenome agreed with the extended sequencing
2359 of the short nad3 single gene marker based on tissue from colonies as well
2360 as eDNA, while revealing additional population genetic diversity. Based on
2361 these findings we designed a set of short markers covering numerous
2362 mitochondrial gene regions that is suitable for population genetics directly
2363 from sampled colonies or from eDNA.

Chapter 4: *Ecological similarities of fish, invertebrate and parasite communities: concordance highlighted by eDNA, conventional count data and ecological traits*

KEYWORDS: parasite eDNA, myxozoa, host communities, biodiversity, ecological traits, concordance patterns

Abstract

2364 Similarities between local communities reflect shared ecology.
2365 Concordantly, parasites are shaped by but also impact communities of
2366 ecosystems they exist in. Using eDNA metabarcoding of myxozoan
2367 parasites, extensive conventional count data of freshwater fish and
2368 (macro-)invertebrates together with ecological traits of these species we
2369 compare local community biodiversity of several river systems to
2370 understand if we can trace host and parasite co-occurrence through
2371 shared ecology without the need to directly assess parasitic host
2372 dependency and infection rates. Classifying ecosystems through
2373 combined ecological traits rather than simple species abundance or
2374 composition overcomes shortcomings of the conventional count data in
2375 taxonomic identification and without losing differentiation between sites.
2376 Through pairwise comparison of species and trait abundance and
2377 composition of parasites, fish and invertebrates we derive concordance
2378 patterns. Fish and invertebrates show good to strongly significant
2379 similarities in community structure, while parasites are only weakly
2380 concordant in the comparisons. Parasite ecology might partly mirror fish
2381 and invertebrate biodiversity but is likely overshadowed by their complex
2382 life history and unresolved host and parasite dependencies.

Introduction

Communities of organisms are structured and filtered by both biotic interactions and abiotic environmental conditions (Leach et al. 2016). In riverine ecosystems, species distributions are directly influenced by local environmental variables such as flow velocity, substrate type or riparian shading (Ward 1998, Tockner and Ward 1999, Altermatt et al. 2013), and indirectly by ecosystem-level processes like nutrient cycling and trophic interactions (Jacquet et al. 2022, Talluto et al. 2024). These environmental drivers shape local community composition, leading to variation in biodiversity across sites. Such variation can be described by concepts like species turnover, where species composition changes between locations, or nestedness, where less diverse communities represent subsets of more diverse ones (Altermatt 2013, Altermatt et al. 2020). While these patterns are often examined along continuous spatial gradients, such as up- to downstream sites (Carraro et al. 2023), they can also emerge across environmentally distinct but spatially unconnected sites.

Parasites with complex life cycles, such as those within the Phylum Myxozoa, rely on multiple host species to complete their development (Holzer et al. 2018, Naldoni et al. 2019). This ecological dependency suggests that parasite distributions may be shaped by the composition of both host communities and the surrounding environment. When community structure of parasites and hosts varies in a similar way across landscapes, this is referred to as community concordance. Concordant patterns may result from shared responses to environmental filtering, overlapping ecological niches, or the constraints of host-parasite specificity (Blasco-Costa et al. 2013, Berkhout et al. 2020, Luo et al. 2022). Studying concordance among taxonomic groups can therefore provide insights into the ecological and evolutionary processes that drive community assembly, particularly in systems where direct observation of interactions is limited.

2412 Comparative analyses of parasite, fish and invertebrate communities are
2413 relatively rare, largely due to the challenges of detecting internal parasites
2414 and the limited scope of conventional monitoring approaches (Hutchins et
2415 al. 2021, Duval et al. 2024). This is especially true for parasites with
2416 complex life cycles that involve multiple host types. Environmental DNA
2417 metabarcoding now provides a way to detect these parasites in aquatic
2418 systems without direct observation. eDNA approaches have been used to
2419 identify a wide range of parasites across diverse host groups, offering new
2420 opportunities to explore their distribution patterns and ecological
2421 relationships (Huver et al. 2015, Hartikainen et al. 2016, Carraro et al.
2422 2017, Rusch et al. 2018, Marshall et al. 2022, Oredalen et al. 2022).

2423 This study focuses on myxozoan parasites, which infect a wide range of fish
2424 species including cyprinids, eel, sticklebacks and salmonids, as well as
2425 invertebrates such as annelids and freshwater bryozoans (Holzer et al.
2426 2018, Naldoni et al. 2019). Most myxozoans have a two-stage life cycle
2427 which alternates between fish and an invertebrate host and are found
2428 across many freshwater systems (Holzer et al. 2018). Although they are a
2429 natural part of aquatic ecosystems, some species can become harmful
2430 under certain conditions (Grabner and El-Matbouli 2009, Svavarsdóttir
2431 2016), such as high host density (Duval et al. 2024) or elevated
2432 temperatures (Bettge et al. 2009a, Bettge et al. 2009b, Strepparava et al.
2433 2018). For example, *Tetracapsuloides bryosalmonae*, the agent of
2434 Proliferative Kidney Disease (PKD), can cause high mortality in salmonid
2435 populations (Hutchins et al. 2021).

2436 I analysed the community composition of fish, invertebrate and myxozoan
2437 parasites at 13 freshwater sites across the UK, each characterised by
2438 different environmental conditions and catchment features. Fish and
2439 invertebrate data were collected from the decades long databases
2440 documenting the Environment Agency (EA) surveys which employ
2441 electrofishing of fish and kick net sampling of macroinvertebrates. Due to
2442 limited taxonomic resolution in the survey data, we used an ecological

2443 trait database (freshwaterecology.info) to group species by shared
2444 characteristics, such as feeding strategy, habitat preference and life history
2445 traits. This approach reduces the impact of taxonomic uncertainty and
2446 allows us to investigate whether ecological traits explain community
2447 patterns across sites.

2448 The aim of this study is to evaluate the degree of concordance between
2449 host and parasite communities across environmentally variable riverine
2450 sites. We investigate whether fish and invertebrate communities show
2451 similar spatial patterns in composition, which would suggest that they are
2452 shaped by shared environmental drivers (Grunberg 2021). We then
2453 examine whether myxozoan parasite communities, detected through
2454 eDNA metabarcoding, exhibit distribution patterns that align with those of
2455 their hosts, particularly fish, indicating possible ecological or
2456 transmission-based dependencies.

2457 In addition, we assess the extent to which environmental variables explain
2458 variation in ecological trait-based classifications of host and parasite
2459 communities across sites. If host traits are predictive of parasite
2460 community composition, this would suggest that ecological function,
2461 rather than taxonomic identity, plays a key role in shaping parasite
2462 distributions. Such a finding would highlight the potential of trait-based
2463 approaches for inferring parasite presence in systems where direct
2464 observation or full host data is limited.

2465 We hypothesise that fish and invertebrate communities exhibit
2466 concordant spatial patterns in their composition, reflecting shared
2467 responses to environmental gradients such as flow rate, substrate and
2468 habitat complexity. Because myxozoan parasites depend on their hosts for
2469 the completion of their life cycles, we further hypothesise that parasite
2470 community composition will be concordant with host community
2471 structure, particularly that of fish. This would indicate that parasite
2472 distributions are influenced by the presence, abundance or ecological
2473 traits of host species. Finally, we hypothesise that trait-based

2474 representations of host communities will effectively predict patterns of
2475 parasite distribution, suggesting that ecological function and host traits
2476 play a key role in shaping parasite communities. This would indicate that
2477 parasite presence can be inferred from host ecological characteristics
2478 even when species-level identification is limited or incomplete.

Methods

2479 *Collection of eDNA and parasite metabarcoding*

2480 Multiple sites on 13 different rivers and streams from eight catchments
2481 across the United Kingdom were sampled (Fig. 1, Supplementary Table 1).
2482 One site up- and downstream of wastewater outfalls in the rivers were
2483 sampled with two replicates, except for the River Dove
2484 (Staffordshire/Derbyshire, UK) and River Cherwell
2485 (Northamptonshire/Oxfordshire, UK) where 4 sites were sampled in total
2486 and the River Erewash (Derbyshire/Nottinghamshire, UK) with 15 sites.
2487 eDNA was collected, extracted and analysed as outlined in Chapter 1.
2488 Parasite presence was assessed through semi-quantitative
2489 metabarcoding targeting myxozoans (see Chapter 1 methods) and
2490 replicate results pooled. We downloaded the bulk EA site dataset using the
2491 EA Ecology & Fish Data Explorer
2492 (<https://environment.data.gov.uk/ecology/explorer/>) for FW Fish and River
2493 Invertebrates surveys and reduced the data down to the relevant sites,
2494 dates, taxonomy and counts. Organisms that were observed but are
2495 outside the survey scope (e.g. plants) were removed. The year 2020 saw a
2496 marked reduction in surveys due to the Covid19 pandemic and was
2497 therefore also removed. Since we focus on broadscale biodiversity of >100
2498 species and not individual species changes, we averaged counts over the
2499 last 10 years species to account for significant between-year variance.

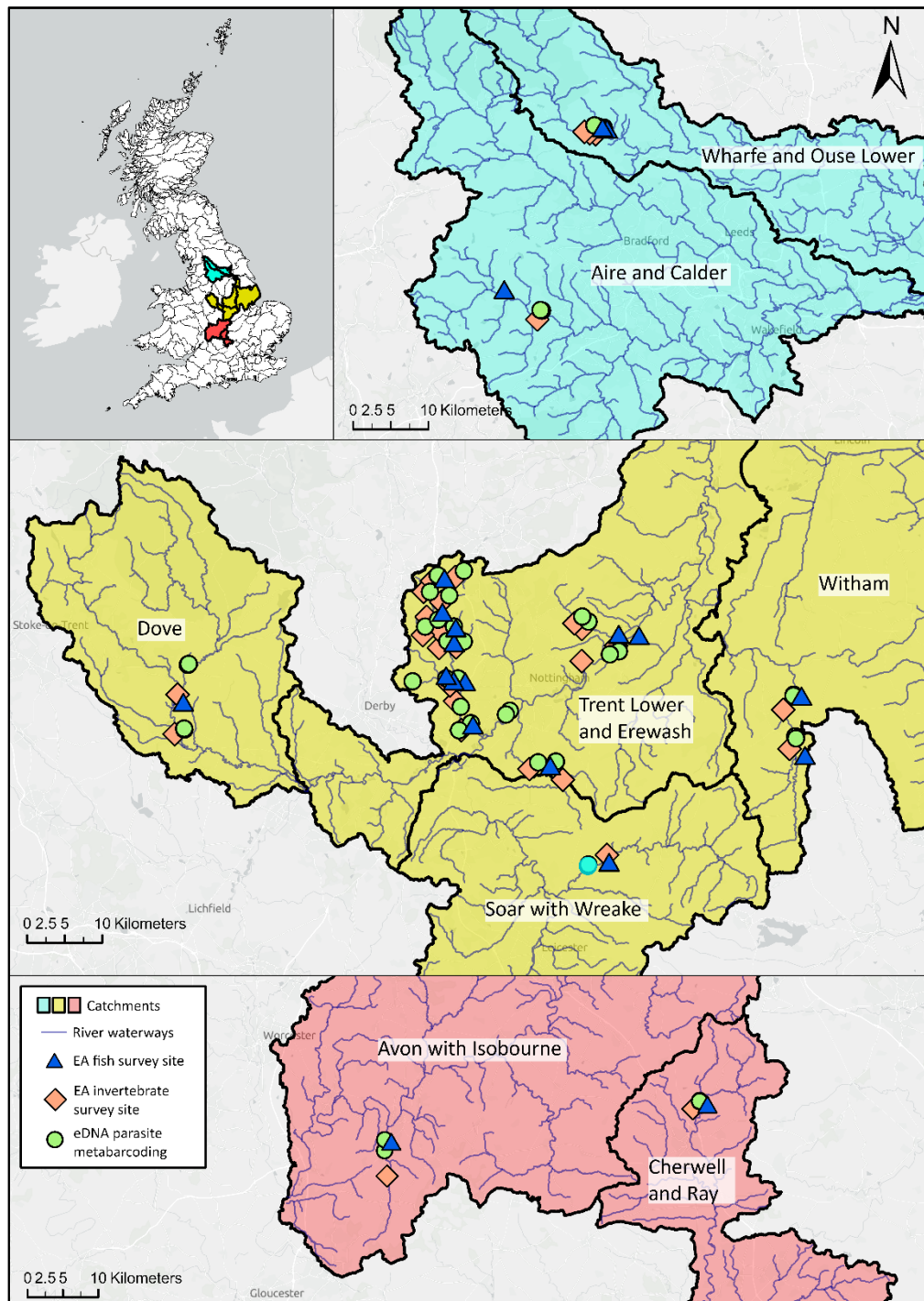


Figure 1: Map of the eight sampled catchments Wharfe and Ouse Lower, Aire and Calder, Witham, Trent Lower and Erewash, Dove, Soar with Wreake, Avon with Isobourne and Cherwell and Ray. Green circles show eDNA parasite metabarcoding sites, blue triangles for EA fish survey and orange diamonds for EA invertebrate survey sites.

2500 The closest, available EA survey sites on freshwater fish and invertebrates
2501 were matched to our eDNA sampling sites. All eDNA sites were paired with
2502 a unique survey site. In a few cases no unique invertebrate or fish survey
2503 sites could be paired with a unique eDNA site. In these cases, eDNA sites
2504 shared the closest survey sites (below 3 km). Some sites lacked a close EA
2505 survey site for either fish or invertebrates and were therefore excluded in
2506 the respective comparative analysis. The EA data varies considerably in the
2507 taxonomic level of identification. In numerous cases organisms of
2508 potentially the same species were recorded separately, ranging from order
2509 to genus and species level identifications, while other taxonomic groups
2510 such as annelids lacked lower-level taxonomic identification completely.
2511 We evaluated if we could associate the EA findings with ecological traits,
2512 then simplify and collapse the taxonomic rankings into traits (Fig. 2). Doing
2513 so we convert abundance and composition of species (or entities of
2514 various taxonomic levels) into abundance and composition of traits
2515 without the taxonomic ambiguity of the original data resulting in ecology-
2516 based instead of taxonomy-based biodiversity descriptors. For trait
2517 assignment, we assumed that closely related taxa share ecological traits,
2518 allowing trait assignment to taxa even with ambiguous identification.
2519 Groups lacking lower-level taxonomic resolution were assigned traits
2520 based on the trait profile of the group (cluster). We expect the uncertainty
2521 introduced by this process to be smaller than the uncertainty inherent in
2522 the raw taxonomic data, thereby improving ecological interpretation
2523 without losing meaningful variation.

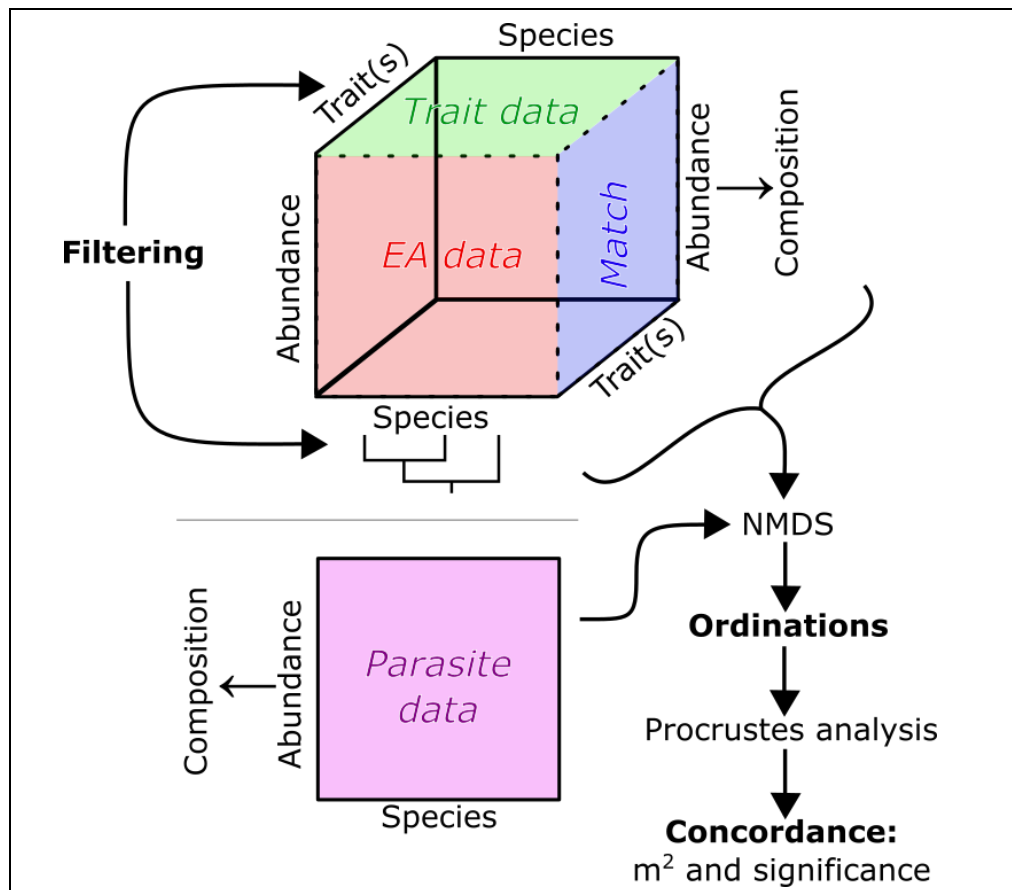


Figure 2: Data processing and analysis of EA survey count data, ecological species traits and parasite metabarcoding results.

2524 *Selection of traits and species*

2525 The freshwaterecology.info database contains ‘autecological
 2526 characteristics, ecological preferences and biological traits as well as
 2527 distribution patterns of more than 20,000 European freshwater organisms
 2528 belonging to fish, macro-invertebrates, macrophytes, phytobenthos and
 2529 phytoplankton’ including occurrence and region, habitat and life and body
 2530 variables (Schmidt-Kloiber and Hering 2015). We refer to those variables
 2531 as ‘traits’ from here on. We extracted all freshwater fish and invertebrate
 2532 species from the database. Only species with occurrence in the UK were
 2533 retained. Species with no trait entries were removed along with traits that
 2534 were not connected to habitats (e.g. reproductive age or size). Since we
 2535 have no life history or life stage information from the EA sampling, we
 2536 aimed to retain just generally applicable, constant traits such as feeding or
 2537 habitat preference.

Not all taxonomic groups were equally well represented. For invertebrates: Species count per Class or Order ranged from one for Araneae, Nematomorpha and Nemertini to over 2,000 for Coleoptera (2126) and Diptera (2345). To allow for better representation of smaller groups we removed the high species number orders like Coleoptera and Diptera in the initial assessment of usable traits. Proportion of missing traits ranged from 63% (Nemertini) to 99% (Hydrachnidia) while Coleoptera had generally the best described species with missing value proportion as low as 33%. Most traits had around 80% missing values. Selection of fish species was easier as they were overall better represented with missing trait values ranging from 0% (Esocidae, Nemacheilidae, Siluridae and Thymallidae) to 47% (Perciformes: Odontobutidae) with overall much lower species numbers than when compared to invertebrates. After filtering for completeness and representativeness, the final dataset comprised 637 invertebrate species characterized by 38 traits and 132 fish species described by 19 traits. Parasite traits consist of putative host species and infection site rates collected from literature (Fiala et al. 2015, Holzer et al. 2018, Liu et al. 2019, Lisnerová et al. 2023).

FAMD clustering, NMDS and Procrustes

Factor Analysis of Mixed Data (FAMD) was used to reduce dimensionality of mixed trait data and identify clusters of ecologically similar taxa. These clusters facilitated the assignment of traits to taxonomic groups in the EA data. Traits were summarized per species, genus, subfamily, family and phyla or class as well as the cluster grouping. After assigning the traits to the taxa in the EA data, we calculated the presence/absence and abundance of the traits on each site (Fig. 1).

NMDS ordinations visualised community composition patterns, while Procrustes analysis quantified concordance between trait and parasite community structures. The NMDS ordinations were calculated for the original species count (abundance) and presence/absence (composition) data and for the newly generated trait abundance and composition data

2569 using the package ‘vegan’ (with settings: distance = ‘bray’, ‘monoMDS’, k =
2570 2, maxit = 300, try = 100, trymax = 1000) (Oksanen et al. 2013). To assess
2571 the scaling, we used stress plots and their R² fit (non-metric and linear) as
2572 well as goodness of fit. The different ordinations were compared with
2573 Procrustes analysis including the m² and significance values.

Results

2574 *NMDS clustering of fish and invertebrate traits*

2575 Using ecological traits from the EA freshwater fish survey data, all fish
2576 species in the ecological trait database grouped into three broad clusters
2577 (Fig. 3a). The subset of fish species detected at our sampling sites, mostly
2578 identified to species level, formed three distinct clusters corresponding to
2579 the major clusters in the ordination with full data. In addition, three single
2580 species in this reduced dataset formed additional “clusters” due to
2581 dissimilarity of their traits (eel - *Anguilla anguilla*; pike - *Esox lucius*; three-
2582 spined stickleback - *Gasterosteus aculeatus*) (Fig. 3b). Eels are
2583 ecologically demanding species with a complex life history regarding
2584 reproduction and migration. Pike are the largest fish species in this study.
2585 They prefer slower and shallower waters with ample cover to hunt prey.
2586 Stickleback show elaborate courtship and brood care. They can be found
2587 in variety of salt, brackish and freshwater habitats from sheltered bays and
2588 marshes to rivers and lakes. The multi-species clusters were characterised
2589 as: 1) slow-water cyprinids and perch, 2) faster-water cyprinids and 3)
2590 migratory, fast-water salmonids and bullhead.

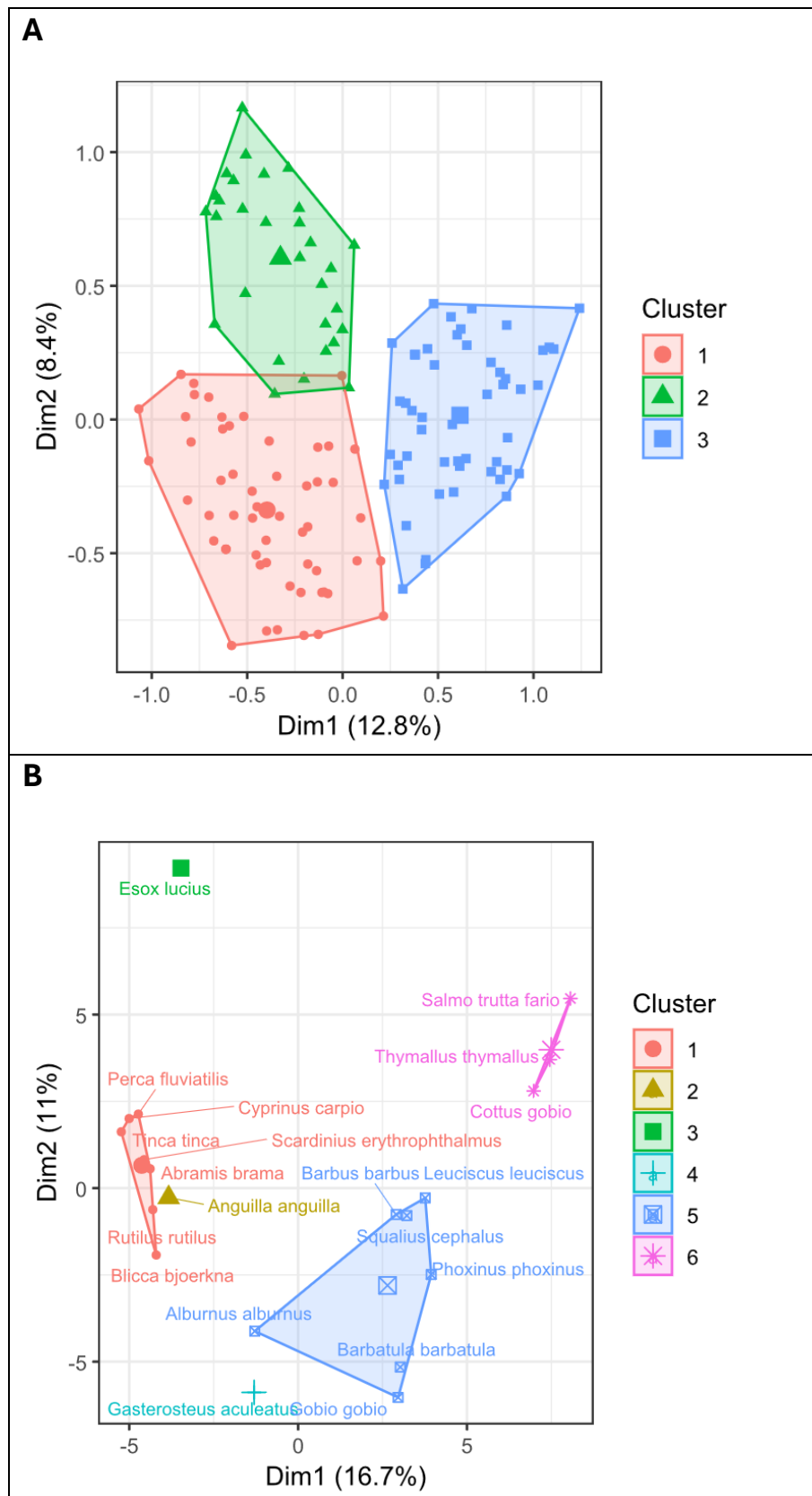


Figure 3: A) Clustering of ecological traits for all fish species in the ecological trait database (clusters indicated by colour). B) Subset of fish present in the sampling sites with subclusters and species names indicated by colours (colours do not indicate same clusters in A and B).

2591 The number of invertebrate taxa found in the EA database for our sampling
2592 sites ranged between 22 to 211 taxonomic entities. While counts ranged
2593 from 1 to 1000 individuals, most invertebrate taxonomic entities were rare
2594 (1-2 records per site). Insects, mostly caddis flies (Trichoptera), mayflies
2595 (Ephemeroptera) and true flies (Diptera) were the most common insect
2596 group, however beetles (Coleoptera) were the most species rich group,
2597 followed by caddis flies (Fig 4A). For non-insect invertebrates, molluscs,
2598 crustaceans and annelids were the most common (Fig 4B). Relatively few
2599 taxa with species level identifications were retained in the data included in
2600 this study, thus necessitating converting all taxonomic levels to traits (Fig
2601 5A-C shows the reduction in data with filtration by taxonomic accuracy of
2602 identification). Bivalves, gastropods and crustaceans were usually
2603 identified to species level (Fig.5C), but annelids (Hirudinea, Oligochaeta
2604 and Polychaeta) nearly completely lacked lower taxonomic level
2605 identification (Fig. 5C).

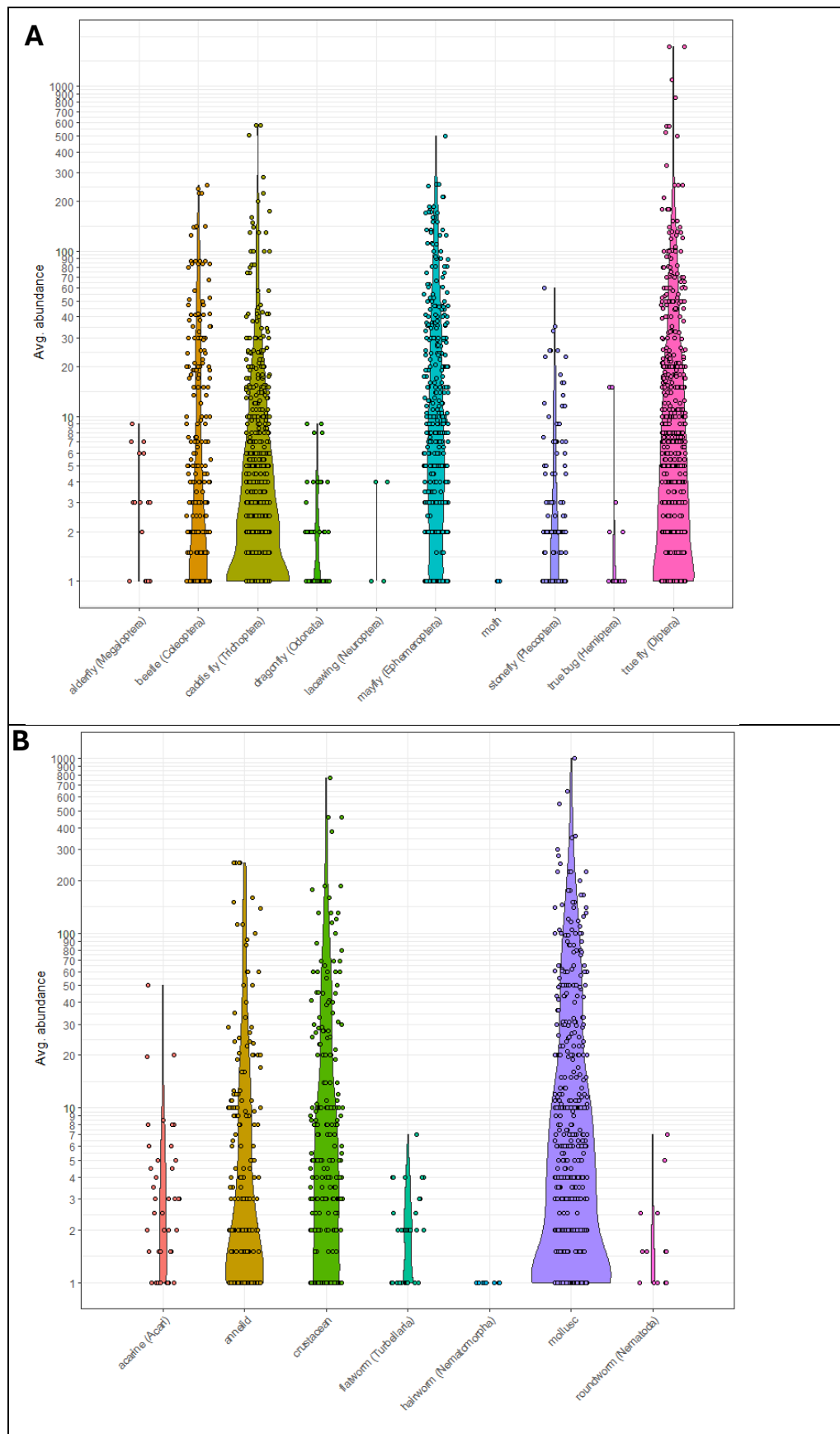
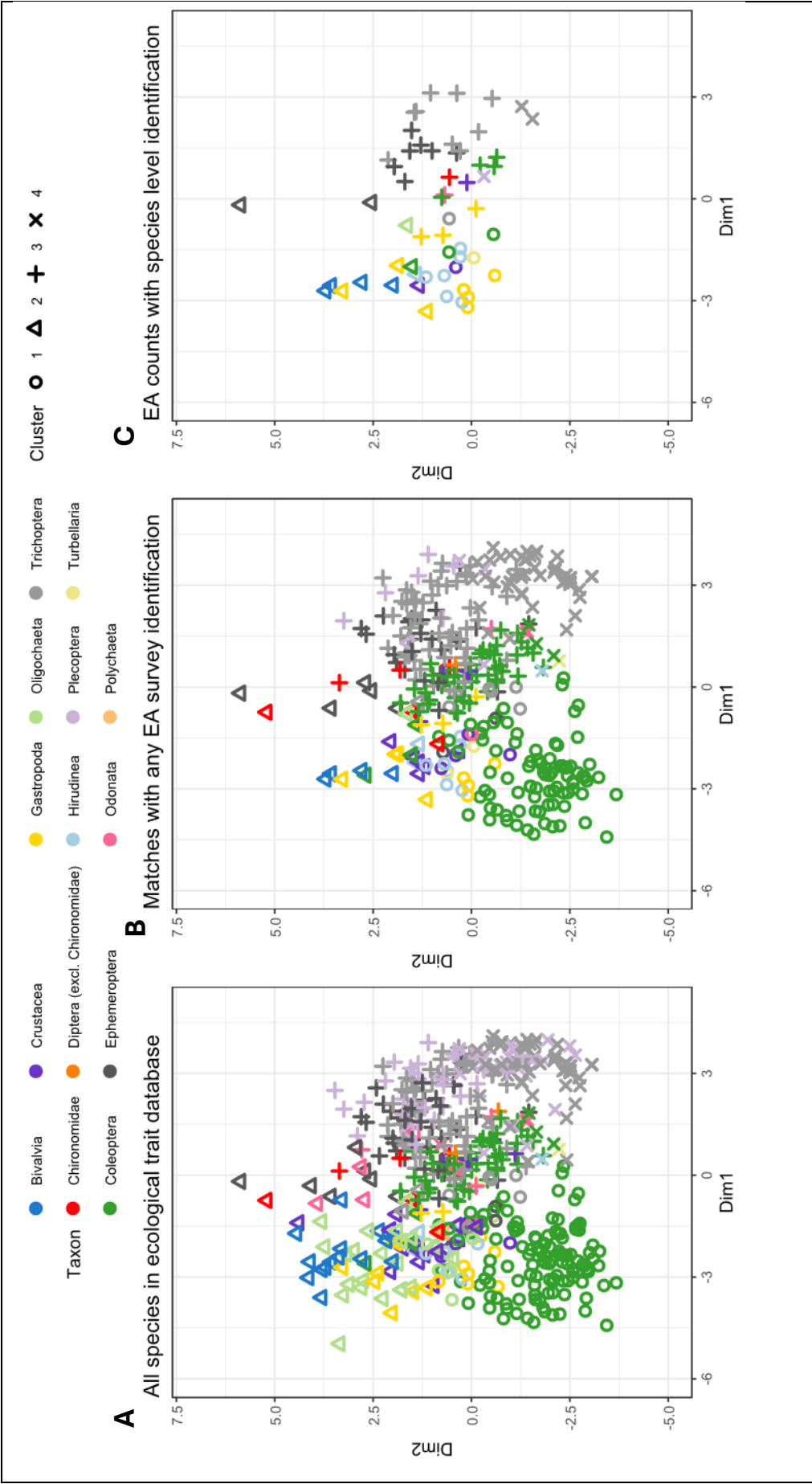


Figure 4: 10-year average of insect (A) and non-insect invertebrate (B) abundance from EA sampling sites.



↑ **Figure 5:** NMDS ordinations based on ecological traits of A) all invertebrate species found in the invertebrate ecological trait database, B) subset of species found in the full EA database and C) species with a species level identification in the EA database. Shape indicates trait cluster association and colour indicates taxonomy.

2606 Annelids are hosts of several myxozoan species and other host groups
2607 such as bryozoans do not occur in the EA data due to sampling method
2608 limitations. Trait cluster 2 contains most annelids and aligns most with
2609 bryozoan traits, but bryozoan species were not included in the analysis due
2610 to missing data. Correlations of oligochaetes and trait cluster abundance
2611 (including cluster 2) with parasite abundance were non-significant in the
2612 range of $R = [-0.19, 0.29]$ and $p = [0.1, 0.31]$ (Fig. 6).

2613 *Procrustes comparisons of NMDS ordinations*

2614 The NMS ordinations of species and trait composition (presence/absence)
2615 and abundance (count data) were compared using pairwise Procrustes
2616 analysis, within fish, invertebrates and parasites, and across the groups.
2617 Lower m^2 values (sum of squared differences of pairwise ordinations)
2618 indicate higher concordance between ordinations, with values below 0.50
2619 considered significant (Heino 2010), and 0 indicating a perfect match
2620 between the ordinations.

2621 Fish species composition and abundance ($m^2 = 0.37^*$, Fig. 8a vs. b and
2622 Table 1) was strongly concordant, as was the case for invertebrates ($m^2 =$
2623 0.06^* , Fig. 8a vs. b and Table 1), suggesting that overall patterns of
2624 presence/absence and relative abundance were aligned within each
2625 group. This indicates that species occurring more frequently across sites
2626 also tended to occur in higher numbers, reflecting consistent ecological
2627 structuring within both the fish and invertebrate communities.

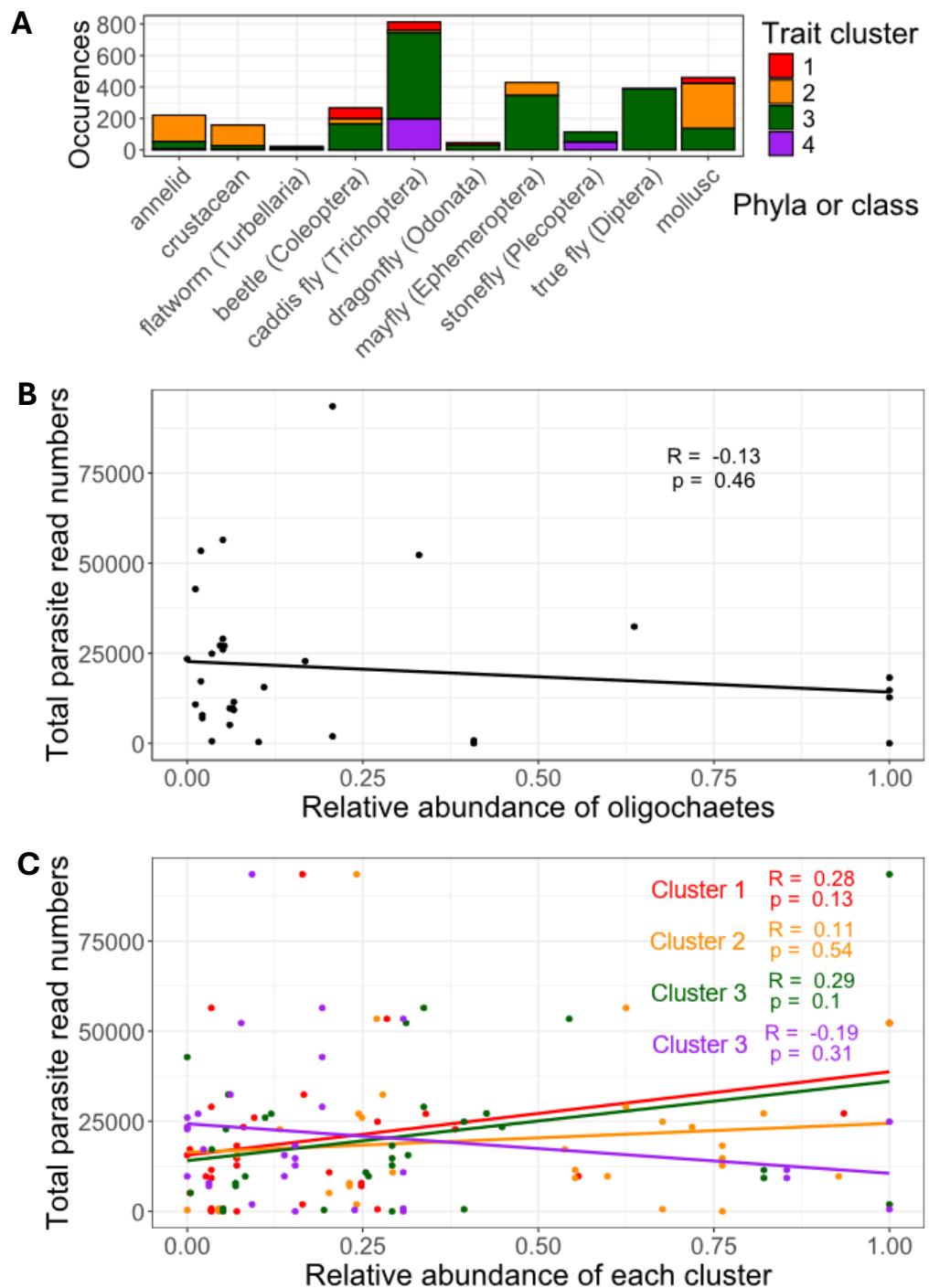


Figure 6: A) Occurrences of invertebrate species sorted by phylum to class taxonomy and trait clusters (colour), B) Total parasite read numbers of the eDNA metabarcoding in relation to relative abundance of oligochaetes on the sampling sites with correlation (R) and p-value (p), C) Total parasite read numbers of the eDNA metabarcoding in relation to relative abundance of the trait clusters (colours) on the sampling sites with correlation (R) and p-value (p).

2628 A strong concordance was found between fish species composition and
 2629 trait composition ordinations ($m^2 = 0.15^*$, Fig. 8a vs. b and Table 1),
 2630 whereas invertebrate species composition and trait composition showed
 2631 no concordance ($m^2 > 0.95$, Fig. 9a vs. b and Table 1), indicating that the
 2632 ordination patterns based on species-level data were similar to those
 2633 based on traits in fish but not in invertebrates. In fish, the species
 2634 abundance and trait composition ordinations showed a weak
 2635 concordance ($m^2 = 0.50^*$, Fig. 8b vs. c and Table 1), while in invertebrates
 2636 the concordance was stronger and significant ($m^2 = 0.26$, Fig. 9a vs. c and
 2637 Table 1), indicating greater similarity between the respective ordination
 2638 structures in the invertebrate dataset.

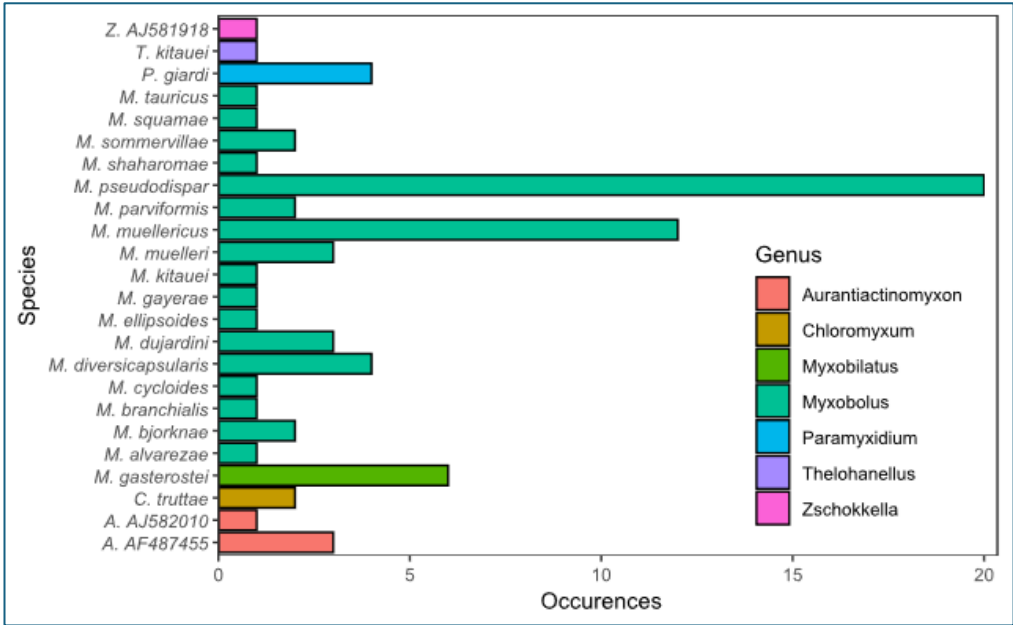


Figure 7: Myxozoan parasite diversity from eDNA metabarcoding with total occurrences on the tested sites and species of the parasite genera.

2639 24 parasite species of seven myxozoan genera were found through eDNA
 2640 metabarcoding on the sampled sites (Fig. 7). Myxobolus pseudodispar was
 2641 the most widespread parasite. Parasite species abundance and
 2642 composition were moderately concordant to parasite trait abundance and
 2643 composition, respectively ($m^2 = 0.44^*$ Fig 10a vs. b, and 0.46^* Fig. 10c vs. d
 2644 and Table 2). Other comparisons show only weak concordance (>0.68).
 2645 Because the parasite detection was not successful in all sites, some of the

2646 NMDS ordinations and Procrustes analyses were omitted due to sample
2647 size (Table 2). Finally, the physical habitat parameters and descriptors
2648 differentiated the sites in ordination similarly when using the fish species
2649 abundance and composition and weaker when using fish traits (Table 2),
2650 however, no significant concordance was detected in invertebrate
2651 composition (Table 2) or trait ordinations when compared to the
2652 environmental ordinations (Table 2).

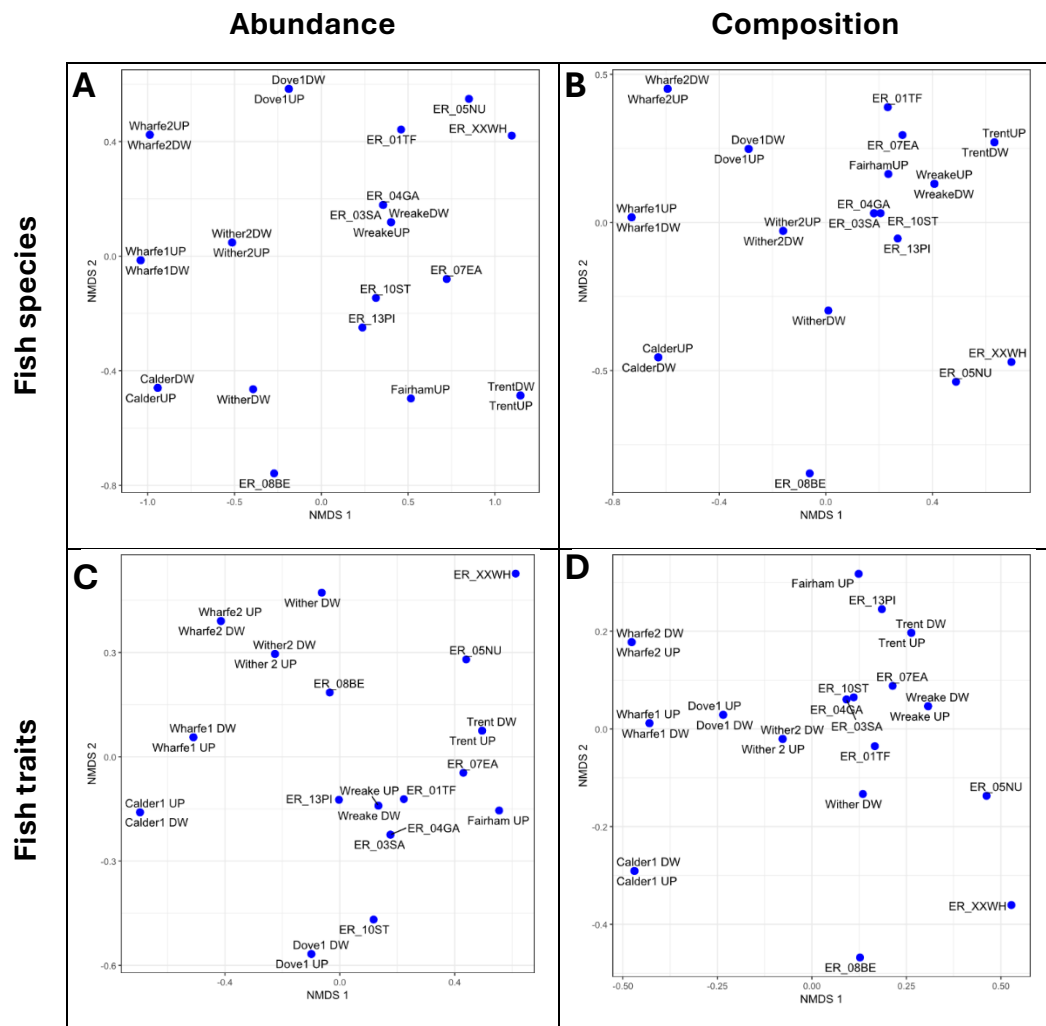


Figure 8: Non-metric multidimensional scaling (NMDS) graphs of fish species and trait abundance and composition on all sites. A) Fish species abundance, B) Fish species composition, C) Fish trait abundance, D) Fish trait composition.

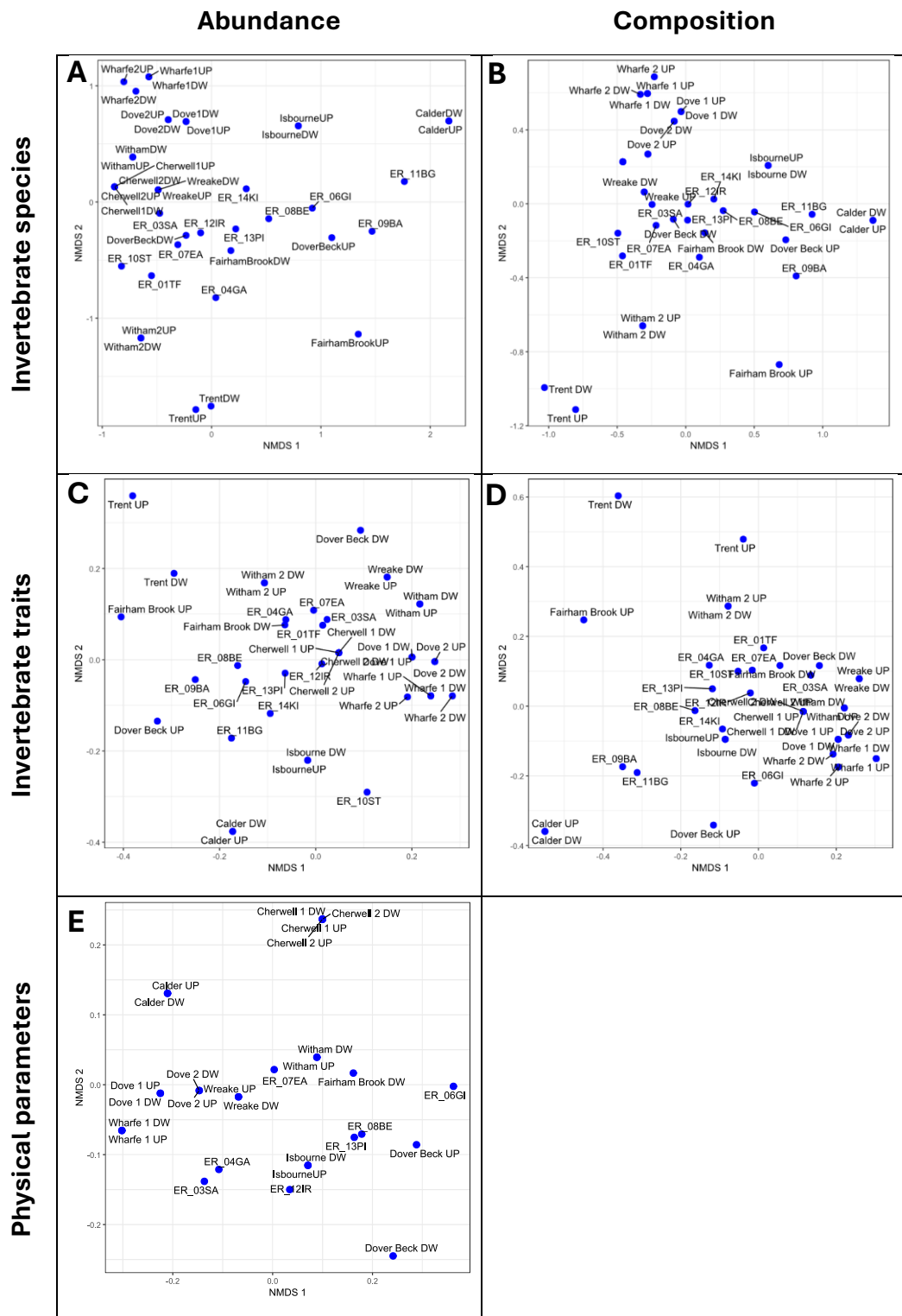
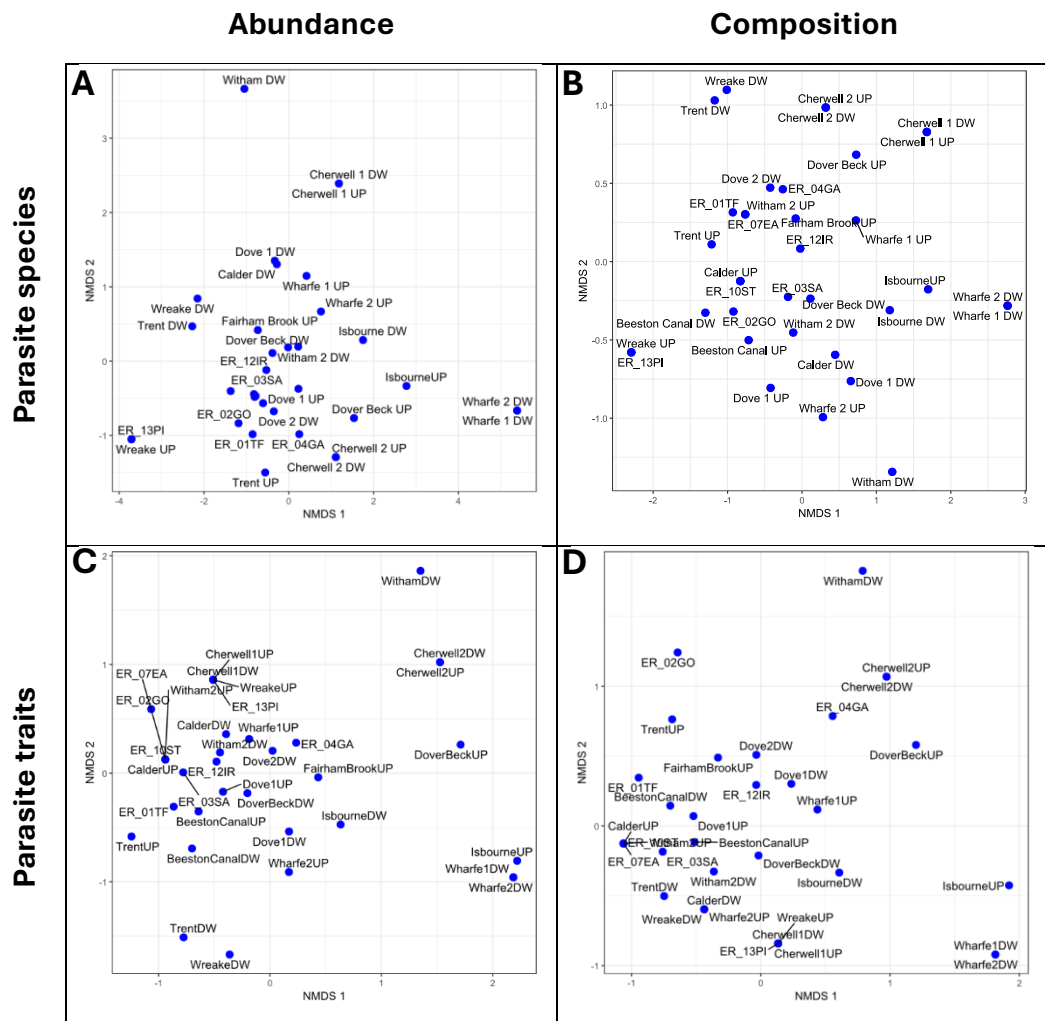


Figure 9: Non-metric multidimensional scaling (NMDS) graphs of invertebrate species and trait abundance and composition as well as physical site parameters on all sites. A) Invertebrate species abundance, B) Invertebrate species composition, B) Invertebrate traits abundance, D) Invertebrate traits composition, E) Physical parameters.



2653 *Concordance across hosts and parasites*

Invertebrate trait ordinations showed significant concordance with fish traits ($m^2 = 0.41 - 0.66$, Table 1), indicating that sites with similar invertebrate trait compositions tend to have similar fish community structures. Both invertebrate species and trait composition and abundance demonstrated fair concordance with fish species composition (0.41^* to 0.48^*), and moderate concordance with fish trait composition (0.50 to 0.66). These results suggest overlap in the ecological characteristics of fish and invertebrate communities across sites.

2662 Physical site characteristics aligned strongly with fish species abundance
2663 and composition ($m^2 = 0.33^*$ and 0.36^*), and showed moderate
2664 concordance with fish trait composition ($m^2 = 0.52^*$), indicating that
2665 environmental conditions are closely associated with fish community
2666 patterns. Conversely, physical parameters showed low concordance with
2667 both invertebrate species and trait data ($m^2 > 0.8^*$), suggesting that
2668 invertebrate community patterns are less directly related to the measured
2669 abiotic factors.

2670 Parasite species abundance displayed weak to moderate concordance
2671 with fish species and trait data ($m^2 = 0.61^*$ to 0.82^*), but these association
2672 were not statistically significant. Parasite trait ordinations showed no
2673 concordance with invertebrate species and trait data. The only significant
2674 association involving parasites was weak concordance between parasite
2675 species abundance and invertebrate species composition ($m^2 = 0.48^*$),
2676 implying a limited relationship between parasite presence and
2677 invertebrate community structure. However, these patterns were generally
2678 inconsistent and constrained by the available parasite data (Table 2).

Table 1: Procrustes analysis of ordination comparison including fish, invertebrate and physical site parameters as m^2 with better concordance indicated by blue shading and significant values in bold.

	Fish				Inv			
	Species		Trait		Species		Trait	
	Abun.	Comp.	Ab.	Comp.	Abun.	Comp.	Abun.	Comp.
Fish	Species	Abundance						
		Composition						
	Trait	Abundance	0.54					
		Composition	0.50	0.15				
Inv	Species	Abundance	0.43	0.44				
		Composition	0.44	0.46				
	Trait	Abundance	0.43	0.48				
		Composition	0.41	0.45				
Site	Physical		0.33	0.36	0.71	0.52	0.93	0.93
							0.80	0.93

Table 2: Procrustes analysis of ordination comparison of parasite parameters internally, against fish and invertebrates as m² with better concordance indicated by blue shading and significant values in bold.

			Parasite			
			Species		Trait	
			Abun.	Comp.	Abun.	Comp.
Parasite	Species	Abundance				
		Composition	0.44			
	Trait	Abundance	0.68	0.72		
		Composition	0.76	0.74	0.46	
Fish	Species	Abundance	0.69			
		Composition	0.63			
	Trait	Abundance	0.82			
		Composition	0.61			
Inv	Species	Abundance	0.73		0.94	
		Composition	0.55		0.94	
	Trait	Abundance	0.76		0.95	
		Composition	0.80		0.94	

Discussion

Community ecology predicts that taxonomically distinct groups sharing similar environmental requirements will exhibit concordant spatial patterns in composition and abundance, reflecting the influence of shared environmental filters, dispersal constraints, and biotic interactions (Carraro et al. 2023). Cross-taxon concordance is therefore expected when communities respond in parallel to broad habitat gradients, and weaker when taxon-specific factors override shared drivers (Carraro et al. 2020b). In our study, we found strong and consistent concordance between fish and invertebrate communities at both species and trait levels, suggesting that common ecological structuring operates across these groups, despite differences in life history, mobility, and species richness. By contrast, parasite communities showed weak and inconsistent concordance with either taxonomic group, indicating that parasite distributions are shaped by specific host-parasite interactions, beyond the ecological functions that drive free-living communities.

An important methodological choice in this study was converting species data to trait-based representations, which was used effectively to reduce the noise associated with ambiguous taxonomic identifications, particularly for invertebrates. The noise reduction was achieved by collapsing multiple taxonomic levels. For example, in the EA database, the occurrence of likely the same freshwater clam species, e.g. *Sphaerium corneum*, could be recorded mostly to species level or anywhere between genus (*Sphaerium*) to family (*Sphaeriidae*), resulting in recording of false rare taxa, whereas in trait-based analyses these taxa would be captured under the same ecological trait space. Traits thus capture functional aspects of species ecology, meaning that trait similarity should correspond to similarity in habitat use. For instance, the abundance of a species with a burrowing trait can be taken as the abundance of “burrowing”, which represents the sites ecological conditions without requiring perfect species-level resolution. This reduces the level of

2709 taxonomic expertise required, but also reduces within-group variation, and
2710 we saw that invertebrate species composition and trait composition were
2711 no longer directly comparable ($m^2 > 0.80$).

2712 Regardless of the taxonomic constraints, and the trade-off associated with
2713 trait-based analysis, most comparison of species abundance,
2714 composition, and trait abundance and composition, yielded good
2715 concordance across fish and invertebrates ($m^2 < 0.5-0.6$). An exception was
2716 invertebrate species composition versus traits composition, which was
2717 not significant. Invertebrate species composition matched fish species
2718 and traits as well as it matched invertebrate traits, indicating that the data
2719 could be reduced to a representative, shared ecology structure with high
2720 and often significant similarity. The concordance between fish and
2721 invertebrate ordinations reflects parallel responses to habitat gradients,
2722 with, for example, higher flow sites supporting fish such as trout and eel,
2723 alongside invertebrates with clinging or burrowing traits, and low flow sites
2724 favouring lentic associated fish, such as pike together with sediment-
2725 preferring invertebrates.

2726 Interpreting these patterns requires considering how abundance data are
2727 shaped by sampling biases, such as uneven sampling effort, variation in
2728 detectability, taxonomic uncertainties and skewed distributions of rare
2729 species (Magurran and Henderson 2003). Reducing data to presence-
2730 absence can increase the influence of rare species and potentially smooth
2731 differences among sites (Matthews and Whittaker 2015, Callaghan et al.
2732 2024), yet, in the Procrustes analyses we observed significant
2733 concordance ($m^2 < 0.50$) between species abundance and composition for
2734 all major comparisons.

2735 Fish communities were characterised by three distinct trait-based
2736 clusters, with pike, eel and sticklebacks as outliers. Most fish traits were
2737 linked to physical fluvial characteristics such as flow rate and water
2738 quality, with fewer traits related to diet, morphology or reproduction. Fewer
2739 fish than invertebrate species, and less trait overlap between distantly

2740 related fish species led to clearer clustering and greater site separation in
2741 ordinations involving fish than those involving invertebrate taxa. Fish
2742 species composition matched trait composition very closely ($m^2 = 0.15^*$)
2743 and showed the strongest concordance with physical site characteristics
2744 ($m^2 = 0.36^*$ for composition and 0.33^* for abundance), while physical
2745 variables did not align with fish traits ($m^2 > 0.50$). This contrast suggests
2746 that while species composition reflects habitat filtering, traits may capture
2747 a broader set of ecological aspects not directly measured by our
2748 environmental variables.

2749 Invertebrate species composition matched species abundance
2750 exceptionally well, although both were affected by incomplete taxonomic
2751 resolution. Trait ordinations reduced complexity and noise, producing
2752 similar concordance with fish ordinations as invertebrate species did (m^2
2753 $\approx 0.50^*$ for species and 0.52^* for traits). Physical site characteristics
2754 showed no concordance with any invertebrate ordination, suggesting that
2755 the measured physical variables may be less predictive of invertebrate
2756 than fish distributions. As with fish, this indicates that physical metrics
2757 alone cannot fully capture the local ecological variation.

2758 Neither oligochaetes nor the trait cluster 2 which contains annelids and
2759 best aligns with bryozoans (which were not observed in the EA invertebrate
2760 surveys) correlate with parasite abundance (Fig. 6). (Potential) host
2761 occurrence does not predict co-occurrence with parasites. The assembly
2762 of clusters, traits and species abundance is more informative than
2763 potential host group occurrences. Parasite ordinations showed significant
2764 but moderate concordance and a weak alignment with ordinations of free-
2765 living taxa. The strongest match was between parasite abundance and
2766 invertebrate species composition, although this was still modest (m^2
2767 $=0.61^*$). Patchy parasite occurrence, likely due to complex life cycles,
2768 variable host specificity from specialists to generalists (Agnew and Koella
2769 1999, Locke et al. 2013, Bellis et al. 2021, Hasegawa and Koizumi 2021,
2770 Ivers and Jha 2023), and the absence of intermediate hosts at some sites,

2771 reduced the variation available for cross-taxon comparisons. These
2772 findings align with previous work showing difficulty in predicting parasite
2773 diversity from host distributions alone (Grunberg 2021). These results
2774 suggest further that parasite distributions are shaped in part by
2775 environmental constraints, dispersal processes, but with a strong
2776 influence of host community composition, which leads to mosaic
2777 occurrence patterns that are not clearly aligned with ecological drivers of
2778 free-living communities of species.

2779 A key limitation for interpreting parasite patterns here is the poor
2780 taxonomic resolution of the bryozoan and annelid hosts that the myxozoan
2781 parasites analysed here infect. Many annelids were only identified to
2782 coarse taxonomic levels, meaning any subtle differences in their
2783 ecological traits (e.g., type and depth of sediment burrowing, specific
2784 feeding mode and diet preferences, level of tolerance to low oxygen,
2785 demography and life-history traits) could not be explicitly linked to parasite
2786 occurrence. Bryozoans are not captured by the sampling method
2787 employed by the EA and thus are completely missing. This means that
2788 many of the filter-feeding traits, and hydrology related habitat preference
2789 traits were not captured in the analysis. Consequently, correlation
2790 between potential host and parasite abundance is poor (Fig. 6). Improving
2791 eDNA-based identification of annelids and metabarcoding of bryozoans
2792 could therefore be critical for understanding whether host ecological traits
2793 can predict myxozoan community composition. Such an approach could
2794 provide indirect insights into parasite life cycles and co-occurrence
2795 patterns, even when direct host–parasite links are difficult to observe.

2796 While I show that eDNA is an effective tool for detecting parasites in river
2797 systems (Rusch et al. 2018, Duval et al. 2024, see Chapter 1), relating
2798 parasite diversity to host biodiversity requires integrating host-parasite
2799 association data and accounting for intermediate host presence.

2800 Relationships between parasite pressure and environmental factors have
2801 been demonstrated using eDNA (Carraro et al. 2018, Robinson et al.

2802 2022, Schuster et al. 2022). The fact that any host-parasite concordance
2803 was detectable here suggests that broad scale biodiversity patterns may
2804 indirectly reflect parasite occurrence, but strong predictive relationships
2805 remain elusive without more complete host taxonomic and occurrence
2806 data.

2807 Conclusion

2808 Our findings show that fish and invertebrate communities exhibit strong
2809 cross-taxon concordance in both species and trait patterns, reflecting
2810 shared environmental structuring, whereas parasite communities display
2811 weaker and more variable alignment with either host group. This suggests
2812 that predicting parasite distributions from general biodiversity metrics is
2813 possible but limited by the complexity of parasite life cycles and the
2814 specificity of host–parasite relationships. In the case of myxosporeans, the
2815 main hosts are annelids, yet these were poorly resolved in our invertebrate
2816 dataset, as are the parasite traits, especially regarding host specificity
2817 (Lisnerová et al. 2020, Lisnerová et al. 2023). Improving eDNA-based
2818 identification of annelids and bryozoans, and the ecological traits of
2819 myxozoans, would be a critical step toward determining whether host
2820 ecology can reliably predict parasite occurrence and would offer indirect
2821 insights into parasite life cycles.

2822 Theoretical and realised parasite niches are shaped by a combination of
2823 local habitat conditions, ecosystem-scale processes and host community
2824 composition. eDNA provides an efficient tool for detecting parasites and
2825 mapping potential disease hotspots, particularly when complemented by
2826 conventional surveillance. Integrating these approaches could enable the
2827 tracing of host and parasite co-occurrence patterns over time and space.
2828 In the context of climate change and conservation/rewilding efforts to
2829 improve habitats, understanding these dynamics becomes increasingly
2830 important as parasites are both integral to biodiversity (Okamura et al.
2831 2018) and capable of influencing ecosystem balance (Hutchins et al.
2832 2021). Future work should therefore prioritise improving host taxonomic

2833 resolution to strengthen the ecological basis for predicting parasite
2834 distributions in freshwater systems.

Bibliography

- 2835 Abd-Elfattah, A., I. Fontes, G. Kumar, H. Soliman, H. Hartikainen, B. Okamura,
2836 and M. El-Matbouli. 2014. Vertical transmission of *Tetracapsuloides*
2837 *bryosalmonae* (Myxozoa), the causative agent of salmonid proliferative
2838 kidney disease. *Parasitology* **141**:482-490.
- 2839 Adams, C. I. M., M. Knapp, N. J. Gemmell, G.-J. Jeunen, M. Bunce, M. D.
2840 Lamare, and H. R. Taylor. 2019. Beyond biodiversity: Can
2841 environmental DNA (eDNA) cut it as a population genetics tool? *Genes*
2842 **10**:192.
- 2843 Agersnap, S., W. B. Larsen, S. W. Knudsen, D. Strand, P. F. Thomsen, M.
2844 Hesselsøe, P. B. Mortensen, T. Vrålstad, and P. R. Møller. 2017.
2845 Monitoring of noble, signal and narrow-clawed crayfish using
2846 environmental DNA from freshwater samples. *PLOS ONE* **12**:e0179261.
- 2847 Agnew, P., and J. C. Koella. 1999. Life history interactions with environmental
2848 conditions in a host–parasite relationship and the parasite's mode of
2849 transmission. *Evolutionary Ecology* **13**:67-91.
- 2850 Ahmed, M., E. Kayal, and D. V. Lavrov. 2024. Mitochondrial DNA of the
2851 demosponge *Acanthella acuta*: Linear architecture and other unique
2852 features. *Genome Biology and Evolution* **16**:evae168.
- 2853 Altermatt, F. 2013. Diversity in riverine metacommunities: a network
2854 perspective. *Aquatic Ecology* **47**:365-377.
- 2855 Altermatt, F., C. J. Little, E. Mächler, S. Wang, X. Zhang, and R. C. Blackman.
2856 2020. Uncovering the complete biodiversity structure in spatial
2857 networks: the example of riverine systems. *Oikos* **129**:607-618.
- 2858 Altermatt, F., M. Seymour, and N. Martinez. 2013. River network properties
2859 shape α -diversity and community similarity patterns of aquatic insect
2860 communities across major drainage basins. *Journal of Biogeography*
2861 **40**:2249-2260.
- 2862 Anderson, C. L., E. U. Canning, and B. Okamura. 1999. Molecular data implicate
2863 bryozoans as hosts for PKX (Phylum Myxozoa) and identify a clade of
2864 bryozoan parasites within the Myxozoa. *Parasitology* **119**:555-561.
- 2865 Andres, K. J., S. A. Sethi, D. M. Lodge, and J. Andrés. 2021. Nuclear eDNA
2866 estimates population allele frequencies and abundance in
2867 experimental mesocosms and field samples. *Molecular Ecology* **30**:685-
2868 697.
- 2869 Andrews, K. R., M. De Barba, M. A. Russello, and L. P. Waits. 2018. Advances in
2870 using non-invasive, archival, and environmental samples for population
2871 genomic studies. Pages 63–99 *Population Genomics: Wildlife*.
- 2872 Angst, P., C. Ameline, C. R. Haag, F. Ben-Ami, D. Ebert, and P. D. Fields. 2022.
2873 Genetic drift shapes the evolution of a highly dynamic metapopulation.
2874 *Molecular Biology and Evolution* **39**:msac264.
- 2875 Atkinson, C. L., C. C. Vaughn, K. J. Forshay, and J. T. Cooper. 2013. Aggregated
2876 filter-feeding consumers alter nutrient limitation: consequences for
2877 ecosystem and community dynamics. *Ecology* **94**:1359-1369.

2878 Atkinson, S. D., J. L. Bartholomew, and T. Lotan. 2018. Myxozoans: Ancient
2879 metazoan parasites find a home in phylum Cnidaria. *Zoology* **129**:66-
2880 68.

2881 Azarian, C., S. Foster, F. Devloo-Delva, and P. Feutry. 2020. Population
2882 differentiation from environmental DNA: Investigating the potential of
2883 haplotype presence/absence-based analysis of molecular variance.
2884 *Environmental DNA* **3**:541-552.

2885 Baeza, J. A., and F. J. García-De León. 2022. Are we there yet? Benchmarking
2886 low-coverage nanopore long-read sequencing for the assembling of
2887 mitochondrial genomes using the vulnerable silky shark *Carcharhinus*
2888 *falciformis*. *BMC Genomics* **23**:320.

2889 Bailey, C., J. W. Holland, C. J. Secombes, and C. Tafalla. 2020. A portrait of the
2890 immune response to proliferative kidney disease (PKD) in rainbow
2891 trout. *Parasite Immunology* **42**:e12730.

2892 Baker, C. S., D. Steel, S. Nieu Kirk, and H. Klinck. 2018. Environmental DNA
2893 (eDNA) from the wake of the whales: Droplet digital PCR for detection
2894 and species identification. *Frontiers in Marine Science* **5**:133.

2895 Balard, A., M. Baltazar-Soares, C. Eizaguirre, and M. J. Heckwolf. 2024. An
2896 epigenetic toolbox for conservation biologists. *Evolutionary*
2897 *Applications* **17**:e13699.

2898 Barbarossa, V., J. Bosmans, N. Wanders, H. King, M. F. P. Bierkens, M. A. J.
2899 Huijbregts, and A. M. Schipper. 2021. Threats of global warming to the
2900 world's freshwater fishes. *Nature Communications* **12**:1701.

2901 Barnes, M. A., W. L. Chadderton, C. L. Jerde, A. R. Mahon, C. R. Turner, and D.
2902 M. Lodge. 2021. Environmental conditions influence eDNA particle size
2903 distribution in aquatic systems. *Environmental DNA* **3**:643-653.

2904 Barton, N. H., and B. Charlesworth. 1984. Genetic Revolutions, Founder
2905 Effects, and Speciation. *Annual Review of Ecology and Systematics*
2906 **15**:133-164.

2907 Bass, D., K. W. Christison, G. D. Stentiford, L. S. J. Cook, and H. Hartikainen.
2908 2023. Environmental DNA/RNA for pathogen and parasite detection,
2909 surveillance, and ecology. *Trends in Parasitology* **39**:285-304.

2910 Bellis, E. S., C. M. McLaughlin, C. W. Depamphilis, and J. R. Lasky. 2021. The
2911 geography of parasite local adaptation to host communities. *Ecography*
2912 **44**:1205-1217.

2913 Berkhout, B. W., M. K. Borregaard, R. Brandl, M. Brändle, D. M. Dehling, C. Hof,
2914 R. Poulin, and D. W. Thieltges. 2020. Host assemblage and environment
2915 shape β -diversity of freshwater parasites across diverse taxa at a
2916 continental scale. *Global Ecology and Biogeography* **29**:38-49.

2917 Berman, H. M. 2000. The Protein Data Bank. *Nucleic Acids Research* **28**:235-
2918 242.

2919 Bernt, M., A. Braband, B. Schierwater, and P. F. Stadler. 2013. Genetic aspects
2920 of mitochondrial genome evolution. *Molecular phylogenetics and*
2921 *evolution* **69**:328-338.

2922 Bettge, K., H. Segner, R. Burki, H. Schmidt-Posthaus, and T. Wahli. 2009a.
2923 Proliferative kidney disease (PKD) of rainbow trout: temperature- and

2924 time-related changes of *Tetracapsuloides bryosalmonae* DNA in the
 2925 kidney. *Parasitology* **136**:615-625.

2926 Bettge, K., T. Wahli, H. Segner, and H. Schmidt-Posthaus. 2009b. Proliferative
 2927 kidney disease in rainbow trout: time- and temperature-related renal
 2928 pathology and parasite distribution. *Diseases of Aquatic Organisms*
 2929 **83**:67-76.

2930 Bilton, D. T., J. R. Freeland, and B. Okamura. 2001. Dispersal in Freshwater
 2931 Invertebrates. *Annual Review of Ecology and Systematics* **32**:159-181.

2932 Bittrich, S., J. Segura, J. M. Duarte, S. K. Burley, and Y. Rose. 2024. RCSB protein
 2933 Data Bank: exploring protein 3D similarities via comprehensive
 2934 structural alignments. *Bioinformatics* **40**:btac370.

2935 Blasco-Costa, I., A. V. Koehler, A. Martin, and R. Poulin. 2013. Upstream-
 2936 downstream gradient in infection levels by fish parasites: a common
 2937 river pattern? *Parasitology* **140**:266-274.

2938 Bonen, L., and J. Vogel. 2001. The ins and outs of group II introns. *Trends in*
 2939 *Genetics* **17**:322-331.

2940 Bradshaw, P. C., A. Rath, and D. C. Samuels. 2005. Mitochondrial-encoded
 2941 membrane protein transcripts are pyrimidine-rich while soluble
 2942 protein transcripts and ribosomal RNA are purine-rich. *BMC Genomics*
 2943 **6**:136.

2944 Bushnell, J. H., and K. S. Rao. 1974. Dormant or quiescent stages and structures
 2945 among the ectoprocta: Physical and chemical factors affecting viability
 2946 and germination of statoblasts. *Transactions of the American*
 2947 *Microscopical Society* **93**:524-543.

2948 Bustin, S. A. 2010. Why the need for qPCR publication guidelines?—The case
 2949 for MIQE. *Methods* **50**:217-226.

2950 Bylemans, J., E. M. Furlan, D. M. Gleeson, C. M. Hardy, and R. P. Duncan. 2018.
 2951 Does size matter? An experimental evaluation of the relative
 2952 abundance and decay rates of aquatic environmental DNA.
 2953 *Environmental Science & Technology* **52**:6408-6416.

2954 Cable, J., I. Barber, B. Boag, A. R. Ellison, E. R. Morgan, K. Murray, E. L. Pascoe,
 2955 S. M. Sait, A. J. Wilson, and M. Booth. 2017. Global change, parasite
 2956 transmission and disease control: lessons from ecology. *Philosophical*
 2957 *Transactions of the Royal Society B: Biological Sciences* **372**:20160088.

2958 Callaghan, C. T., L. Santini, R. Spake, and D. E. Bowler. 2024. Population
 2959 abundance estimates in conservation and biodiversity research. *Trends*
 2960 *in Ecology & Evolution* **39**:515-523.

2961 Carim, K. J., T. M. Wilcox, M. Anderson, D. J. Lawrence, M. K. Young, K. S.
 2962 McKelvey, and M. K. Schwartz. 2016. An environmental DNA marker
 2963 for detecting nonnative brown trout (*Salmo trutta*). *Conservation*
 2964 *Genetics Resources* **8**:259-261.

2965 Carraro, L., and F. Altermatt. 2024. eDITH: An R-package to spatially project
 2966 eDNA-based biodiversity across river networks with minimal prior
 2967 information. *Methods in Ecology and Evolution* **15**:806-815.

2968 Carraro, L., E. Bertuzzo, E. A. Fronhofer, R. Furrer, I. Gounand, A. Rinaldo, and
 2969 F. Altermatt. 2020a. Generation and application of river network

2970 analogues for use in ecology and evolution. *Ecology and Evolution*
 2971 **10**:7537-7550.
 2972 Carraro, L., E. Bertuzzo, L. Mari, I. Fontes, H. Hartikainen, N. Strepparava, H.
 2973 Schmidt-Posthaus, T. Wahli, J. Jokela, M. Gatto, and A. Rinaldo. 2017.
 2974 Integrated field, laboratory, and theoretical study of PKD spread in a
 2975 Swiss prealpine river. *Proceedings of the National Academy of Sciences*
 2976 **114**:11992-11997.
 2977 Carraro, L., R. C. Blackman, and F. Altermatt. 2022. Modelling eDNA transport
 2978 in river networks reveals highly resolved spatio-temporal patterns of
 2979 freshwater biodiversity. Cold Spring Harbor Laboratory.
 2980 Carraro, L., R. C. Blackman, and F. Altermatt. 2023. Modelling environmental
 2981 DNA transport in rivers reveals highly resolved spatio-temporal
 2982 biodiversity patterns. *Scientific Reports* **13**:8854.
 2983 Carraro, L., H. Hartikainen, J. Jokela, E. Bertuzzo, and A. Rinaldo. 2018.
 2984 Estimating species distribution and abundance in river networks using
 2985 environmental DNA. *Proceedings of the National Academy of Sciences*
 2986 **115**:11724-11729.
 2987 Carraro, L., E. Mächler, R. Wüthrich, and F. Altermatt. 2020b. Environmental
 2988 DNA allows upscaling spatial patterns of biodiversity in freshwater
 2989 ecosystems. *Nature Communications* **11**:3585.
 2990 Carraro, L., J. B. Stauffer, and F. Altermatt. 2020c. How to design optimal eDNA
 2991 sampling strategies for biomonitoring in river networks. **3**:157-172.
 2992 Carroll, E. L., M. W. Bruford, J. A. Dewoody, G. Leroy, A. Strand, L. Waits, and
 2993 J. Wang. 2018. Genetic and genomic monitoring with minimally
 2994 invasive sampling methods. *Evolutionary Applications* **11**:1094-1119.
 2995 Chang, E. S., M. Neuhof, N. D. Rubinstein, A. Diamant, H. Philippe, D. Huchon,
 2996 and P. Cartwright. 2015. Genomic insights into the evolutionary origin
 2997 of Myxozoa within Cnidaria. *Proceedings of the National Academy of*
 2998 *Sciences* **112**:14912-14917.
 2999 Chen, Y., F. Nie, S.-Q. Xie, Y.-F. Zheng, Q. Dai, T. Bray, Y.-X. Wang, J.-F. Xing, Z.-
 3000 J. Huang, D.-P. Wang, L.-J. He, F. Luo, J.-X. Wang, Y.-Z. Liu, and C.-L. Xiao.
 3001 2021. Efficient assembly of nanopore reads via highly accurate and
 3002 intact error correction. *Nature Communications* **12**:60.
 3003 Clifton-Hadley, R. S., D. Bucke, and R. H. Richards. 1984. Proliferative kidney
 3004 disease of salmonid fish: a review. *Journal of Fish Diseases* **7**:363-377.
 3005 Clifton-Hadley, R. S., D. Bucke, and R. H. Richards. 1987. A study of the
 3006 sequential clinical and pathological changes during proliferative kidney
 3007 disease in rainbow trout, *Salmo gairdneri* Richardson. *Journal of Fish*
 3008 *Diseases* **10**:335-352.
 3009 Corse, E., C. Tougard, G. Archambaud-Suard, J. F. Agnès, F. D. Messu
 3010 Mandeng, C. F. Bilong Bilong, D. Duneau, L. Zinger, R. Chappaz, C. C. Y.
 3011 Xu, E. Megléc, and V. Dubut. 2019. One-locus-several-primers: A
 3012 strategy to improve the taxonomic and haplotypic coverage in diet
 3013 metabarcoding studies. *Ecology and Evolution* **9**:4603-4620.
 3014 Crane, L. C., J. S. Goldstein, D. W. Thomas, K. S. Rexroth, and A. W. Watts. 2021.
 3015 Effects of life stage on eDNA detection of the invasive European green

3016 crab (*Carcinus maenas*) in estuarine systems. *Ecological Indicators*
3017 **124**:107412.

3018 Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A.
3019 Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021.
3020 Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008.

3021 De Barba, M., C. Miquel, S. Lobréaux, P. Y. Quenette, J. E. Swenson, and P.
3022 Taberlet. 2017. High-throughput microsatellite genotyping in ecology:
3023 improved accuracy, efficiency, standardization and success with low-
3024 quantity and degraded DNA. *Molecular Ecology Resources* **17**:492-507.

3025 De Kinkelin, P., M. Gay, and S. Forman. 2002. The persistence of infectivity of
3026 *Tetracapsula bryosalmonae*-infected water for rainbow trout,
3027 *Oncorhynchus mykiss* (Walbaum). *Journal of Fish Diseases* **25**:477-482.

3028 Defontaine, A., F. M. Lecocq, and J. N. Hallet. 1991. A rapid miniprep method
3029 for the preparation of yeast mitochondrial DNA. *Nucleic Acids Research*
3030 **19**:185-185.

3031 Deiner, K., M. A. Renshaw, Y. Li, B. P. Olds, D. M. Lodge, and M. E. Pfrender.
3032 2017. Long-range PCR allows sequencing of mitochondrial genomes
3033 from environmental DNA. *Methods in Ecology and Evolution* **8**:1888-
3034 1898.

3035 Deiner, K., J.-C. Walser, E. Mächler, and F. Altermatt. 2015. Choice of capture
3036 and extraction methods affect detection of freshwater biodiversity
3037 from environmental DNA. *Biological Conservation* **183**:53-63.

3038 Dellaporta, S. L., A. Xu, S. Sagasser, W. Jakob, M. A. Moreno, L. W. Buss, and B.
3039 Schierwater. 2006. Mitochondrial genome of *Trichoplax adhaerens*
3040 supports Placozoa as the basal lower metazoan phylum. *Proceedings*
3041 *of the National Academy of Sciences* **103**:8751-8756.

3042 Demarche, M. L., D. F. Doak, and W. F. Morris. 2019. Incorporating local
3043 adaptation into forecasts of species' distribution and abundance under
3044 climate change. *Global Change Biology* **25**:775-793.

3045 Donath, A., F. Jühling, M. Al-Arab, S. H. Bernhart, F. Reinhardt, P. F. Stadler, M.
3046 Middendorf, and M. Bernt. 2019. Improved annotation of protein-
3047 coding genes boundaries in metazoan mitochondrial genomes. *Nucleic*
3048 *Acids Research* **47**:10543-10552.

3049 Doorenspleet, K., L. Jansen, S. Oosterbroek, O. Bos, P. Kamermans, M. Janse,
3050 E. Wurz, A. Murk, and R. Nijland. 2021. High resolution species
3051 detection: accurate long read eDNA metabarcoding of North Sea fish
3052 using Oxford Nanopore sequencing. Page PPR426843 in E. PMC, editor.

3053 Dufresnes, C., T. Déjean, S. Zumbach, B. R. Schmidt, L. Fumagalli, P. Ramseier,
3054 and S. Dubey. 2019. Early detection and spatial monitoring of an
3055 emerging biological invasion by population genetics and environmental
3056 DNA metabarcoding. *Conservation Science and Practice* **1**.

3057 Dugal, L., L. Thomas, M. Reinholdt Jensen, E. E. Sigsgaard, T. Simpson, S.
3058 Jarman, P. F. Thomsen, and M. Meekan. 2021. Individual haplotyping
3059 of whale sharks from seawater environmental DNA. *Molecular Ecology*
3060 *Resources* **22**:56-65.

3061 Duval, E., S. Blanchet, E. Quéméré, L. Jacquin, C. Veyssiére, and G. Loot. 2024.
3062 When does a parasite become a disease? eDNA unravels complex host-

3063 pathogen dynamics across environmental stress gradients in wild
 3064 salmonid populations. *Science of The Total Environment* **946**:174367.
 3065 Egeter, B., J. Veríssimo, M. Lopes-Lima, C. Chaves, J. Pinto, N. Riccardi, P. Beja,
 3066 and N. A. Fonseca. 2022. Speeding up the detection of invasive bivalve
 3067 species using environmental DNA: A Nanopore and Illumina
 3068 sequencing comparison. *Molecular Ecology Resources*.
 3069 Elbrecht, V., T. W. A. Braukmann, N. V. Ivanova, S. W. J. Prosser, M. Hajibabaei,
 3070 M. Wright, E. V. Zakharov, P. D. N. Hebert, and D. Steinke. 2019.
 3071 Validation of COI metabarcoding primers for terrestrial arthropods.
 3072 *PeerJ* **7**:e7745.
 3073 Elbrecht, V., and F. Leese. 2015. Can DNA-based ecosystem assessments
 3074 quantify species abundance? Testing primer bias and biomass—
 3075 Sequence relationships with an innovative metabarcoding protocol.
 3076 *PLOS ONE* **10**:e0130324.
 3077 Elbrecht, V., and F. Leese. 2017. Validation and development of COI
 3078 metabarcoding primers for freshwater macroinvertebrate
 3079 bioassessment. *Frontiers in Environmental Science* **5**:11.
 3080 Elbrecht, V., E. E. Vamos, K. Meissner, J. Aroviita, and F. Leese. 2017. Assessing
 3081 strengths and weaknesses of DNA metabarcoding-based
 3082 macroinvertebrate identification for routine stream monitoring.
 3083 *Methods in Ecology and Evolution* **8**:1265-1275.
 3084 Elbrecht, V., E. E. Vamos, D. Steinke, and F. Leese. 2018. Estimating
 3085 intraspecific genetic diversity from community DNA metabarcoding
 3086 data. *PeerJ* **6**:e4644.
 3087 Falkenberg, M., and C. M. Gustafsson. 2020. Mammalian mitochondrial DNA
 3088 replication and mechanisms of deletion formation. *Critical Reviews in*
 3089 *Biochemistry and Molecular Biology* **55**:509-524.
 3090 Farrell, J. A., L. Whitmore, N. Mashkour, D. R. Rollinson Ramia, R. S. Thomas, C.
 3091 B. Eastman, B. Burkhalter, K. Yetsko, C. Mott, and L. Wood. 2022.
 3092 Detection and population genomics of sea turtle species via
 3093 noninvasive environmental DNA analysis of nesting beach sand tracks
 3094 and oceanic water. *Molecular Ecology Resources* **22**:2471-2493.
 3095 Feist, S., M. Longshaw, E. Canning, and B. Okamura. 2001. Induction of
 3096 proliferative kidney disease (PKD) in rainbow trout *Oncorhynchus*
 3097 *mykiss* via the bryozoan *Fredericella sultana* infected with *Tetracapsula*
 3098 *bryosalmonae*. *Diseases of Aquatic Organisms* **45**:61-68.
 3099 Fiala, I., P. Bartošová-Sojková, and C. M. Whipps. 2015. Classification and
 3100 phylogenetics of Myxozoa. Pages 85-110 *Myxozoan evolution, ecology*
 3101 *and development*. Springer International Publishing.
 3102 Ficetola, G. F., C. Miaud, F. Pompanon, and P. Taberlet. 2008. Species detection
 3103 using environmental DNA from water samples. *Biology Letters* **4**:423-
 3104 425.
 3105 Figuerola, J., Andy, and Thomas. 2005. Invertebrate Eggs Can Fly: Evidence of
 3106 Waterfowl-Mediated Gene Flow in Aquatic Invertebrates. *The*
 3107 *American Naturalist* **165**:274-280.
 3108 Filippenko, D., H. Hartikainen, B. Okamura, and A. Vasemägi. 2014.
 3109 Characterisation of polymorphic microsatellite loci for the bryozoan

3110 *Fredericella sultana*, the primary host of the causative agent of
3111 salmonid proliferative kidney disease. *Conservation Genetics*
3112 *Resources* **6**:481-482.

3113 Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L.
3114 McCraw, and S. J. Gurr. 2012. Emerging fungal threats to animal, plant
3115 and ecosystem health. *Nature* **484**:186-194.

3116 Fonseca, V. G. 2018. Pitfalls in relative abundance estimation using eDNA
3117 metabarcoding. *Molecular Ecology Resources* **18**:923-926.

3118 Fontes, I., H. Hartikainen, J. Holland, C. Secombes, and B. Okamura. 2017.
3119 *Tetracapsuloides bryosalmonae* abundance in river water. *Diseases of*
3120 *Aquatic Organisms* **124**:145-157.

3121 Fossøy, F., H. Brandsegg, R. Sivertsgård, O. Pettersen, B. K. Sandercock, Ø.
3122 Solem, K. Hindar, and T. A. Mo. 2020. Monitoring presence and
3123 abundance of two gyrodactylid ectoparasites and their salmonid hosts
3124 using environmental DNA. *Environmental DNA* **2**:53-62.

3125 Freeland, J. R., V. K. Rimmer, and B. Okamura. 2001. Genetic changes within
3126 freshwater bryozoan populations suggest temporal gene flow from
3127 statoblast banks. *Limnology and Oceanography* **46**:1121-1129.

3128 Freeland, J. R., C. Romualdi, and B. Okamura. 2000. Gene flow and genetic
3129 diversity: a comparison of freshwater bryozoan populations in Europe
3130 and North America. *Heredity* **85**:498-508.

3131 Fremier, A. K., K. M. Strickler, J. Parzych, S. Powers, and C. S. Goldberg. 2019.
3132 Stream transport and retention of environmental DNA pulse releases
3133 in relation to hydrogeomorphic scaling factors. *Environmental Science*
3134 *& Technology* **53**:6640-6649.

3135 Gan, H., M. B. Schultz, and C. M. Austin. 2014. Integrated shotgun sequencing
3136 and bioinformatics pipeline allows ultra-fast mitogenome recovery and
3137 confirms substantial gene rearrangements in Australian freshwater
3138 crayfishes. *BMC Evolutionary Biology* **14**:19.

3139 Geerts, A. N., P. Boets, S. Van Den Heede, P. Goethals, and C. Van Der Heyden.
3140 2018. A search for standardized protocols to detect alien invasive
3141 crayfish based on environmental DNA (eDNA): A lab and field
3142 evaluation. *Ecological Indicators* **84**:564-572.

3143 Giebner, H., K. Langen, S. J. Bourlat, S. Kukowka, C. Mayer, J. J. Astrin, B. Misof,
3144 and V. G. Fonseca. 2020. Comparing diversity levels in environmental
3145 samples: DNA sequence capture and metabarcoding approaches using
3146 18S and COI genes. *Molecular Ecology Resources* **20**:1333-1345.

3147 Giesselmann, P., B. Brändl, E. Raimondeau, R. Bowen, C. Rohrandt, R. Tandon,
3148 H. Kretzmer, G. Assum, C. Galonska, R. Siebert, O. Ammerpohl, A.
3149 Heron, S. A. Schneider, J. Ladewig, P. Koch, B. M. Schuldt, J. E. Graham,
3150 A. Meissner, and F.-J. Müller. 2019. Analysis of short tandem repeat
3151 expansions and their methylation state with nanopore sequencing.
3152 *Nature Biotechnology* **37**:1478-1481.

3153 Gissi, C., F. Iannelli, and G. Pesole. 2008. Evolution of the mitochondrial
3154 genome of Metazoa as exemplified by comparison of congeneric
3155 species. *Heredity* **101**:301-320.

3156 Gómez-Rodríguez, C., M. J. T. N. Timmermans, A. Crampton-Platt, and A. P.
3157 Vogler. 2017. Intraspecific genetic variation in complex assemblages
3158 from mitochondrial metagenomics: comparison with DNA barcodes.
3159 Methods in Ecology and Evolution **8**:248-256.

3160 Gorgoglione, B., M. H. Kotob, and M. El-Matbouli. 2016. Migrating zooids allow
3161 the dispersal of *Fredericella sultana* (Bryozoa) to escape from
3162 unfavourable conditions and further spreading of *Tetracapsuloides*
3163 *bryosalmonae*. Journal of Invertebrate Pathology **140**:97-102.

3164 Gorički, Š., D. Stanković, A. Snoj, M. Kuntner, W. R. Jeffery, P. Trontelj, M.
3165 Pavičević, Z. Grizelj, M. Năpăruș-Aljančič, and G. Aljančič. 2017.
3166 Environmental DNA in subterranean biology: range extension and
3167 taxonomic implications for *Proteus*. Scientific Reports **7**:45054.

3168 Gould, M. P., C. M. Bosworth, S. McMahon, S. Grandhi, B. T. Grimmer, and T.
3169 Laframboise. 2015. PCR-free enrichment of mitochondrial DNA from
3170 human blood and cell lines for high quality next-generation DNA
3171 Sequencing. PLOS ONE **10**:e0139253.

3172 Grabner, D. S., and M. El-Matbouli. 2009. Comparison of the susceptibility of
3173 brown trout (*Salmo trutta*) and four rainbow trout (*Oncorhynchus*
3174 *mykiss*) strains to the myxozoan *Tetracapsuloides bryosalmonae*, the
3175 causative agent of proliferative kidney disease (PKD). Veterinary
3176 Parasitology **165**:200-206.

3177 Griffin, J. E., E. Matechou, A. S. Buxton, D. Bormpoudakis, and R. A. Griffiths.
3178 2020. Modelling environmental DNA data; Bayesian variable selection
3179 accounting for false positive and false negative errors. Journal of the
3180 Royal Statistical Society: Series C (Applied Statistics) **69**:377-392.

3181 Grunberg, R. L. 2021. Spatial and seasonal variation in host community
3182 structure is weakly concordant with patterns of parasite composition.
3183 Freshwater Biology **66**:303-316.

3184 Hamann, L., and A. Blanke. 2022. Suspension feeders: diversity, principles of
3185 particle separation and biomimetic potential. Journal of the Royal
3186 Society Interface **19**:20210741.

3187 Harper, L. R., L. Lawson Handley, A. I. Carpenter, M. Ghazali, C. Di Muri, C. J.
3188 Macgregor, T. W. Logan, A. Law, T. Breithaupt, D. S. Read, A. D.
3189 McDevitt, and B. Hänfling. 2019. Environmental DNA (eDNA)
3190 metabarcoding of pond water as a tool to survey conservation and
3191 management priority mammals. Biological Conservation **238**:108225.

3192 Hartikainen, H., D. Bass, A. G. Briscoe, H. Knipe, A. J. Green, and B. Okamura.
3193 2016. Assessing myxozoan presence and diversity using environmental
3194 DNA. International Journal for Parasitology **46**:781-792.

3195 Hartikainen, H., A. Gruhl, and B. Okamura. 2014. Diversification and repeated
3196 morphological transitions in endoparasitic cnidarians (Myxozoa:
3197 Malacosporea). Molecular phylogenetics and evolution **76**:261-269.

3198 Hartikainen, H., and J. Jokela. 2012. Characterisation of polymorphic
3199 microsatellite loci in the freshwater bryozoan *Fredericella sultana*.
3200 Conservation Genetics Resources **4**:475-477.

- 3201 Hartikainen, H., and B. Okamura. 2015. Ecology and evolution of
3202 malacosporean-bryozoan interactions. Pages 201-216 *Myxozoan*
3203 *evolution, ecology and development*. Springer.
- 3204 Hasegawa, R., and I. Koizumi. 2021. Relative importance of host-dependent
3205 versus physical environmental characteristics affecting the distribution
3206 of an ectoparasitic copepod infecting the mouth cavity of stream
3207 salmonid. *Ecological Research* **36**:1015-1027.
- 3208 Hatton-Ellis, T. W., L. R. Noble, and B. Okamura. 1998. Genetic variation in a
3209 freshwater bryozoan. I: Populations in the Thames basin, UK. *Molecular*
3210 *Ecology* **7**:1575-1585.
- 3211 Hay, E. M., R. Poulin, and F. Jorge. 2020. Macroevolutionary dynamics of
3212 parasite diversification: A reality check. *Journal of Evolutionary Biology*
3213 **33**:1758-1769.
- 3214 Hedrick, R. P., E. Macconnell, and P. De Kinkelin. 1993. Proliferative kidney
3215 disease of salmonid fish. *Annual Review of Fish Diseases* **3**:277-290.
- 3216 Heino, J. 2010. Are indicator groups and cross-taxon congruence useful for
3217 predicting biodiversity in aquatic ecosystems? *Ecological Indicators*
3218 **10**:112-117.
- 3219 Hengherr, S., and R. O. Schill. 2011. Dormant stages in freshwater bryozoans—
3220 An adaptation to transcend environmental constraints. *Journal of*
3221 *Insect Physiology* **57**:595-601.
- 3222 Hintikka, S., J. E. L. Carlsson, and J. Carlsson. 2022. The bacterial hitchhiker's
3223 guide to COI: Universal primer-based COI capture probes fail to exclude
3224 bacterial DNA, but 16S capture leaves metazoa behind. *Metabarcoding*
3225 *and Metagenomics* **6**:e80416.
- 3226 Hirayama, I. T., L. Wu, and T. Minamoto. 2024. Stability of environmental DNA
3227 methylation and its utility in tracing spawning in fish. *Molecular Ecology*
3228 *Resources* **24**:e14011.
- 3229 Holzer, A. S., P. Bartošová-Sojková, A. Born-Torrijos, A. Lövy, A. Hartigan, and
3230 I. Fiala. 2018. The joint evolution of the Myxozoa and their alternate
3231 hosts: A cnidarian recipe for success and vast biodiversity. *Molecular*
3232 *Ecology* **27**:1651-1666.
- 3233 Hopper, G. W., I. S. González, J. R. Bucholz, J. D. Lozier, and C. L. Atkinson. 2023.
3234 Different components of biodiversity mediate two ecosystem functions
3235 derived from naturally assembled filter-feeder communities. *Oikos*
3236 **2023**:e09699.
- 3237 Huchon, D., A. Szitenberg, S. Shefer, M. Ilan, and T. Feldstein. 2015.
3238 Mitochondrial group I and group II introns in the sponge orders
3239 Agelasida and Axinellida. *BMC Evolutionary Biology* **15**:1-14.
- 3240 Hunter, M. E., J. A. Ferrante, G. Meigs-Friend, and A. Ulmer. 2019. Improving
3241 eDNA yield and inhibitor reduction through increased water volumes
3242 and multi-filter isolation techniques. *Scientific Reports* **9**:5259.
- 3243 Hutchins, P. R., A. J. Sepulveda, H. Hartikainen, K. D. Staigmiller, S. T. Opitz, R.
3244 M. Yamamoto, A. Huttinger, R. J. Cordes, T. Weiss, L. R. Hopper, M. K.
3245 Purcell, and B. Okamura. 2021. Exploration of the 2016 Yellowstone
3246 River fish kill and proliferative kidney disease in wild fish populations.
3247 *Ecosphere* **12**:e03436.

3248 Hutchins, P. R., A. J. Sepulveda, R. M. Martin, and L. R. Hopper. 2018. A probe-
3249 based quantitative PCR assay for detecting *Tetracapsuloides*
3250 *bryosalmonae* in fish tissue and environmental DNA water samples.
3251 Conservation Genetics Resources **10**:317-319.

3252 Huver, J. R., J. Koprivnikar, P. T. J. Johnson, and S. Whyard. 2015. Development
3253 and application of an eDNA method to detect and quantify a
3254 pathogenic parasite in aquatic ecosystems. Ecological Applications
3255 **25**:991-1002.

3256 Ivers, N. A., and S. Jha. 2023. Biogeography, climate, and land use create a
3257 mosaic of parasite risk in native bumble bees. Science of The Total
3258 Environment **868**:161545.

3259 Jacquet, C., L. Carraro, and F. Altermatt. 2022. Meta-ecosystem dynamics drive
3260 the spatial distribution of functional groups in river networks. Oikos
3261 **2022**:e09372.

3262 Jenkins, H. L., R. Graham, J. S. Porter, L. M. Vieira, A. C. S. De Almeida, A. Hall,
3263 A. O'Dea, S. E. Coppard, and A. Waeschenbach. 2022. Unprecedented
3264 frequency of mitochondrial introns in colonial bilaterians. Scientific
3265 Reports **12**:10889.

3266 Jensen, M. R., E. E. Sigsgaard, S. Liu, A. Manica, S. S. Bach, M. M. Hansen, P. R.
3267 Møller, and P. F. Thomsen. 2021. Genome-scale target capture of
3268 mitochondrial and nuclear environmental DNA from water samples.
3269 Molecular Ecology Resources **21**:690-702.

3270 Jerde, C. L. 2021. Can we manage fisheries with the inherent uncertainty from
3271 eDNA? Journal of Fish Biology **98**:341-353.

3272 Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. "Sight-
3273 unseen" detection of rare aquatic species using environmental DNA.
3274 Conservation Letters **4**:150-157.

3275 Jeunen, G. J., M. Knapp, H. G. Spencer, M. D. Lamare, H. R. Taylor, M. Stat, M.
3276 Bunce, and N. J. Gemmell. 2019. Environmental DNA (eDNA)
3277 metabarcoding reveals strong discrimination among diverse marine
3278 habitats connected by water movement. Molecular Ecology Resources
3279 **19**:426-438.

3280 Jo, T., H. Murakami, R. Masuda, and T. Minamoto. 2020. Selective collection of
3281 long fragments of environmental DNA using larger pore size filter.
3282 Science of The Total Environment **735**:139462.

3283 Jo, T., H. Murakami, S. Yamamoto, R. Masuda, and T. Minamoto. 2019. Effect
3284 of water temperature and fish biomass on environmental DNA
3285 shedding, degradation, and size distribution. Ecology and Evolution
3286 **9**:1135-1146.

3287 Jo, T., and H. Yamanaka. 2022. Meta-analyses of environmental DNA
3288 downstream transport and deposition in relation to hydrogeography in
3289 riverine environments. Freshwater Biology **67**:1333-1343.

3290 Kakehashi, R., S. Ito, K. Yasui, C. Kambayashi, S. Kanao, and A. Kurabayashi.
3291 2022. Amplification and sequencing of the complete mtDNA of the
3292 endangered bitterling, *Acheilognathus longipinnis* (Cyprinidae), using
3293 environmental DNA from aquarium water. Journal of Ichthyology
3294 **62**:280-288.

3295 Karin, B. R., S. Arellano, L. Wang, K. Walzer, A. Pomerantz, J. M. Vasquez, K.
3296 Chatla, P. H. Sudmant, B. H. Bach, L. L. Smith, and J. A. McGuire. 2022.
3297 MtGenome for the cost of a gene: Long-amplicon PacBio and Nanopore
3298 sequencing of hundreds of full mitochondrial genomes. Europe
3299 PMC:PPR542302.

3300 Keck, F., M. Couton, and F. Altermatt. 2022. Navigating the seven challenges
3301 of taxonomic reference databases in metabarcoding analyses.
3302 Molecular Ecology Resources **23**:742-755.

3303 Kimura, M. 1980. A simple method for estimating evolutionary rates of base
3304 substitutions through comparative studies of nucleotide sequences.
3305 Journal of Molecular Evolution **16**:111-120.

3306 King, C. E., R. Debruyne, M. Kuch, C. Schwarz, and H. N. Poinar. 2009. A
3307 quantitative approach to detect and overcome PCR inhibition in
3308 ancient DNA extracts. BioTechniques **47**:941-949.

3309 King, J. L., B. L. LaRue, N. M. Novroski, M. Stoljarova, S. B. Seo, X. Zeng, D. H.
3310 Warshauer, C. P. Davis, W. Parson, A. Sajantila, and B. Budowle. 2014.
3311 High-quality and high-throughput massively parallel sequencing of the
3312 human mitochondrial genome using the Illumina MiSeq. Forensic
3313 Science International: Genetics **12**:128-135.

3314 Kirtane, A., D. Wieczorek, T. Noji, L. Baskin, C. Ober, R. Plosica, A. Chenoweth,
3315 K. Lynch, and L. Sassoubre. 2021. Quantification of Environmental DNA
3316 (eDNA) shedding and decay rates for three commercially harvested fish
3317 species and comparison between eDNA detection and trawl catches.
3318 Environmental DNA **3**:1142-1155.

3319 Kobayashi, G., H. Itoh, and S. Kojima. 2022. Mitogenome of a stink worm
3320 (Annelida: Traviidae) includes degenerate group II intron that is also
3321 found in five congeneric species. Scientific Reports **12**:4449.

3322 Kolmogorov, M., D. M. Bickhart, B. Behsaz, A. Gurevich, M. Rayko, S. B. Shin,
3323 K. Kuhn, J. Yuan, E. Pevikov, T. P. L. Smith, and P. A. Pevzner. 2020.
3324 metaFlye: scalable long-read metagenome assembly using repeat
3325 graphs. Nature Methods **17**:1103-1110.

3326 Kolmogorov, M., K. J. Billingsley, M. Mastoras, M. Meredith, J. Monlong, R.
3327 Lorig-Roach, M. Asri, P. A. Jerez, L. Malik, R. Dewan, X. Reed, R. M.
3328 Genner, K. Daida, S. Behera, K. Shafin, T. Pesout, J. Prabakaran, P.
3329 Carnevali, J. Yang, A. Rhie, S. W. Scholz, B. J. Traynor, K. H. Miga, M.
3330 Jain, W. Timp, A. M. Phillippy, M. Chaisson, F. J. Sedlazeck, C.
3331 Blauwendraat, and B. Paten. 2023. Scalable Nanopore sequencing of
3332 human genomes provides a comprehensive view of haplotype-
3333 resolved variation and methylation. Nature Methods **20**:1483-1492.

3334 Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. Assembly of long,
3335 error-prone reads using repeat graphs. Nature Biotechnology **37**:540-
3336 546.

3337 Krehenwinkel, H., M. Fong, S. Kennedy, E. G. Huang, S. Noriyuki, L. Cayetano,
3338 and R. Gillespie. 2018. The effect of DNA degradation bias in passive
3339 sampling devices on metabarcoding studies of arthropod communities
3340 and their associated microbiota. PLOS ONE **13**:e0189188.

3341 Krehenwinkel, H., M. Wolf, J. Y. Lim, A. J. Rominger, W. B. Simison, and R. G.
3342 Gillespie. 2017. Estimating and mitigating amplification bias in
3343 qualitative and quantitative arthropod metabarcoding. *Scientific*
3344 *Reports* **7**:17668.

3345 Kumar, G., A. Abd-Elfattah, and M. El-Matbouli. 2015. Identification of
3346 differentially expressed genes of brown trout (*Salmo trutta*) and
3347 rainbow trout (*Oncorhynchus mykiss*) in response to *Tetracapsuloides*
3348 *bryosalmonae* (Myxozoa). *Parasitology Research* **114**:929-939.

3349 Kumar, G., J. E. Eble, and M. R. Gaither. 2020a. A practical guide to sample
3350 preservation and pre-PCR processing of aquatic environmental DNA.
3351 *Molecular Ecology Resources* **20**:29-39.

3352 Kumar, G., R. Ertl, J. L. Bartholomew, and M. El-Matbouli. 2020b. First
3353 transcriptome analysis of bryozoan *Fredericella sultana*, the primary
3354 host of myxozoan parasite *Tetracapsuloides bryosalmonae*. *PeerJ*
3355 **8**:e9027.

3356 Kumar, G., R. Ertl, J. L. Bartholomew, and M. El-Matbouli. 2020c.
3357 Transcriptome analysis elucidates the key responses of bryozoan
3358 *Fredericella sultana* during the development of *Tetracapsuloides*
3359 *bryosalmonae* (Myxozoa). *International Journal of Molecular Sciences*
3360 **21**:5910.

3361 Kumar, G., R. Ertl, F. Nilsen, J. L. Bartholomew, and M. El-Matbouli. 2021. Data
3362 of de novo transcriptome assembly of the myxozoan parasite
3363 *Tetracapsuloides bryosalmonae*. *Data in Brief* **35**:106831.

3364 Kumar, G., E. Farrell, A. M. Reaume, J. A. Eble, and M. R. Gaither. 2022. One
3365 size does not fit all: Tuning eDNA protocols for high- and low-turbidity
3366 water sampling. *Environmental DNA* **4**:167-180.

3367 Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary
3368 Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology*
3369 *and Evolution* **33**:1870-1874.

3370 Lacoursière-Roussel, A., K. Howland, E. Normandeau, E. K. Grey, P.
3371 Archambault, K. Deiner, D. M. Lodge, C. Hernandez, N. Leduc, and L.
3372 Bernatchez. 2018. eDNA metabarcoding as a new surveillance
3373 approach for coastal Arctic biodiversity. *Ecology and Evolution* **8**:7763-
3374 7777.

3375 Lambowitz, A. M., and M. Belfort. 2015. Mobile bacterial group II introns at
3376 the crux of eukaryotic evolution. *Mobile DNA III*:1209-1236.

3377 Landínez-García, R. M., J. C. Narváez, and E. J. Márquez. 2020. Population
3378 genetics of the freshwater fish *Prochilodus magdalenae*
3379 (Characiformes: Prochilodontidae), using species-specific
3380 microsatellite loci. *PeerJ* **8**:e10327.

3381 Laroche, O., S. A. Wood, L. A. Tremblay, J. I. Ellis, G. Lear, and X. Pochon. 2018.
3382 A cross-taxa study using environmental DNA/RNA metabarcoding to
3383 measure biological impacts of offshore oil and gas drilling and
3384 production operations. *Marine Pollution Bulletin* **127**:97-107.

3385 Lavrov, D. V., and W. Pett. 2016. Animal mitochondrial DNA as we do not know
3386 it: mt-Genome organization and evolution in nonbilaterian lineages.
3387 *Genome Biology and Evolution* **8**:2896-2913.

- 3388 Lawson Handley, L., D. S. Read, I. J. Winfield, H. Kimbell, H. Johnson, J. Li, C.
3389 Hahn, R. Blackman, R. Wilcox, R. Donnelly, A. Szitenberg, and B.
3390 Hänfling. 2019. Temporal and spatial variation in distribution of fish
3391 environmental DNA in England's largest lake. *Environmental DNA* **1**:26-
3392 39.
- 3393 Leach, K., W. I. Montgomery, and N. Reid. 2016. Modelling the influence of
3394 biotic factors on species distribution patterns. *Ecological Modelling*
3395 **337**:96-106.
- 3396 Lear, G., I. Dickie, J. Banks, S. Boyer, H. L. Buckley, T. R. Buckley, R. Cruickshank,
3397 A. Dopheide, K. M. Handley, and S. Hermans. 2018. Methods for the
3398 extraction, storage, amplification and sequencing of DNA from
3399 environmental samples. *New Zealand Journal of Ecology* **42**:10-50A.
- 3400 Li, J., L. J. Lawson Handley, D. S. Read, and B. Hänfling. 2018. The effect of
3401 filtration method on the efficiency of environmental DNA capture and
3402 quantification via metabarcoding. *Molecular Ecology Resources*
3403 **18**:1102-1114.
- 3404 Li, X., and Y. Jiang. 2024. Research progress of group II intron splicing factors
3405 in land plant mitochondria. *Genes* **15**:176.
- 3406 Lisnerová, M., P. Blabolil, A. Holzer, P. Jurajda, and I. Fiala. 2020. Myxozoan
3407 hidden diversity: the case of *Myxobolus pseudodispar* Gorbunova,
3408 1936. *Folia Parasitologica* **67**:1-12.
- 3409 Lisnerová, M., A. Holzer, P. Blabolil, and I. Fiala. 2023. Evaluation and
3410 optimization of an eDNA metabarcoding assay for detection of
3411 freshwater myxozoan communities. *Environmental DNA* **5**:312-325.
- 3412 Liu, M., L. J. Clarke, S. C. Baker, G. J. Jordan, and C. P. Burridge. 2020. A practical
3413 guide to DNA metabarcoding for entomological ecologists. *Ecological*
3414 *Entomology* **45**:373-385.
- 3415 Liu, T., and A. M. Pyle. 2021. Discovery of highly reactive self-splicing group II
3416 introns within the mitochondrial genomes of human pathogenic fungi.
3417 *Nucleic Acids Research* **49**:12422-12432.
- 3418 Liu, Y., A. Lövy, Z. Gu, and I. Fiala. 2019. Phylogeny of Myxobolidae (Myxozoa)
3419 and the evolution of myxospore appendages in the Myxobolus clade.
3420 *International Journal for Parasitology* **49**:523-530.
- 3421 Locke, S. A., J. D. McLaughlin, and D. J. Marcogliese. 2013. Predicting the
3422 similarity of parasite communities in freshwater fishes using the
3423 phylogeny, ecology and proximity of hosts. *Oikos* **122**:73-83.
- 3424 Longshaw, M., R.-M. Le Deuff, A. F. Harris, and S. W. Feist. 2002. Development
3425 of proliferative kidney disease in rainbow trout, *Oncorhynchus mykiss*
3426 (Walbaum), following short-term exposure to *Tetracapsula*
3427 *bryosalmonae* infected bryozoans. *Journal of Fish Diseases* **25**:443-449.
- 3428 Loos, L. M., and R. Nijland. 2020. Biases in bulk: DNA metabarcoding of marine
3429 communities and the methodology involved. *Molecular Ecology*.
- 3430 Loose, M., S. Malla, and M. Stout. 2016. Real-time selective sequencing using
3431 nanopore technology. *Nature Methods* **13**:751-754.
- 3432 Lugg, W. H., J. Griffiths, A. R. Rooyen, A. R. Weeks, and R. Tingley. 2018.
3433 Optimal survey designs for environmental DNA sampling. *Methods in*
3434 *Ecology and Evolution* **9**:1049-1059.

- 3435 Luo, M., S. Wang, S. Saavedra, D. Ebert, and F. Altermatt. 2022. Multispecies
3436 coexistence in fragmented landscapes. *Proceedings of the National*
3437 *Academy of Sciences* **119**:e2201503119.
- 3438 Macaulay, S., A. R. Ellison, P. Kille, and J. Cable. 2022. Moving towards
3439 improved surveillance and earlier diagnosis of aquatic pathogens: From
3440 traditional methods to emerging technologies. *Reviews in Aquaculture*
3441 **14**:1813-1829.
- 3442 Macé, B., R. Hocdé, V. Marques, P. E. Guerin, A. Valentini, V. Arnal, L. Pellissier,
3443 and S. Manel. 2022. Evaluating bioinformatics pipelines for population-
3444 level inference using environmental DNA. *Environmental DNA* **4**:674-
3445 686.
- 3446 Magurran, A. E., and P. A. Henderson. 2003. Explaining the excess of rare
3447 species in natural species abundance distributions. *Nature* **422**:714-
3448 716.
- 3449 Mant, R., G. Moggridge, and D. Aldridge. 2011. Biofouling by bryozoans,
3450 Cordylophora and sponges in UK water treatment works. *Water*
3451 *Science and Technology* **63**:1815-1822.
- 3452 Mant, R. C., G. D. Moggridge, and D. C. Aldridge. 2013. Control of the biofouling
3453 bryozoan, *Plumatella repens*, using pulsed chlorine treatment.
3454 *International Journal of Environmental Science and Technology*
3455 **10**:199-208.
- 3456 Marshall, N. T., and C. A. Stepien. 2019. Invasion genetics from eDNA and
3457 thousands of larvae: A targeted metabarcoding assay that distinguishes
3458 species and population variation of zebra and quagga mussels. *Ecology*
3459 *and Evolution* **9**:3515-3538.
- 3460 Marshall, N. T., and C. A. Stepien. 2020. Macroinvertebrate community
3461 diversity and habitat quality relationships along a large river from
3462 targeted eDNA metabarcode assays. *Environmental DNA* **2**:572-586.
- 3463 Marshall, W. L., T. Macwilliam, K. Williams, H. Reinholt, H. Vanvliet, D. New, M.
3464 Mills, and D. Morrison. 2022. Detection of *Kudoa thyrsites* (Myxozoa)
3465 eDNA by real-time and digital PCR from high seawater volumes. *Journal*
3466 *of Fish Diseases* **45**:1403.
- 3467 Martins, F. M. S., M. Porto, M. J. Feio, B. Egeter, A. Bonin, S. R. Q. Serra, P.
3468 Taberlet, and P. Beja. 2020. Modelling technical and biological biases
3469 in macroinvertebrate community assessment from bulk preservative
3470 using multiple metabarcoding markers. *Molecular Ecology* **30**:3221-
3471 3238.
- 3472 Mascolo, C., M. Ceruso, P. Sordino, G. Palma, A. Anastasio, and T. Pepe. 2017.
3473 Development of a method to extract and amplify the complete
3474 mitogenome of some Sparidae species. *Italian Journal of Food Safety*
3475 **6**:75-77.
- 3476 Mascolo, C., M. Ceruso, P. Sordino, G. Palma, A. Anastasio, and T. Pepe. 2019.
3477 Comparison of mitochondrial DNA enrichment and sequencing
3478 methods from fish tissue. *Food Chemistry* **294**:333-338.
- 3479 Matthews, T. J., and R. J. Whittaker. 2015. On the species abundance
3480 distribution in applied ecology and biodiversity management. *Journal*
3481 *of Applied Ecology* **52**:443-454.

- McGrath, K. C., S. R. Thomas-Hall, C. T. Cheng, L. Leo, A. Alexa, S. Schmidt, and P. M. Schenk. 2008. Isolation and analysis of mRNA from environmental microbial communities. *Journal of Microbiological Methods* **75**:172-176.
- McKee, A. M., S. F. Spear, and T. W. Pierson. 2015. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biological Conservation* **183**:70-76.
- Michel, F., U. Kazuhiko, and O. Haruo. 1989. Comparative and functional anatomy of group II catalytic introns — a review. *Gene* **82**:5-30.
- Minamoto, T. 2022. Environmental DNA analysis for macro-organisms: species distribution and more. *DNA Research* **29**:dsac018.
- Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. ColabFold: making protein folding accessible to all. *Nature Methods* **19**:679-682.
- Miyata, K., Y. Inoue, Y. Amano, T. Nishioka, T. Nagaike, T. Kawaguchi, O. Morita, M. Yamane, and H. Honda. 2022. Comparative environmental RNA and DNA metabarcoding analysis of river algae and arthropods for ecological surveys and water quality assessment. *Scientific Reports* **12**:19828.
- Monge, O., D. Dumas, and I. Baus. 2020. Environmental DNA from avian residual saliva in fruits and its potential uses in population genetics. *Conservation Genetics Resources* **12**:131-139.
- Morales-Castilla, I., P. Pappalardo, M. J. Farrell, A. A. Aguirre, S. Huang, A.-L. M. Gehman, T. Dallas, D. Gravel, and T. J. Davies. 2021. Forecasting parasite sharing under climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**:20200360.
- Morris, D. J., and A. Adams. 2007. Sacculogenesis and sporogony of *Tetracapsuloides bryosalmonae* (Myxozoa: Malacosporea) within the bryozoan host *Fredericella sultana* (Bryozoa: Phylactolaemata). *Parasitology Research* **100**:983-992.
- Morris, D. J., D. C. Morris, and A. Adams. 2002. Development and release of a malacosporean (Myxozoa) from *Plumatella repens* (Bryozoa: Phylactolaemata). *Folia Parasitologica* **49**:25-34.
- Naldoni, J., E. A. Adriano, A. Hartigan, C. Sayer, and B. Okamura. 2019. Malacosporean myxozoans exploit a diversity of fish hosts. *Parasitology* **146**:968-978.
- Nicholls, T. J., and M. Minczuk. 2014. In D-loop: 40years of mitochondrial 7S DNA. *Experimental Gerontology* **56**:175-181.
- Nosek, J., and Ľ. Tomáška. 2003. Mitochondrial genome diversity: evolution of the molecular architecture and replication strategy. *Current Genetics* **44**:73-84.
- Novikova, O., and M. Belfort. 2017. Mobile group II introns as ancestral eukaryotic elements. *Trends in Genetics* **33**:773-783.
- Okamura, B., A. Gruhl, and J. L. Bartholomew. 2015. An Introduction to Myxozoan evolution, ecology and development. Pages 1-20 *Myxozoan evolution, ecology and development*. Springer International Publishing.

3529 Okamura, B., A. Hartigan, and J. Naldoni. 2018. Extensive uncharted
3530 biodiversity: the parasite dimension. *Integrative and Comparative*
3531 *Biology* **58**:1132–1145.

3532 Okamura, B., H. Hartikainen, H. Schmidt-Posthaus, and T. Wahli. 2011. Life
3533 cycle complexity, environmental change and the emerging status of
3534 salmonid proliferative kidney disease. *Freshwater Biology* **56**:735-753.

3535 Okamura, B., H. Hartikainen, and J. Trew. 2019. Waterbird-mediated dispersal
3536 and freshwater biodiversity: General insights from bryozoans. *Frontiers*
3537 *in Ecology and Evolution* **7**:29.

3538 Okamura, B., and T. S. Wood. 2002. Bryozoans as hosts for *Tetracapsula*
3539 *bryosalmonae*, the PKX organism. *Journal of Fish Diseases* **25**:469-475.

3540 Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. G. O'Hara, G. Simpson, P.
3541 Solymos, H. Stevens, and H. Wagner. 2013. Multivariate analysis of
3542 ecological communities in R: vegan tutorial. R package version 1.7.

3543 Olds, B. P., C. L. Jerde, M. A. Renshaw, Y. Li, N. T. Evans, C. R. Turner, K. Deiner,
3544 A. R. Mahon, M. A. Brueseke, P. D. Shirey, M. E. Pfrender, D. M. Lodge,
3545 and G. A. Lamberti. 2016. Estimating species richness using
3546 environmental DNA. *Ecology and Evolution* **6**:4214-4226.

3547 Oredalen, T. J., T. A. Mo, A. Jenkins, N. Haugan, and M. Sæbø. 2022. Use of
3548 environmental DNA to detect the myxozoan endoparasite
3549 *Tetracapsuloides bryosalmonae* in large Norwegian lakes.
3550 *Environmental DNA* **4**:1294-1322.

3551 Orr, R. J. S., M. N. Haugen, B. Berning, P. Bock, R. L. Cumming, W. K. Florence,
3552 M. Hirose, E. Di Martino, M. H. Ramsfjell, M. M. Sannum, A. M. Smith,
3553 L. M. Vieira, A. Waeschenbach, and L. H. Liow. 2019. A genome-
3554 skimmed phylogeny of a widespread bryozoan family, Adeonidae. *BMC*
3555 *Evolutionary Biology* **19**:235.

3556 Ostroumov, S. A. 2005. Some aspects of water filtering activity of filter-feeders.
3557 *Hydrobiologia* **542**:275-286.

3558 Palmer, J. D., and J. M. Logsdon. 1991. The recent origins of introns. *Current*
3559 *Opinion in Genetics & Development* **1**:470-477.

3560 Parro, V., M. Moreno-Paz, and E. González-Toril. 2007. Analysis of
3561 environmental transcriptomes by DNA microarrays. *Environmental*
3562 *Microbiology* **9**:453-464.

3563 Parsons, K. M., M. Everett, M. Dahlheim, and L. Park. 2018. Water, water
3564 everywhere: environmental DNA can unlock population structure in
3565 elusive marine species. *Royal Society Open Science* **5**:180537.

3566 Peloquin, J. J., K. B. David Mc, and E. G. Platzer. 1993. Rapid Miniprep Isolation
3567 of Mitochondrial DNA from Metacestodes, and Free-Living and
3568 Parasitic Nematodes. *The Journal of Parasitology* **79**:964-967.

3569 Perry, W. B., M. Seymour, L. Orsini, I. B. Jâms, N. Milner, F. Edwards, R. Harvey,
3570 M. De Bruyn, I. Bista, K. Walsh, B. Emmett, R. Blackman, F. Altermatt, L.
3571 Lawson Handley, E. Mächler, K. Deiner, H. M. Bik, G. Carvalho, J.
3572 Colbourne, B. J. Cosby, I. Durance, and S. Creer. 2024. An integrated
3573 spatio-temporal view of riverine biodiversity using environmental DNA
3574 metabarcoding. *Nature Communications* **15**:4372.

Pinfield, R., E. Dillane, A. K. W. Runge, A. Evans, L. Mirimin, J. Niemann, T. E. Reed, D. G. Reid, E. Rogan, F. I. P. Samarra, E. E. Sigsgaard, and A. D. Foote. 2019. False-negative detections from environmental DNA collected in the presence of large numbers of killer whales (*Orcinus orca*). *Environmental DNA* **1**:316-328.

Pochon, X., A. Zaiko, L. M. Fletcher, O. Laroche, and S. A. Wood. 2017. Wanted dead or alive? Using metabarcoding of environmental DNA and RNA to distinguish living assemblages for biosecurity applications. *PLOS ONE* **12**:e0187636.

Pont, D., M. Rocle, A. Valentini, R. Civade, P. Jean, A. Maire, N. Roset, M. Schabuss, H. Zornig, and T. Dejean. 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Scientific Reports* **8**:10361.

Poretzky, R. S., S. Gifford, J. Rinta-Kanto, M. Vila-Costa, and M. A. Moran. 2009. Analyzing gene expression from marine microbial communities using environmental transcriptomics. *Journal of visualized experiments: JoVE*:1086.

Poulin, R., and H. S. Randhawa. 2015. Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology* **142**:S6-S15.

Pulkkinen, K., L. R. Suomalainen, A. F. Read, D. Ebert, P. Rintamäki, and E. T. Valtonen. 2010. Intensive fish farming and the evolution of pathogen virulence: the case of columnaris disease in Finland. *Proceedings of the Royal Society B: Biological Sciences* **277**:593-600.

Quispe-Tintaya, W., R. R. White, V. N. Popov, J. Vijg, and A. Y. Maslov. 2013. Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. *BioTechniques* **55**:133-136.

Quispe-Tintaya, W., R. R. White, V. N. Popov, J. Vijg, and A. Y. Maslov. 2015. Rapid Mitochondrial DNA Isolation Method for Direct Sequencing. Pages 89-95. Springer New York.

Rackham, O., and A. Filipovska. 2022. Organization and expression of the mammalian mitochondrial genome. *Nature Reviews Genetics* **23**:606-623.

Ramón-Laca, A., K. M. Nichols, K. M. Parsons, R. Gallego, D. R. Vieites, I. Doadrio, and A. Riesgo Gil. 2025. Prospects of environmental nucleic acids (eDNA and eRNA) as multipurpose biomonitoring tools for fisheries management. 389419.

Ramón-Laca, A., R. Gallego, and K. M. Nichols. 2023. Affordable de novo generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments. *Molecular Ecology Resources* **23**:818-832.

Reid, A. J., A. K. Carlson, I. F. Creed, E. J. Eliason, P. A. Gell, P. T. J. Johnson, K. A. Kidd, T. J. Maccormack, J. D. Olden, S. J. Ormerod, J. P. Smol, W. W. Taylor, K. Tockner, J. C. Vermaire, D. Dudgeon, and S. J. Cooke. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews* **94**:849-873.

Reinholdt Jensen, M., E. Egelyng Sigsgaard, S. Agersnap, J. Jessen Rasmussen, A. Baattrup-Pedersen, P. Wiberg-Larsen, and P. Francis Thomsen. 2021.

3622 Seasonal turnover in community composition of stream-associated
3623 macroinvertebrates inferred from freshwater environmental DNA
3624 metabarcoding. *Environmental DNA* **3**:861-876.

3625 Renshaw, M. A., B. P. Olds, C. L. Jerde, M. M. McVeigh, and D. M. Lodge. 2015.
3626 The room temperature preservation of filtered environmental DNA
3627 samples and assimilation into a phenol–chloroform–isoamyl alcohol
3628 DNA extraction. *Molecular Ecology Resources* **15**:168-176.

3629 Richter, S., F. Schwarz, L. Hering, M. Böggemann, and C. Bleidorn. 2015. The
3630 utility of genome skimming for phylogenomic analyses as
3631 demonstrated for glycerid relationships (Annelida, Glyceridae).
3632 *Genome Biology and Evolution* **7**:3443-3462.

3633 Robinson, C. V., T. M. Uren Webster, J. Cable, J. James, and S. Consuegra. 2018.
3634 Simultaneous detection of invasive signal crayfish, endangered white-
3635 clawed crayfish and the crayfish plague pathogen using environmental
3636 DNA. *Biological Conservation* **222**:241-252.

3637 Robinson, H. E., J. D. Alexander, J. L. Bartholomew, S. L. Hallett, N. J. Hetrick, R.
3638 W. Perry, and N. A. Som. 2022. Using a mechanistic framework to
3639 model the density of an aquatic parasite *Ceratonova shasta*. *PeerJ*
3640 **10**:e13183.

3641 Ros, A., J. Baer, T. Basen, C. Chucholl, E. Schneider, R. Teschner, and A. Brinker.
3642 2021. Current and projected impacts of the parasite *Tetracapsuloides*
3643 *bryosalmonae* (causative to proliferative kidney disease) on Central
3644 European salmonid populations under predicted climate change.
3645 *Freshwater Biology* **66**:1182-1199.

3646 Ros, A., H. Schmidt-Posthaus, and A. Brinker. 2022. Mitigating human impacts
3647 including climate change on proliferative kidney disease in salmonids
3648 of running waters. *Journal of Fish Diseases* **45**:497-521.

3649 Ruiz, E., F. Leprieur, G. Sposito, M. Lüthi, M. Schmidlin, J. Panfili, L. Pellissier,
3650 and C. Albouy. 2025. Environmental DNA epigenetics accurately
3651 predicts the age of cultured fish larvae. *Ecology and Evolution*
3652 **15**:e70645.

3653 Rusch, J. C., H. Hansen, D. A. Strand, T. Markussen, S. Hytterød, and T. Vrålstad.
3654 2018. Catching the fish with the worm: a case study on eDNA detection
3655 of the monogenean parasite *Gyrodactylus salaris* and two of its hosts,
3656 Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus*
3657 *mykiss*). *Parasites & Vectors* **11**:333.

3658 Saito, T., and H. Doi. 2020. Environmental DNA degradation simulation from
3659 water temperature and DNA fragment length: A meta-analysis
3660 approach. Cold Spring Harbor Laboratory.

3661 Sanches, T. M., and A. D. Schreier. 2020. Optimizing an eDNA protocol for
3662 estuarine environments: Balancing sensitivity, cost and time. *PLOS ONE*
3663 **15**:e0233522.

3664 Sansom, B. J., and L. M. Sassoubre. 2017. Environmental DNA (eDNA) shedding
3665 and decay rates to model freshwater mussel eDNA transport in a river.
3666 *Environmental Science & Technology* **51**:14244-14253.

3667 Sato, H., Y. Sogo, H. Doi, and H. Yamanaka. 2017. Usefulness and limitations of
3668 sample pooling for environmental DNA metabarcoding of freshwater
3669 fish communities. *Scientific Reports* **7**:14860.

3670 Schenekar, T. 2022. The current state of eDNA research in freshwater
3671 ecosystems: are we shifting from the developmental phase to standard
3672 application in biomonitoring? *Hydrobiologia*:1263–1282.

3673 Schmidt-Kloiber, A., and D. Hering. 2015. www.freshwaterecology.info – An
3674 online tool that unifies, standardises and codifies more than 20,000
3675 European freshwater organisms and their ecological preferences.
3676 *Ecological Indicators* **53**:271-282.

3677 Schmidt-Posthaus, H., E. Schneider, N. Schölzel, R. Hirschi, M. Stelzer, and A.
3678 Peter. 2021. The role of migration barriers for dispersion of
3679 Proliferative Kidney Disease—Balance between disease emergence
3680 and habitat connectivity. *PLOS ONE* **16**:e0247482.

3681 Schuster, A., J. V. Lopez, L. E. Becking, M. Kelly, S. A. Pomponi, G. Wörheide, D.
3682 Erpenbeck, and P. Cárdenas. 2017. Evolution of group I introns in
3683 Porifera: new evidence for intron mobility and implications for DNA
3684 barcoding. *BMC Evolutionary Biology* **17**:82.

3685 Schuster, C. J., M. L. Kent, J. Peterson, and J. L. Sanders. 2022. Multi-state
3686 occupancy model estimates probability of detection of an aquatic
3687 parasite using environmental DNA: *Pseudoloma neurophilia* in
3688 zebrafish aquaria. *The Journal of Parasitology* **108**:527-538.

3689 Sepulveda, A. J., J. Schabacker, S. Smith, A. C. Robert, G. Luikart, and S. J. Amish.
3690 2019. Improved detection of rare, endangered and invasive trout in
3691 using a new large-volume sampling method for eDNA capture.
3692 *Environmental DNA* **1**:227-237.

3693 Sethi, S. A., W. Larson, K. Turnquist, and D. Isermann. 2019. Estimating the
3694 number of contributors to DNA mixtures provides a novel tool for
3695 ecology. *Methods in Ecology and Evolution* **10**:109-119.

3696 Seymour, M., F. K. Edwards, B. J. Cosby, I. Bista, P. M. Scarlett, F. L. Brailsford,
3697 H. C. Glanville, M. De Bruyn, G. R. Carvalho, and S. Creer. 2021.
3698 Environmental DNA provides higher resolution assessment of riverine
3699 biodiversity and ecosystem function via spatio-temporal nestedness
3700 and turnover partitioning. *Communications Biology* **4**:512.

3701 Shogren, A. J., J. L. Tank, S. P. Egan, D. Bolster, and T. Riis. 2019. Riverine
3702 distribution of mussel environmental DNA reflects a balance among
3703 density, transport, and removal processes. *Freshwater Biology*
3704 **64**:1467-1479.

3705 Shtolz, N., and D. Mishmar. 2023. The metazoan landscape of mitochondrial
3706 DNA gene order and content is shaped by selection and affects
3707 mitochondrial transcription. *Communications Biology* **6**:93.

3708 Shu, L., A. Ludwig, and Z. Peng. 2020. Standards for Methods Utilizing
3709 Environmental DNA for Detection of Fish Species. *Genes* **11**:296.

3710 Sidstedt, M., P. Rådström, and J. Hedman. 2020. PCR inhibition in qPCR, dPCR
3711 and MPS—mechanisms and solutions. *Analytical and Bioanalytical*
3712 *Chemistry* **412**:2009-2023.

- 3713 Sieber, N., H. Hartikainen, and C. Vorburger. 2020. Validation of an eDNA-
3714 based method for the detection of wildlife pathogens in water.
3715 Diseases of Aquatic Organisms **141**:171-184.
- 3716 Sigsgaard, E. E., M. R. Jensen, I. E. Winkelmann, P. R. Møller, M. M. Hansen,
3717 and P. F. Thomsen. 2020. Population-level inferences from
3718 environmental DNA—Current status and future perspectives.
3719 Evolutionary Applications **13**:245-262.
- 3720 Sigsgaard, E. E., I. B. Nielsen, S. S. Bach, E. D. Lorenzen, D. P. Robinson, S. W.
3721 Knudsen, M. W. Pedersen, M. A. Jaidah, L. Orlando, E. Willerslev, P. R.
3722 Møller, and P. F. Thomsen. 2017a. Population characteristics of a large
3723 whale shark aggregation inferred from seawater environmental DNA.
3724 Nature Ecology & Evolution **1**:0004
- 3725 Sigsgaard, E. E., I. B. Nielsen, H. Carl, M. A. Krag, S. W. Knudsen, Y. Xing, T. H.
3726 Holm-Hansen, P. R. Møller, and P. F. Thomsen. 2017b. Seawater
3727 environmental DNA reflects seasonality of a coastal fish community.
3728 Marine Biology **164**:128.
- 3729 Spens, J., A. R. Evans, D. Halfmaerten, S. W. Knudsen, M. E. Sengupta, S. S. T.
3730 Mak, E. E. Sigsgaard, and M. Hellström. 2017. Comparison of capture
3731 and storage methods for aqueous microbial eDNA using an optimized
3732 extraction protocol: advantage of enclosed filter. Methods in Ecology
3733 and Evolution **8**:635-645.
- 3734 Stat, M., M. J. Huggett, R. Bernasconi, J. D. Dibattista, T. E. Berry, S. J. Newman,
3735 E. S. Harvey, and M. Bunce. 2017. Ecosystem biomonitoring with eDNA:
3736 metabarcoding across the tree of life in a tropical marine environment.
3737 Scientific Reports **7**:12240.
- 3738 Stelzer, M., J. Ord, S. Neyrinck, J. Steiner, H. Hartikainen, R. Brys, and H.
3739 Schmidt-Posthaus. 2024. Optimal detection protocol of
3740 *Tetracapsuloides bryosalmonae* by environmental DNA: A comparison
3741 of qPCR and ddPCR approaches. Environmental DNA **6**:e501.
- 3742 Stepien, C. A., M. R. Snyder, and A. E. Elz. 2019. Invasion genetics of the silver
3743 carp *Hypophthalmichthys molitrix* across North America:
3744 Differentiation of fronts, introgression, and eDNA metabarcode
3745 detection. PLOS ONE **14**:e0203012.
- 3746 Stoeckle, B. C., S. Beggel, R. Kuehn, and J. Geist. 2021. Influence of stream
3747 characteristics and population size on downstream transport of
3748 freshwater mollusk environmental DNA. Freshwater Science **40**:191-
3749 201.
- 3750 Strepparava, N., H. Segner, A. Ros, H. Hartikainen, H. Schmidt-Posthaus, and T.
3751 Wahli. 2018. Temperature-related parasite infection dynamics: the
3752 case of proliferative kidney disease of brown trout. Parasitology
3753 **145**:281-291.
- 3754 Strickler, K. M., A. K. Fremier, and C. S. Goldberg. 2015. Quantifying effects of
3755 UV-B, temperature, and pH on eDNA degradation in aquatic
3756 microcosms. Biological Conservation **183**:85-92.
- 3757 Sudhagar, A., M. El-Matbouli, and G. Kumar. 2020. Identification and
3758 Expression Profiling of Toll-Like Receptors of Brown Trout (*Salmo*

3759 *trutta*) during Proliferative Kidney Disease. International Journal of
3760 Molecular Sciences **21**:3755.

3761 Sudhagar, A., R. Ertl, G. Kumar, and M. El-Matbouli. 2019. Transcriptome
3762 profiling of posterior kidney of brown trout, *Salmo trutta*, during
3763 proliferative kidney disease. Parasites & Vectors **12**.

3764 Svavarsdóttir, F. R. 2016. Proliferative kidney disease (PKD) in Icelandic fresh
3765 water. Distribution and prevalence of *Tetracapsuloides bryosalmonae*
3766 and its effect on salmonid populations in Iceland.

3767 Syrová, E., M. Palíková, J. Mendel, V. Seidlová, I. Papežíková, H. Schmidt-
3768 Posthaus, K. Somerlíková, H. Minářová, L. Mareš, I. Mikulíková, J.
3769 Pikula, and J. Mareš. 2020. Field study indicating susceptibility
3770 differences between salmonid species and their lineages to
3771 proliferative kidney disease. Journal of Fish Diseases **43**:1201-1211.

3772 Székely, D., N. L. Corfixen, L. L. Mørch, S. W. Knudsen, M. L. McCarthy, J.
3773 Teilmann, M. P. Heide-Jørgensen, and M. T. Olsen. 2021.
3774 Environmental DNA captures the genetic diversity of bowhead whales
3775 (*Balaena mysticetus*) in West Greenland. Environmental DNA **3**:248-
3776 260.

3777 Szitenberg, A., C. Rot, M. Ilan, and D. Huchon. 2010. Diversity of sponge
3778 mitochondrial introns revealed by cox1 sequences of Tetillidae. BMC
3779 Evolutionary Biology **10**:288.

3780 Taberlet, P., and G. Luikart. 1999. Non-invasive genetic sampling and individual
3781 identification. Biological journal of the linnean society **68**:41-55.

3782 Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling:
3783 look before you leap. Trends in Ecology & Evolution **14**:323-327.

3784 Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis
3785 by DNA polymorphism. Genetics **123**:585-595.

3786 Talluto, M., R. Del Campo, E. Estévez, F. Altermatt, T. Datry, and G. Singer.
3787 2024. Towards (better) fluvial meta-ecosystem ecology: a research
3788 perspective. npj Biodiversity **3**:3.

3789 Templeton, A. R. 1980. The theory of speciation via the founder principle.
3790 Genetics **94**:1011-1038.

3791 Thaling, B., K. Deiner, L. R. Harper, H. C. Rees, R. C. Blackman, D. Sint, M.
3792 Traugott, C. S. Goldberg, and K. Bruce. 2021a. A validation scale to
3793 determine the readiness of environmental DNA assays for routine
3794 species monitoring. Environmental DNA **3**:823-836.

3795 Thaling, B., D. Kirschner, Y. Pütz, C. Moritz, R. Schwarzenberger, J.
3796 Wanzenböck, and M. Traugott. 2021b. Lateral and longitudinal fish
3797 environmental DNA distribution in dynamic riverine habitats.
3798 Environmental DNA **3**:305-318.

3799 Tockner, K., and J. Ward. 1999. Biodiversity along riparian corridors. Large
3800 rivers **11**:293-310.

3801 Tops, S., and B. Okamura. 2003. Infection of bryozoans by *Tetracapsuloides*
3802 *bryosalmonae* at sites endemic for salmonid proliferative kidney
3803 disease. Diseases of Aquatic Organisms **57**:221-226.

3804 Trevisan, B., D. M. C. Alcantara, D. J. Machado, F. P. L. Marques, and D. J. G.
3805 Lahr. 2019. Genome skimming is a low-cost and robust strategy to

assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. *PeerJ* **7**:e7543.

Troth, C. R., M. J. Sweet, J. Nightingale, and A. Burian. 2021. Seasonality, DNA degradation and spatial heterogeneity as drivers of eDNA detection dynamics. *Science of The Total Environment* **768**:144466.

Tsuji, S., A. Maruyama, M. Miya, M. Ushio, H. Sato, T. Minamoto, and H. Yamanaka. 2020. Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Molecular Ecology Resources* **20**:1248-1258.

Tsuji, S., M. Miya, M. Ushio, H. Sato, T. Minamoto, and H. Yamanaka. 2018. Evaluating intraspecific genetic diversity of a fish population using environmental DNA: An approach to distinguish true haplotypes from erroneous sequences. *Europe PMC*:PPR57359

Turner, C. R., M. A. Barnes, C. C. Y. Xu, S. E. Jones, C. L. Jerde, and D. M. Lodge. 2014. Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods in Ecology and Evolution* **5**:676-684.

Turon, X., A. Antich, C. Palacín, K. Præbel, and O. S. Wangensteen. 2020. From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications* **30**:e02036.

Tzafesta, E., and M. Shokri. 2025. The combined negative effect of temperature, UV radiation and salinity on eDNA detection: A global meta-analysis on aquatic ecosystems. *Ecological Indicators* **176**:113669.

Uchii, K., H. Doi, H. Yamanaka, and T. Minamoto. 2017a. Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution* **7**:8515-8522.

Uchii, K., H. Doi, H. Yamanaka, and T. Minamoto. 2017b. Distinct seasonal migration patterns of Japanese native and non-native genotypes of common carp estimated by environmental DNA. *Ecology and Evolution* **7**:8515-8522.

Urban, L., A. Miller, D. Eason, D. Vercoe, M. Shaffer, S. Wilkinson, J. Guhlin, P. Dearden, G.-J. Jeunen, and N. Gemmell. 2022. Genomic monitoring of the critically endangered Kākāpō by real-time targeted Nanopore sequencing of environmental DNA. *SSRN*:3977260.

Uthicke, S., M. Lamare, and J. R. Doyle. 2018. eDNA detection of corallivorous seastar (*Acanthaster* cf. *solaris*) outbreaks on the Great Barrier Reef using digital droplet PCR. *Coral Reefs* **37**:1229-1239.

Vallès, Y., K. M. Halanych, and J. L. Boore. 2008. Group II introns break new boundaries: Presence in a bilaterian's genome. *PLOS ONE* **3**:e1488.

van Nouhuys, S. 2009. Metapopulation ecology. *eLS*.

Veeraragavan, S., M. Johansen, and I. G. Johnston. 2024. Evolution and maintenance of mtDNA gene content across eukaryotes. *Biochemical Journal* **481**:1015-1042.

Von Ammon, U., S. A. Wood, O. Laroche, A. Zaiko, S. D. Lavery, G. J. Inglis, and X. Pochon. 2019. Linking environmental DNA and RNA for improved

3852 detection of the marine invasive fanworm *Sabella spallanzanii*.
3853 Frontiers in Marine Science **6**:621.

3854 Wadgymar, S. M., M. L. DeMarche, E. B. Josephs, S. N. Sheth, and J. T.
3855 Anderson. 2022. Local adaptation: Causal agents of selection and
3856 adaptive trait divergence. Annual Review of Ecology, Evolution, and
3857 Systematics **53**:87-111.

3858 Waits, L. P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of
3859 identity among genotypes in natural populations: cautions and
3860 guidelines. Molecular Ecology **10**:249-256.

3861 Waits, L. P., and D. Paetkau. 2005. Non-invasive genetic sampling tools for
3862 wildlife biologists: A review of applications and recommendations for
3863 accurate data collection. Journal of Wildlife Management **69**:1419-
3864 1433.

3865 Ward, J. V. 1998. Riverine landscapes: Biodiversity patterns, disturbance
3866 regimes, and aquatic conservation. Biological Conservation **83**:269-
3867 278.

3868 Weinstein, S. B., and A. M. Kuris. 2016. Independent origins of parasitism in
3869 Animalia. Biology Letters **12**:20160324.

3870 Weitemier, K., B. E. Penaluna, L. L. Hauck, L. J. Longway, T. Garcia, and R. Cronn.
3871 2021. Estimating the genetic diversity of Pacific salmon and trout using
3872 multigene eDNA metabarcoding. Molecular Ecology **30**:4970-4990.

3873 Wood, Z. T., A. Lacoursière-Roussel, F. LeBlanc, M. Trudel, M. T. Kinnison, C.
3874 Garry McBrine, S. A. Pavey, and N. Gagné. 2021a. Spatial heterogeneity
3875 of eDNA transport improves stream assessment of threatened salmon
3876 presence, abundance, and location. Frontiers in Ecology and Evolution
3877 **9**:650717.

3878 Wood, Z. T., A. Lacoursière-Roussel, F. LeBlanc, M. Trudel, M. T. Kinnison, C.
3879 Garry McBrine, S. A. Pavey, and N. Gagné. 2021b. Spatial heterogeneity
3880 of edna transport improves stream assessment of threatened salmon
3881 presence, abundance, and location. Frontiers in Ecology and Evolution
3882 **9**.

3883 Xu, L., T. Liu, K. Chung, and A. M. Pyle. 2023. Structural insights into intron
3884 catalysis and dynamics during splicing. Nature **624**:682-688.

3885 Yoshitake, K., T. Yoshinaga, C. Tanaka, N. Mizusawa, M. S. Reza, A. Tsujimoto,
3886 T. Kobayashi, and S. Watabe. 2019. HaCeD-Seq: a Novel Method for
3887 Reliable and Easy Estimation About the Fish Population Using
3888 Haplotype Count from eDNA. Marine Biotechnology **21**:813-820.

3889 Yu, X., H. Yang, J. Liu, Y. Qi, L. Sun, and X. Tian. 2022. A strategy for a high
3890 enrichment of insect mitochondrial DNA for mitogenomic analysis.
3891 Gene **808**:145986.

3892 Zaiko, A., U. Von Ammon, J. Stuart, K. F. Smith, R. Yao, M. Welsh, X. Pochon,
3893 and H. A. Bowers. 2022. Assessing the performance and efficiency of
3894 environmental DNA /RNA capture methodologies under controlled
3895 experimental conditions. Methods in Ecology and Evolution **13**:1581-
3896 1594.

3897 Zhao, B., P. M. van Bodegom, and K. Trimbos. 2021. The particle size
3898 distribution of environmental DNA varies with species and
3899 degradation. *Science of The Total Environment* **797**:149175.

3900 Zhao, C., and A. M. Pyle. 2017. The group II intron maturase: a reverse
3901 transcriptase and splicing factor go hand in hand. *Current Opinion in*
3902 *Structural Biology* **47**:30-39.

3903 Zilius, M., A. Samuiloviene, R. Stanislauskienė, E. Broman, S. Bonaglia, R.
3904 Meškys, and A. Zaiko. 2021. Depicting temporal, functional, and
3905 phylogenetic patterns in estuarine diazotrophic communities from
3906 environmental DNA and RNA. *Microbial Ecology* **81**:36-51.

3907 Zimmerly, S., and C. Semper. 2015. Evolution of group II introns. *Mobile DNA*
3908 **6**:7.

3909 Zizka, V. M. A., M. Weiss, and F. Leese. 2020. Can metabarcoding resolve
3910 intraspecific genetic diversity changes to environmental stressors? A
3911 test case using river macrozoobenthos. *Metabarcoding and*
3912 *Metagenomics* **4**:e51925.

3913

Appendices

3914 Chapter 1 – Supplementary

3915 Sampling kit preparation & guide

3916 **Field sampling box:**

- 3917 2x Permanent **markers**
- 3918 1x Trash **bag** to collect used kits
- 3919 1x **Cooler** for sample tubes and tissue samples
- 3920 1x Paper **towels**
- 3921 1x **Disinfectant**
- 3922 Enough **kits** for all sites + 1 negative control

3923 **Kit content (in plastic zip-top bag):**

- 3924 1x Pair of lab **gloves**
- 3925 1x Plastic **syringe** (Luer lock)
- 3926 1x Plastic **tweezers**
- 3927 1x Plastic (zip-top) bag for **sample tubes**
- 3928 2x **Filter holders** (with Luer lock and rubber O-ring, empty)
- 3929 1x **3 Filter papers**, size 5 μm (extra bag)
- 3930 1x **3 Filter papers**, size 0.45 μm (extra bag)
- 3931 3x **Screw-top tubes** for samples (filled with ~200 μl small glass
- 3932 beads and 1ml of Longmire's buffer)

3933

3934 **Preparation:**

- 3935 1. Clean used plastics (syringe, tweezers and filter holders; all taken
- 3936 apart) with diluted bleach (10%) for 20min and rinse thrice, final
- 3937 wash with double-distilled water (Watch out for the O-rings!)
- 3938 2. Put in big plastic box with some paper towels on the bottom, leave
- 3939 with lid slightly ajar on shelf
- 3940 3. Let dry completely (24-48h)
- 3941 4. Dry off any leftover water with paper towels
- 3942 5. UV for 20mins in hood (Filter holders with the inside facing
- 3943 up/towards the UV lamp)

- 3944 6. Assemble syringes and filter holders in hood and immediately
3945 transfer to plastic zip bag for kits
- 3946 7. Move to eDNA lab (kit prep bench), clean bench top and use fresh
3947 gloves
- 3948 8. Fill screw-top tubes with glass beads (can use half-filled 0.1ml
3949 tube as scoop), 1m filtered Longmire's buffer
- 3950 9. Add 3 tubes in small plastic bag to kit
- 3951 10. Add 3 filter papers to a small plastic bag, label (!) and put into kit
3952 bag for each size
- 3953 11. Add gloves to kit bag
- 3954 12. Label finished sampling kit if necessary
- 3955 **Use:**
- 3956 1. Open kit, take out gloves and put them on
- 3957 2. Put filter papers with the tweezer in the filter holders (remember
3958 size and order of filters!)
- 3959 3. Take water (50ml) with syringe and push through filter column
- 3960 4. Repeat until 250ml put through both filter (5x)
- 3961 5. Remove 0.45µm filter and transfer to sample tube
- 3962 6. Fill 5µm filter with an additional 250ml (total 500ml) and transfer to
3963 the same sample tube
- 3964 7. Clearly label tubes (best to do lids and tube sides)!
- 3965 8. Repeat for all 3 replicates
- 3966 9. Put sample tubes in the former filter paper holding bags
- 3967 10. Label bag and close securely, move to cooler
- 3968 11. Put leftovers (tweezers, filter holders, syringe and gloves) into
3969 sample kit bag and put away
- 3970 For field negatives:
- 3971 1. Bring 500ml of double-distilled water in glass bottle from lab (filled
3972 on sampling day or shortly before) and extra kit
- 3973 2. Use as kit as normal
- 3974
- 3975 eDNA extraction using GuHCl and Longmire's buffer (after
3976 Qiagen DNeasy Blood & Tissue protocol)
- 3977 **General preparation:**
- 3978 1. **Reagents:** 8M GuHCl, Ethanol (100%), Proteinase K, AW1 & AW2
3979 buffer, AE buffer from Qiagen DNeasy Blood & Tissue kit
- 3980 2. **1% Distel** and **ethanol** for cleaning
- 3981 3. Preheat thermomixer
- 3982 4. Take out samples, document (photo + list)
- 3983 5. Make codes corresponding to sampling info

- 3984 6. Prepare and label:
- 3985 a. Bead tubes
- 3986 b. 2 tubes:
- 3987 i. First extract (2ml)
- 3988 ii. Final voucher (1.5ml)
- 3989 c. Columns (top + bottom)
- 3990 DNA extraction:
- 3991 1. Vortex whole sample rack for 10mins or use tissue lyzer
- 3992 o Programs with 8.000 – 10.000 rpm and 3 cycles of 20s
- 3993 shaking/40s rest
- 3994 o Avoid heating the probes too much!
- 3995 2. Spin down tubes in a centrifuge at below 6000rpm for 1min at a
- 3996 time to break foam
- 3997 3. Add _____ **µl proteinase K**, 10 µl proteinase K per 100 µl initial
- 3998 buffer volume (estimate)
- 3999 4. Vortex
- 4000 5. Incubate in thermomixer @ 56°C for at least 1h or overnight
- 4001 (shaking)
- 4002 6. Vortex for 15s
- 4003 7. Pipette all liquid from bead tubes into 2ml tube (amount ~1ml)
- 4004 (first extract)
- 4005 8. *Optional: Wash bead tube with the GuHCl and ethanol of next step*
- 4006 *and add to 2ml tube (first extract)*
- 4007 **Estimate volume. Better add too much GuHCl and ethanol than**
- 4008 **too little!**
- 4009 9. Add _____ **µl 8M GuHCl**, 50 µl 8M GuHCl per 100 µl initial
- 4010 volume in tube
- 4011 10. Add _____ **µl ethanol (100%)**, 100 µl ethanol (100%) per 100 µl
- 4012 initial volume in tube
- 4013 *or top of with 100% ethanol to 2ml*
- 4014 **CALCULATE STEP 8 BEFORE STEP 7 and BEFORE ADDING 8M**
- 4015 **GuHCl!!!**
- 4016 Solution might become cloudy or precipitate clear white.
- 4017 11. Vortex
- 4018 12. Spin down if necessary
- 4019 13. FIXATION: Pipet 500 µl on column, centrifuge at 8000 rpm (6000 x
- 4020 g) for 1min, discard flow-through (dry discard tube)
- 4021 14. *Repeat with remaining liquid*
- 4022 15. FIXATION: Pipet 500 µl on column, centrifuge at 8000 rpm
- 4023 (6000 x g) for 1min, discard flow-through (dry discard tube)
- 4024 **If first round of centrifuging does not pull all of the liquid**
- 4025 **through, repeat centrifuging.**
- 4026 16. WASH: Add **500 µl AW1 buffer**, centrifuge at 8000 rpm (6000 x g)
- 4027 for 1min, discard flow-through (dry discard tube)

- 4028 17. WASH: Add **500 µl AW2 buffer**, centrifuge at 14,000 rpm (20,000 x
4029 g) for 3min, discard flow-through and bottom tube
4030 18. Check: Final voucher
4031 19. ELUTION: Place column into final voucher
4032 20. Pipet **100 µl AE buffer** directly on column membrane, make sure
4033 to wet it completely
4034 21. Incubate in thermomixer for 5min @ 37°C (shaking)
4035 22. Centrifuge for 1min @ 8000 rpm (6000 x g)
4036 23. **KEEP FLOW-THROUGH!**
4037 24. *Repeat*:
4038 25. Pipet **100 µl AE buffer** directly on column membrane, make
4039 sure to wet it completely
4040 26. Incubate in thermomixer for 5min @ 37°C (shaking)
4041 27. Centrifuge for 1min @ 8000 rpm (6000 x g)
4042 28. **KEEP FLOW-THROUGH!**
4043 29. Store final voucher in freezer, check labels

4044 **Table 1:** River Erewash sampling sites

Site Code	Site Number	Site Name	Latitude	Longitude
ER_1TF	1	Toton Fields	52.90581208	-1.266035464
ER_2GO	2	Golden Brook	52.897457	-1.289596
ER_3SA	3	Sandiacre	52.9256233	-1.285903818
ER_4GA	4	Gallows	52.95947865	-1.290722109
ER_5NU	5	Nut Brook	52.95957405	-1.313933422
ER_XXWH	XX	West Hallam	52.95661842	-1.380615098
ER_6GI	6	Gilt Brook	53.00362047	-1.278592181
ER_7EA	7	Eastwood	53.00415799	-1.311155807
ER_8BE	8	Beuvale Brook	53.02166183	-1.297625048
ER_9BA	9	Bailey Brook	53.021761	-1.354877
ER_10ST	10	Stoney	53.02996834	-1.328752584
ER_11BG	11	Bagthorpe Brook	53.05887867	-1.306328395
ER_12IR	12	Ironville	53.0633666	-1.344551355
ER_13PI	13	Pinxton	53.082972	-1.328839
ER_14KI	14	Kirkby	53.08818294	-1.276745255

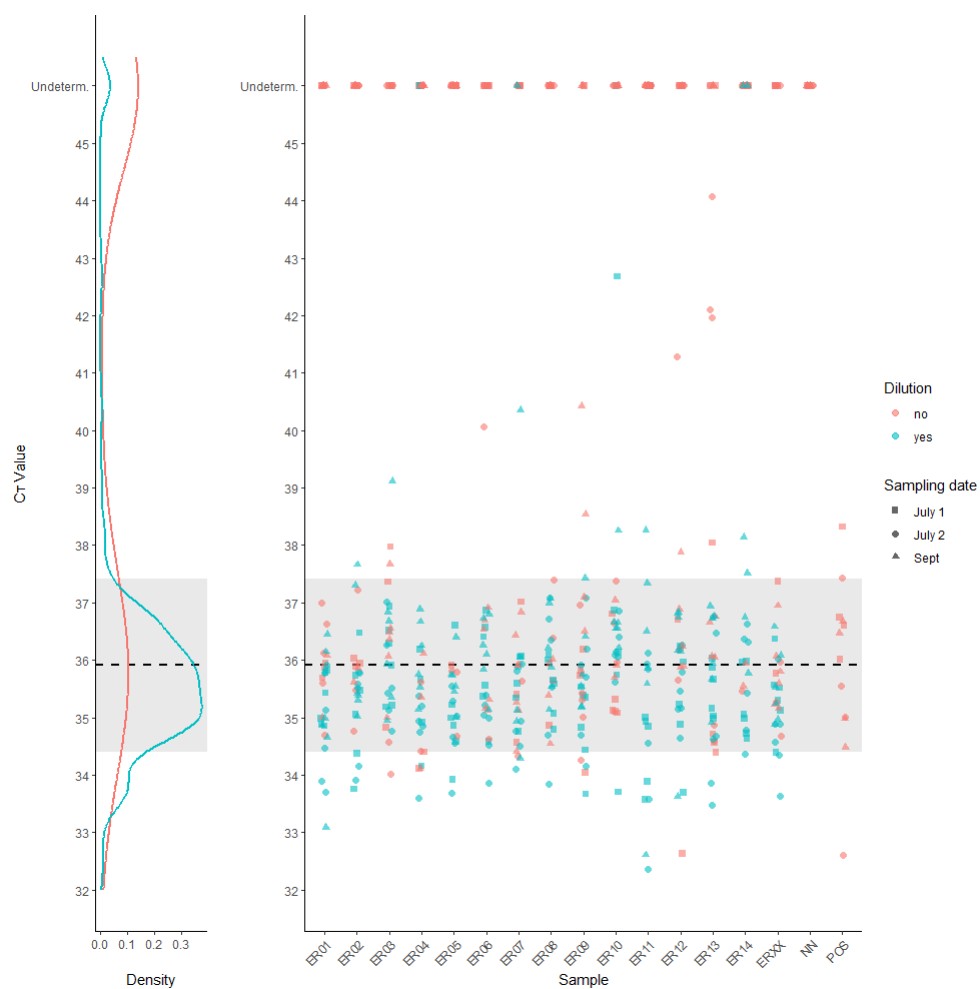


Figure 1: Inhibition Positive Control (expected Ct = 36.1) of all eDNA samples from both sampling dates with and without 1:2 dilution.

4045 Chapter 2 - Supplementary

4046 Step-by-Step: Rapid direct sequencing, de-novo assembly and
 4047 annotation of complete mitogenomes using long read
 4048 technology with application in the freshwater bryozoan
 4049 *Fredericella sultana*

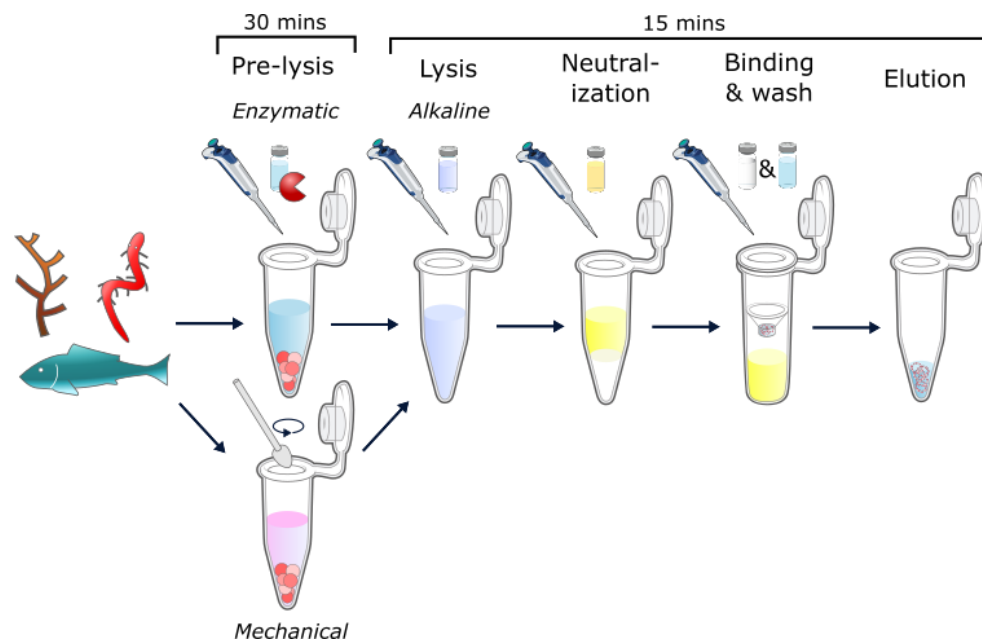


Figure 1: DNA extraction using a commercial kit like NEB Monarch® Spin Plasmid Miniprep Kit

4050 **Step1. DNA extraction**

4051 The user should begin with fresh tissue sample of approximately 2 mm²
 4052 (pin head size) or 25mg (more than 100mg lead to poor yields in our
 4053 experience). After testing extensively, we recommend using a plasmid
 4054 miniprep kit, for example, Monarch® Spin Plasmid Miniprep Kit (NEB).
 4055 Care must be taken if the suspected size of the mitogenome is larger than
 4056 the size limits of the extraction kit (commonly <25 kbp). Plasmid prep kits
 4057 rely on an alkaline and SDS-based lysis buffer to digest or disrupt
 4058 (bacterial) cells. A high-salt neutralization buffer is then added to the
 4059 lysate to denature and precipitate cell debris as well as longer
 4060 (chromosomal) DNA molecules (facilitated by high, chaotropic guanidine
 4061 hydrochloride concentration). RNA is removed through RNase A

4062 digestion. Any shorter molecules (up to 25kb length according to the
4063 manufacturer) remain largely dissolved and can bind to membrane of the
4064 spin column, whereas larger unbound fragments are washed away before
4065 bound DNA is eluted. Gentle handling and wide-bore tips are
4066 recommended throughout to avoid shearing and excessive co-extraction
4067 of chromosomal DNA.

4068 Place tissue in a 1.5ml Eppendorf tube with 180µl Longmire's buffer tube
4069 and 20µl Proteinase K solution (20mg/µl or 600 U/µl). Using a tapered
4070 plastic pestle, ground or smear the tissue against the side wall of the tube
4071 and incubate for 30min at 56°C to break down the tissue. Check that
4072 tissue is completely lysed before proceeding. If not, increase incubation
4073 time after gentle mixing.

4074 Follow manufacturers protocol starting with the addition of the lysis
4075 buffer step (e.g. Monarch Lysis Buffer (blue)). To improve yields, it is
4076 possible to increase spin times by x2 for improved pelleting. We
4077 recommend warming the elution buffer (e.g. the Monarch EY buffer) to
4078 50°C before the elution step and increasing incubation time to 10min.

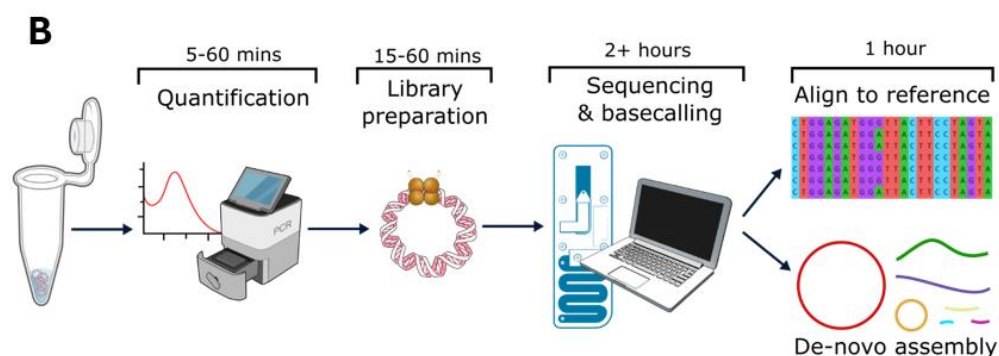
Monarch® Spin Plasmid Miniprep:

- 1) Addition of 200µl **Lysis Buffer** (B2 **blue**), invert gently 5-6 times and incubate 1min. DO NOT VORTEX.
- 2) Neutralize with 400µl **Neutralization Buffer** (B3, **yellow**), invert until uniformly yellow, incubate 2mins. DO NOT VORTEX.
- 3) Spin down 5mins @ 16.000xg. (Repeat if not fully pelleted.
Optional: Transfer supernatant to new tube and then repeat spin with it to remove any remaining debris)
- 4) Transfer **supernatant** to column and spin 1min. Discard flow-through.
- 5) Wash with 200µl **Wash Buffer 1 (BZ)**. (Optional incubate for 5mins). Spin 1min. Discard flow-through.
- 6) Wash with 400ul **Wash Buffer 2 (WZ)**. Spin 1min. Discard flow-through.
- 7) Transfer column to new 1.5ml tube. Add 30-40ul **Elution Buffer (EY)** (warmed to 50°C) and incubate for 5-10min. Spin 1min.

4079 Quantify DNA to determine yield using e.g. Qubit™ dsDNA HS

4080 Quantification Assay (Thermofisher) and check for purity on microvolume

4081 spectrophotometers (e.g. Biochrom BioDrop). Target yield useful for
4082 library preparation was set at 50ng in 9µl. In cases where this was not
4083 met, the DNA was concentrated using AMPure XP beads (Beckman
4084 Coulter) using manufacturers protocol and 1:3 DNA:bead ratio. We aimed
4085 for a higher than necessary concentration to account for loss. Increasing
4086 the concentration 5 - 10x is possible.



4087

4088 **Figure 2:** Direct Nanopore sequencing and de-novo assembly workflow

4089 Step 2. **Library prep – Rapid barcoding kit (Oxford Nanopore, V14**
 4090 **Chemistry)**

4091 We followed the recommended protocol for Rapid sequencing DNA V14 –
 4092 Plasmid sequencing (SQK-RBK114.24 or SQK-RBK114.96) with x2
 4093 increased incubation and elution times for the bead clean-up before
 4094 Rapid Adaptor addition to maximise long-fragment retention.

4095 The input volume according to the manual should be 50ng in 9µl, however
 4096 we still obtained good sequencing results down to ~25ng of input
 4097 amount. If you wish to load more of your sample increase the volume and
 4098 DNA amount accordingly (this will reduce the space for other samples).

4099 Prepare all the kits components according to the manual.

4100 Add 1µl unique Rapid Barcode to each 9 µl DNA sample in a 0.2ml PCR
 4101 tube, mix and spin down.

4102 Incubate in a thermocycler for 2 minutes at 30°C and then 2 minutes at
 4103 80°C, cool on ice and combine in 1.5ml Eppendorf tube.

4104 Add 10 µl AMPure XP beads per sample/barcode to the tube and incubate
 4105 for at least 10 minutes at room temperature on a rotator mixer. Increased
 4106 to incubation times lead to better recovery in our experience, extend
 4107 incubation up to 1 hour for low input sequencing.

4108 Spin down and pellet on magnet. Keep on magnet, remove supernatant
 4109 and wash twice with 1.5ml of fresh 80% ethanol.

- 4110 Air dry for 30s and resuspend with 15 µl Elution Buffer (EB) off the magnet.
- 4111 Incubate for at least 10 minutes at room temperature.
- 4112 Pellet beads on magnet. Quantify 1 µl of elute using a Qubit fluorometer.
- 4113 Transfer 11 µl of the elute into a new 1.5ml Eppendorf tube.
- 4114 Dilute the Rapid Adapter (RA) by mixing 1.5 µl of Rapid Adapter (RA) and
- 4115 3.5 µl of Adapter Buffer (ADB).
- 4116 Add 1 µl of this mix to the barcoded DNA and incubate for 5 minutes at
- 4117 room temperature.
- 4118 Store prepared library on ice until loading.
- 4119
- 4120 **Step 3. Direct sequencing on MinION**
- 4121 Make sure your sequencing PC and MinKNOW are up to date and ready to
- 4122 go. Disable sleep functions on the PC.
- 4123 Prepare all the kits components according to the manual.
- 4124 Prepare the flow cell priming mix by combining and mixing:
- 4125 1,170 µl Flow Cell Flush (FCF)
- 4126 5 µl Bovine Serum Albumin (BSA) at 50 mg/ml
- 4127 30 µl Flow Cell Tether (FCT)
- 4128 Insert the flow cell into the MinION device. Ensure flow cell sit securely in
- 4129 the device.
- 4130 Open the priming port on the flow cell. Draw back a small volume (20-30
- 4131 µl) by using a P1000 pipette set to 200 µl, inserting the tip firmly into the
- 4132 port and draw back slowly using the pipette setting wheel until volume
- 4133 enters the tip.
- 4134 Load of 800 µl of the priming mix into the priming port, avoid air bubbles.
- 4135 Wait 5 minutes.

4136 Prepare the library from Step 2 for loading by combining:

4137 37.5 µl Sequencing Buffer (SB)

4138 25.5 µl Library Beads (LIB) mixed immediately before use

4139 12 µl DNA library

4140 Complete flow cell priming by gently lifting the SpotON sample port cover

4141 and load 200 µl of priming mix into the priming port (NOT the SpotON

4142 sample port), avoid air bubbles.

4143 Mix the library gently and dropwise add 75 µl of library to SpotON sample

4144 port. Ensure drop flowed into the port before adding the next.

4145 Replace the SpotON sample port and close the priming port. Add the light

4146 shield to the flow cell (covering the entire top section of the flow cell).

4147 Close the MinION device and setup the sequencing run on MinKNOW.

4148 Basecalling while sequencing can be skipped/disabled which reduces

4149 processing times while also reducing the hardware requirements of the

4150 computer used. We will basecall later with dorado in Step 4. Run the

4151 sequencing as long as there are sufficient active pores available and

4152 ample reads are generated. A run time of 24h is usually sufficient, for

4153 single samples a couple of hours might be enough. After 48h the library

4154 and flow cell will be exhausted. Flow cells can be reused. Follow the Flow

4155 Cell Wash Kit protocol to prepare the flow cell for storage after the

4156 sequencing run. Optional: Recover your library.

4157 **Step 4 - Basecalling**

4158 Retrieve the reads from MinKNOW. You will require the pod5 files from the

4159 'all_pod5' folder.

4160 Transfer to a suitable storage location, on a server or locally.

4161 Conduct basecalling using dorado basecaller (

4162 <https://github.com/nanoporetech/dorado>) with sup model and modified

4163 bases. Nanopore offers different open source basecalling options: the

4164 established standard software guppy (included and used in MinKNOW)
4165 and the new improved program dorado. Dorado is recommended here as
4166 it has far greater performance than guppy Dorado does not natively run
4167 while sequencing through MinKNOW (as guppy does). However, dorado
4168 requires a CUDA environment to run and we assume user installation is
4169 available. For easier handling and to be able to call modified bases
4170 (methylation) saving readings to pod5 format (default) is necessary (in
4171 place of the old fast5 format).
4172 If run on a high-performance cluster or similar, users should adapt the
4173 header of the batch file accordingly, see example below. Ensure
4174 necessary modules are installed and available to load.

```
#!/bin/bash
#SBATCH --chdir=/gpfs01/home/"user"/
#SBATCH --partition=ampereq
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --gres=gpu:1
#SBATCH --mem=116g
#SBATCH --time=24:00:00
#SBATCH --mail-user= "user_email"
#SBATCH --mail-type=END,FAIL

# Script for basecalling Nanopore data with GPU acceleration
# through Dorado
# 20/07/2024

# Load the relevant module
#module load SRA-Toolkit/3.0.3-gompi-2023a
module load dorado-uon/binary/0.7.3
module load R-uoneasy/4.3.3-gfbb-2023b
module load minimap2

# Set file paths and names
POD5_DIR="quick_mtgenome_ER2024/no_sample/20240730_1738_M
N41698_FAW14401_d629c83d/pod5/" # Path to pod5 files from
working directory

BAM_FILE="quick_mtg_er2024_CpG_sup" # Name of your bam file only,
without .bam ending

REF_SEQ="ref_fs_mito.mmi" # Full reference name, with ending
```

```

# Run basecaller, with modified bases (model sup or hac and
5mC_5hmC or 5mCG_5hmCG) and kit-name
dorado basecaller sup,5mCG_5hmCG --device cuda:0 "$POD5_DIR" --
kit-name SQK-RBK114-24 > "${BAM_FILE}.bam"

dorado summary "${BAM_FILE}.bam"> "${BAM_FILE}_sum.tsv"

# OPTIONAL: Demultiplex into separate bams - PRE-ALIGNMENT: Code
needs adaptation to loop through all bam files
#dorado demux --output-dir "${BAM_FILE}_barcodes" --no-classify
"${BAM_FILE}.bam"

#End file

```

4175 Using the sup model is recommended if computing power and time is of
4176 no concern. Otherwise use hac. Modified bases (methylation) can be
4177 specified with the options 4mC_5mC, 5mCG_5hmCG, 5mC_5hmC, 6mA.
4178 Possible combinations depend on the run setup. 5mCG_5hmCG is not
4179 compatible with 4mC_5mC or 5mC_5hmC. Methylation is saved as flags
4180 in the bam file and are analogous to bisulfite sequencing. Converting the
4181 bam files into fasta/fastq files might remove some of the flags and read
4182 association needed for downstream processing and analysis.

4183 Optional: Alignment to mitochondrial and nuclear references with
4184 minimap2 through dorado aligner, sorted and indexed with samtools.

4185 De-novo assembly with flye, identification with BLAST and alignment of all
4186 reads to assembly

4187 The produced alignment was demultiplexed with dorado demux. (Pooled
4188 sample) coverage and consensus calculated with samtools coverage and
4189 samtools consensus.

4190 Alignments (bam files) can be visualized in IGV (2.18.0), bandage (0.8.1)
4191 for the assembly plot, Snapgene for the MITOS2 annotation, or DnaSP 6
4192 and RStudio for more complex plots (coverage, expression etc.).

4193 **Optional: Step 5 – Alignment to reference**

4194 Processed and basecalled reads can be aligned with minimap2 through
4195 dorado aligner. This is strongly recommended as it preserves headers
4196 containing read group (= barcode) and other information necessary in
4197 subsequent further analysis. A reference is required (fasta file).
4198 Alignments are sorted and indexed with samtools. Reducing aligned bam
4199 files to aligned reads only greatly reduces file size. Generating a
4200 consensus for re-aligning, annotation etc. (per barcode) is also possible.
4201 Alignments can be visualized with IGV (recent releases expand available
4202 tools). Methylation can be further analysed with bedMethyl. Other
4203 options are available.

```
#!/bin/bash
#SBATCH --chdir=/gpfs01/home/"user"/
#SBATCH --partition=ampereq
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=1
#SBATCH --gres=gpu:1
#SBATCH --mem=116g
#SBATCH --time=24:00:00
#SBATCH --mail-user="user_email"
#SBATCH --mail-type=END,FAIL

# Script for basecalling Nanopore data with GPU acceleration
# through Dorado
# 20/07/2024

# Load the relevant module
#module load SRA-Toolkit/3.0.3-gompi-2023a
module load dorado-uon/binary/0.7.3
module load R-uoneasy/4.3.3-gfbf-2023b
module load minimap2

# Set file paths and names
POD5_DIR="quick_mtgenome_ER2024/no_sample/20240730_1738_M
N41698_FAW14401_d629c83d/pod5/" # Path to pod5 files from
working directory

BAM_FILE="quick_mtg_er2024_CpG_sup" # Name of your bam file only,
without .bam ending

REF_SEQ="ref_fs_mito.mmi" # Full reference name, with ending
```

```

# Run basecaller, with modified bases (model sup or hac and
5mC_5hmC or 5mCG_5hmCG) and kit-name
dorado basecaller sup,5mCG_5hmCG --device cuda:0 "$POD5_DIR" --
kit-name SQK-RBK114-24 > "${BAM_FILE}.bam"
dorado summary "${BAM_FILE}.bam"> "${BAM_FILE}_sum.tsv"

# OPTIONAL: Demultiplex into separate bams - PRE-ALIGNMENT: Code
needs adaptation to loop through all bam files
#dorado demux --output-dir "${BAM_FILE}_barcodes" --no-classify
"${BAM_FILE}.bam"

# Reference, skip if no reference available (useful against whatever is
available to check for successful sequencing)
minimap2 -d $REF_SEQ fs_er_consensus.fa

# Mapping
# Stand-alone minimap2 removes information from reads
#minimap2 -ax map-ont ref_long_reads.mmi
quick_mtg_er2024_mod/fastq/quick_mtg_er2024_mod.fastq >
all_to_long_reads.sam

dorado aligner "$REF_SEQ" "${BAM_FILE}.bam" >
"${BAM_FILE}.aligned.bam"

#Aligned reads only
samtools view -b -F 4 "${BAM_FILE}.aligned.bam" >
"${BAM_FILE}.only.bam"

# Sorting
#samtools view -S -b all_to_long_reads.sam > all_to_long_reads.bam
#for minimap2 output only
samtools sort "${BAM_FILE}.aligned.bam" -o
"${BAM_FILE}.sorted.bam"
samtools index "${BAM_FILE}.sorted.bam"

# OPTIONAL: Demultiplex into separate bams - POST-ALIGNMENT
#dorado demux --kit-name SQK-RBK114-24 --output-dir
"${BAM_FILE}_barcodes" --no-classify "${BAM_FILE}.sorted.bam" --no-
trim --sort-bam

# Coverage
samtools coverage "${BAM_FILE}.sorted.bam" -o "${BAM_FILE}.depth"

# bedMethyl: Methylation, needs modkit installation
./dist/modkit pileup "${BAM_FILE}.sorted.bam"
"pileup_${BAM_FILE}.bed --log-filepath pileup_.log

```

```
#Consensus
samtools consensus -f fasta "${BAM_FILE}.sorted.bam" -o
"${BAM_FILE}_consensus.fa"

#End file
```

4204

4205 **Step 6 – De novo assembly**

4206 Flye (assembly)

4207 Flye is a de novo assembler which is suggested to be the most applicable
4208 for long error-prone single-molecule sequencing such as Nanopore. It is
4209 very easy to use: Raw ONT reads are assembled and polished consensus
4210 contigs (linear or circular) with length and coverage information are
4211 returned. Edge graphs of the assembly can be visualized with Bandage.
4212 Flye needs local installation on the HPC. Other assemblers like SPAdes or
4213 MitoHiFi were tested, but don't work well with Nanopore data and failed
4214 to assembly or while producing an assembly they failed to recover the
4215 complete mitogenome. Contig lengths were also generally shorter than
4216 the flye assembled contigs. Assemblers or pipelines designed for long-
4217 read Nanopore or PacBio data from bacteria like Strainly, Hybracter and
4218 Unicycler match closer to our requirements and could be promising but
4219 have not been tested as flye worked 'out of the box'. As SPAdes,
4220 Hybracter and Unicycler are capable to hybrid assembly by addition of
4221 short read data such as from the published Fs RNA-seq (and are
4222 discussed below). Further processing with programs such as Tricycler or
4223 quickmerge where not necessary but might be useful to improve tricky
4224 mitogenomic assemblies or to uncover more nuclear chromosomal data.
4225 From the metagenomic sequencing data flye was more than capable to
4226 produce a circular assembly that appears covers the complete
4227 mitogenome while including variable and repeat regions with fairly little
4228 error as outlined below. Latest Nanopore sequencing and basecalling
4229 with dorado and the sup model should enable the usage of high-quality

4230 reads (--nano-hq). The metagenome option for mixed samples (--meta
4231 and --keep-haplotypes) and can be disabled.

```
#[ADD module load: r and python]

git clone https://github.com/fenderglass/Flye
cd Flye
make

python bin/flye

# Sometimes are not uniquely named, code below ensures fly runs
# complaining about duplicated read names
module load Biopython/1.81-foss-2023a
python
from Bio import SeqIO
unique_reads = {}
with open("FS_ONT_106.only.fastq", "r") as infile,
open("FS_ONT_106.unique.fastq", "w") as outfile:
    for record in SeqIO.parse(infile, "fastq"):
        if record.id in unique_reads:
            record.id += "_dup" # Adjust as needed to ensure uniqueness
            record.description = ""
            unique_reads[record.id] = record
        SeqIO.write(record, outfile, "fastq")
exit()

bedtools bamtofastq -i FS_ONT_106.only.bam -fq
FS_ONT_106.only.fastq

# Regular reads, 20% error rate
flye --nano-raw [path] --out-dir [path] --threads [int] --iterations [int] --
meta --keep-haplotypes

# High quality reads
flye --nano-hq [path] --out-dir [path] --threads [int] --iterations [int] --
meta --keep-haplotypes

# Results can be visualised with bandage outside (.gfa file)
# assembly_info.txt contains contig and coverage information
# assembly.fasta contains the contig sequences
```

4232

4233 **Step 7 – BLAST of assembly against database**

4234 To identify origins of the different contigs from contaminated or mixed
4235 samples we mapped the complete assembly contig collection against the

4236 core BLAST database (needs installation by user). This of course requires
 4237 suitable references to be present on BLAST (or in GenBank). If there are
 4238 no direct references they might get an incorrect assignment or none.
 4239 When matches are close enough and suitable references are available
 4240 assignments to related species might be produced. Conserved motif in
 4241 BLAST/Genbank sequences - on and off target - might be (falsely)
 4242 associated. Initial setup and settings of the blastn command might lead
 4243 to (considerably) different results. Especially e-value, bitscore, pident and
 4244 alignment length play an important role. Output of the blast results in
 4245 tabular output format 6 (as below) is recommended:

4246 **BLASTn tabular output format 6**

4247 **Column headers:** qseqid sseqid pident length mismatch gapopen qstart
 4248 qend sstart send evalue bitscore

4249 **qseqid** query or source (gene) sequence id

4250 **sseqid** subject or target (reference genome) sequence id

4251 **pident** percentage of identical positions

4252 **length** alignment length (sequence overlap)

4253 **mismatch** number of mismatches

4254 **gapopen** number of gap openings

4255 **qstart** start of alignment in query

4256 **qend** end of alignment in query

4257 **sstart** start of alignment in subject

4258 **send** end of alignment in subject

4259 **evalue** expect value

4260 **bitscore** bit score

4261 Trying out different threshold setting for these variables is encouraged. E-
 4262 value and bitscore can be set as a hard threshold (for example: e-value <
 4263 e-10) as these are reference independent. Since the other variables - at
 4264 least to some extent depend on the sequence lengths, initial quality and
 4265 sequencing technique, copy number or natural (inherent) variance

4266 (genetic diversity) – sorting and filtering based on these can be difficult.
4267 Strict settings might be useful if accurate and closest identification is
4268 necessary for a conservative approach. However, this will lead to a larger
4269 number of sequences remaining unidentified. Conversely, a relaxed
4270 approach with less strict settings might impact subsequent analysis by
4271 erroneous introduction or overinflation of inaccurate or false
4272 associations. This can in part be avoided if single gene or marker
4273 databases (16S, COI) or a combination of them are used with appropriate
4274 (tested) threshold settings for reliable identification. Naturally, off-target
4275 sequence not aligning with any of the database references will remain
4276 unidentified. This might be most contigs generated from sequencing. For
4277 de novo discovery of valid sequences all-vs-all matching might be more
4278 useful but is error-prone. For secure and reliable identification of target
4279 species and associated microbiome contigs specific curated single gene
4280 database are probably advisable.

4281 If you want to search for specific sequences blasting against selected
4282 sequences is also possible. Full mitogenomes should be reliable
4283 identifiable through short gene-sized markers or full mitogenomes of the
4284 target or related species. Alternatively, all contigs can be used as
4285 reference for minimap2 or other aligners to map the original Nanopore
4286 reads to check the assembly or, reversely, a known reference (genome)
4287 can be used to align the contigs against. The coverage of the assembly
4288 contigs is a useful indicator of the proportion of target reads sequenced
4289 and assembled. However, some potentially alignments might be removed
4290 through processing and polishing by flye. The assembler only takes
4291 sequences > 1kb into account. Minimum overlap might have an impact
4292 on performance, contig number and coverage.

```
# BLAST align
```

```
# Load modules
```

```
module load R-uoneasy/4.3.3-gfbf-2023b
```

```
module load BLAST+/2.14.1-gompi-2023a
```

```
module load dorado-uoneasy/0.5.1-foss-2022a-CUDA-11.7.0
```

```

module load minimap2

# Database, single gene alternatives below, for the core_nt BLAST
database check their website
#cd blast_spec_db
#wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/16S_ribosomal_RNA.tar.gz
#tar -xzvf 18S_fungal_sequences.tar.gz
#tar -xzvf 16S_ribosomal_RNA.tar.gz
#tar -xzvf Kelly_16S.tar.gz

#for file in *.tar.gz; do
#  tar -xzvf "$file"
#  # Check disk usage and continue if there's enough space
#  df -h /gpfs01/home/stxfm6 .
#done
#cd ..

# BLAST search
blastn -outfmt 6 -query out_flye/assembly.fasta -db
blast_core_db/core_nt -evaluate 1e-10 -num_threads 96 -out
blastb_core_flye.txt

# Best hits
sort -k1,1 -k12,12nr blastb_core_flye.txt | sort -u -k1,1 --merge >
blastb_core_flye.hits.out

# Number of hits
cut -f2 blastb_core_flye.hits.out | sort | uniq -c | sort -nr >
blastb_core_flye.counts.txt

cut -f1 blastb_core_flye.txt > query_ids_flye.txt
sort query_ids_flye.txt | uniq -c > unique_reads_flye.txt
wc -l unique_reads_flye.txt

# Making BLAST database from FASTA, for check against hand selected
references
#makeblastdb -in all_tb_seq.fasta -dbtype nucl -out tb_blast_db

# BLAST search
#blastn -outfmt 6 -query quick_mtg_er2024_mod_hac.fasta -db
tb_blast_db -evaluate 1e-20 -num_threads 96 -out blastb_tb.txt

# End of file

```

4293 **Step 8 - Annotation**

4294 *Circularization:* Flye should produce a full circular assembly of
4295 mitogenome assuming it is not linear, has enough coverage and it is not
4296 hindered by complex regions and repeats. An advantage of long reads is
4297 that they span and assemble over these regions more easily without
4298 issue. Polishing steps of assemblers potentially hinder circularisation.
4299 Circularisation can also be manually checked with aligners if enough long
4300 reads were generated, but this should not be necessary. If sequencing
4301 depth is high enough there is a decent chance that multiple (near)
4302 complete mitogenome spanning reads were generated. Alternatively, long
4303 individual reads should align to both ends of the reference/assembly or go
4304 far beyond one end if the reference is incomplete. Specific programs for
4305 circularization are also available (such as Strainy, Hybracter and
4306 Unicycler), but have not been tested here as flye already produced a
4307 circular contig for the mitogenome.

4308 Aligning all reads against the mitogenome assembly and building a
4309 consensus sequence leads to better results in our experience, however,
4310 with high enough coverage the assembly contig can also be used directly.
4311 A consensus sequence will increase coverage and sequence quality
4312 because the flye assembler only includes reads above 1000bp.

```
### Consensus

BAM_FILE="initial_mito_test" # Name of your bam file only, without
                             .bam ending

REF_SEQ="contig_282.mmi" # Mitogenome contig

## ALIGNMENT
# Reference
# Contig from assembly
samtools faidx initial_mito_assembly/assembly.fasta contig_282 >
contig_282.fasta

minimap2 -d $REF_SEQ contig_282.fasta
dorado aligner "$REF_SEQ" "${BAM_FILE}.bam" >
"${BAM_FILE}.aligned2.bam"
```

```

#Aligned reads only
samtools view -b -F 4 "${BAM_FILE}.aligned2.bam" >
"${BAM_FILE}.only2.bam"

# Sorting
samtools sort "${BAM_FILE}.only2.bam" -o "${BAM_FILE}.sorted2.bam"
samtools index "${BAM_FILE}.sorted2.bam"

# Coverage
samtools coverage "${BAM_FILE}.sorted2.bam" -o
"${BAM_FILE}.depth2"

#Consensus
samtools consensus -f fasta "${BAM_FILE}.sorted.bam" -o
fs_initial_mito.fa

# End of file

```

4313 We used Galaxy (usegalaxy.eu or usegalaxy.org) to annotate the
 4314 mitogenome assembly. Upload the fasta file of your mitogenome contig
 4315 and annotate with MITOS2 (2.1.9) with the settings according to your
 4316 sequenced organisms. Changing the advanced options including the
 4317 protein coding and ncRNA gene prediction can help to produce a better
 4318 annotation depending on the organism. Increasing the overlaps and
 4319 lowering the quality of limits can lead to a gap free and more uniform gene
 4320 structure. It is recommended to check gene sequences or ORF (including
 4321 start and stop codons) manually. MITOS2 produces a tabular gff file as
 4322 well as annotation graphs.

4323 RepeatMasker (4.1.5) and equicktandem (5.0.0.1) can be used to identify
 4324 repeat regions.

4325 **Optional: Step 9 – Transcriptomic data**

4326 Transcriptomic data can be very useful to check for expression profiles
 4327 along the de-novo generated mitogenome assembly and to determine
 4328 sensible gene structure and order. Any gaps or irregularities in the profiles
 4329 might indicate suboptimal assembly.

- 4330 Check on genbank if RNAseq data is available
- 4331 Download the SRA data
- 4332 fastq files (forward and reverse were available) are aligned with bowtie2
- 4333 and default settings. Alignments are converted to BAM format, sorted,
- 4334 indexed and coverage summaries obtained with samtools.
- 4335 Alignment can be visually checked with IGV (2.18.0)

```
### Transcriptome mapping

# Load modules
module load R-uoneasy/4.3.2-gfbf-2023a
module load SRA-Toolkit/3.0.3-gompi-2023a
module load FastQC
module load STAR
module load Bowtie2

# Define the reference index basename
REFERENCE_INDEX="assembly1_flye"

# Index reference
if [ ! -f "${REFERENCE_INDEX}.1.bt2" ]; then
    bowtie2-build assembly.fasta $REFERENCE_INDEX
fi

# Define the directory containing the FASTQ files
FASTQ_DIR="transcriptome_mapping/fastq"

# Define the output directory for SAM and BAM files
OUTPUT_DIR="srr_aligned_assembly"

# Create the output directory if it doesn't exist
mkdir -p $OUTPUT_DIR

# Loop through all _1.fastq files in the FASTQ directory
for READ1 in $FASTQ_DIR/*_1.fastq; do
    # Get the base name of the file (e.g., SRR9089782)
    BASENAME=$(basename $READ1 _1.fastq)

    # Define the paired read file (_2.fastq)
    READ2="$FASTQ_DIR/${BASENAME}_2.fastq"

    # Define the output SAM file
    SAM_FILE="$OUTPUT_DIR/${BASENAME}_paired.sam"
```

```

# Define the output BAM file
BAM_FILE="$OUTPUT_DIR/${BASENAME}_paired.bam"

# Define the sorted BAM file

SORTED_BAM_FILE="$OUTPUT_DIR/${BASENAME}_paired.sorted.bam
"

# Check if the sorted BAM file exists
if [ ! -f "$SORTED_BAM_FILE" ]; then
    # Align the reads with Bowtie2
    bowtie2 -x $REFERENCE_INDEX -1 $READ1 -2 $READ2 -S $SAM_FILE

    # Convert SAM to BAM
    samtools view -S -b $SAM_FILE > $BAM_FILE # -F 4 for aligned reads
only

    # Sort the BAM file
    samtools sort $BAM_FILE -o $SORTED_BAM_FILE

    # Index the sorted BAM file
    samtools index $SORTED_BAM_FILE

        # Coverage
        samtools coverage $SORTED_BAM_FILE -o ${BASENAME}.depth

    # Optionally, remove intermediate files to save space
    rm $SAM_FILE $BAM_FILE
else
    echo "$SORTED_BAM_FILE already exists, skipping alignment for
$BASENAME."
fi
done

# End of file

```

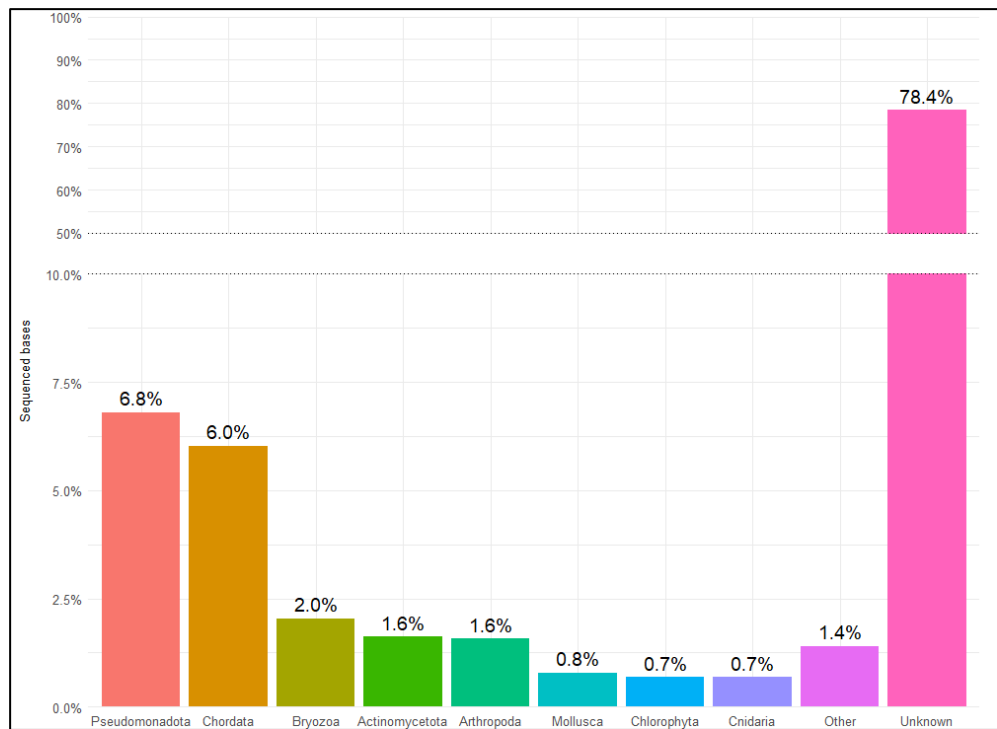


Figure 3: Example of BLAST results from a read subset identifying closest BLAST database match

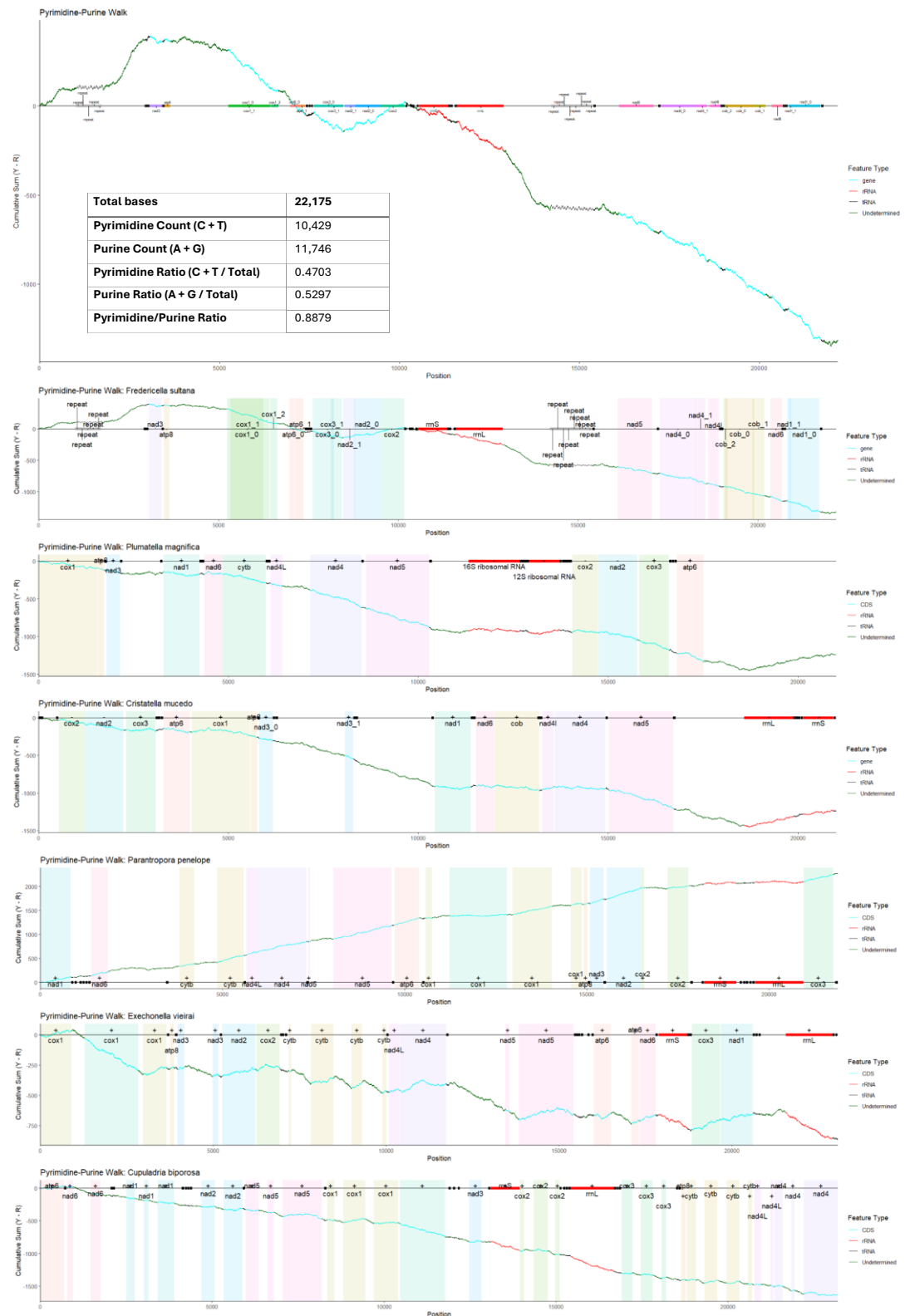


Figure 4: YR graphs of all bryozoans with mitochondrial Group II introns

4336 Chapter 4 – Supplementary

Sample ID	site_name	site_name_2	river_name	eDNA_lat	eDNA_long	Inv_Site_ID	Last_Sample_Year	Fish_Site_ID	Last_Sample_Year3
1-W-DW-1_S3	Witham DW 1	Witham DW	Witham	52.8845000	-0.625559000	55420	2023	5880	2023
1-W-DW-2_S4	Witham DW 2	Witham DW	Witham	52.8845000	-0.625559000	55420	2023	5880	2023
1-W-UP-1_S1	Witham UP 1	Witham UP	Witham	52.8819940	-0.624039000	55420	2023	5880	2023
1-W-UP-1_S88	Witham UP 2	Witham UP	Witham	52.8819940	-0.624039000	55420	2023	5880	2023
1-W-UP-2_S2	Witham UP 3	Witham UP	Witham	52.8819940	-0.624039000	55420	2023	5880	2023
13-BC-DW-1_S38	Beeston Canal DW 1	Beeston Canal DW	Beeston Canal	52.9202110	-1.188753000				
13-BC-DW-2_S39	Beeston Canal DW 2	Beeston Canal DW	Beeston Canal	52.9202110	-1.188753000				
13-BC-UP-1_S36	Beeston Canal UP 1	Beeston Canal UP	Beeston Canal	52.9157100	-1.196553000				
13-BC-UP-2_S37	Beeston Canal UP 2	Beeston Canal UP	Beeston Canal	52.9157100	-1.196553000				
14-FB-DW-1_S42	Fairham Brook DW 1	Fairham Brook DW	Fairham Brook	52.8581960	-1.134133000	53157	2015	3657	2012
14-FB-UP-1_S40	Fairham Brook UP 1	Fairham Brook UP	Fairham Brook	52.8588160	-1.098942000	47075	2011	3657	2012
14-FB-UP-2_S41	Fairham Brook UP 2	Fairham Brook UP	Fairham Brook	52.8588160	-1.098942000	47075	2011	3657	2012
15-WV-DW-1_S45	Wreake DW 1	Wreake DW	Wreake	52.7334380	-1.038573000	51670	2023	3628	2018
15-WV-DW-2_S46	Wreake DW 2	Wreake DW	Wreake	52.7334380	-1.038573000	51670	2023	3628	2018

15-WV-UP-1_S43	Wreake UP 1	Wreake UP	Wreak e	52.73 52170 00	- 1.0375 04000 0	5167 0	2023	3628	2018
15-WV-UP-2_S44	Wreake UP 2	Wreake UP	Wreak e	52.73 52170 00	- 1.0375 04000 0	5167 0	2023	3628	2018
16-TR-DW-1_S53	Trent DW 1	Trent DW	Trent	52.98 95020 00	- 0.9713 93000 0	2028 20	2022	4197 3	2023
16-TR-DW-2_S54	Trent DW 2	Trent DW	Trent	52.98 95020 00	- 0.9713 93000 0	2028 20	2022	4197 3	2023
16-TR-UP-1_S51	Trent UP 1	Trent UP	Trent	52.98 61340 00	- 0.9894 49000 0	2028 20	2022	4197 3	2023
16-TR-UP-2_S52	Trent UP 2	Trent UP	Trent	52.98 61340 00	- 0.9894 49000 0	2028 20	2022	4197 3	2023
17-DB-DW-1_S57	Dover Beck DW 1	Dover Beck DW	Dover Beck	53.02 55090 00	- 1.0308 80000 0	5183 8	2020	3746 7	2011
17-DB-DW-2_S58	Dover Beck DW 2	Dover Beck DW	Dover Beck	53.02 55090 00	- 1.0308 80000 0	5183 8	2020	3746 7	2011
17-DB-UP-1_S55	Dover Beck UP 1	Dover Beck UP	Dover Beck	53.03 17890 00	- 1.0432 14000 0	4891 7	2016	3746 7	2011
17-DB-UP-2_S56	Dover Beck UP 2	Dover Beck UP	Dover Beck	53.03 17890 00	- 1.0432 14000 0	4891 7	2016	3746 7	2011
2-W-DW-1_S6	Witham 2 DW 1	Witham 2 DW	Witha m	52.93 42520 00	- 0.6270 81000 0	1961 75	2023	5698 3	2016
2-W-DW-2_S7	Witham 2 DW 2	Witham 2 DW	Witha m	52.93 42520 00	- 0.6270 81000 0	1961 75	2023	5698 3	2016
2-W-UP-1_S5	Witham 2 UP 1	Witham 2 UP	Witha m	52.93 29860 00	- 0.6274 39000 0	1961 75	2023	5698 3	2016
3-D-DW-1_S10	Dove 1 DW 1	Dove 1 DW	Dove	52.90 03480 00	- 1.8325 51000 0	5213 9	2014	6894 4	2017
3-D-DW-2_S11	Dove 1 DW 2	Dove 1 DW	Dove	52.90 03480 00	- 1.8325 51000 0	5213 9	2014	6894 4	2017

3-D-UP-1_S8	Dove 1 UP 1	Dove 1 UP	Dove	52.90 15870 00	- 1.8329 09000 0	5213 9	2014	6894 4	2017
3-D-UP-1_S89	Dove 1 UP 2	Dove 1 UP	Dove	52.90 15870 00	- 1.8329 09000 0	5213 9	2014	6894 4	2017
3-D-UP-2_S9	Dove 1 UP 3	Dove 1 UP	Dove	52.90 15870 00	- 1.8329 09000 0	5213 9	2014	6894 4	2017
4-D-DW-1_S14	Dove 2 DW 1	Dove 2 DW	Dove	52.97 78330 00	- 1.8255 15000 0	4755 1	2017	6894 4	2017
4-D-DW-2_S15	Dove 2 DW 2	Dove 2 DW	Dove	52.97 78330 00	- 1.8255 15000 0	4755 1	2017	6894 4	2017
4-D-UP-1_S12	Dove 2 UP 1	Dove 2 UP	Dove	52.97 88300 00	- 1.8235 44000 0	4755 1	2017	6894 4	2017
4-D-UP-2_S13	Dove 2 UP 2	Dove 2 UP	Dove	52.97 88300 00	- 1.8235 44000 0	4755 1	2017	6894 4	2017
5-CU-DW-1_S18	Cherwell 1 DW 1	Cherwell 1 DW	Cherwell	52.11 19830 00	- 1.3161 47000 0	1817 06	2021	6554 4	2018
5-CU-DW-2_S19	Cherwell 1 DW 2	Cherwell 1 DW	Cherwell	52.11 19830 00	- 1.3161 47000 0	1817 06	2021	6554 4	2018
5-CU-UP-1_S16	Cherwell 1 UP 1	Cherwell 1 UP	Cherwell	52.11 19830 00	- 1.3161 47100 0	1817 06	2021	6554 4	2018
5-CU-UP-1_S90	Cherwell 1 UP 2	Cherwell 1 UP	Cherwell	52.11 19830 00	- 1.3161 47100 0	1817 06	2021	6554 4	2018
5-CU-UP-2_S17	Cherwell 1 UP 3	Cherwell 1 UP	Cherwell	52.11 19830 00	- 1.3161 47100 0	1817 06	2021	6554 4	2018
6-CT-DW-1_S22	Cherwell 2 DW 1	Cherwell 2 DW	Cherwell	52.11 05060 00	- 1.3163 24000 0	1817 06	2021	6554 4	2018
6-CT-DW-2_S23	Cherwell 2 DW 2	Cherwell 2 DW	Cherwell	52.11 05060 00	- 1.3163 24000 0	1817 06	2021	6554 4	2018
6-CT-UP-1_S20	Cherwell 2 UP 1	Cherwell 2 UP	Cherwell	52.11 11190 00	- 1.3165 81100 0	1817 06	2021	6554 4	2018

6-CT-UP-2_S21	Cherwell 2 UP 2	Cherwell 2 UP	Cherwell	52.11 11190 00	- 1.3165 81100 0	1817 06	2021	6554 4	2018
7-WH-DW-2_S27	Wharfe 1 DW 1	Wharfe 1 DW	Wharfe	53.92 93470 00	- 1.7920 56000 0	1000	2021	5992 3	2014
7-WH1-DW-1_S26	Wharfe 1 DW 2	Wharfe 1 DW	Wharfe	53.92 93470 00	- 1.7920 56000 0	1000	2021	5992 3	2014
7-WH1-UP-1_S24	Wharfe 1 UP 1	Wharfe 1 UP	Wharfe	53.92 94010 00	- 1.7968 64000 0	1000	2021	5992 3	2014
7-WH1-UP-1_S91	Wharfe 1 UP 2	Wharfe 1 UP	Wharfe	53.92 94010 00	- 1.7968 64000 0	1000	2021	5992 3	2014
8-I-DW-1_S49	Isbourne DW 1	Isbourne DW	Isbourne	52.05 00570 00	- 1.9676 95500 0	4696 3	2015	5659	2013
8-I-UP-1_S47	Isbourne UP 1	Isbourne UP	Isbourne	52.06 39210 00	- 1.9672 74100 0	4696 3	2015	5659	2013
8-I-UP-2_S48	Isbourne UP 2	Isbourne UP	Isbourne	52.06 39210 00	- 1.9672 74100 0	4696 3	2015	5659	2013
8-WH2-DW-1_S30	Wharfe 2 DW 1	Wharfe 2 DW	Wharfe	53.93 06680 00	- 1.8043 73000 0	1950 72	2019	5992 3	2014
8-WH2-DW-2_S31	Wharfe 2 DW 2	Wharfe 2 DW	Wharfe	53.93 06680 00	- 1.8043 73000 0	1950 72	2019	5992 3	2014
8-WH2-UP-1_S28	Wharfe 2 UP 1	Wharfe 2 UP	Wharfe	53.93 19080 00	- 1.8108 29000 0	1950 73	2019	3762	2014
8-WH2-UP-2_S29	Wharfe 2 UP 2	Wharfe 2 UP	Wharfe	53.93 19080 00	- 1.8108 29000 0	1950 73	2019	3762	2014
9-CA-DW-1_S34	Calder DW 1	Calder DW	Calder	53.71 06920 00	- 1.9176 92000 0	1167	2012	5186 3	2021
9-CA-DW-2_S35	Calder DW 2	Calder DW	Calder	53.71 06920 00	- 1.9176 92000 0	1167	2012	5186 3	2021
9-CA-UP-1_S32	Calder UP 1	Calder UP	Calder	53.71 13110 00	- 1.9205 64000 0	1167	2012	5186 3	2021

9-CA-UP-2_S33	Calder UP 2	Calder UP	Calder	53.71 13110 00	- 1.9205 64000 0	1167	2012	5186 3	2021
ER-01-P1_S59		ER_01TF	Erewham	52.90 58120 85	- 1.2660 35464 4	1951 12		3651	2019
ER-01-P2_S74		ER_01TF	Erewham	52.90 58120 85	- 1.2660 35464 4	1951 12		3651	2019
ER-02-P1_S60		ER_02GO	Erewham	52.89 74570 00	- 1.2895 96000 0				
ER-02-P2_S75		ER_02GO	Erewham	52.89 74570 00	- 1.2895 96000 0				
ER-03-P1_S61		ER_03SA	Erewham	52.92 56232 95	- 1.2859 03818 3	4845 9		4197 0	2019
ER-03-P2_S76		ER_03SA	Erewham	52.92 56232 95	- 1.2859 03818 3	4845 9		4197 0	2019
ER-04-P1_S62		ER_04GA	Erewham	52.95 94786 53	- 1.2907 22109 5	4777 3		4197 0	2019
ER-04-P2_S77		ER_04GA	Erewham	52.95 94786 53	- 1.2907 22109 5	4777 3		4197 0	2019
ER-05-P1_S63		ER_05NU	Erewham	52.95 95740 53	- 1.3139 33421 7	4999 4		5610 4	2016
ER-05-P2_S78		ER_05NU	Erewham	52.95 95740 53	- 1.3139 33421 7	4999 4		5610 4	2016
ER-06-P1_S64		ER_06GI	Erewham	53.00 36204 67	- 1.2785 92181 4	1549 65			
ER-06-P2_S79		ER_06GI	Erewham	53.00 36204 67	- 1.2785 92181 4	1549 65			
ER-07-P1_S65		ER_07EA	Erewham	53.00 41579 92	- 1.3111 55806 9	4892 6		3505 0	2022
ER-07-P2_S80		ER_07EA	Erewham	53.00 41579 92	- 1.3111 55806 9	4892 6		3505 0	2022

ER-08-P1_S66		ER_08B E	Erewa sh	53.02 16618 35	- 1.2976 25048 0	5193 7		6470 4	2015
ER-08-P2_S81		ER_08B E	Erewa sh	53.02 16618 35	- 1.2976 25048 0	5193 7		6470 4	2015
		ER_09B A	Erewa sh	53.02 17610 00	- 1.3548 77000 0	5367 2			
		ER_09B A	Erewa sh	53.02 17610 00	- 1.3548 77000 0	5367 2			
ER-10-P1_S68		ER_10S T	Erewa sh	53.02 99683 37	- 1.3287 52583 8	4839 1		3653	2022
ER-10-P2_S82		ER_10S T	Erewa sh	53.02 99683 37	- 1.3287 52583 8	4839 1		3653	2022
ER-11-P1_S69		ER_11B G	Erewa sh	53.05 88786 73	- 1.3063 28394 6	1569 95			
ER-11-P2_S83		ER_11B G	Erewa sh	53.05 88786 73	- 1.3063 28394 6	1569 95			
ER-12-P1_S70		ER_12IR	Erewa sh	53.06 33665 96	- 1.3445 51354 7	4968 5		3654	2022
ER-12-P2_S84		ER_12IR	Erewa sh	53.06 33665 96	- 1.3445 51354 7	4968 5		3654	2022
ER-13-P1_S71		ER_13PI	Erewa sh	53.08 29720 00	- 1.3288 39000 0	4978 1		3654	2022
ER-13-P2_S85		ER_13PI	Erewa sh	53.08 29720 00	- 1.3288 39000 0	4978 1		3654	2022
ER-14-P1_S72		ER_14KI	Erewa sh	53.08 81829 42	- 1.2767 45254 6	1552 23			
ER-14-P2_S86		ER_14KI	Erewa sh	53.08 81829 42	- 1.2767 45254 6	1552 23			
ER-XX-P1_S73		ER_XXW H	Erewa sh	52.95 66184 2	- 1.3806 15098			7012 3	2017
ER-XX-P2_S87		ER_XXW H	Erewa sh	52.95 66184 2	- 1.3806 15098			7012 3	2017

