Parasitology

cambridge.org/par

Lurking in the water: testing eDNA metabarcoding as a tool for ecosystem-wide parasite detection

Leighton J. Thomas, Marin Milotic, Felix Vaux and Robert Poulin

Department of Zoology, University of Otago, P.O. Box 56, Dunedin, 9054, New Zealand

Research article

Cite this article: Thomas LJ, Milotic M, Vaux F, Poulin R (2022). Lurking in the water: testing eDNA metabarcoding as a tool for ecosystem-wide parasite detection. *Parasitology* **149**, 261–269. https://doi.org/10.1017/S0031182021001840

Received: 1 September 2021
Revised: 14 October 2021
Accepted: 17 October 2021
First published online: 28 October 2021

Key words:
Biodiversity monitoring; eDNA; freshwater; metabarcoding; Nematoda; platyhelminthes

Author for correspondence:
Leighton J. Thomas,
E-mail: leighton.thomas@otago.ac.nz

Abstract

In the light of global biodiversity change and emerging disease, there is an urgent need to establish efficient monitoring programmes of parasites in aquatic ecosystems. However, parasite identification is time-consuming, requires a high degree of taxonomic expertize and in general requires lethal sampling. The use of environmental DNA methodology to identify parasites has the potential to circumvent these limitations. This study evaluates the use of eDNA metabarcoding to detect the presence of all species of nematode and platyhelminth parasites in two New Zealand lakes. We developed two novel metabarcoding primer pairs targeting a region of cytochrome oxidase I gene (COI) specific to platyhelminths and nematodes. We successfully detected parasite DNA in both lakes. Platyhelminth DNA yield was in general greater than nematode DNA yield. This most likely results from the larger biomass of the former quantified using traditional methods, or the presence of free-swimming life stages in the life cycle of many platyhelminths. By using eDNA, we did not detect all expected parasite families revealed through traditional methods, likely due to a lack of sequencing data available from public databases such as GenBank. As such, genetic resources need to include full reference sequences if parasitology is to truly harness eDNA to characterize and monitor parasite biodiversity in natural systems.

Introduction

The use of environmental DNA (or eDNA) is rapidly becoming a routine tool for biodiversity monitoring and promises to be a valuable tool for parasitology (Harper *et al.*, 2019). The identification of parasites to species level *via* traditional methods is time-consuming and requires a high degree of expertize. Unfortunately, the number of experts with the necessary knowledge and skills may be in decline (Poulin and Presswell, 2016). This taxonomic impediment could be partly circumvented by using DNA barcoding, which does not require taxon-specific knowledge. Furthermore, traditional methods of detecting, identifying and monitoring parasites generally require lethal sampling and dissection of host organisms. Not only does such sampling raise ethical and conservation concerns, but it is labour intensive, and it can easily fail to detect parasites that occur only in a small fraction of hosts – leading to false-negative results and a common failure to detect cryptic species (Beveridge and Gasser, 2014).

Most studies aiming to detect parasite eDNA to date have focused on single species detection methodologies, such as targeted qPCR using species-specific primers and probes, and such techniques have been shown to be more sensitive than metabarcoding approaches (Harper *et al.*, 2018). Species-specific eDNA methods have been shown to be more cost and time effective compared to traditional methods; for example using species-specific PCR primers to detect the pathogenic trematode *Ribeiroia ondatrae* (Huver *et al.*, 2015), qPCR designed for the detection of the pathogenic protozoan *Chilodonella hexasticha* in fish farms (Bastos Gomes *et al.*, 2017) and the nematode *Pseudocapillaria tomentosa* in aquaria water (Norris *et al.*, 2020). On the other hand, broad-scale metabarcoding studies have been conducted for biodiversity surveys focusing on identifying higher taxonomic groups which reduces time spent in the field compared to traditional methods. This approach is useful when DNA is likely to be at a high concentration, for instance using bat blood (Dario *et al.*, 2017) and primate faeces (Gogarten *et al.*, 2020). In addition, a recent broad-scale metabarcoding study investigated platyhelminths in marine and freshwater environments; by placing data into a molecular phylogeny, novel clades were identified which could have been missed when using traditional methodologies (Mitsi *et al.*, 2019).

Before the use of eDNA metabarcoding becomes a routine tool in parasitology, several questions must be addressed. Here, we will compare the species recovered using eDNA methodology against an independent benchmark: a recently completed, comprehensive and detailed data set compiled using traditional survey methodology providing precise data of the parasite diversity found in two sites in the Otago region (South Island, New Zealand): Tomahawk Lagoon and Lake Waihola. The objective of this study is to assess the use of eDNA for monitoring parasites at the ecosystem level. The reliability of eDNA metabarcoding as a proxy for traditional methods of parasite sampling is yet to be established a fact that prevents its acceptance as the method of choice.

CAMBRIDGE UNIVERSITY PRESS

Materials and methods

Field sampling

Sampling was conducted in two lakes, Tomahawk Lagoon (45.901°S, 170.545°E) and Lake Waihola (46.018°S, 170.082°E) where a robust estimate of abundance and biomass of all metazoan parasites has previously been obtained based on extensive surveys and dissections of large samples from all fish and invertebrate species (greater than 1 mm in length) (Lagrue and Poulin, 2016). Sampling was conducted towards the end of the austral summer (February 2020) when water temperatures were high and cercarial abundance (and thus eDNA) was expected to be at its maximum, therefore increasing the chance of parasite detection. The two lakes share many species of helminth parasites, although prevalence and intensity of infection are generally greater in Tomahawk than Waihola, resulting in higher parasite biomass (Table 1). Sampling for eDNA (see below) was conducted at two collection sites, corresponding exactly with two of the three sampling sites used by Lagrue and Poulin (2016)

Water samples were collected in 1 L bottles that were previously soaked in 10% bleach to remove any potential contaminant DNA and then rinsed with molecular grade water to remove traces of bleach. At each of the two collection sites per lake, surface water was taken by partially submerging the bottles below the water surface; bottom samples were also taken by opening a bottle that was fully submerged and touching the bottom (about 0.5 m deep). This procedure was repeated, so that 3 L (each in a separate bottle) of surface water and 3 L of bottom water were obtained from each sampling point (3 surface samples + 3 bottom samples, ×2 sites = 12 samples per lake). Disposable latex gloves were worn during sampling and changed between sites. The water samples were put on ice prior to being returned to the laboratory where they were immediately filtered.

The water samples were filtered through a 1.0 μm nitrocellulose filter using a manifold pump [following the recommendations of Jeunen *et al.* (2019)]. Filter funnels and forceps used to handle the filters were washed with 10% bleach and rinsed with distilled water between samples. When a filter became blocked with lake material, it was replaced and all filters used for the same sample were then pooled for DNA extraction. Negative filtration controls for each field site were added by filtering 1 L distilled water taken out into the field on the sampling day and thereafter treated the same way as genuine sample bottles. All filtering work was performed in a PCR-free clean room.

Primer design and in silica evaluation

Separate primers were developed for the parasitic phyla found in the study sites: Nematoda and Platyhelminthes. Primers were not designed for acanthocephalans because only a single species has been previously detected, in only one of the two lakes, and at low prevalence (Lagrue and Poulin, 2016). Mitochondrial COI sequences for these phyla were downloaded from BOLD and GenBank and clustered into operational taxonomic units (OTUs) using the R package *PrimerMineR* v0.18 (Elbrecht and Leese, 2017); subsequently, specific taxa were selected for each phylum based on what has previously been found in the sampling locations (Table 1). Sequences were aligned in Geneious Prime 2020 (http://www.geneious.com) using MAFFT v7.017 (Katoh and Standley, 2013). The consensus sequence generated in Geneious was used to design degenerate primers. Conserved regions were identified by eye in Geneious, primers were then designed for these conserved regions. *PrimerMineR* was then used to evaluate each primer against alignments for each phylum using the default settings.

DNA extraction and library preparation

The DNA was extracted from the filters using a Qiagen DNeasy Blood & Tissue Kit following the manufacturer's instructions. The PCR reactions were initially inhibited, so DNA extractions were diluted (1:10) using molecular grade water, which allowed for successful PCR amplification. We prepared the amplicon libraries *via* a two-step PCR strategy. The first-step PCR was conducted using MyTaq™ HS and our phylum-specific primers, with attached Illumina adapters (Table 2). The reaction mix included 5 μL of MyTaq Reaction Buffer, 2 μL of 10 μM forward and reverse primers, 2 μL of template DNA, 0.25 μL of MyTaq HS DNA polymerase and 16 μL of molecular grade water to make a final reaction volume of 25 μL. Negative control for each field site was included using the same reaction mix but with 2 μL of molecular grade water instead of DNA. The thermocycling conditions were 95°C for 5 min, 38 cycles of 95°C for 30 s, 51°C for 30 s and 72°C for 2 min and then a final extension at 72°C for 7 min for one cycle. PCR products were visualized on a 1.5% agarose gel in a 1 × TAE buffer and the remaining PCR product was purified separately with magnetic beads as described in Rohland and Reich (2012). 2 μL of purified DNA was used in the second-step PCR (barcoding step), where we used a Nextera XT index kit to attach dual indices to each PCR product of the first-step PCR. Here, the reaction mix was the same as the first-step PCR, except the primers were Nextera XT Index Kit primers instead of phylum-specific primers. The thermocycling conditions for the second stage PCR were 95°C for 3 min, eight cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, and then a final extension at 72°C for 5 min. The PCR product was purified using magnetic beads (Rohland and Reich, 2012) and the DNA concentration of the purified products was measured using a Qubit HS assay. The PCR products were normalized for DNA concentration and pooled prior to sequencing.

Library quality control and MiSeq run

The pooled library concentration was quantified by Qubit HS assay; the library size was checked on a PerkinElmer LabChip® GX Touch HT instrument using the DNA High Sensitivity LabChip® Assay. qPCR was used to check if the library was suitable for MiSeq run and to help determine the final loading volume for the library. 10 μL of the 2 nM pooled library was denatured to single-stranded DNA with 10 μL 0.2 N NaOH (pH > 12.5) by mixing and incubating at room temperature for 5 min. Illumina PhiX Control v3 was diluted to 2 nM with 10 mM Tris pH 8.5 with 0.1% Tween 20, then denatured with 10 μL 0.2 N NaOH (pH >12.5) by mixing and incubating at room temperature for 5 min. The library was diluted to desired molarity and PhiX was diluted to 12.5 pM with ice-cold hybridization buffer (HT-1). Finally, 800 μL of the pooled library and 200 μL of PhiX were combined to give a calculated spike of 20% PhiX. The combined library was loaded into a thawed Illumina MiSeq V2 Nano kit cartridge for sequencing on the Illumina MiSeq platform using 250 bp paired-end reads.

All data generated by Illumina sequencing was QC checked using the following protocols: FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), FastQscreen (Wingett and Andrews, 2018), SolexaQA + + (Cox *et al.*, 2010). Reads were mapped back to the PhiX control library to identify the known SNPs, using BWA (Li and Durbin, 2010), samtools (Li *et al.*, 2009) and Varscan (Koboldt *et al.*, 2009).

Metabarcoding analysis

Metabarcoding reads recovered by paired-end sequencing were first aligned using the Illumina MiSeq analysis software under

Table 1. Trematode and cestode species, along with their total average biomass (mg m^{-2}) across all hosts in the life cycle, found in Lake Waihola and Tomahawk Lagoon using traditional approaches, i.e. lethal dissection of hosts captured in each system and morphological identification (data from Lagrue and Poulin, 2016)

Phylum	Family	Species	Lake Waihola	Tomahawk Lagoon
Platyhelminthes	Strigeidae	*Apatemon* sp.	0.49	183.56
	Lepocreadiidae	*Stegodexamene anguillae*	0.65	26.62
	Cryptogonimidae	*Telogaster opisthorchis*	0.24	2.87
	Opecoelidae	*Coitocaecum parvum*	4.69	1.70
	Zoogonidae	*Deretrema* sp.	<0.01	–
	Microphallidae	*Microphallus livelyi*	5.40	20.66
	Microphallidae	*Maritrema poulini*	10.63	1.13
	Microphallidae	*Microphallus* sp.	–	0.05
	Notocotylidae	*Notocotylus* sp.	0.01	–
		Plagiorchioid sp.	0.08	1.91
		Aporocotylid sp. I	0.01	–
		Pronocephaloid sp. I	0.22	7.65
		Pronocephaloid sp. IV	0.05	0.87
		Strigeid-like sp.	–	0.02
		Acaudate xiphidocercaria	0.11	–
		Gymnocephalous cercaria I	–	0.02
		Gymnocephalous cercaria II	–	0.14
		Virgulate cercaria	–	0.01
		Trematode sp. A	0.02	–
		Trematode sp. B	0.01	–
		Trematode sp. X	0.84	3.39
		Cestode sp.	<0.01	–
Nematoda	Dioctophymidae	*Eustrongylides* sp.	0.06	4.40
	Hedruridae	*Hedruris spinigera*	0.85	–
	Dracunculidae	*Anguillicoloides* sp.	0.09	–
	Anisakidae	Nematode sp. C	0.04	<0.01
		Nematode sp. X	0.34	–

the default settings. To eliminate low-quality sequences, only reads matching 100% to Illumina adaptors, index barcodes and template specific oligonucleotides identified using Geneious were retained for downstream analysis. Forward and reverse reads were joined to obtain longer and higher-quality reads using PEAR v0.9.6 (Zhang *et al.*, 2014). Denoising, chimaera checking and clustering through dereplication was achieved using DADA2 (Zhang *et al.*, 2014). This step produced a list of identified Amplicon Sequence Variants (ASVs). To assign taxonomic information to the ASVs, each sequence was aligned against the NCBI GenBank nucleotide database using blastn (Altschul *et al.*, 1990).

Results

PCR amplification results

Two metabarcoding primer pairs were developed, one targeting platyhelminths (250 bp fragment) and the other nematodes (300 bp fragment). Both primer pairs targeted a region of the cytochrome c oxidase subunit I (COI). Both primer pairs were shown to amplify a clear band of the expected size range. All of the environmental samples produced a band of expected length in the expected region when amplified with nematode primers, whereas only three samples from Tomahawk Lagoon amplified a band using the platyhelminth primers. Only the environmental samples that produced a band for the first round of PCR were retained for further processing. Bands were not produced for any of the negative control PCRs but these were samples retained for further processing.

Sequencing and alignment results

Total number of raw reads was 328 990 (Table 3) after quality filtering, 299 836 reads were retained for further analyses. Platyhelminth primers generated a total of 73 547 reads [46 984 reads on-target and 26 563 off-target organisms (here off-target refers to non-Platyhelminth organisms or contamination)], while the nematode primers generated a total of 62 675 reads [1489 on-target and 61 186 off-target organisms (here off-target refers to non-Nematode organisms or contamination)].

Further analysis of the reads produced by the Platyhelminth primers revealed 24 unique ASVs in Tomahawk lagoon, nine of which aligned to non-target organisms (Supplementary Material), and 15 of which aligned to platyhelminth sequences on GenBank. Two of these 15 ASVs aligned with two families within class Cestoda, and the remainder with five families within

Table 2. Metabarcoding primers designed in this study.

Primer name	Target taxon	Approximate expected product size (bp)	Locus specific sequence (Illumina adapter in bold)
Platy_369.1_F	Platyhelminthes	250	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**ATGATHTTYTTYTTYYTDATGCC
Platy_179.1_R	Platyhelminthes	250	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GGRTAAAANGTYCAHCCHAC
Nematode_972_F	Nematoda	300	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GCTGCWACWATRRTTATTGC
Nematode_1221_R	Nematoda	300	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AAATWCCAWAHACAGCHCC

Table 3. Total number of reads after de-multiplexing (input reads) for each sample produced by Illumina sequencing, including the number of reads retained after each filtering step: denoising with DADA2, followed by removal of off-target sequences or contamination

Sample ID	Input reads	DADA2 filtering	Unfiltered platyhelminth reads	Filtered platyhelminth reads	Unfiltered nematode reads	Filtered nematode reads
Negative-Filer-1-Tomahawk-S33	66 126	59 524	16 076	0	0	0
Negative-Filter-1-Tomahawk-S22	337	177	0	0	0	0
Negative-Filter-2-Waihola_S23	0	0	0	0	0	0
Negative-Filter-2-Waihola-S34	8257	7524	0	0	0	0
PCR-Negative-S24	13	1	0	0	0	0
PCR-Negative-S35	35 983	31 284	0	0	0	0
platy-Tomahawk-1-1-BOTTOM-S25	35 283	28 199	26 914	16 843	NA	NA
platy-Tomahawk-1-2-BOTTOM-S26	17 375	16 534	15 612	15 433	NA	NA
platy-Tomahawk-1-3-SURFACE-S27	17 697	15 832	14 945	14 708	NA	NA
Tomahawk-1-2-BOTTOM-S14	10 199	9117	NA	NA	9113	0
Tomahawk-1-2-SURFACE-S13	3081	2648	NA	NA	2614	0
Tomahawk-1-3-BOTTOM-S16	4759	4429	NA	NA	4402	297
Tomahawk-1-3-SURFACE-S15	1088	743	NA	NA	742	0
Tomahawk-2-1-SURFACE-S17	3191	2876	NA	NA	2876	336
Tomahawk-2-2-BOTTOM-S19	5158	4081	NA	NA	4076	0
Tomahawk-2-2-SURFACE-S18	2597	1914	NA	NA	1886	468
Tomahawk-2-3-BOTTOM-S21	4063	3886	NA	NA	3886	0
Tomahawk-2-3-SURFACE-S20	591	450	NA	NA	450	107
Waihola-1-1-BOTTOM-S2	2837	2512	NA	NA	2510	0
Waihola-1-1-SURFACE-S1	5551	4867	NA	NA	4747	0
Waihola-1-2-BOTTOM-S4	3172	2829	NA	NA	2826	0
Waihola-1-2-SURFACE-S3	4160	3666	NA	NA	3662	281
Waihola-1-3-BOTTOM-S6	3939	3468	NA	NA	3468	0
Waihola-1-3-SURFACE-S5	4033	3708	NA	NA	3707	0
Waihola-2-1-BOTTOM-S8	2277	1742	NA	NA	1742	0
Waihola-2-1-SURFACE-S7	6021	5117	NA	NA	5090	0
Waihola-2-2-BOTTOM-S10	1553	1081	NA	NA	1082	0
Waihola-2-2-SURFACE-S9	2297	1813	NA	NA	1803	0
Waihola-2-3-BOTTOM-S12	4	2	NA	NA	2	0
Waihola-2-3-SURFACE-S11	2821	1992	NA	NA	1991	0
Total	254 463	222 016	73 547	46 984	62 675	1489
Mean	8482	7400	8172	5220	2321	55
Standard Deviation	14 130	12 489	10 300	7849	2161	128

class Trematoda (Table 4). The three environmental samples from Tomahawk Lagoon yielded ASVs that were assigned to Platyhelminth species (4–6 ASVs per sample; Table 4).

A total of 306 ASVs were amplified from the Tomahawk Lagoon and Lake Waihola samples, 299 of which aligned to non-target organisms (Supplementary Material), but seven aligned to a

Table 4. Results (from 57 471 reads) for the platyhelminth primers tested on three samples from Tomahawk Lagoon. The listed ASVs successfully aligned to platyhelminth sequences on NCBI GenBank using blastn

	Number of ASV reads			Top blastn result				
	Tomahawk samples							
ASV Table ID	1–1 Bottom	1–2 Bottom	1–3 Surface	GenBank accession	*e*-value	Class	Family	Species
P-ASV1	8979	0	0	MT131822.1	4.00×10^{-42}	Cestoda	Diphyllobothriidae	*Spirometra erinaceieuropaei*
P-ASV2	0	0	6469	MK840618.1	3.00×10^{-31}		Hymenolepididae	*Hymenolepis diminuta*
P-ASV3	0	5701	0	KJ477485.1	2.00×10^{-28}	Trematoda	Clinostomidae	*Clinostomum* sp.
P-ASV4	5198	0	0		5.00×10^{-29}			
P-ASV5	0	0	684	KU156806.1	2.00×10^{-28}			
P-ASV6	0	811	0	KP721418.1	3.00×10^{-31}			*Euclinostomum heterostomum*
P-ASV7	0	4401	0	KM538076.1	1.00×10^{-36}		Heterophyidae	*Acetodextra amiuri*
P-ASV8	0	3942	0	MF438097.1	2.00×10^{-39}			*Apophallus donicus*
P-ASV9	0	0	3932		5.00×10^{-35}			
P-ASV10	0	0	3090	KM880011.1	8.00×10^{-45}		Himasthlidae	*Acanthoparyphium spinulosum*
P-ASV11	1606	0	0		1.00×10^{-48}			
P-ASV12	1060	0	0	KM880010.1	7.00×10^{-46}			
P-ASV13*	0	578	0	LC438938.1	7.00×10^{-46}		Philophthalmidae	*Parorchis* sp.
P-ASV14*	0	0	375		3.00×10^{-50}			
P-ASV15	0	0	158	MT185366.1	2.00×10^{-33}		Proterodiplostomidae	*Proterodiplostomum* sp.

The top blastn result is listed for each sequence, although these sequences are not necessarily exact sequence matches indicating successful genus- or species-level identification. ASVs sequenced from negative control samples were excluded, as were ASVs for non-target organisms such as fishes and mammals. No ASVs were sequenced from nematodes using the platyhelminth primers. ASV IDs (matching to Table IDs) and nucleotide sequences for all ASVs are provided in the Supplementary Material. P-ASV13 and P-ASV14 are marked with an asterisk (*), as an ASV amplified from a negative control sample aligned to the same sequence on GenBank. Rows are shaded to aid reading.

Table 5. Results (from 62 675 reads) for the nematode primers tested on nine samples from Tomahawk Lagoon and 12 samples from Lake Waihola

ASV Table ID	N-ASV1	N-ASV2	N-ASV3	N-ASV4	N-ASV5	N-ASV6	N-ASV7*
Tomahawk Lagoon							
1-2-BTM-S14	0	0	0	0	0	0	0
1-2-SFC-S13	0	0	0	0	0	0	0
1-3-BTM-S16	0	0	297	0	0	0	0
1-3-SFC-S15	0	0	0	0	0	0	0
2-1-SFC-S17	336	0	0	0	0	0	0
2-2-BTM-S19	0	0	0	0	0	0	0
2-2-SFC-S18	0	468	0	0	0	0	0
2-3-BTM-S21	0	0	0	0	0	0	0
2-3-SFC-S20	0	0	0	107	0	0	0
Lake Waihola							
1-1-BTM-S2	0	0	0	0	0	0	0
1-1-SFC-S1	0	0	0	0	172	1651	320
1-2-BTM-S4	0	0	0	0	0	0	0
1-2-SFC-S3	281	0	0	0	0	0	0
1-3-BTM-S6	0	0	0	0	0	0	0
1-3-SFC-S5	0	0	0	0	0	0	0
2-1-BTM-S8	0	0	0	0	0	0	0
2-1-SFC-S7	0	0	0	0	0	0	0
2-2-BTM-S10	0	0	0	0	0	0	0
2-2-SFC-S9	0	0	0	0	0	0	0
2-3-BTM-S12	0	0	0	0	0	0	0
2-3-SFC-S11	0	0	0	0	0	0	0
Top blastn result							
GenBank Accession	NC_031648.1				LR215992.1	KM880011.1	LC438938.1
e-value	2×10^{-117}	5.00×10^{-112}	9.00×10^{-116}	2.00×10^{-117}	3.00×10^{-32}	$7.00 \times 10^{-4}6$	2.00×10^{-47}
Phylum	Nematoda				Platyhelminthes		
Class	Chromadorea				Cestoda	Trematoda	
Family	Anisakidae				Hymenolepididae	Himasthlidae	Philophthalmidae
Species	*Contracaecum ogmorhini*				*Hymenolepis microstoma*	*Acanthoparyphium spinulosum*	*Parorchis* sp.

The listed ASVs successfully aligned to nematode or platyhelminth sequences on NCBI GenBank using blastn. The top blastn result is listed for each sequence, although these sequences are not necessarily exact sequence matches indicating successful genus- or species-level identification. ASVs sequenced from negative control samples were excluded, as were ASVs for non-target organisms such as fishes and mammals. ASV IDs (matching to Table IDs) and nucleotide sequences for all ASVs are provided in the Supplementary Material. N-ASV7 is marked with an asterisk (*), as an ASV amplified from a negative control sample aligned to the same sequence on GenBank. Columns are shaded to aid reading.

Table 6. Numbers of genetic sequences available on GenBank and BOLD for the mitochondrial COI gene region for the parasite taxa that were identified using traditional methodology

Phylum	Order	Family	Clusters per order	BOLD	NCBI	Mitogenomes
Platyhelminthes	Strigeidida	Strigeidae	104	595	770	3
	Plagiorchiida	Lepocreadiidae	57	69	116	0
		Cryptogonimidae		2	53	0
		Opecoelidae		24	432	0
		Zoogonidae		45	45	0
		Microphallidae		34	179	0
	Echinostomida	Notocotylidae	0	0	0	0
Nematoda	Ascaridida	Rhabditida	106	374	361	50
		Dioctophymatida		89	130	0
	Spirurida	Hedruridae	0	0	0	0
	Camallanida	Dracunculidae	19	114	100	3

The clusters per taxonomic order are the number of remaining genetic sequences once the package *PrimerMineR* has dereplicated and clustered sequences.

mixture of nematode and platyhelminth sequences on GenBank (Table 5). Four of the seven ASVs aligned to the same COI sequence for the nematode species *Contracaecum ogmorhini,* and the remaining three sequences aligned to three records representing three species of platyhelminths, belonging to one family within class Cestoda and two families within class Trematoda (Table 5). Overall, 4/9 and 2/12 environmental samples from Tomahawk Lagoon and Lake Waihola respectively yielded parasite ASVs (nematode or platyhelminth; Table 5).

Twenty ASVs were amplified from the negative control (molecular grade water and negative PCR control) samples using both sets of metabarcoding primers simultaneously (Supplementary Material). One of these ASVs (amplified from a Tomahawk Lagoon negative control) was also amplified from a Tomahawk Lagoon environmental sample using the platyhelminth primers. This ASV aligned with a sequence for the platyhelminth *Parorchis* sp. (LC438938.1; *e*-value: 7×10^{-46}). Three ASVs amplified from the environmental samples using both sets of primers (P-ASV13, P-ASV14, and N-ASV7) also aligned to the same sequence on GenBank [marked with an asterisk (*) in Tables 4 and 5]. No other ASVs amplified from the negative control samples appeared in the environmental sample results. Between the two primer datasets, two ASVs amplified from Lake Waihola and Tomahawk Lagoon aligned to sequences for separate species of *Hymenolepis* platyhelminths.

Discussion

With climate change and environmental degradation likely to affect the distribution, prevalence, or severity of diseases in freshwater ecosystems (Altizer *et al.*, 2013), there is an urgent need to establish efficient monitoring programmes to track changes in the local diversity, prevalence and abundance of aquatic parasites. However, traditional methods to survey parasites, i.e. through dissections of hosts, should be replaced by more efficient and ethical sampling methods. Here, we evaluated the usefulness of environmental DNA as a tool to capture total parasite diversity in lake ecosystems from a few water samples.

Our platyhelminth primers successfully amplified platyhelminth DNA from samples collected from Tomahawk Lagoon. In contrast, the nematode primers amplified parasite DNA (nematode or platyhelminth) from 4/9 and 2/12 samples from Tomahawk Lagoon and Lake Waihola, respectively. The platyhelminth primers were more specific for the amplification of platyhelminth DNA (9/24 of amplified ASVs for non-target organisms), whereas the nematode primers appeared to be far less specific (likely due to the higher level of degeneracy in the primer design) and amplified a wide range of organisms (299/306 ASVs for non-target organisms; Supplementary Material). Using both sets of primers, many ASVs aligned with the same sequences on GenBank despite none of the ASVs being identical. This result may reflect limited taxonomic representation in the GenBank database: some phyla found at the study site do not have any COI sequences available on NCBI or BOLD (Table 6). Furthermore, many of the species identified in the lakes using the traditional methodology have not been previously sequenced and so they are absent from the GenBank database. Alternatively, the primers may have amplified similar but distinct mitochondrial haplotypes – either within the same species or divided across closely related taxa. A necessary step to fully utilize eDNA methodologies for routine parasite biodiversity assessment is to first build a reference database for every parasite of interest. Therefore, at this stage, this means going back to traditional parasitology methods to obtain, identify and then sequence each parasite, before eDNA can become a routine monitoring tool.

The nematode primers yielded fewer reads for the successfully aligned ASVs, compared to the platyhelminth primers (Tables 4 and 5). This pattern was probably affected by the amplification of many more ASVs under the less specific nematode primers, which would have led to competition for resources during PCR amplification. Overall, to assess parasite diversity using an eDNA methodology, we recommend designing primers that have a higher level of specificity towards more recently derived clades, rather than less specific primers that may universally amplify a wide range of non-target organisms.

Currently, we were unable to match the ASVs obtained here to the expected species (identified using traditional methods), but this is likely a problem that could be resolved with a complete reference database. However, four ASVs from a Lake Waihola sample aligned with sequences of *Contracaecum ogmorhini*, belonging to the family Anisakidae (amplified using the nematode primers; Table 5). These ASVs may represent a nematode species recorded in the taxonomic survey of the lakes, which was identified to the family level within Anisakidae and found within yellow-eye mullet *Aldrichetta forsteri* (Supplementary Material). Alternatively, these ASVs could actually represent *Contracaecum ogmorhi* or *O. mirounga*, as these species have been recorded as parasites for New Zealand pinnipeds (Lehnert *et al.*, 2019; Shamsi, 2019)

and New Zealand fur seals *Arctocephalus forsteri* are known to occasionally enter Lake Waihola.

An ASV amplified from a negative control sample from Tomahawk Lagoon aligned to a *Parorchis* sp. sequence also present in the environmental sample results (Tables 4 and 5). It is therefore possible that three ASVs amplified from the environmental samples represent contamination. However, none of these ASVs were identical and therefore the shared top blastn hit may reflect limited taxonomic diversity in the GenBank database.

The platyhelminth primers were only able to amplify DNA from Tomahawk Lagoon, and this is likely due to the much greater platyhelminth biomass found in this lake through traditional sampling methods (Table 1). In general, the number of ASVs was greater for platyhelminths than nematodes, and this may be due to differing life histories. Trematodes have a free-swimming life stage, and it is, therefore, possible that the concentration of trematode DNA in the water was higher than nematode DNA, because nematodes are contained within the host organism and they do not have a free-swimming stage. Indeed, even cestode eDNA was captured at a similar frequency as nematodes, and both of these parasites are internal, with no free-swimming stage. Nevertheless, our results highlight that eDNA can be useful for capturing both strictly internal parasites, as well as parasites that have free-swimming stages.

In conclusion, while we were able to detect parasite DNA from a range of parasite taxa, there was a large mismatch between the parasites identified using the eDNA methodology and those found using traditional methodologies. Importantly, we used negative controls at all stages of the study, which is a step that is often recommended but not always implemented in eDNA studies. Successful amplification in negative control samples is quite common in eDNA studies, especially at the field sampling stage, and the majority of eDNA metabarcoding studies use amplification in negative control samples to determine the level of background noise (Sepulveda *et al.*, 2020). Here, we have used amplification in the negative controls to identify false-positive results that were likely the product of sample contamination. Using eDNA methodology did not reveal any of the parasite families we expected to recover, this is most likely due to the lack of genetic data available for New Zealand parasites in general. A stumbling block of the present study is the lack of genetic information for New Zealand parasites sequences available in genetic databases. If eDNA is to be used as a monitoring tool, we recommend the use of custom metabarcodes designed with target taxa in mind and most importantly, we recommend building a custom genetic database comprised of the organisms of interest. As suggested by Collins *et al.* (2019) such a genetic database would be constructed using whole mitogenomes, which would open up the possibility of using different gene regions – in turn improving specificity and species identification and possibly reducing the problems associated with non-target amplification.

Supplementary material. The supplementary material for this article can be found at: https://doi.org/10.1017/S0031182021001840

Data. Jonined reads are available at: https://www.researchgate.net/publication/357776533_joined_reads

Acknowledgements. Eirik Haugstvedt Henriksen assisted with field work. Tania King and Bruce Roberts provided advice with laboratory procedures. We thank the two anonymous reviewers for their careful and detailed suggestions on earlier drafts have greatly improved this manuscript.

Author contributions. Conceptualization: RP and LJT. Methodology: LJT and MM. Formal analysis and Investigation: LJT, MM, FV. Writing: All authors. Funding acquisition: RP.

Financial support. This research was funded by a University of Otago Research Grant to R. Poulin.

Conflict of interest. The authors declare there are no conflicts of interest.

Ethical standards. Not applicable.

References

Altizer S, Ostfeld RS, Johnson PTJ, Kutz S and Harvell CD (2013) Climate change and infectious diseases: from evidence to a predictive framework. *Science (New York, N.Y.)* **341**, 514–519.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Bastos Gomes G, Hutson KS, Domingos JA, Chung C, Hayward S, Miller TL and Jerry DR (2017) Use of environmental DNA (eDNA) and water quality data to predict protozoan parasites outbreaks in fish farms. *Aquaculture* **479**, 467–473.

Beveridge I and Gasser RB (2014) Diversity in parasitic helminths of Australasian marsupials and monotremes: a molecular perspective. *International Journal for Parasitology* **44**, 859–864.

Collins RA, Bakker J, Wangensteen OS, Soto AZ, Corrigan L, Sims DW, Genner MJ and Mariani S (2019) Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution* **10**, 1985–2001.

Cox MP, Peterson DA and Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485.

Dario MA, Moratelli R, Schwabl P, Jansen AM and Llewellyn MS (2017) Small subunit ribosomal metabarcoding reveals extraordinary trypanosomatid diversity in Brazilian bats. *PLOS Neglected Tropical Diseases* **11**, e0005790.

Elbrecht V and Leese F (2017) PrimerMiner: an R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution* **8**, 622–626.

Gogarten JF, Calvignac-Spencer S, Nunn CL, Ulrich M, Saiepour N, Nielsen HV, Deschner T, Fichtel C, Kappeler PM, Knauf S, Müller-Klein N, Ostner J, Robbins MM, Sangmaneedet S, Schülke O, Surbeck M, Wittig RM, Sliwa A, Strube C, Leendertz FH, Roos C and Noll A (2020) Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Molecular Ecology Resources* **20**, 204–215.

Harper LR, Handley LL, Hahn C, Boonham N, Rees HC, Gough KC, Lewis E, Adams IP, Brotherton P, Phillips S and Hänfling B (2018) Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (Triturus cristatus). *Ecology and Evolution* **8**, 6330–6341.

Harper LR, Buxton AS, Rees HC, Bruce K, Brys R, Halfmaerten D, Read DS, Watson HV, Sayer CD, Jones EP, Priestley V, Mächler E, Múrria C, Garcés-Pastor S, Medupin C, Burgess K, Benson G, Boonham N, Griffiths RA, Lawson Handley L and Hänfling B (2019) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia* **826**, 25–41.

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

Jeunen G-J, Knapp M, Spencer HG, Taylor HR, Lamare MD, Stat M, Bunce M and Gemmell NJ (2019) Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecology and Evolution* **9**, 1323–1335.

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

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK and Ding L (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics (Oxford, England)* **25**, 2283–2285.

Lagrue C and Poulin R (2016) The scaling of parasite biomass with host biomass in lake ecosystems: are parasites limited by host resources? *Ecography* **39**, 507–514.

Lehnert K, Poulin R and Presswell B (2019) Checklist of marine mammal parasites in New Zealand and Australian waters. *Journal of Helminthology* **93**, 649–676.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079.

Mitsi K, Arroyo AS and Ruiz-Trillo I (2019) A global metabarcoding analysis expands molecular diversity of Platyhelminthes and reveals novel early-branching clades. *Biology Letters* **15**, 20190182.

Norris L, Lawler N, Hunkapiller A, Mulrooney DM, Kent ML and Sanders JL (2020) Detection of the parasitic nematode, *Pseudocapillaria tomentosa*, in zebrafish tissues and environmental DNA in research aquaria. *Journal of Fish Diseases* **43**, 1087–1095.

Poulin R and Presswell B (2016) Taxonomic quality of species descriptions varies over time and with the number of authors, but unevenly among parasitic taxa. *Systematic Biology* **65**, 1107–1116.

Rohland N and Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22**, 939–946.

Sepulveda AJ, Hutchins PR, Forstchen M, Mckeefry MN and Swigris AM (2020) The elephant in the lab (and field): contamination in aquatic environmental DNA studies. *Frontiers in Ecology and Evolution* **8**, 440.

Shamsi S (2019) Parasite loss or parasite gain? Story of Contracaecum nematodes in antipodean waters. *Parasite Epidemiology and Control* **4**, e00087.

Wingett SW and Andrews S (2018) FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research* **7**. doi: 10.12688/f1000research.15931.2

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